Integrated Software Tools for Functional Genomics

Thomas Ferrin

Patricia Babbitt

Conrad Huang

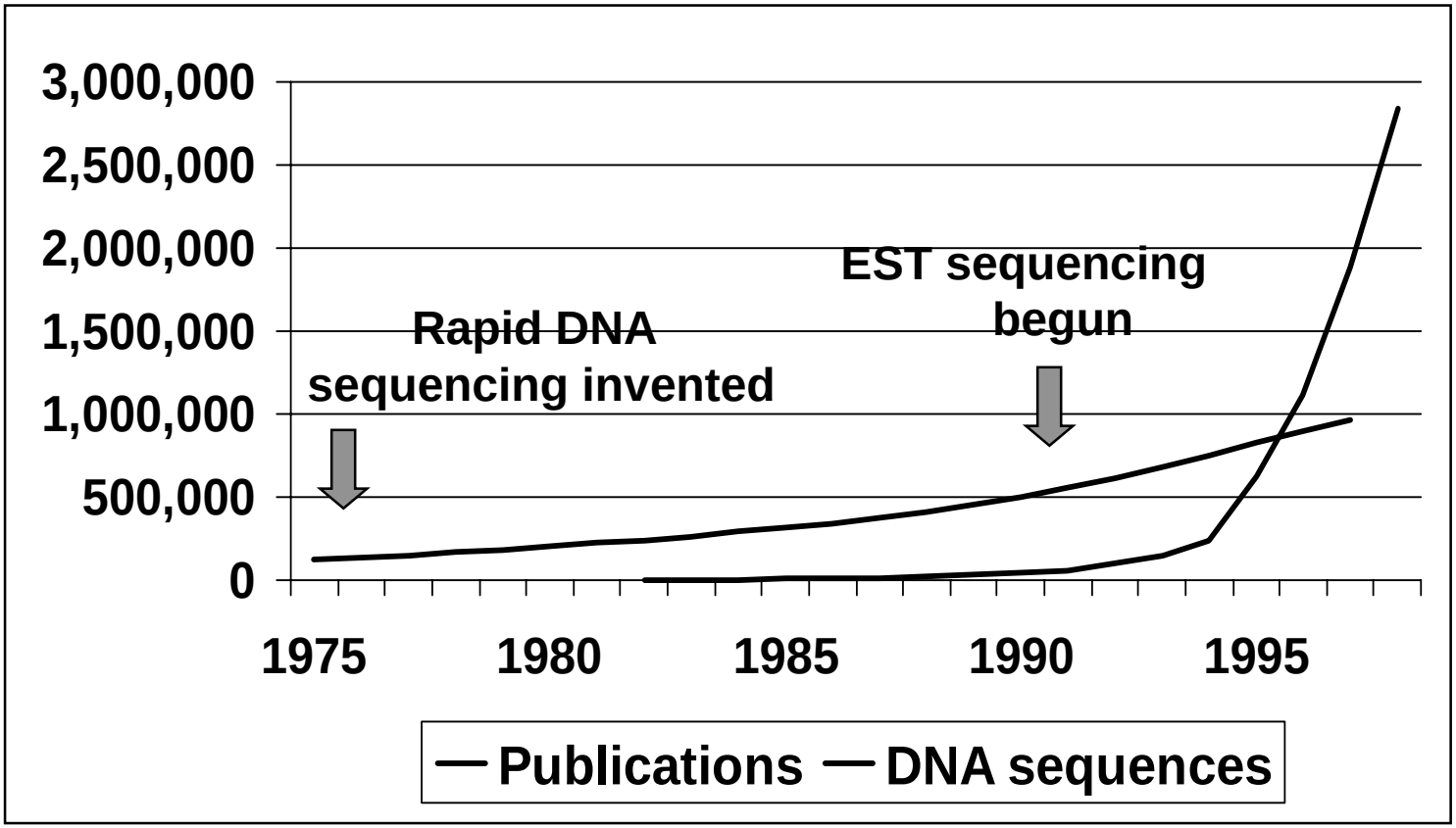
Teri Klein

University of California, San Francisco

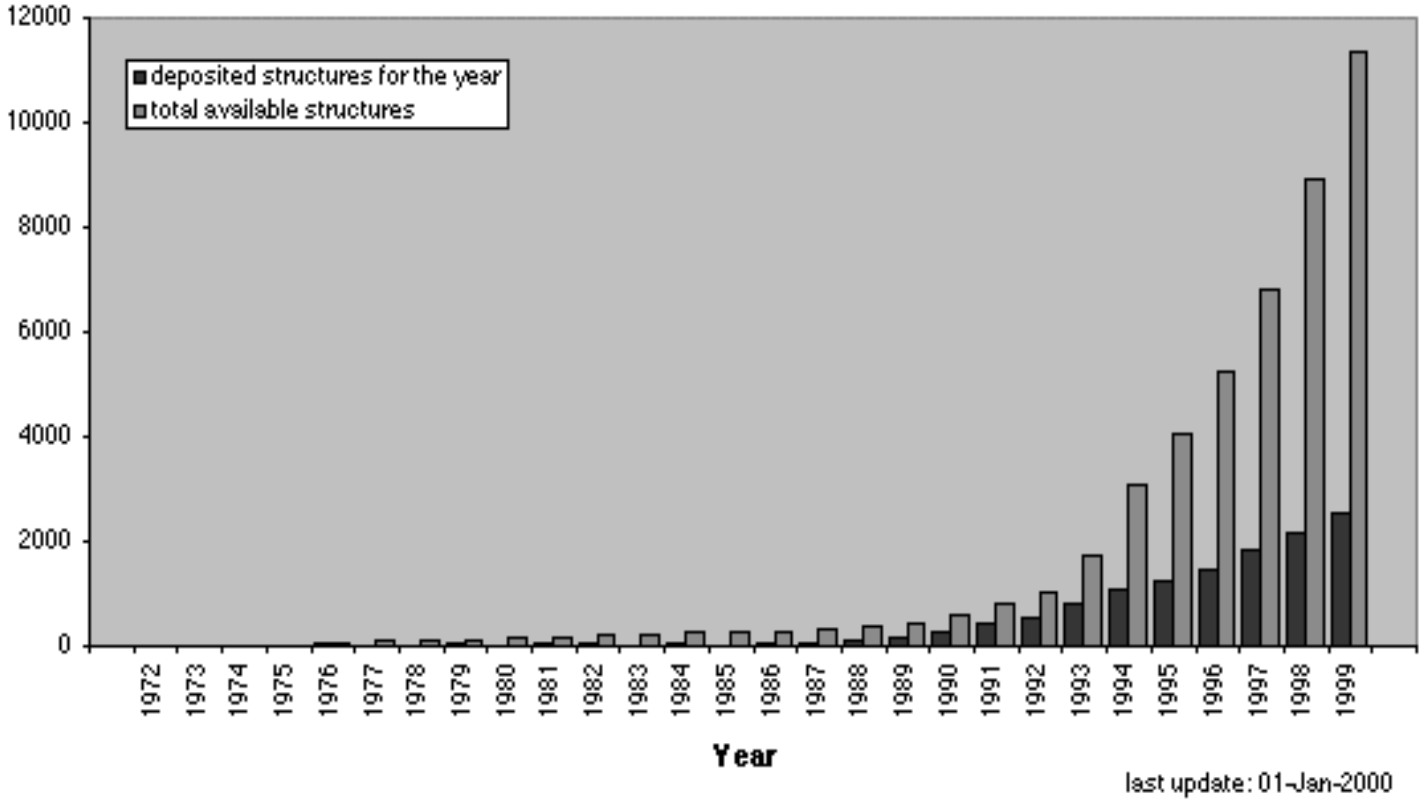
“It’s sink or swim as a tidal wave of data approaches”

Tony Reichhardt
Nature 399:517-520 10 June 1999

The Growing Gap in Functional Knowledge



Growth in Protein Structures

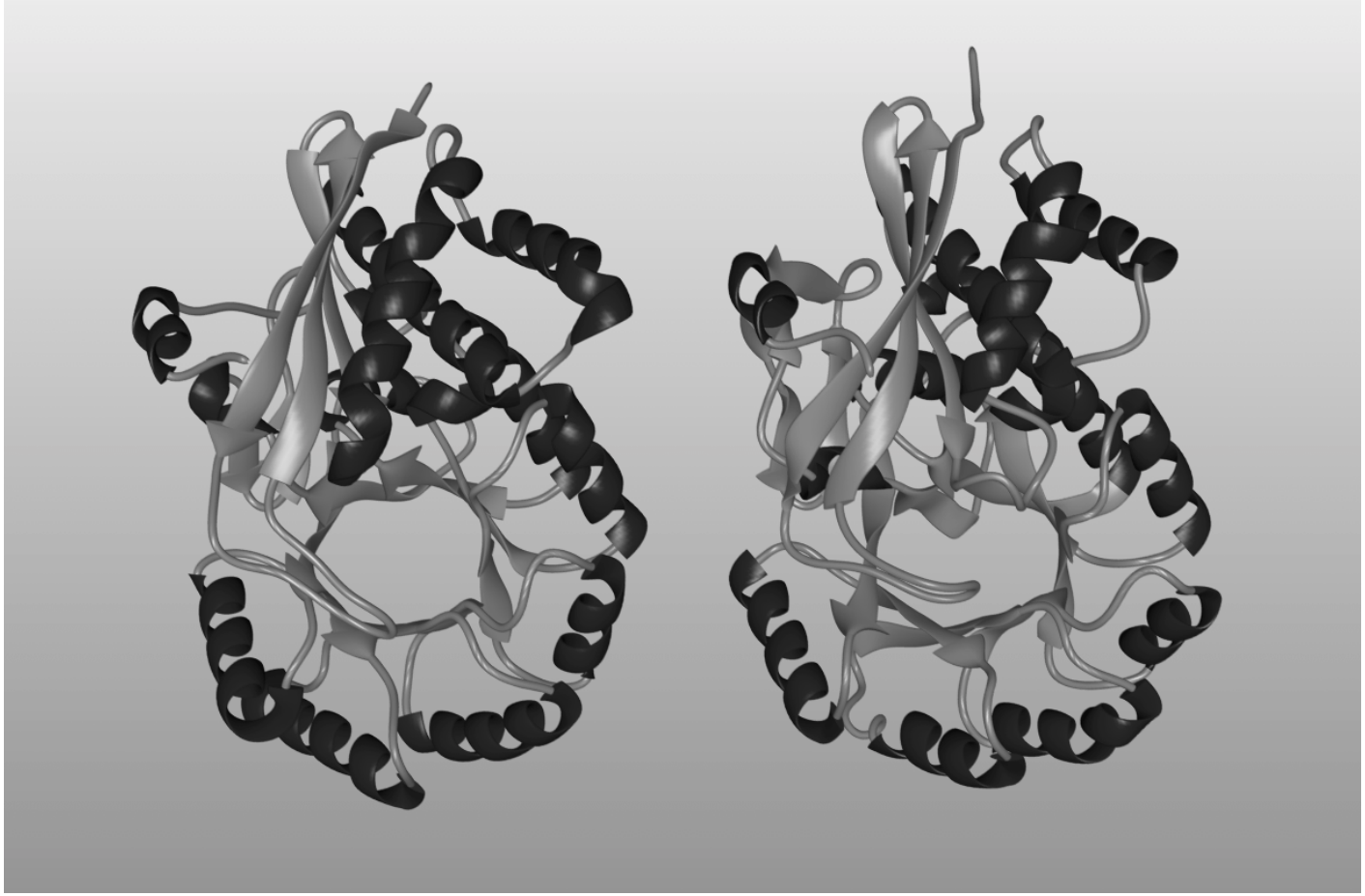


Sequence -> Structure -> Function

Challenges:

- Prediction of structure from sequence
- Prediction of function from sequence
- Understanding of evolutionary changes
- Engineering of specialized function
- Applications in pharmacogenomics and pharmacogenetics
- . . .

Stereo pairs ?



Tools for Comparative Protein Studies

MinRMS - exhaustive search for all plausible structural alignments of two proteins

AlignPlot - interactive exploration of structural alignments

MSFviewer - integrates sequence and structure space

Chimera - extensible 3-D molecular modeling system

MinRMS

Find all plausible alignments between two protein structures (experimentally-determined or modeled) using root-mean-square difference of coordinates of alpha-carbons.

- **RMSD metric easy to interpret**
- **Avoids “single best alignment” problem**
- **Avoids need for parameters**
- **Finds reasonable alignments even for apparently dis-similar structures**

MinRMS Algorithm

Two step process:

- **1. Rotate & translate the two structures to bring similarly shaped regions into close proximity;**
- **2. With the two proteins fixed at a particular relative position, select corresponding alpha-carbon atoms between the proteins which minimizes the intermolecular RMSD.**

Apply a dynamic programming algorithm to find best matches for different numbers of amino acid residues

Algorithm runs in O(n^5) time

- **For two 300-residue proteins requires ~1 hour on a fast workstation**

MinRMS Output

Large table containing, for each structure alignment:

- **Number of matched residues**
- **RMSD for the alignment**
- **Longest distance between any pair of matched residues**
- **Levitt & Gerstein similarity score, -log(P)**
- **Transformation matrix for aligning the structures**

AlignPlot

Used to examine MinRMS output for alignments of interest

- RMSD vs. Number of matched residue pairs
  - Useful for examining trade-off between number of matched residues and global superposition
- Orientation clusters
  - Reduces hundreds of alignments into a few representative groups
- Sequence vs. sequence histogram
  - Provides easy identification of patterns such as secondary structure

MSFviewer

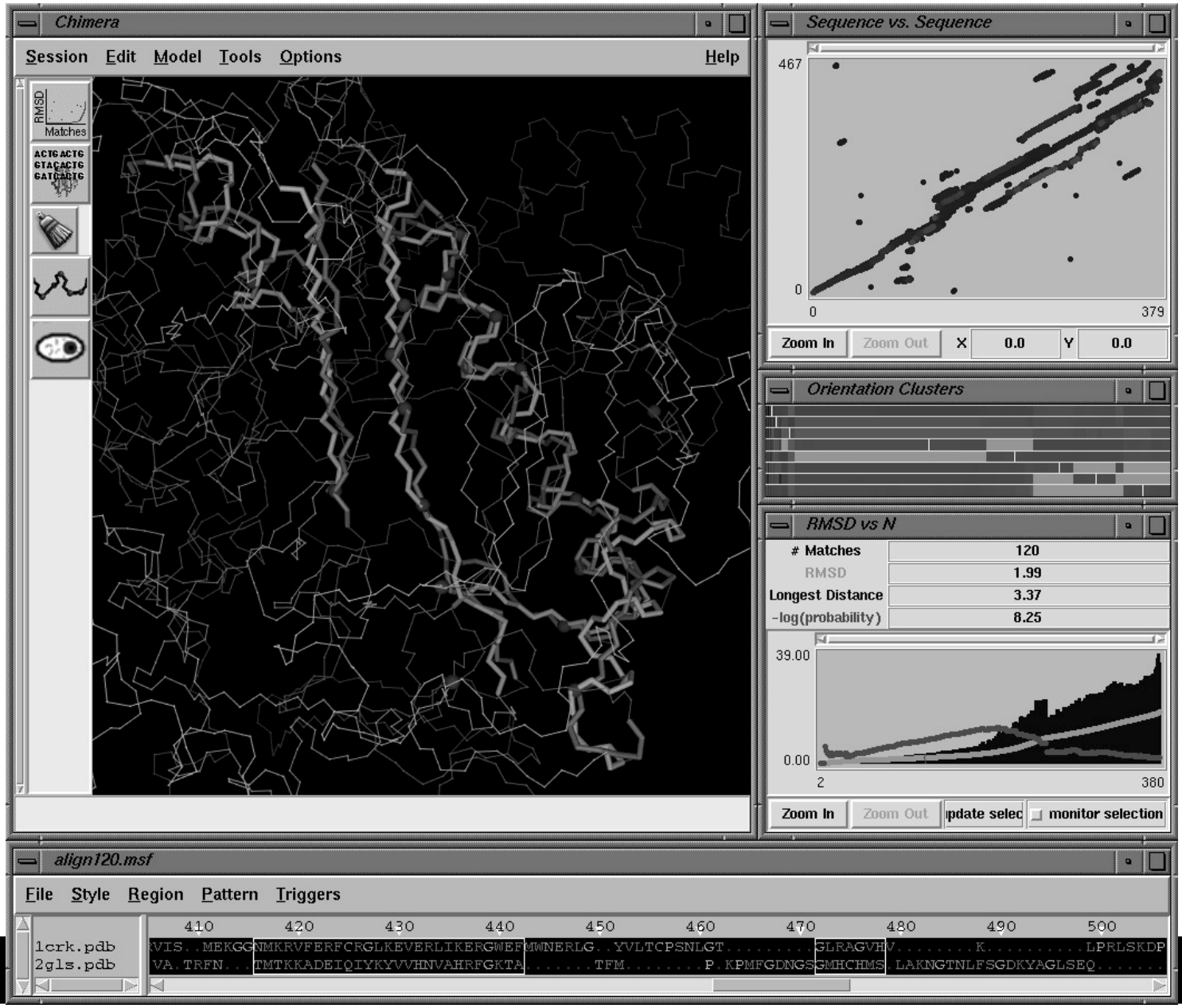
Displays multiple sequence alignments from common alignment programs

- Groups of residues in the alignment can be selected
- Corresponding residues in the structure also get highlighted
- Allows user to facile interface to sequence space

Chimera

Molecular visualization system providing:

- Interactive manipulation of multiple molecular structures
- Real-time rendering of models in several formats
  - e.g. ball-and-stick, ribbons, molecular surfaces
- Support for non-molecular objects
  - e.g. points, vectors, markers, spheres, cylinders, polygons
- Command line compatibility with MidasPlus
- Extensible functionality without access to source code
- Use of standard APIs ensure portability to many platforms
  - Windows 95/98/NT/2000, Compaq, SGI, Linux, …


Chimera
Session Edit Model Tools Options Help
RMSD
Matches
Sequence vs. Sequence
467
0
0
379
Zoom In
Zoom Out
X
0.0
Y
0.0
Orientation Clusters
RMSD vs N
Matches
120
RMSD
1.99
Longest Distance
3.37
-log(probability)
8.25
39.00
0.00
2
380
Zoom In
Zoom Out
pdate selec
monitor selection
align120.msf
File Style Region Pattern Triggers
410 420 430 440 450 460 470 480 490 500
1crk.pdb
2gls.pdb

Chimera's Extensibility

Use of Python programming language as Chimera's command language provides for both complex command "scripts" and user-written extensions

- **True programming language allows for user commands to contain such constructs as iterative loops and conditional execution with full access to internal data structures**
- **Widely available Python libraries provide for custom GUIs**
  - **e.g. menus, dialog boxes, custom graphics**
- **Python's interpreted language provides for dynamic run-time linking**
  - **Don't need access to source code to add new features**
  - **New modules "linked in" when Chimera executes**

Chimera Extensions

Extensions are just groups of one or more cooperating processes

- AlignPlot, MSFviewer, MidasPlus Command Interpreter are all implemented as extensions
- Extensions can maintain their won state and have their own graphical user interface
- Extensions can be ancillary to Chimera or Chimera can be invoked by another program to provide interactive graphical output

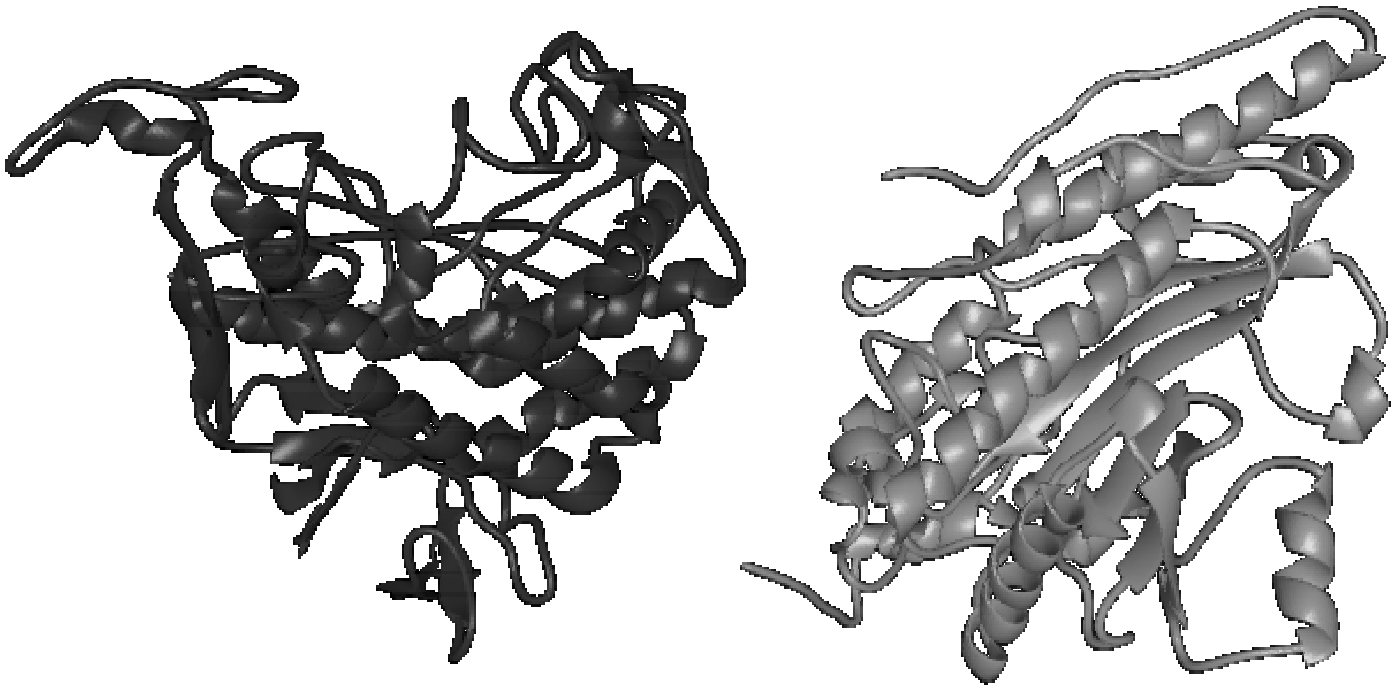
Example Study

Structural comparison of glutamine synthetase (GS) and creatine kinase (CK)

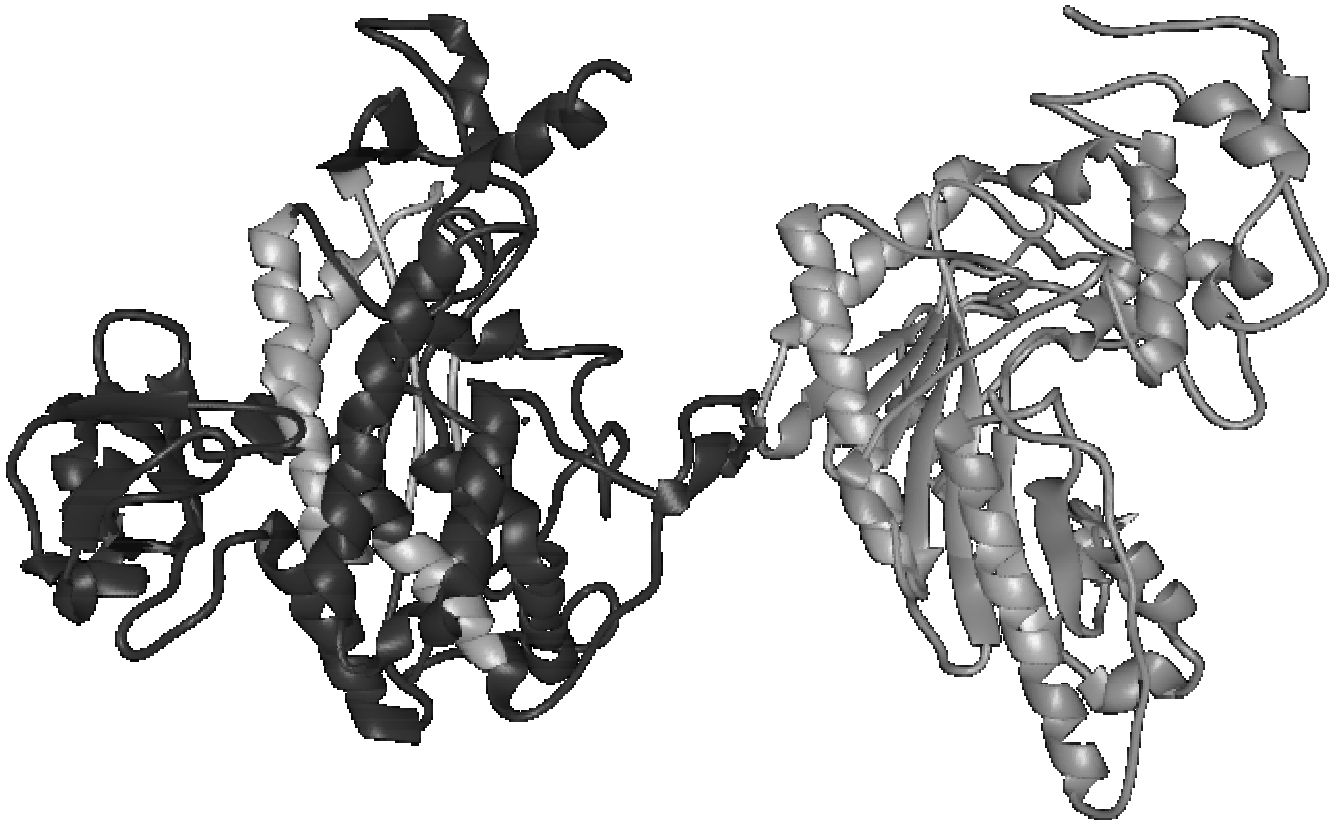
- **GS: 468 residues, PDB entry 2gls**
- **CK: 380 residues, PDB entry 1crk**
- **No significant sequence similarity, both have multimeric forms, proposed similar tertiary structures, and catalyze similar reactions**

GS and CK catalysis

Glutamine synthetase and creatine kinase



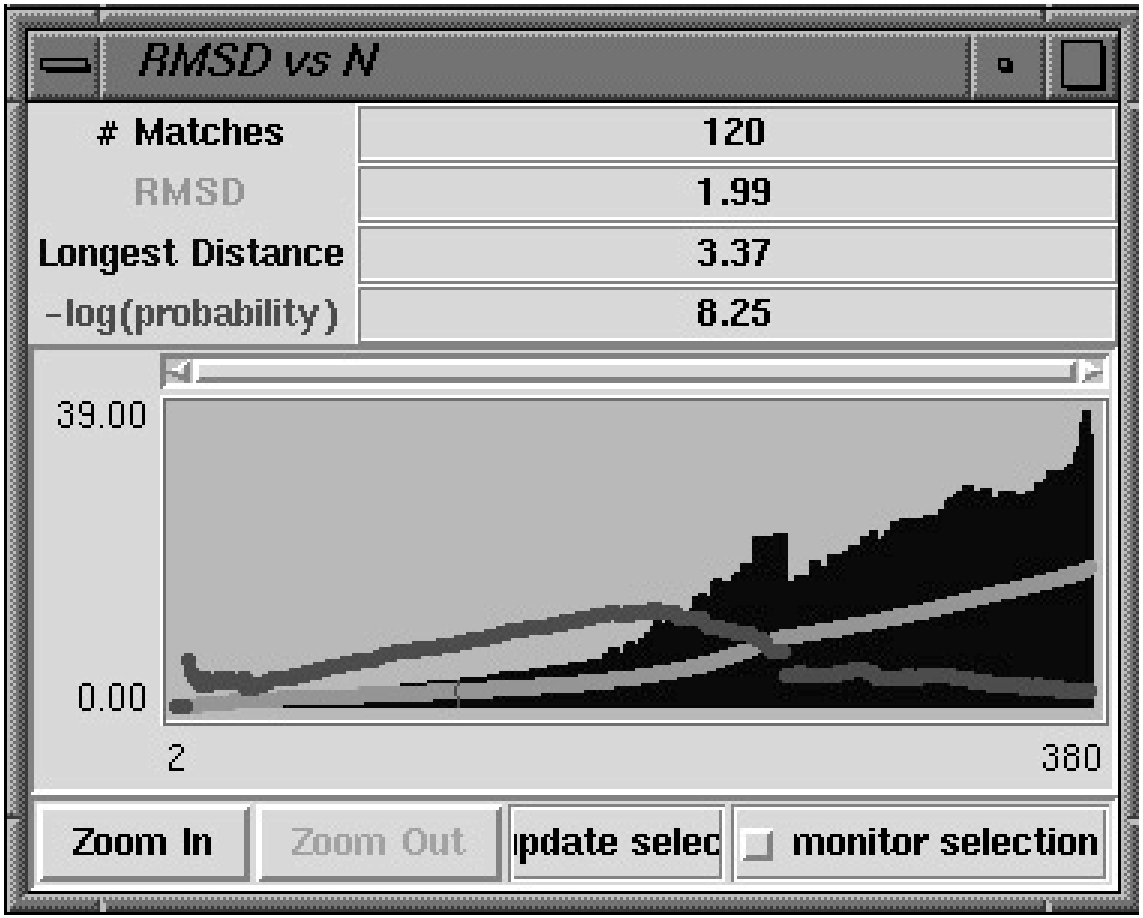
After MinRMS alignment



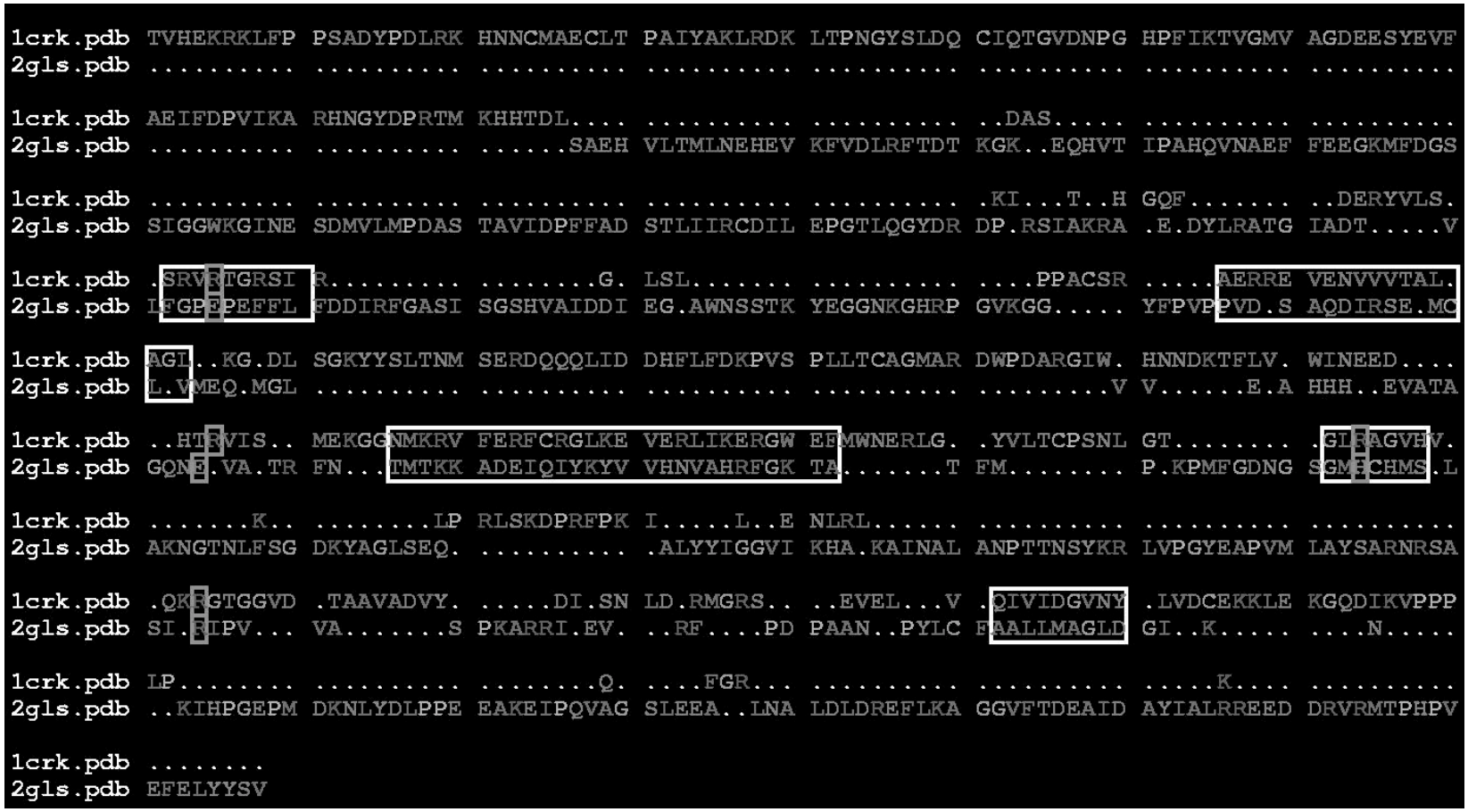
Glutamine synthetase

Creatine kinase

AlignPlot GUI



Resulting structure-based sequence alignment



Live Demonstration

Disclaimer: Anything that can go wrong will do so in direct proportion to the number of people watching

Recent developments

Re-engineering of a natural enzyme with new catalytic function

- **Alan Fersht & coworkers at Cambridge Centre for Protein Engineering**
- **Convert activity of indole-3-glycerol phosphate synthase (IGPS) into that of phosphoribosylanthranilate isomerase (PRAI)**
- **See C&E News February 21, 2000**
- ***Nature* 403,** 617 (2000)

Acknowledgements

Co-authors

- Prof. Patricia Babbitt, Prof. Teri Klein, Dr. Conrad Huang

National Center for Research Resources

- P41-RR01081

Department of Energy

- DE-FG03-96ER62269

National Institutes of Health

- AR17323

Additional Information

See UCSF Computer Graphics Laboratory web site:

http://www.cgl.ucsf.edu/chimera