# Protein structure

**Primary structure – sequence**

PAH   GQETSYIEDNSNQNGAISLIFSLKEEVGALAKVLRLFEENDINLTHIESRPSRLNKDEYEFFTYLDKRTKPVLGSIIKSLRNDIGATVHELSRDKEKNTV
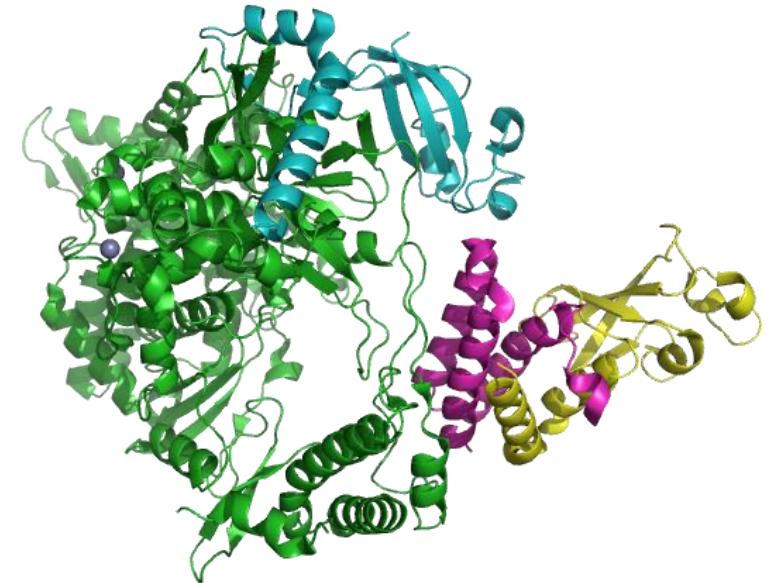
**Secondary structure – α-helix, β-strand, coil**

GQETSYIEDNSNQNGAISLIFSLKEEVGALAKVLRLFEENDINLTHIESRPSRLNKDEYEFFTYLDKRTKPVLGSIIKSLRNDIGATVHELSRDKEKNTV
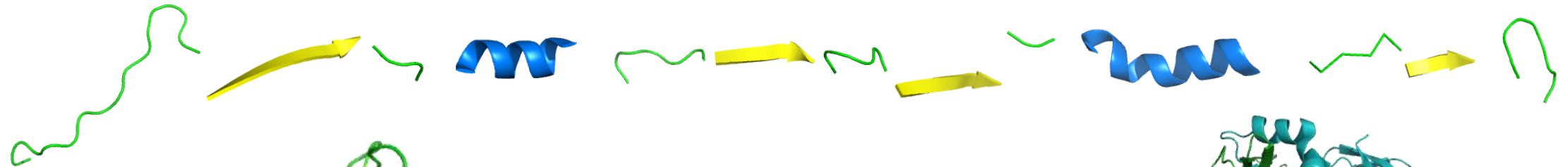
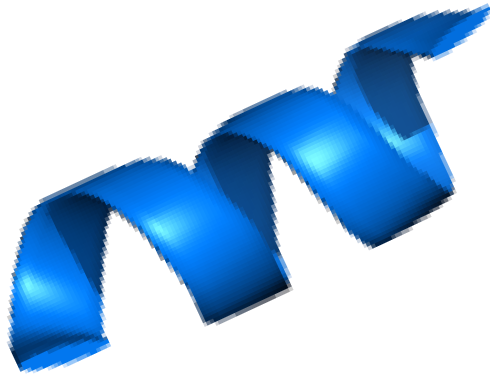**Tertiary structure – folded**
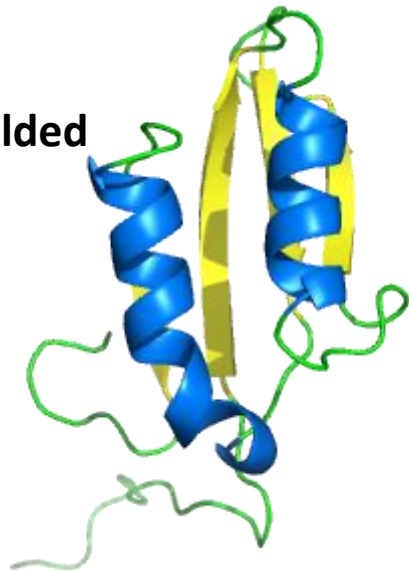
**Quaternary structure – protein complex**

- # PSIPRED – predict secondary structure!
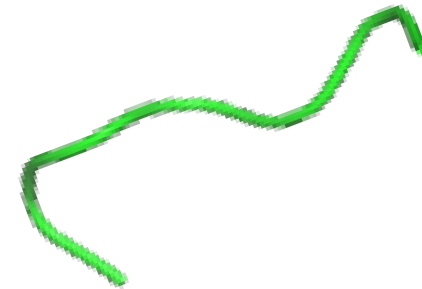
3 states:

Alpha helix                Beta strand             Coil

*How does it work?*
PSIPRED is an artificial neural network (machine learning) trained with lots of sequence data with known secondary structure. PSIPRED has learned to recognize the sequence patterns that form a specific secondary structure.

# To the server!
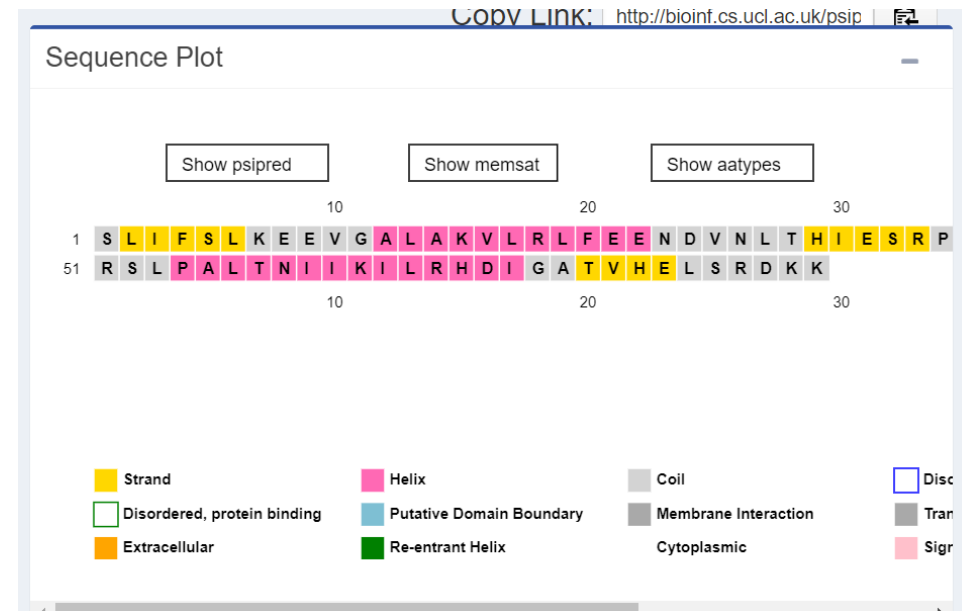
LOOP  HELIX  STRAND
LOOP  HELIX  STRAND

Experiment  SLIFSLKEEVGALAKVLRLFEENDVNLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKK

PSIPRED    SLIFSLKEEVGALAKVLRLFEENDVNLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKK

36  41  46  51  56  61  66  71  76  81

SLIFSLKEEVGALAKVLRLFEENDVNLTHIESRPSRLKKDEYEFFTHLDKR

1  S L I F S L K E E V G A L A K V L R L F E E N D V N L T H I E S R P S R L K K D E Y E F F T H L D K  50

51  R S L P A L T N I I K I L R H D I G A T V H E L S R D K K  79

40  50

86  91  96  101  106  111

KRSLPALTNIIKILRHDIGATVHELSRDKKK

Strand
Disordered, protein binding
Extracellular

Helix
Putative Domain Boundary
Re-entrant Helix

Coil
Membrane Interaction
Cytoplasmic

Disordered
Transmembrane Helix
Signal Peptide

```
>Binary (1 = correct, 0 = incorrect)
0111110111011111111111110111110111101111111111110000000111111111111110110011001111
>Experiment
SLIFSLKEEVGALAKVLRLFEENDVNLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKK
>PSIPRED
SLIFSLKEEVGALAKVLRLFEENDVNLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKK
```

Number of 1: 61 correct

Number of 0: 18 incorrect

Accuracy = 61/(61+18)=77.2%

For Helix

1=True Negative          2= True Positive          3 = False Negative          4 = False Positive

```
>4state
1111111111132222222222231111111111111111111111111111333332222222222222241111111111111
>Experiment
SLIFSLKEEVGALAKVLRLFEENDVNLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKK
>PSIPRED
SLIFSLKEEVGALAKVLRLFEENDVNLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKK
```

$$Specificity = \frac{TN}{TN + FP}$$

$$Sensititity = \frac{TP}{TP + FN}$$

For Helix

1=True Negative => 47  2= True Positive => 24  3 = False Negative => 7 4 = False Positive => 1

```
>Binary
111111111132222222222231111111111111111111111111111133333222222222224111111111111
>Experiment
SLIFSLKEEVGALAKVLRLFEENDVNLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKK
>PSIPRED
SLIFSLKEEVGALAKVLRLFEENDVNLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKK
```

$$Specificity = \frac{TN}{TN + FP}$$

TN = 47, FP = 1 → Specificity = 47/48 = 97.9%  (of those that should not be predicted to be helix, almost none are. This means very few false positives.)

$$Sensititity = \frac{TP}{TP + FN}$$

TP = 24, FN = 7 → Sensitivity = 24/31 = 77.4%  (of those that should be predicted as helix, a few were missed. This means there were false negatives)

While the specificity is great, the sensitivity can be improved. Typically, a predictor must be benchmarked against a much larger dataset, but these values are in the "ballpark" for how accurate PSIPRED is (approaching 80-84% accuracy).