A Collaboratory for Interactive Molecular Modeling

Project Overview

Resource for Biocomputing, Visualization, and Informatics

University of California, San Francisco

Resource objectives:

To create and apply computational and visualization methods for solving a wide range of genomic and molecular recognition problems within the complex sequence→structure→function triad.

- Characterization and interpretation of genomic data, including knowledge discovery and transfer.
- Development, application, and dissemination of analysis methodologies and tools in computational structural biology.
- Theoretical and applied research in understanding how protein structures deliver function.

Sequence → Structure → Function

Challenges:

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

Potential for major impact on…

- Drug discovery
- Prediction of drug response
- Avoidance of toxic effects in many individuals

Collaboratory goals:

Develop a collaboratory environment for carrying out interactive three-dimensional molecular modeling studies for research and training

- Multiple scientists at remote locations to interactively manipulate shared, complex three-dimensional molecular models (as similar as possible to “face-to-face” collaborations)
- Full semantics for the modification of an object by any collaborative participant
  - access to the object's data, not just the object's graphical representation
  - individual participants can perform operations privately first, then present results in collaborative session

Motivation

Solid evidence of benefit of "face-to-face" scientific collaboration through a common visually-based work environment

- 1,000's of multiple author publications
- direct observation of interaction among visiting scientists
- direct reports that "collaboration promotes new ideas"

Desire to extend environment to remote scientists

- promote more frequent and spontaneous collaboration sessions
- encourage new collaborations

Desire to utilize remote collaborative environments in new ways

- Training
- Improved feedback for software development

Typical scenario

Input commands (typed keywords or menu selections) on one workstation, display and execute these commands on all other participating collaborative workstations

- Displayed molecules move in tandem on all workstations simultaneously in response to input from any participant

Provide independent control for each participant

- separately shaped or colored mouse cursors for highlight interesting facets of a molecular model
- interactive control of rotations, translations and scaling
- “private” window for testing ideas
- participants can join and/or leave an on-line session at any time

Why interactive graphics?

Qualitative

Generates pictures

Quality vs. interactive real-time

Value lies not in numbers themselves, but insights gained

Idea generator

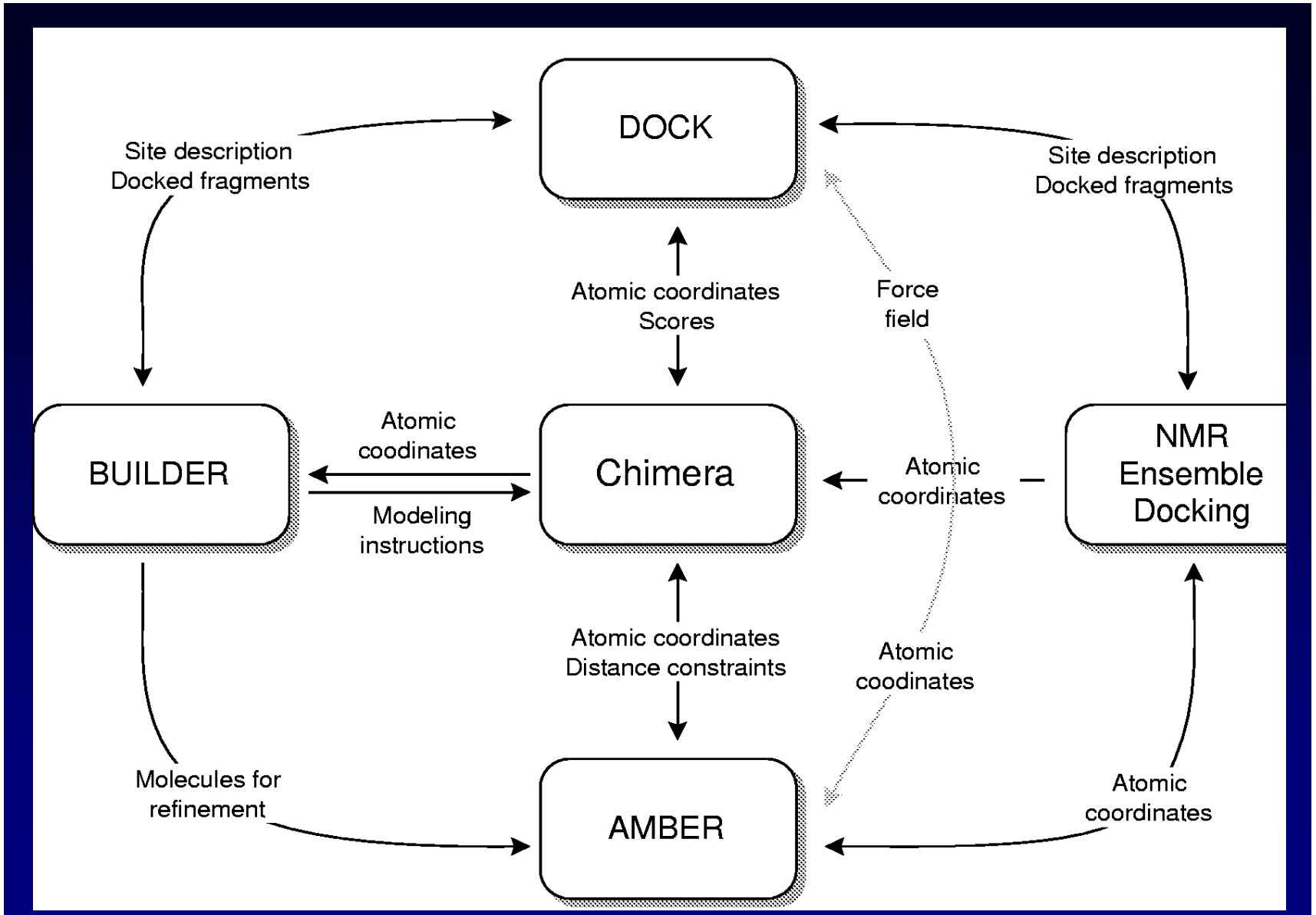
Complementary to experimental methods

- X-ray crystallography
- NMR spectroscopy
- Mass spectroscopy

Structure-Based Drug Design

Facile and comprehensive system for determining the structures of proteins and nucleic acids in solution and designing new ligands and drugs

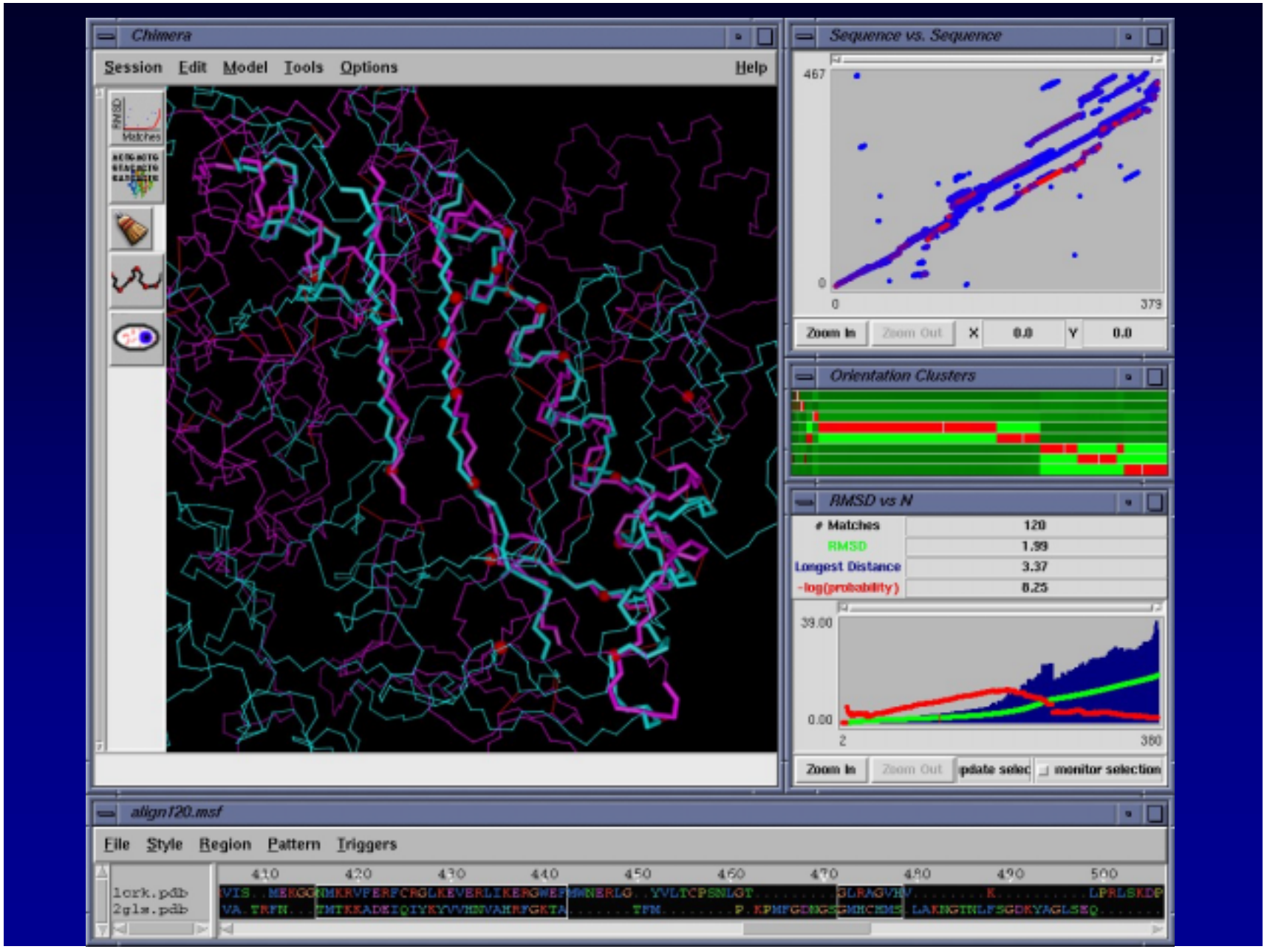
- Integrated Software Tools for Structure-Based Drug Design Applications [figure 1]


DOCK
Site description
Docked fragments
Site description
Docked fragments
Atomic coordinates
Scores
Force
field
BUILDER
Atomic
coodinates
Modeling
instructions
Chimera
Atomic
coordinates
NMR
Ensemble
Docking
Atomic coordinates
Distance constraints
Atomic
coodinates
Molecules for
refinement
AMBER
Atomic
coordinates

UCSF's Chimera

Molecular visualization system providing:

- Interactive manipulation of multiple molecular structures
- Real-time rendering of models in several formats
  - ball-and-stick, ribbons, molecular surfaces...
  - Takes advantage of available high-performance graphics hardware
- Support for non-molecular objects
  - points, vectors, spheres, cylinders, polygons...
- Both menu and command line modes of interaction
  - Command emulator for compatibility with MidasPlus
  - "Power users" prefer command line based system
- Extensible functionality without access to source code through a high-level, dynamically-loadable programming language (Python)
  - Allows ideas for new functionality to be quickly tested
- Use of standard APIs ensure portability to many platforms
  - Windows 98/NT/2000, Compaq Alpha, 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
1ork.pdb
2glm.pdb

Research plan

Phased implementation approach

- 1:1
- 1:N
- N:N

Multiple testbed research projects

- Molecular Mechanisms of Mutagenesis and DNA Repair: Recognition of Damaged DNA
- Drug specificity of Dihydrofolate Reductase
- Structural Aspects of type 1 Collagen in imperfecta Osteogenesis
- Feedback to Chimera developers

Training at two levels

- active learning for local and remote Chimera users
- active learning of molecular recognition and drug design principles for biologists and chemists

Technologies employed

Chimera

- Collaboratory being implemented as extension to basic system

COTS desktop videoconferencing

- Prefer standards-based, platform independent tools (e.g. vic and rat)
- High quality audio imperative, low frame rate video just fine

Low-latency networking

- Initial session startup may require high bandwidth, but otherwise only modest bandwidth needs envisioned (this needs to be evaluated)
  - Internet 2?
  - Campus local area network?
  - DSL/Cable modem?

Networking issues

Parameters affecting network performance

- bandwidth
- latency
- delay variance
- connection quality
- multicast support
- quality-versus-price tradeoffs

Collaboratory data connection characteristics

- lower bandwidth requirements than videoconferencing (especially audio)
- need high reliability (I.e. low packet loss)
- bursts of activity will by typical

Project status

Funding began in August 1999

Project staffing has been a challenge

Design and initial implementation recently completed

- demonstration at Saturday afternoon session

Initial deployment to selected user research community during winter 2000

Testbed projects and evaluation to follow

Acknowledgements

People

- Conrad Huang
- David Konerding
- Greg Couch
- Eric Pettersen
- Al Conde
- Teri Klein
- Patsy Babbitt
- Elaine Meng

Funding

- NCRR P41-RR01081

Additional information

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