Name: _____

MCB 100A/ChemC130 Midterm 1 2009

Please write your name on the first page.

1. Find the letter below that best matches the following statements. Use a letter only once. **(20 pts.)**

A. hydrogen bonds
B. most important structural restraints determined using NMR
C. wide and shallow in RNA duplex
D. flips backbone into the structure
E. reverses chain direction in 4 residues
F. stacking
G. Watson-Crick base pairs display different H-bonding pattern
H. C3' endo sugar pucker
I. measure of quality of a sequence alignment
J. genetic code
K. measure of quality of a NMR structure
L. all tRNAs
M. depends on temperature, denaturant concentration and buried surface area
N. ionic interaction
O. often recognized by hydrogen bonds from side chains in a helix
P. fold
Q. extended protein strand
R. measure of quality of a crystal structure

i) **__I___** E-value
ii) **__B__** through-space interactions
iii) **__J__** requires faithful charging of tRNAs with the correct amino acid
iv) **__M__** hydrophobic effect
v) **__N__** falls off as 1/distance
vi) **__L___** cloverleaf secondary structure
vii) **__E___** β-turn
viii) **__F___** metal-ion core
ix) **__P___** shared by homologous proteins
x) **__O___** major groove

2a. List three different advantages of a DNA genome compared to a RNA genome. **(9 pts.)**

1. **More chemically stable**
2. **Made by error correcting polymerase (replicated more accurately).**
3. **U can be recognized as DNA damage.**
4. **Major grove of DNA is accessible for sequence specific recognition.**
5. **Can be longer (because of 1 and 2).**

Partial credit:
-Describe DNA is B form but don't explain the effect of this, i.e. larger major groove, larger P to P distance, etc.
-Saying less error in DNA, but not being specific about DNA polymerase vs RNA polymerase or the ability to recognize DNA damage like U
-NO points for just saying DNA is more stable
-NO points for saying DNA is double helix because RNA can also adopt double helix structure

Name: _____

2b. The A-minor motif occurs frequently in folded RNA structures. What is the general role of an A-minor motif? **(5 pts.)**

It brings together three chains, two of which form a base pair.

Partial credit:
-Only mention 3 stands or 3 chains
-To form/stabilize tertiary structure
-Stabilization of RNA structure

2c. List two reasons that DNA strands could not form an A-minor motif. **(4 pts.)**

1. **No 2' OH in DNA for ribose H-bonding (in ribose zipper).**
2. **Minor groove in B DNA too narrow and deep to fit A base.**

Partial credit:
-Describe DNA is B form but don't explain consequences of this, i.e narrow minor groove

3. The sequence of part of a helix in the folded core of your favorite protein (YFP) is: RMELLKAAIEGD.

3a. Propose a conservative amino acid substitution that you would expect to only minimally disrupt the stability or structure of the protein. Draw the chemical structure of the starting amino acid and the new residue, and give two reasons why this mutation is likely to preserve the folded structure. **(6 pts.)**

Many choices. Best would be a chemically similar replacement for a surface residue:
K > R, R > K, E > Q, or N-terminal R > N (helix cap)

Best choices preserve 1) polarity/hydrophobicity/charge and 2) size/length
Credit NOT given for "similar shape," "isomers."
Only partial credit given for the following answers:
Ala → Val, since Val is larger than Ala
Ala → Gly, Though they have similar size and polarity, Gly is often a helix breaker and the question indicates that this is part of a helix
1 point deducted if structures have hydrogens missing from or added to atomic formulas, or if only sidechains are drawn (question asks for complete AA)

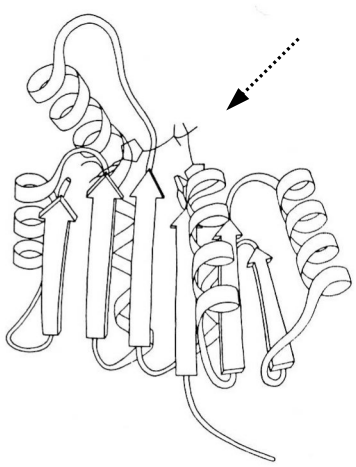
3b. The program BLAST assigns a score to define quantitatively the degree of conservation of each identity and each amino acid substitution. What is the name of the table of such scores? What do the numbers in the table represent? **(6 pts.)**

1. **Substitution matrix or BLOSUM 62**
2. **Relative frequencies of replacements sequence blocks in homologous proteins with sequence identities less than a chosen threshold.**

BLOSUM 62 calculates only the FREQUENCY of substitution of one amino acid for another based on information in a sequence database. How favorable one substitution is over another is one interpretation of this data, and is not directly what these numbers represent. Only partial credit was given if the answer focused too much on the "favorability," of substitutions.

Name: _____

4. Here is Jane Richardson's "ribbon diagram" showing the arrangement of secondary structures and loops in alcohol dehydrogenase (ADH):



4a. What is the importance of secondary structure to protein folding? **(5 pts.)**

Partners H-bond donors and acceptors in the backbone.
Partial credit also given for helping to optimize the position of hydrophobic and hydrophilic residues due to place hydrophobic sections toward the interior and hydrophilic sections toward the exterior.

4b. What type of β-sheet is found in ADH? **(3 pts.)**

parallel

4c. What is the fold class of ADH? **(3 pts.)**

α/β

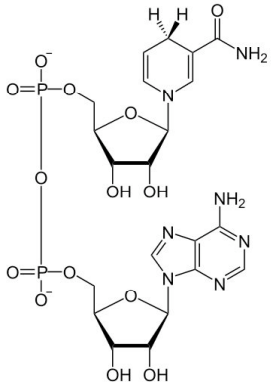
4d. Why do the helices tend to align with the strands of the sheet? **(5 pts.)**

To fill up the space between the helix and the (curved) sheet. Or
To maintain close-packing of the helix on the sheet. Or
To match the curvature of the sheet and the shape of the helix.

Partial credit given for maximizing favorable interactions, particularly hydrophobic/hydrophobic or hydrophilic/hydrophilic interactions, burial of surface area, hydrophobic effect, van der Waals interactions between sidechains.

Please note: although there may be some minimal hydrogen bonding interactions between certain sidechains, the backbone of an alpha helix DOES NOT hydrogen bond to the backbone of a beta strand. Alpha helices are defined by intramolecular hydrogen bonds between an amide nitrogen of residue n and the carbonyl group of residue n+4. Thus, the backbone in an alpha helix is already participating in hydrogen bonds, and cannot hydrogen bond to the backbone of a beta strand.

4e. The stick drawing (dashed arrow) in the ribbon diagram shows the bound cofactor, NADH (chemical structure on the right). What feature of helices, such as the one second-from-the-right in the ribbon diagram, promotes binding of NADH to the structure? **(4 pts.)**



Partial positive charge (or positive end of the helix dipole) complements the negative charge on the phosphates.

Partial credit for positive charge instead of partial positive, or mentioning backbone hydrogen bond donors/acceptors exposed on ends

4f. Do you think RNA polymerase could make NADH? Why or why not? **(4 pts.)**

Name: _____

No.
Wrong linkage between bases. Or: One base is not A, U, C or G. Or: RNA polymerase doesn't accept the nicotinamide base. Or: No template.

5. Lisa is studying a 32 kDa, 296-amino-acid terpene synthase (TPS) required for virulence of *M. tuberculosis.* The enzyme produces a novel diterpene product *in vivo*, so she knows it's active.

The best hit in a BLAST search against the sequences of proteins in the Protein Data Bank is:

```
CopII coat protein     Length=748
Expect = 0.13
 Identities = 16/42 (38%), Positives = 24/42 (57%), Gaps = 6/42 (14%)

Query  7    KEFLDLPLVSVAEIVRCRGPKVSVFPFDG---TRRWFHLECN  45
            K+ + LP+V+ + IVRCR  +  + PF      RRW   +CN
Sbjct  70   KDLVQLPVVTSSTIVRCRSCRTYINPFVSFLDQRRW---KCN  108
```

5a. What can Lisa conclude about the fold of TPS? Briefly explain your answer. **(8 pts.)**

There is no known structure that is homologous. The PDB contains the available structures, and the E-value is below the 0.001 threshold for homologous sequences.

Partial credit for folds of TPS and CopII not being similar. This is true, but the main point is that there is *no known structure* that is similar. Partial credit for E value too high without stating the threshold of 0.001.

5b. Lisa compared the TPS sequence to the database of non-redundant sequences and found the match on the next page with a "hypothetical protein" from a single-cell eukaryote called *Dictyostelium discoideum*.

She showed the results to her colleague Larry who said, "Aha, that means that the two proteins have the same fold and the same function!" Lisa says, "Not so fast. The sequence identity is pretty low." Larry said, "OK. What other computational technique could we use to test if both proteins are terpene synthases?" What would you advise Larry? What method should they use, and what result would they expect find if both proteins have the same enzymatic function? **(6 pts.)**

Use a multiple sequence alignment (MSA) of TPS enzymes to see if both sequences contain conserved functional motifs.

4 pts.:Fold recognition (PHYRE or 3D-1D profile) to reveal structural homology to known TPS enzymes.

2 pts.: Rosetta or secondary structure prediction to see if the proteins have the same fold or a known fold.

```
Expect = 2e-32
 Identities = 89/320 (27%), Positives = 158/320 (49%), Gaps = 35/320 (10%)

Query  7    KEFLDLPLVSVAEIVRCR--GPKVSVFPFDGTRRWFHLE----------------------  43
            +EF  L    +++I+  R       V+ +DGTRR + +E
Sbjct  11   QEFNKLTDNEISKIINSRLNNCNTMVYAYDGTRRSYLIENTISKLQTNGIHNNKCKFTGK  70

Query  44   ---CNPQYDDYQQAALRQSIRILKMLFEHGIETVISPIFSDDLLDRGDRYI---VQALEG  97
                  YDDY + A+ + +  L M+F+HGI+T++ P++   L DRG  Y+   ++ L G
Sbjct  71   DEKTTIDYDDYCKTAISKLLFDLVMMFKHGIKTIVYPMWFCTLEDRGPEYLPKFIKYLSG  130
```

Name: _____

```
Query  98   MALLANDEEILSFYKEHEVHVLFYGDYKKRLPSTAQGAAVVKSFDDLTISTSSNTEHRLC  157
            +  L  +E ++  YKE  + V+FYG+Y K L        ++++F+ +   T  N  H +
Sbjct  131  LKALLENETLVKLYKECGIRVIFYGEYIKLL-ERGNDPILLETFNKIMELTKDNISHTIL  189

Query  158  FGVFGNDAAESVAQFSISWNETHGKPPTRREIIEGYYGEYVDKADMFIGFGRFSTFDFPL  217
            FG    + ++++ + SI + E +   PT+ ++I+ YYG  VD+   ++GF RFST   P+
Sbjct  190  FGTTIQEPSQTIIENSIDFFEKYNYRPTKNQLIKKYYGVDVDQVSFYLGFDRFSTDGRPI  249

Query  218  LSS--GKTSLYFTVAPSYYMTETTLRRILYDHIYLRHFRPKPDYSAMSADQLNVLRNRYR  275
              S  G   LY+T++P  Y ++   R++L+D +Y R      +Y     D + +++  Y
Sbjct  250  YISDKGNEDLYYTISPHSYFSKINFRKVLFDKLYCRSNTNAKEYELKLTD-IEMMKEFYE  308

Query  276  AQPDRVFGVGCV--HDGIWF  293
                 V G+G V  H   W+
Sbjct  309  NNSTNVMGLGNVNPHGNYWY  328
```

5c. Lisa decides to determine the structure of the *M. tuberculosis* TPS. What experimental technique should she use and what is the most important reason she should use that technique? **(6 pts.)**

X-ray crystallography.
The protein is too big to easily determine the structure using NMR.

Partial credit:
-Gave a valid reason other than size (if you said x-ray)
-If you said NMR, the maximum points you could get was 4 (3 pts for NMR and 1 pts for a valid reason). However, if you said NMR for an invalid reason, you would not receive any points.

5d. Once she generates a finished set of atomic coordinates using her method of choice, what are two criteria she could use to make sure the structural coordinates contain no serious errors? **(6 pts.)**

1. Rfree < 0.3 and within 0.1 of R. or NMR models are similar to each other.
2. Good covalent geometry.
3. No interpretable difference electron density. or Few or no violations of distance restraints
4. No disallowed main-chain dihedral angles.
5. Most residues in rotameric conformations.
6. Structure is consistent with biochemical data.

Partial credit:
If part C answer was X-ray
-If you say to use low R values but don't list cut-offs
-If you say to use the different R values, but forget to mention any one of them, i.e. if you said R < 0.2 and Rf < 0.3 but don't say Rf within 0.1 of R
-a lot of people said R value for one reason and Rf value for their other reason, but these criteria fall into the same group, so if you gave these reasons you would not get all the points

If part C answer was NMR
-saying rmsd but not giving a cut off value or giving a cutoff value that wasn't in the range of 0.8-1.5.
-(NO points) if you said R values, because this is for X-ray technique
-saying that you want similar structures and giving rmsd cut-off falls into the same category

If you gave an incorrect answer for part c, i.e a non-experimental technique, but still gave valid criteria to evaluate that method, points were awarded. Also, the question asked for criteria to evaluate if the structure contained any errors, not if you had the best structure, i.e. high resolution is not a correct answer for criteria