cealign: A Structure Alignment Plugin for PyMol

Jason Vertrees

April 6, 2007

Contents

1 Introduction

1.1 Origins

cealign[1] is a structure alignment plugin for PyMol. It is implemented in Python and C.

1.2 Implementation

The code is relatively quick — at least on my machine — and a typical alignment takes less than 1 second. I also wrote a pure-Python version of the very same program, but never released it because it was about an order of magnitude slower that the C/Python mixed code.

1.3 Thanks

I'd like to thank Ilya Shindyalov for helpful comments on the CE algorithm. I'd like to thank Bosco K. Ho for his discussion about the SVD solution to the problem. I'd also like to thank MIT for their Open Course Ware [2] online classes. I watched the entire Linear Algebra series, which was applied here.

I'd like to thank the testers:

- Bryan Sutton
- Shiven Shandilya
- Jouni Valiaho
- Robert Campbell

1.4 To Do

1. The original CE alignment algorithm calls for two round of optimization. I did not yet implement those two rounds; the RMSD improvement can be significant – up to about 2.0Å! This should be added soon.
2. More/Cleaner Documentation.
3. Good test scripts.

[1] Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. **Protein Eng.** 1998 Sep;11(9):739-47. PMID: 9796821

[2] http://ocw.mit.edu/

2 Install

2.1 Requirements

To install *cealign* you'll need the following:

- Modern Linux/Unix machine (Windows/Mac testers needed)
- PyMol 0.93+
- Python 2.4+
- Numpy (tested with 1.0.1)

2.2 Quick Install

To install *cealign*, simply do the following:

```
> tar -jzvf cealign-VERSION.tar.bz2
> cd cealign-VERSION
> sudo python setup.py install
```

You need to edit your /.pymolrc file (or Mac/Windows equivalent) if you PyMol to automatically load the *cealign* extension. Add the following to your .pymolrc file, or simply run the two scripts in the *cealign* directory called, "cealign.py" and "qkabsch.py".

```
run cealign-VERSION/cealign.py
run cealign-VERSION/qkabsch.py
```

3 Using

Once installed, one can perform an alignment using the following syntax:

```
cealign SEL1, SEL2
```

where SEL1 is the first selection to align, and SEL2 is the second selection. Any PyMol selection should be valid. For example all of the following are syntactically valid:

```
# align two proteins
cealign 1ggz, 1cll

# align residue ranges on two proteins
cealign 1cbs and i. 6-55, 1hmt and i. 6-90

# align certain chains on two proteins
cealign 1OKE and c. A, 1S6N
```

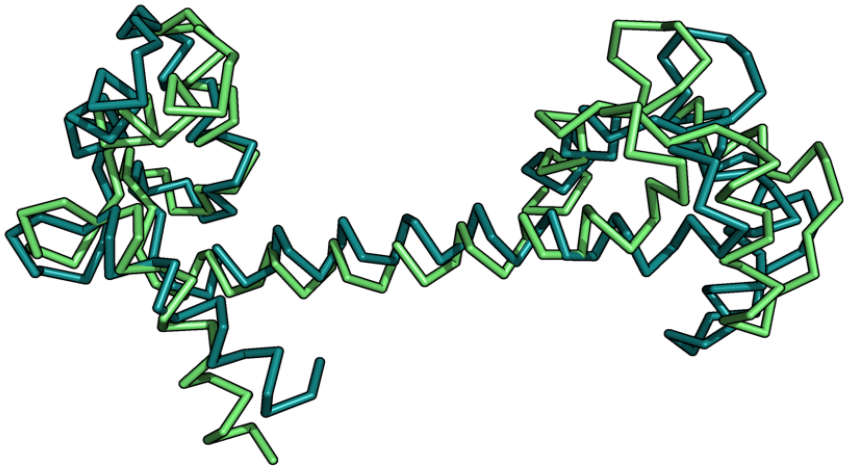

Figure 1: 1CLL vs. 1GGZ

4 Examples

For a more thorough set up example, please see *cealign*'s home page on the PyMOLWiki. There are examples ranging from easy to difficult with images and updates for each revision of this program.

1CLL and 1GGZ. Calmodulins. Results are 128 aligned residues to an RMSD of 4.244Å.

```
cealign 1cll, 1ggz
```

1KAO and 1CTQ. Results are 152 residues to an RMSD of 1.575Å.

```
cealign 1KAO, 1CTQ
```

1OKE and 1S6N. Now, for a more difficult alignment. Viral capsid E-proteins. Results are 96 aligned residues to an RMSD of 2.260Å.

```
cealign 1OKE, 1S6N
```

5 References

1. CE Align on the PyMol Wiki

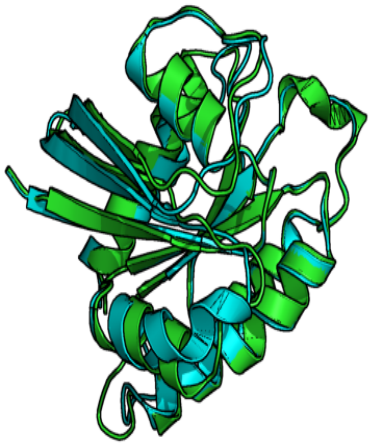

Figure 2: 1CTQ to 1KAO

Figure 3: 1OKE vs. 1S6N. Aligned 96 residues to an RMSD of 2.260Å.

(http://www.pymolwiki.org/index.php/Cealign)

2. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. **Protein Eng.** 1998 Sep;11(9):739-47. PMID: 9796821 [PubMed - indexed for MEDLINE]

3. Jia Y, Dewey TG, Shindyalov IN, Bourne PE. A new scoring function and associated statistical significance for structure alignment by CE. **J Comput Biol.** 2004;11(5):787-99. PMID: 15700402 [PubMed - indexed for MEDLINE]

4. Pekurovsky D, Shindyalov IN, Bourne PE. A case study of high-throughput biological data processing on parallel platforms. **Bioinformatics.** 2004 Aug 12;20(12):1940-7. Epub 2004 Mar 25. PMID: 15044237 [PubMed - indexed for MEDLINE]

5. Shindyalov IN, Bourne PE. An alternative view of protein fold space. **Proteins.** 2000 Feb 15;38(3):247-60. PMID: 10713986 [PubMed - indexed for MEDLINE] The CE Alignment Server

   (http://cl.sdsc.edu/)

6. More on structural alignments

   (http://en.wikipedia.org/wiki/Protein_structural_alignment)

6 License & Copyright