

Sara Kemmler 5760949

Robin Bonkaß 5769588

| 1 | 2 | 3 | $\Sigma$ |
|---|---|---|----------|
|   |   |   |          |

## Übungsblatt Nr. 12

(Abgabetermin 28.07.22)

### Aufgabe 1

SOV und  $Q_3$  für  $S1_{pred}$ :

$$S_e = S_{obs} = \overbrace{CCC}^{se_1} \underbrace{EEEEEE}_{se_2} \overbrace{CCC}^{se_3} \underbrace{EEEEEEE}_{se_4} \overbrace{CC}^{se_5}$$

$$S_p = S_{pred} = \underbrace{CCCC}_{sp_1} \overbrace{EEE}^{sp_2} \underbrace{CCCCCC}_{sp_3} \overbrace{EEEE}_{sp_4} \underbrace{CCC}_{sp_5}$$

$$S(C) = \{(se_1, sp_1), (se_3, sp_3), (se_4, sp_4), (se_5, sp_5)\}$$

$$S(E) = \{(se_2, sp_2), (se_4, sp_4)\}$$

$$S'(C) = \{ \}$$

$$S'(E) = \{ \}$$

$$N(C) = 3 + 3 + 2 = 8$$

$$N(E) = 6 + 6 = 12$$

$$N = 20$$

|              | minov | maxov | $\delta$ | $len(se_i)$ |
|--------------|-------|-------|----------|-------------|
| $se_1, sp_1$ | 3     | 4     | 1        | 3           |
| $se_2, sp_2$ | 3     | 6     | 1        | 6           |
| $se_3, sp_3$ | 3     | 6     | 1        | 3           |
| $se_4, sp_4$ | 4     | 6     | 2        | 6           |
| $se_5, sp_5$ | 2     | 3     | 1        | 2           |

$$SOV_1 = \left[ \frac{100}{20} \cdot \left( \frac{3+1}{4} \cdot 3 + \frac{3+1}{6} \cdot 6 + \frac{3+1}{6} \cdot 3 + \frac{4+2}{6} \cdot 6 + \frac{2+1}{3} \cdot 2 \right) \right]$$

$$= 5 \cdot (3 + 4 + 2 + 6 + 2) = 5 \cdot 17 = 85 \%$$

$$Q_3(S1_{pred}) = \frac{15}{20} \cdot 100\% = 75\%$$

SOV und  $Q_3$  für  $S2_{pred}$ :

$$S_e = S_{obs} = \overbrace{CCC}^{se_1} \underbrace{EEEEEE}_{se_2} \overbrace{CCC}^{se_3} \underbrace{EEEEEE}_{se_4} \overbrace{CC}^{se_5}$$

$$S_p = S2_{pred} = \underbrace{CCC}_{sp_1} \underbrace{CEE}_{sp_2} \underbrace{EEE}_{sp_3} \underbrace{CCC}_{sp_4} \underbrace{EEE}_{sp_5} \underbrace{C}_{sp_6} \underbrace{EE}_{sp_7} \underbrace{C}_{sp_8} \underbrace{C}_{sp_9}$$

$$S(C) = \{(se_1, sp_1), (se_3, sp_5), (se_5, sp_{11})\}$$

$$S(E) = \{(se_2, sp_2), (se_2, sp_4), (se_4, sp_6), (se_4, sp_8), (se_4, sp_{10})\}$$

$$S'(C) = \{sp_3, sp_7, sp_9\}$$

$$S'(E) = \{\}$$

$$N(C) = 3 + 3 + 2 = 8$$

$$N(E) = 6 + 6 + 6 + 6 + 6 = 30$$

$$N = 38$$

|                 | minov | maxov | $\delta$ | $len(se_i)$ |
|-----------------|-------|-------|----------|-------------|
| $se_1, sp_1$    | 3     | 3     | 0        | 3           |
| $se_3, sp_5$    | 2     | 4     | 1        | 3           |
| $se_5, sp_{11}$ | 2     | 2     | 0        | 2           |
| $se_2, sp_2$    | 1     | 6     | 0        | 6           |
| $se_2, sp_4$    | 3     | 6     | 1        | 6           |
| $se_4, sp_6$    | 2     | 7     | 1        | 6           |
| $se_4, sp_8$    | 1     | 6     | 0        | 6           |
| $se_4, sp_{10}$ | 1     | 6     | 0        | 6           |

$$SOV_2 = \left[ \frac{100}{38} \cdot \left( \frac{3+0}{3} \cdot 3 + \frac{2+1}{4} \cdot 3 + \frac{2+0}{2} \cdot 2 + \frac{1+0}{6} \cdot 6 \right. \right. \\ \left. \left. + \frac{3+1}{6} \cdot 6 + \frac{2+1}{7} \cdot 6 + \frac{1+0}{6} \cdot 6 + \frac{1+0}{6} \cdot 6 \right) \right]$$

$$= \frac{100}{38} \cdot (3 + \frac{9}{4} + 2 + 1 + 4 + \frac{18}{7} + 1 + 1)$$

$$= \frac{100}{38} (16,82) = 44,26\%$$

$$Q_3(S2_{pred}) = \frac{15}{20} \cdot 100 = 75\%$$

Die beiden Werte für  $Q_3$  sind gleich, obwohl die beiden vorhergesagten Strukturen unterschiedlich sind. Das liegt daran, dass bei der Berechnung von  $Q_3$  keine Nachbarschaftsbeziehungen mit einfließen. Die Berechnung von SOV hingegen beruht auf dem Vergleich von Paaren von Sekundärstruktursegmenten. Dadurch ist die Berechnung genauer, weshalb man für die unterschiedlichen Vorhersagen auch unterschiedliche Werte berechnet. Da bei der ersten Vorhersage ein deutlich höherer Wert berechnet wurde entspricht diese Vorhersage mehr der beobachteten Realität.

8

## Aufgabe 2

a)

Die Vorhersage der Sekundärstruktur durch PHD befindet sich in der Datei `secStructure_PHD.txt`

b)

Die Vorhersage der Sekundärstruktur durch Proteus2 befindet sich in der Datei `secStructure_proteus.txt`

c)

Das Python Programm, welches sich in der Datei `Sara_Kemmler_Robin_Bonkass_A12.py` befindet, kann mit folgendem Befehl `python3 Sara_Kemmler_Robin_Bonkass_A12.py -ftrue trueSecStructure.txt -fphd secStructure_PHD.txt -fproteus2 secStructure_proteus.txt` ausgeführt werden. Zum kopieren:

```
python3 Sara_Kemmler_Robin_Bonkass_A12.py -ftrue trueSecStructure.txt -fphd secStructure_PHD.txt -fproteus2 secStructure_proteus.txt
```

Die Ausgabe erfolgt in der Konsole und sieht wie folgt aus:

```
1 Q3 of the true secondary structure and the PHD prediction:
2 75.16339869281046%
3
4 Q3 of the true secondary structure and the Proteus2 prediction:
5 88.56209150326796%
```

d)

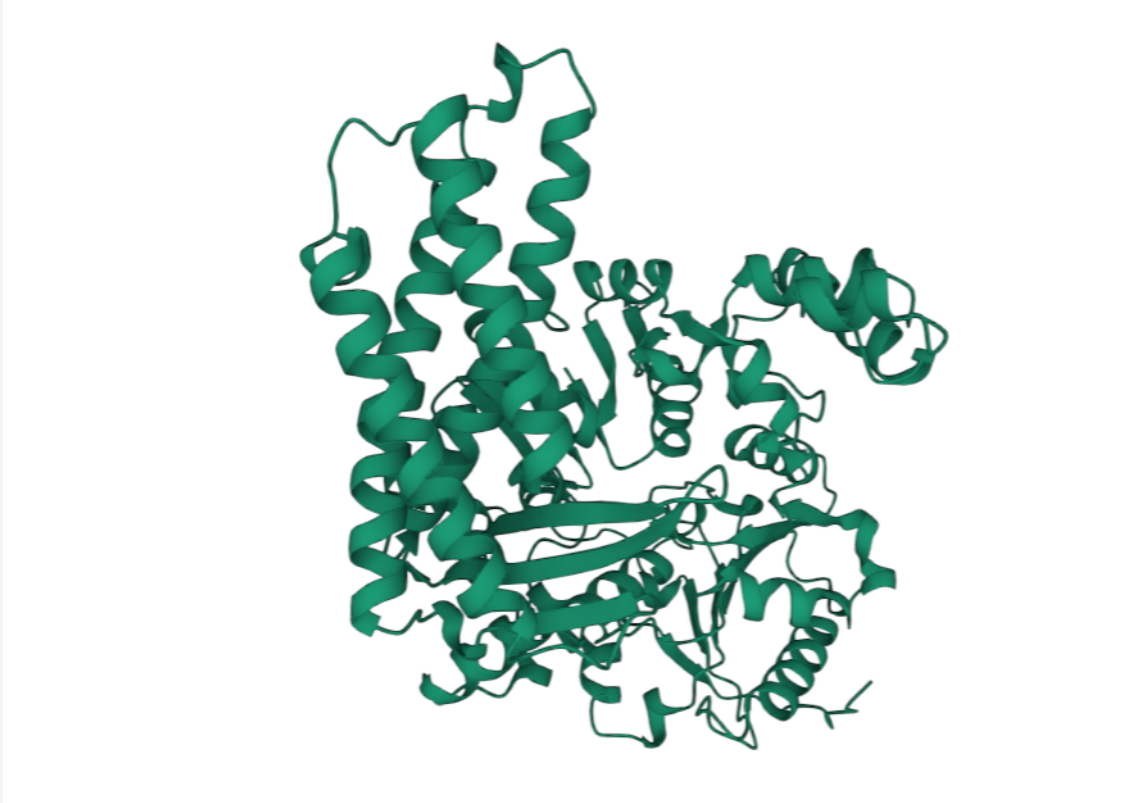
Da der  $Q_3$ -Wert bei der Vorhersage durch Proteus2 größer ist, kann man annehmen, dass diese Methode genauer ist. Da bei der Berechnung von  $Q_3$  jedoch keine Nachbarschaftsbeziehungen berücksichtigt werden, kann man sich nicht zu 100% sicher sein, ob Proteus2 genauer als PHD ist.

8

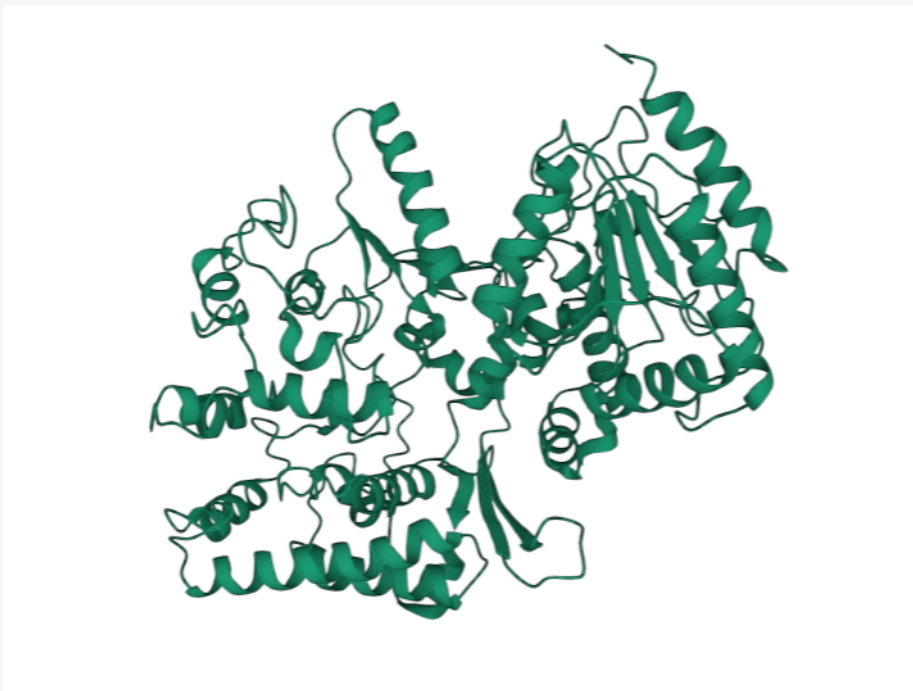
## Aufgabe 3

a)

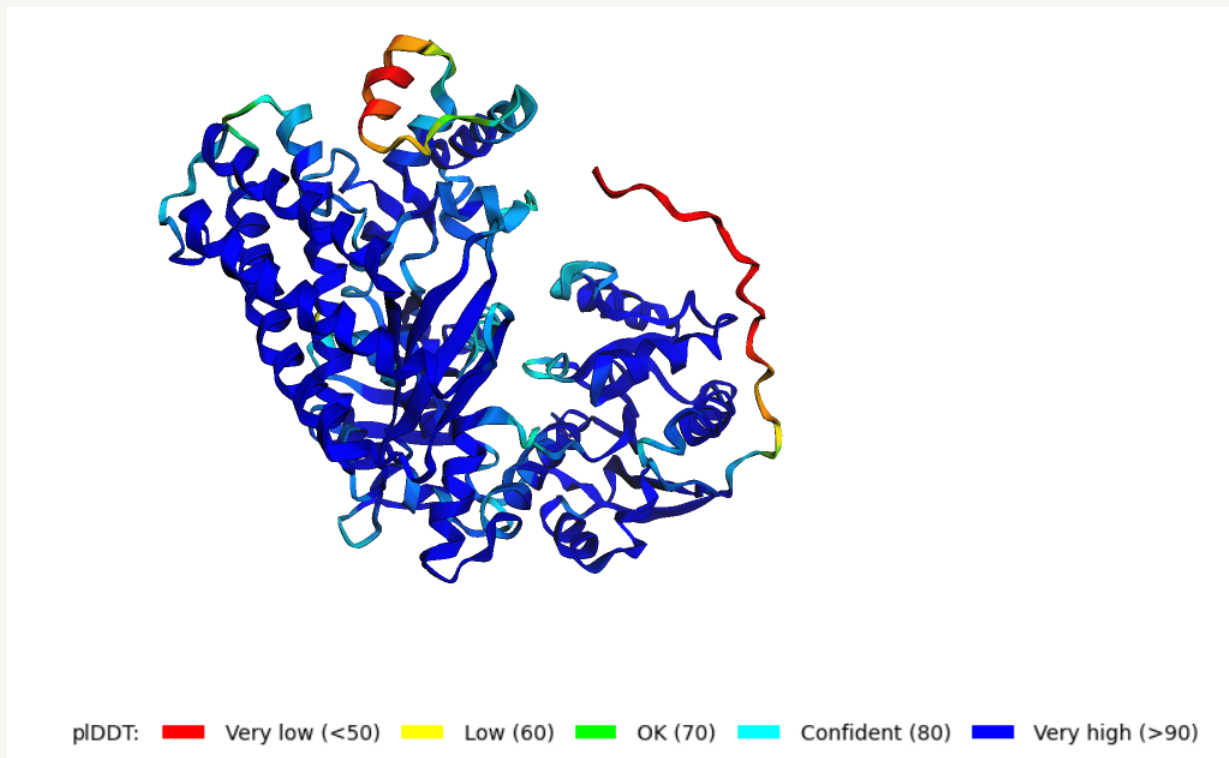
3D-Struktur erstellt mit dem phyre-Server:



3D-Struktur erstellt mit iTasser:



3D-Struktur erstellt mit AlphaFold2:



Da die Strukturen aus unterschiedlichen Winkeln gescreenshotet wurden, kann man nur sehr schwer Gemeinsamkeiten erkennen. Allerdings kann man beobachten, dass sich in ihrer Struktur jeweils dreimal drei Faltblätter in näherer Nachbarschaft befinden.

Funktion des Proteins: Man kann erkennen, dass sich in der Mitte des Proteins eine Kerbe befindet in der sich beispielsweise ein anderes Protein oder die DNA/RNA anlagern kann. Das Protein kann dann in Interaktion mit seinem Liganden treten.