Explanation of Analytical Cune Grading

This was an announced topic on detection and the paper was clearly about sensing. Many people simply described in words the data on the figures (that received half the points). Very few students talked about sensitivity and selectivity based on the data (remaining points).
CRIB for the biochemistry cume held on October 23, 2010

An important area in biochemistry is to be able to obtain information from biological databases. NCBI (The National Center for Biotechnology Information) at NIH provides tools for accessing and retrieving data from protein-sequence and nucleotide-sequence databases (http://www.ncbi.nlm.nih.gov/).

Because of data produced by large-scale DNA-sequencing projects, it is no longer necessary to determine the amino acid sequence of an entire polypeptide chain in a protein. Instead, researchers rely on "conceptual translation" of known mRNA sequences.

Technically, it is difficult to determine the nucleotide sequence of RNA. The current strategy is to obtain a mixture of processed mRNAs from various organisms and various cell-types. Large-scale studies use reverse-transcriptase to make a complementary DNA (cDNA) copy of each mRNA. By following standard procedures, researchers prepare cloned double-stranded cDNA. They employ automated systems for sequencing thousands and thousands purified-cloned cDNAs. The sequences are routinely deposited in GenBank.

Conceptual translation of cDNAs relies on the known dictionary of the genetic code, to deduce the amino acid sequence of a polypeptide chain from its corresponding cDNA. To facilitate data retrieval and interpretation, GenBank offers a non-redundant listing of cDNA sequences. The non-redundant listing is known as RefSeq.

NCBI provides an engine (Known as BLAST) for identifying and retrieving sequences of interest (http://blast.ncbi.nlm.nih.gov/). BLAST offers several programs for searching the sequence databases. For example: protein blast (pblast ) searches the protein database using as query either the accession number or the amino acid sequence of a protein; nucleotide blast searches a nucleotide database using as query either an accession number or a nucleotide sequence; tblastn searches translated nucleotide database using a protein sequence as query.

1 A. (20 points)

One can use the short-amino-acid-sequences obtained for catalytase, by mass spectrometry, to search the protein database for sequences that match the short-sequences. To do so, one can perform pblast. In pblast, the sequence of each fragment should be used separately. If one hits the same protein for each fragment, then one can be confident that the right protein has been identified.

The file reporting the deduced sequence of a polypeptide chain is information-rich. It includes the accession number for the file, the length of the polypeptide chain, as well its predicted amino acid sequence deduced from a corresponding cDNA. In this file, one can also obtain link to articles that have studied the protein. One might find that the function of the protein is unknown.

1 B. (15 points)

As described above, cDNA is the complementary DNA sequence of an mRNA. To identify the cDNA for catalytase, one can perform tblastn to search translated-nucleotide-database using the amino acid sequence of catalytase as query. Analysis of results of tblastn could identify cDNAs corresponding to amino-acid sequences obtained from mass spectrometry.
1 C. (15 points)

As the protein-file, the cDNA sequence file is information-rich. It includes the accession number for the file, the length of the cDNA, and the coding-sequence (start and end of translation, as well as the conceptual translation-product). In this file, one can also obtain link to articles reporting the isolation of cDNA and articles describing the known functions of the corresponding protein.

D. (20 points)

Genes in higher eukaryotes are not contiguous. The coding sequences are interrupted by non-coding segments known as introns. Protein-coding genes are initially transcribed by RNA polymerase II producing precursor mRNAs. The RNA splicing apparatus (spliceosome) processes the precursor mRNAs to produce the mRNA molecules used by ribosomes for protein synthesis. Alternative splicing (or differential splicing) is a process by which the exons in pre-mRNA are connected in multiple ways. The resulting different mRNAs are referred to as mRNA variants. The corresponding cDNAs are referred to as variants.

2A. (10 points)

Because human beings are so complex, it would be reasonable to expect than the human genome may contain greater than 100,000 protein-coding genes.

2B. (20 points)

Differential splicing of coding exons in mRNA precursors would produce mRNAs that encode distinct but related proteins (for details see Vogt and Vogt). As a consequence of differential splicing, from transcripts of 22,000 genes, the cell can produce a variety of proteins (much much greater than 100,000)!
Phosphorous reagents are required in many of synthetic reactions. This examination will cover one of the most prevalent phosphorous reagents, PPh₃, and its uses recent articles in J. Am. Chem. Soc.

1.) (7 points) The $^{13}$C NMR of PPh₃ in CDCl₃ is provided below. Label all of the peaks in the $^{13}$C NMR spectrum using the letters A-D for carbons of the structure.

![Diagram of PPh₃ structure]

<table>
<thead>
<tr>
<th>Peak (ppm)</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>137.26</td>
<td>A</td>
</tr>
<tr>
<td>137.12</td>
<td>A</td>
</tr>
<tr>
<td>133.62</td>
<td>B</td>
</tr>
<tr>
<td>133.62</td>
<td>B</td>
</tr>
<tr>
<td>128.73</td>
<td>D</td>
</tr>
<tr>
<td>128.56</td>
<td>C</td>
</tr>
<tr>
<td>128.47</td>
<td>C</td>
</tr>
</tbody>
</table>

2.) (3 points) The molecular formula for PPh₃ is C₁₈H₁₈P, but there are seven peaks in the $^{13}$C NMR spectrum. Provide a brief explanation.

The phenyl rings are equivalent

P splits C, so the 4 carbon

split to 7 peaks
3.) (4 points) In the total synthesis of (+)-fastigine, the Shair group prepared an alkyl azide and treated the synthetic intermediate with PPh₃ (J. Am. Chem. Soc. 2010, 132, 9594-9595). Draw the product and clearly depict any stereochemistry.

4.) (8 points) When an alkyl azide is treated with PPh₃ and water, an amine forms. Provide a mechanism for the transformation below.
5. (6 points) In the synthesis of (+)-lepadiformine C, the Rychnovsky group prepared a chiral diol and treated the synthetic intermediate with PPh₃ and NBS (J. Am. Chem. Soc. 2010, 132, 9591-9593). Draw the product of this reaction and clearly depict any stereochemistry.

6. (8 points) In the synthesis of phostriecin, the Boger group prepared a chiral aldehyde and treated it with PPh₃ and CBr₄ to generate a dibromoolefin (J. Am. Chem. Soc. 2010, 132, 2157-2159). Provide a mechanism for this transformation:
7.) (4 points) In the synthesis of (+)-maitotoxin, the Nicolaou group prepared a chiral aldehyde and treated it with a reagent derived from PPh₃ (J. Am. Chem. Soc. 2010, 132, 9900-9593). Draw the product and clearly depict any stereochemistry.

8.) (4 points) How can the reagent, Ph₃P=CHCO₂Me, be prepared? Show all of the reagents and reactions.

9.) (4 points) Again, in the synthesis of (+)-maitotoxin, the Nicolaou group prepared a chiral aldehyde and treated it with a reagent derived from PPh₃ (J. Am. Chem. Soc. 2010, 132, 9900-9593). Draw the product and clearly depict any stereochemistry.
9.) (4 points) In the synthesis of apiyukrodinone, the Danishefsky group prepared a mixture of epimeric alcohols and treated the synthetic intermediate with PPh₃, I₂, and imidazole (J. Am. Chem. Soc. 2010, 132, 9567-9569). Draw the product of this reaction and clearly depict any stereochemistry.

10.) (8 points) When an alcohol is treated with PPh₃, DEAD, and a carboxylic acid, an ester forms. Provide a mechanism for the transformation below and clearly depict any stereochemistry.