ProMOL: Simplification and Increased Functionality of PyMOL

By Brett Hanson, Charlie Westin, Paul Craig, and Len Slatest

Molecular Modeling Programs

- Growth in Structural Biology data
- Challenges in analysis of this data
- Molecular modeling programs to cut down time and expense

PyMOL

- Created by Warren DeLano
- Open Source
- User-sponsored
- Extensible by the Python programming language
- Can create plugin programs that use PyMOL's abilities. ProMOL is a plugin.
- Powerful analytical tools and graphics
- Complex syntactical commands

Simplification of PyMOL

- Long syntactical commands can be replaced by buttons
- A Graphical User Interface (GUI) helps to facilitate simplification
- ProMOL provides ready access to dozens of PyMOL tools

Simplification of PyMOL

- To enable the roving function seen next requires the following commands:

```
Hide everything
Remove hydro
Set roving_detail,1
Set roving_origin,1
Set stick_radius, 0.3
Set roving_sticks, 4
Set roving_polar_contacts, 8
```

Motif Prediction

- Motif prediction using a novel approach
- Conserved spatial and structural recognition
- Does not look at sequence
- Enables the prediction of enzyme catalytic sites quickly and easily

ProMOL Motif Definitions

- PyMOL uses the PDB file format
- Three-dimensional position of each atom
- Motif definitions rely on this format and PyMOL's selection algebra

ProMOL Motif Definitions

- Requires a repository of literature entries
- The Catalytic Site Atlas (CSA)
- http://www.ebi.ac.uk/thornton-srv/databases/CSA/

ProMOL Motif Definitions

- Template proteins from the CSA
- Residue atom cross measurements
- Specificity from PyMOL's selection algebra

ProMOL Motif Definitions

```
cmd.select('gln1', 'name ne2 and resn gln within 7 of (name ne2 and resn his)')
cmd.select('gln2', 'name cd and resn gln within 6.7 of (name ce1 and resn his)')
cmd.select('gln3', 'name ce1 and resn gln within 7 of (name nd1 and resn his)')
cmd.select('gln4', 'name ne2 and resn gln within 5.5 of (name cb and resn cys)')
cmd.select('gln5', 'name ce1 and resn gln within 6.7 of (name sg and resn cys)')
cmd.select('gln', 'byres gln1 and byres gln2 and byres gln3 and byres gln4 and byres gln5')
cmd.select('his1', 'name ne2 and resn his within 7 of (name ne2 and gln)')
cmd.select('his2', 'name ce1 and resn his within 6.7 of (name cd and gln)')
cmd.select('his3', 'name nd1 and resn his within 7 of (name oe1 and gln)')
cmd.select('his4', 'name ce1 and resn his within 5.7 of (name cb and resn cys)')
cmd.select('his5', 'name nd1 and resn his within 6 of (name sg and resn cys)')
cmd.select('his', 'byres his1 and byres his2 and byres his3 and byres his4 and byres his5')
cmd.select('cys1', 'name cb and resn cys within 5.7 of (name ce1 and his)')
cmd.select('cys2', 'name sg and resn cys within 6 of (name nd1 and his)')
cmd.select('cys3', 'name cb and resn cys within 5.5 of (name ne2 and gln)')
cmd.select('cys4', 'name sg and resn cys within 6.7 of (name ce1 and gln)')
cmd.select('cys', 'byres cys1 and byres cys2 and byres cys3 and byres cys4')
cmd.select('paplike', 'gln(his(cys))')
```

ProMOL Motif Definitions

- Selection algebra provides specific requirements for motifs to be returned.
- To be shown, amino acids must fit all selection requirements
- Small motif distances increase specificity
- Large motif distances decrease specificity

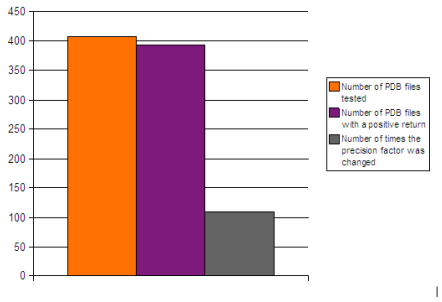
ProMOL Motif Definitions

- PDB files exist at different angstrom resolutions
- This affects motif distance measurements
- Can accommodate for this
- Append variable coefficient in front of measurements

Testing Motifs

- The CSA allowed motifs to be tested on homologous proteins
- Each tested on 15 different homologues
- Diversify homologues tested
- Species variability

Testing Motifs



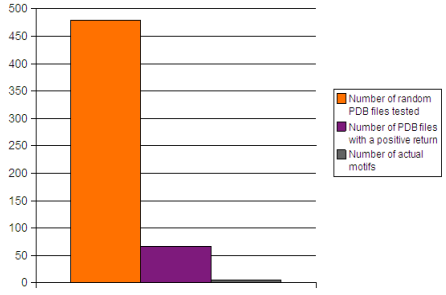
Testing Motifs

- 96% success rate
- 28% required range modification
- 68% of motifs accurately predicted with no modification

Testing Motifs

- False positives
- Random PDB file downloader
- Each motif tested on 15 random PDB files
- No PDB files composed of nucleic acids

Testing Motifs



Faster Motif Making

- New plugin called Motif Maker.
- Automates most of the process.
- Saves time, reduced human error.

Current Problems

- Catalytic sites with only one or two residues
- Catalytic sites with multiple identical residues
- Catalytic sites with large distances between residues

Summary

- Novel method of protein function determination
- Uses extend beyond enzymes
- Any small molecular pattern can be investigated
- Alleviate time and money expenditure

The Future

- Phi and Psi angles as requirements
- Secondary Structure requirements
- Support for multiple identical residues

Acknowledgments

- Department Of Energy
- Brookhaven National Laboratory
- Rochester Institute of Technology
- NIH 1R15GM078077-01
- NSF DUE 0402408