Ultimate Secondary Structure Assignment Method

Projet Long

Manon Curaudeau

17th January 2020

M2BI - Université Paris Diderot

Context

Secondary structure assignment:

- Necessary step in the analysis of 3D protein structures

Context

Secondary structure assignment:

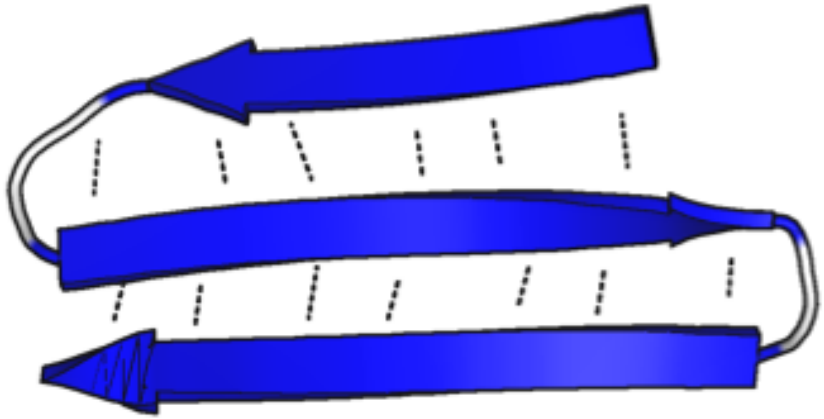
- Necessary step in the analysis of 3D protein structures

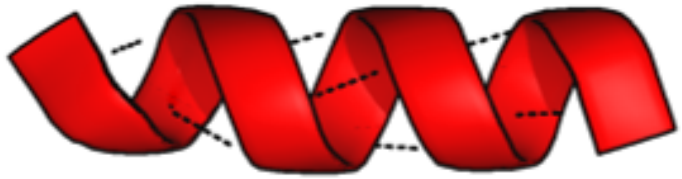
- Most popular software is DSSP (Kabsch and Sander, 1983)

Context

Secondary structure assignment:

- Necessary step in the analysis of 3D protein structures
- Most popular software is DSSP (Kabsch and Sander, 1983)
- Helices (α, π, and 3_{10}), β–strands, β–bridges, turns, and bends





Context

Other methods

Context

Other methods

ProMotif

- β-turns
- β-hairpins
- β-buldges

(Hutchinson and Thornton, 1994)

Context

Other methods

ProMotif

- β-turns
- β-hairpins
- β-buldges

(Hutchinson and Thornton, 1994)

Helanal

- Linear
- Curved
- Kinked

(Kumar and Bansal, 1998; Bansal et al., 2000)

Ultimate Secondary Structure Assignment Method

Combining and centralising the preexisting tools to implement a more generic approach

Programming

UltimateSSAM divided into several modes, each of them performing a secondary structure assignment method.

All scripts were written in Python 3.7

```
$ python3 ssam.py mode -i input -o output
```

Programming

Read input files. Both PDB (`.pdb`) and PDBx/mmCIF (`.cif`)

- Atomic coordinates
- Atom object instance
- Residue object instance
- Chain: list of Residue object instances
- List of chains

DSSP MODE

```
$ python3 ssam.py dssp -i input -o output
```

DSSP MODE

```
$ python3 ssam.py dssp -i input -o output
```

$$E = q_1 q_2 \left(\frac{1}{r_{ON}} + \frac{1}{r_{CH}} - \frac{1}{r_{OH}} - \frac{1}{r_{CN}}\right) * f$$

DSSP MODE

```
$ python3 ssam.py dssp -i input -o output
```

$$E = q_1 q_2 \left(\frac{1}{r_{ON}} + \frac{1}{r_{CH}} - \frac{1}{r_{OH}} - \frac{1}{r_{CN}}\right) * f$$

- Elementary Hydrogen Bond Patterns
- Cooperative Hydrogen Bond Patterns
- Eight different conformational states

DSSPCOMPARE MODE

Comparison of the output of the `dssp` mode with the output if DSSP for a given protein

```
$ python3 ssam.py dsspcompare -i input -o output -oc
    output-compare
```

DSSPCOMPARE MODE

Comparison of the output of the `dssp` mode with the output if DSSP for a given protein

```
$ python3 ssam.py dsspcompare -i input -o output -oc
    output-compare
```

Percentage of matches for the eight conformational states

Percentage of matches for the three classes: H = HGI, E = EB, and C = STC

Datasets

Three datasets of 100 proteins each

- One chain

Datasets

Three datasets of 100 proteins each

- One chain
- Multiple chains

Datasets

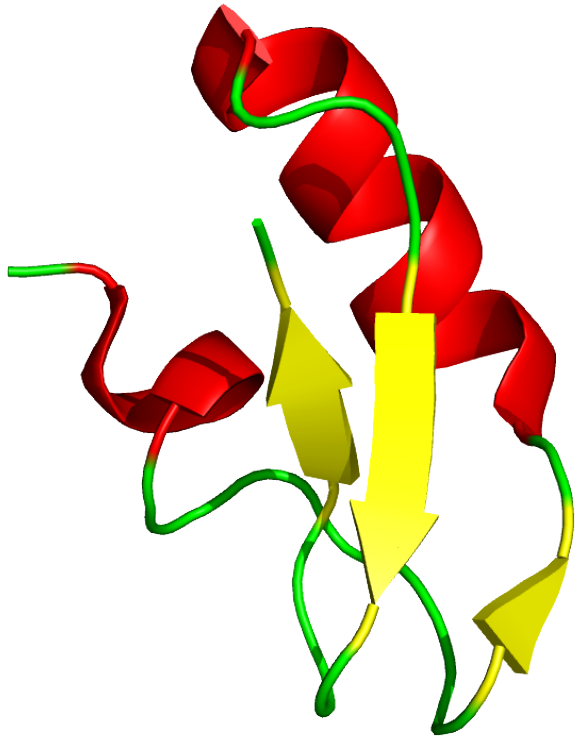
Three datasets of 100 proteins each

- One chain
- Multiple chains
- Random proteins

Results - 1PX9

CnErg1 Ergtoxin (42 residues) synthesised by *Centruroides noxius*

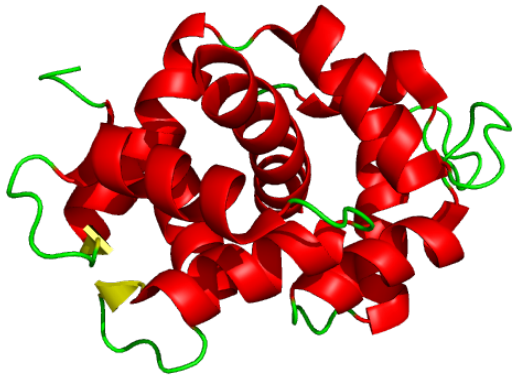
- one α-helice
- one 3_{10}-helice
- one β-sheet with two antiparallel β-strands



Results - 2L7L

Calmodulin-binding domain of calmodulin kinase I (170 residues in 2 chains) synthesised by *Homo sapiens*

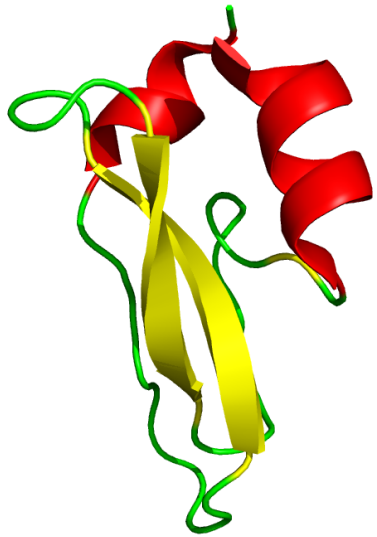
- nine α-helices
- one antiparallel β-strand



Results - 6PTI

Pancreatic trypsin inhibitor (58 residues) synthesised by *Bos taurus*

- one α-helice
- one 3_{10}-helice
- one β-sheet with two antiparallel β-strands



Comparison with DSSP

	Matches between DSSP and UltimateSSAM			
	With infile hydrogen atoms		With added hydrogen atoms	
Dataset	Percentage (%)	Standard deviation	Percentage (%)	Standard deviation
One chain	97.65	2.68	98.82	2.24
Multiple chains	96.09	4.57	96.83	4.41
Random proteins	–	–	97.62	2.96

Provisional budget

	Matches between DSSP and UltimateSSAM					
	Helices (H = HGI)		Sheets (E = BE)		Coils (C = STC)	
Dataset	Percentage (%)	Standard deviation	Percentage (%)	Standard deviation	Percentage (%)	Standard deviation
One chain	97.25	12.042	97.59	3.96	99.39	1.41
With added hydrogen	99.95	0.47	97.96	3.18	99.97	0.34
Multiple chains	99.37	2.19	86.68	25.43	99.42	1.35
With added hydrogen	99.94	0.43	87.02	25.53	99.99	0.06
Random proteins	99.80	1.47	94.17	11.31	99.934	0.42

Conclusion

- Good performances compared to DSSP

Conclusion

- Good performances compared to DSSP
- Add missing functions: residue uncertainty, multiple models, negative residues, etc.

Conclusion

- Good performances compared to DSSP
- Add missing functions: residue uncertainty, multiple models, negative residues, etc.
- ProMotif (Hutchinson and Thornton, 1994) and Helanal (Bansal et al., 2000)

Conclusion

- Good performances compared to DSSP
- Add missing functions: residue uncertainty, multiple models, negative residues, etc.
- ProMotif (Hutchinson and Thornton, 1994) and Helanal (Bansal et al., 2000)
- PDBx/mmCIF

Thank you for your attention

References 1

Kabsch, W. and Sander, C., 1983. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen–Bonded and Geometrical Features. *Biopolymers: Original Research on Biomolecules*, 22(12), pp.2577–2637.

Hutchinson, E.G. and Thornton, J.M., 1994. A revised set of potentials for β-turn formation in proteins. *Protein Science*, 3(12), pp.2207–2216. doi : 10.1002/pro.5560031206.

Kumar, S. and Bansal, M., 1998. Geometrical and sequence characteristics of α-helices in globular proteins. *Biophysical Journal*, 75(4), pp.1935–1944. doi : 10.1016/S0006-3495(98)77634-9.

Bansal, M., Kumart, S. and Velavan, R., 2000. Helanal: a program to characterize helix geometry in proteins. *Journal of Biomolecular Structure and Dynamics*, 17(5), pp.811–819. doi : 10.1080/07391102.2000.10506570.