

# Sequence alignments

**Tutorial by Dr. René Staritzbichler**

<http://proteinformatics.org> -> Teaching -> Download course material

Workshop 10.8.2020

# Comparing sequences

- The chain of amino acids can be written as string or sequence
- Biochemical properties of proteins are represented in their sequence
- The sequence determines structure and function
- **Similar sequence means similar structure and often similar function**
- Starting point of many modelling techniques
- Many modelling techniques depend crucially on the quality of sequence alignment

# Applications

- Here, we use it to repair the downloaded PDB
  - Erase extra sequences
  - Model missing loops
  - Reverse mutations
- Homology modeling
- Understanding similarities / differences between proteins
  - Identifying important positions in the sequence
  - E.g. all GPCRs share 1 absolutely conserved amino acid, which is crucial to maintain the overall topology / fold

# Get sequence

## 0) Go to course directory

- `'cd course_MD/day_1/homology'`

## 1) Download from PDB website

- go to <https://www.rcsb.org> and download 3sn6.fasta and 3sn6.pdb
- move the downloaded file to '3sn6\_pdb\_web.fa' in current directory

## 2) Extract from coordinate section in .pdb file

- go to <https://github.com/reneDominik/toolbox> and download pdb\_sequences.py
- Execute: `'$ python3 pdb_sequences.py 3sn6.pdb > 3sn6_pdb_coor.fa'`

## 3) Download from uniprot website

- this is the wildtype sequence as found in the organism
- move it to 3sn6\_uniprot.fa
- for the simulations we want a model with this sequence

# Sequence from pdb coordinates

>3sn6:A

TEDORNEEKAOREANKKIEKOLQDKOVYRATHRLLLLGAGESGKSTIVKOKATKVODIKNNLKEA IETIVAAMS NLVPPVELANPENQFRVDYILS  
VMNVPDFDFPPEFYEHAKALWEDEGVRACYERSNEYQLIDCAQYFLDKIDVIKODDYVPSDODLLRCRVSGIFETKFOVDKVNFMFDVGGORD  
ERRKWIOCFNDVTAIIFVASSSYNMTNRLOEALNLFKSIWNNRWLRTISVILFLNKQDLLAEKVLGKSKIEDYFPEFARYTTPEDATPEPGEDPRV  
TRAKYFIRDEFRLISTASGDGRHYCYPHFTCAVDTENIRRVFNDCRDIIQRMHLRQYELL

>3sn6:B

QSELDQLRQEA EQLKNOIRDARKACADATLSOITNNIDPVGRIQMRTTRTLRGHLAKIYAMHWGTD S RLLVSASODGKLIWDSYTTNKVHAIPLR  
SSWVMTCA YAPSGNYVACGGLDNICSIYNLKTREGNVRVSRELAGHTGYLSCCRFLDDNOIVTSSGDTTCALWDIETGOOTTTFTGHTGDVMSL  
SLAPDTRLFVSGACDASAKLWDVREGMCROTFTGHESDINAICFFPNGNAFATGSDDATCRLFDL RADQELMTYSHDNIICGITSVSFSKSGRLLL  
AGYDDFNCNVWDALKADRAGVLAGHDNRV SCLGVTDDGMAVATGSWDSFLKIWN

>3sn6:G

NTASIAQARKLVEQLKMEANIDRIKVS KAAADLMAYCEAHAKEDPLLTPVPA SENPFR

>3sn6:R

NIFEMLRIDEGLRLKIYKDTEGYTTIGIGHLLTKSPSLNAAKSELDKAIGRNTNGVITKDEAEKLFNODVDAAVRGILRNAKLPVYDSLDAVRAALI  
NMVFOMGETGVAGFTNSLRMLLOOKRWDEAAVNLA KSRWYNQTPNRAKRVITTFRTGTWDAEVWVVGMGIVMSLIVLAIVFGNVLVITAIKFER  
LOTVTNYFITSLACADLVMGLAVVPFGAAHILTKTWTFGNFWCEFWTSIDVLCVTAS IETLCVIAVDRYFAITSPFKYQSLLTKNKARVIILMVWIVSG  
LTSFLPIQMHWYROEA INCYAEETCCDFFTNOAYAIASSIVSFYVPLVIMV FVYSRVFQEAQRQLQKIDKSEGRCLKEHKALKTLGIIMGTFTLCWLPF  
FIVNIVHVIQDNLIRKEVYILLNWIGYVNSGFNPLIYCRSPDFRIAFQELLC

>3sn6:N

OVOLQESGGGLVOPGGSLRLSCAASGFTFSNYKMNWVRQAPGKGLEWVSDISQSGASISYTG SVKGRFTISRDNANTLYLQMNSLKPEDTAV  
YYCARCPAPFTRDCFDVTSTTYAYRGQGTQVTVSS

# Align membrane sequences with AlignMe

- Alignments match equal or similar motifs to each other
- Membrane proteins are notoriously difficult to align
- AlignMe is optimized for but not limited to membrane proteins
- Go to website by using search machine or clicking <http://www.bioinfo.mpg.de/AlignMe/>
- Select ,Sequence to sequence alignment‘
- Compare the three sequences (in pairs)
- Select ,Fast‘ mode and submit
  
- Download results and move to current directory
- Why are all three different? (Guess)
- Which pair is the one we need if we want to get a model for the wildtype sequence?

# Understanding the alignment

- Perfect match: underlined by '\*'
- Somewhat matches: still written on top of each other
- Mismatch: '-'

Alignment algorithms consist of two parts

- Search method (dynamic programming matrix)
- Scoring scheme: match or mismatch

Task:

- What would you conclude from the alignment? (discussion)

# The AlignMe scoring scheme

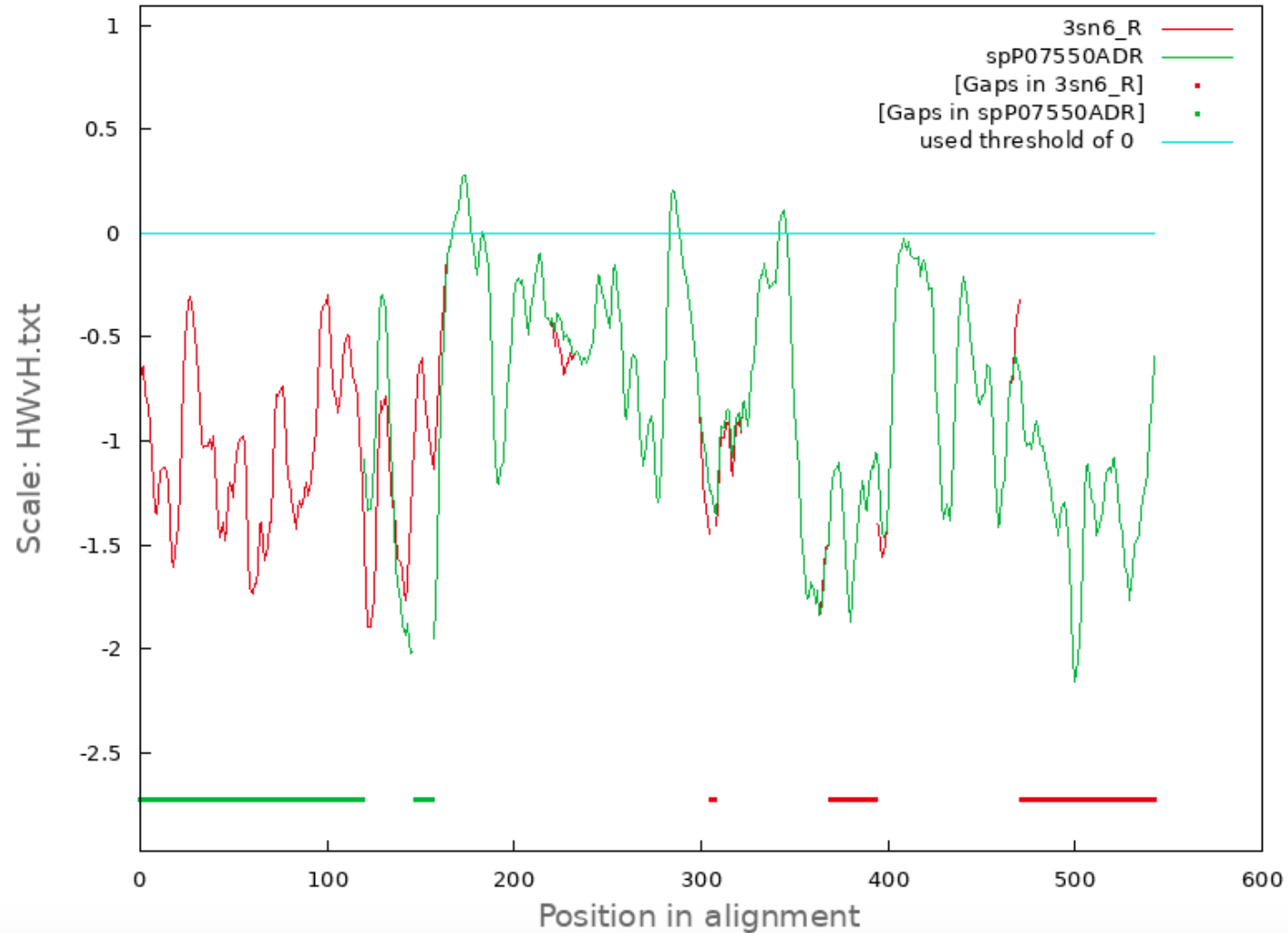
- Mismatch: gap ('-') penalty
  - Opening a gap versus extending it
  - Gap in a terminus versus a gap somewhere within the protein
  - Gaps in helices versus gaps in loops
- Similarity
  - Classical: evolutionary exchange probability (BLOSUM, PAM), TM specific
  - Position specific scoring matrices
  - AlignMe uses additionally biochemical properties and structural features
    - Hydrophobicity
    - Secondary structure
    - Transmembrane regions
- Anchors
  - Fix pairs of amino acids (e.g. AA15 from seq1 will be aligned to AA7 from seq2)



# BLOSUM Substitution matrix (evolution)

<b>Ala</b>	4																				
<b>Arg</b>	-1	5																			
<b>Asn</b>	-2	0	6																		
<b>Asp</b>	-2	-2	1	6																	
<b>Cys</b>	0	-3	-3	-3	9																
<b>Gln</b>	-1	1	0	0	-3	5															
<b>Glu</b>	-1	0	0	2	-4	2	5														
<b>Gly</b>	0	-2	0	-1	-3	-2	-2	6													
<b>His</b>	-2	0	1	-1	-3	0	0	-2	8												
<b>Ile</b>	-1	-3	-3	-3	-1	-3	-3	-4	-3	4											
<b>Leu</b>	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4										
<b>Lys</b>	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5									
<b>Met</b>	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5								
<b>Phe</b>	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6							
<b>Pro</b>	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7						
<b>Ser</b>	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4					
<b>Thr</b>	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5				
<b>Trp</b>	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11			
<b>Tyr</b>	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7		
<b>Val</b>	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	
	<b>Ala</b>	<b>Arg</b>	<b>Asn</b>	<b>Asp</b>	<b>Cys</b>	<b>Gln</b>	<b>Glu</b>	<b>Gly</b>	<b>His</b>	<b>Ile</b>	<b>Leu</b>	<b>Lys</b>	<b>Met</b>	<b>Phe</b>	<b>Pro</b>	<b>Ser</b>	<b>Thr</b>	<b>Trp</b>	<b>Tyr</b>	<b>Val</b>	

# Example of biochemical similarity



# A more challenging example

- Download sequence of human cannabinoid receptor 1
- Align it with the  $\beta 2$  wildtype sequence using AlignMe's fast mode
- Repeat with PS mode
  
- Discuss the alignment in comparison the previous examples
- Discuss the differences between Fast and PS mode