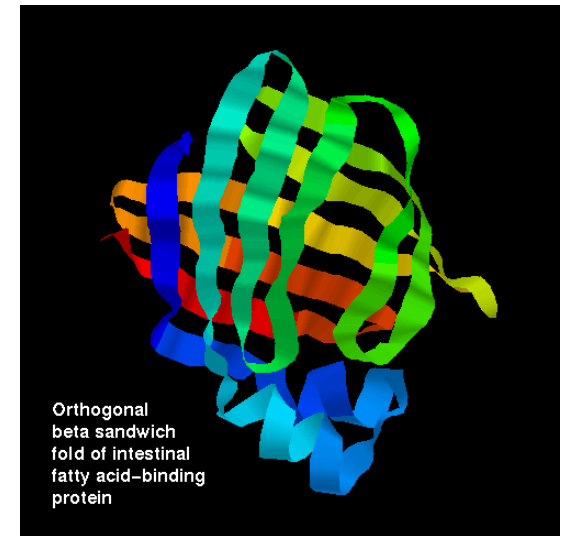
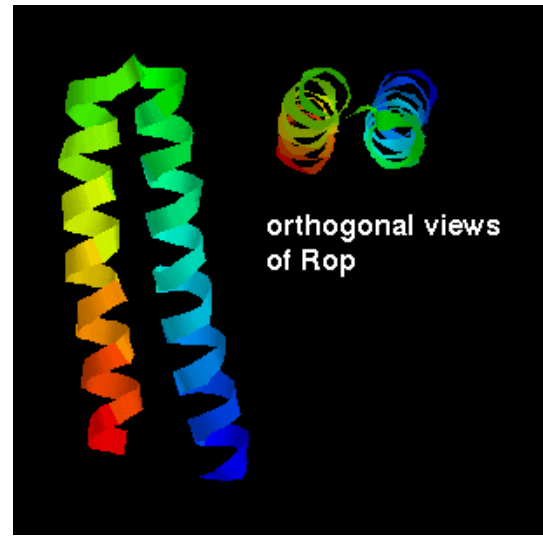


3D Structure

Visualizing, Comparing, Classifying



David Wishart
Athabasca 3-41
david.wishart@ualberta.ca

Outline & Objectives*

- **Visualization Programs**
- **Vectors & Matrices**
- **Difference Distance Matrices**
- **Molecular Superposition**
- **Measuring Superposition**
- **Classifying 3D Structures**

PDB Viewers

The image displays a screenshot of the RCSB PDB Jmol Viewer interface. The browser address bar shows the URL: <http://www.rcsb.org/pdb/explore/jmol.do?structureId=2TRX&bionumber=1>. The page title is "RCSB PDB : Jmol Viewer".

The main content area features the title: **CRYSTAL STRUCTURE OF THIOREDOXIN FROM ESCHERICHIA COLI AT 1.68 ANGSTROMS RESOLUTION** with the PDB ID **2trx**. Below the title, it indicates "Jmol Version 11.6". The central visualization shows a ribbon representation of the protein structure, colored in shades of blue, green, and yellow, with a ball-and-stick model of a ligand (likely a disulfide bond) shown in red and white.

The left sidebar contains navigation and utility links:

- Home**: News & Publications, Policies, FAQ, Contact, Feedback, About Us
- Deposition**: All Deposit Services, Electron Microscopy, NMR, Validation Server, BioSync Beamline, Related Tools
- Search**: Advanced Search, Latest Release, Latest Publications, Sequence Search, Ligand Search, Unreleased Entries, Browse Database, Histograms
- Explorer**: Last Structure: 2TRX
- Tools**: File Downloads, File Formats, Services: RESTful | SOAP, Widgets, Compare Structures
- Education**: Looking at Structures, Molecule of the Month, Educational Resources

At the bottom left, there are options to "View in Jmol", "SimpleViewer", "Protein Workshop", and "Other Viewers". A status message at the bottom center reads "Jmol script terminated".

Jmol*

- **Java-based program**
- **Open source applet and application**
 - **Compatible with Linux, MacOS, Windows**
- **Menus access by clicking on Jmol icon on lower right corner of applet**
- **Works with all major web browsers**
 - **Internet Explorer (Win32)**
 - **Mozilla/Firefox (Win32, OSX, *nix)**
 - **Safari (Mac OS X) and Opera 7.5.4**

WebMol*

The screenshot displays the WebMol interface for a distance matrix plot of the protein 2trx. The main window is titled "WebMol - Distance Matrix Plot: 2trx".

Distance Matrix Plot: A heatmap showing pairwise distances between residues. The x and y axes are labeled with residue numbers: 1(1), 216(108), and 27A<. A color scale at the bottom indicates distances from 0Å (red) to 27Å (blue). A red diagonal line represents the self-distances. A note below the plot states: "Move or drag the mouse pointer inside the map and view the structure! |= connectivity gap".

3D Molecular Structure: A 3D ribbon representation of the protein structure, colored by distance from the selected residue. The structure is shown in a dark background with various colored lines representing different distances. Labels like "16", "10", "121(13)D", "4", "2", "6", "8", and "9" are visible on the structure.

Control Panel: Located on the right side, it includes dropdown menus for "MainCh", "Color", and "Surface". There are checkboxes for "Labels", "HetAt", and "HOH", and a "Stereo" button. Navigation buttons include "<", ">", "Rock", "Select", and "Focus". A "Measure" dropdown menu is also present, with options "DMOff", "Rama", and "Trace".

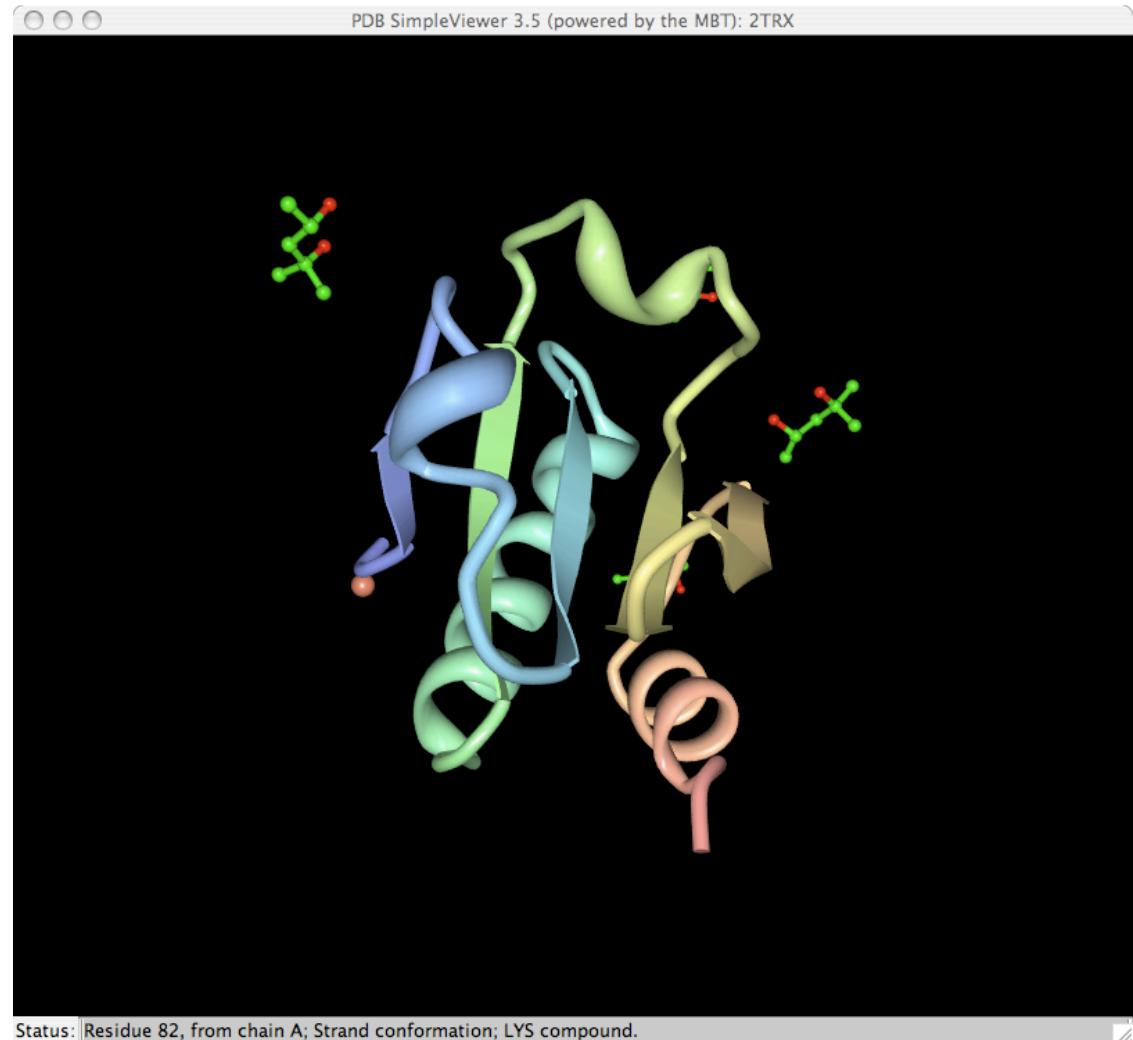
Bottom Panel: Contains a search bar with "2trx" and buttons for "Open", "Print", "C'nP", "ResetSlab", "Center", "Control", "Info", "Help", "[?]", and "[]". Below this are links for "Help interacting with WebMol" and "Simple Interaction Guide (requires flash)". The "2trx" logo and "© RCSB Protein Data Bank" are also visible.

Taskbar: The Windows taskbar at the bottom shows the Start button, a "Welcome to WebMol" window, a "Microsoft PowerP..." window, an "RCSB Protein Data Ba..." window, and the "WebMol - Distance M..." window. The system tray on the right shows the time as 5:05 PM.

WebMol*

- **Both a Java Applet and a downloadable application**
- **Offers many tools including distance, angle, dihedral angle measurements, detection of steric conflicts, interactive Ramachandran plot, diff. distance plot**
- **Compatible with most Java (1.3+) enabled browsers including:**
 - Internet Explorer
 - Safari on Mac OS
 - Mozilla 1.6/Firefox on Linux (Redhat 8.0)

PDB SimpleViewer



Requires Java WebStart (~30 sec install)

Chime*

Insulin: Structure of a Protein Hormone - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Location: http://c4.cabrillo.cc.ca.us/projects/insulin_tutorial/tutorial/index.html What's Related

The Insulin Tutorial Introduction

Reset Model

Insulin, the hormone that regulates blood glucose levels, is a small protein containing 51 amino acids.

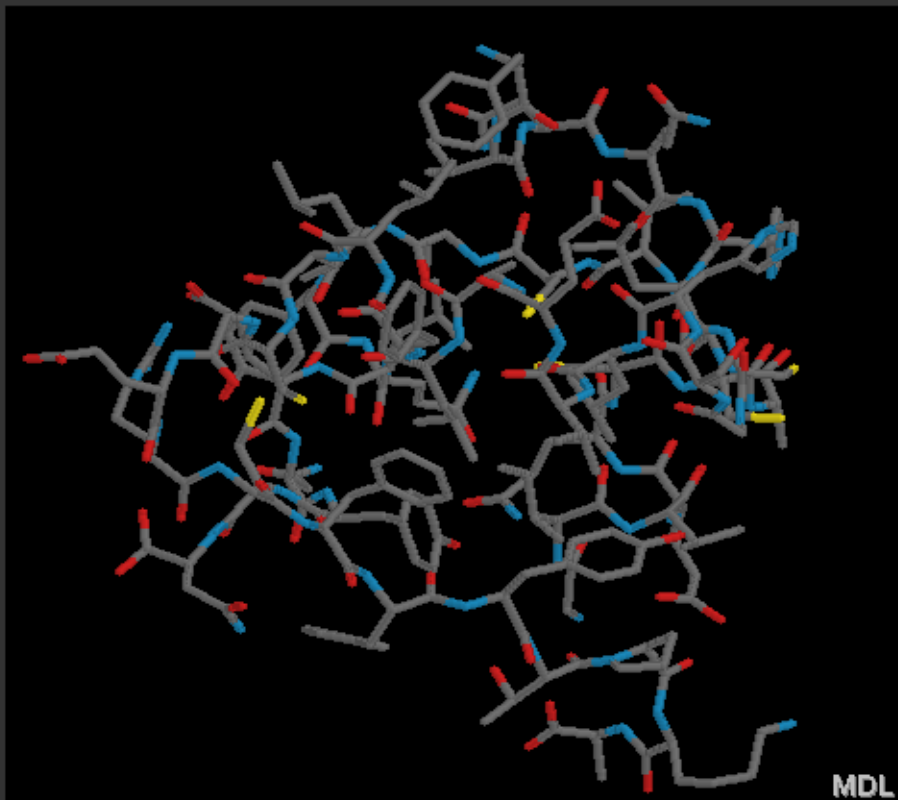
- ▶ See insulin's [primary structure](#).

A single insulin molecule consists of 2 polypeptide chains, **A (21 amino acids)** and **B (30 amino acids)**.

The structure of each chain can best be understood by simplifying the picture. In this view you see only the **backbone**, the repeating central structure of each polypeptide chain.

Protruding from the polypeptide backbone are the **side chains** which are different for each kind of amino acid.

- ▶ See the chemical [backbone structure](#) of a



insulin lighting PC Mac MDL **CHNOS**

Document: Done

Chime*

- http://www.umass.edu/microbio/chime/neccsoft.htm#download_install
- **Among first PDB viewing programs with limited manipulation capacity**
- **Uses Rasmol for its back end source**
- **View both large and small molecules**
- **Browser Plug-in (Like PDF reader)**
- **Interesting from historical perspective (now mostly phased out)**

Protein Explorer (Chime)

Explorer: One Molecule at a Time - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: <http://www.rcsb.org/pdb/pe/explorer/explore.htm?id=3TRX&> What's Related

Protein Explorer

About / © Click ? for Help.

1. Check out the [Tutorial ?](#) and [Tips for Effective Use.](#)
2. [Explore 3TRX](#)

Enter Commands Here

Number of Bridges ... 0
Number of Bonds 1644
Number of Atoms 1629
Number of Groups 105
Brookhaven Code 3TRX

MDL
1,629 atoms selected. Ready

Document: Done

Protein Explorer*

- http://www.umass.edu/microbio/chime/pe_beta/pe/protexpl/
- **Uses Chime or Jmol for its back-end**
- **Very flexible, user friendly, well documented, offers morphing, sequence structure interface, comparisons, context-dependent help, smart zooming, off-line**
- **Browser Plug-in (Like PDF reader)**
- **Compatible with Netscape (Mac & Win)**

QuickPDB

QuickPDB

Sequence: drag or click to select residues | 3D: double click to select residue

```
STRX: _      1  M V K Q I E S K T A F Q E A L D A A G D K L V V V D F S A T W C G P C K M I K P F F H S L S E K Y S M V I F L E V D V D D C Q D V A S E C E
STRX: _      71  V K C T P T P Q F F K K G Q K V G E F S G A N K E K L E A T I N E L V
```

Alpha (HGI) Beta (E)
Other

Polymers:

STRX: _

Secondary Structure ▾

Hydrophobic ▾

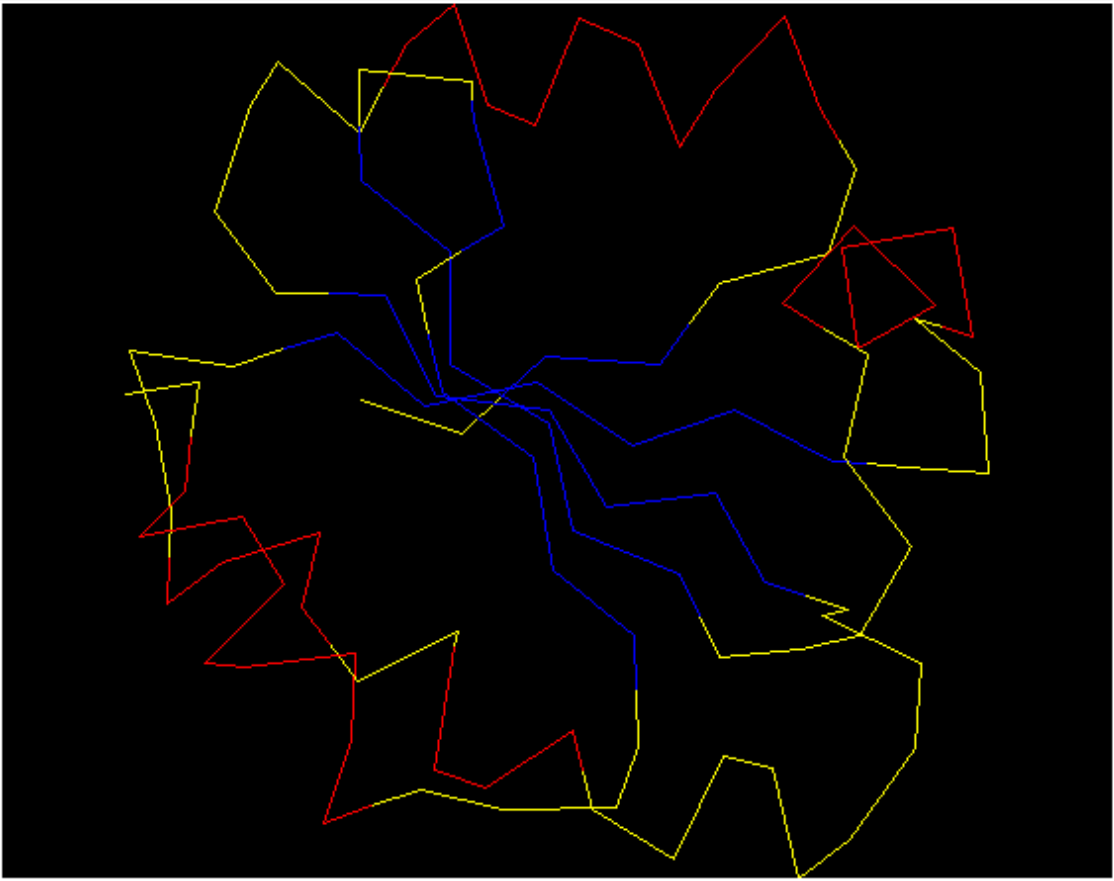
Stereo

Mouse: Rotate ▾

Color: Cyan ▾

Help

Reset Close



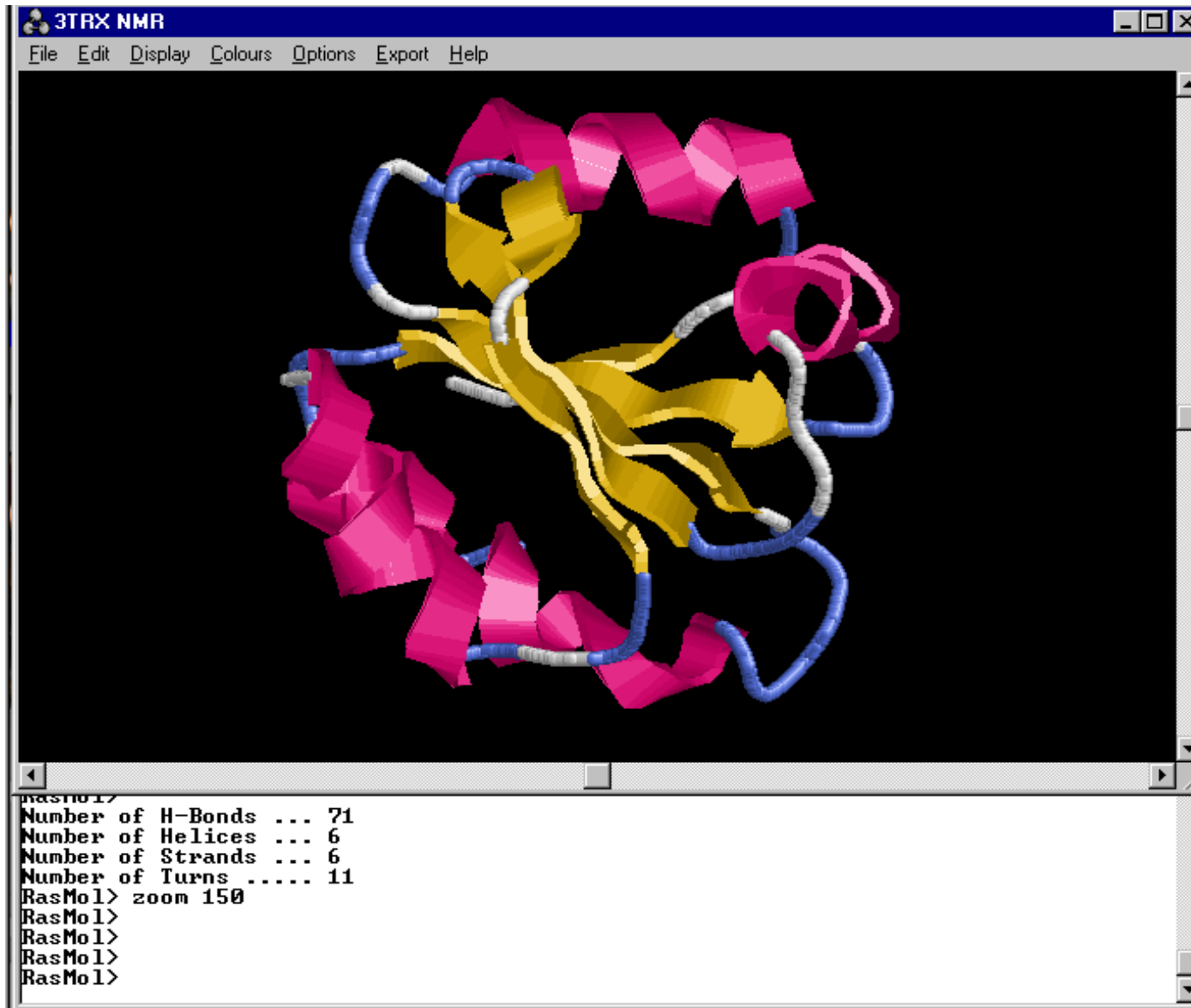
Applet QuickPDB v1.1 (C) 1996-1998 SDSC, by Ilya Shindyalov & Phil Bourne

Unsigned Java Applet Window

Quick PDB*

- <http://www.sdsc.edu/pb/Software.html>
- **Very simple viewing program with limited manipulation and very limited rendering capacity -- Very fast**
- **Java Applet (Source code available)**
- **Compatible with most browsers and computer platforms**

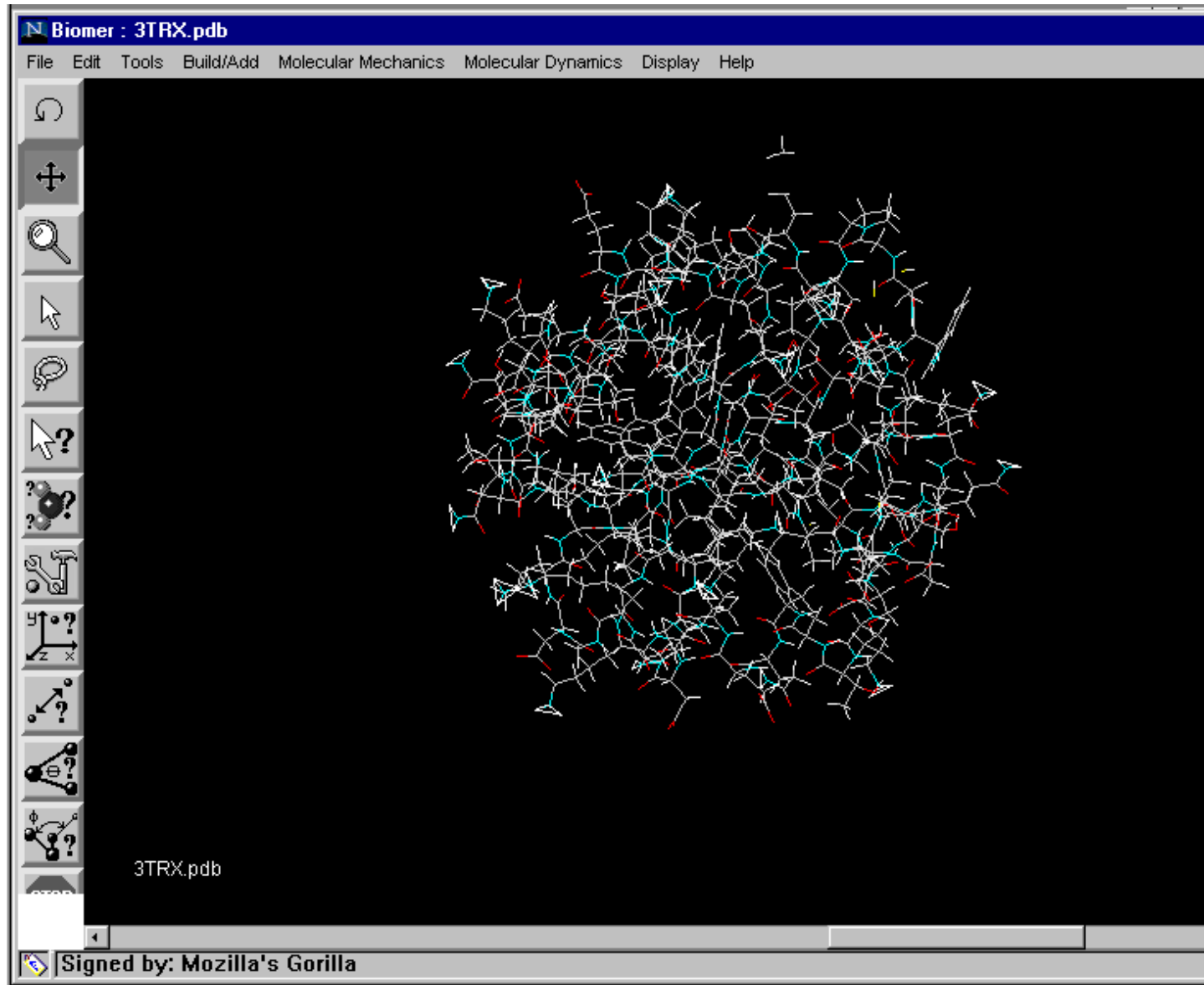
Rasmol



Rasmol*

- <http://www.umass.edu/microbio/rasmol/>
- **Very simple viewing program with limited manipulation capacity, easy to use!**
- **“Grand-daddy” of all visual freeware**
- **Runs as installed “stand-alone” program**
- **Source code available**
- **Runs on Mac, Windows, Linux, SGI and most other UNIX platforms**

B (Biomer)



Biomer (B)

- <http://casegroup.rutgers.edu/Biomer/index.html>
- **Very sophisticated molecular rendering and modelling package for both large and small molecules (kind of rough)**
- **Supports molecular dynamics & En. min**
- **Written in Java (source code available)**
- **Can run as an applet or stand-alone**
- **Compatible on most platforms**

Swiss PDB Viewer*

- <http://spdbv.vital-it.ch/>
- Among most sophisticated molecular rendering, manipulation and modelling packages (commercial or freeware)
- Supports threading, hom. modelling, energy minimization, seq/struc interface
- Stand-alone version only
- Compatible on Mac, Win, Linux, SGI

Swiss PDB Tutorial*

Deep View (Swiss-PdbViewer) Tutorial

http://spdbv.vital-it.ch/TheMolecularLevel/SPVTut/index.html

Department o...ell Biology Login- Depar... of Alberta Audiobaba Music Search Bioinformati... the U of A! Coilgun Basics 2 Pathguide: t...esource list

DeepView Tutorial

[What DeepView Can Do](#)

Gale Rhodes
[Contact Information](#)

Do you know about the [Grameen Foundation?](#)

Quick Links

- [Keyboard Command Modifiers](#)
- [Direct Import of Files](#)
- [User Guide](#)
- [User Manual](#) (pdf)

Contents

[Overview](#)

1. [Getting Started](#)
2. [Windows and Help](#)
3. [Manipulating the Model](#)
4. [Selecting and Displaying](#)
5. [Coloring](#)
6. [Measuring and Labeling](#)

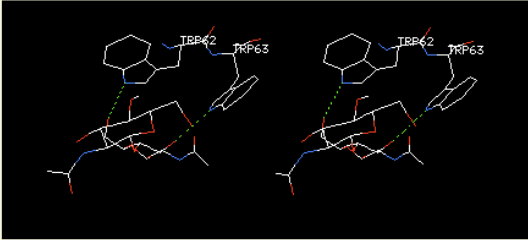
Assessing Your Command of DeepView: [Assignment 1](#)

7. [Mutating and Changing Side-Chain Conformations](#)
8. [Using a Ramachandran Plot](#)
9. [Judging the Quality of Models](#)

The Molecular Level

presents

MOLECULAR MODELING FOR BEGINNERS



Tutorial For DeepView - Swiss-PdbViewer

Revised 2008/07/07 For DeepView Version 4.0

USE MODELS WISELY: READ
[Crystallography Made Crystal Clear](#)
A Guide for Users of Macromolecular Models
Gale Rhodes, 3rd Edition, Academic Press, February 2006.

Thanks to my biochemistry students, Nicolas Guex, and online users of this tutorial for your many helpful suggestions. I invite and appreciate comments, corrections, and suggestions.

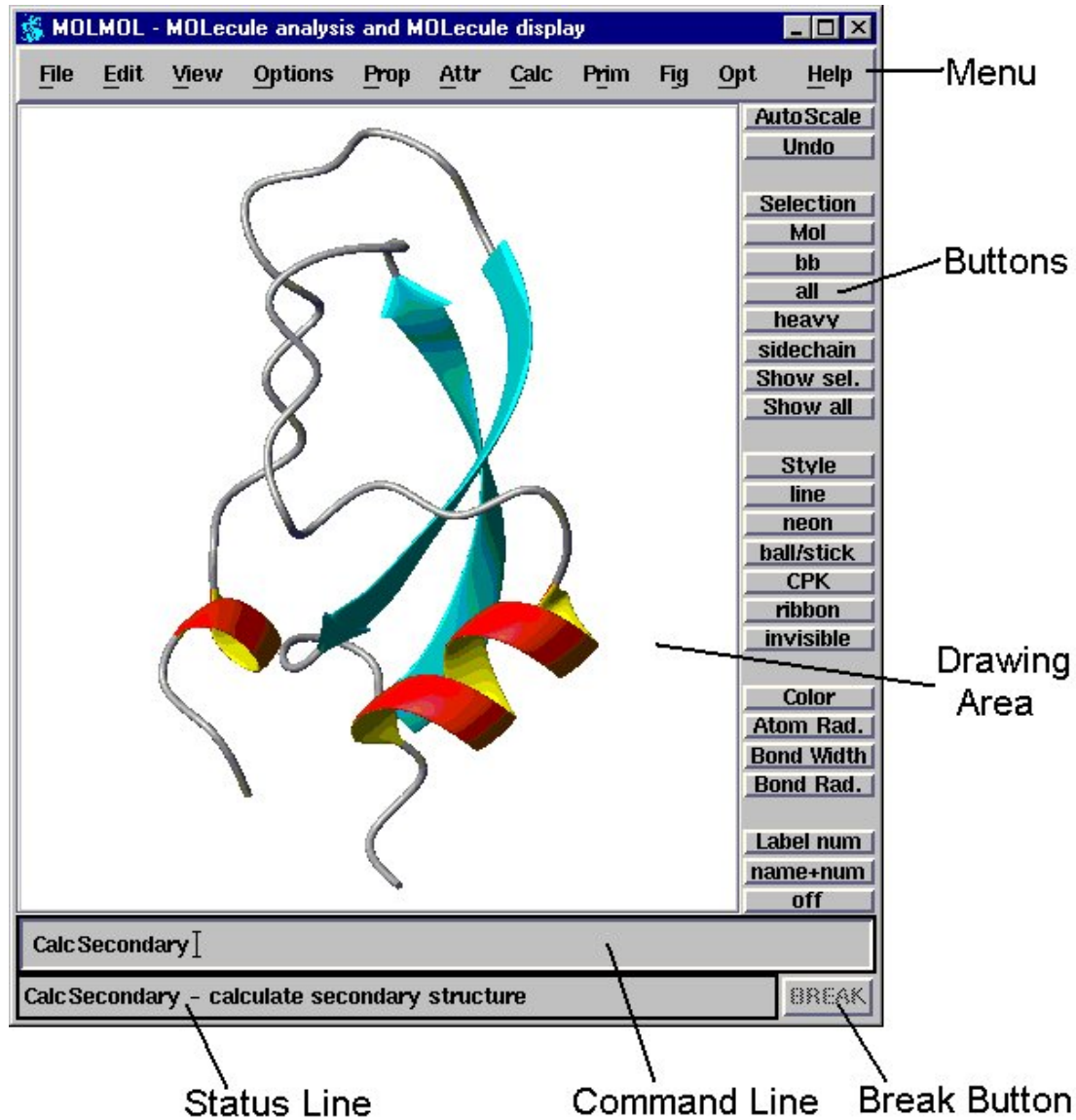
[Get DeepView](#) and [DeepView User Manual](#)

[Join the DeepView Discussion Group](#)

Get Thousands of Models at the [Protein Data Bank](#)

<http://spdbv.vital-it.ch/TheMolecularLevel/SPVTut/index.html>

MolMol



MolMol*

- <http://www.mol.biol.ethz.ch/wuthrich/software/molmol/>
- **Very sophisticated molecular rendering, and manipulation package (among the best graphics of all freeware)**
- **Special focus on NMR compatibility, supports many calculations/plots**
- **Stand-alone version only**
- **Compatible on Win, Unix (nearly all)**

Summary*

Mac Win Unix Rendr SeqView Super E Min Modeling

Rasmol	+	+	+	++	-	-	-	-
Chime	+	+	-	+	-	-	-	-
Prot. Expl.	+	+	-	++	+	+	-	-
Quick PDB	+	+	+	+	+	-	-	-
Biomer	+	+	+	++	-	+	+	+
SwP Viewer	+	+	+	+++	+	+	+	+
MolMol	-	+	+	+++	-	+	-	+

Visualization Hub

Top 5 MolVis Techs

Department o...ell Biology Login- Depar... of Alberta Audiobaba Music Search Bioinformati... the U of A! Coilgun Basics 2 Pathguide: t...esource list

http://www.umass.edu/microbio/chime/top5.htm

Google

MolviZ "Top 5"

The "Top 5" 3D Molecular Visualization Technologies for the rest of us...*

* ...who do not solve protein structures.

I. [Tutorials](#) • II. [Tutorial-Authoring Systems](#) • III. [Explore Any Molecule](#) • IV. [See Protein Evolution](#) • V. [Animate Molecules in Powerpoint](#)

top5.molviz.org
collected by [Eric Martz](#) for **MolviZ.Org**
Last updated: Nov-2008

NSF
Made with Jmol

Students

- [Tutorials](#) show you how the 3D structures of **DNA, hemoglobin, and many other** macromolecules support their functions.
- [Find and explore](#) the 3D structure of **any** macromolecule easily.
- [Identify](#) the parts of any 3D protein molecule that **mutate** most slowly, because they must be **conserved** in order to perform crucial functions.
- Easily make 3D macromolecules [rotate in your Powerpoint® slides](#).

Educators

- Want to **eliminate Chime** from your classes? The solution is here! See [below](#) ...
- **Project** the following resources in lectures, **offer** them to your students, or **assign** them.
- Some resources are targeted specifically to **high school** level; most are suitable for **college** and **graduate** levels.
- Ready-made [Tutorials](#) on 3D structures of **DNA, hemoglobin, lipid bilayers, water, and many other** macromolecules.
- Some tutorials include **challenge questions** (answers on request).
- [Customize interactive molecular views easily](#) to show in class, and for your students to rotate, zoom, and admire.
- [Find and explore](#) the 3D structure of **any** macromolecule easily.
- [Identify](#) the parts of any 3D protein molecule that **mutate** most slowly (because they must be **conserved** [see [Gallery](#)] in order to perform its functions) or most rapidly (to support function, e.g. escape of influenza hemagglutinin from immunity).
- Easily make 3D macromolecules [rotate in your Powerpoint® slides](#), and your students can do the same.

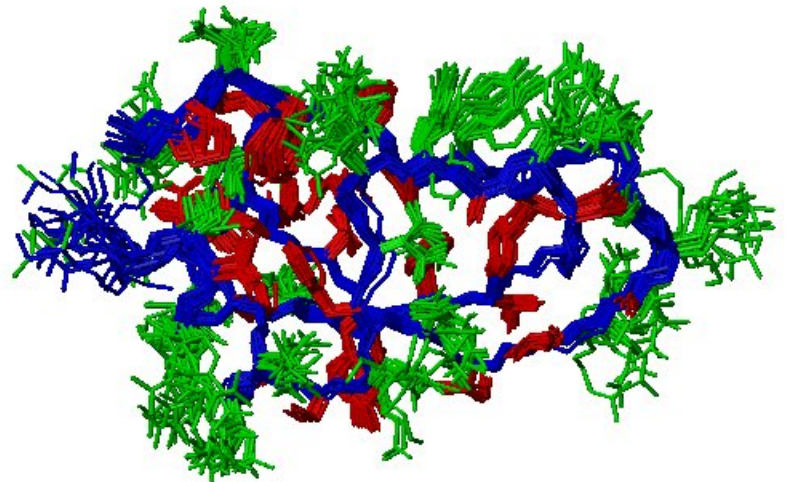
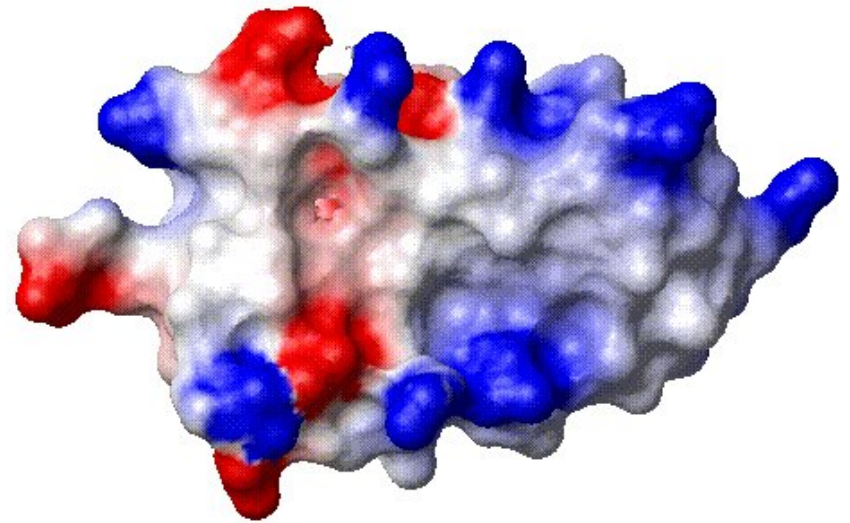
Researchers

- [Find and explore](#) the key structural features of **any** macromolecule, easily.
- [Locate](#) the positions of crucial residues.
- See [noncovalent bonds](#) between any moiety and the remainder of the structure.
- Annotate interactive macromolecular structures, without learning scripting languages, for **journal supplementary materials** or your **lab group's website** at [Proteopedia.Org](#), a structural biology wiki with scene-authoring tools.
- [Identify functional regions](#) by coloring amino acids with their levels of evolutionary conservation -- in minutes, totally automatically if you wish. Locate conserved residues, e.g. for **functional mutagenesis** studies.
- Make [publication-quality molecular figures](#) with ease, highlighting specific residues or regions, and customizing rendering and colors.
- Easily make 3D macromolecules [rotate/animate in your Powerpoint® slides](#).

<http://www.umass.edu/microbio/chime/top5.htm>

Graphics Formats

- GIF
- JPEG
- PNG
- TIFF (Tag Image)
- BMP
- EPS
- PS
- RGP (SGI)



Graphics Formats*

- **GIF (Graphical Interchange Format)**
 - pronounced “JIF”
 - introduced in 1987 by CompuServe
 - handles 8 bit colour (256 colours)
 - lossy compression (up to 10 X)
 - best for drawings, simple B+W or colour diagrams, images with hard edges
 - supported by Perl graphics library (GD.pm)
 - supports animation & transparency

Graphics Formats*

- **JPEG (Joint Photographic Experts Group)**
 - pronounced “JAY-peg”
 - exploits eye’s poor perception of small changes in colour variation
 - handles 24 bit colour (1.6 million colours)
 - allows adjustable lossy compression
 - best for colour pictures of real objects with varied colour, shadow, fuzzy edges
 - among most common web image formats

Graphics Formats*

- **PNG (Portable Network Graphics)**
 - designed to replace GIF and TIFF
 - supports lossless compression
 - supports 24 bit, grayscale and 8 bit
 - supports transparency & interlacing
 - offers better compression than GIF (15%)
 - supported by new GD.pm Perl library
 - problems with many early browsers in viewing PNG (now fixed)

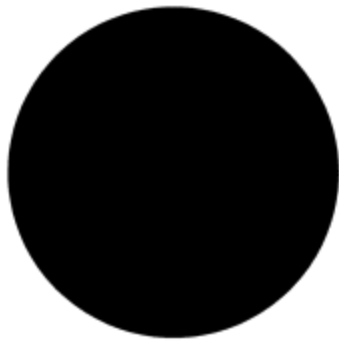
PovRay (www.povray.org)

The screenshot shows the homepage of the POV-Ray website, titled "POV-Ray - The Persistence of Vision Raytracer". The browser address bar shows "http://www.povray.org/". The page features a navigation menu with links for Download, Hall Of Fame, Docs, FAQ, Resources, Community, Support, Wiki, Lib, and Search. The main content is organized into several sections:

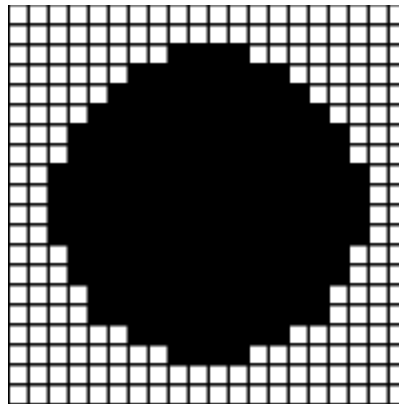
- Welcome:** A paragraph introducing the Persistence of Vision Raytracer as a high-quality, free tool for creating stunning three-dimensional graphics. It is available in official versions for Windows, Mac OS/Mac OS X, and i86 Linux. The source code is also available for those wanting to do their own ports.
- Download and Navigation:** A section providing navigation instructions, including links to the download page and a beta-test page for SMP and multi-core systems.
- Contacting Us:** A section providing contact information and links to the license page and privacy policy.
- What's New:** A section with three news items:
 - Bishop3D Distribution Release:** A Windows-based dedicated modeller for POV-Ray with SDL import, texture editor, and native keyframe animation, now reaching its first distribution release. It has been several years in the making and has a promising future with a new development cycle of enhancements and additions now beginning. Comments and discussion are encouraged on the Bishop3D forum. Download it now to give it a try! (July 19, 2009)
 - LionSnake 1.7.14:** LionSnake modeller has received a number of new features and updates including more advanced subdivision previews and face editing... (July 14, 2009)
 - Bishop3D reaches Release Candidate milestone:** (Text partially obscured)
- Hall of Fame:** A section featuring a rendered image of a highway scene titled "Autobahn WIP". Below the image is a link to "See more images... [dialup version]".
- POV-Ray Posters:** A section with two small rendered images.

The website has a blue and yellow color scheme and a vertical sidebar on the right with the text "THE PERSISTENCE OF VISION" and "RAYTRACER".

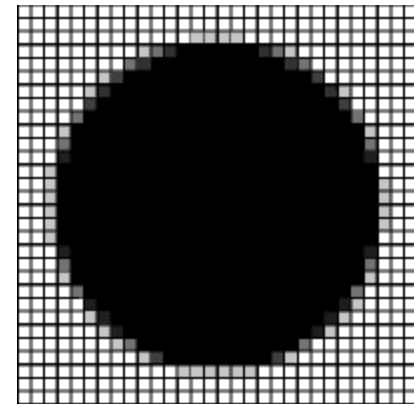
Aliasing & Antialiasing*



True Image



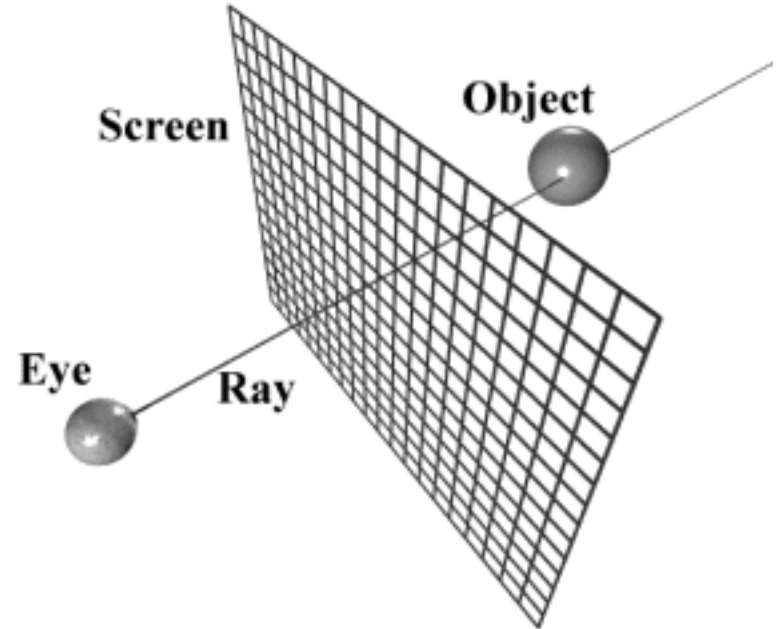
Aliased Image



Anti-aliased Image

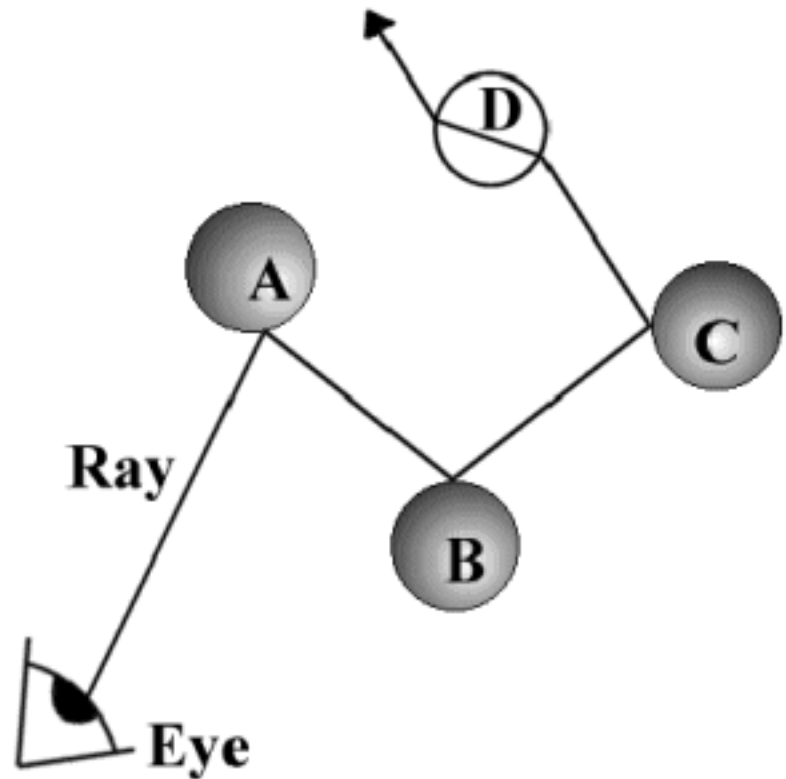
Ray Casting (from 3D to 2D)*

- Ray = beam of light
- For each pixel on screen, cast ray from eye thru pixel
- Test every object in scene to see if ray intersects object
- Each ray intersection nearest to eye is made visible, color pixel



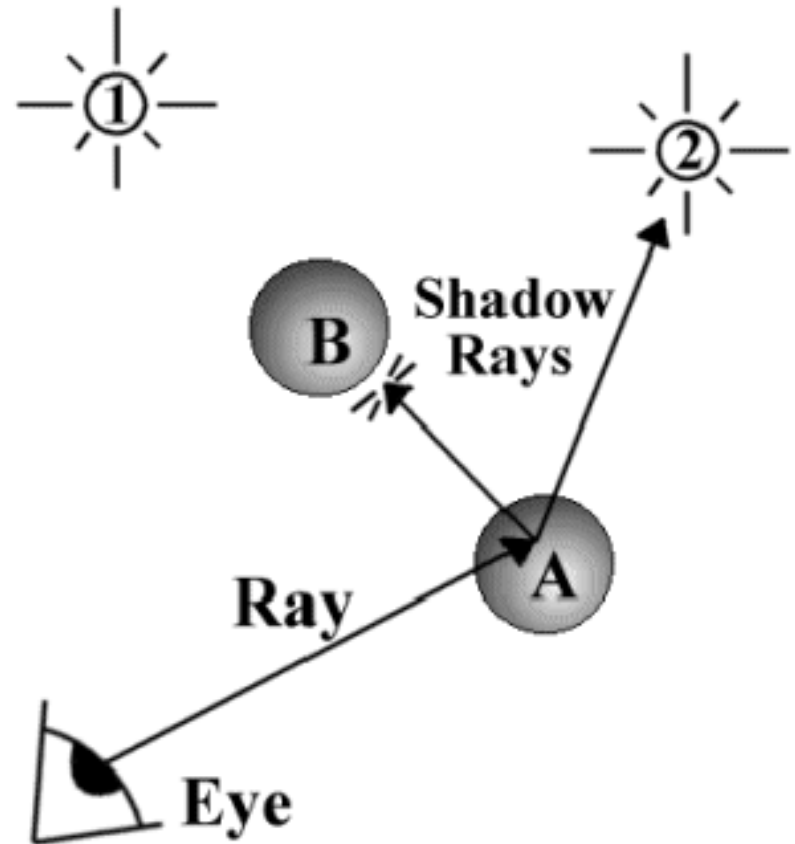
Ray Tracing & Reflection*

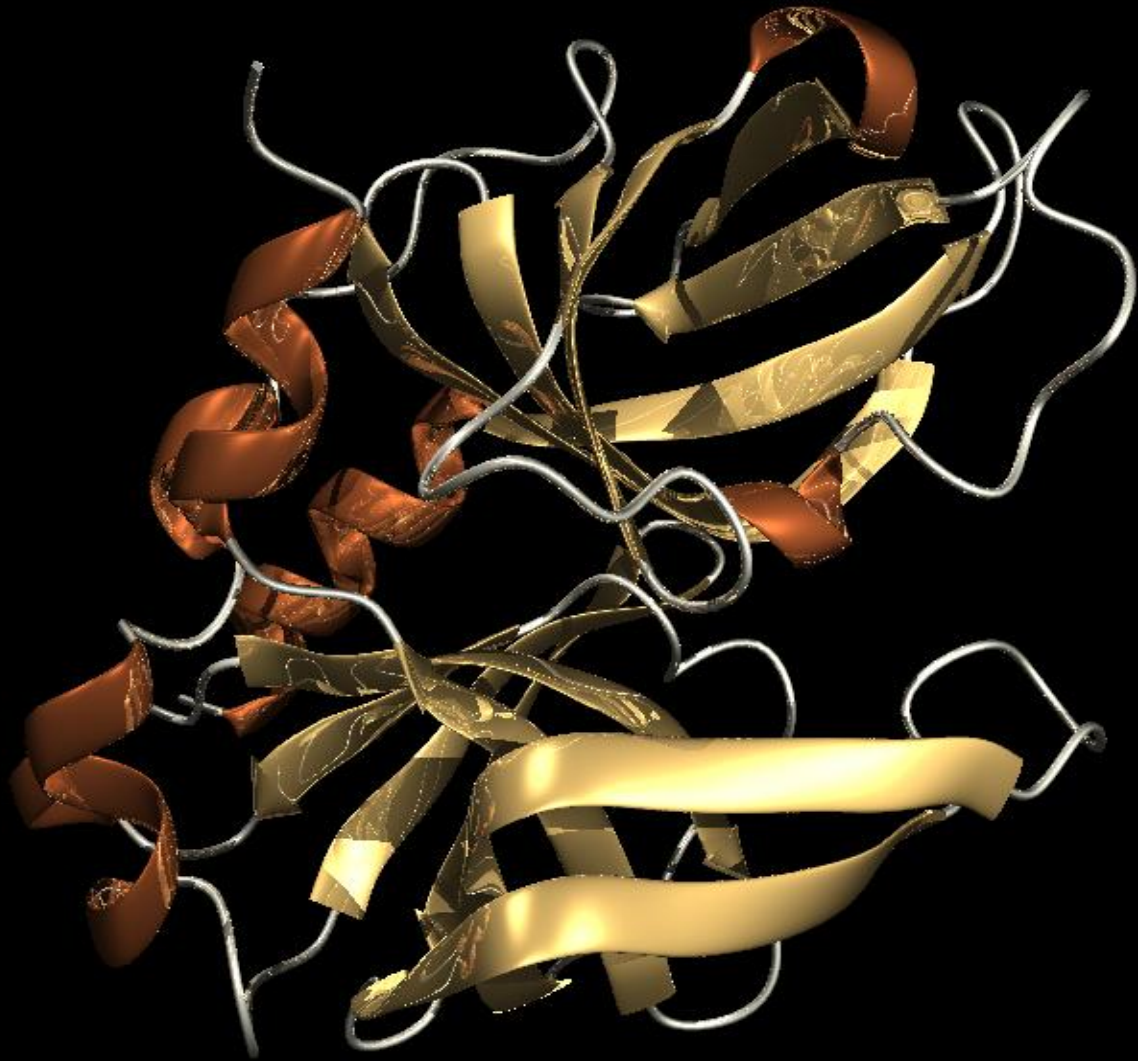
- Used to determine surface appearance
- Begins with ray casting, determine intersects, then recursively sends 2ndary rays to see which objects reflect, which are transparent, which absorb, etc.



Shadowing*

- Uses ray tracing algorithm
- Sends out 2ndary rays towards light sources to see if opaque objects are in the way, if so, then surface is in shadow
- “shadow feelers”



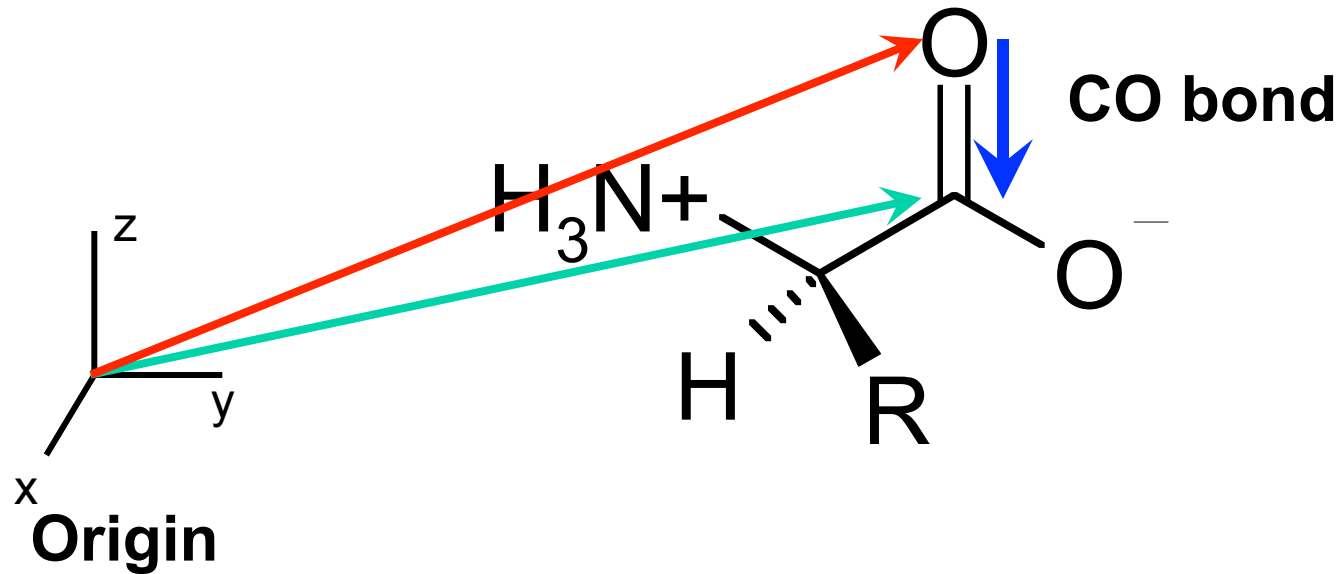


HAV-3C Protease - Alan Gibbs

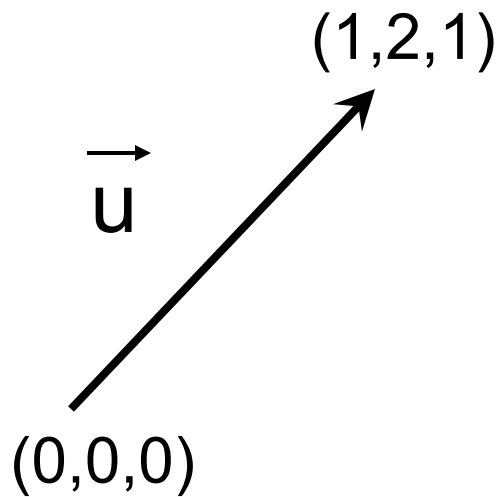
Outline

- **Visualization Programs**
- **Vectors & Matrices**
- **Difference Distance Matrices**
- **Molecular Superposition**
- **Measuring Superposition**
- **Classifying 3D Structures**

Vectors Define Bonds and Atomic Positions

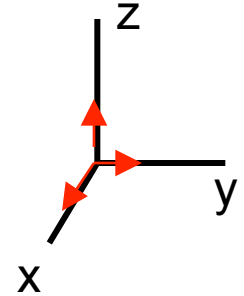


Review – Vectors*



$$\vec{u} = 1\hat{i} + 2\hat{j} + 1\hat{k}$$

$$\vec{u} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$



$$|\vec{u}| = \sqrt{(1-0)^2 + (2-0)^2 + (1-0)^2} = \sqrt{6}$$

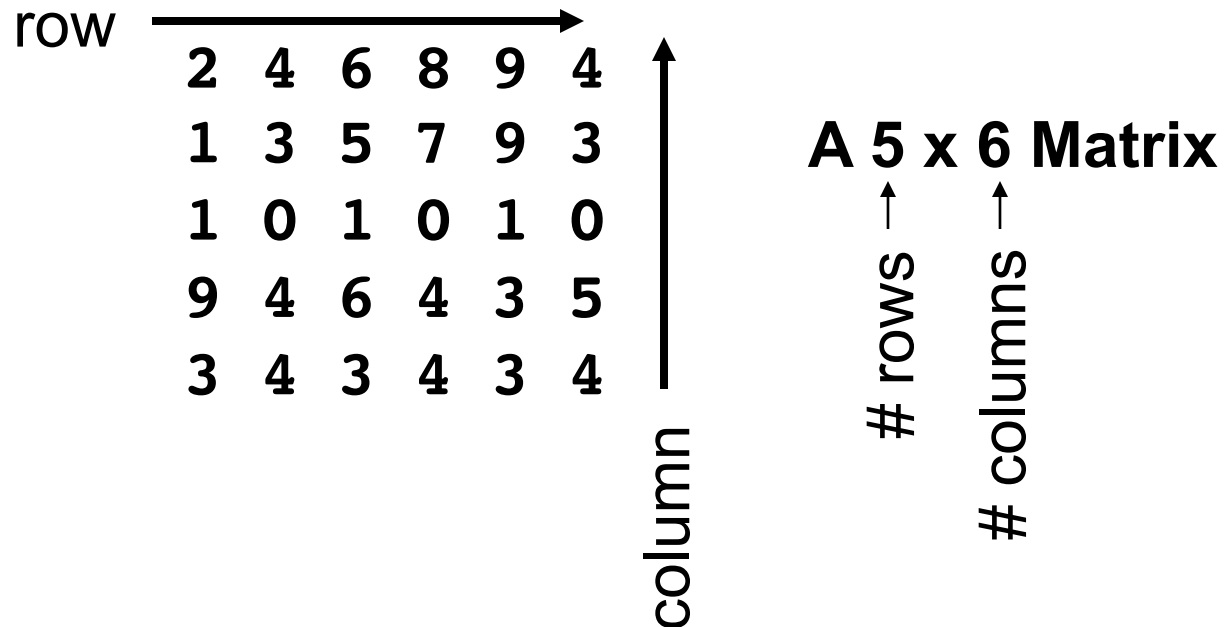
Vectors have a length & a direction

Review - Vectors

- **Vectors can be added together**
- **Vectors can be subtracted**
- **Vectors can be multiplied (dot or cross or by a matrix)**
- **Vectors can be transformed (resized)**
- **Vectors can be translated**
- **Vectors can be rotated**

Matrices*

- A matrix is a table or “array” of characters
- A matrix is also called a tensor of “rank 2”



Different Types of Matrices

$$\begin{bmatrix} 2 & 4 & 6 & 8 & 9 & 4 \\ 1 & 3 & 5 & 7 & 9 & 3 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 9 & 4 & 6 & 4 & 3 & 5 \\ 3 & 4 & 3 & 4 & 3 & 4 \\ 3 & 6 & 7 & 9 & 1 & 0 \end{bmatrix}$$

A square
Matrix

$$\begin{bmatrix} 2 & 4 & 6 & 8 & 9 & 4 \\ 4 & 3 & 5 & 7 & 9 & 3 \\ 6 & 5 & 1 & 0 & 1 & 0 \\ 8 & 7 & 0 & 4 & 3 & 5 \\ 9 & 9 & 1 & 3 & 3 & 4 \\ 4 & 3 & 0 & 5 & 4 & 0 \end{bmatrix}$$

A symmetric
Matrix

$$\begin{bmatrix} 1 \\ 3 \\ 5 \\ 9 \\ 7 \\ 3 \end{bmatrix}$$

A column
Matrix
(A vector)

Different Types of Matrices*

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{C} & \mathbf{D} & \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} & \mathbf{I} & \mathbf{J} & \mathbf{K} & \mathbf{L} \\ \mathbf{M} & \mathbf{N} & \mathbf{O} & \mathbf{P} & \mathbf{Q} & \mathbf{R} \\ \mathbf{S} & \mathbf{T} & \mathbf{U} & \mathbf{V} & \mathbf{W} & \mathbf{X} \end{bmatrix}$$

**A rectangular
Matrix**

$$\begin{bmatrix} \cos\theta & \sin\theta & 0 \\ \sin\theta & -\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

**A rotation
Matrix**

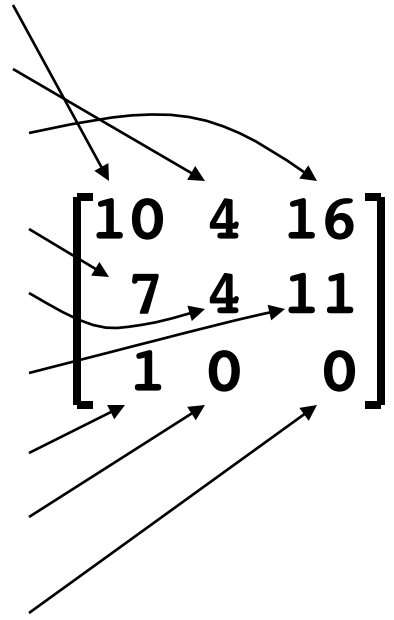
$$[2 \ 4 \ 6 \ 8 \ 9]$$

**A row
Matrix
(A vector)**

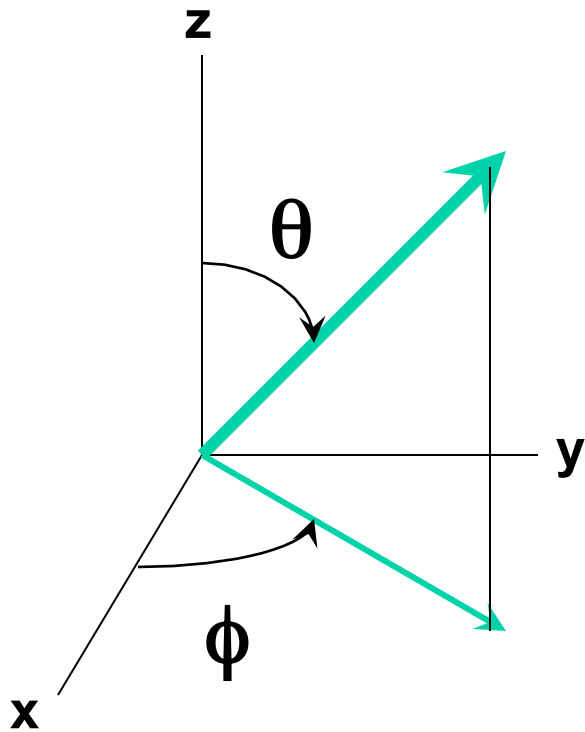
Review - Matrix Multiplication

$$\begin{bmatrix} 2 & 4 & 0 \\ 1 & 3 & 1 \\ 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 3 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\begin{aligned} &2x1 + 4x2 + 0x0 \\ &2x0 + 4x1 + 0x1 \\ &2x2 + 4x3 + 0x0 \\ &1x1 + 3x2 + 1x0 \\ &1x0 + 3x1 + 1x1 \\ &1x2 + 3x3 + 1x0 \\ &1x1 + 0x2 + 0x0 \\ &1x0 + 0x1 + 0x1 \\ &1x2 + 0x3 + 0x0 \end{aligned}$$



Rotation*



$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \quad \text{Rotate about x}$$

$$\begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{Rotate about z}$$

Rotation*

Counterclockwise about x Counterclockwise about z

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

$$\begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

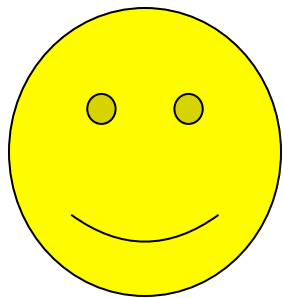
Clockwise about x

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}$$

Clockwise about z

$$\begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

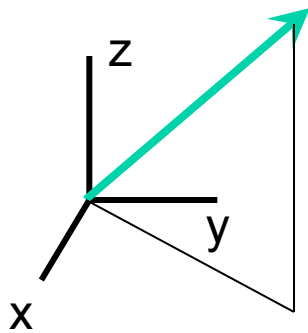
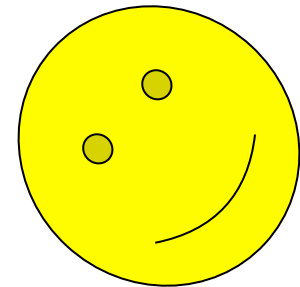
Rotation



X

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}$$

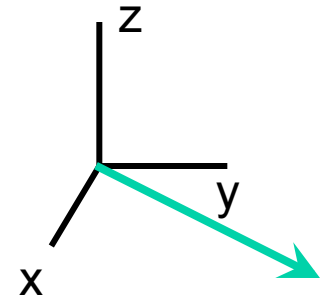
=



X

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}$$

=



Rotation (Detail)*

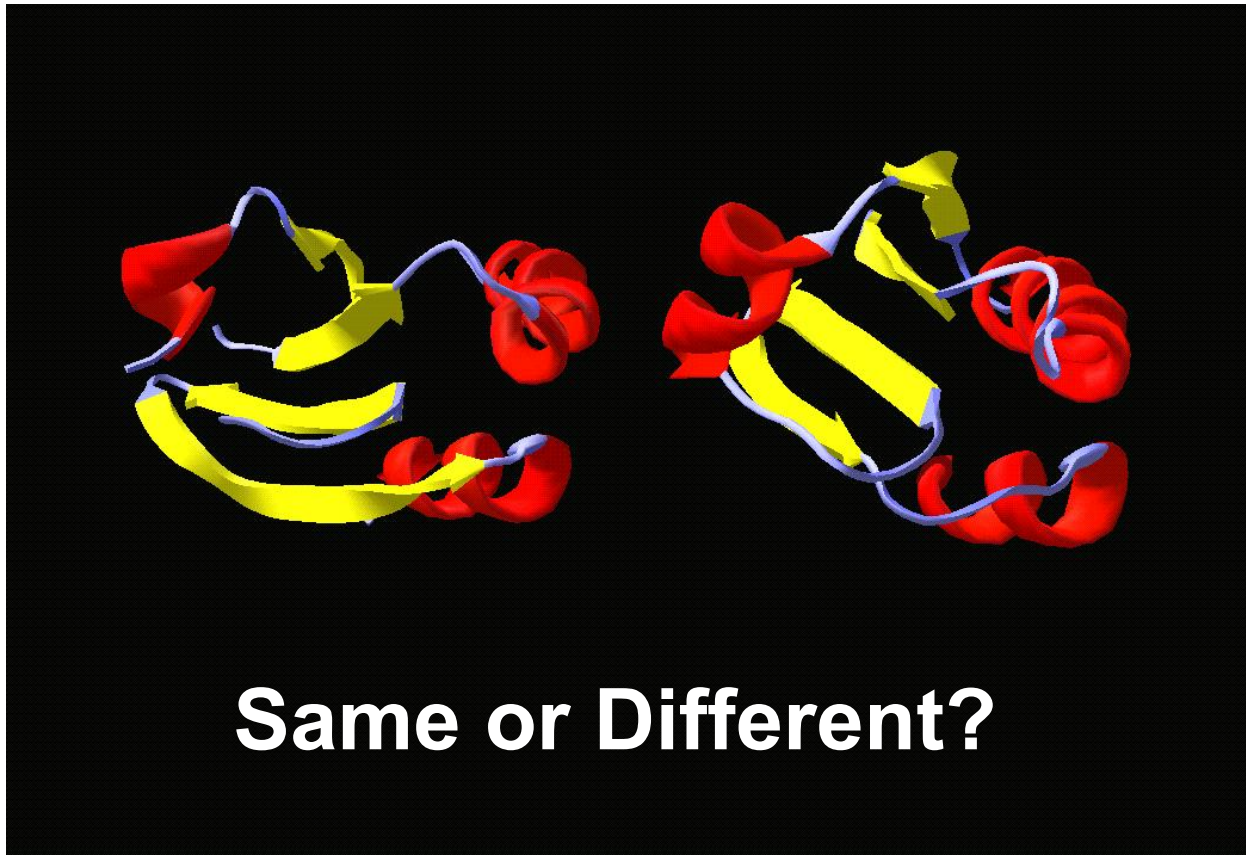
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \mathbf{X} \begin{img alt="Diagram of a 3D coordinate system with x, y, and z axes. A red arrow points from the origin into the 3D space, representing a vector in the original coordinate system." data-bbox="515 335 675 555"/> = \begin{img alt="Diagram of a 3D coordinate system with x, y, and z axes. A red arrow points from the origin along the y-axis, representing the rotated vector." data-bbox="730 345 890 545"/>$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ \cos\theta + \sin\theta \\ -\sin\theta + \cos\theta \end{bmatrix}$$

Comparing 3D Structures

- **Visual or qualitative comparison**
- **Difference Distance Matrices**
- **Superimposition or superposition**
- **Root mean square deviation (RMSD)**
- **Subgraph isomorphisms (Ullman's algorithm)**
- **Combinatorial extension (CE)**

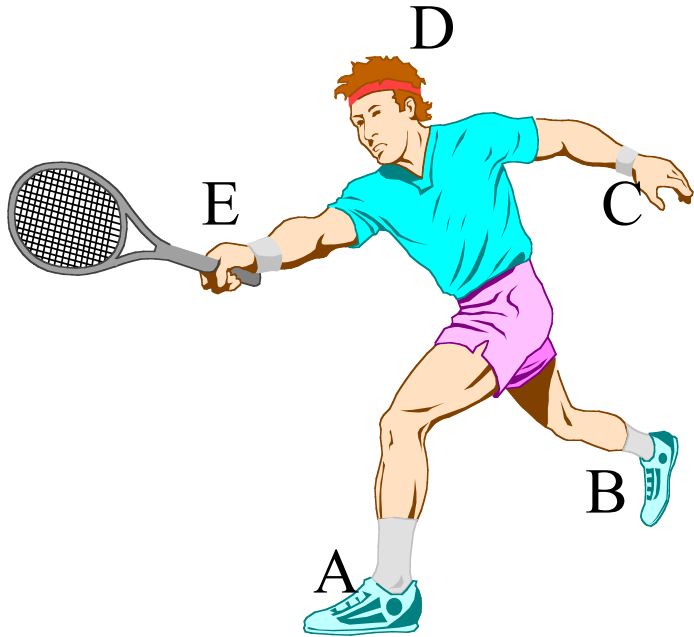
Qualitative Comparison



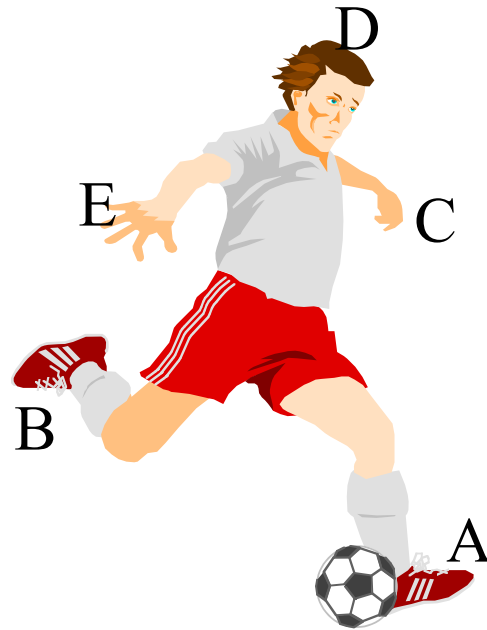
Outline

- **Visualization Programs**
- **Vectors & Matrices**
- **Difference Distance Matrices**
- **Molecular Superposition**
- **Measuring Superposition**
- **Classifying 3D Structures**

Difference Distance Matrix*



Object A



Object B

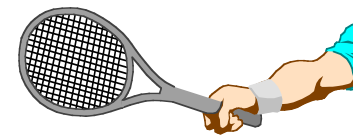
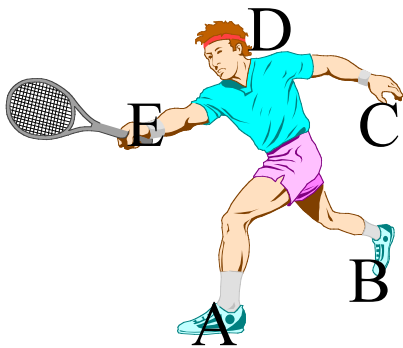
Difference Distance Matrix*

	A	B	C	D	E
A	0	4	5	6	4
B		0	4	6	7
C			0	3	6
D				0	3
E					0

	A	B	C	D	E
A	0	4	5	6	4
B		0	4	5	3
C			0	3	5
D				0	3
E					0

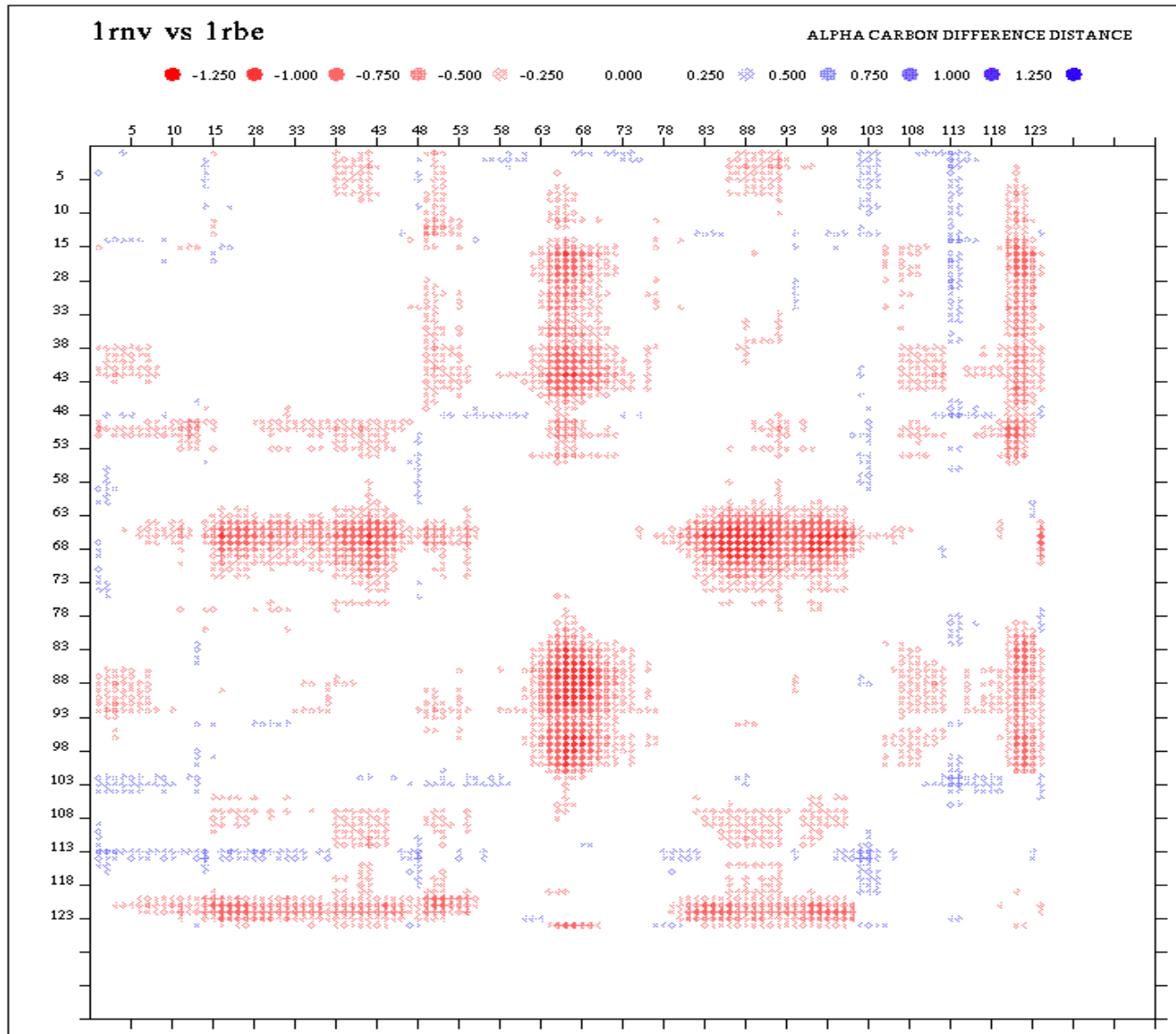
 =

	A	B	C	D	E
A	0	0	0	0	0
B		0	0	1	4
C			0	0	1
D				0	0
E					0



Hinge motion

Difference Distance Matrix



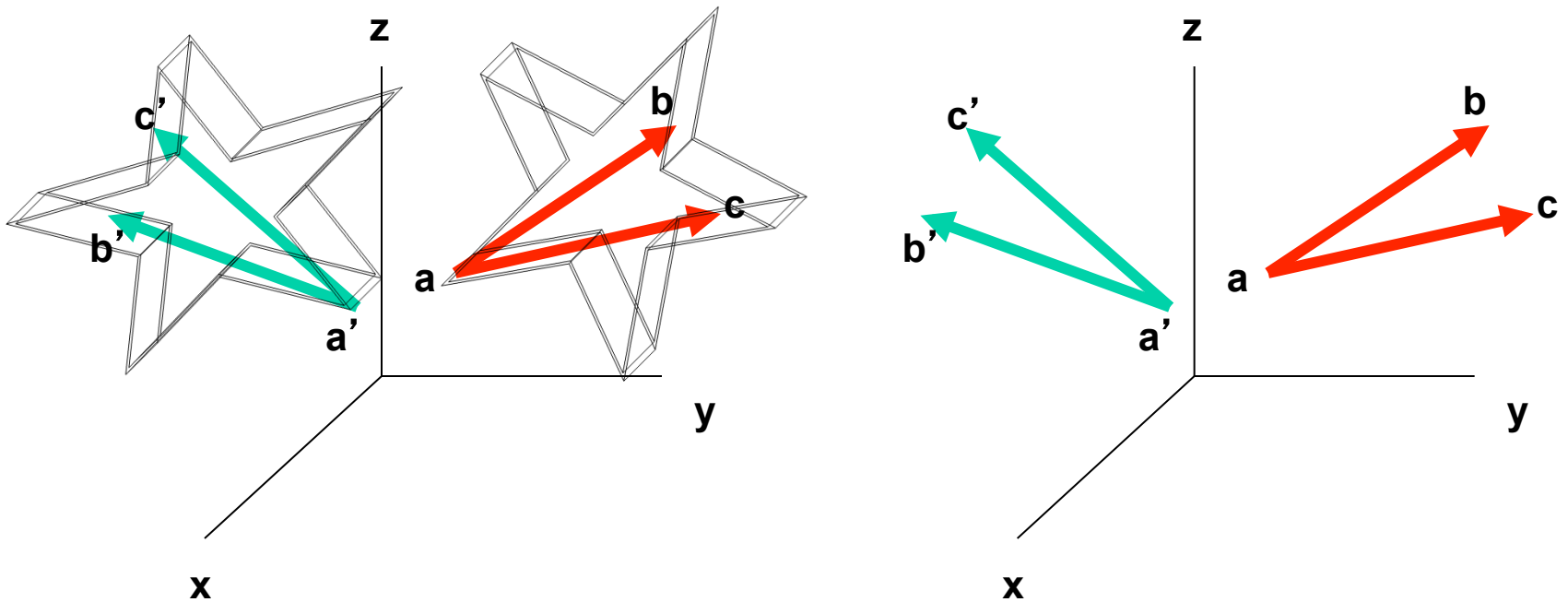
Difference Distance Matrices or DDM' s*

- **Simplest method to perform structural comparisons**
- **Requires no transformations, no rotations or superpositions**
- **Very effective at identifying “hinge” motions or localized changes**
- **Produces a visually pleasing, quantitative measure of similarity**

Superposition*

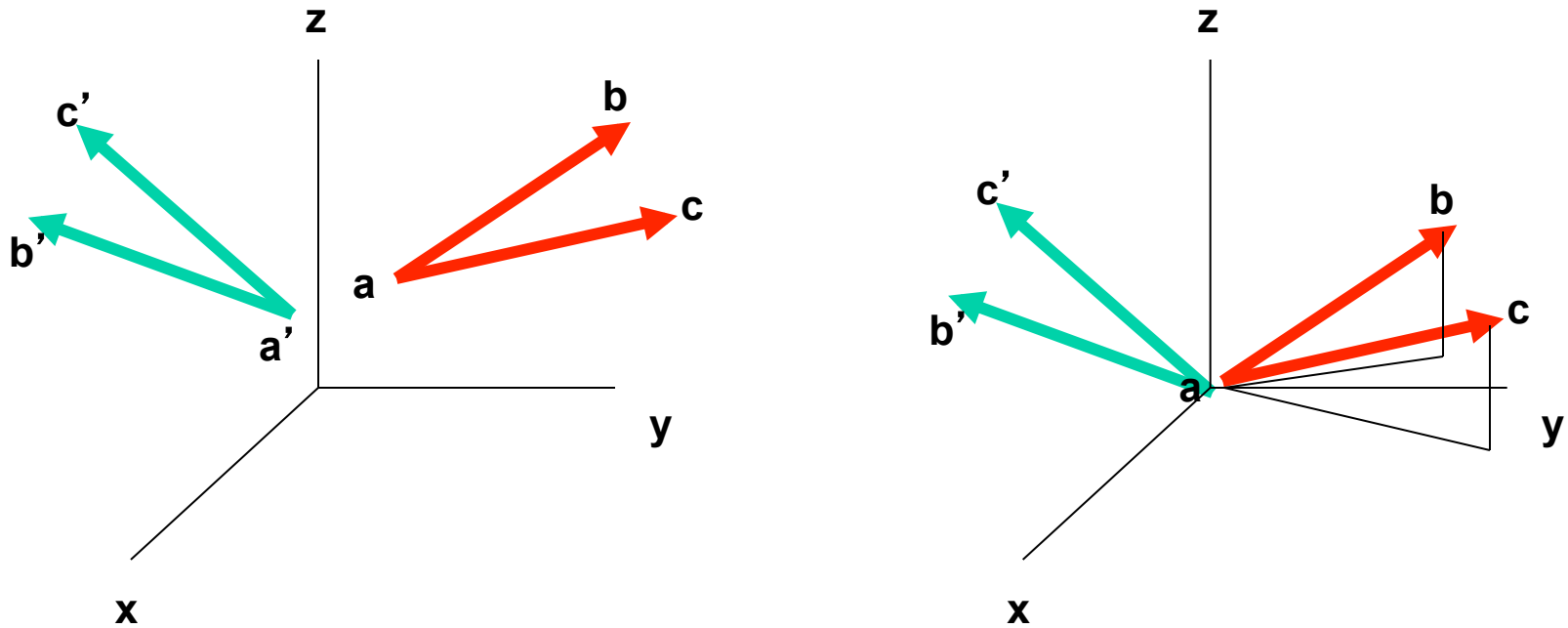
- **Objective is to match or overlay 2 or more similar objects**
- **Requires use of translation and rotation operators (matrices/vectors)**
- **Recall that every three dimensional object can be represented by a plane defined by 3 points**

Superposition*



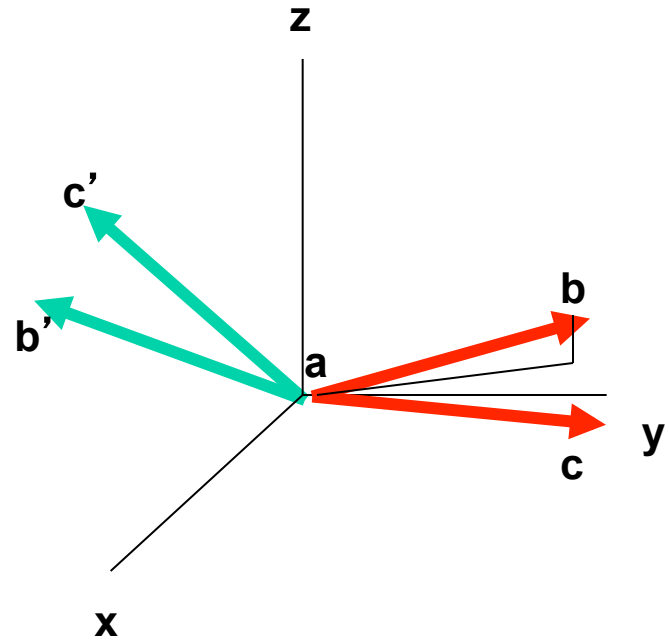
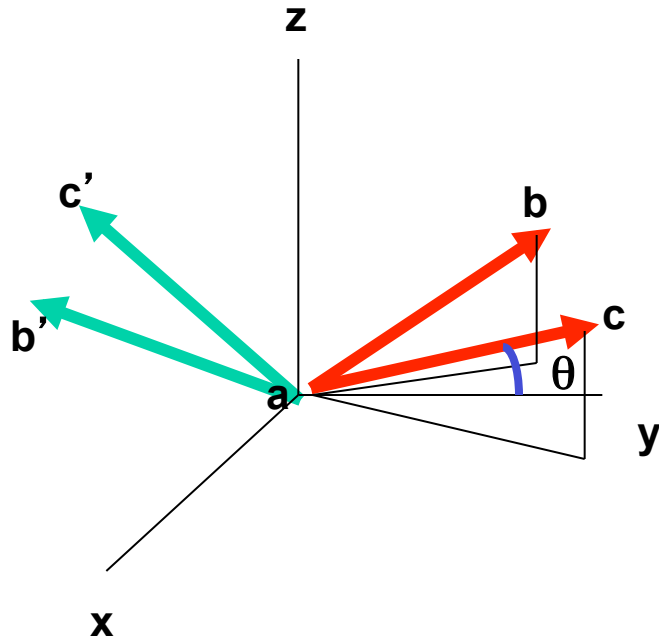
Identify 3 “equivalence” points in objects to be aligned

Superposition



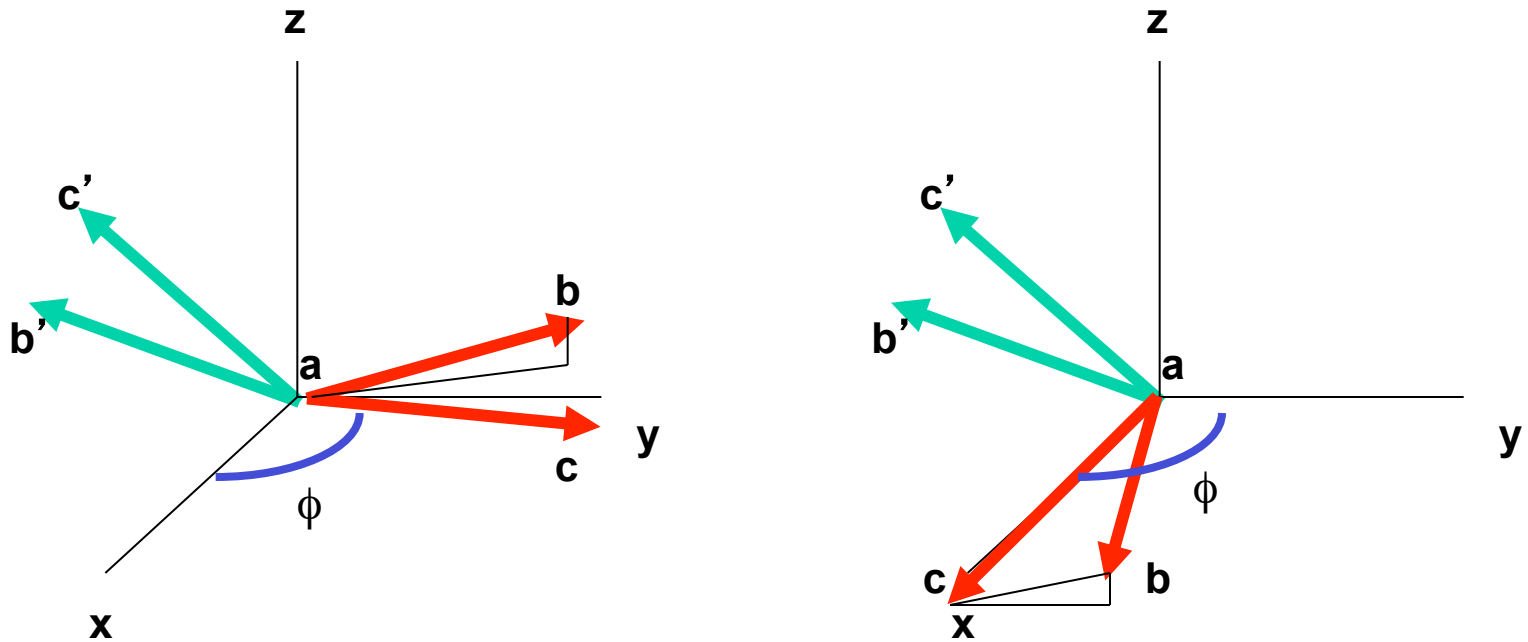
Translate points a, b, c and a', b', c' to origin

Superposition



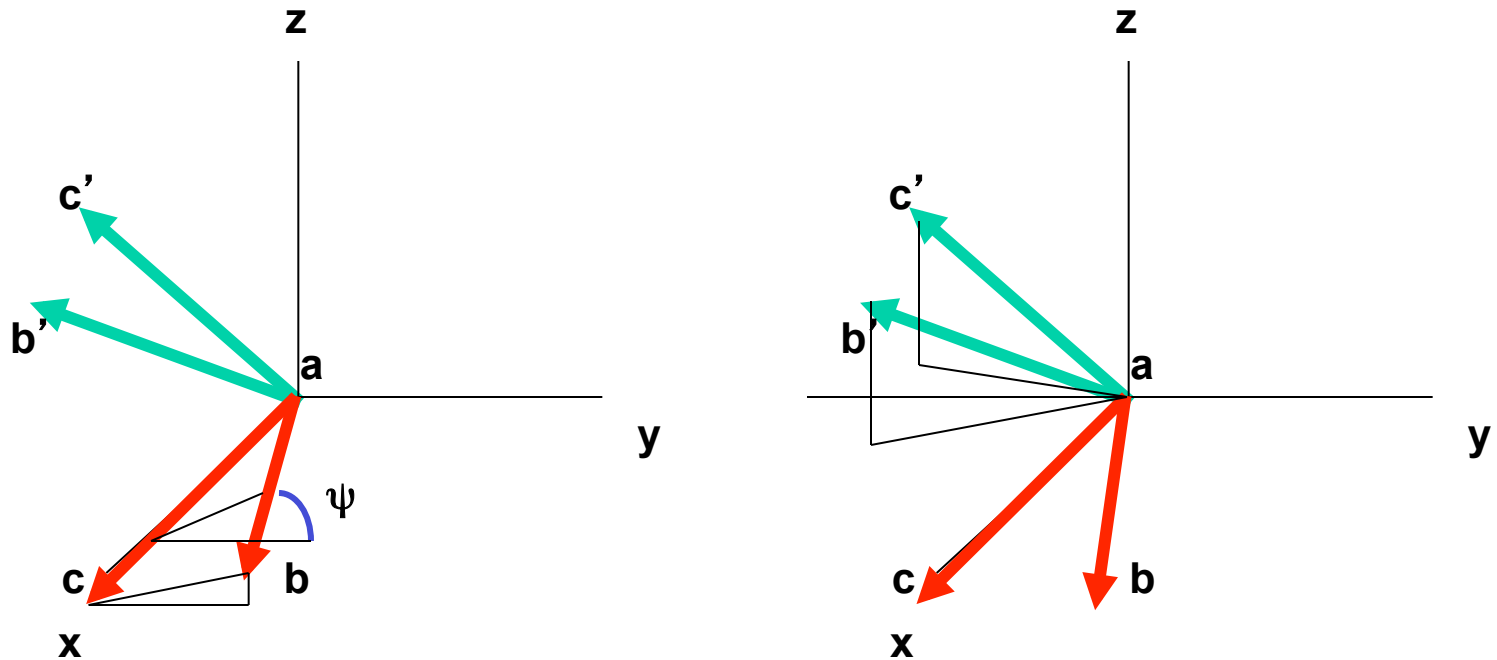
Rotate the a,b,c plane clockwise by θ about x axis

Superposition



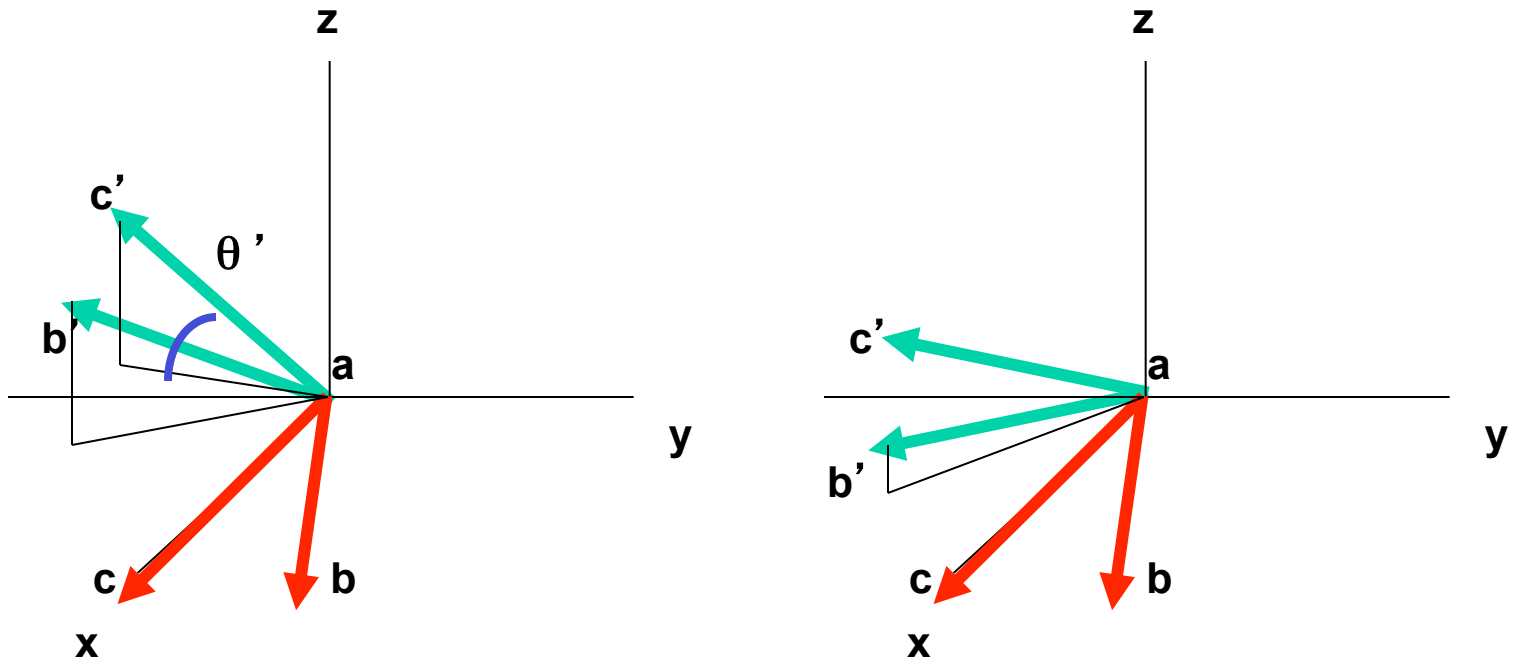
Rotate the a,b,c plane clockwise by ϕ about z axis

Superposition



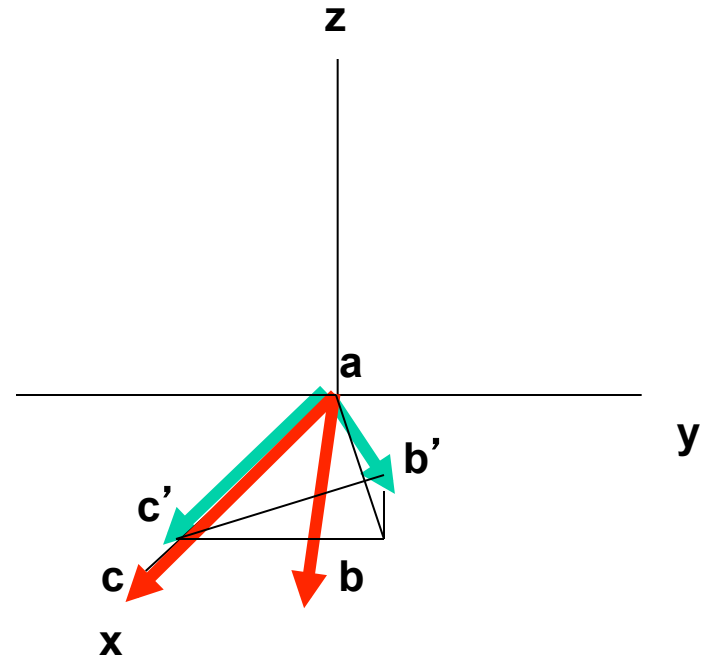
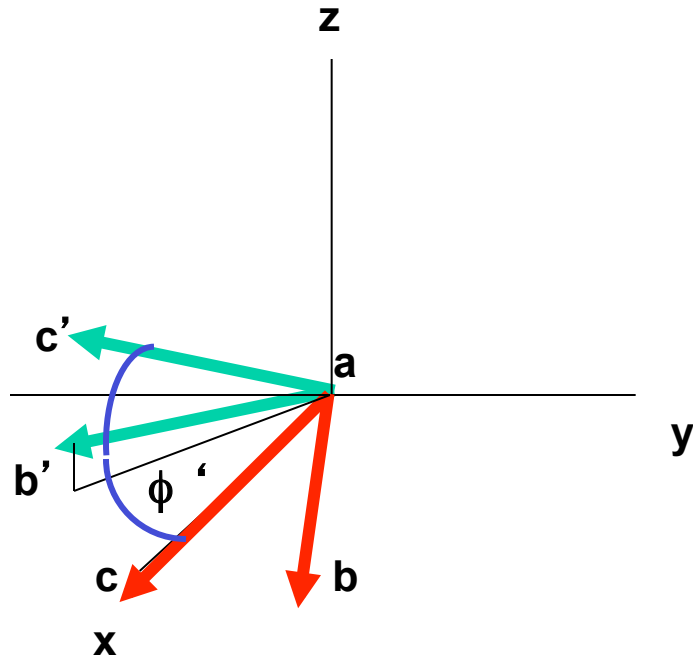
Rotate the a,b,c plane clockwise by ψ about x axis

Superposition



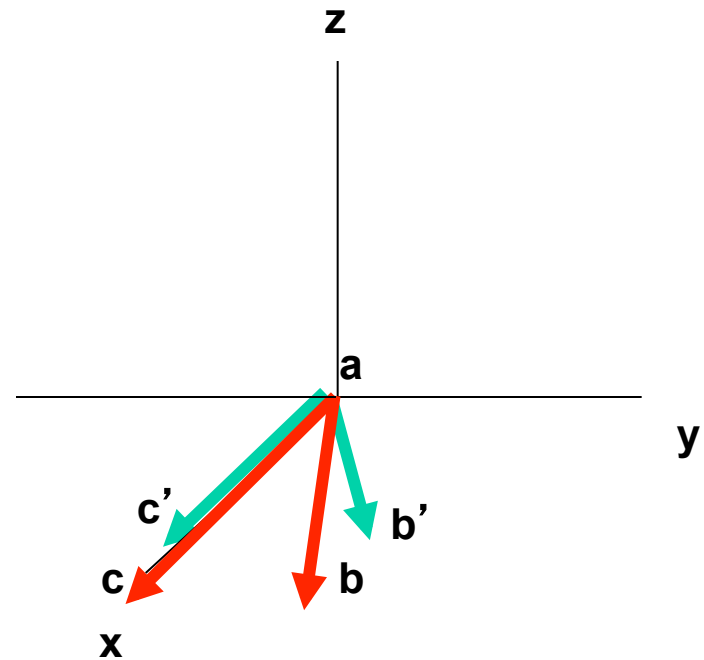
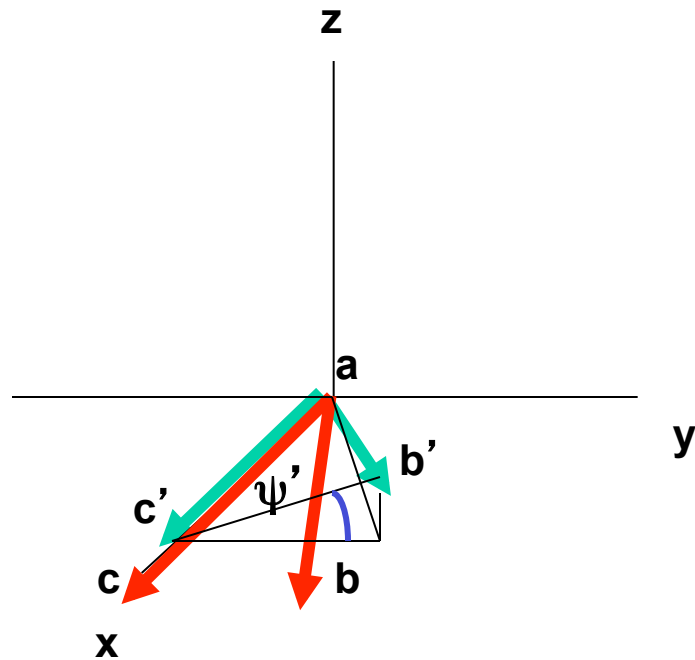
Rotate the a', b', c' plane anticlockwise by θ' about x axis

Superposition



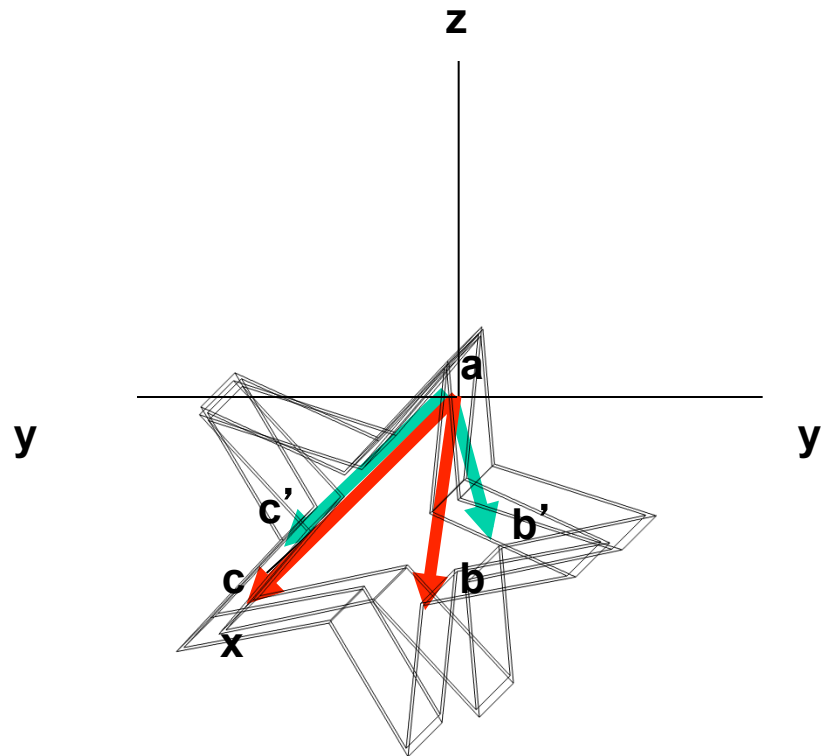
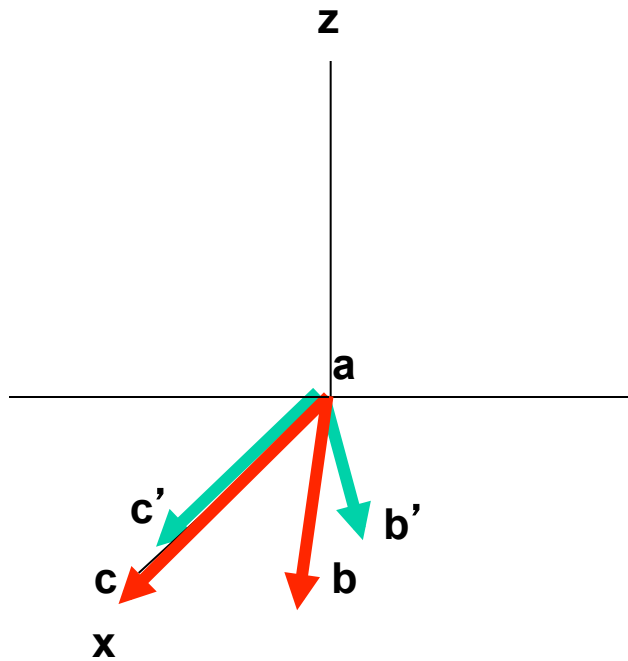
Rotate the a' , b' , c' plane anticlockwise by ϕ' about z axis

Superposition



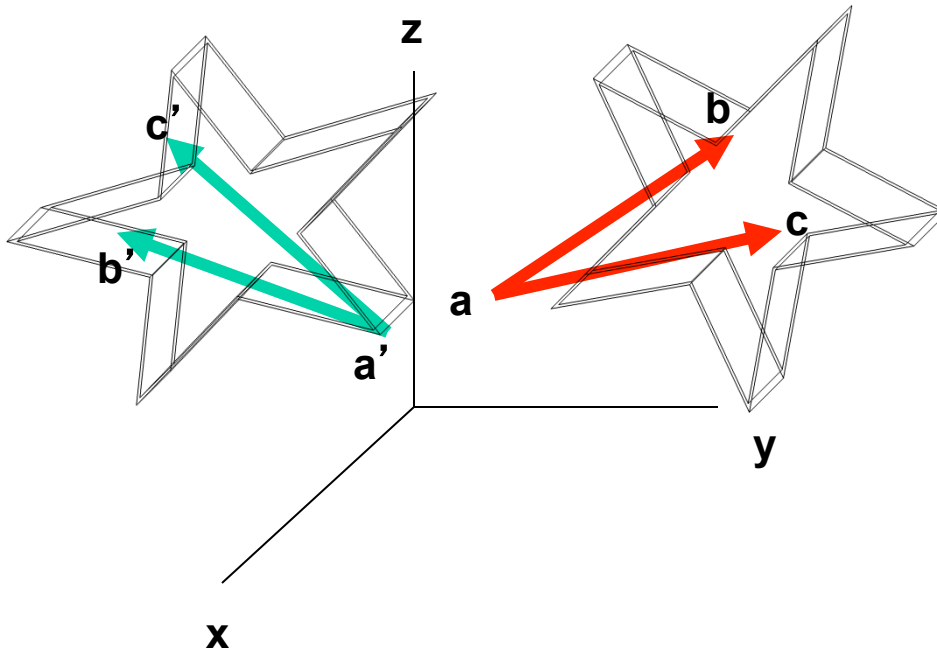
Rotate the a' , b' , c' plane clockwise by ψ' about x axis

Superposition

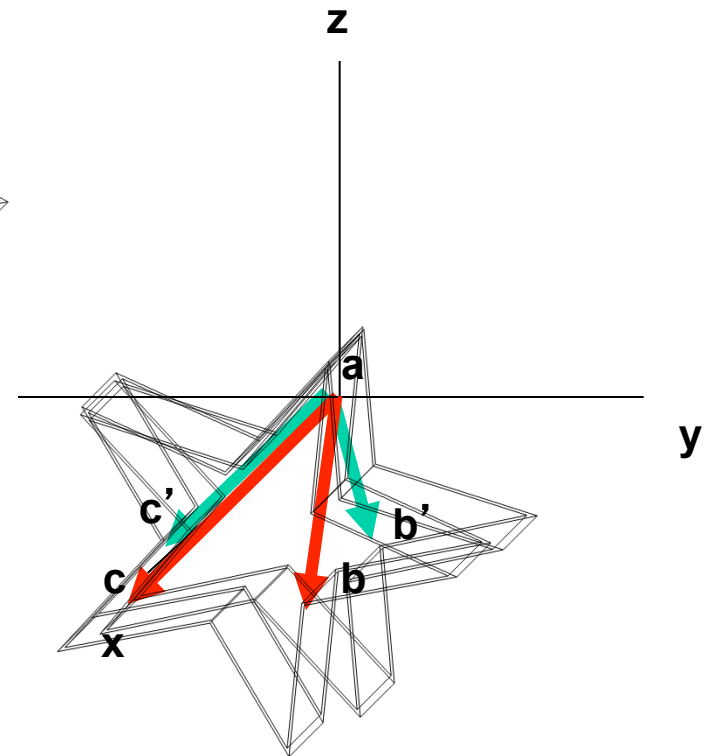


Apply all rotations and translations to remaining points

Superposition

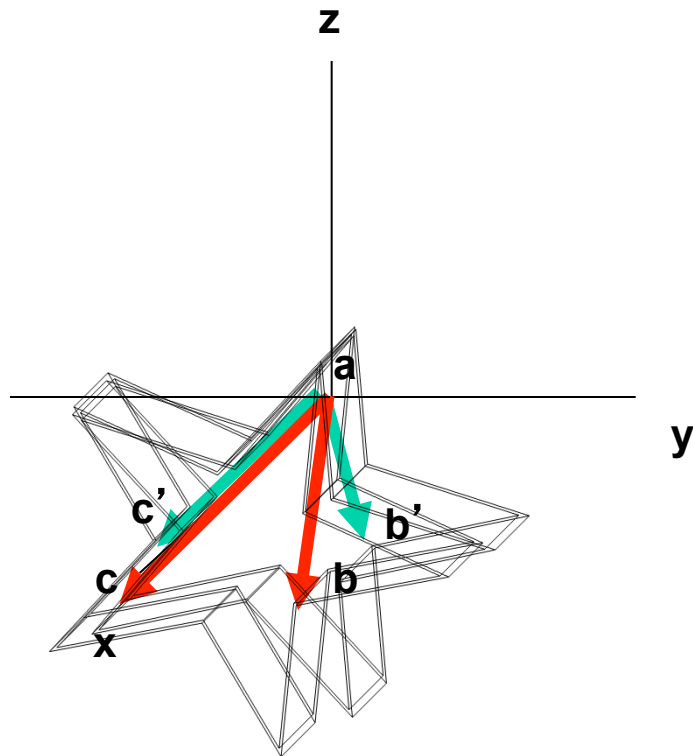


Before

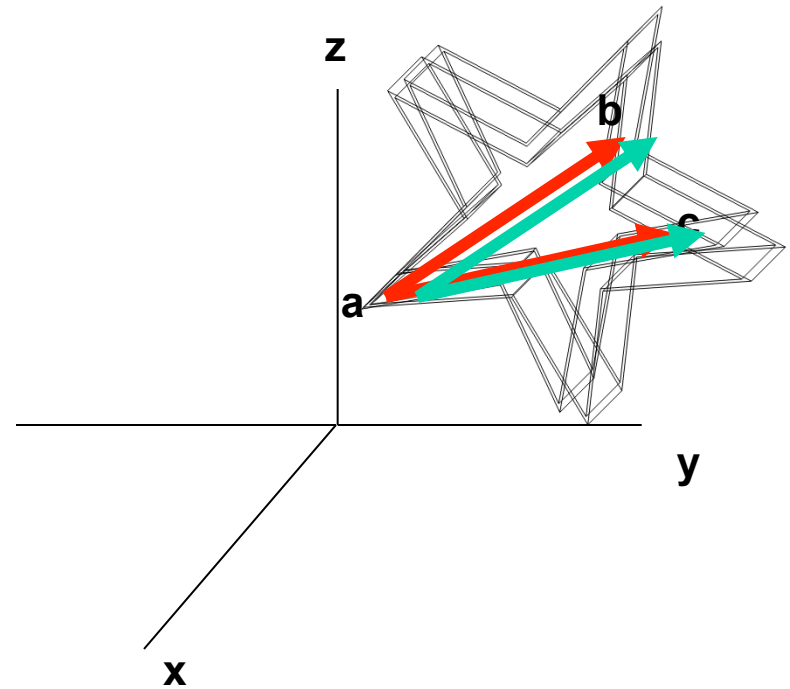


After

Returning to the “red” frame



Before



After

Returning to the “red” frame*

- **Begin with the superimposed structures on the x-y plane**
- **Apply counterclockwise rot. By ψ**
- **Apply counterclockwise rot. By ϕ**
- **Apply counterclockwise rot. By θ**
- **Apply red translation to red origin**

Just do things in reverse order!

Shortcomings*

- **Requires some initial assumptions regarding the anchoring points for superposition**
- **Anchoring points can't always *a priori* be known or easily calculated**
- **It “privileges” the first point “a” over “c” which is in turn privileged over “b”**

More General Approaches*

- **Monte Carlo or Genetic Algorithms**
- **Matrix methods using least squares or conjugate gradient minimization (McLachlan/Kabsch)**
- **Lagrangian multipliers**
- **Rotation Angle Methods**
- **Quaternion-based methods (fastest)**

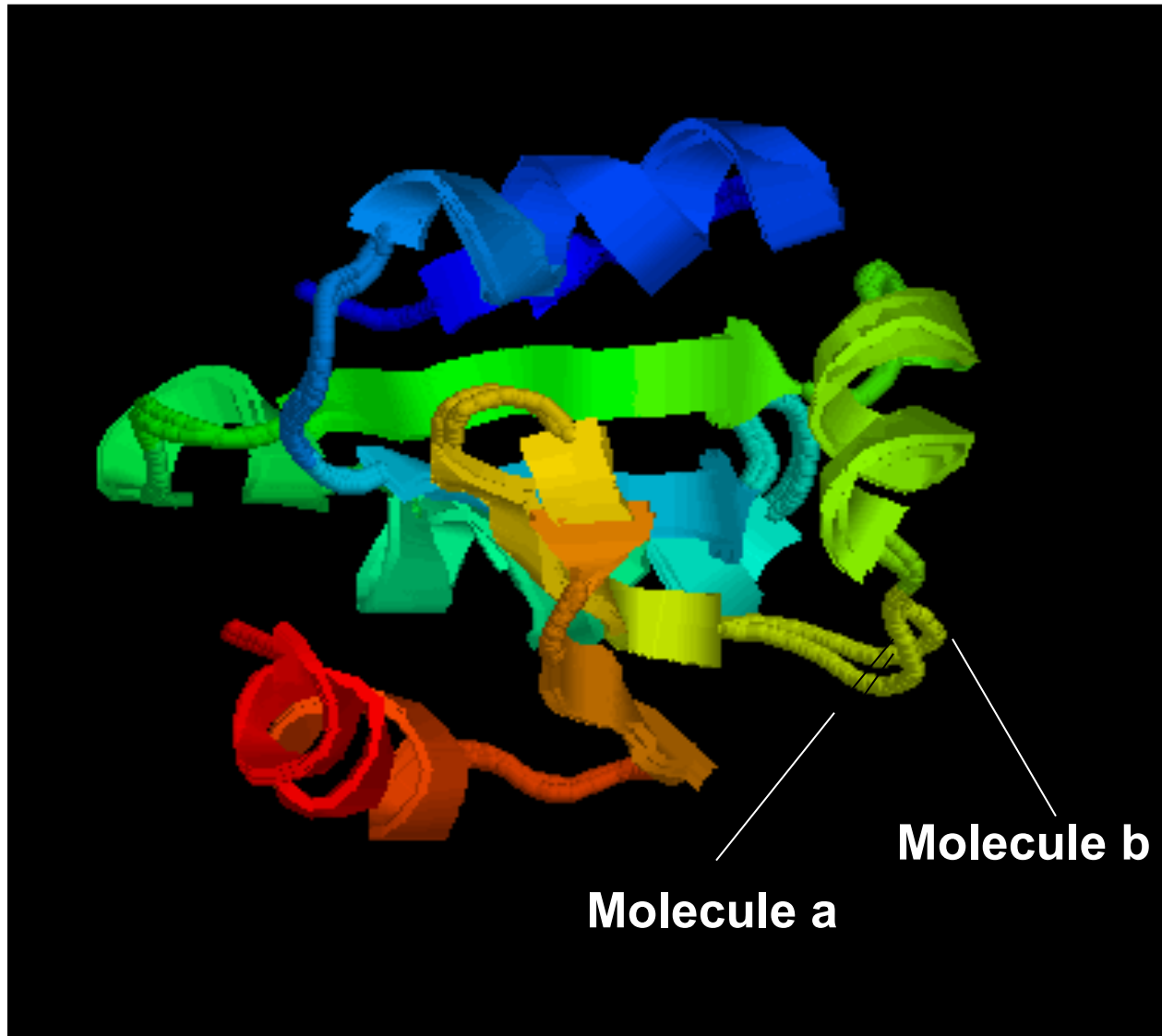
Superposition – Applications*

- **Ideal for comparing or overlaying two or more protein structures**
- **Allows identification of structural homologues (CATH and SCOP)**
- **Allows loops to be inserted or replaced from loop libraries (comparative modelling)**
- **Allows side chains to be replaced or inserted with relative ease**

Outline

- **Visualization Programs**
- **Vectors & Matrices**
- **Difference Distance Matrices**
- **Molecular Superposition**
- **Measuring Superposition**
- **Classifying 3D Structures**

Measuring Superpositions

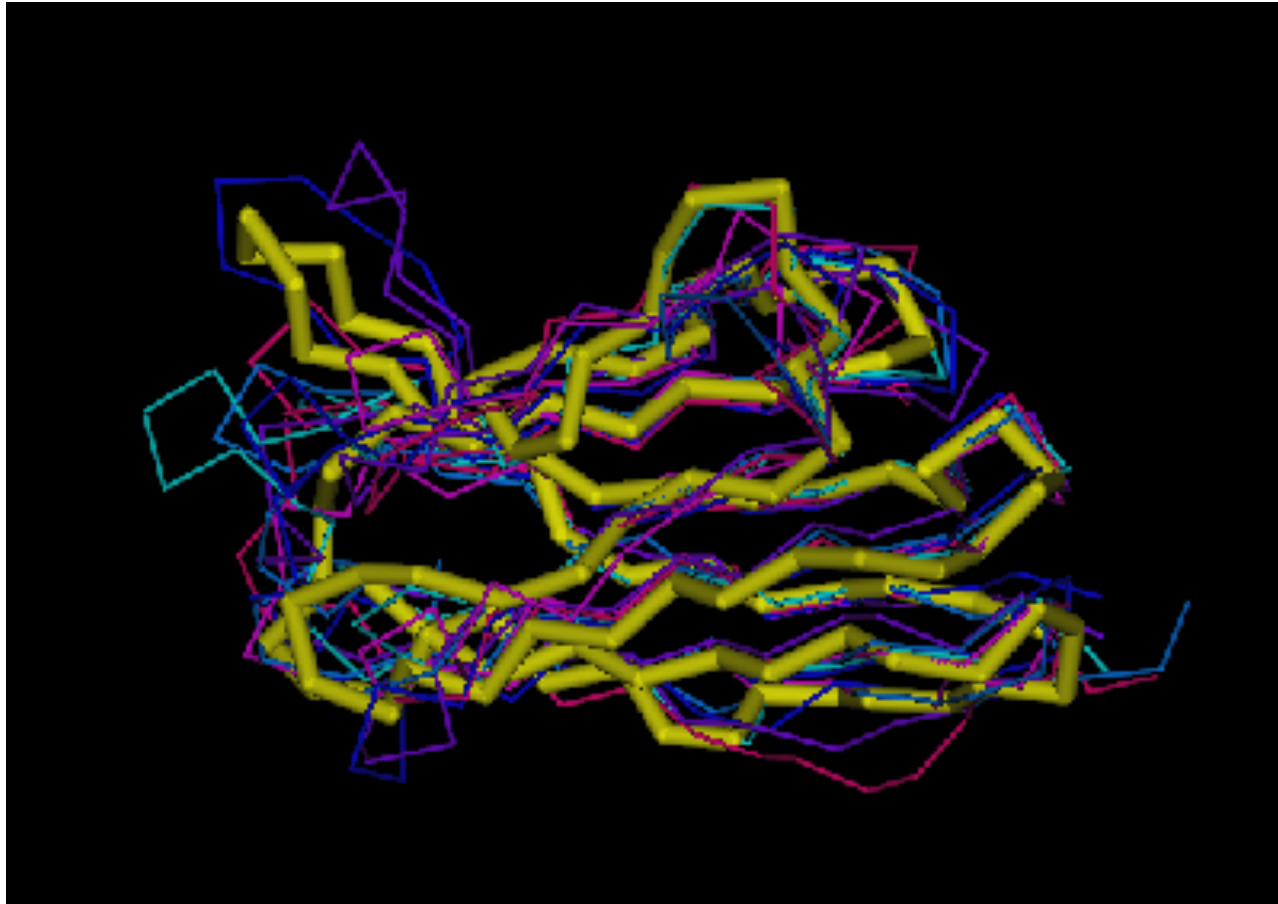


RMSD - Root Mean Square Deviation*

- Method to quantify structural similarity - same as standard deviation
- Requires 2 superimposed structures (designated here as “a” & “b”)
- N = number of atoms being compared

$$\text{RMSD} = \frac{\sqrt{\sum_i (x_{ai} - x_{bi})^2 + (y_{ai} - y_{bi})^2 + (z_{ai} - z_{bi})^2}}{\sqrt{N}}$$

Superpositions for Multiple Structures



RMSD - For Multiple Structures*

- Requires multiple superimposed structures over a single “averaged” structure $(\bar{x}, \bar{y}, \bar{z})$
- N = number of atoms being compared
- M = number of structures superimposed

$$\text{RMSD} = \sum_a \left\{ \frac{\sqrt{\sum_i (x_{ai} - \bar{x}_i)^2 + (y_{ai} - \bar{y}_i)^2 + (z_{ai} - \bar{z}_i)^2}}{\sqrt{N}} \right\}$$

M

RMSD without Superposition*

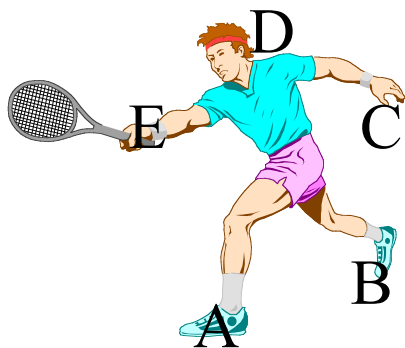
	A	B	C	D	E
A	0	4	5	6	4
B		0	4	6	7
C			0	3	6
D				0	3
E					0

-

	A	B	C	D	E
A	0	4	5	6	4
B		0	4	5	3
C			0	3	5
D				0	3
E					0

=

	A	B	C	D	E
A	0	0	0	0	0
B		0	0	1	4
C			0	0	1
D				0	0
E					0

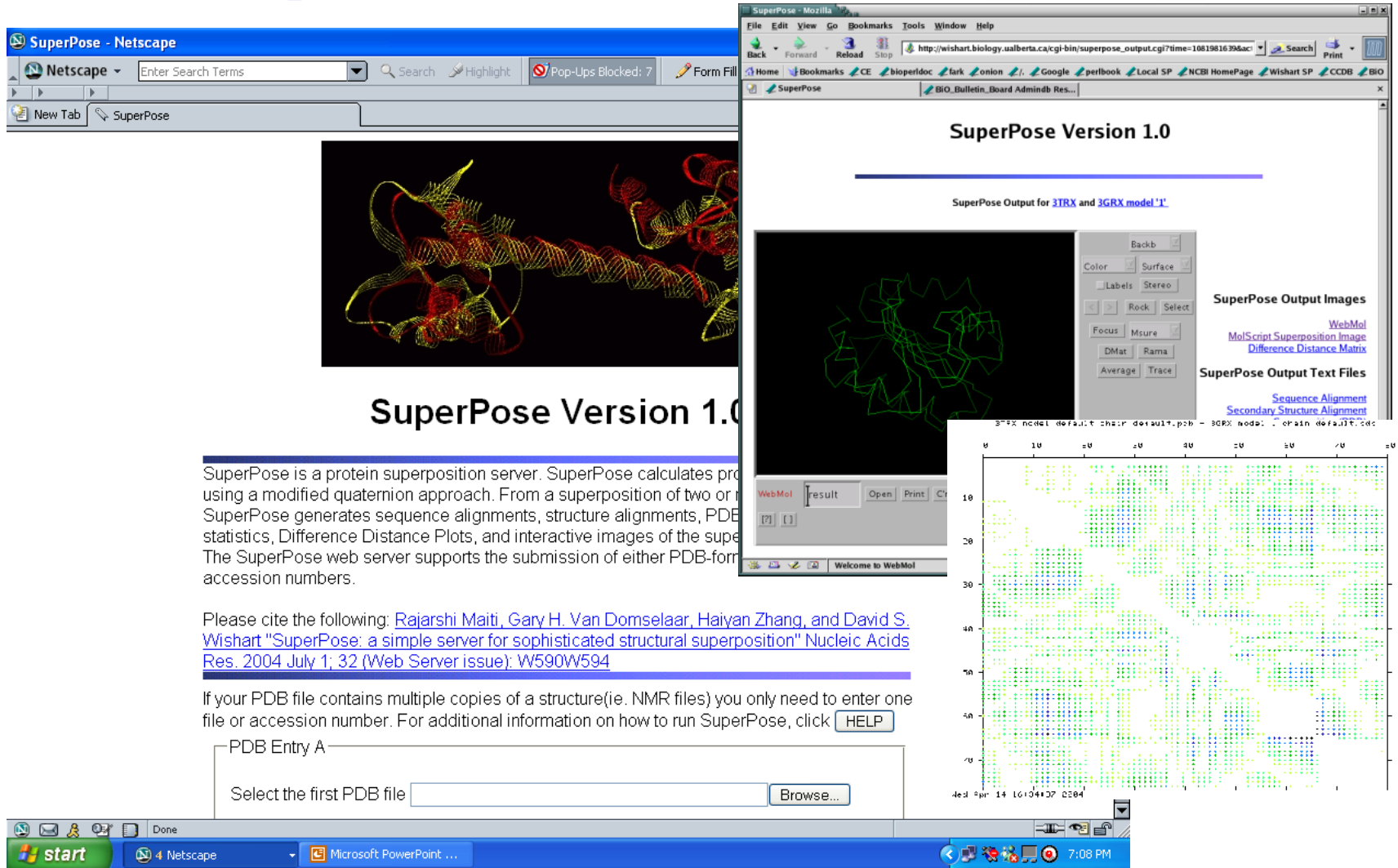


$$\text{RMS} = \frac{1+4+1}{\sqrt{10}} = 1.89$$

RMSD*

- **0.0-0.5 Å → Essentially Identical**
- **<1.5 Å → Very good fit**
- **< 5.0 Å → Moderately good fit**
- **5.0-7.0 Å → Structurally related**
- **> 7.0 Å → Dubious relationship**
- **> 12.0 Å → Completely unrelated**

SuperPose Web Server



SuperPose Version 1.0

SuperPose Output for **3TRX** and **3GRX** model.''

SuperPose Output Images

- [WebMol](#)
- [MolScript Superposition Image](#)
- [Difference Distance Matrix](#)

SuperPose Output Text Files

- [Sequence Alignment](#)
- [Secondary Structure Alignment](#)

SuperPose is a protein superposition server. SuperPose calculates protein superpositions using a modified quaternion approach. From a superposition of two or three protein structures, SuperPose generates sequence alignments, structure alignments, PDB statistics, Difference Distance Plots, and interactive images of the superposition. The SuperPose web server supports the submission of either PDB-format or accession numbers.

Please cite the following: [Rajarshi Maiti, Gary H. Van Domselaar, Haiyan Zhang, and David S. Wishart "SuperPose: a simple server for sophisticated structural superposition" Nucleic Acids Res. 2004 July 1; 32 \(Web Server issue\): W590W594](#)

If your PDB file contains multiple copies of a structure (i.e. NMR files) you only need to enter one file or accession number. For additional information on how to run SuperPose, click [HELP](#)

PDB Entry A: _____

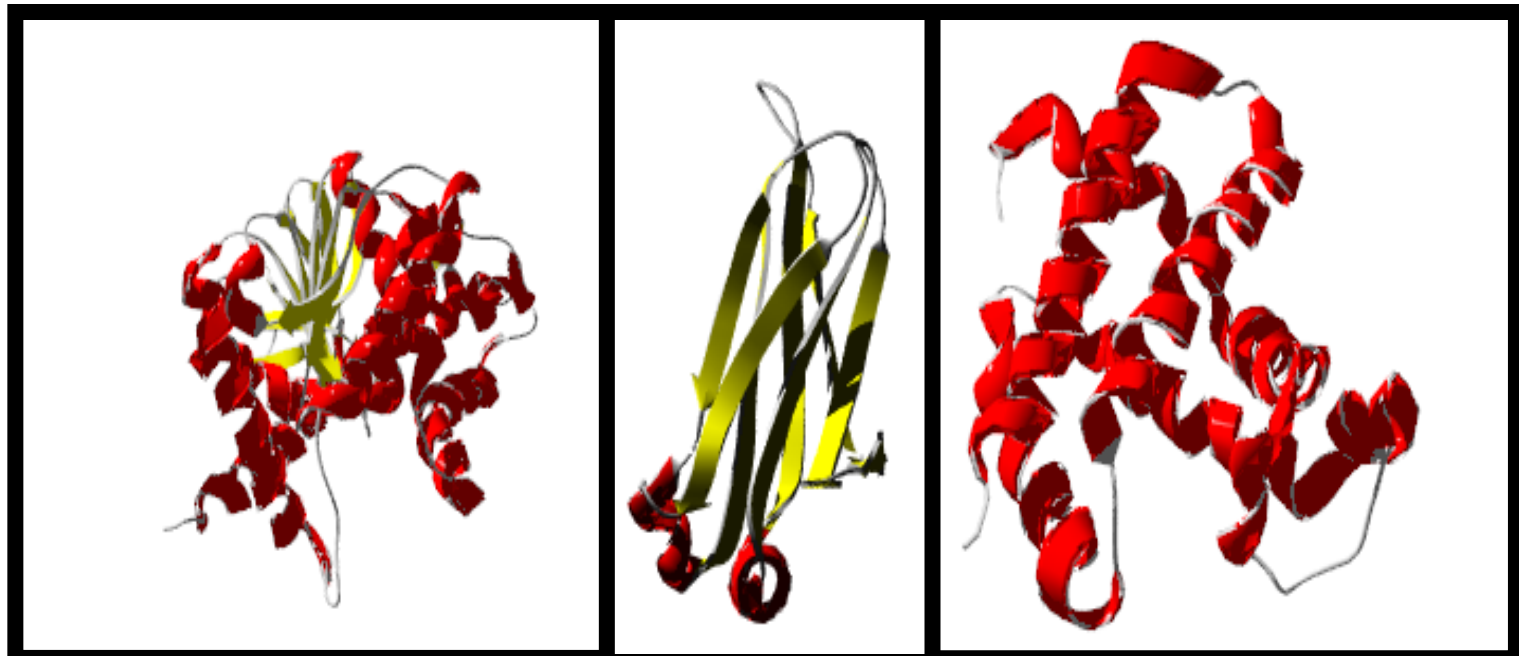
Select the first PDB file [Browse...](#)

<http://wishart.biology.ualberta.ca/SuperPose/>

Outline

- **Visualization Programs**
- **Vectors & Matrices**
- **Difference Distance Matrices**
- **Molecular Superposition**
- **Measuring Superposition**
- **Classifying 3D Structures**

Classifying Protein Folds*

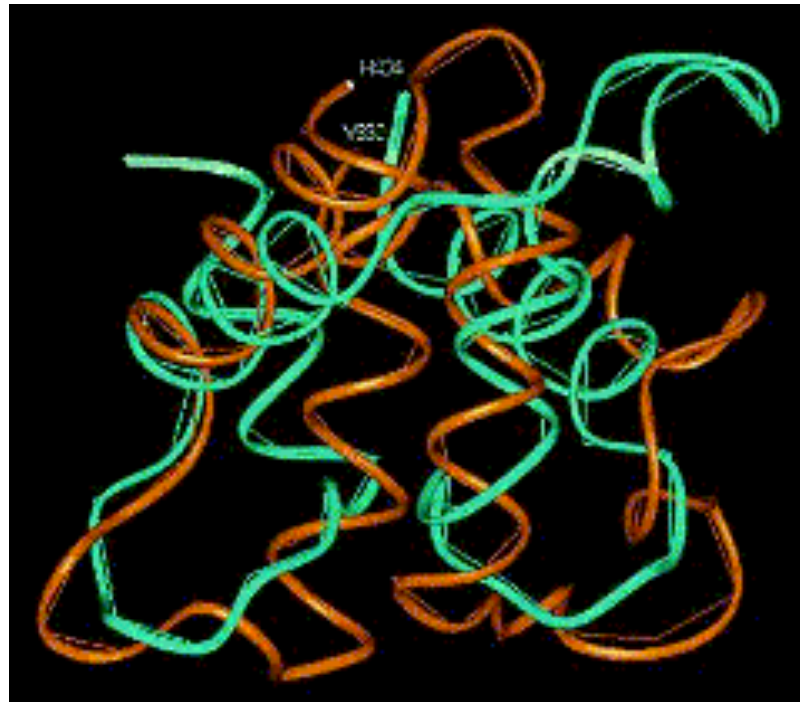


Lactate
Dehydrogenase:
Mixed α / β

Immunoglobulin
Fold: β

Hemoglobin B
Chain: α

Detecting Unusual Relationships



Similarity between Calmodulin and Acetylcholinesterase

Classifying Protein Folds

RCSB PDB - 2TRX Annotations Report

http://www.pdb.org/pdb/explore/derivedData.do?structureId=2TRX

Most Visited Getting Started Latest Headlines

RCSB PDB - 2TRX Annotations Re...

Summary Sequence **Annotations** Link Similarity 3D Similarity Literature Biol. & Chem

CRYSTAL STRUCTURE OF THIOREDOXIN FROM ESCHERICHIA COLI AT 1.68 ANGSTROMS RESOLUTION

Annotation data related to this entry.

Domain Annotation: SCOP Classification (version 1.75)

Domain Info	Class	Fold	Superfamily	Family
d2trxa_	Alpha and beta proteins (a/b)	Thioredoxin fold	Thioredoxin-like	Thioltransfer
d2trxb_	Alpha and beta proteins (a/b)	Thioredoxin fold	Thioredoxin-like	Thioltransfer

Domain Annotation: CATH Classification (version 3.4.0)

Domain	Class	Architecture	Topology
2trxA00	Alpha Beta	3-Layer(aba) Sandwich	Glutar
2trxB00	Alpha Beta	3-Layer(aba) Sandwich	Glutar

Protein Family Annotation: PFAM Classification

Chain	PFAM Accession	PFAM ID	Description
A	PF00085	Thioredoxin	Thioredoxin
B	PF00085	Thioredoxin	Thioredoxin

Gene Product Annotation: GO Terms

Polymer	Molecular Function	Biological Process	Cellular Component
THIOREDOXIN (2TRX:A,B)	<ul style="list-style-type: none"> protein binding protein disulfide 	<ul style="list-style-type: none"> transport electron transport chain 	<ul style="list-style-type: none"> none

SCOP: Structural Classification of Proteins

2 Structural Classification of Proteins

Welcome to SCOP: Structural Classification of Proteins. 1.75 release (June 2009)

38221 PDB Entries. 1 Literature Reference. 110800 Domains. (excluding nucleic acids and theoretical models)

Folds, superfamilies, and families [statistics here](#).
[New folds superfamilies families](#).
[List of obsolete entries and their replacements](#).

Authors: Alexey G. Murzin, John-Marc Chandonia, Antonina Andreeva, Dave Howorth, Loredana Lo Conte, Bartlett G. Ailey, Steven E. Brenner, Tim J. P. Hubbard, and Cyrus Chothia. scop@mrc-lmb.cam.ac.uk

Reference: Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536-540. [PDF]

Recent changes are described in: Lo Conte L., Brenner S. E., Hubbard T.J.P., Chothia C., Murzin A. (2002). SCOP database in 2002: refinements accommodate structural genomics. *Nucl. Acid Res.* 30(1), 264-267. [PDF].
 Andreeva A., Howorth D., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2004). SCOP database in 2004: refinements integrate structure and sequence family data. *Nucl. Acid Res.* 32:D226-D229. [PDF].
 Andreeva A., Howorth D., Chandonia J.-M., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2007). Data growth and its impact on the SCOP database: new developments. *Nucl. Acids Res.* 2008 36: D419-D425; doi:10.1093/nar/gkm993 [PDF].

Access methods

- Enter SCOP at the [top of the hierarchy](#)
- [Keyword search of SCOP entries](#)
- [SCOP parseable files](#)
- [All SCOP releases and reclassified entry history](#)
- [pre-SCOP - preview of the next release](#)
- SCOP domain sequences and pdb-style coordinate files ([ASTRAL](#))
- Hidden Markov Model library for SCOP superfamilies ([SUPERFAMILY](#))
- Structural alignments for proteins with non-trivial relationships ([SISYPHUS](#))
- [Online resources](#) of potential interest to SCOP users

SCOP [mirrors](#) around the world may speed your access.

Done

SCOP Database

SCOP: Structural Classification of Proteins

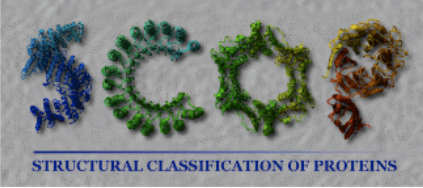
http://scop.mrc-lmb.cam.ac.uk/scop/

Department of Biology Login - Department of Alberta Audiobaba Music Search Bioinformatics the U of A! Coilgun Basics 2 Pathguide: t...esource list

Structural Classification of Proteins

Welcome to **SCOP: Structural Classification of Proteins**.
1.75 release (June 2009)

38221 PDB Entries. 1 Literature Reference. 110800 Domains. (excluding nucleic acids and theoretical models).
Folds, superfamilies, and families [statistics here](#).
[New folds superfamilies families](#).
[List of obsolete entries and their replacements](#).



Authors. Alexey G. Murzin, John-Marc Chandonia, Antonina Andreeva, Dave Howorth, Loredana Lo Conte, Bartlett G. Ailey, Steven E. Brenner, Tim J. P. Hubbard, and Cyrus Chothia. scop@mrc-lmb.cam.ac.uk

Reference: Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536-540. [\[PDF\]](#)

Recent changes are described in: Lo Conte L., Brenner S. E., Hubbard T.J.P., Chothia C., Murzin A. (2002). SCOP database in 2002: refinements accommodate structural genomics. *Nucl. Acid Res.* 30(1), 264-267. [\[PDF\]](#),
Andreeva A., Howorth D., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2004). SCOP database in 2004: refinements integrate structure and sequence family data. *Nucl. Acid Res.* 32:D226-D229. [\[PDF\]](#), and
Andreeva A., Howorth D., Chandonia J.-M., Brenner S.E., Hubbard T.J.P., Chothia C., Murzin A.G. (2007). Data growth and its impact on the SCOP database: new developments. *Nucl. Acids Res.* 2008 36: D419-D425; doi:10.1093/nar/gkm993 [\[PDF\]](#).

Access methods

- Enter SCOP at the [top of the hierarchy](#)
- [Keyword search of SCOP entries](#)
- [SCOP parseable files](#)
- [All SCOP releases and reclassified entry history](#)
- [pre-SCOP - preview of the next release](#)
- SCOP domain sequences and pdb-style coordinate files ([ASTRAL](#))
- Hidden Markov Model library for SCOP superfamilies ([SUPERFAMILY](#))
- Structural alignments for proteins with non-trivial relationships ([SISYPHUS](#))

- [Online resources](#) of potential interest to SCOP users

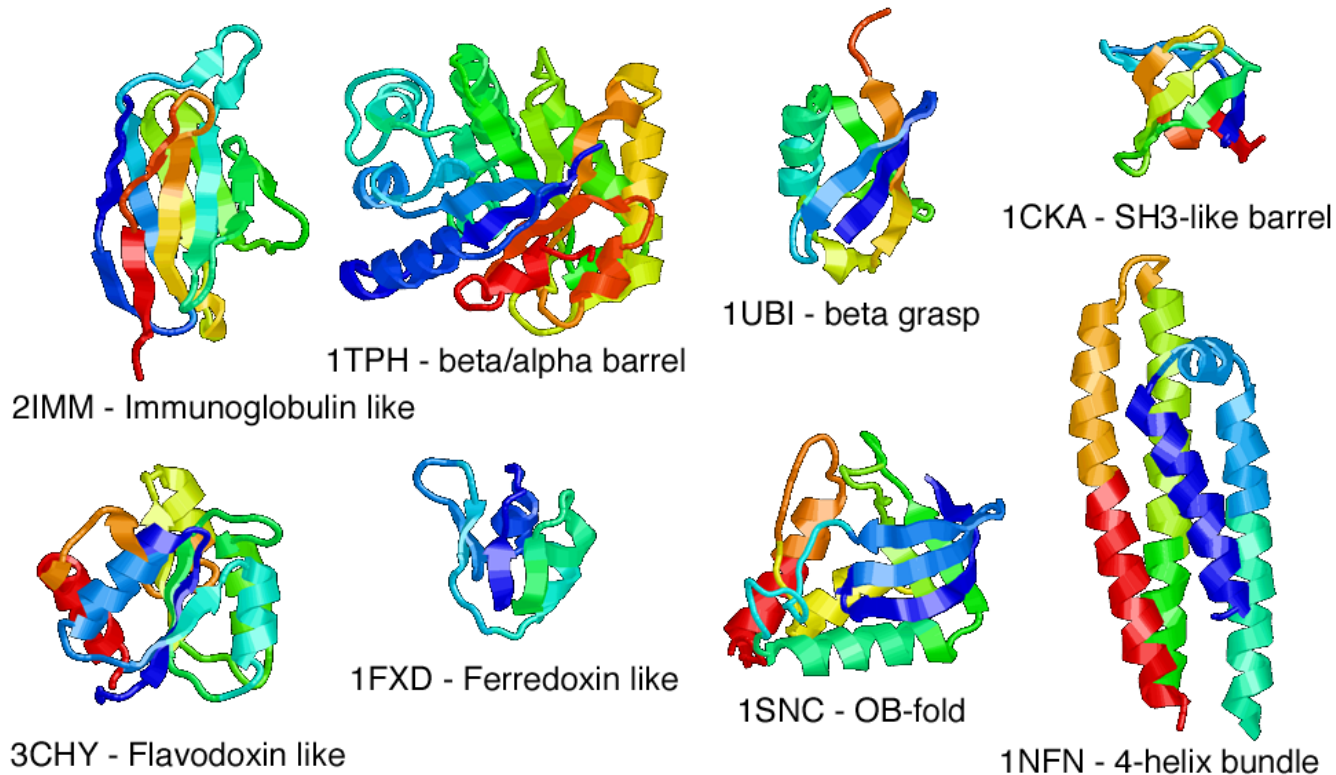
SCOP [mirrors](#) around the world may speed your access.

<http://scop.mrc-lmb.cam.ac.uk/scop>

SCOP

- **Class** folding class derived from secondary structure content
- **Fold** derived from topological connection, orientation, arrangement and # 2° structures
- **Superfamily** clusters of low sequence ID but related structures & functions
- **Family** clusters of proteins with seq ID > 30% with v. similar struct. & function

SCOP Structural Classification



The eight most frequent SCOP superfolds

The CATH Database

The screenshot shows a web browser window displaying the CATH Protein Structure Classification Database. The browser's address bar shows the URL <http://www.cathdb.info/>. The page features a dark blue header with the CATH logo and navigation links for Home, Search, Documentation, Tools, and Download. Below the header is a search bar with 'Search' and 'Clear' buttons. The main content area includes a 'Home' link, a 'Welcome to CATH' section with a detailed description of the database, a 'New in CATH v3.4' section listing recent updates, and a 'New in Gene3D 10.2' section listing new features. At the bottom, there are three columns of links: 'Using CATH' (Search, Browse, Download, Tutorials), 'CATH Tools' (Find My Sequence, Find My Structure, Linking to CATH), and 'About CATH' (Release Statistics, Glossary, CATH Team, References).

CATH: Protein Structure Classification Database – Prof. Orengo's Bioinformatics Group at UCL, London, UK

<http://www.cathdb.info/>

Resources » [CATH](#) [Gene3D](#) [FuncNet](#)

CATH

PROTEIN STRUCTURE CLASSIFICATION

[Home](#) | [Search](#) | [Documentation](#) | [Tools](#) | [Download](#)

[Home](#)

Welcome to CATH

CATH is a manually curated classification of protein domain structures. Each protein has been chopped into structural domains and assigned into homologous superfamilies (groups of domains that are related by evolution). This classification procedure uses a combination of automated and manual techniques which include computational algorithms, empirical and statistical evidence, literature review and expert analysis.

[Search the CATH database >>](#)

[Find out more about CATH >>](#)

New in CATH v3.4

CATH v3.4 is built from 104,238 PDB chains. We have added the following data since v3.3:

- 49 folds (total 1,282)
- 163 superfamilies (total 2,549)
- 1,311 sequence families (total 11,330)
- 24,232 domains (total 152,920)

[Download CATH data >>](#)

New in Gene3D 10.2

Gene3D 10.2 (released Sep 2011) uses CATH domains to provide 16,118,154 structural annotations for 14,963,305 protein sequences. The latest release also offers a number of new features:

- Interaction Network Data
- Interactive Graphical Representations
- Genome Comparisons

[Goto Gene3D >>](#)

Using CATH

- [Search](#)
- [Browse](#)
- [Download](#)
- [Tutorials](#)

CATH Tools

- [Find My Sequence](#)
- [Find My Structure](#)
- [Linking to CATH](#)

About CATH

- [Release Statistics](#)
- [Glossary](#)
- [CATH Team](#)
- [References](#)

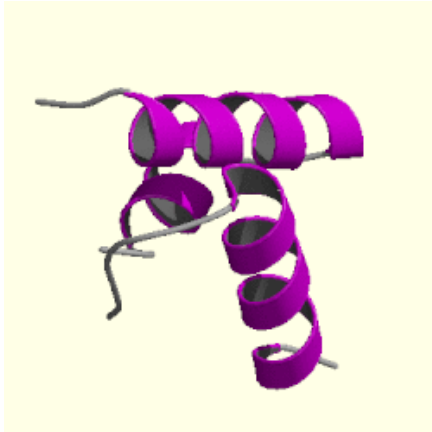
Done

<http://www.cathdb.info>

CATH

- **Class [C]** derived from secondary structure content (automatic)
- **Architecture (A)** derived from orientation of 2° structures (manual)
- **Topology (T)** derived from topological connection and # 2° structures
- **Homologous Superfamily (H)** clusters of similar structures & functions

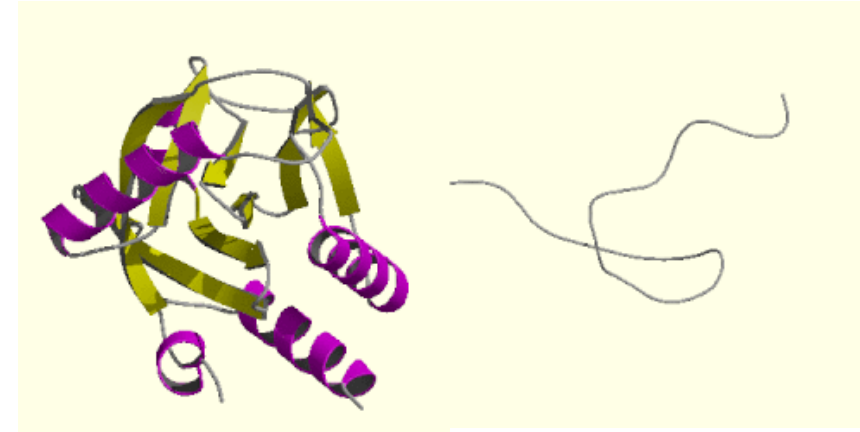
CATH - Class



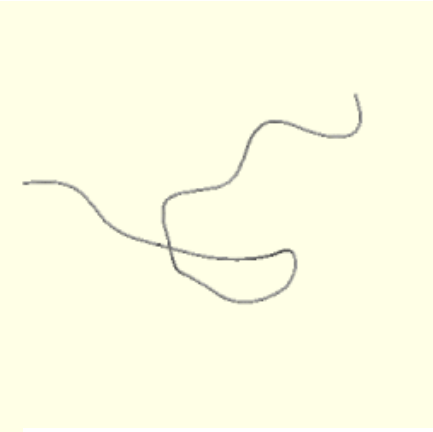
**Class 1:
Mainly Alpha**



**Class 2:
Mainly Beta**



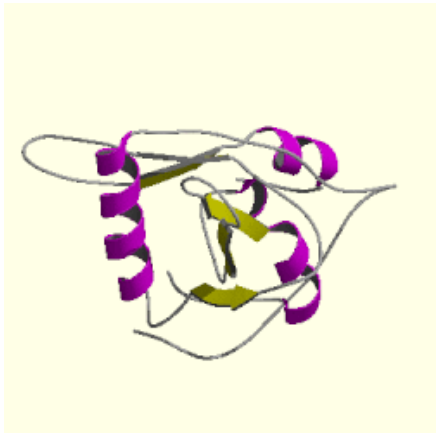
**Class 3:
Mixed Alpha/
Beta**



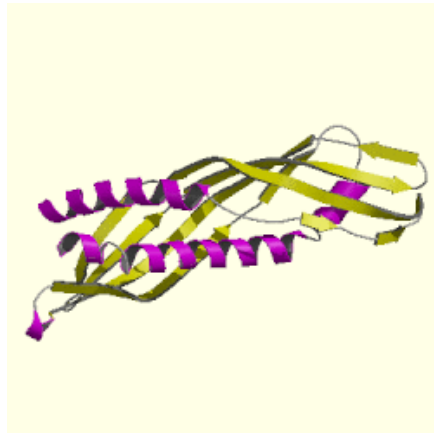
**Class 4:
Few Secondary
Structures**

Secondary structure content (automatic)

CATH - Architecture



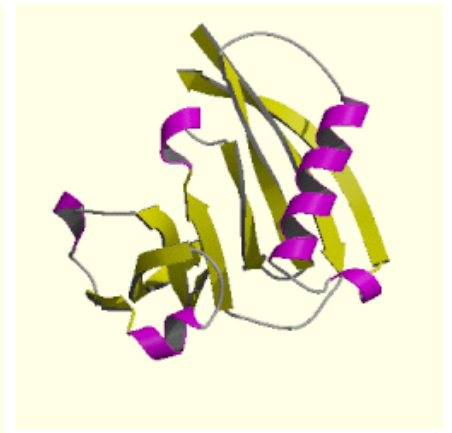
Roll



Super Roll



Barrel



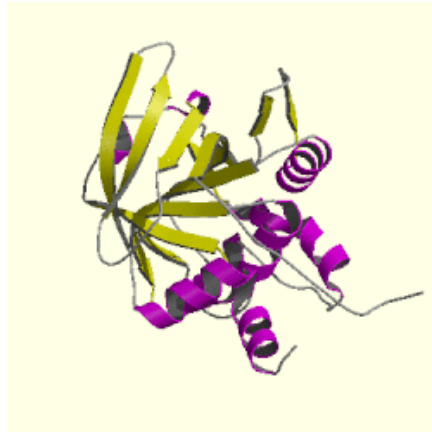
**2-Layer
Sandwich**

Orientation of secondary structures (manual)

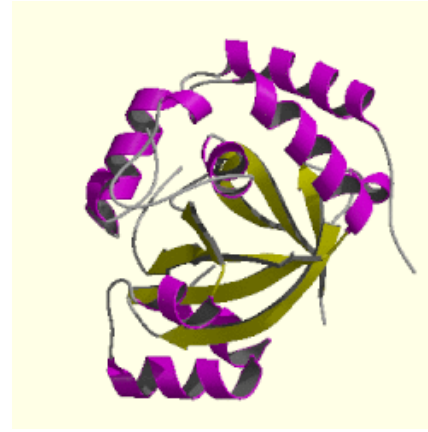
CATH - Topology



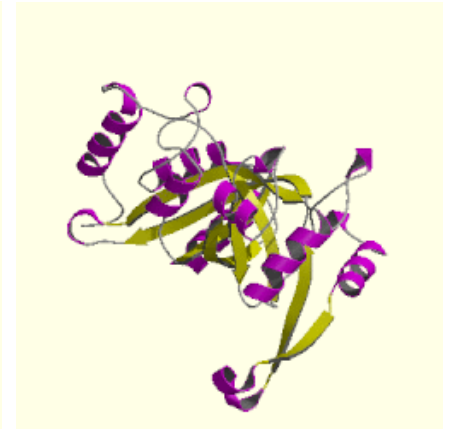
L-fucose Isomerase



Serine Protease



**Aconitase,
domain 4**



TIM Barrel

Topological connection and number of secondary structures

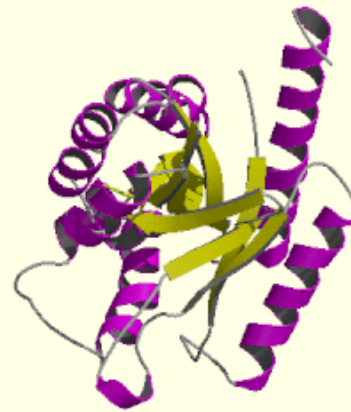
CATH - Homology



Alanine racemase



**Dihydropterotate (DHP)
synthetase**



**FMN dependent
fluorescent
proteins**



**7-stranded
glycosidases**

Superfamily clusters of similar structures & functions

Other Servers/Databases

- **Dali** - http://ekhidna.biocenter.helsinki.fi/dali_server/
- **VAST** - <http://www.ncbi.nlm.nih.gov/Structure/VAST/vast.shtml>
- **Matras** - <http://biunit.aist-nara.ac.jp/matras/>
- **CE** - <http://cl.sdsc.edu/ce.html>
- **TopMatch** - <http://topmatch.services.came.sbg.ac.at/>
- **PDBsum** - <http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/>

CE Search

CE CALCULATE CHAIN AGAINST PDB Calculate structural neighbors for one protein uploaded by the user against the PDB.

Enter a filename of a PDB formatted file containing the polypeptide chain for which you are seeking neighbors. Optionally specify your email address for notification of results and hit the **Search Database** button.

User File: Browse... ? Use structure representatives ?

Notify by e-mail (optional): ?

Time required to search structure neighbors depends on the level of similarity between query structure and known structures in PDB, size of the structure and load on the server and may vary from 5 minutes to 10 hours.

Option "Use structure representatives" (default) is advised for faster search. In this case structure neighbors for query structure will be derived from representative ? structure if it exists for a given query structure. Otherwise, full search will be performed. In both cases search is performed on a set of representative structures and then all other structure neighbors associated with representatives are added.

CE	FIND	ALL	REPRESENTATIVES	
	CALCULATE	TWO CHAINS	CHAIN AGAINST PDB	
	DOWNLOAD	SOFTWARE	DATABASES	
	RESULTS	SUBDOMAINS	PROTEIN KINASES	ESTERASES, LIPASES

<http://cl.sdsc.edu/ce/all-to-all/1-to-all.html>

CE Search

Structure Neighbors for Chain 2HHB:C - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Mail Print Mail Stop SnagIt

Address http://cl.sdsc.edu/ce_scratch/ata23376.html Go Links SnagIt

Chain 4HHB:A is represented by chain 2HHB:C. All structure neighbors are provided for representing chain.

Selected 52 chains Criteria used: Z-Score>4.0 RMSD<5.0Å

Sorted by: Z-Score

GET ALIGNMENT

Use the checkboxes to select one or more chains to align with 2HHB:C. In the ID column select the ID to get further information on that structure. In the ID column select "Neighbors" to display the structure neighbors matching that ID.

?	ID	Z-Score	RMSD (Å)	Seq. (%)	Aligned/ Size	Gap	Exp.	Name
	2HHB:C	Representative					X-Ray	HEMOGLOBIN (DEOXY)
<input checked="" type="checkbox"/>	1C7C:A Neighbors	6.9	0.2	100.0	141 / 283	0	X-Ray	MOL_ID: 1; MOLECULE: DEOXYHEMOGLOBIN (ALPHA CHAIN); CHAIN: A; ENGINEERED: YES; M
<input type="checkbox"/>	2MB5: Neighbors	6.6	1.6	26.2	141 / 165	6	X-Ray	MYOGLOBIN (CARBONMONOXYMYOGLOBIN) (NEUTRON STUDY)
<input type="checkbox"/>	1VRE:A Neighbors	6.3	2.2	20.3	138 / 149	9	X-Ray	MOL_ID: 1; MOLECULE: GLOBIN, MONOMERIC COMPONENT M-IV; CHAIN: A; SYNONYM: GHM4;
<input type="checkbox"/>	1ITH:B Neighbors	6.2	2.1	12.8	133 / 141	10	X-Ray	HEMOGLOBIN (CYANOMET)
<input type="checkbox"/>	1MBA: Neighbors	6.2	2.3	19.0	137 / 147	11	X-Ray	MYOGLOBIN (MET) (\$P*H 7.0)
<input type="checkbox"/>	1HLM: Neighbors	6.1	2.3	26.7	131 / 161	8	X-Ray	HEMOGLOBIN (CYANO-MET) (SEA CUCUMBER)

Done Internet

start CBW Proteom... Proteomics2006 2 Microsoft ... RCSB PDB : 5... SCOP: Protein... Structure Nei... 12:20 AM

Summary

- Many different tools and formats to visualize 3D structure – *learn how to use at least one of them*
- Visualization on computers is mostly about matrix and vector manipulation
- Structure comparison also requires the use of linear algebra
- Protein structures can be compared and aligned – just like sequences