Visualization and Data Analysis with VIDA

Joe Corkery
OpenEye Scientific Software
OpenEye

- Small software company
- Efficient large scale 3D computations
- Tools for managing computed data
VIDA

- Primarily a data visualization and navigation tool
- Makes large data sets accessible to the user
- Interfaces with data generation tools
Overview

- Philosophy
- A Brief History of VIDA
- Challenges
- Examples
- Future
Philosophy

- Million molecule viewer
- Multiple simultaneous views of data
- All (reasonable) platforms
- Configurable for different end users/uses
- Basic access to all OpenEye functionality
The Million Molecule Viewer

- Large scale modeling in memory
  - ~1-2K per molecule
  - Memory is cheap (2 GB of RAM - ~ $400, why not?)
  - Faster and cheaper hardware makes real-time database wide operations feasible

- Why?
  - Analyze/search your entire corporate database
  - View all your docking/simulation results
  - Expand virtual libraries
Multiple Simultaneous Views of Data

- 3D structures – molecules, surfaces, grids, proteins, boxes, ellipsoids
- 2D depictions
- Line notation
- Chemically-aware spreadsheet
  - no 64k row limit
- Data graphs
- Clustering
All (Reasonable) Platforms

- Built using Qt multi-platform GUI toolkit
- 3D displays through OpenGL
- Laptops and Desktops

- Currently supported:
  - Windows *.*
  - Linux
  - SGI Irix
  - Mac OS X
  - Compaq Alpha Tru64

- Under Development
  - IBM AIX
Technological Balkanization

- Multiple platforms within individual companies
- Different groups of people use different platforms (e.g. modelers using SGIs and chemists using PCs, CEOs using Macs)
“Es mi VIDA”

- Target Audiences
  - Modelers
  - Chemists
  - Crystallographers

- Scripted
  - Internal command language
  - RasMol / GRASP interpreters
  - Python wrappers
  - Java wrappers

- Read from networks (via URLs) / databases
- Plug in your own code
Access to OpenEye

• Data generation available
  – Cheminformatics (e.g. SMARTS matching)
  – Physical property calculation (FILTER)
  – Poisson-Boltzman electrostatics (ZAP)
  – Structure generation (OMEGA)
  – Shape matching (ROCS / shape toolkit)
  – Docking (FRED)
Overview

- Philosophy
- A Brief History of VIDA
- Challenges
- Examples
- Future
A Brief History of VIDA

- Development began as a part-time project in November, 1999.
- Born out of a need to easily visualize large numbers of molecules.
- Early releases were well received and its potential recognized.
- Became a full-time development project in June, 2000.
History cont.

• Spread quickly amongst our friends and customers.
  – Acquired many cheerleaders with lots of useful feedback

• Early design decision to target a broad audience
  – Initially chemists and modelers
  – Expanded to biologists and crystallographers
  – Designed to be easy enough for everyone to use

• Growing up into a simple, yet powerful interface to support large scale modeling.
Present

- Currently being used by both modelers and chemists alike in many companies and universities around the world.

- Large scale deployment as the primary visualization platform has been proposed and/or begun in many companies as well.
Overview

- Philosophy
- A Brief History of VIDA
- Challenges
- Examples
- Future
Challenges

• Large scale visualization
  – Enormous amounts of data
  – Real-time usability

• 3D for everyone
  – Platform independence
  – stereo
  – 2D
  – Accessibility
Accomplishments

- Large scale modeling
  - Memory usage
  - File I/O and multiple formats

- Mixing physics and chemistry

- Integrated 2D & 3D
Initial Limitations

- **3D Building**

- **Components**
  - Made the initial decision that nobody would really want this available as a library
  - Assumed that anybody who wanted a library version would be willing to use Qt.

- **Scripting**
  - Did not anticipate the need and desire for a scriptable application
Components

- Support for custom GUI plug-ins
- Molecular property calculation plug-ins
Overview

- Philosophy
- A Brief History of VIDA
- Challenges
- Examples
- Future
Advanced Examples

- Present pre-setup information via the web
- Generate script files to visualize docking output
VIDA 1.0: Currently Available

Booth 170/172

OpenEye Scientific Software
Overview

- Philosophy
- A Brief History of VIDA
- Challenges
- Examples
- Future
Future

- Advanced scripting through Python.
- Direct relational database access
- More components and plug-ins
- John Barnard’s clustering toolkit
- Advanced visualization components with AVS
Conclusion

- Powerful tool for data interrogation
- Capable of handling virtually any size data set
- Integration of 2D structure, 3D structure, associated data
- Available on multiple platforms
- Customizable and scriptable
Acknowledgements

OpenEye
- Mark McGann
- Anthony Nicholls
- Roger Sayle
- Geoff Skillman
- Matt Stahl

Fan Club
- Juan Alvarez
- Jeff Blaney
- Jonas Bostrom
- Ken Brameld
- Paul Charifson
- Claudio Chuaqui
- Donovan Chin
- Jaouad Elbahraoui
- Andrew Grant
- Brian Goldman
- James Haigh
- John Irwin
- Man-Ling Lee
- Dave Lorber
- Ken Lind
- Demetri Moustakas
- Steve Muchmore
- Connie Oshiro
- Govinda Rao
- Bob Tolbert
- John Van Drie
- Pat Walters

Contributors
- Wolf Dietrich-Ihlenfeldt
- Brian Maly