Bioinformatica A (MOL075) 2013-2014 Naam

De hoeveelheid witte ruimte die onder elke vraag is vrijgelaten geeft aan hoeveel ruimte ik denk dat je maximaal nodig hebt voor de beantwoording. Houd de antwoorden svp kort. Denk steeds eerst rustig na en schrijf de antwoorden netjes en goed geformuleerd op. Denk vooral ook aan de 10 geboden voor de bioinformatica.

Je mag antwoorden in het Nederlands, Engels, of Duits. Duitse studenten mogen, als ze een woord niet kennen in het Nederlands rustig een Duits woord gebruiken in het midden van een Nederlandse zin. Schrijf voor de zekerheid je naam rechtsboven op elk blaadje dat je inlevert.

Vul eerst het aminozuur formulier in (en houd het bij de hand bij het beantwoorden van de overige vragen!). Het ingevulde formulier moet je inleveren want het telt als 1 vraag.

1) Vul het aminozuur formulier in.

2) Voorspel de structuur van VTVTGPVTVT en leg uit hoe je tot het antwoord bent gekomen.

Teken de hairpin en je ziet dat fobe en fiele residuen tegenover elkaar staan, dat kan/mag niet. Dus dit ding bestaat niet, of blijft loop.

3) Align CWEALALLAELALAAMKGSTPNGS met CWEALALLLEALMRGTTPNGG

??hhhhhhhhhhhhhhhh-------

CWEALALLLEALMR---GTTPNGG

??hhhhhhhhhhhh----------

CW natuurlijk op CW. Voorspel/align twee helices. Gap aan einde van helix.

4) Waarin schuilt de grote kracht van de SwissProt database? En waarin de grote zwakte?

+ Well annotated en curated, vooral belangrijk zijn de vele hyperlinks

- Arbeidsintensief, en dus nooit up-to-date

5) Wat is het verschil tussen een SNP en een mutatie die vermeld staat in OMIM?

SNP is Single Nucleotide Polymorphism (komt in > 1% bevolking voor anders heet het rare mutation; officieel) en kan dus niks doen, of onschuldige dingen zoals haarkleur ‘regelen’. OMIM (Online Mendelian Inheritance in Man) is database met ziekteverwekkende mutaties.

6) Verzin de meest alanine-rijke sequentie die voldoet aan het Prosite patroon:

[LI]-{L}-[W]-x(1,2)-[LI]

L A W AA I

7) Toen ik vorig jaar mijn peptide X tegen SwissProt runde met BLAST had mijn beste hit een E-score van 7.4\*10-13. Vorige week leverde diezelfde search een E-score van 7.5\*10-13. Hoe kan dat nou? (Hier zat typefout in deze vraag bij iedereen goed gerekend…)

Na een jaar zitten er meer sequenties in de database en is de kans dat iets bij toeval gevonden wordt dus groter, en de significantie van iets wat je vindt dus kleiner.

8) Ik heb een kort humaan peptide gesequenced: STGSPDALMQPMIKAGTNNS.

Voorspel de secundaire structuur van dit peptide. Misschien heb je er iets aan dat het overeenkomstige peptide uit de Andes Emu als sequentie heeft: GASGPAEIIRALLQQGSSPA

STGSPDALMQPMIKAGTNNS Lijkt helix in het midden, maar de P zit in de weg.

----hhhhhh?hhhh-----

GASGPAEIIRALLQQGSSPA Deze homoloog is echt nette helix, dus de eerste dan ook.

----hhhhhhhhhhh-----

9) a) Waarom komen Gly, Pro, Asn, Ser, Asp, vaker dan andere residuen voor in β-turns?

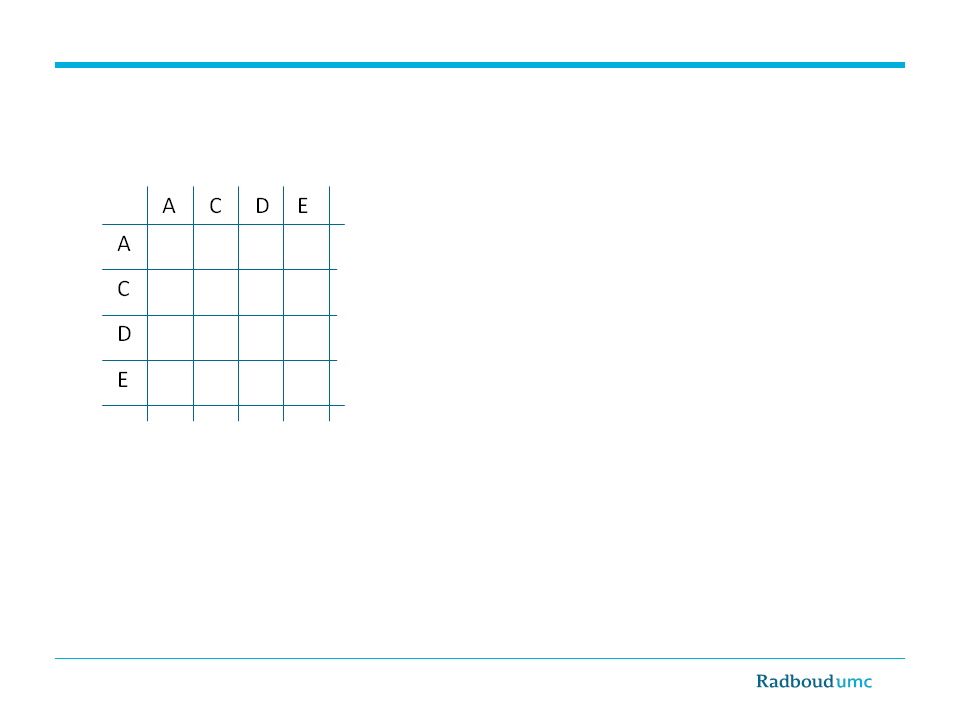
Gly flexibel. Pro rigide, maar in goede vorm voor b-turn. Andere drie compenseren energie verlies door turn-vorming met entropy van water door H-bonds met locale backbone.

b) En waarom is dit eigenlijk belangrijke informatie voor een bioinformaticus die voornamelijk geïnteresseerd is in de overdracht van informatie tussen databases en zijn/haar eiwit van interesse?

Dit help bij de correctie van alignments.

10) a) Wat is een Dayhoff matrix?

Om scores te geven aan gealignde paren van aa.

b) Waarvoor wordt een Dayhoff matrix zoal gebruikt?

In BLAST en in alignment programmas (Clustal bijv).

c) Rechts zie je de linker-boven hoek van een Dayhof matrix.

Vul svp de 16 missende getallen in.

Op diagonaal bijv 2 6 2 3

Dan A-C en D-E bijvoorbeeld 1 en A of C tegen D of E -1. Meer detaillering

Zou nog kunnen (A is iets dichter bij E dan bij D, bijv), maar hoeft niet.

11) Waarom is het capping effect van negatieve residuen aan de N-terminale kant van de helix sterker dan het effect van positieve residuen aan de C-terminale kant?

Residuen steken uit helix weg in N-terminale richting. Interactie met N-term lading komt dus natuurlijker dan die met C-term lading. (Niemand had bedacht dat negatieve residuen kleiner zijn dan positieve residuen, en er dus minder entropy van zijketen verloren gaat aan N-term (+) mkant.

12) Als ik aan het oppervlak van een eiwit een leucine muteer naar een aspartaat, wat verwacht je dan in termen van eiwit stabiliteit? Verklaar je antwoord.

In ongevouwen en gevouwen vorm bevinden beide zich volledig in water. Er verandert dus niets extra bij eiwit vouwing in een van beide gevallen. Het maakt dus niks uit.

13) GPNANGPALLEILSLIAEAAQALAGGNQDDEA Kan op precies 1 plek gefosforileerd worden door het kinase kin3a. GGLEAAKLASSAASAAELLAGDNKKKW kan door ditzelfde kinase kin3a ook op precies 1 plek gefosforileerd worden. Omcirkel de twee residuen die gefosforileerd kunnen worden, en verklaar je antwoord. (Denk hierbij goed aan de tien geboden…).

GPNANGPALLEIL**S**LIAEAAQALAGGNQDDEA

------hhhhhhhhhhhhhhhhhh--------

GGLEAAKLA**S**SAASAAELLAGDNKKKW

--hhhhhhhhhhhhhhhhhh-------

De tien geboden zeggen: 1) voorspel twee secundaire structure. 2) align de sequenties (ahv structuur). 3) Denk na over de vraag. Op S na komen alle residu typen meer dan 1 keer voor, dus je wordt al een beetje naar S toegeduwd als je vergeten bent dat fosforileren op S, T, of Y gebeurd (maar vnl op S!).

14) Er zitten geen vrije cysteines in: SSNPDLCLQACAKGGSGALEACIALCARLGSSN. Welke cysteines vormen bruggen met elkaar? Verklaar je antwoord.

SSNPDLCLQACAKGGSGALEACIALCARLGSSN

---hhhhhhhhhh----hhhhhhhhhhhh----

De Cyssen zitten op i en i+3 in beide helices, en dus boven elkaar. Er zijn nu twee mogelijkheden 1-2 en 3-4 of 1-4 en 2-3 (cross-helix, dus). ¾ van de punten voor 1 van de twee EN voorspelde structuur, de volle mep als beide alternatieven genoemd zijn.

15) Wat is een accession code en wat hebben we er eigenlijk aan?

Wordt bij depositie van data in database aan die data gekoppeld en verandert nooit meer. Met accession code kun je data altijd weer terugvinden.

16) We praten in bioinformatica vaak over overdracht van informatie tussen homologe sequenties. Wat betekent het eigenlijk als twee sequenties homoloog zijn?

Gemeenschappelijke voorouder 9in evolutie).

17) Align CWTVTVTGPTVTICW met CWTVTVGPVTICW

CWTVTVTGPTVTICW Teken twee hairpins en verwijder het uiteinde in

CWTVTVG--PVTICW de onderste.

18) De volgende woorden horen steeds paarsgewijs bij elkaar. Maak alle paartjes. Pas op, sommige woorden kunnen met meerdere woorden paren, maar je mag/moet alle woorden precies 1 keer gebruiken. alignment alpha beta bond capping carbon chou database entropy fasman gap helix hydrogen information pdb penalty proteinstructure sequence strand swissprot transfer water

gap penalty entropy water   
chou fasman sequence alignment   
swissprot database pdb proteinstructure    
hydrogen bond alpha carbon   
beta strand helix capping   
transfer information

Nogal wat mensen begonnen met alpha-helix en hadden dan bij carbon-capping niet de moed om opnieuw te beginnen.

19) Welke cysteine wordt makkelijk geoxideerd in NHSADACALMLALLSAACALMIFALLLGTSKKT (en verklaar je antwoord)?

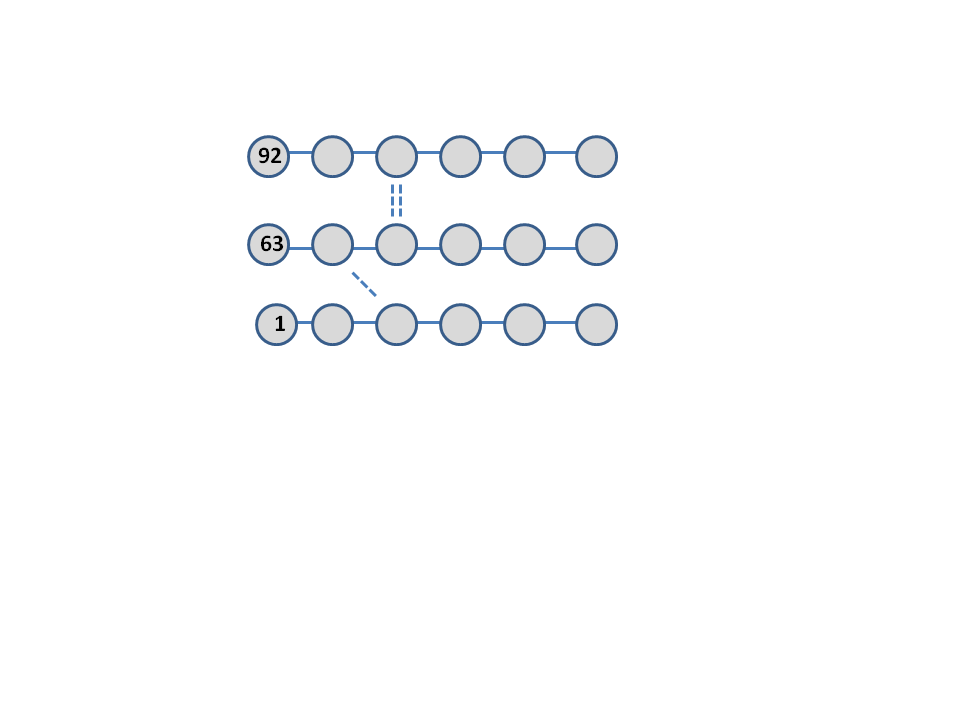
Het rode stuk is een TM helix. De eerste C zit dus nog een beetje aan de buitenkant.

20) Aan welke cysteine kan een spin-label gebonden worden in: SNGPAELCLKACLCALQALGGKS (ps, een spinlabel is een water oplosbaar molecuul dat meerdere ring systemen bevat en waar met microgolven een aangeslagen toestand in te genereren valt).

SNGPA**E**LCL**K**ACLCAL**Q**ALGGKS

---hhhhhhhhhhhhhhhh----

Helix heeft binnen en buiten kant. Onderstreepte C zit aan buiten kant



21) In het plaatje rechts staat iedere cirkel voor een ander residu in het eiwit proteohypotheticus.

Dichte lijntjes geven aan welke residuen aan elkaar zitten.

Gebroken lijntjes geven waterstof bruggen weer.

a) Vul alle residu nummers in (doe dit svp eerst op een kladblaadje want corrigeren gaat natuurlijk erg lastig).

b) Voeg alle missende waterstof bruggen toe.

c) In dit kleine deel van het eiwit zitten maar liefst 5 (vijf) prolines. Kun je svp de prolines omcirkelen?

Te lastig zonder pen en papier, sorry. Prolines in onderste en bovenste strands op plekken die geen Hbonds maken ivm dat pro geen H op bb-N heeft.

22) We hebben het in de cursus steeds maar gehad over overdracht van informatie. Kun je vijf radicaal verschillende vormen van informatie geven die van SwissProt naar jouw eiwit van interesse getransfereerd zouden kunnen worden? Bijvoorbeeld:

Domeinen

Post-translationele modificaties

Secundaire structuur

Metal binding

Active site

TM helices

23) In een nieuw gesequenced eiwit vinden we met bioinformatica tools eigenlijk niks bijzonders, behalve dat het motief N-{P}-[ST]-{G}-C-x(2)-C maar liefst vier keer voorkomt. Wat moeten we nog wel even doen (en hoe) voordat we dit gaan publiceren?

Eerst BLAST om homologen te vinden. Kijken of het geconserveerd is, indien ja, dan vervolg experimenten of misschien zelfs wel meteen publiceren.