

TØ 5 – Proteinstruktur 2

Table of contents

1 Opgave 1. Cis-peptider	2
1.1 Beregn cis Xaa-Pro frekvens per protein	2
1.2 Beregn andre cis-peptidbindingers frekvens	2
2 Opgave 2. CATH	3
2.1 Identificer klassen med flest domæner	3
2.2 Find barrel-arkitekturer i klasserne	3
2.3 Find barrel-arkitekturen med flest domæner	3
2.4 Beskriv klassisk β -barrel foldning	3
2.5 Sammenlign CATH:3.20.20 og CATH:2.40.10	3
3 Opgave 3. Proteinfoldning	4
3.1 Undersøg dss-kommandoen	4
3.2 Tæl domæner i EF-Tu	4
3.3 Angiv foldningsklasse for hvert domæne	4
3.4 Tæl β - α - β -motiver i domæne 1	4
3.5 Tæl β -hairpins i domæne 2	4
3.6 Beskriv foldning af domæne 3	4
3.7 Identificer nukleotidbindende domæne	4
4 Opgave 4. Sekvens og struktur	5
4.1 Bestem foldningsklasse for PGP-I	5
4.2 Beskriv PGP-I's kvaternære struktur	5
4.3 Tegn et helix-hjul for helixen	5
4.4 Identificer helixens hydrofobe side	5
4.5 Lav script med surface-visualisering	6
4.6 Tegn topologisk foldningsdiagram	6
4.7 Diskuter katalytisk triade og switchpoint	6
5 Opgave 5. Skattejagt	7
5.1 Find de to proteiners funktion	7
5.2 Find β -galactosidasens molekylvægt	7
5.3 Find PDB-kode og accessionsnummer	7
5.4 Find myoglobins subcellulære placering	7
5.5 Find kromosomlokalisering for proteinerne	7
5.6 Beskriv posttranslationelle modifikationer	7
5.7 Find β -galactosidasens K_M og V_{\max}	7
5.8 Find active site aminosyrer	7
5.9 Find isoelektrisk punkt for proteinerne	7
5.10 Find sekvenser med mere end 50 % identitet	8
5.11 Find allosteriske effektorer af expression	8

1 Opgave 1. Cis-peptider

Højtopløste strukturer i Protein Data Bank blev gennem søgt for *cis*-peptidbindinger. Her fandt man at det gennemsnitlige protein består af 270 aminosyrer samt at frekvensen af prolinrester er 4.8%. Man fandt desuden at frekvensen af *cis* Xaa-Pro (altså forekomsten af et *cis*-peptid lige før prolin ud af alle aminosyrer) er 5.2% og at frekvensen af *cis*-peptider for andre (ikke-proline) rester er 0.029%.

1.1 Beregn *cis* Xaa-Pro frekvens per protein

Baseret på ovenstående, hvor mange *cis* Xaa-Pro bindinger vil du i gennemsnit forvente at finde i hvert protein? Hvis mindre end én, hvor mange proteiner ville man skulle kigge i gennem før man fandt én *cis* Xaa-Pro?

1.2 Beregn andre *cis*-peptidbindingers frekvens

Hvordan ser det ud for andre *cis*-peptider (altså ikke involverende proline)?

2 Opgave 2. CATH

Der findes flere forskellige klassifikationer af observerede proteinstrukturer:

CATH (<http://www.cathdb.info/>)

SCOP (<http://scop2.mrc-lmb.cam.ac.uk/>)

DALI (<http://ekhidna2.biocenter.helsinki.fi/dali/>)

I denne opgave vil vi bruge CATH, der klassificerer domæner i fire niveauer: Foldningsklasse (Class), Arkitektur, Topology og Homology.

Gå til CATH hjemmesiden og vælg “Browse” i øverste linie.

2.1 Identificer klassen med flest domæner

Hvilken af de fire klasser indeholder flest domæner? Diskutér.

Expandér klasserne og kig efter arkitekturer der betegnes som “barrel”.

2.2 Find barrel-arkitekturer i klasserne

I hvor mange klasser og i hvilke klasser findes der barrel-arkitekturer?

2.3 Find barrel-arkitekturen med flest domæner

Hvilken barrel-arkitektur indeholder flest domæner?

2.4 Beskriv klassisk β -barrel foldning

Beskriv foldningen i den klassiske β -barrel.



Tip

Se figur 1.58 i Petsko & Ringe.

2.5 Sammenlign CATH:3.20.20 og CATH:2.40.10

Hvilke typer barrels tilhører foldningerne CATH:3.20.20 og CATH:2.40.10? Beskriv forskellen mellem de to foldninger.

3 Opgave 3. Proteinfoldning

I denne opgave skal vi bruge PyMOL til at kigge på strukturen af elongeringsfaktor EF-Tu.

Kig på filen `EF-Tu-folding.pml`, nogle linier nede i scriptet vil du finde linjen; `dss /left`

3.1 Undersøg `dss`-kommandoen

Hvilken funktion har kommandoen `dss` (Hint: se <https://pymolwiki.org/index.php/Category:Commands>)

Copy/Paste scriptet fra `EF-Tu-folding.pml` ind i PyMOL.

Tryk **F1** for at få et overblik over strukturen.

3.2 Tæl domæner i EF-Tu

Hvor mange domæner består proteinet af? Forklar hvordan det ses.

3.3 Angiv foldningsklasse for hvert domæne

Angiv foldningsklasse for hvert enkelt domæne.

3.4 Tæl β - α - β -motiver i domæne 1

Tryk **F2** for at undersøge domæne 1. Hvor mange β - α - β motiver består domæne 1 af?

3.5 Tæl β -hairpins i domæne 2

Tryk **F3** for at undersøge domæne 2. Hvor mange β -hairpins findes der i domæne 2?

3.6 Beskriv foldning af domæne 3

Tryk **F4** for at undersøge domæne 3. Beskriv foldningen af domæne 3.

3.7 Identificer nukleotidbindende domæne

Hvilket domæne vil du forvente binder nukleotid? Begrund dit svar.

4 Opgave 4. Sekvens og struktur

PyMOL-scripting opgave: I denne opgave skal i lære at bruge surface visualiseringen til strukturel analyse.

Visse hormoner af peptidtypen har en pyroglutamylrest som N-terminal, hvorved hormonet er beskyttet mod nedbrydning af sædvanlige aminopeptidaser. Pyroglutamyl peptidase I (PGP-I) er en protease der kan spalte peptidbindingen efter en N-terminal pyroglutamyl rest. PGP-I er således væsentlig for kontrollen af disse hormonsignaler. Enzymet findes dog også hos bakterier. Formålet hos bakterier er ukendt, men det er blevet foreslået at det skulle være del af en forsvarsstrategi mod nedbrydning af bakterien i fordøjelsessystemet hos dyr. Gastrin, der stimulerer udskillelse af fordøjelsesenzymer, er netop et hormon af peptidtypen med en N-terminal pyroglutamylrest. Cystein proteaser har en katalytisk triade bestående af Cys, His og i dette tilfælde Glu, men i andre cysteine proteaser ses i nogen tilfælde en Asp i stedet for Glu. I PGP-I fra *Bacillus amyloliquefaciens* har de tre aminosyrerester sekvensnumrene E81, C144 og H168.

Strukturen af *B.amyloliquefaciens* PGP-I (PDB-ID: 1AUG) kan findes i Protein Data Bank.

4.1 Bestem foldningsklasse for PGP-I

Hvilken foldningsklasse tilhører PGP-I og hvor mange domæner består protomeren af.

4.2 Beskriv PGP-I's kvaternære struktur

Beskriv PGP-I's kvaternære struktur og beskriv hvilke(n) symmetri(er) der er mellem de enkelte protomerer. (Hint: Petsko og Ringe, side 40-45)

Cysteine fra den katalytiske triade i PGP-I er N-terminal i en helix som har sekvensen

FVCNHLFYGLMDEISRHHHP

Anvend EMBOSS (f.eks. på <http://emboss.bioinformatics.nl/cgi-bin/emboss/pepwheel>) til at tegne et helix hjul.

4.3 Tegn et helix-hjul for helixen

Diskuter fordelingen af aminosyrer på hjulet med hensyn til deres fysisk/kemiske egenskaber.

4.4 Identificer helixens hydrofobe side

Hvilken side af helixen forventer du vender ind mod kernen af proteinet?

4.5 Lav script med surface-visualisering

PyMOL Info

Surface kommandoen bruges ofte i situationer, hvor man vil vise pasform interaktioner mellem ligand og protein. Funktionen viser Connolly overfladen, også kaldet den tilgængelige overflade, som viser alle de teoretisk mulige steder vand kan befinde sig omkring proteinet. Da afgrænsningen sættes ved vandets oxygenatoms centrum, vil det sige, at der godt kan ske overlap mellem ligandens og proteinets Connolly overflade, men kun svarende til oxygens radius. I kan læse mere om surface funktionen og Connolly overfladen på; https://en.wikipedia.org/wiki/Accessible_surface_area.

Lokaliser helixen i strukturen og vis kun den i surface repræsentation og farvelæg efter atom type. Du vil bemærke, at hvis du har lavet en selektion, der omfatter helixen, er det stadigvæk kun overfladen mod solventet, der bliver vist. For at få vist overfladen af hele helixen er du nødt til at definere et nyt object kun bestående af helixen.

Lav et script, der definerer helixen som et nyt objekt, lav en passende visualisering og gem scenen som F1. Se nu på overfladen af helixen (show surface) og undersøg hvordan overfladen ind mod kernen af proteinet adskiller sig fra den side der vender mod solvent og sammenlign med helix hjulet fra tidligere. Vender helixen som forventet jvf spørgsmål 4?

4.6 Tegn topologisk foldningsdiagram

Tegn et topologisk foldningsdiagram og lokalisér “switch point” (jvf P&R side 38-39). Lokalisér også switchpoint i PyMOL og lav en scene kaldet F2, der visualiserer dette.

Tip

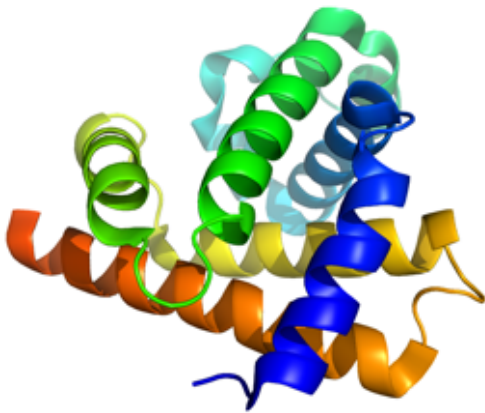
Brug `spectrum` kommandoen (`spectrum count, rainbow, 1AUG`) på én af kæderne.

4.7 Diskuter katalytisk triade og switchpoint

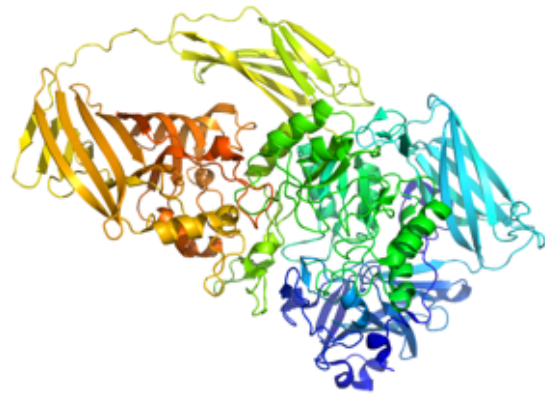
Diskuter placeringen af den katalytiske triade i forhold til “switch point” og det forventede substrat bindingssite. Lav en ny scene, F3, som viser active site og dens placering ift. Switchpoint.

5 Opgave 5. Skattejagt

Vi har i en tidligere opgave stiftet bekendtskab med følgende to proteiner:



(a) Human myoglobin



(b) *E. coli* β -galactosidase

Figure 1: De to proteiner

Brug dit kendskab til online databaser til at besvare følgende spørgsmål

5.1 Find de to proteiners funktion

Hvad er funktionen af de to proteiner i cellen og i hvilke sammenhænge optræder de?

5.2 Find β -galactosidasens molekylvægt

Hvad er molekylvægten af β -galactosidase og hvor mange kæder består det aktive enzym af?

5.3 Find PDB-kode og accessionsnummer

Hvilken PDB-kode (struktur) og accessionsnummer er registreret for myoglobin?

5.4 Find myoglobins subcellulære placering

Hvilken subcellulære position findes myoglobin i?

5.5 Find kromosomlokalisering for proteinerne

Benævn chromosompositionen for proteinerne. Forklar hvorfor β -galactosidase er lidt anderledes.

5.6 Beskriv posttranslationelle modifikationer

Beskriv de to proteiners posttranscriptionelle modifikationer.

5.7 Find β -galactosidasens K_M og V_{\max}

Hvad er K_M -værdien for β -galactosidase overfor det stærkest bindende substrat, man har målt? Hvad er V_{\max} for enzymet med dette substrat?

5.8 Find active site aminosyrer

Find aminosyrerne der er den del af active site for begge proteiner.

5.9 Find isoelektrisk punkt for proteinerne

Hvad er det isoelektriske punkt (pI) for henholdsvis myoglobin og beta-galactosidase?

5.10 Find sekvenser med mere end 50 % identitet

Hvor mange sekvenser i Uniprot har mere end 50 % sekvens lighed med beta-galactosidase?

5.11 Find allosteriske effektorer af expression

Find allosteriske effektorer af expression for β -galactosidase.

