Name (please print)

Protein Bioinformatics PH260.655

Final Exam

=> take-home questions
=> open-note
=> please use either
* the Word-file to type your answers
or
* print out the PDF and hand-write your answers

=> due: May 20th, 2010, 3:30 pm

GOOD LUCK!

In HW#5, you learned about the importance of the Pentose-Phosphate Pathway and one of its key enzymes for *Agrobacterium tumefaciens*. The other key enzyme of this pathway which was detected highly abundant in the pH 7 mass spectrometry data set is **Transketolase**.

You become curious about the nature of this protein and start to explore it a bit more, e.g. in order to have enough structure-function related information for subsequent cloning / mutagenesis / purification experiments:

Please retrieve the protein sequence of Transketolase (TktA) from *A. tumefaciens C58* in fasta format from BioCyc.

1) While looking the protein sequence up in BioCyc, you become aware of the gene reaction schematic, and you see that *Agrobacterium* actually harbors three genes associated with a transketolase function: Tkt, tktB and tktA.

They are thought to have originated from gene duplication, but while Tkt and TktB are located on the main, circular chromosome, TktA is located on a second, linear chromosome.

Please circle the correct expression below:

1a) tkt and tktB are orthologs paralogs ?

1b) tktB and tktA are orthologs paralogs ?

2) Now find out about some of TktA's physical properties, choosing an appropriate proteomic tools program, e.g. from the ExPASy server.

2a) Molecular weight:

2b) pI:

2c) Average hydropathy:

2d) Please briefly rank the value you found in 2c by marking one of the words below:

hydrophobic amphiphilic hydrophilic

3) Of course, you need to know whether there is a crystal structure known for your protein or at least for a homolog:
Search for structures using the PDB‘s advanced search option for FASTA / BLAST sequences.
3a) Please briefly explain the significance of e-values.

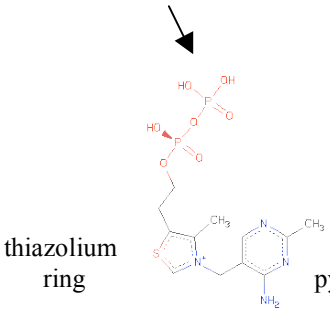
3b) Which e-value cut-off do you have to choose in order to obtain less than 30 results?

4) Alas, there is no Transketolase structure from *Agrobacterium* in the database. Instead, you decide to have a closer look at one of the *E.coli* structures with a good e-value, 2R5N, because it contains several ligands.
=> Please download the fasta sequence associated with 2R5N and the PDB file of 2R5N.

4a) There is 1 polymers and 2 chains reported in the 2R5N structure. What does that mean for TktA's oligomerization state and the number of assigned domains (see ”Derived Data” section)?

4b) Look at the ”Derived Data” section and summarize in general terms or with the help of a schematic what SCOP, CATH and Pfam classifications tell you about the enzyme‘s overall folding architecture and functional domains.

=> Please state in simple terms which domains are predicted to be involved in binding the cofactor TPP (= Thiamine pyrophosphate = Thiamine diphosphate = ThDP) ?



4c) Derive the annotated domain boundaries from the "Sequence" section (use the SCOP annotation) and list the respective residue numbers and names below:

5) You find one page of an old publication about Transketolase function, showing the multiple sequence alignment below. You vaguely recall that several entirely conserved Histidine residues are important for catalysis, but you cannot look the paper up because the library server is on maintenance till the end of the week.

```
CLUSTAL 2.0.12 multiple sequence alignment
2R5N_Transketolase      DAVQKAKSGHPGAPMGMADIAEVLWRDFLKHNPQNPSWADRDRFVLSNGHGSMLIYSLLH 76
Y. pestis               DAVQKAKSGHPGAPMGMADIAEVLWRDYLNHNPTNPHWADRDRFVLSNGHGSMLIYSLLH 76
V. P. aeroginosa        DAVQKANSGHPGAPMGMADIAEVLWRDYMQHNPSNPQWANRDRFVLSNGHGSMLIYSLLH 76
A. tumefaciens          DAVEKANSGHPGLPMGAADVATVLFTRYLKFDPKAPLWADRDRFVLSAGHGSMLLYSLLY 79
M. tuberculosis         DAVQKVGNGHPGTAMSLAPLAYTLFQRTMRHDPSDTHWLGRDRFVLSAGHSSLTLYIQLY 120
                        *.*:*. .**** .*. * :* .*:   :..:*  . * .******* **.*: :*  *:

2R5N_Transketolase      LTGYD-LPMEELKNFRQLHSKTPGHPEVGYTAGVETTTGPLGQGIANAVGMAIAEKTLAA 135
Y. pestis               LTGYD-LPMEELKNFRQLHSKTPGHPEYGYTAGVETTTGPLGQGIANAVGFAIAERTLGA 135
V. cholerae             LSGYE-LSIDDLKNFRQLHSKTPGHPEYGYAPGIETTTGPLGQGITNAVGMAIAEKALAA 164
P. aeroginosa           LTGYD-LGIEDLKNFRQLNSRTPGHPEYGYTAGVETTTGPLGQGIANAVGMALAEKVLAA 135
A. tumefaciens          LTGYEDMTIDEIKRFRQFGSKTAGHPEYGHATGIETTTGPLGQGIANAVGMAIAERKLEE 139
M. tuberculosis         LGGFG-LELSDIESLRTWGSKTPGHPEFRHTPGVEITTGPLGQGLASAVGMAMASRYERG 179
                        * *:  : :.::: :*   *:*.****  ::.*:* ********::.***:*:*.:

2R5N_Transketolase      QFN---RPGHDIVDHYTYAFMGDGCMMEGISHEVCSLAGTLKLGKLIAFYDDNGISIDGH 192
Y. pestis               QFN---RPGHDIVDHHTYAFMGDGCMMEGISHEVCSLAGTMKLGKLTAFYDDNGISIDGH 192
V. cholerae             QFN---KPGHDIVDHFTYVFMGDGCLMEGISHEACSLAGTLGLGKLIAFWDDNGISIDGH 221
P. aeroginosa           QFN---RDGHAVVDHYTYAFLGDGCMMEGISHEVASLAGTLRLNKLIAFYDDNGISIDGE 192
A. tumefaciens          EF------GSDLQSHFTYVLCGDGCLMEGISHEAIALAGHLKLNKLVLFWDDNNITIDGE 193
M. tuberculosis         LFDPDAEPGASPFDHYIYVIASDGDIEEGVTSEASSLAAVQQLGNLIVFYDRNQISIEDD 239
                         *      *    .*. *.: .** : **:: *. :**.   *.:*  *:* * *:*:..

2R5N_Transketolase      VEGWFTDDTAMRFEAYGWHVIRDIDGHDAASIKRAVEEARAVTDKPSLLMCKTIIGFGSP 252
Y. pestis               VEGWFTDDTAARFEAYGWHVVRGVDGHNADSIKAAIEEAHKVTDKPSLLMCKTIIGFGSP 252
V. cholerae             VEGWFSDDTPKRFEAYGWHVIPAVDGHDADAINAAIEAAKAETSRPTLICTKTIIGFGSP 281
P. aeroginosa           VHGWFTDDTPKRFEAYGWQVIRNVDGHDADEIKTAIDTARKS-DQPTLICCKTVIGFGSP 251
A. tumefaciens          VGLSDSTDQIARFQAVHWNTIR-VDGHDPDAIAAAIEAAQKS-DRPTFIACKTVIGFGAP 251
M. tuberculosis         TNIALCEDTAARYRAYGWHVQEVEGGENVVGIEEAIANAQAVTDRPSFIALRTVIGYPAP 299
                        .      *   *:.*  *:.    .*.:   *  *:  *:   .:*:::  :*:**: :*

2R5N_Transketolase      NKAGTHDSHGAPLGDAEIALTREQLGWKY-APFEIPSEIYAQWDAKEAGQAK-ESAWNEK 310
Y. pestis               NKAGTHDSHGAPLGEAEVAATREALGWKY-PAFEIPQDIYAAWDAKEAGKAK-EAAWNEK 310
V. cholerae             NKAGSHDCHGAPLGNDEIKAAREFLGWEH-APFEIPADIYAAWDAKQAGASK-EAAWNEK 339
P. aeroginosa           NKQGKEECHGAPLGADEIAATRAALGWEH-APFEIPAQIYAEWDAKETGAAQ-EAEWNKR 309
A. tumefaciens          NKQGTHKVHGNPLGAEEIAAARKSLNWEA-EAFVIPEDVLDAWRLAGLRSTKTRQDWEAR 310
M. tuberculosis         NLMDTGKAHGAALGDDEVAAVKKIVGFDPDKTFQVREDVLTHTRGLVARGKQAHERWQLE 359
                        *  .. . ** .**  *:  .:  :.:.   .* :  ::           : .  *: .

2R5N_Transketolase      FAAYAKAYPQEAAEFTRRMKGEMPSDFDAKAKEFIAKLQANPAKIASRKASQNAIEAFGP 370
Y. pestis               FAAYAKAYPELAAEFKRRVSGELPANWAVESKKFIEQLQANPANIASRKASQNALEAFGK 370
V. cholerae             FAAYAKAYPAEAAEYKRRVAGELPANWEAATSEIIANLQANPANIASRKASQNALEAFGK 399
P. aeroginosa           FAAYQAAHPELAAELLRRLKGELPADFAEKAAAYVADVANKGETIASRKASQNALNAFGP 369
A. tumefaciens          LEATETAK---KAEFKRRFAGDLPGNFDSSIDAFKKKIIENNPTVATRKASEDSLEVING 367
M. tuberculosis         FDAWARREPERKALLDRLLAQKLPDGWDADLP----HWEPGSKALATRAASGAVLSALGP 415
                        : *         *   * .  .:* .:          .       :*:* **   :..:.

2R5N_Transketolase      LLPEFLGGSADLAPSNLTLWSGSKAIN----------EDAAGNYIHYGVREFGMTAIANG 420
Y. pestis               VLPEFLGGSADLAPSNLTIWSGSKSLS----------DDLAGNYIHYGVREFGMSAIMNG 420
V. cholerae             LLPEFMGGSADLAPSNLTMWSGSKSLTA---------EDASGNYIHYGVREFGMTAIING 450
P. aeroginosa           LLPELLGGSADLAGSNLTLWKGCKGVSA---------DDAAGNYVFYGVREFGMSAIMNG 420
A. tumefaciens          ILPEMVGGSADLTPSNNTKTSQMKSITP---------TDFSGRYLHYGIREHGMAAAMNG 418
M. tuberculosis         KLPELWGGSADLAGSNNTTIKGADSFGPPSISTKEYTAHWYGRTLHFGVREHAMGAILSG 475
                         ***: ******: ** *  .  ...            .  *. :.:*:**..* *  .*

2R5N_Transketolase      ISLHGGFLPYTSTFLMFVEYARNAVRMAALMKQRQVMVYTHDSIGLGEDGPTHQPVEQVA 480
Y. pestis               IALHGGFIPYGATFLMFVEYARNAVRMAALMKIRSVFVYTHDSIGLGEDGPTHQPVEQMA 480
V. cholerae             IALHGGFVPYGATFLMFMEYARNAMRMAALMKVQNIQVYTHDSIGLGEDGPTHQPVEQIA 510
P. aeroginosa           VALHGGFIPYGATFLIFMEYARNAVRMSALMKQRVLYVFTHDSIGLGEDGPTHQPIEQLA 480
A. tumefaciens          IALHGGLIPYAGGFLIFSDYCRPSIRLAALMGIRVVHVLTHDSIGVGEDGPTHQPVEQIA 478
M. tuberculosis         IVLHGPTRAYGGTFLQFSDYMRPAVRLAALMDIDTIYVWTHDSIGLGEDGPTHQPIEHLS 535
                        : ***   .* . ** * :* * ::*::***    : * ******:*********:*:::
```

5a) Explain the symbols underneath the alignment:

•

*

:

5b) List all completely conserved Histidines. Please include their residue numbers corresponding to the 2R5N sequence.

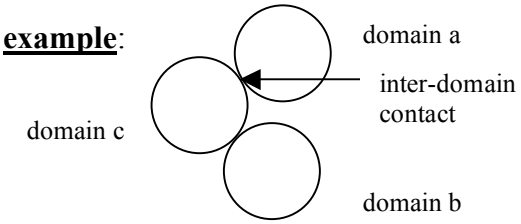
Now you would like to visualize your findings in 3D. Please open 2R5N in Jmol.

=> HINT: It might be helpful to use the "select" commands in the Jmol scrip command line, as shown in the Extra-homework set posted under May 18th.

6) From your answer in 4c, color the functional domains according to the assignment by SCOP. It is okay to color both chains.

Please draw a crude schematic of the overall protein architecture (comprising both chains) using oval- or egg-like shapes for each domain.
Please indicate inter-domain contacts.

example:



7) Examine the structure more closely and identify the different types of ligands by mousing oven them and comparing them with the Ligand Component section on the PDB structure summary page.
=> Select all ethanediol molecules and hide everything else.
(Hint: Ligands are considered "Hetero"atoms in Jmol ...)

=> How many ethanediol molecules do you see? Where could they have originated from?

You might want to change their color to "translucent" and select again all atoms.
The important ligands should now be better visible.

Using the center and zoom tools in Jmol, navigate the structure to a position from where you have a good overview of the structural environment of a sugar and a TPP ligand in one of the active sites.

8) Now, you want to distinguish between conserved Histidines in the active site and elsewhere in the structure:
=> Choose a clear representation to display all conserved His-residues found in 5b and note their residue numbers.
=> Hint: A good strategy could be again to use the command line selection tools.

8a) Which Histidines are NOT in the active site?

8b) List those Histidines which ARE in the active site and also indicate which domain they are located on.

8c) Focus on the Histidine residue that interacts closely with the phosphate moiety of the sugar and note the His-residue number.

8d) Measure the shortest distance between an atom of this His and the sugar. What is the distance?

8e) What kind of bond is formed between the sugar and this Histidine residue?

9) There is a calcium atom present in the active site.

9a) What atoms are bound to the calcium that are not part of the protein?

9b) Which molecule do these atoms belong to?

==========> PLEASE TRUN OVER <==========

10) For Extra credit:
Homology Modelling of TktA from *Agrobacerium tumefaciens* using Swiss Model.

Having dealt in so much detail with a homologous structure from *E. coli*, you would now like to get an impression of how a 3D-model of the *Agrobacterium* enzyme might look like.

http://swissmodel.expasy.org/ => Automated Mode

Submit a modelling request using your email address and *A. tumefaciens'* TktA protein sequence. You will be notified by email when the process is finished (typically takes 30-60 min.). To retrieve the result, you will need to enter a user code which will be sent to your email address.

=> What structure did Swiss model choose as a homology modeling template. Please list the PDB code.

=> Did you encounter this structure code already somewhere in the course of the exam??