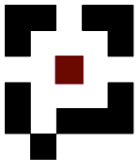
# Python-based Tools and Web Services for Structural Bioinformatics

Randy Heiland, Charles Moad
*IU Pervasive Technology Labs*
Sean Mooney, *IU School of Medicine*
{heiland,cmoad}@indiana.edu
sdmooney@iupui.edu

pervasivetechnologylabs
AT INDIANA UNIVERSITY

# Outline

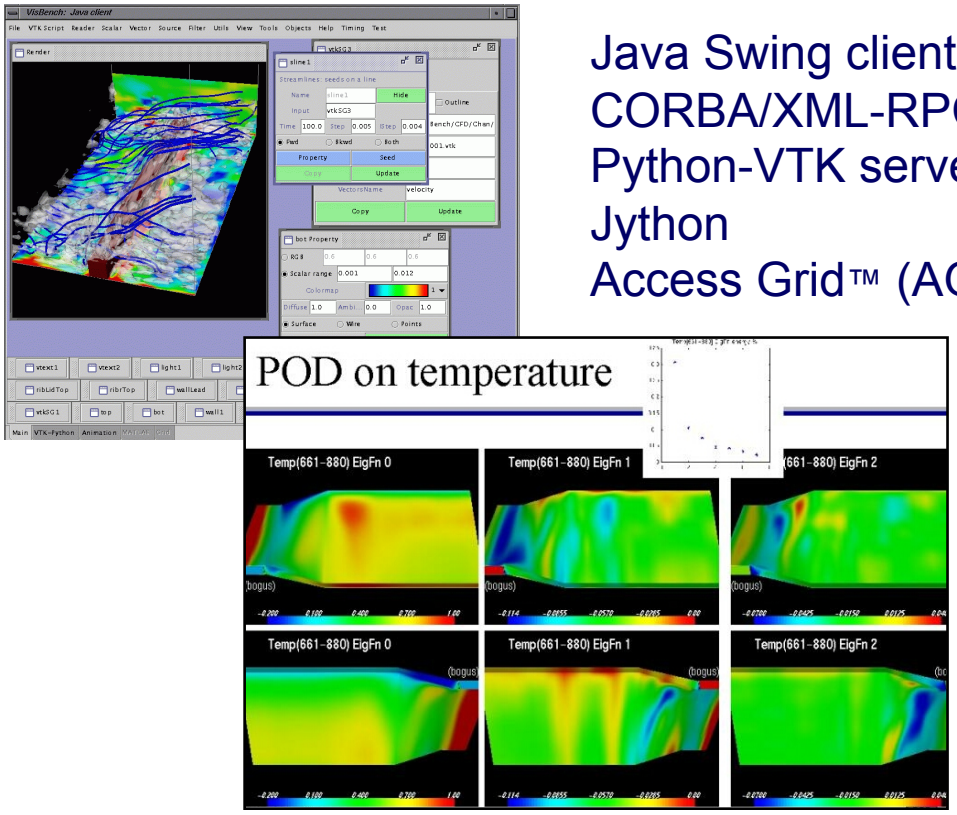
- Past Python-related work (at NCSA, no proteomics)
- Indiana University/IUPUI: Pervasive Tech Labs, Center for Computational Biology and Bioinformatics
- Intro to Structural Bioinformatics
- Tools/Services for Mutation Data
  - Vis tools (UCSF Chimera, PyMOL)
  - Web Services (Axis, Pywebsvcs/SOAPpy)
- Future work


pervasivetechnologylabs
AT INDIANA UNIVERSITY

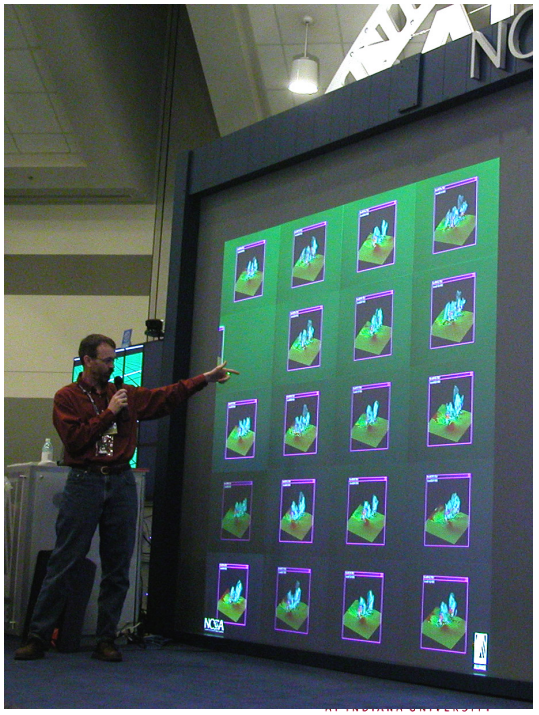
# Past Python-related work

(RH at UIUC/NCSA ’97-’03)

- Python-wrapped VTK [+ pyMPI] for [cluster-based] SciVis
- VisBench project: client-server vis & analysis




Java Swing client
CORBA/XML-RPC
Python-VTK server
Jython
Access Grid™ (AG2)



# Indiana University; IUPUI

- Pervasive Technology Labs at IU – six labs pervasive.iu.edu (~1999), sda.iu.edu (2003)
  - Help grow the IT economy in Indiana via collaborations in academia and industry

- Center for Computational Biology and Bioinformatics - Mooney Lab compbio.iupui.edu/mooney (2003)
  - Characterize the structural elements that enable protein function
  - Understand the effects of genomic variation on the proteome

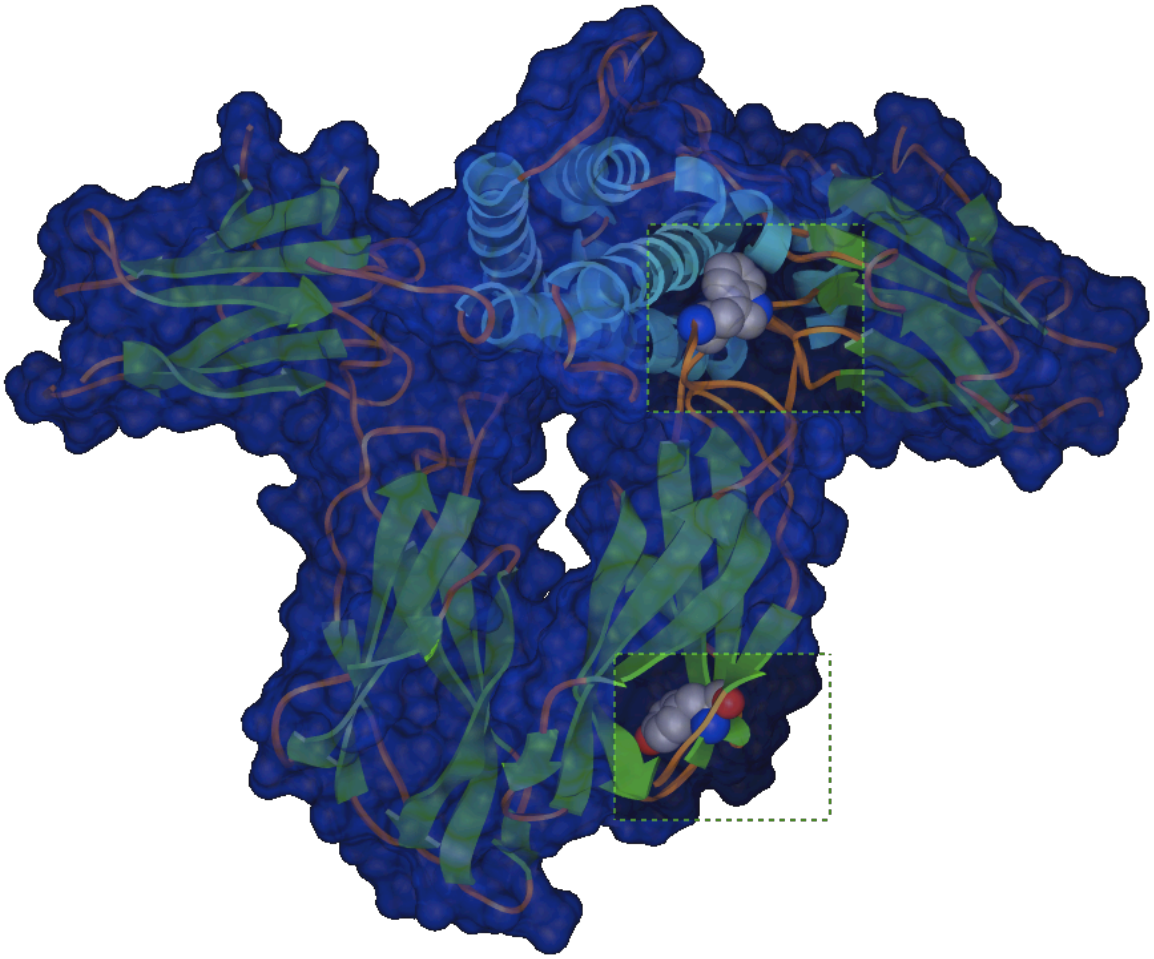
pervasivetechnologylabs
AT INDIANA UNIVERSITY

# Some terminology

- Cell contains genome = complete set of DNA
- DNA = sequence of ATCG nucleotides
- Genes = specific seqs that encode instructions for making proteins
- Protein = molecules of (20) amino acids that perform much of life's function
- Proteome = set of all proteins in a cell
- Proteomics = study of protein's structure & function
- Bioinformatics = Biology + CS + IT

# Intro to Structural Bioinformatics

- Protein Data Bank now contains more than 26,000 structures
- Annotation of structural data is a challenging and relevant problem



pervasivetechnologylabs
AT INDIANA UNIVERSITY

# Protein visualization tools

- UCSF Chimera (www.cgl.ucsf.edu/chimera)
- PyMOL (pymol.sourceforge.net)

- Python-based tools for interactive visualization of protein 3-D structure (& 1-D sequence)
- Each provides a Python-based API for writing extensions


pervasivetechnologylabs
AT INDIANA UNIVERSITY

# Web Services

- Any service that is:
  - available over the Internet
  - uses XML messaging
  - independent of OS & pgming lng
- XML messaging:
  - XML-RPC, SOAP, HTTP post/get
- WSDL: Web Svcs Description Lng

For MutDB:

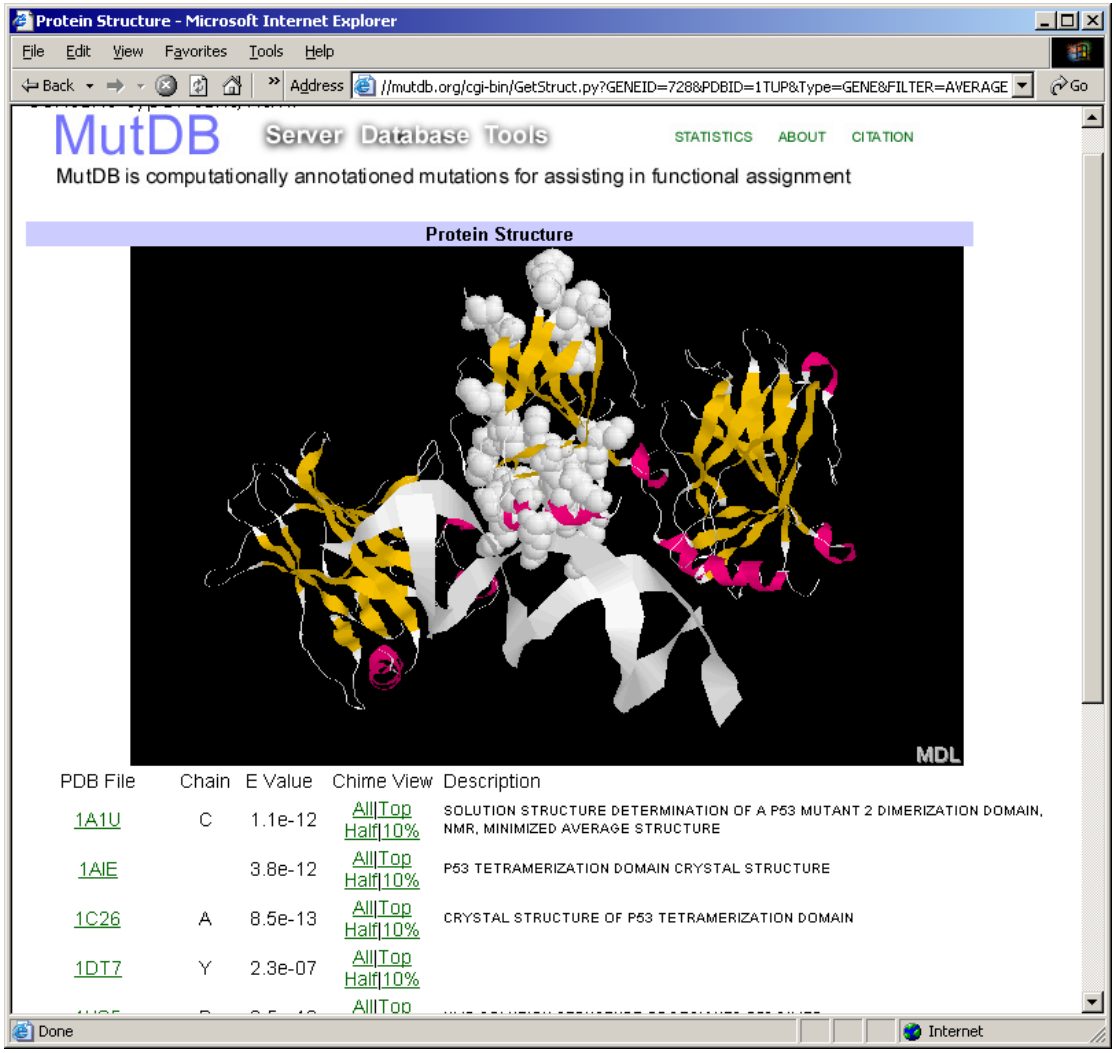
- Apache Axis (ws.apache.org/axis)
- PyWebSvcs/SOAPpy (pywebsvcs.sf.net)


pervasivetechnologylabs
AT INDIANA UNIVERSITY

# Examples of Bio Web Services

- New PDB (pdbbeta.rcsb.org/pdb)
  - alpha.rcsb.org/jboss-net/services/pdbWebService?wsdl
- KEGG (www.genome.jp/kegg/soap)
- biomoby.org
- Google ‘bio web services’

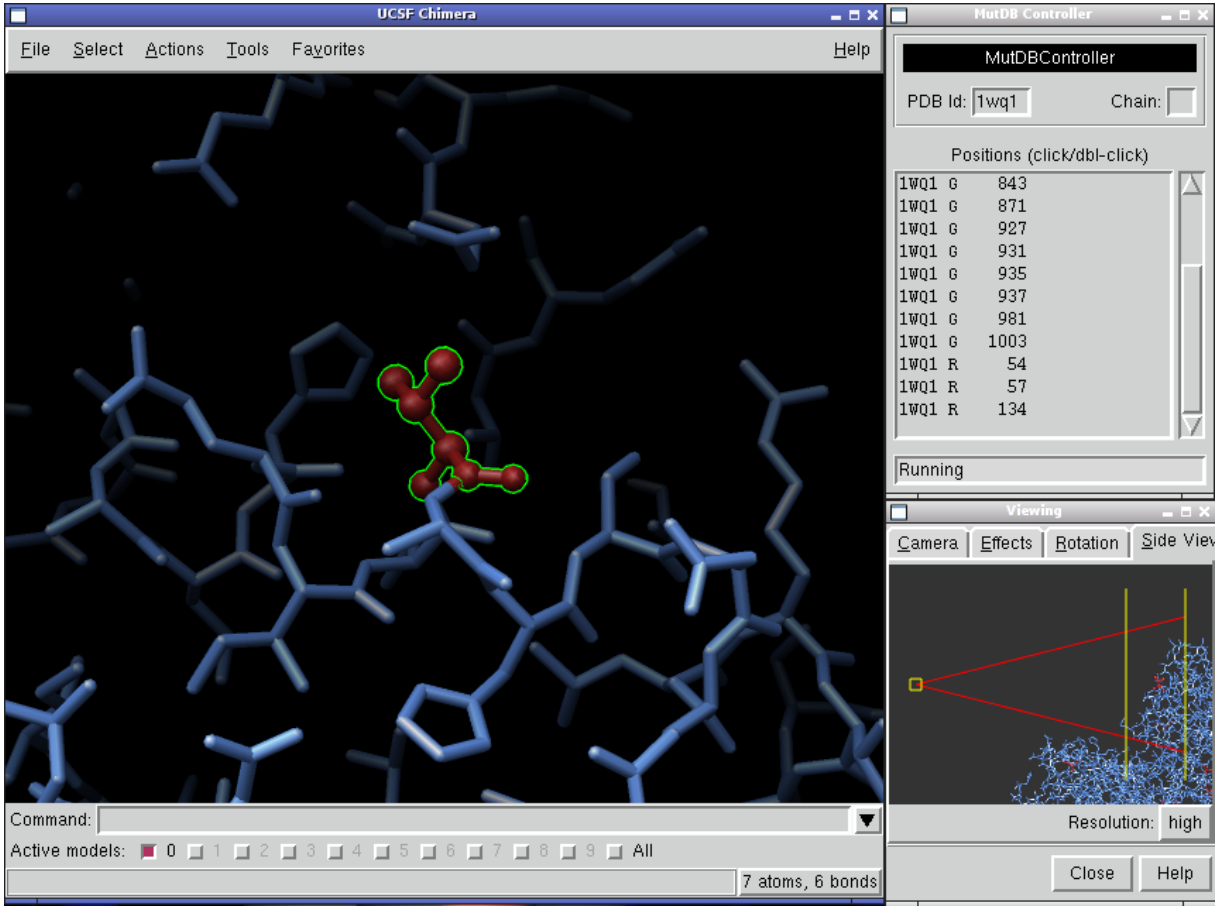
# MutDB (http://www.mutdb.org)




MutDB provides structural annotations for disease-associated mutations and single nucleotide polymorphisms (SNPs)

pervasivetechnologylabs
AT INDIANA UNIVERSITY

# Structural Mutation Service

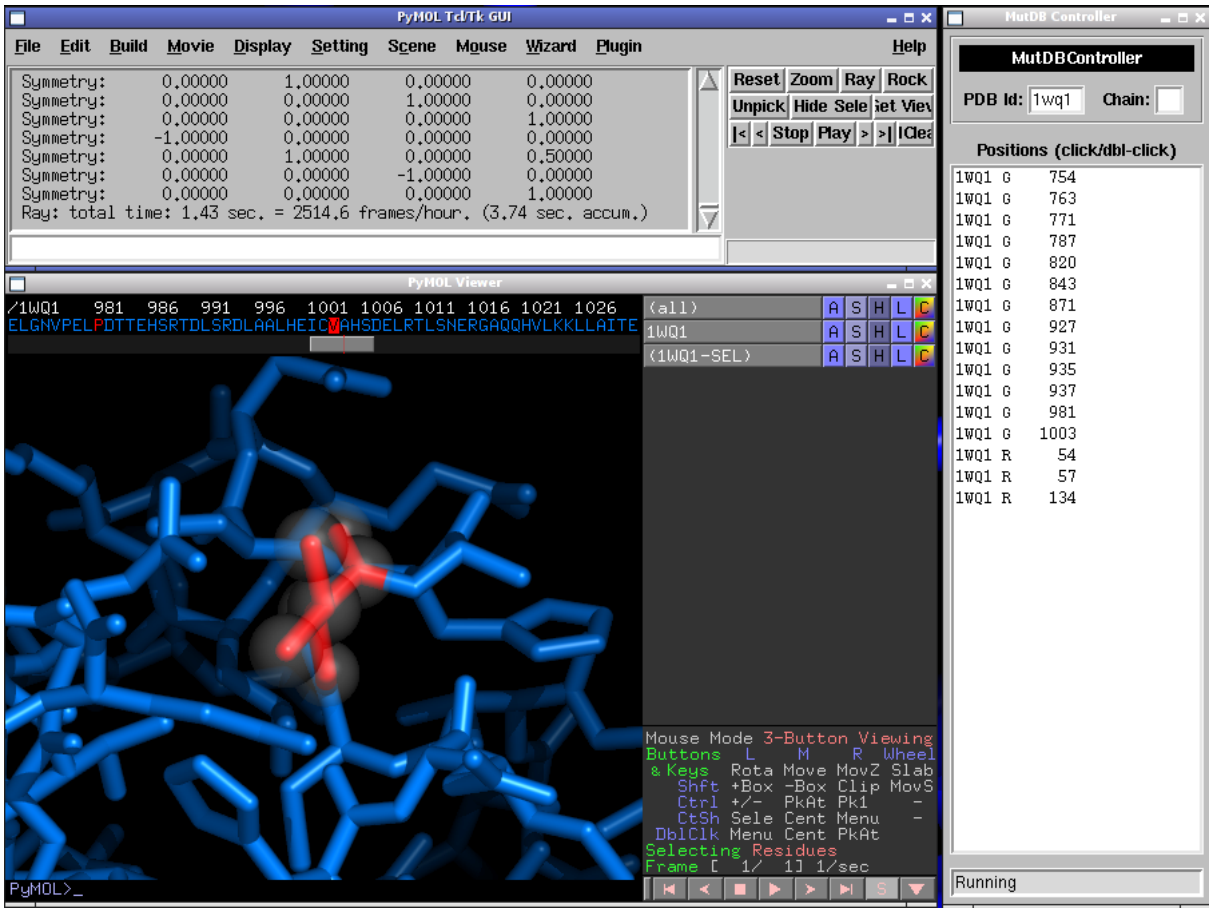



- Mutations on MutDB are mapped to protein structure
- Extension in Chimera queries MutDB

UCSF Chimera extension

pervasivetechnologylabs
AT INDIANA UNIVERSITY

# PyMol Extension

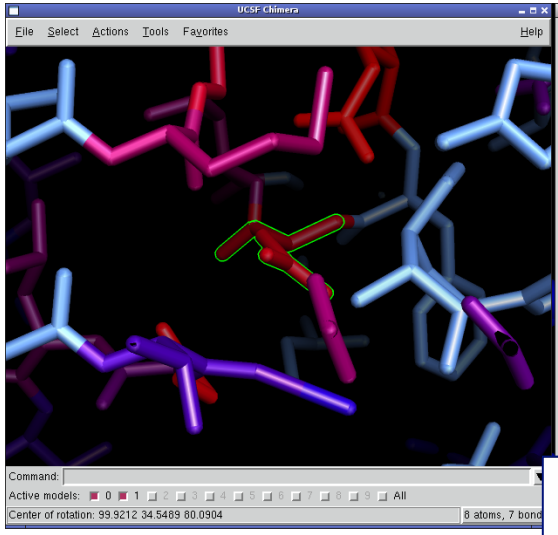



Controller window identifies mapped mutation positions which are highlighted structurally

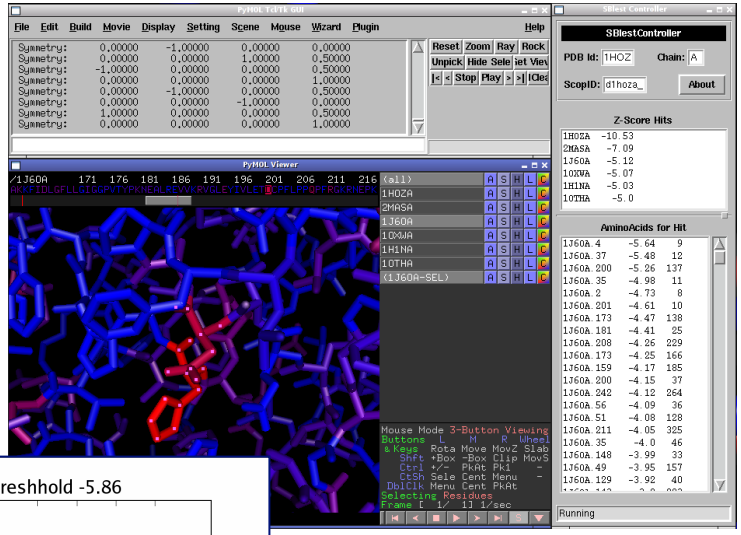
pervasivetechnologylabs
AT INDIANA UNIVERSITY

# Future work

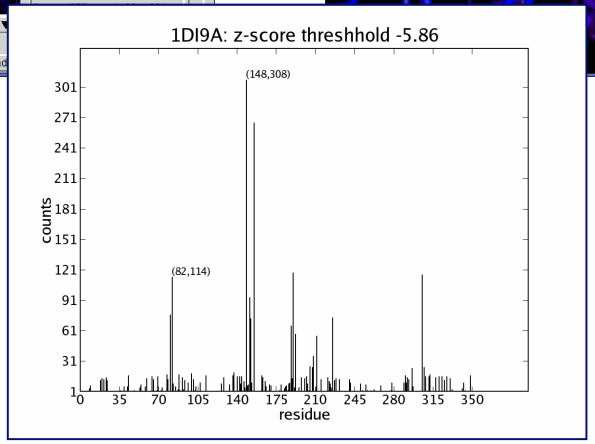
- Web services for identifying regions of structural similarity between a query protein and a database of protein structures




Chimera




PyMOL




matplotlib

# Acknowledgements & Ref

The Indiana Genomics Initiative (INGEN) and the Pervasive Technology Labs of Indiana University are supported in part by Lilly Endowment Inc.

S.D. Mooney and R.B. Altman, “MutDB: annotating human variation with functionally relevant data”. *Bioinformatics. 2003 Sep 22;19(14):1858-1860*


pervasivetechnologylabs
AT INDIANA UNIVERSITY