Roadmap

v1.3 Codename: *Monkeys in a Barrel*

Structure

Sequence

Big Picture Hypothesis

"Fold subsequence duplication is a statistically significant mode of protein domain evolution."

Trefoil Broom-Doxey

TIM Barrel Ma

Snip SSEs at Coils

Best Fit Euclidean Solid; Simplex Algo

PSIBlast Representative Snipped SSEs in CDD

Foil-Sensitive MSA
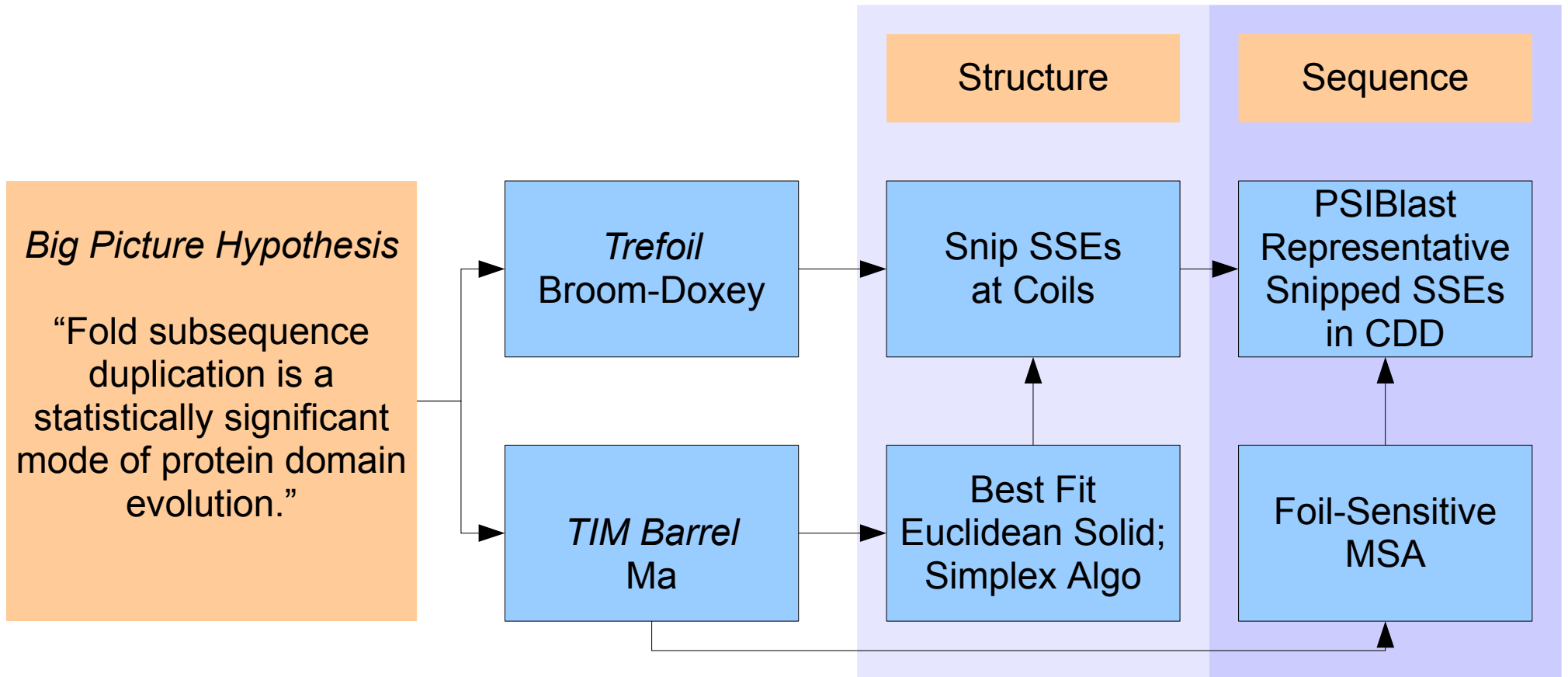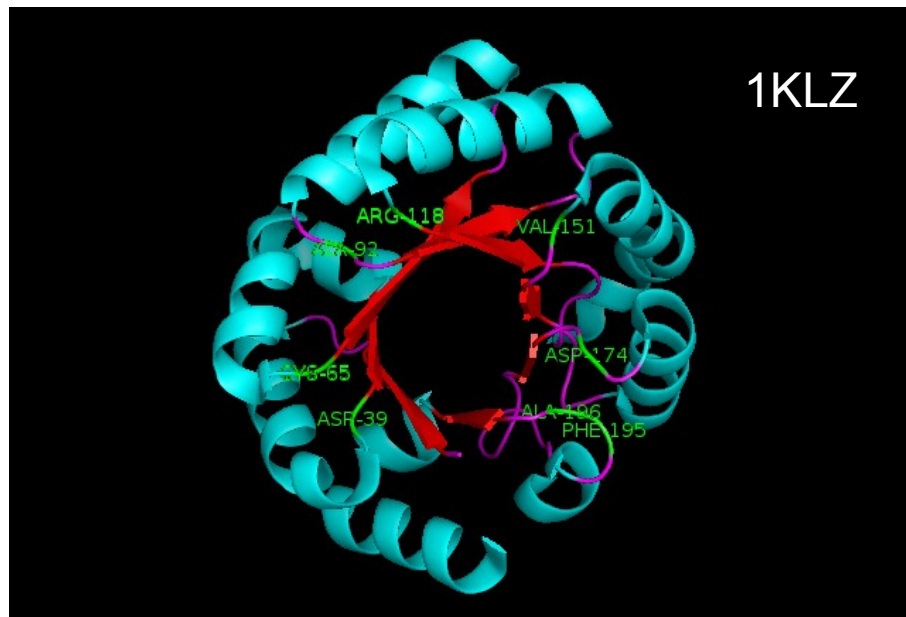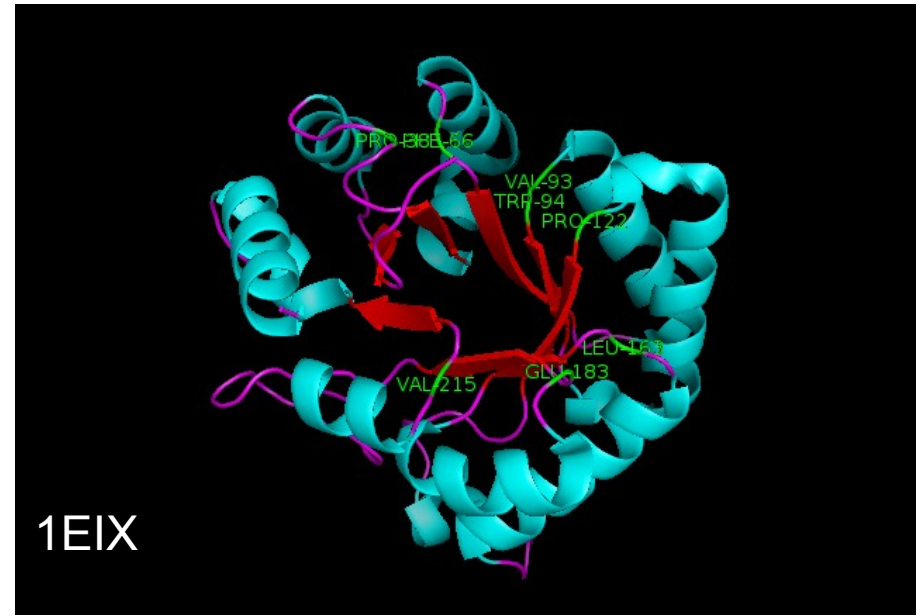
Structural Analysis Method

Best Fit Solids Approach

Breaking apart TIM barrels

More obvious ones

Breaking apart TIM barrels
Less obvious ones

Doxey's hint: Parse SSEs

Parse: To derive meaning from sequence
SSE: Secondary Structural Element

- ## Our operating definition of a SSE

  - A single subsequence or substructure that is repeated in a single translation unit

  - A unit foil in n-foils

- ## TIM Barrel SSE = (β-a)

- ## Trefoil SSE = (β-β-β-o-β)

- ## i.e. Parse the calculated secondary structure and break apart each SSE.

Problem with Parsing
## TIM barrel SSEs aren't always regular

```
>1GQN_A
1; 3; 3; Coil; ___;
4; 22; 19; Beta; EEE__EEE_____EEEEEE;
23; 26; 4; Coil; ___;
27; 37; 11; Alpha; HHHHHHHHHHH;
38; 43; 6; Coil; _____;
44; 48; 5; Beta; EEEEE;
49; 71; 23; Alpha; HHH_____HHHHHHHHHHHHHH;
72; 76; 5; Coil; ____;
77; 83; 7; Beta; EEEE__E;
84; 86; 3; Alpha; HHH;
87; 89; 3; Coil; ___;
90; 90; 1; Beta; E;
91; 93; 3; Coil; ___;
94; 107; 14; Alpha; HHHHHHHHHHHHHH;
108; 111; 4; Coil; ____;
112; 116; 5; Beta; EEEEE;
117; 133; 17; Alpha; HHH_HHHHHHHHHHHHH;
134; 136; 3; Coil; ___;
137; 144; 8; Beta; EEEEEEEE;
145; 150; 6; Coil; _____;
151; 163; 13; Alpha; HHHHHHHHHHHHH;
164; 167; 4; Coil; ____;
168; 173; 6; Beta; EEEEEE;
174; 177; 4; Coil; ____;
178; 194; 17; Alpha; HHHHHHHHHHHHHHHHH;
195; 200; 6; Coil; _____;
201; 204; 4; Beta; EEEE;
205; 209; 5; Coil; _____;
210; 219; 10; Alpha; HHHHH_HHHH;
220; 222; 3; Coil; ___;
223; 238; 16; Beta; EEE_E_____E;
239; 250; 12; Alpha; HHHHHHHHHHHH;
251; 252; 2; Coil; __;
```

```
>1GVO_A
1; 8; 8; Coil; _____;
9; 36; 28; Beta; EEE__EEE___EEE_____E____E;
37; 38; 2; Coil; __;
39; 47; 9; Alpha; HHHHHHHH;
48; 52; 5; Coil; _____;
53; 73; 21; Beta; EEEEEEEE_____E;
74; 76; 3; Coil; ___;
77; 92; 16; Alpha; HHHHHHHHHHHHHHHH;
93; 96; 4; Coil; ___;
97; 102; 6; Beta; EEEEEE;
103; 109; 7; Coil; _____;
110; 116; 7; Alpha; HHH_HHH;
117; 119; 3; Coil; ___;
120; 150; 31; Beta; EE_____EEEEE_____EEEEE____EE;
151; 152; 2; Coil; __;
153; 172; 20; Alpha; HHHHHHHHHHHHHHHHHHHH;
173; 176; 4; Coil; ___;
177; 182; 6; Beta; EEEEEE;
183; 186; 4; Coil; ___;
187; 229; 43; Alpha; HHHHHH_____HHHH_HHHHHHHHHHHHH__HHH;
230; 243; 14; Beta; EEEEE_____E__E;
244; 248; 5; Coil; _____;
249; 262; 14; Alpha; HHHHHHHHHHHHHH;
263; 266; 4; Coil; ___;
267; 277; 11; Beta; EEEEE__E__E;
278; 282; 5; Coil; _____;
283; 290; 8; Alpha; HHHHHHHH;
291; 295; 5; Coil; _____;
296; 300; 5; Beta; EEEEE;
301; 304; 4; Coil; ___;
305; 313; 9; Alpha; HHHHHHHH;
314; 318; 5; Coil; _____;
319; 322; 4; Beta; EEEE;
323; 323; 1; Coil; _;
324; 349; 26; Alpha; HHHHH____HHHHHHH_____HHH;
350; 364; 15; Coil; _____;
```

PDBs Contain More Data!

How do we dope the parse with a 3D barrel fit?
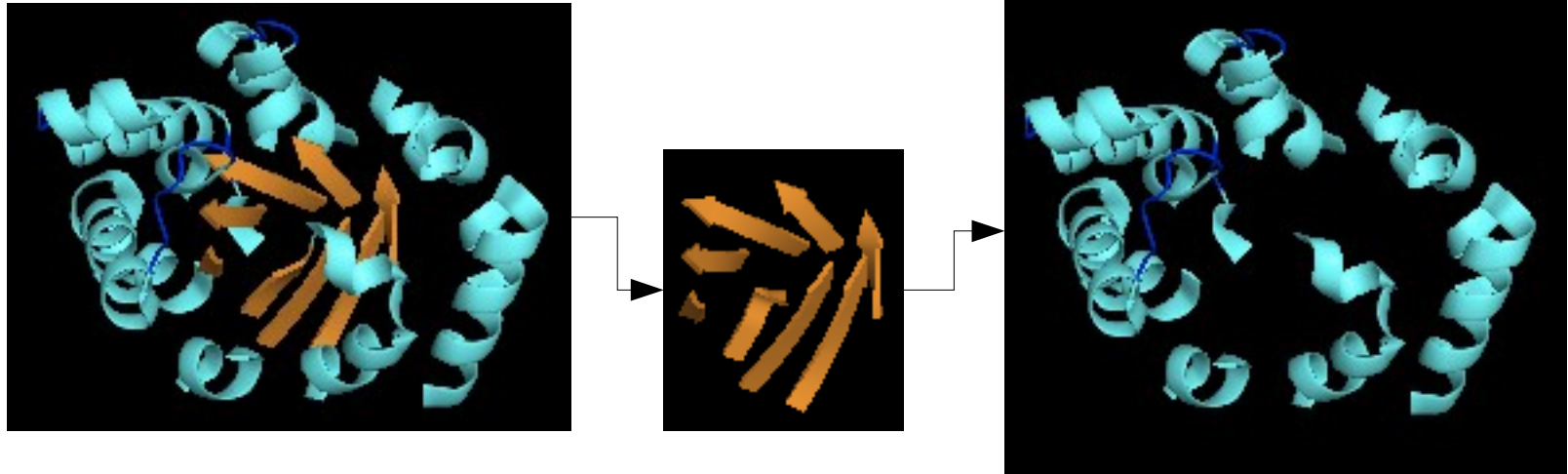
Algorithm:
- Select all residues belonging to beta sheets from a PDB
- Find the sheets that belong to the best fit barrel
  - For each combination of sheets:
    - Perform simplex algorithm to fit sheets to cylinder
  - Best fit cylinder has the smallest residuals
- Expand selection to discover residues belonging to helices

Advantages and Drawbacks

- Advantage:

  - Method can be extended to trefoil and other objects that can be projected to two-space

  - An extension exists to extend this to things that can't be projected – like sandwiches

- Drawback:

  - Unable to cope with 1YBE, our broken friend.

Presentation Midpoint

Any questions before moving onto Foil-Sensitive MSA?

- **Structure Progress**

  - Learned the math needed

  - Now need to code this

  - Now need to adapt PDB data to fit the algorithm

- **Sequence Progress**

  - Deployed code to create guide tree

  - Deployed code to create profiles

  - Deployed nearest neighbour code

  - Now need to adapt Needleman-Wunsch algorithm for our partial profiles

End of Structure Half of Presentation