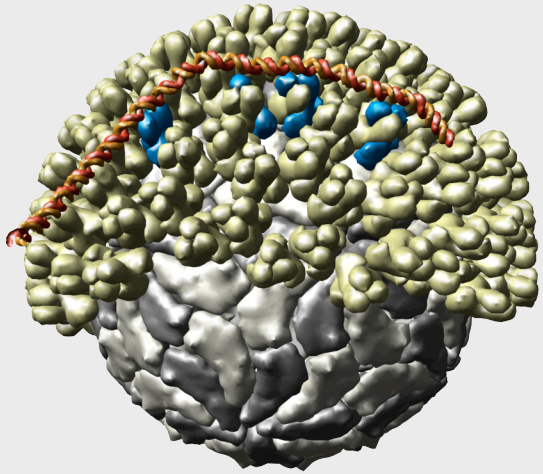
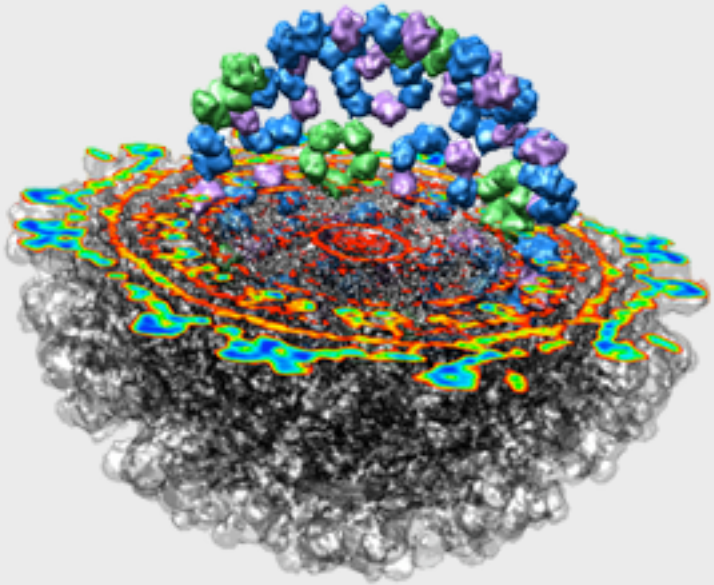
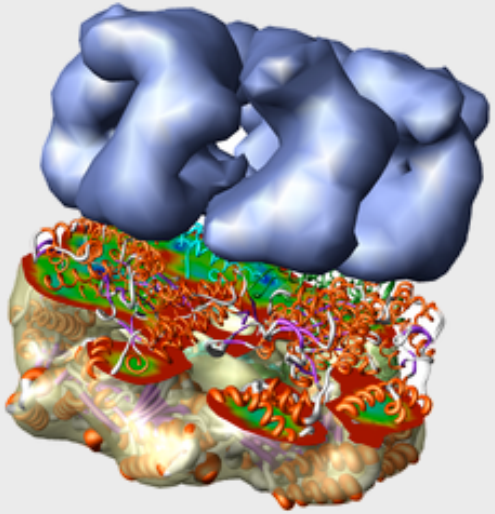
Visualization Tools for Complex Biological Structures



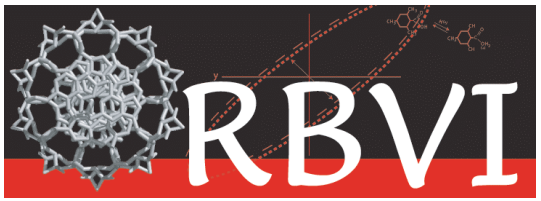




Tom Ferrin

Depts. of Pharmaceutical Chemistry and Biopharmaceutical Sciences
Resource for Biocomputing, Visualization, and Informatics
University of California at San Francisco

UCSF



Resource for Biocomputing, Visualization, and Informatics

We are an NIH Biomedical Technology Resource Center, funded by the National Center for Research Resources. We create innovative computational and visualization-based data analysis methods and algorithms, turn these into easy-to-use software tools which we distribute to the scientific community, and apply these tools for solving a wide range of genomic and molecular recognition problems within the complex sequence → structure → function triad.

National Center for Research Resources

Sample application areas

Insight into molecular structure and function:

- Protein engineering
- Drug design
- Biomaterials design
- Annotation of protein function from sequence and structure

Gene annotation, characterization, and interpretation:

- Pharmacogenetics - understanding and prediction of variation in drug response due to genetic factors
- Mouse gene knock-outs for modeling of human disease

Outline

This talk:

- Chimera overview
- Introductory demo
- Bluetongue virus demo
- Myosin fitting demo

This afternoon:

- Tom Goddard's "hands-on" Chimera exercise - Visualizing volume data from single particle EM reconstructions

"It's sink or swim as a tidal wave of data approaches"

Tony Reichhardt
Nature 399:517-520
June 1999

Petabyte (1,000 terabytes)

Exabyte (1,000 petabytes)

Zettabyte (1,000 exabytes)

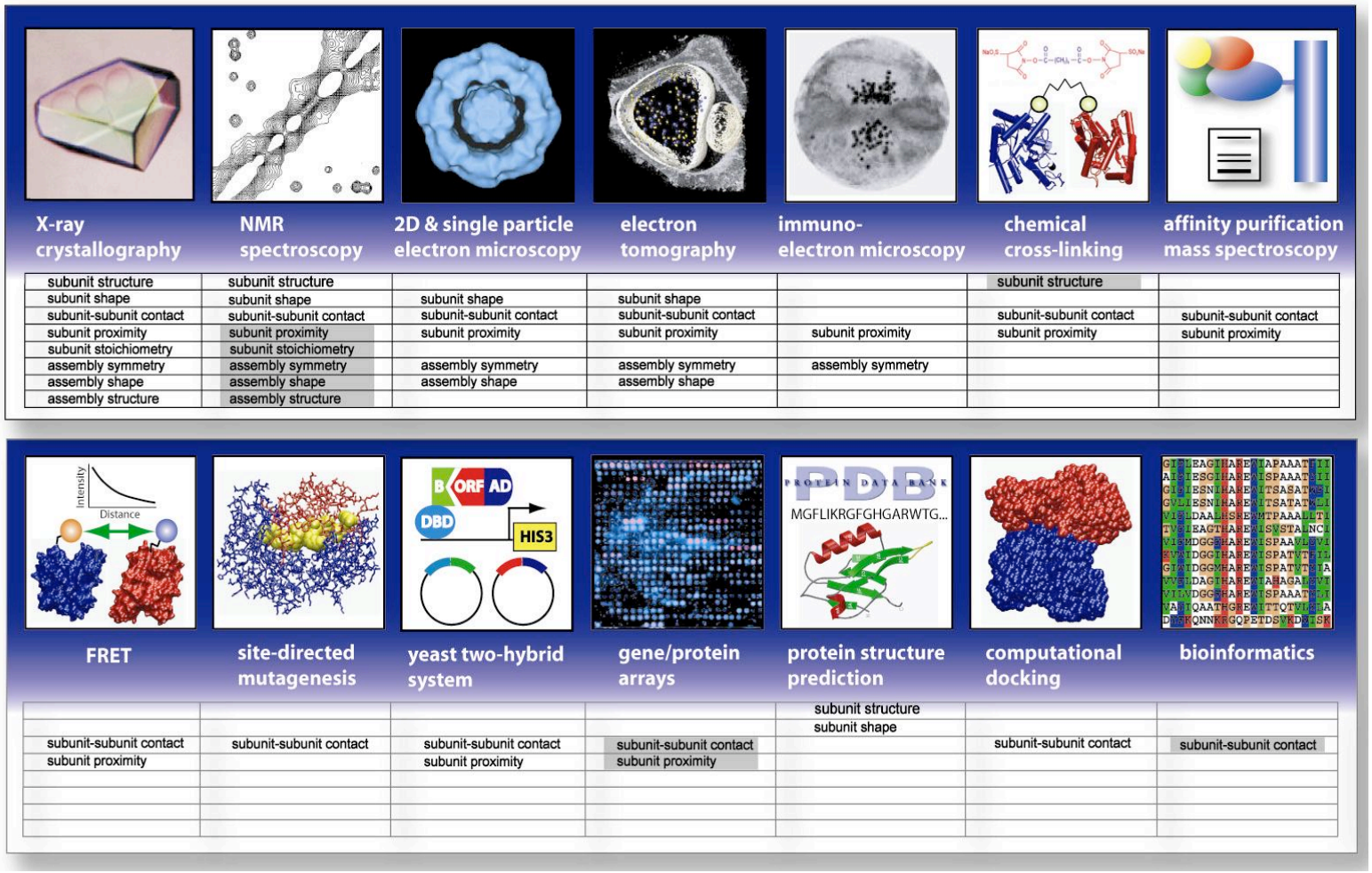
Yottabyte (1,000 zettabytes)

Determining the Structures of Proteins and Assemblies

Structural information from…

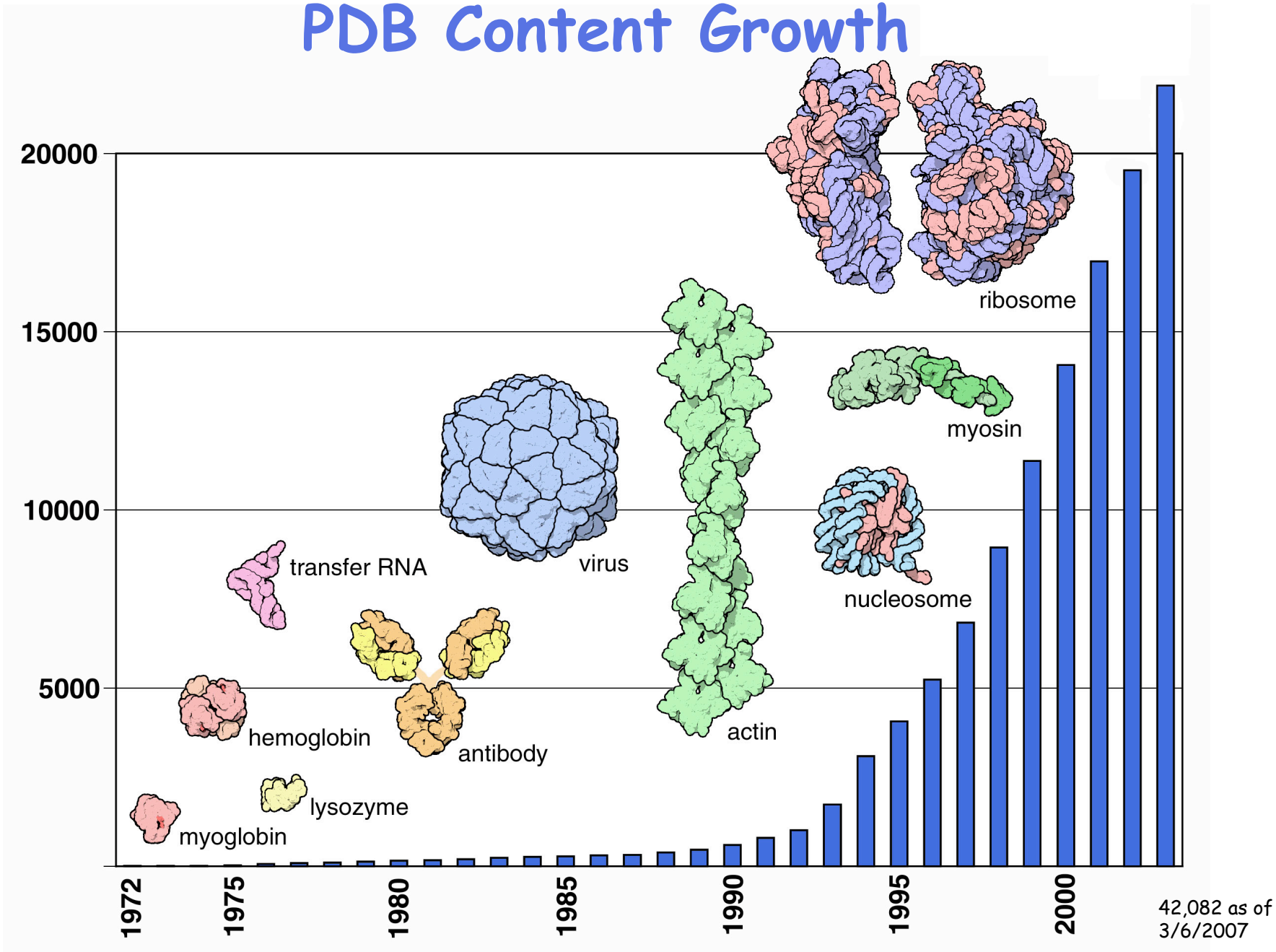
source: measurement and models

resolution: low or high resolution

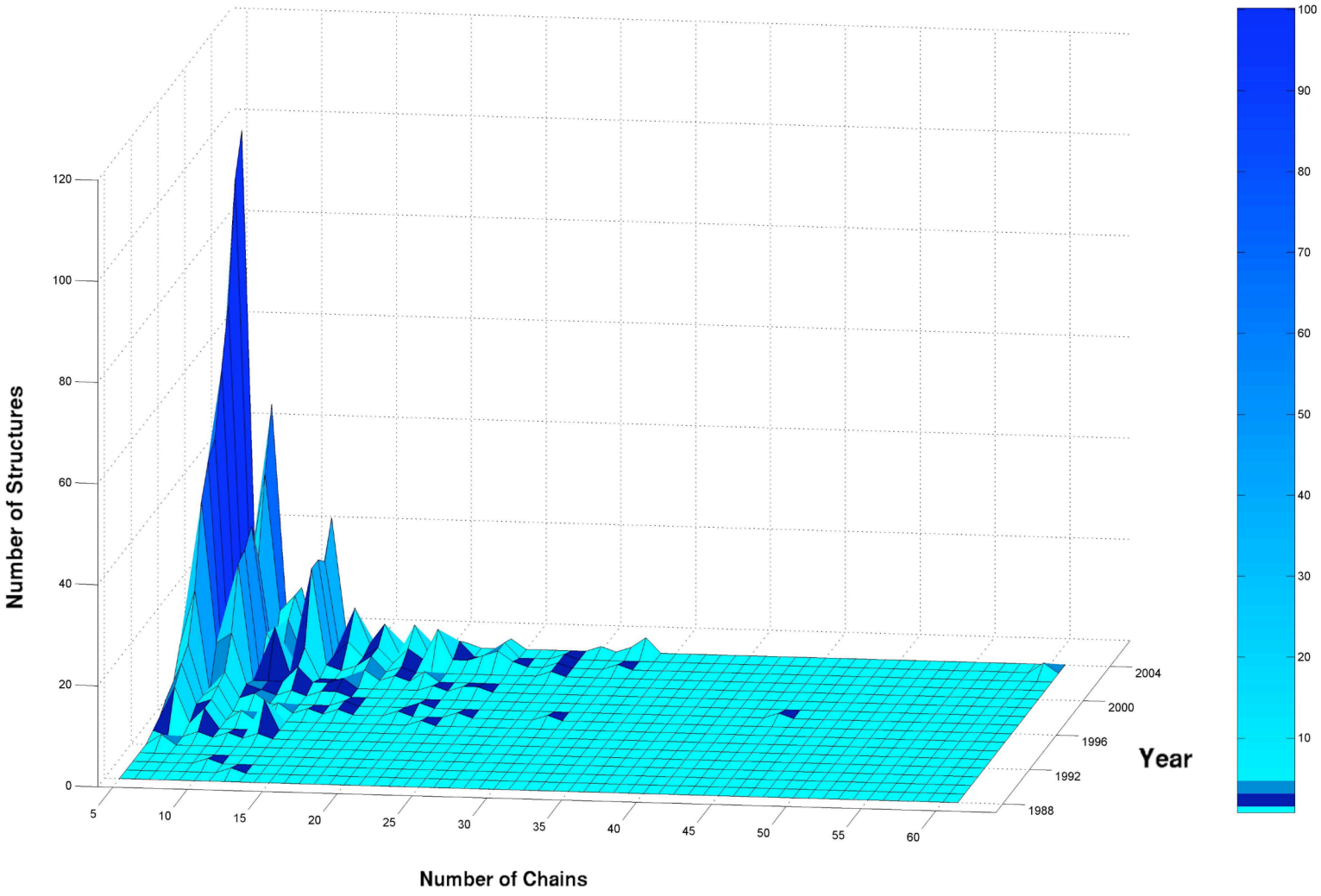


Sali, Earnest, Glaeser, Baumeister. From words to literature in structural proteomics. Nature 422, 216-225, 2003.

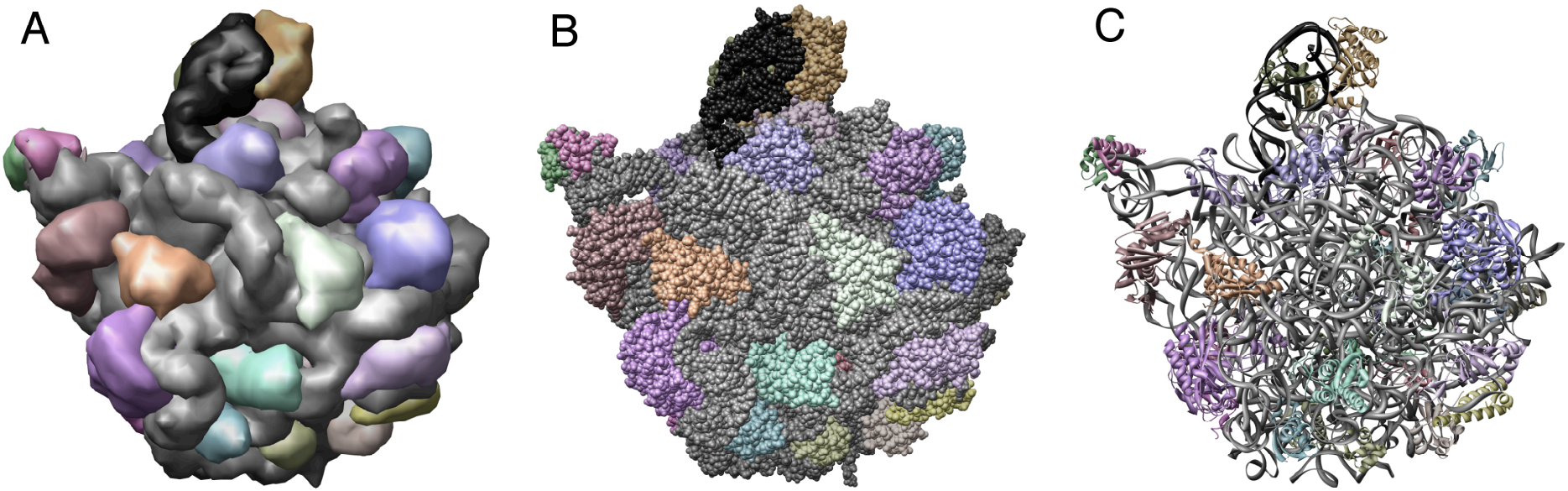
PDB Content Growth



Not just more structures, but increasing complex ones too



Effective visualization requires careful attention to user needs, computer capabilities, computer-human interaction, psychology…

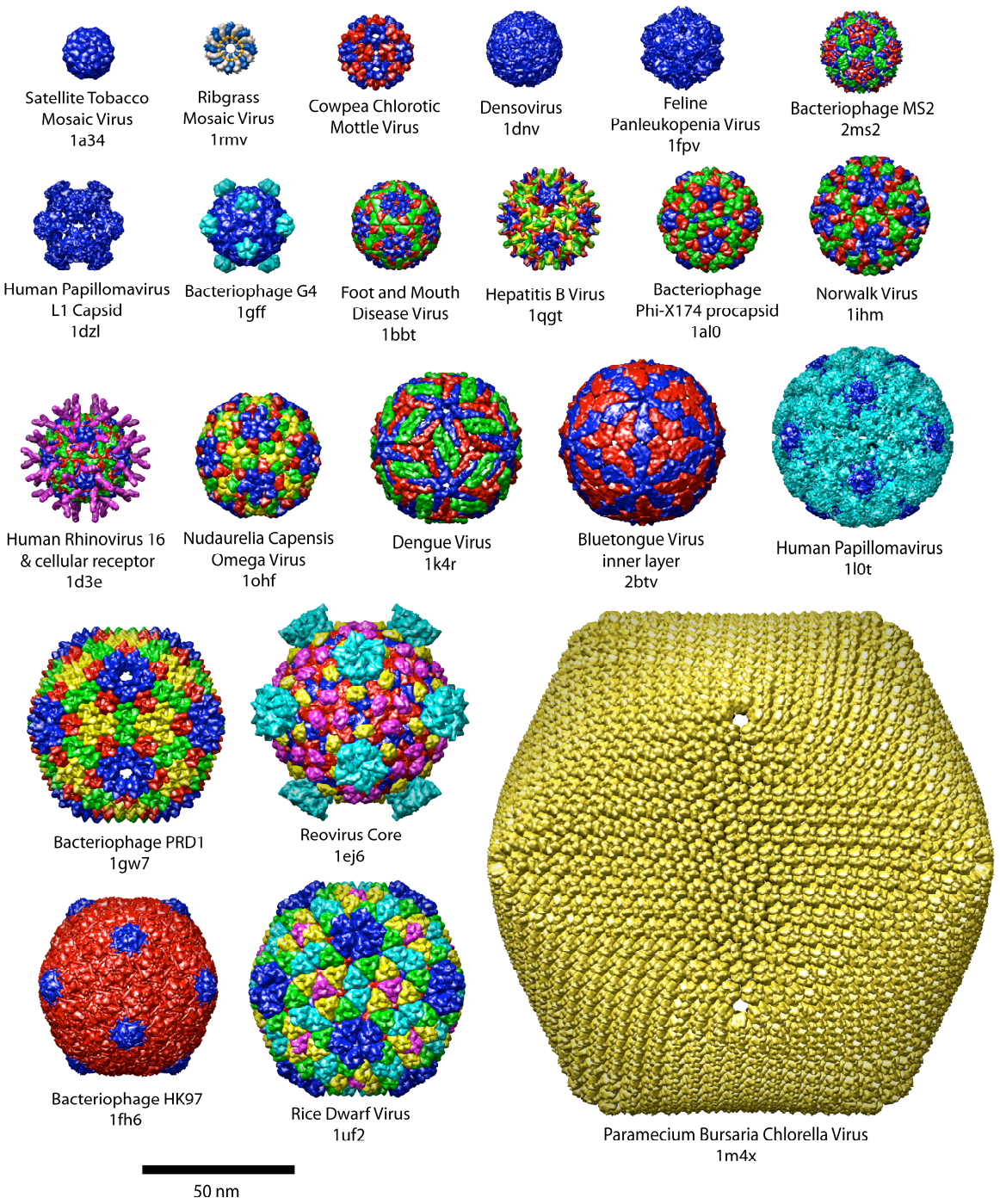


Protein-RNA Contacts in the Large Ribosomal Subunit

A - Low resolution surface depictions of 27 proteins contacting 23S RNA (gray) and 5S RNA (black)

B - Atoms as spheres - pebbly surface reduces effectiveness of 3-D lighting cues

C - Ribbon display style - detailed protein-RNA interactions clear only if view restricted to small pieces of the assembly


Satellite Tobacco Mosaic Virus 1a34
Ribgrass Mosaic Virus 1rmv
Cowpea Chlorotic Mottle Virus
Densovirus 1dnv
Feline Panleukopenia Virus 1fpv
Bacteriophage MS2 2ms2
Human Papillomavirus L1 Capsid 1dzl
Bacteriophage G4 1gff
Foot and Mouth Disease Virus 1bbt
Hepatitis B Virus 1qgt
Bacteriophage Phi-X174 procapsid 1al0
Norwalk Virus 1ihm
Human Rhinovirus 16 & cellular receptor 1d3e
Nudaurelia Capensis Omega Virus 1ohf
Dengue Virus 1k4r
Bluetongue Virus inner layer 2btv
Human Papillomavirus 1l0t
Bacteriophage PRD1 1gw7
Reovirus Core 1ej6
Bacteriophage HK97 1fh6
Rice Dwarf Virus 1uf2
Paramecium Bursaria Chlorella Virus 1m4x
50 nm

UCSF Chimera – an Extensible Molecular Modeling System

Chimera is an extensible interactive 3-D modeling system designed to allow developers to quickly incorporate novel visualization algorithms and analysis tools

Chimera runs on laptops/desktops and takes maximum advantage of low-cost, state-of-the-art graphics chips

- $500 today buys you 3-D interactive graphics capabilities that cost $20,000 five years ago
- Platforms supported: Windows 98/2000/XP/Vista, Mac OS X, Linux, SGI, HP Alpha

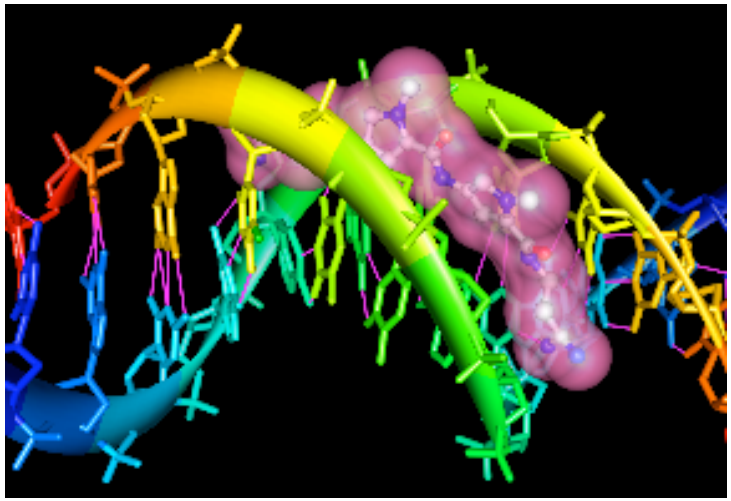
Chimera has extensive documentation for users and developers to enable effective scientific studies to be accomplished rapidly and with a "low entry barrier"

Available from our Research Resource Center web site after simple "click to accept" license agreement

Chimera's Built-in Features

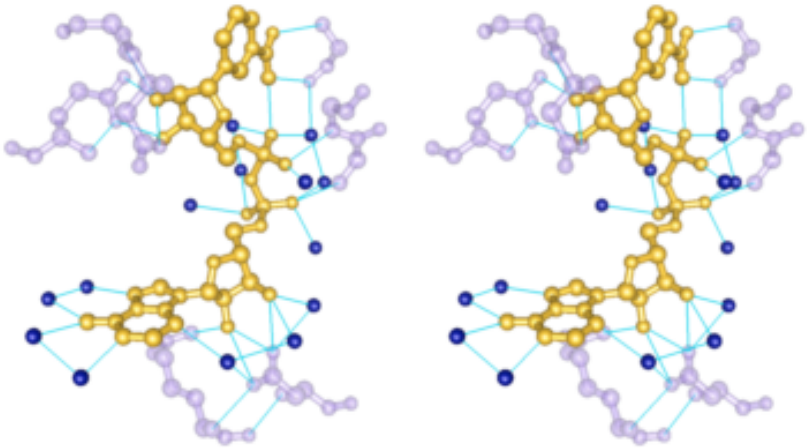
Molecular Graphics:

- interactively manipulate stick, ball-and-stick, CPK, ribbon representations, and molecular surfaces
- highly intuitive model translation, scaling, and rotation
- interactive color editing
- ability to save high resolution images for presentation and publication
- stereo viewing



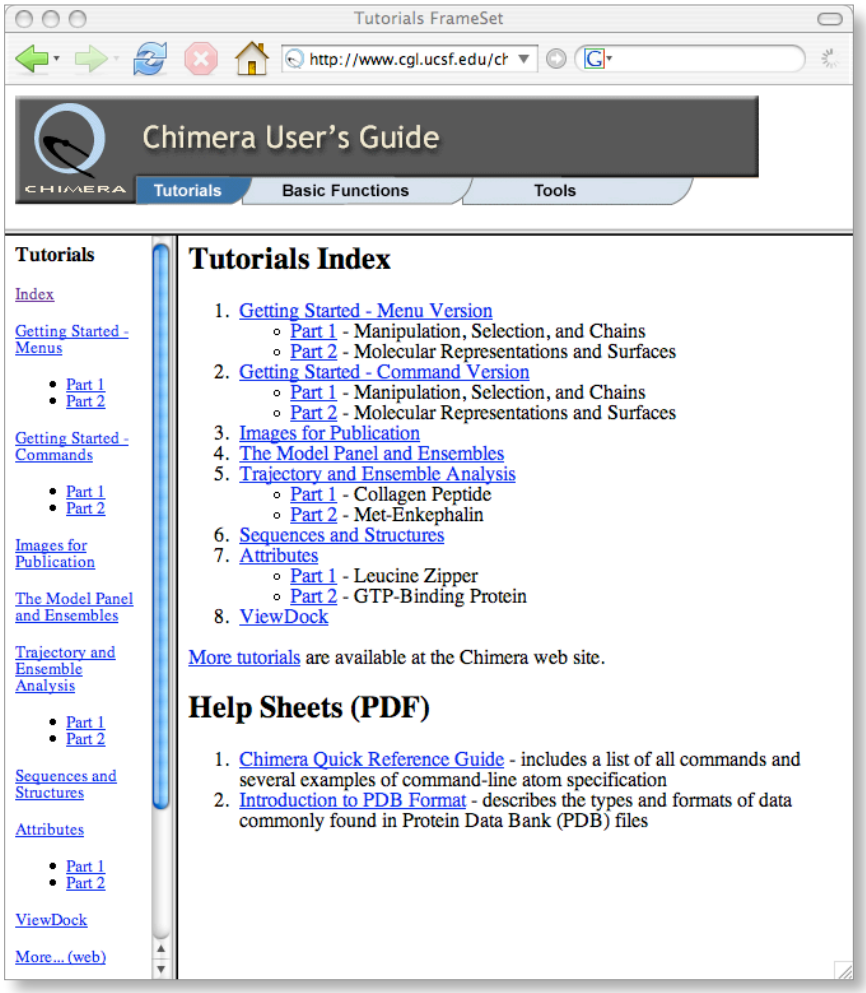
Chemical Knowledge:

- determination of atom types in arbitrary molecules
- ability to add hydrogen atoms
- high-quality hydrogen bond identification
- selection of atoms/bonds by element, atom type, functional group, and amino acid category
- interactive bond rotation, distance, and angle measurements



Extensive Documentation

- User's guide
  - Authoritative description of all Chimera functionality
- Tutorials
  - Overview of basic features for displaying and manipulating structures so beginners can get up-to-speed quickly
- Self-running demos
  - Allow you to sit back and watch any of several modeling session scenarios
- Periodic workshops
  - One- or two-day events with lectures & hands-on training sessions



Additional information: UCSF Chimera - A Visualization System for Exploratory Research and Analysis, *J. Comp. Chem.*, 25(13):1605-12, 2004.

Users need high quality software, well tested and well documented – or else…

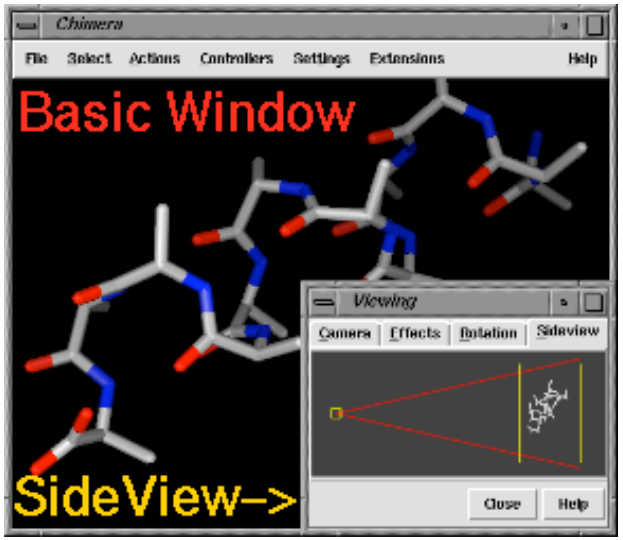
But such software is very time consuming to develop, often 10x the time the initial version requires*

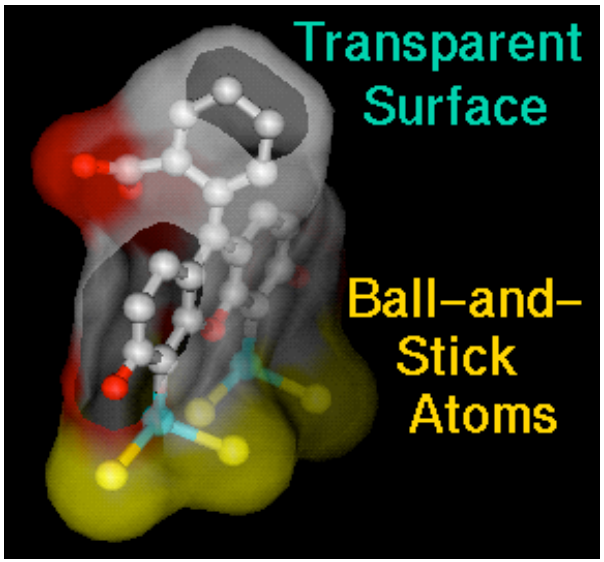
* "The Mythical Man-Month: Essays on Software Engineering" by Frederick P. Brooks, 1975

Chimera's Programmability/Extensibility

Chimera is designed to allow developers to quickly incorporate novel algorithms and analysis tools

- Extensions can control standard Chimera user interface features (e.g. camera, help, menus, toolbar), as well as create their own custom graphical user interfaces
- Extensions are written in the Python programming language
  - Python is easy to learn, even for novice programmers
  - Python is object-oriented and provides features needed for development of complex codes
  - ~30 extensions written to date

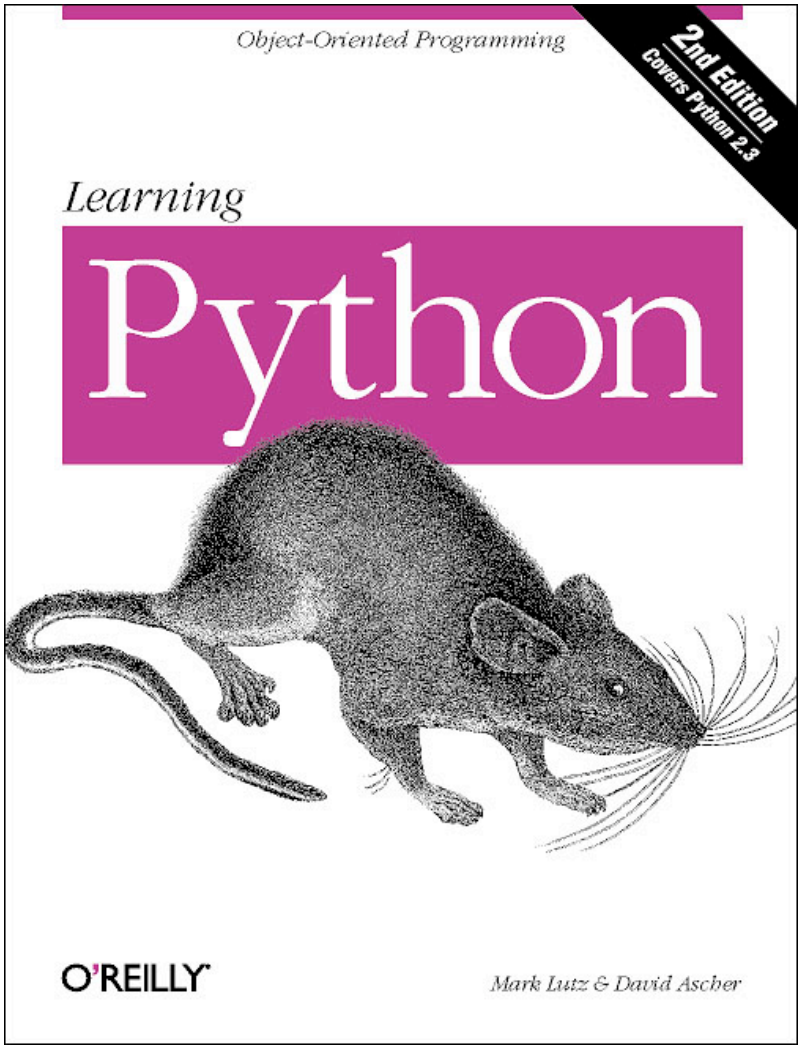




Learn more about Python

Python Web site:
www.python.org

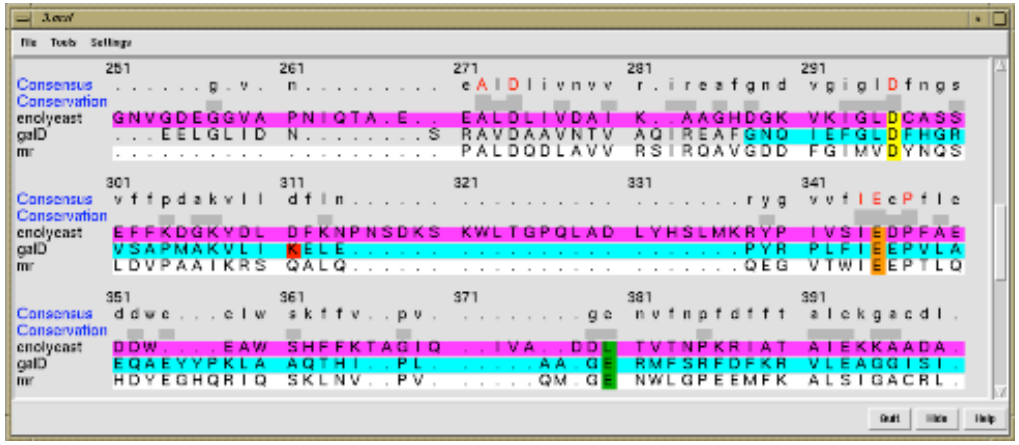
Good book:
"Learning Python"
by Mark Lutz & David Ascher
Available from
www.oreilly.com/catalog/lpython2

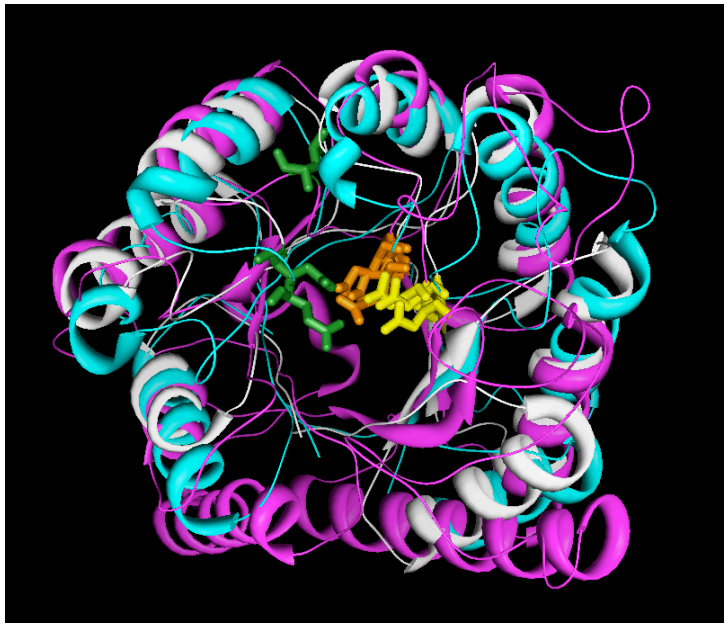


Sample Chimera Extension

Multalign Viewer

- Simultaneously displays multiple protein sequence alignments and corresponding structure superpositions, calculates and displays consensus sequence and conservation histogram, and highlights corresponding regions in both sequence and structure space

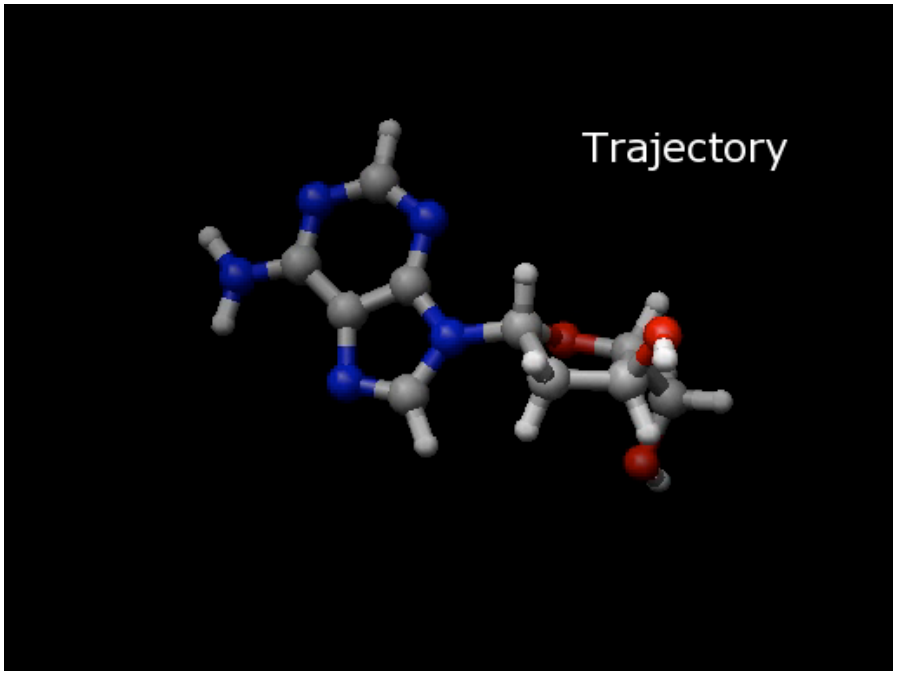




Sample Chimera Extension

Molecular Dynamics Trajectories

- All built-in Chimera analysis and display capabilities also work with trajectories. Support is provided for a number of common programs: AMBER, CHARMM, GROMOS, MMTK, NAMD, PDB, and X-PLOR.



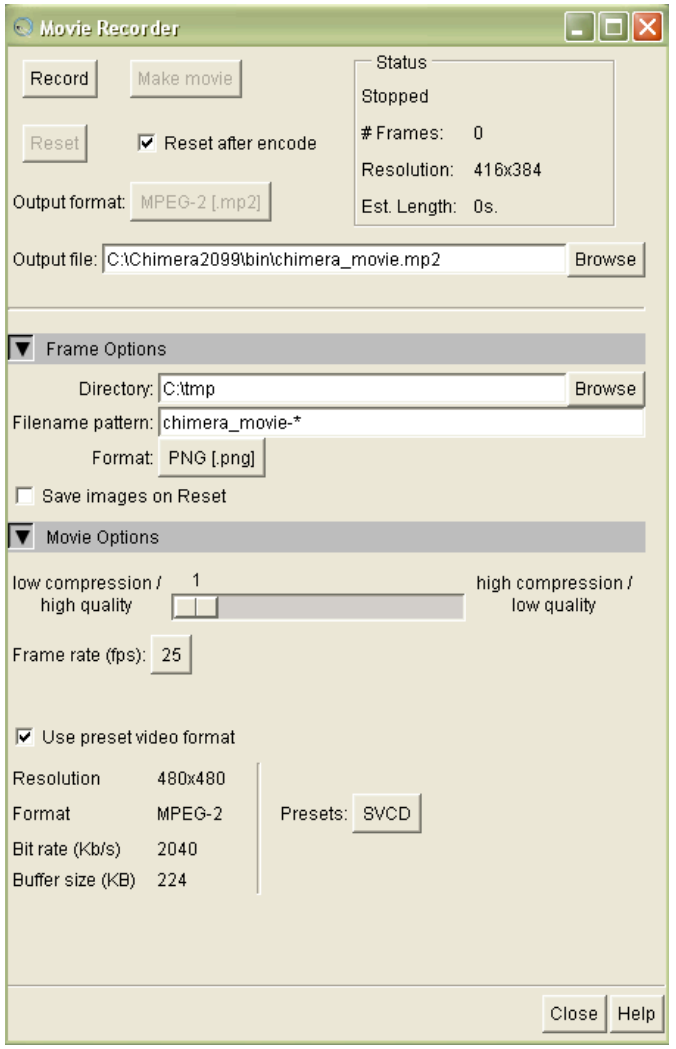
Sample Chimera Extension

Movie Recorder

Capture image frames from Chimera and assemble these into a movie file

Formats supported:

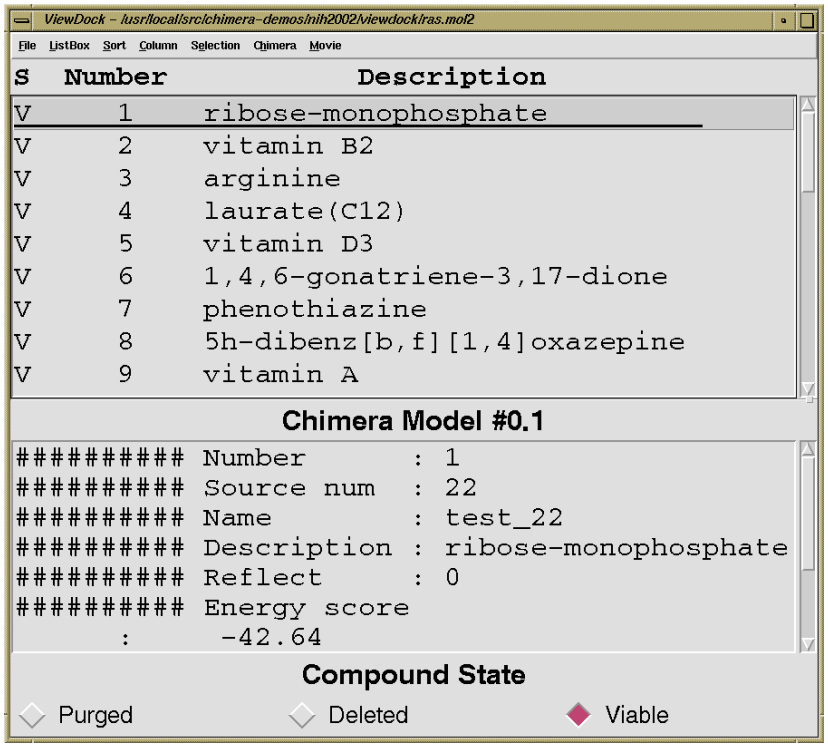
- MPEG-1
- MPEG-2
- MPEG-4
- Quicktime

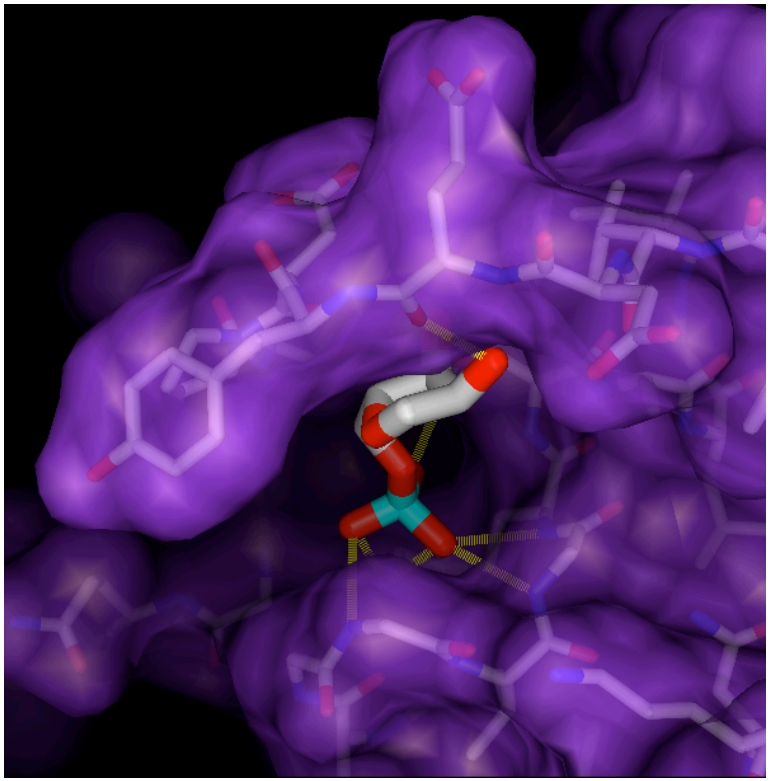


Sample Chimera Extension

ViewDock

- Facilitates selection of promising drug candidates found with the UCSF DOCK program

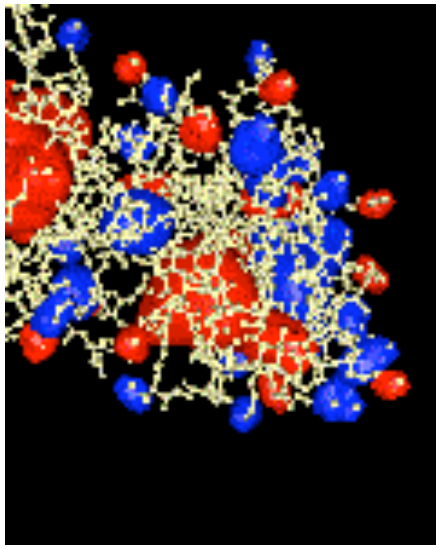


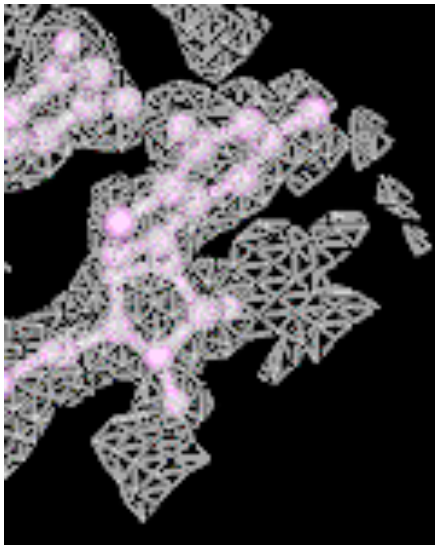


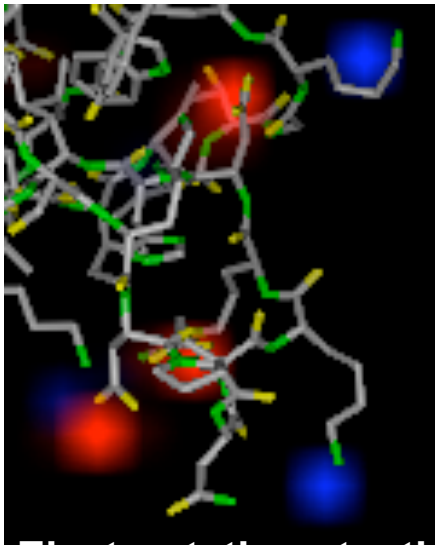
Sample Chimera Extension

Volume Viewer

- An extension for visualizing three-dimensional numerical data sets such as x-ray or cyro-EM density maps



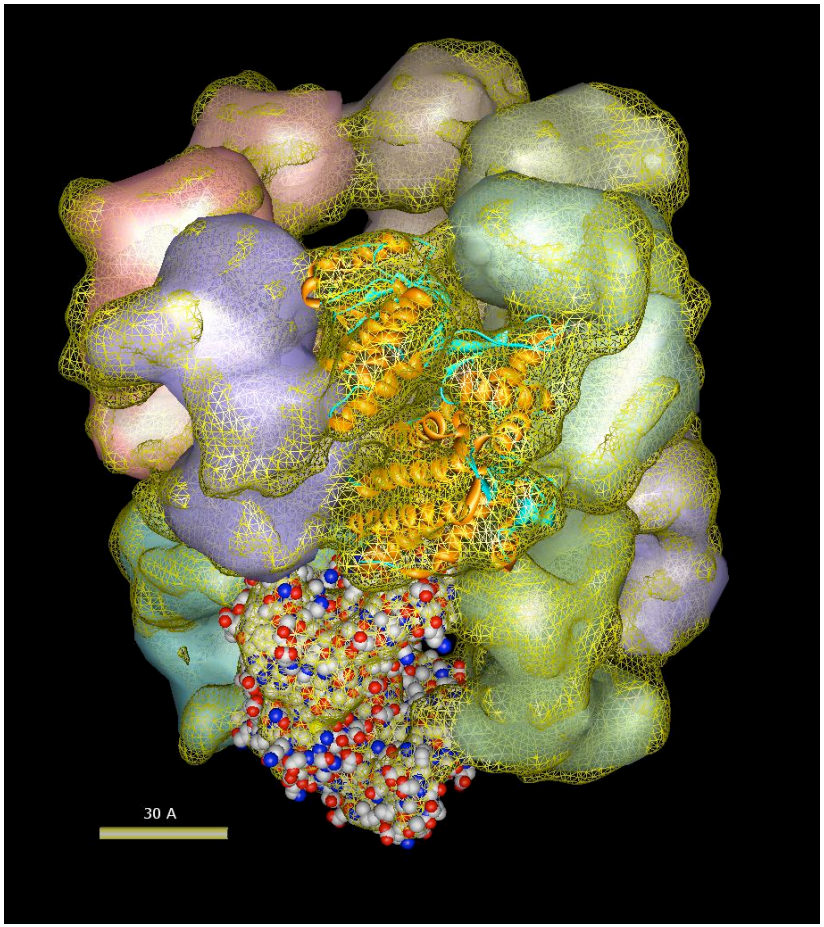




Sample Chimera Extension

Multiscale Modeling

- useful for exploring models of large molecular complexes
- combines volume visualization and atomic resolution capabilities
- example systems include viruses and chromosomes
- GroEL model: 14 copies of the monomeric GroEL crystal structure docked to a 10.3 Å electron microscope map

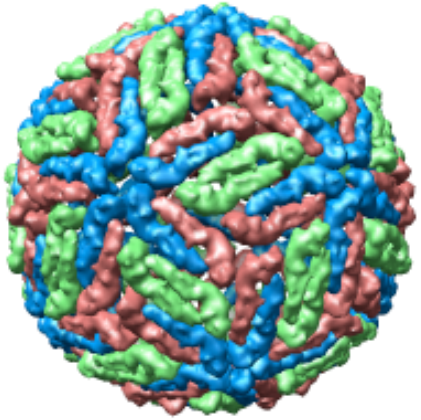


Additional information: Software Extensions to UCSF Chimera for Interactive Visualization of Large Molecular Assemblies, *Structure,* 13(3):473-82, 2005.

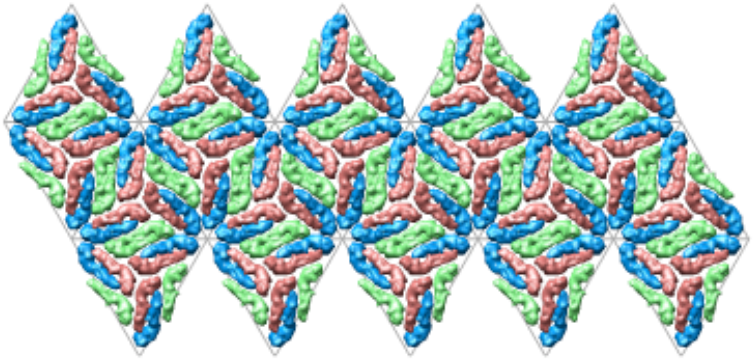
Sample Chimera Extension

Flatten Icosahedron

The 20 triangles making up the icosahedral capsid are laid out in a plane.
The flat view can be printed and folded into a paper icosahedron model.
Construction time: 30 minutes Cost: $1 Extension: 254 lines of Python



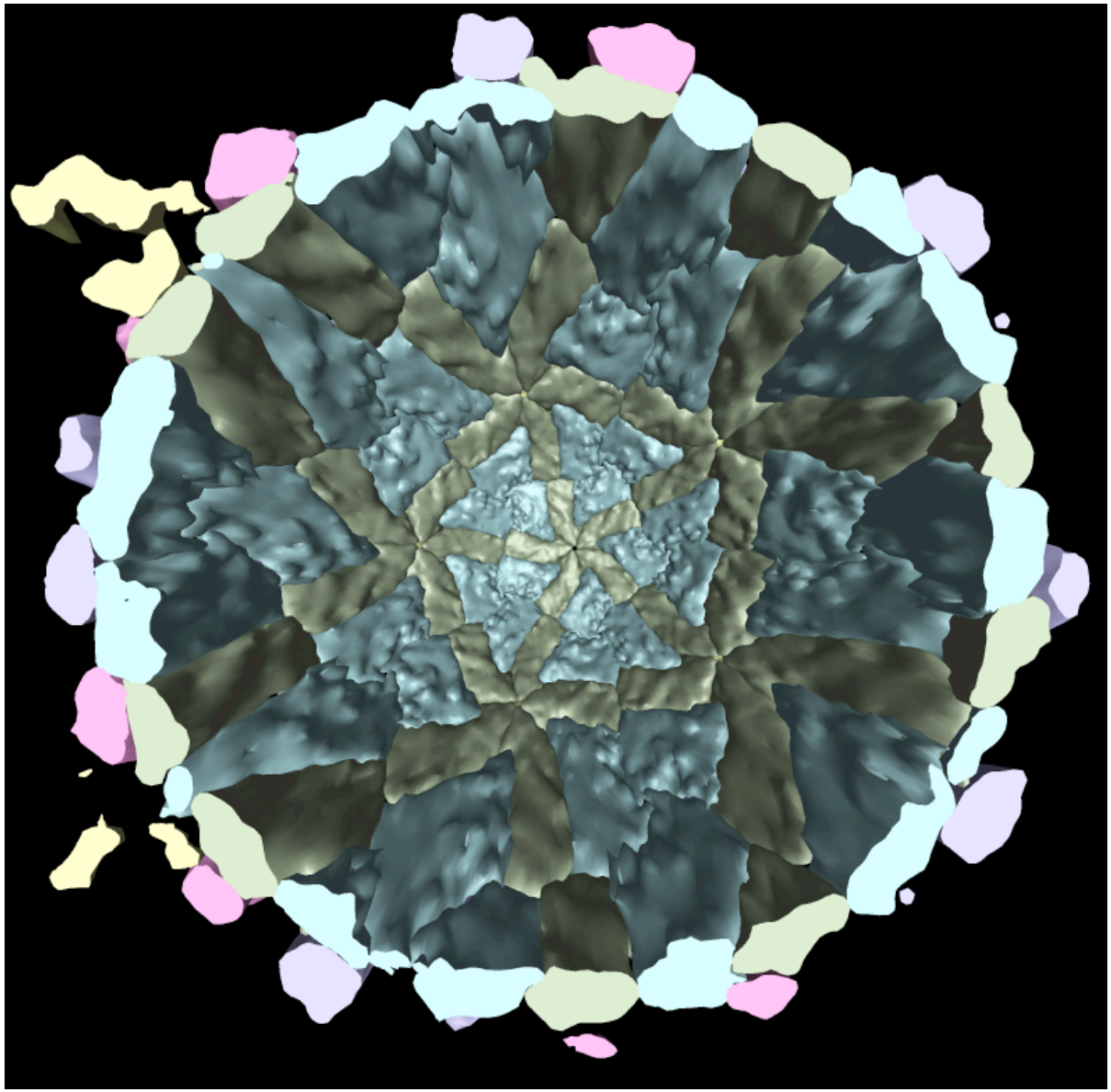
Dengue virus



Flattened

Paper model

Sometimes implementing new ideas just requires access to the Python interpreter layer…



Fish-eye view of reovirus (1ej6).

Chimera Demonstration #1

Introduction to molecular representation and basic use

Files used: 2gbp.pdb

Features illustrated:

- Opening files
- Selecting and displaying atoms, bonds, and surfaces
- Manipulating models: rotate, translate, scale, clip
- Display styles: wireframe, sticks, balls&sticks, CPK
- Command line

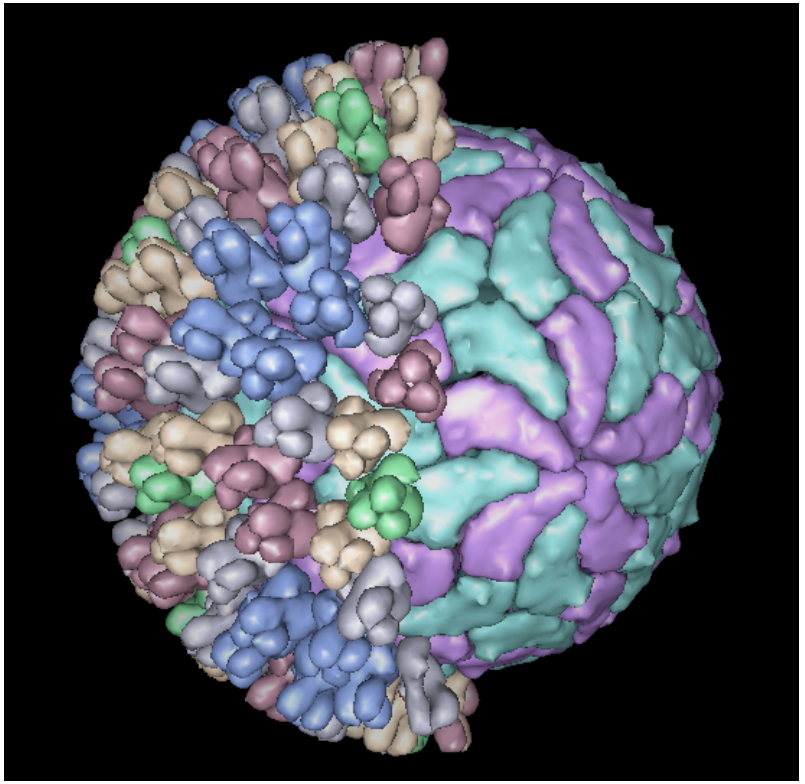
Additional information: UCSF Chimera - A Visualization System for Exploratory Research and Analysis, *J. Comp. Chem.*, 25(13):1605-1612, 2004.

[Introductory Demo]

Chimera Demonstration #2

Bluetongue Virus Core

- PDB structure 2btv by David Stuart's lab (Nature 395: 470-478, 1998)
- Full particle 700 Å diameter, 3.5 Å resolution, 1000 crystals and $3x10^6$ atoms (no hydrogens)
- Multiscale extension makes 60 copies of unit cell
  - –Outer layer: 260 trimers of VP7 protein in 5 symmetry classes
  - –Inner layer: 60 dimers of VP3
- Extension focuses on hierarchical structure relationships and their selection and display



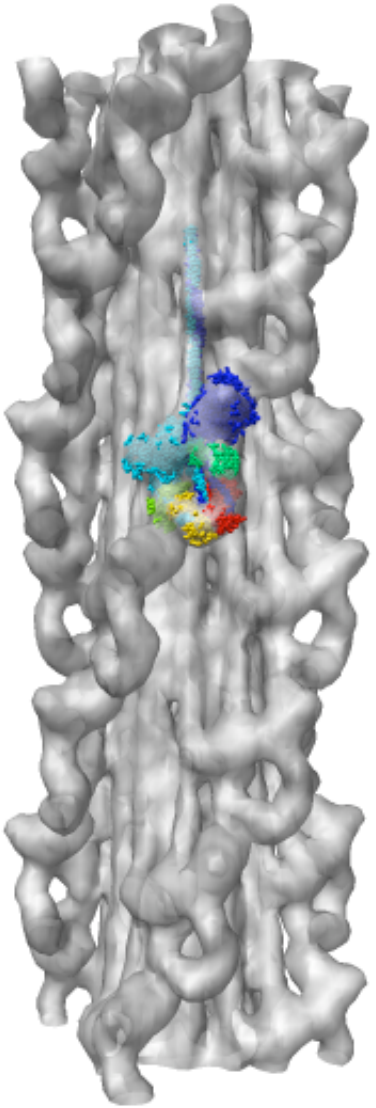
Additional information: Software Extensions to UCSF Chimera for Interactive Visualization of Large Molecular Assemblies, *Structure*, 13(3):473-482, 2005.

[Bluetongue Virus Demo]

Chimera Demonstration #3

Atomic model of a myosin filament:

- Collaboration with Ed Egelman at Univ. of Virginia Health Sciences Center
- Map and model fitting provided by Roger Craig's lab at Univ. of Massachusetts Medical School
- ~2.5nm resolution cryo-EM map 2x better than anything previous
- High res. map allows unambiguous fitting of myosin atomic models
- Fitting reveals intermolecular contacts that may be important for maintaining the relaxed muscle state



Additional information: J.L. Woodhead *et al.*, *Nature,* 436:1195-9, August 2005.

[Myosin Fitting Demo]

Summary

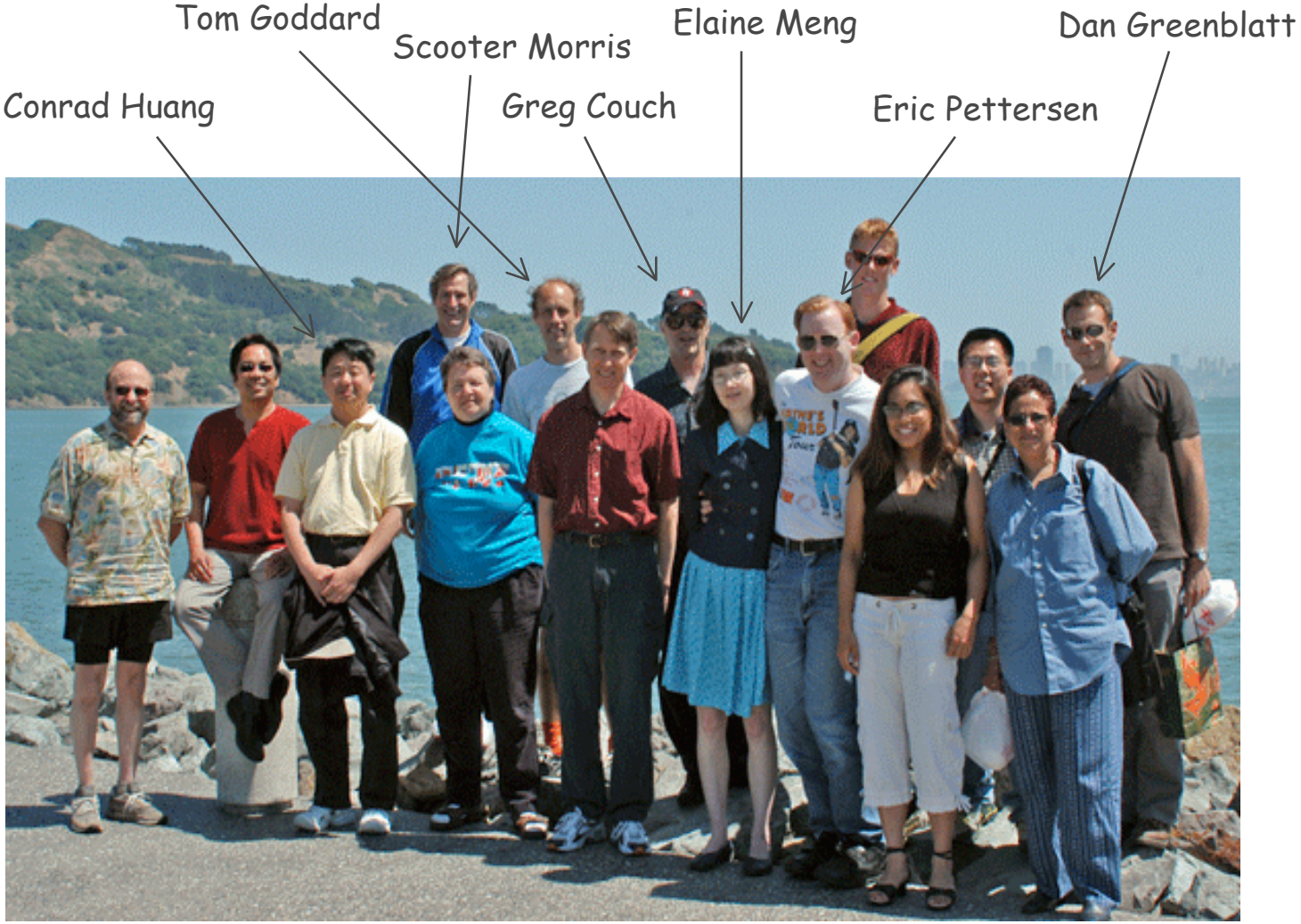
The volume and complexity of available biological data have grown enormously in recent years, requiring increasingly sophisticated visualization and analysis tools in order to understand the underlying life processes.

Effective visualization, especially interactive visualization, must leverage the latest advances in computing and computer graphics.

Enabling others to easily extend the functionality of software results in greater innovation and productivity than a single individual or group can provide alone.

It is much more time consuming (and therefore expensive) to create robust, well-documented, and easy to use software than most people think.

Chimera development team



Funding: NIH NCRR (P41 RR01081)