**TØ 12 - Coronavirus**

[**1. Protease 1**](#_Toc83996336)

[**2. Protease dimer 2**](#_Toc83996337)

[**3. Protease alignment 3**](#_Toc83996338)

[**4. RdRp og Remdesivir 5**](#_Toc83996339)

[**5. Spike og ACE2 6**](#_Toc83996340)

[**6. Frameshift element 8**](#_Toc83996341)

## 1. Protease

I denne opgave skal I lave et PyMOL script til at undersøge SARS-CoV-2 hovedprotease 3CL (PDB-ID: 6YB7). Proteinet består af tre distinkte domæner; domæne I, som består af alpha-beta struktur, domæne II, som består af et beta sheet og domæne III, som kun består af alpha-helicer.

1. Lav en scene, kaldet F1, der tydeligt viser proteinet sekundære struktur og en scene, kaldet F2, der viser hver af de tre domæner (I, II og III) i hver sin farve.   
   Definér kort et domæne. Hvor mange helicer er der i domæne III?
2. Lav en scene, kaldet F3, som tydeligt viser forløbet fra N- til C-terminal. Hint: kig på [PyMOL-wiki](https://pymolwiki.org/index.php/Spectrum) og se hvordan man bruger spectrum.  
   I hvilket domæne binder N-terminus?
3. Lav en scene, kaldet F4, der viser enzymets active site med aminosyrerne Glu166, Cys145 og His41. Hvor på proteinet er enzymets active site lokaliseret?
4. Hvad er afstanden mellem S-atomet på Cys og NE2 på His i det aktive site? Hvilken type binding er det? Lav en scene, kaldet F5, som viser dette.

## 2. Protease dimer

I denne opgave skal I bygge videre på jeres PyMOL script fra forrige opgave. 3CL proteasen, der også kaldes Nsp5, er en homodimer og I skal i denne opgave se på hvordan dimerisering af enzymet er nødvending for den katalytiske aktivitet.

**PyMOL-hint:** *For at se dimer versionen af 3CL fra PDB databasen kan I hente en "biologisk samling" ved at bruge fetch med type=pdb1. En "*[*biologisk samling*](https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/biological-assemblies)*" består af de asymmetriske enheder (ASU) gemt i forskellige states. For at lave de to monomerer til individuelle objekter bruges split\_states, hvorefter kædenavn for den ene monomer kan ændres med alter. Dette vil blive dækket i mere detalje i TØ 21 om strukturmetoder, så for nu skal I bare anvende følgende kode i jeres eget script:*

reinit  
fetch 6yb7, type=pdb1, async=0  
split\_states 6yb7  
alter 6yb7\_0002, chain='B'  
remove 6yb7  
remove solvent  
remove resn DMS

**PyMOL-hint***: Når man i sit script gerne vil gå tilbage til en defineret scene, f.eks. F2, skal man bruge kommandoen scene F2, recall.*

1. Lav en scene, kaldet F6, der viser proteasen som en dimer. Brug F2 repræsentationen af den ene monomer og farv den anden grå.  
   Hvilke(t) af de tre domæner er involveret i dimerisering?
2. Hvor er active site lokaliseret i relation til dimer interface? Lav en scene, kaldet F7, der viser dette.
3. Identificér en interaktion af hver af følgende typer omkring E166: hydrogenbinding, saltbro, pi-stacking, hydrofobe interaktioner. Lav en scene, kaldet F8, der viser dette.
4. Forskerne der har studeret denne struktur skriver at "Dimerisering af enzymet er nødvending for den katalytiske aktivitet" (Zhang et al., 2020, doi: [**10.1126/science.abb3405**](https://doi.org/10.1126/science.abb3405)). Kan du forklare hvorfor på baggrund af din analyse af interaktionerne?
5. Hvordan tror du den første Nsp5 dimer dannes lige efter infektion med virus? Hint: Se Chen et al., 2010 (doi: [**10.1007/s13238-010-0011-4**](https://dx.doi.org/10.1007%2Fs13238-010-0011-4)).

## 3. Protease alignment

I denne opgave skal vi først lave sekvens-alignment af Nsp5:

* Download text filen Nsp5-align.fasta.txt, som findes på denne uge på Brightspace, der indeholder et multiple alignment af følgende 10 Nsp5 sekvenser:
  1. SARS CoV-2 (severe acute respiratory syndrome coronavirus type 2)
  2. SARS-CoV (severe acute respiratory syndrome coronavirus)
  3. MERS (Middle East respiratory syndrome coronavirus)
  4. PEDV (Porcine Epidemic Diarrhea Virus)
  5. TGEV (Porcine Transmissible Gastroenteritis Virus)
  6. HKU4 (Tylonycteris bat coronavirus HKU4)
  7. A59 (Mouse hepatitis virus A59)
  8. HKU1 (Human coronavirus HKU1)
  9. FIPV (Feline infectious peritonitis virus)
  10. NL63 (Human coronavirus NL63)
* Åben også protease struktur for SARS-CoV (PDB-ID: 1Q2W), SARS-CoV-2 (PDB ID: 6YB7) og MERS (PDB-ID: 4YO9) Nsp5.   
  Vi arbejder først på SARS-CoV-2 kæde A, så gem resten indtil videre. Gem dette under scene F1.
* SARS-CoV (PDB-ID: 1Q2W) og MERS (PDB-ID: 4YO9) er allerede navngivet som to kæder, men SARS-CoV-2 (PDB ID: 6YB7) er kun angivet som én kæde. Derfor skal vi bruge split\_states, navngive kæden med alter kommandoen, og kombinere dem til et objekt igen med create:

split\_states 6yb7  
alter 6yb7\_0002, chain='B'  
delete 6yb7  
create 6yb7, (6yb7\_0001 or 6yb7\_0002)  
delete 6yb7\_\*

* Kør programmet colCons (colCons.zip findes på denne uge på Brightspace), som I gjorde det i TØ 7 opgave 5, og brug Nsp5-align.fasta fil som input. Kald scenen F2.

1. Er den indre eller ydre del mest bevaret?
2. Er Cys145, His41 og Glu166, som er vigtige for det aktive site, bevaret? Lav en scene kaldet F3, som viser dette.

I næste omgang laver vi struktur-alignment af Nsp5:

* Align strukturerne for SARS-CoV-2 (PDB-ID: 6YB7) med hhv. SARS-CoV (PDB-ID: 1Q2W) og MERS (PDB-ID: 4YO9). Kald alignment af SARS-CoV-2 og SARS-CoV scene F4 og aligment af SARS-CoV-2 og MERS for scene F5.

1. Hvad er RMSD mellem de to par af strukturer? Hvilken protease-struktur er mest forskellig fra SARS-CoV-2?
2. Identificér active site og sammenlign det i de to alignede sekvenser. Lav en scene, kaldet F6, som viser dette. Hint: active site i MERS har ikke samme residue numre (resi) som SARS-CoV og SARS-CoV-2.

## 4. RdRp og Remdesivir

SARS-CoV-2 har en RNA-afhængig RNA-polymerase (RdRp), der hjælper med at kopiere det 30 kb store RNA-genom. Via en speciel template-hoppe-mekanisme kan RdRp også danne kortere mRNAer, der udtrykker strukturelle og hjælpe proteiner.

* Åben PDB-filen 6yyt i PyMOL og farvelæg nsp7 (lilla), nsp8 (blå) og nsp12 (cyan). Hint: For at finde ud af hvilke kæder der svarer til hvilke proteiner kan du gå ind på [**RCSB.org**](http://RCSB.org) og søge på 6YYT.

1. Hvilken RNA-streng er template-strengen og hvilken er produkt-strengen? Hvilken vej bevæger RdRp sig i forhold til RNA dobbelthelicen?

Da vores celler ikke syntetiserer RNA fra RNA er RdRp et attraktivt mål for udvikling af antivirale stoffer, men de tager langtid at udvikle og sikkerhedsgodkende. Remdesivir er en polymerase-hæmmer der blev udviklet mod hepatitis og Ebola, men som ikke var særlig effektiv. Det er nu godkendt til brug i nødstilfælde mod SARS-CoV-2.

* Åben PDB-filen 7BV2 i PyMOL, der indeholder RdRp bundet til Remdesivir. Farvelæg template-streng (rød) og produkt-streng (gul). Vis proteinerne i "ribbon" og RNA og Remdesivir som sticks. Zoom ind på active site.

1. Beskriv hvilke cofaktorer, der medvirker til syntesen.
2. Beskriv den molekylære struktur af Remdesivir og hvordan det hæmmer RNA syntesen.

## 5. Spike og ACE2

*Følgende er en eksamensopgave, der tester jeres kunnen indenfor dette emne. Den kan derfor være mere udfordrende end I er vant til.*  
Spike-proteinet fra ny coronavirus (SARS-CoV-2) består af S1 domænet, der binder til ACE2, og S2 domænet, der rummer fusionsmaskineriet. På figuren herunder er de enkelte aminosyrer farvelagt ud fra deres variation mellem et betydeligt antal forskellige Spike-sekvenser:

Diagram, map

Description automatically generated

1. Forklar hvad skalaen fra "Conserved" til "Variable" betyder og hvorfor S1 domænet er mest variabelt?

S1 domænet indeholder ACE2-bindingsdomænet. Herunder vises Clustal O sekvensalignment af receptor-bindingsmotiv (RBM) for SARS-CoV, SARS-CoV-2 og tæt beslægtet coronavirus fra flagermus:

CLUSTAL O(1.2.4) multiple sequence alignment

SARS-CoV NTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCT-PPALNCYWPLNDYG 59

SARS-CoV-2 NSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYG 60

bat-SL-RaTG13 NSKHIDAKEGGNFNYLYRLFRKANLKPFERDISTEIYQAGSKPCNGQTGLNCYYPLYRYG 60

\*:.::\*:. \*\*:\*\* \*\* :\*:.:\*:\*\*\*\*\*\*\*. :. ...\*\*. .:\*\*\*:\*\* \*\*

SARS-CoV FYTTTGIGYQPY 71

SARS-CoV-2 FQPTNGVGYQPY 72

bat-SL-RaTG13 FYPTDGVGHQPY 72

\* \* \*:\*:\*\*\*

# Percent Identity Matrix - created by Clustal2.1

1: SARS-CoV 100.00 50.70 53.52

2: SARS-CoV-2 50.70 100.00 76.39

3: bat-SL-RaTG13 53.52 76.39 100.00

1. Hvad betyder symbolerne \* : og . angivet under sekvenserne? Hvilken similaritets-score har en S til T og en I til L mutation ifølge Blossum-62 substitutionsmatricen angivet i Stryer Fig. 6.9? Hvilke af de tre sekvenser ligner mest hinanden?

Spike-proteinet består af to dele, S1 og S2. Proteinets hoved udgør S1-delen af spike-proteinet. Receptor-bindingsdomænet (RBD) har to konformationer - en lukket og en åben. I den åbne konfiguration peger domænets ”finger” opad, så den kan berøre ACE2.

Vi skal nu kigge på RBM domænets interaktion med ACE2 receptoren. Åben **Spike-ACE2.pml** som ligger under denne uge på brightspace (det tager lidt tid at loade og aligne strukturerne).

Scene **F1** viser Spike med RBM-domæne i lukket form (PDB-ID: 6VXX), **F2** Spike hvor et af RBM-domænerne er åbnet (PDB-ID: 6VSB), **F3** viser binding af RBM til ACE2 (PDB-ID: 6M17), og endelig zoomer **F4** ind på RBM-domænet.

I den såkaldte engelske variant af SARS-CoV-2 (B.1.1.7) findes følgende ændringer: H69-V70 deletion, Y144-145 deletion, N501Y, A570D, P681H, T716I, S982A, og D1118H. Specielt N501Y og D1118H menes at være vigtig for at B.1.1.7 er mere smitsom.

1. Hvor er N501Y placeret i den ikke-bundne (F1) hhv. bundne form (F3/F4) af Spike? Hvordan kan Tyr-substitutionen resultere i mere effektiv binding?
2. Hvor er D1118H-mutationen placeret og hvordan kan den tænkes at påvirke Spike-proteinets evne til at inficere en celle?

## 6. Frameshift element

*Følgende er en eksamensopgave, der tester jeres kunnen indenfor dette emne. Den kan derfor være mere udfordrende end I er vant til.*  
Umiddelbart efter coronavirus trænger ind i en celle translateres to tredjedele af RNA-genomet til polyproteinerne **pp1a** og **pp1ab**. Polyproteinet pp1ab translateres som en forlængelse af pp1a, hvor en RNA-struktur (*frameshift stimulation element*, FSE) bremser ribosomet og i visse tilfælde fører til et -1 skifte i læserammen, så ribosomet undgår et stopkodon og kan fortsætte translationen af det længere protein. I figuren herunder vises: (a) RNA-strukturdiagram for SARS-CoV-2 FSE vist med stem farver, 3D-struktur vist som cartoon med (b) stem farver og (c) regnbue-farver fra 5' (blå) til 3' (rød).

A picture containing diagram

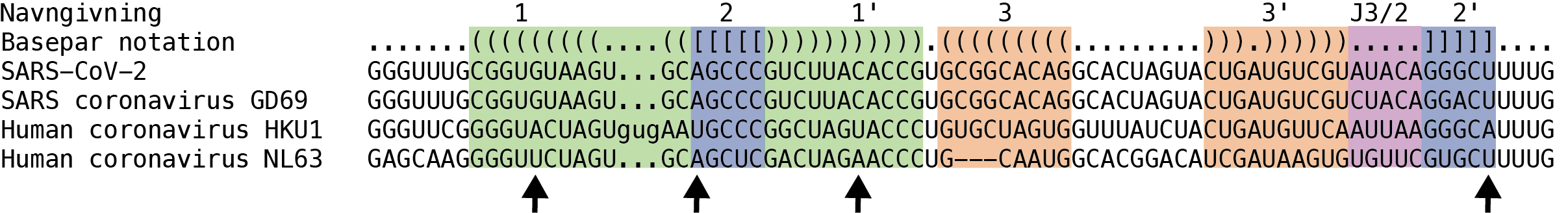
Description automatically generated

1. Beskriv RNA-strukturen af elementerne Stem1, Stem2, Stem3 og J3/2.

3D-strukturen i panel b og c ovenfor giver en mulig forklaring på hvordan ribosomet bremses og foretager et -1 skifte i læserammen. Strukturen kan studeres nærmere i PyMOL ved at hente PDB-filen 6XRZ og følge RNA-kædens forløb.

1. Giv en strukturel / mekanistisk forklaring på hvordan FSE kan blokere ribosomet og skubbe det én position tilbage i læserammen.

Nedenfor ses et sekvens-alignment af FSE-elementet fra SARS-CoV2 med sekvenser fra lignende vira med eksempler på co-varierende baseændringer i Stem1 og Stem2 markeret med røde og fede bogstaver:



1. Hvilke forskellige co-varierende basepar er der på de med pil markerede kolonner i Stem1 og Stem2? Hvorfor ses typisk deletioner/indsættelser på 3 baser i denne RNA-struktur?