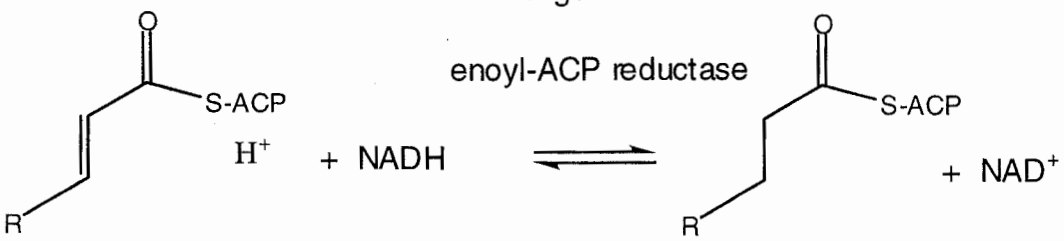
Problem Set 1, 5.08 Feb. 4, due Fri Feb 18

1. This problem has been designed to reintroduce you to protein structure, the active site of enzymes, and the importance of distances in thinking about binding and catalysis. In this age of genomics and proteomics, this problem set will also introduce you to standard methods used to find new proteins with similar functions based on detailed understanding of the structure and chemistry of a single protein. In module 2 on polyketide synthases and non-ribosomal polypeptide synthases, you will see that our understanding of the enzymes involved in fatty acid biosynthesis has played a major role in deciphering function from sequence alone. Since you have already covered fatty acid biosynthesis in 5.07 and we will revisit this pathway in lecture 3, we will use an enzyme from this pathway as an example.

   You will be looking at InhA from *Mycobacterium tuberculosis* an enoyl-ACP (acyl carrier protein) reductase that catalyzes the step in fatty acid biosynthesis shown in Figure 1. This enzyme is a member of FAS

Figure 1



(fatty acid synthase) II family that requires long fatty acyl thio ester substrates for the purpose of making mycolic acids, an essential and unique component of the cell wall of *M. tuberculosis*. Isoniazid has been for 40 years and continues to be a main line antibiotic for the treatment of *M. tuberculosis* infections. This drug is known to inhibit mycolic acid biosynthesis and one of its targets is InhA. In 1999, TB killed more adults than all other infectious diseases combined based on data from the WHO.

   The structure of InhA (1bvr.pdb) has been placed in the module 1 folder in the 5.08 locker. Pull up the structure using the methods discussed in your first recitation section. Note while the protein is a homotetramer, we have removed three of the subunits to make your analysis easier. You will be looking at a monomer.

Questions:

i. Draw a plumbing diagram indicating the secondary structure (all helices and sheets). Indicate with the residue number, where the helix or strand begins and ends.

ii. There are two substrates in the active site of this protein: NAD^+ (nicotinamide adenine dinucleotide (one of two cofactors you learned about in 5.07 that plays an essential role in oxidation/reduction reactions) and trans-2-hexadecenoyl-(N-acetylcysteamine). The coenzyme A ester (look this up in your text book) is replaced with $S(CH_2)_2NHCOCH_3$ (NAC). First think about the chemistry catalyzed by this enzyme. This is essential to thinking about the relative positions of the two substrates and whether their positions are chemically interesting in this structure. Write a mechanism for the reduction process showing the role of the $NADH/NAD^+$ cofactor. Why was NAD^+ used for the crystallization?

iii. Look at the business end of your cofactor and draw the interactions between this end and the fatty acid substrate where the chemistry occurs. Indicate all chemically interesting distances between the two substrates and the residues in the active site cavity that could be mechanistically important. Given what you see, postulate what side chains of the amino acids in the active site region might be important in the reduction of the NAC-unsaturated fatty acid thio ester.

iv. Look at the binding region for the fatty acid substrate. Describe in general the environment, that is, the predominant nature of the amino acid side chains. What is unusual about the conformation of the fatty acid substrate?

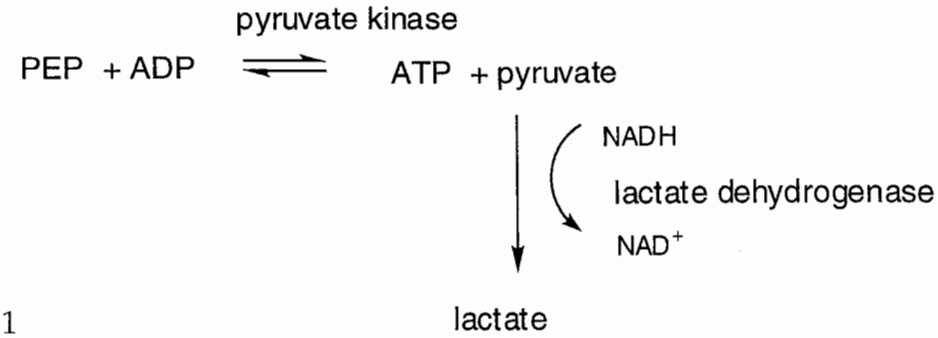
2. Now go to the http://www.ncbi.nlm.nih.gov/ site and search the data base (proteins) for enoyl-ACP reductases. Pull up the sequences for *E. coli* (FAS I, it works on shorter fatty acids); *Helicobacter pylori*; mycobacterium tuberculosis; and *Mycobacterium smegmatis*. The first two are members of FAS I enzyme family that work on fatty acid thioesters up to C16 and the second two are members of FAS II family that work on fatty acid thioesters C16-C26. Put the sequences in the appropriate format and then align these sequences using the program ClustalW (http://www.ebi.ac.uk/clustalw/) [To place your sequence in the

correct format click on help and go to your sequences. They will give you an example of the FASTA format]. In your structure analysis above (iii), you should have picked out F149, K165 and Y158 within the active site. Are these residues conserved in your sequence alignments? The substrate (fatty acid thioester) binding site in *M. tuberculosis* has been reported by the crystallographers to involve residues 196 to 219. From your analysis of the structure above, is this assignment correct? If so, then what might the ClustalW aligments you have carried out tell you about the two classes of FAS I and II proteins?
Structure based sequence alignments of a large number of reductases (> 50 are now available) will reveal very little sequence identity among all members. The conserved residues, however, are almost always informative about catalysis or regulation or even protein-protein interactions.

3. In 5.07 you studied the enzymes in the glycolysis pathway and their regulation. Given that biochemists have studied this pathway for decades, one might think that data available from *in vitro* studies could be used to model glycolysis *in vivo*. Recent studies of Teusink et al (Eur. J. Biochem 267, 5313-24 (2000)) found that despite measuring the concentrations of all of the metabolites and each enzyme in this pathway under the same growth conditions, that they had problems predicting the flux of metabolites through the pathway.

   i. The concentrations of the metabolites were measured using multiple quenching methods. For example the concentrations of G-6-P, F-6-P, F-1,6-P2, ATP, ADP, AMP, NAD etc could be measured by quenching in cold perchloric acid. The extracts were neutralized and these particular metabolites quantitated enzymatically. In general, what types of errors might be associated with these types of measurements?

   ii. The enzymatic parameters used in their calculations (kcat, kcat/Km etc) are measured in 1 mL volumes in a cuvette. For example, in the case of pyruvate kinase (eqn 1), the Vmax used in their analysis was determined to be 340 μmoles/min/mg of protein using a spectrophotometric assay in which the amount of NADH oxidized was monitored by a decrease in A340 nm ($\varepsilon 340 = 6.2 \times 10^3\ M^{-1}cm^{-1}$). The molecular weight of pyruvate kinase is 50 KDa. They observed a change in absorbance at 340 nm of 1 OD in 1 min.



Eq. 1

Previous investigators have measured the concentrations of all of the enzymes in the glycolysis pathways in many different organisms. In general the concentrations range from 0.1 to 1 mM.

Given the information in ii provide an explanation for why their approach to measured enzyme activity might be problematic. Hint, think about the detailed conditions under which they are making their measurements in vitro compared to the conditions in vivo.