MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Overview of this Module

This module has been developed for use in a variety of courses that incorporate a Malate Dehydrogenase related CURE activity: a research project to explore protein structure-function relationships. This module forms the basis of a wide variety of CURE containing courses including a full CURE where the whole course is designed around an overall research project, or a "Module" CURE where approximately 6 weeks of class work within a longer course are devoted to the "CURE". In a full CURE this module represents approximately 40% of the overall class time and is combined with several other "Modules" depending upon the goals of the research. In a modular CURE, the approaches described here may be followed by a limited series of wet lab experiments (for example, construction of the mutant, expression and purification and basic characterization of purity and specific activity), culminating in a brief presentation while in a full semester CURE would involve a wider variety of "hypothesis driven" experiments and a more substantial final presentation and discussion.

In all cases the Hypothesis Development and Proposal Module has three specific learning goals: i) Find, use and present relevant primary literature, protein sequences, structures and bioinformatics tools ii) Understand the various roles that non-covalent interactions may play in the structure and function of an enzyme, and. iii) Create/develop and present a testable and falsifiable hypothesis and appropriate experiments to interrogate the hypothesis.

Each of these learning goals is assessed using a specific rubric that can be customized to suit whatever level class is involved. The first goal, "Find, use and present relevant primary literature, protein sequences, structures and bioinformatics tools" is best assessed using a written or oral presentation (or both) of the background to the project. In the case of the second learning goal "Understand the various roles that non-covalent interactions may play in the structure and function of an enzyme" the rubric is in the form of a series of questions that can be used in a quiz in a pre-post test manner. The final goal,

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

"Create/develop and present a testable and falsifiable hypothesis and appropriate experiments to interrogate the hypothesis" again lends itself to assessment using a written or oral (or both) "proposal" In the longer versions of the CURE, this final goal can be assessed reiteratively using both a proposal and a final report format (in essence a pre-post format.) as a "midterm: and final component.

What is a good research project?

A good research project has a variety of essential elements: is original, explores an interesting research area , has a testable and falsifiable hypothesis, and has well-designed experiments that can provide believable data that are presented in a clear manner (research proposal), and culminates with a presentation of the rationale, results and conclusions, which often includes modifications of the original hypothesis and suggestions for future experiments.

A good research project:

Is related to a big picture question of interest to the broader scientific community

The Vision and Change initiative identified 5 big ideas in biology, i. Evolution, ii. Structure and Function, iii. Information flow, exchange and storage, iv. Pathways and transformations of energy and matter, and v. Systems, as well as identifying 6 attributes of a science literate student: i. Ability to apply the process of science, ii. Ability to use quantitative reasoning, iii. Ability to use modeling and simulation, iv. Ability to tap into the interdisciplinary nature of science, v. Ability to communicate and collaborate, and vi. Ability to understand the relationship between science and society.

You should be able to relate your research project to one, or more, of the five big ideas of biology. During the course of your project you will use the six attributes of a science literate student, and in particular, you may want to think about how your project may be related to the relationship between science and society – is it in any way related to medicine (drug design etc) or to other societal problems (toxic waste, food supply, biomaterials, socio-economic disparities, etc)

Addresses a foundational issue in a particular area of the relevant science

- in this case, foundational concepts of protein structure-function relationships

What is a good hypothesis?

A scientific approach often depends on a good hypothesis, and in developing ideas for a research project it is well worth considering the attributes of a good hypothesis, and thinking about what it takes to both develop and present a good hypothesis driven research project. It is also worth remembering that it is very difficult to prove a hypothesis correct: science is usually based upon eliminating reasonable alternatives- you can prove a hypothesis wrong, or you can collect evidence that is consistent with the hypothesis.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

A good hypothesis is:

Based upon prior observations

These can be your own preliminary results or they can be others work found in the literature, or often a combination of both- to develop a good hypothesis you need to find out what is already known.

Is original

If the answer to your question is known (ie is already in the scientific literature) it is not original research. You can make a hypothesis that further develops others ideas, but if the answer is known, it is not a hypothesis.

Is testable

Whatever hypothesis you make it must have predictions as to results you will get in experiments in support of the hypothesis

Is Falsifiable

The predictions you can make based upon your hypothesis must give rise to experiments where the outcome can "falsify" (disprove) your hypothesis

What is good experimental evidence?

The quality of your overall research project depends not only in the justification for the project and quality of the hypothesis but perhaps most importantly on the quality of the experimental evidence that you will collect and present.

The experimental evidence that is collected during the project is:

Well recorded

If you do not have an accurate record of your experiments, neither you nor anyone else will be able to reproduce your results, an essential part of the scientific approach is to keep accurate records of your experiments (how they are conducted) and the primary data that they yield. You should be collecting replicates of measurements- record both the primary data and the meta data in your notebook

Appropriately analyzed and presented

The majority of the data you collect will be quantitative, and described by some type of equation based upon a model. You should clearly indicate the equation you are using and your analysis of the data should give you both the parameters of the model as well as some indication as to whether you chose the correct model. You will use various

graphical and tabular presentations of your data and meta data- these should be clearly labeled and easy to read.

Reliable and repeatable, by you, or by others

Careful technique (eg pipetting) will give you the best chance of getting reliable data- replicates will have small error bars (standard error or deviation)- remember 4 and 12 are numbers, 4+/-1 and 12+/-2 are statistically different numbers at a high degree of confidence, 4+/-3 and 12+/-7 could easily be part of a single normal distribution and not significantly different. Usually error bar magnitude decreases as the number of replicates increases- more is better!

Repeatability is when you, or any other competent science student, does the experiment over, you (they) get the same results and draw the same conclusions. This depends on you carefully recording all aspects of your experiment so it can be faithfully reproduced, as well as conditions such as temperature, pH, buffer salt etc

Avoids systematic error and bias

Many experiments employ a lot of repetitive pipetting, or stock solutions that may "age" which can lead to systematic errors being introduced to your experiment- the reagent you made up at the start of the experiment may have degraded over the course of the experiment and the concentration you had planned at the start may be significantly different at the end. The enzyme stock solution you started with may have lost activity during the course of the day. These sort of unintended events can lead to systematic error being introduced into your experiment A well designed experiment will contain design elements that have the potential to reveal whether such problems are occurring.

What is a good presentation?

Clear communication is a valuable skill and during this "CURE" you will have a number of opportunities to hone your presentation skills, both written and oral. 1) Typically you will create a written introduction to your project where you will present a survey of relevant literature- this will include essential background to malate dehydrogenase that is provided in the "essential background to Malate Dehydrogenase" section of this module together with the context of your particular project obtained from your own ideas and literature searching. 2) Once you have used a variety of bioinformatics and molecular visualization tools you will have an opportunity to show the basis for your hypothesis that you will explore during the CURE and how you propose to address the various predictions that arise from your hypothesis. This will involve i) an oral presentation where you will receive feedback from peers and faculty allowing you to refine your ideas and approaches, and ii) a written proposal, which may also be anonymously reviewed by a panel of your peers, and you will receive written feedback and given a chance to revise before turning in for grading. Depending upon the overall goals of the course the written work may involve a draft, feedback and revision , and 3) at the end of the CURE you will present a summary of your goals, approaches, results and conclusions

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

which will again involve i) an oral presentation where you will receive feedback from peers and faculty allowing you to refine your ideas and conclusions, and ii) a final written report that will build off of your earlier written work..

The presentations related to the project (proposal and final report):

Whether a written proposal/report or an oral presentation with powerpoint etc, your presentation should be:

Well organized, and provide the requested information- follow the rubric if you are provided with one!

Clear and Easy to follow, and geared towards the appropriate audience

Visually attractive

Reading nothing but page after page of solid type, however well written, is not as much fun as reading text well sprinkled with white space and informative graphics- same for powerpoint slides- a simple attractive graphic with a few well chosen words or phrases will convey your information more effectively.

Use a combination of words and visual images

Appropriately referenced, and not plagarized.

Spell checked and word checked- its really hard to reed something that has correctly spelled wrong words!

Foundational Concepts of Protein Structure & Function

The concept of structure-Function relationships is one of the central themes of the Vision and Change: A Call for Action, supported by NIH, NSF and HHMI. The American Society for Biochemistry and Molecular Biology has created a Biochemistry and Molecular Biology learning framework, published by CourseSource with a major section titled “Macromolecular Structure Determine Function and Regulation” containing a series of learning goals, listed below:

What factors contribute to the size and complexity of biological macromolecules?

Macromolecules are made up of basic molecular units. They include the proteins (polymers of amino acids), nucleic acids (polymers of nucleotides), carbohydrates (polymers of sugars) and lipids (with a variety of modular constituents). The biosynthesis and degradation of biological macromolecules involves linear polymerization, breakdown steps (proteins, nucleic acids and lipids) and may also involve branching/debranching (carbohydrates). These processes may involve multi-protein complexes (e.g. ribosome, proteasome) with complex regulation.

What factors determine structure?

Covalent and non-covalent bonding govern the three dimensional structures of proteins and nucleic acids which impacts function. The amino acid sequences observed in nature are highly selected for biological function but do not necessarily adopt a unique folded structure. The structure (and hence function) of macromolecules is

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

governed by foundational principles of chemistry such as: covalent bonds and polarity, bond rotations and vibrations, non-covalent interactions, the hydrophobic effect and dynamic aspects of molecular structure. The sequence (and hence structure and function) of proteins and nucleic acids can be altered by alternative splicing, mutation or chemical modification. Sequences (and hence structure and function) of macromolecules can evolve to create altered or new biological activities.

What is the role of noncovalent intermolecular interactions?

The interactions between macromolecules and other molecules rely on the same weak, noncovalent interactions that play the major role in stabilizing the three-dimensional structures of the macromolecules themselves. The hydrophobic effect, ionic interactions and hydrogen bonding interactions are prominent. The structural organization of interacting chemical groups in a binding site or an active site lends a high degree of specificity to these interactions. The specificity and affinity of these interactions are critical to biological function.

How is macromolecular structure dynamic?

Macromolecular structure is dynamic over a wide range of time scales, and the dynamic structural changes, large and small, are often critical for biological function. Small changes can come in the form of localized molecular vibrations that can facilitate the access of small molecules to interior portions of the macromolecule. Large conformational changes can come in the form of the motions of different macromolecular domains relative to each other to facilitate catalysis or other forms of work. Proteins can contain intrinsically unstructured domains. The lack of structure in solution may facilitate a function in which interactions must occur promiscuously with several other molecules. The dynamic structure of macromolecules enables rapid changes that impact the homeostasis of biochemical and molecular biological processes

How is the biological activity of macromolecules regulated?

The biological activity of macromolecules is often regulated in one or more of a variety of hierarchical ways (e.g. inhibitors, activators, modifiers, synthesis, degradation and compartmentalization).

How is structure (and hence function) of macromolecules governed by foundational principles of chemistry and physics?

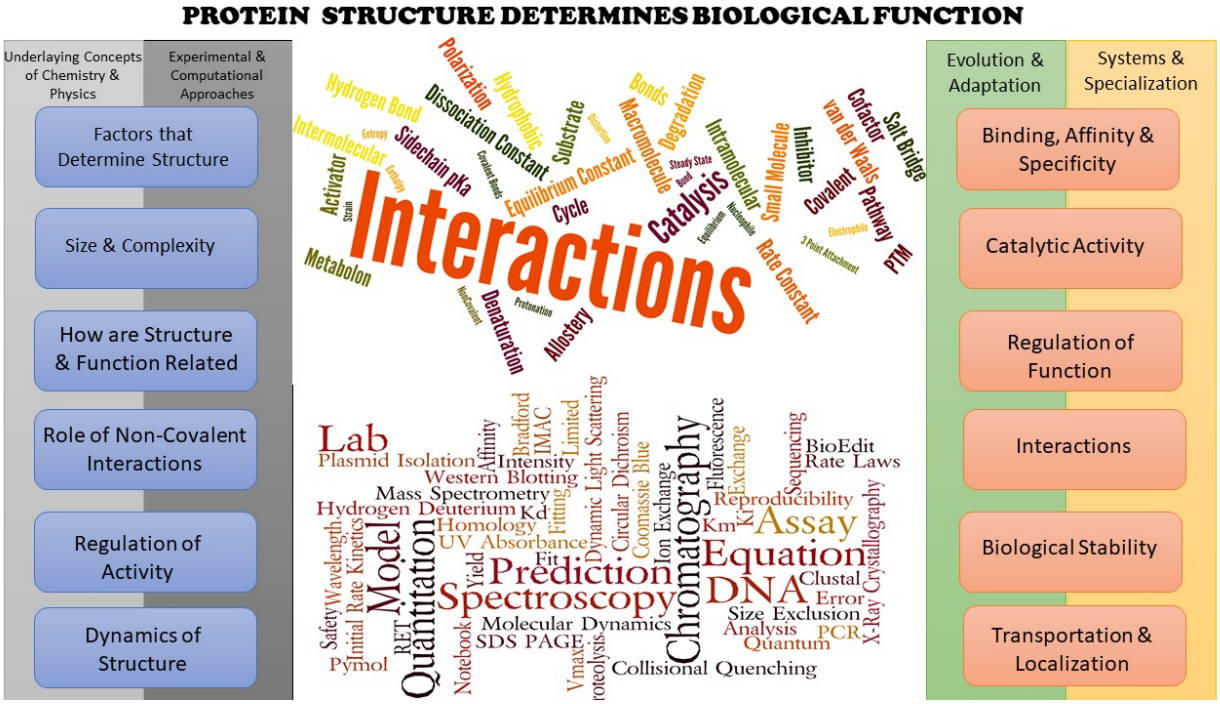
The structure (and hence function) of macromolecules is governed by the foundational principles of chemistry (including covalent bonds and polarity; bond rotations and vibrations; hydrogen bonds and non-covalent interactions; the hydrophobic effect; dynamic aspects of molecular structure; collision theory; transition state theory; rate laws and equilibria; the effects of temperature and structure and chemical reactivity) and physics (including Coulomb's Law; Newton's laws of motion; energy and stability; friction; diffusion; thermodynamics; and the concept of randomness and probability).

How are a variety of experimental and computational approaches used to observe and quantitatively measure the structure, dynamics and function of biological macromolecules?

A variety of experimental and computational approaches can be used to observe and quantitatively measure the structure, dynamics and function of biological macromolecules. Equations can be derived from models and used to predict outcomes or analyze data. Data can be analyzed statistically to assess the correctness of the model and the reliability of the data.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module



These are summarized in the above figure. Just as underlaying concepts of chemistry and physics, and the fact that a wide variety of experimental and computational approaches can be used to examine structure and function of proteins, when considered in the light of other foundational concepts of biological function, Evolution and Adaptation, and Systems and Specialization are over-arching concepts that interface with structure-function relationships.

Overview:

Developing a Hypothesis to explore structure-function relationships in Malate Dehydrogenase

Overall Rationale to a Site Directed Mutagenesis Exploration of Structure-Function Relationships

What sort of things are you changing in a Site Directed Mutagenesis Experiment?

Non-Covalent Interactions!

What Might Non-Covalent Interactions contribute to Protein Structure and Function?

While site directed mutagenesis changes the covalent structure of the protein, the impact on function is mediated by the changes in non-covalent interactions that result and a critical component of developing a testable hypothesis is understanding the various types of roles that non-covalent interactions play in enzyme structure and function. There are two conceptual types of interactions that must be considered, attractive and repulsive and these can be further categorized as coulombic and hydrophobic.

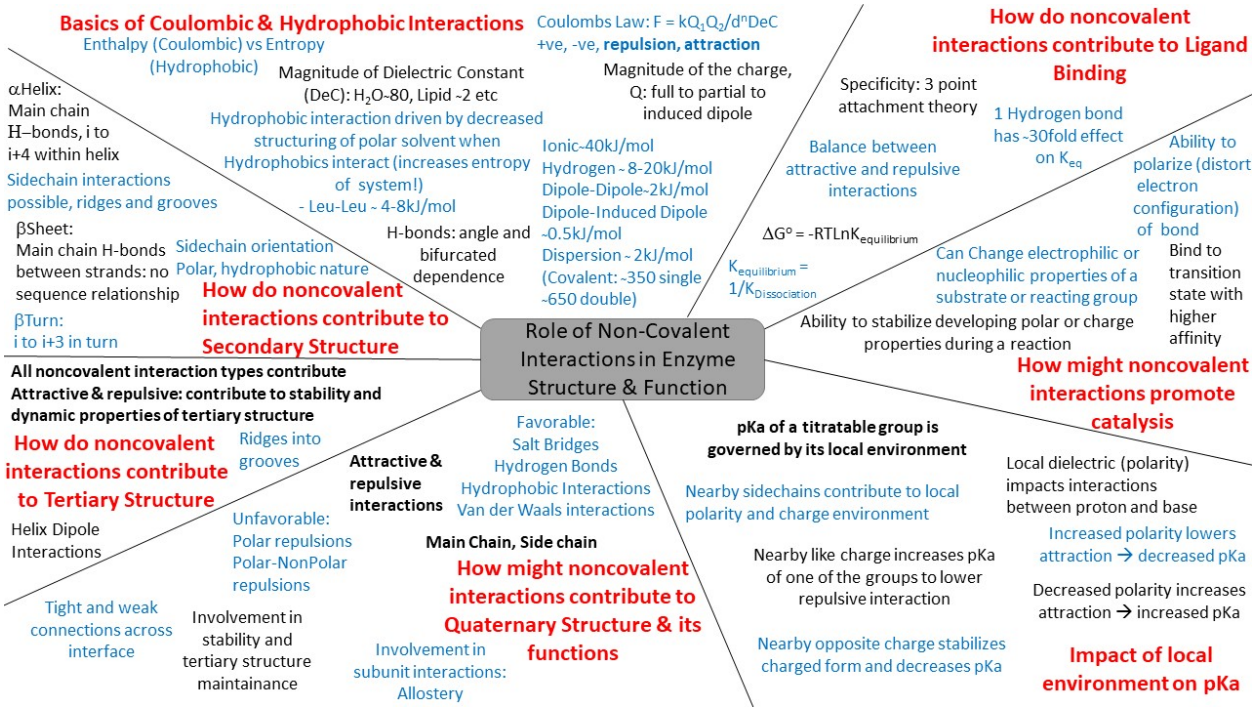
MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Charges in a molecule can be formal charges, for example at pH 7 a carboxyl group usually carries a formal negative charge, or partial charges- polar parts of a molecule usually have a δ negative, δ positive component. Coulombic interactions involving opposite charges are favorable, while those involving like charges are unfavorable. Hydrophobic interactions require a polar solvent to be present and result from "hydrophobic" moieties clustering to minimize the unfavorable change in solvent entropy that would result in the absence of such clustering. A structure that places a polar group (formal or partial charge) in a "hydrophobic" region is unfavorable because polar groups like to interact with other polar groups and are stabilized by such interactions- when they cannot in a nonpolar environment it is not favored since it is a "higher" energy situation.

Non-covalent interactions play critical roles in a variety of aspects of enzyme structure and function including the folding and overall tertiary structure, the quaternary structure as well as secondary structure, all of which involve both favorable and unfavorable components and are related to the dynamic structure of the protein. Non-covalent interactions clearly play critical roles in binding processes, both specificity and affinity, whether it be substrates, inhibitors and regulators or other proteins, in for example a metabolon or as part of a post-translational modification. When it comes to the involvement of non-covalent interactions in the mechanism of catalysis there are a number of components to consider including stabilization of intermediates/transition states to the polarization of bonds in addition to showing favorable binding to the transition state versus the substrates

The mind map below illustrates both the basis of these interactions and summarizes the roles they may play in enzyme structure and function.



MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

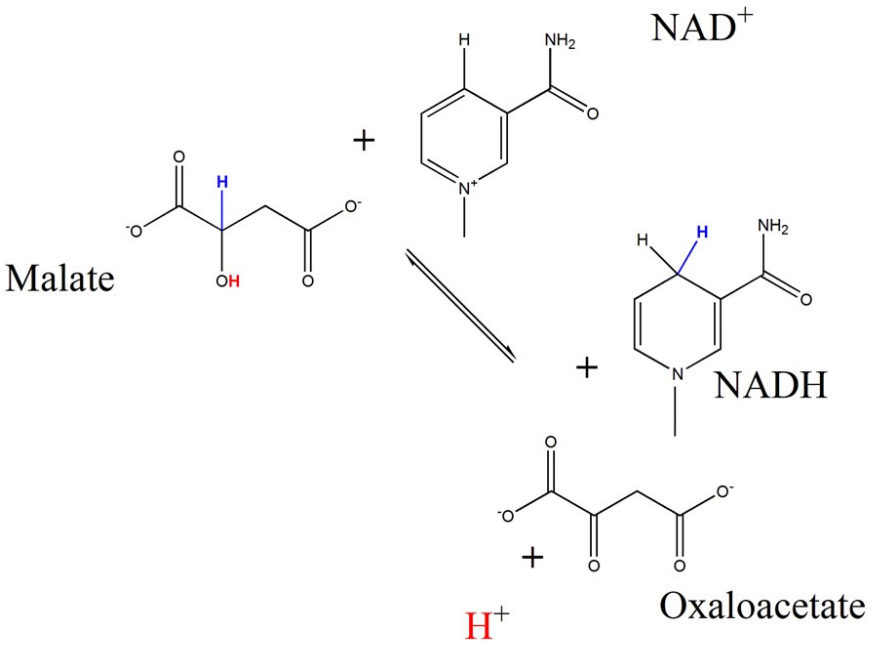
Learning Goal: Find, use and present relevant primary literature, protein sequences, structures and bioinformatics tools

Introduction to Malate Dehydrogenase

Malate Dehydrogenases catalyze the reaction:

Malate + NAD^+ ←→ Oxaloacetate + NADH

Involving a simple hydride transfer from the 2 position of Malate to NAD^+ to form the reduced cofactor NADH. The equilibrium constant for the reaction favors the NAD^+/Malate side of the reaction (1).

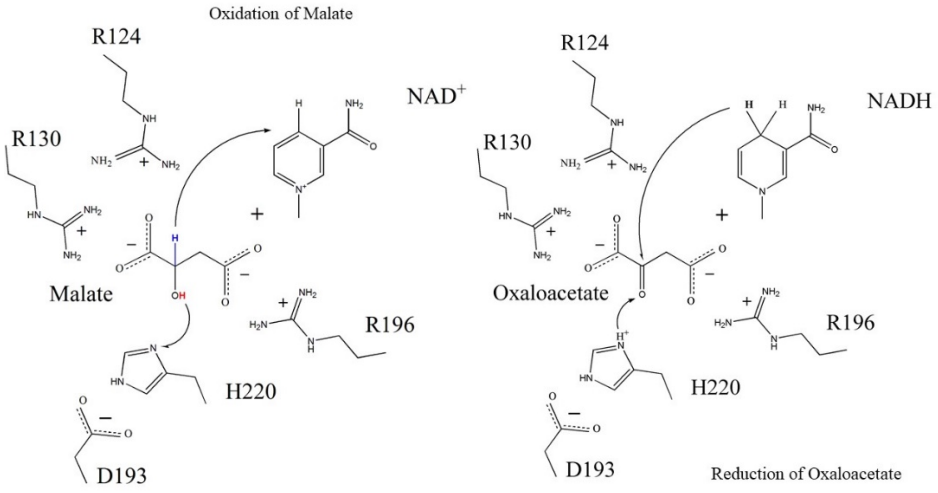


The reaction is thought to involve a base catalyzed abstraction of the proton from the Malate O-H group involving a conserved histidine in the active site of the enzyme. In all known MDHs there is a conserved Aspartate adjacent to the histidine which is thought to increase the basicity of the N: of the histidine ring, enhancing its ability to abstract the proton from the Malate O-H.(2) The malate/oxaloacetate substrate is held in place by the presence of three conserved arginine residues,

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

whose positive charges interact with the negative charges of the malate/oxaloacetate. One of them, (R153 in e Coli) has been the subject of mutagenesis and rescue using chemical biology to probe its involvement in binding and catalysis (3).

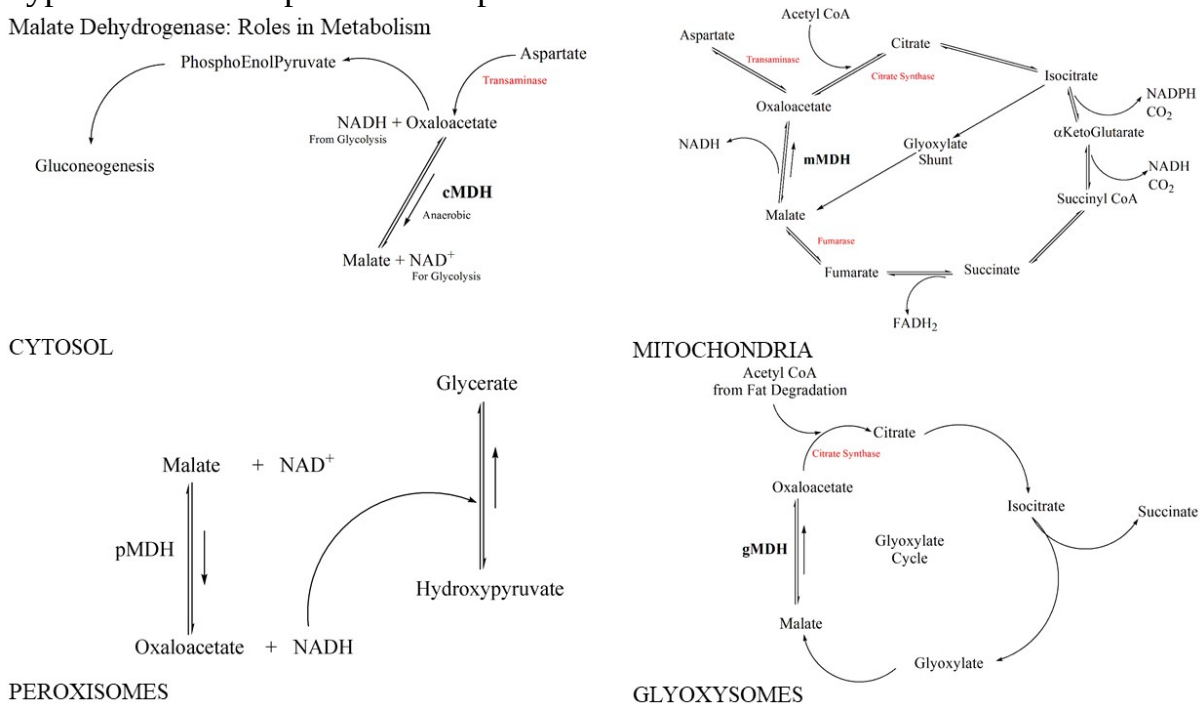


The Hydrogen on the C2 carbon of Malate is transferred as a hydride ion (H-) to the 4-Carbon of the nicotinamide ring of the cofactor NAD^+ to give NADH. During the process a proton is also released to the solvent.

This reaction plays a number of important roles in metabolism, illustrated by a reaction in the mitochondrion in the Tricarboxylic acid cycle, a reaction playing a role in the shuttling of reducing equivalents from the cytosol to the mitochondria, in peroxisomes, and in plants a reaction in the Glyoxysome:(4)

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

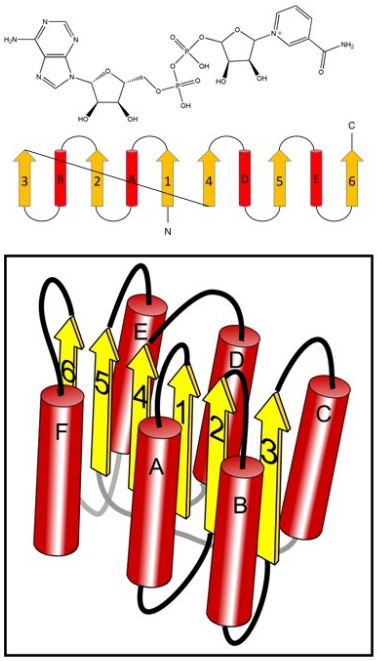


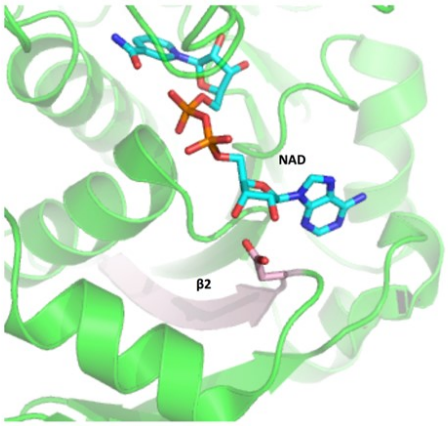
It is clear that there must exist Malate Dehydrogenase in at least two different locations within the cell and in fact there are distinct cytoplasmic MDH [cMDH](5) and a mitochondrial MDH [mMDH](6) isoenzymes in higher eukaryotes which have different amino acid sequences and slightly different three dimensional structures. Peroxisomes [pMDH](7) and in plants, glyoxysomes [gMDH] (8), also have distinct isoenzymes. The organelle forms (mitochondria, glyoxysomes) are synthesized as precursors coded for by nuclear genes, synthesized in the cytosol and transported to the appropriate organelle, guided by a "pre-sequence" of about 40 amino acids that is removed upon import to the organelle (9).

Pig cytoplasmic and mitochondrial malate dehydrogenases were among the first enzyme structures to be determined by X ray crystallography (10,11). Since then the structure of Malate Dehydrogenase from a number of sources have also been determined by X Ray crystallography. In addition to the details of the amino acids that play a role in substrate binding and catalysis, indicated in figure 2, the structure has several notable features. As with many enzymes that bind NADH there is a clear "Rossman fold" (12) associated with cofactor binding consisting of a β–α–β–α–β–β–α–β–α–β secondary structure motif that in enzymes with specificity for NAD(H) has an aspartate at the C-terminal end of the β2 strand that interacts with the Adenine ribose of the cofactor, figure 3.

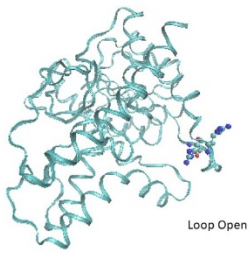
MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

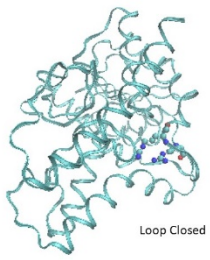
Hypothesis Development & Proposal Module





For substrate (Malate/Oxaloacetate) binding there is a “flexible loop” (13) that contains two of the three arginine residues involved in malate/oxaloacetate binding, and swings in and out of the active site to complete the active complex of the enzyme (figure 4).





The enzyme has quaternary structure and is in most instances a dimer of two identical (in terms of amino acid sequence) polypeptide chains and displays a clear interface between the two subunits. The quaternary structure undergoes a pH dependent dissociation (14), and monomeric forms have been engineered by mutation of specific amino acids at the subunit interface (15). Monomeric forms show little if any activity.

MDH is thought to form loose multienzyme complexes with several other enzymes sharing substrates, so called “metabolons” (16). In particular Aspartate AminoTransferases (which catalyze the transamination of Glutamate and Oxaloacetate to give Aspartate and 2-Oxoglutarate, a key reaction in the the aspartate-malate shuttle, the Glyoxylate Cycle and Gluconeogenesis) (17), Citrate synthase, (which catalyzes the next step in the Tricarboxylic Acid (Krebs) Cycle) (18), and Fumarase, (which catalyzes the preceding reaction in the Krebs cycle)(19) have been suggested to form metabolons with malate dehydrogenase in certain organisms. Such metabolons are thought to either increase the effectiveness of consecutive reactions by “channeling” products from one enzyme to the next one and or allow for coordinated regulation of a pathway.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Given its central role in a number of metabolic pathways, some forms of the enzyme are subject to allosteric regulation by citrate (20), by post-translational modifications such as by lysine acetylation (21) and by transcriptional control mediated through miR-743a (22).

The malate dehydrogenase structure and lactate dehydrogenase structures have many similarities and lactate dehydrogenases are thought to have evolved from malate dehydrogenases (23), a step requiring alteration of substrate specificity to bind lactate/pyruvate in place of malate/oxaloacetate. While many malate dehydrogenases are mesophilic, psychrophilic and thermophilic forms have also been characterized and are often studied to explore mechanisms of adaptation (24). As discussed below in the bioinformatics section the existence of cytosolic and organelle forms of malate dehydrogenase is thought to have arisen from a symbiotic relationship between bacteria (precursors of the organelle forms) and eukaryotes which contained a cytosolic form of malate dehydrogenase

Malate Dehydrogenase has also been used as a model system to examine a variety of aspects of protein folding including tertiary structure, quaternary structure (25,26) and the impact of a precursor sequence on structure (27).

Finally, as with many essential enzymes, malate dehydrogenase is a potential target for drug design in pathogenic organisms such as Mycobacterium tuberculosis (28) and Plasmodium falciparum (29), or tumor tissues depending upon enhanced metabolism (30, 31). Such drug design depends upon exploiting often subtle differences in structure function relationships or developing so called “allosteric drugs” that target flexible regions of the protein required for activity (32).

There remain many unanswered fundamental questions about MDH to be investigated.

Some examples are:

i) Folding, Stability & Oligomeric Structure: most MDHs are dimeric, some are monomers and others form tetramers. Folding and acquisition of oligomeric structure in α/β proteins such as MDH is also little studied. The role of protein dynamics and stability in biological activity is another aspect ripe for further investigation

ii) Substrate specificity and catalytic mechanism. (there exist MDH isoforms with LDH like activity, or NADPH (vs. NADH) affinity). Although the roles of the active site His-Asp diad, and the flexible loop containing 2 of the three active site arginines are frequently assumed, detailed information about their roles in catalysis and substrate binding is lacking. Little is known about the roles of so-called “second sphere” residues

iii) Allosteric regulation (some forms are regulated by citrate inhibition/activation and/or substrate inhibition) and pH, together with the role of Subunit Interactions and the subunit interface .

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Work on the roles of interface residues H90, E256, S266 and L269 is underway in the Bell lab and nearing publication)

iv) Metabolon formation: mammalian mitochondrial MDH (but not cytosolic MDH) can interact with other proteins including citrate synthase and fumarase, while both the cytosolic and mitochondrial forms may interact with aspartate amino transferase

and

v) Adaptation (Salinity, Temperature) and evolution (MDH Isoforms, relationship to LDH)

Literature Cited:

1. Guynn, RW., Gelberg, HJ.,Veech, RL., The Equilibrium Constants of the Malate Dehydrogenase, Citrate Synthase, Citrate Lyase and Acetyl CoA Hydrolysis reactions under physiological conditions, J. Biol. Chem 248, 6957-6965, 1973
2. Birktoft JJ, Banaszak LJ. 1983. The presence of a histidine-aspartic acid pair in the active site of 2-hydroxyacid dehydrogenases. *J Biol Chem*, 258:472-482
3. Jessica K. Bell, Hemant P. Yennawar, S. Kirk Wright, James R. Thompson, Ronald E. Viola, and Leonard J. Banaszak, Structural Analyses of a Malate Dehydrogenase with a Variable Active Site, J. Biol Chem Vol. 276, No. 33, Issue of August 17, pp. 31156–31162, 2001
4. Musrati RA, Kollárová M, Mernik N, Mikulásová D (September 1998). "Malate dehydrogenase: distribution, function and properties". *General Physiology and Biophysics*. 17 (3): 193–210
5. Ding Y, Ma QH., Characterization of a cytosolic malate dehydrogenase cDNA which encodes an isozyme toward oxaloacetate reduction in wheat., Biochimie. 2004 Aug;86(8):509-18
6. Gietl C, Lehnerer M, Olsen *0.* 1990. Mitochondrial malate dehydrogenase from watermelon sequence of complementary DNA clones and primary structure of the higher-plant precursor protein. *Plant Mol Biol,* 14:1019-1030.
7. Asaph B. Cousins, Itsara Pracharoenwattana, Wenxu Zhou, Steven M. Smith, and Murray R. Badge, Peroxisomal Malate Dehydrogenase Is Not Essential for Photorespiration in Arabidopsis But Its Absence Causes an Increase in the Stoichiometry of Photorespiratory CO2 Release, Plant Physiol. Vol. 148, 200, 786-795
8. McAlister-Henn L. 1988. Evolutionary relationships among the malate dehydrogenases. *Trends Biochem Sci* 13:178-181.
9. CHRISTINE GIETL AND BERTOLD HOCK, Organelle-Bound Malate Dehydrogenase Isoenzymes Are Synthesized as Higher Molecular Weight Precursors, Plant Physiol. (1982) 70, 483-487
10. Birktoft JJ, Rhodes G, Banaszak LJ. 1989b. Refined crystal structure of cytoplasmic malate dehydrogenase at 2.5-A resolution. Biochemistry, 28:6065-6089.
11. Birktoft JJ, Fu Z, Carnahan GE, Rhodes G, Roderick SL, Banaszak LJ. Comparison of the molecular structures of cytoplasmic and mitochondrial malate dehydrogenase. *Biochem SOC Trans* 17:301-304

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

12. Buehner M, Ford GC, Moras D, Olsen KW, Rossman MG., D-glyceraldehyde-3-phosphate dehydrogenase: three-dimensional structure and evolutionary significance., Proc Natl Acad Sci U S A. 1973 Nov;70(11):3052-4
13. CHRISTOPHER R. GOWARD AND DAVID J. NICHOLLS, Malate dehydrogenase: A model for structure, evolution, and catalysis, *Protein Science* (1994), 3:1883-1888
14. Chem J and Smith DL., Amide hydrogen exchange shows that malate dehydrogenase is a folded monomer at pH 5, *Protein Science* (2001), 10:1079–1083.
15. Breiter DR, Resnik E, Banaszak LJ., Engineering the quaternary structure of an enzyme: construction and analysis of a monomeric form of malate dehydrogenase from Escherichia coli., Protein Sci. 1994 Nov;3(11):2023-32
16. Wu F, Minteer S., Krebs cycle metabolon: structural evidence of substrate channeling revealed by cross-linking and mass spectrometry., Angew Chem Int Ed Engl. 2015 Feb 2;54(6):1851-4
17. Leonard A. Fahien$, Edward H. Kmiotek, Michael J. MacDonaldg, Barbara Fibich, and Milka Mandic, Regulation of Malate Dehydrogenase Activity by Glutamate, Citrate, &-Ketoglutarate, and Multienzyme Interaction, J. Biol. Chem Vol. 263, No. 22, Issue of August 5, pp. 10687-10697,1988
18. Igor Morgunov and Paul A. Srere, Interaction between Citrate Synthase and Malate Dehydrogenase: SUBSTRATE CHANNELING OF OXALOACETATE, J. Biol. Chem., Vol. 273, No. 45, Issue of November 6, pp. 29540–29544, 1998
19. Sonia BEECKMANS, Edilbert VAN DRIESSCHE and Louis KANAREK, The visualization by affinity electrophoresis of a specific association between the consecutive citric acid cycle enzymes fumarase and malate dehydrogenase, Eur. J. Biochem 183,449 -454 (1 989)
20. Gelpi JL, Dordal A, Montserrat J, Mazo A, Cortes A. 1992. Kinetic studies of the regulation of mitochondrial malate dehydrogenase by citrate. Biochem J 283:289-297.
21. Shimin Zhao, Wei Xu, Wenqing Jiang, Wei Yu, Yan Lin, Tengfei Zhang, Jun Yao, Li Zhou, Yaxue Zeng, Hong Li, Yixue Li, Jiong Shi, Wenlin An, Susan M. Hancock, Fuchu He, Lunxiu Qin, Jason Chin, Pengyuan Yang, Xian Chen, Qunying Lei, Yue Xiong, Kun-Liang Guan, Regulation of Cellular Metabolism by Protein Lysine Acetylation, SCIENCE , 19 FEBRUARY 2010 VOL 327 , 1000-1004
22. Qingli Shi and Gary E. Gibson, Up-regulation of the mitochondrial malate dehydrogenase by oxidative stress is mediated by miR-743a, J Neurochem. 2011 August ; 118(3): 440–448.
23. Madern D (June 2002). "Molecular evolution within the L-malate and L-lactate dehydrogenase super-family". *Journal of Molecular Evolution*. 54(6): 825–40
24. Sundaram TK, Wright IP, Wilkinson AE. 1980. Malate dehydrogenase from thermophilic and mesophilic bacteria. Molecular size, subunit structure, amino acid composition, immunochemical homology, and catalytic activity. Biochemistry 19:2017-2022.
25. Gietl, C, Seidel, C., & Svendsen, I., Plant Glyoxysomal but not mitochondrial malate dehydrogenase can fold without chaperone assistance. Biochimica et Biophysica Acta 1274 (1996) 48-58
26. Bjørk A, Dalhus B, Mantzilas D, Sirevåg R, Eijsink VG., Large improvement in the thermal stability of a tetrameric malate dehydrogenase by single point mutations at the dimer-dimer interface., J Mol Biol. 2004 Aug 27;341(5):1215-26

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

27. Cox B[1], Chit MM, Weaver T, Gietl C, Bailey J, Bell E, Banaszak L., Organelle and translocatable forms of glyoxysomal malate dehydrogenase. The effect of the N-terminal presequence. FEBS J. 2005 Feb;272(3):643-54
28. Vivek Kumar Singh and Indira Ghosh, Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in Mycobacterium tuberculosis, and its application to assessment of drug targets, *Theoretical Biology and Medical Modelling* 2006, 3:27
29. Anupam Pradhan, Abhai K. Tripathi, Prashant V. Desai , Prasenjit K. Mukherjee,Mitchell A. Avery, Larry A. Walker, Babu L. Tekwani, Structure and function of Plasmodium falciparum malate dehydrogenase: Role of critical amino acids in co-substrate binding pocket, Biochimie 91 (2009) 1509–1517
30. Ravi Naik , Hyun Seung Ban, Kyusic Jang, Inhyub Kim, Xuezhen Xu, Dipesh Harmalkar, Seong-Ah Shin, Minkyoung Kim, Bo-Kyung Kim, Jaehyung Park, Bonsu Ku, Sujin Oh,Misun Won, and Kyeong Lee, Methyl 3-(3-(4-(2,4,4-Trimethylpentan-2-yl)phenoxy)-propanamido)benzoate as a Novel and Dual Malate Dehydrogenase (MDH) 1/2 Inhibitor Targeting Cancer Metabolism J. Med. Chem. 2017, 60, 8631-8646
31. Boxi Zhang, Johan Tornmalm, Jerker Widengren, Helin Vakifahmetoglu-Norberg, and Erik Norberg, Characterization of the Role of the Malate Dehydrogenases to Lung Tumor Cell Survival, *Journal of Cancer* 2017, Vol. 8, 2088-2096
32. J. Ellis Bell and ., Jessica K. Bell., Allosteric Proteins and Drug Discovery.,2010 in Burger's Medicinal Chemistry and Drug Discovery, 1, https://doi.org/10.1002/0471266949.bmc028.pub2
33.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

PART 1A

Exploring Structure-Function Relationships in Malate Dehydrogenase I

Using the primary literature and other electronic resources

BACKGROUND LITERATURE

The brief introduction to Malate Dehydrogenase contains a number of references to the primary literature. These articles describe some previous work on Malate Dehydrogenase. You are expected to find additional literature that you will use in your "proposal" and final report.

There are several convenient ways to find appropriate literature. The easiest is to simply read one of the previous papers- it will cite relevant literature that you may find useful. In terms of general background information and references, this is often in the introduction to a published paper- for example, in class we have discussed the fact that glyoxysomal malate dehydrogenase is synthesized on the ribosome as a longer version and the first 36 amino acids are a "signal" for import into the glyoxysome and subsequently removed to give the "mature" protein- the references to this work are cited in the introduction to the paper.

Other Electronic Resources: How to Find General Information about an Enzyme or Enzyme Family

There exist a number of enzyme and protein data bases that collect information about particular enzymes and can be a good source of general information and references about enzymes.

These include:

Brenda (http://www.brenda-enzymes.org/)

Proteopedia (http://proteopedia.org/wiki/index.php/Main_Page)

Enzyme Nomenclature DataBase (http://enzyme.expasy.org/)

Protein Data Base- (https://www.rcsb.org/pdb/home/home.do)

BioCyc Metabolic Pathways etc (http://metacyc.ai.sri.com/PToolsWebsiteHowto.shtml#SearchHelp)

How to Find a Paper

For papers in the Life Sciences PubMed (http://www.ncbi.nlm.nih.gov/pubmed) is probably the most useful source of literature. (It is also great for finding gene and protein sequences and a number of other bioinformatics resources)

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

You can enter key words or phrases, or authors and search. The more specific you are in your search terms the easier it will be to find relevant papers. You can combine key words and authors etc using "and"

For example if you entered the term "ion channels" you would find hundreds of thousands of papers that have something to do with ion channels- by simply adding "structure of" to the search you decrease the returned items by an order of magnitude. If you add a species name as well you reduce even more. If you are looking for references about techniques use the technique name in the search words- for example, add "X Ray Crystallography" to ion channels etc

Once you have a set of citations you can refine by article type and availability, as well as restrict by publication dates.

Which ones do you want to look at- its often easy to tell simply from the title whether it is a relevant paper- you might have to screen 20-100 titles on the list but it doesn't take long. If you click on a title you get the abstract (if available) and link to the whole paper if you filtered using Free Full Text- again, the abstract will often tell you key bits of information about what the paper is about and whether it is relevant to what you are looking for.

Once you have found a relevant paper, finding more is easy- The paper you have found will cite many relevant references and will give you the names of authors working on similar problems.

How to Read a Paper

Once you have found a paper the following steps will help you fulfil the rubric criteria.

Big Picture Aspects

1. Read the title- what does just reading the title tell you?
2. Read the Abstract- what does the abstract tell you?

Start a Mind-Map of the paper based on your answers to these two questions

Read the Introduction

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Before reading the introduction ask yourself the following question: Why are you reading the paper

There are two possible answers:

1. To get an overview of the science- what did they find out about the problem you are interested in.

2. To get information about one or more experimental approaches.

You may of course be interested in both of these aspects

The approach you take to reading the paper depends upon which of these answers is most appropriate- of course if both answers apply you need to combine these approaches.

From the Introduction:

1 What is the context of the paper

2 What work by others is critical to the current paper: identify the names of senior authors on important references

3 Identify 3 critical background references

4 Summarize the Big Picture aspect of the work

5 What is the central hypothesis that is to be tested

From the Methods and Results Sections:

6 Identify preparative experiments

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Understand the basis of the experiments, look at quantitative aspects of the experiments, how much protein is needed etc, how do they establish purity, quantitation etc

7 What are the critical experiments that test the hypothesis

What is the basis for the experiment, what is measured, is it a direct or indirect measurement

8 Which are the most important figures or tables in the paper

What is displayed in the figure, what are the axes- how is the dependent variable measured?, what are the units etc

From the Discussion and Conclusions sections

9 What are the major conclusions reached

10 What evidence are the major conclusions based upon

11 What is the reproducibility of the experimental data and how might this affect the conclusions that will be reached for each experiment

12 What are the controls that are used

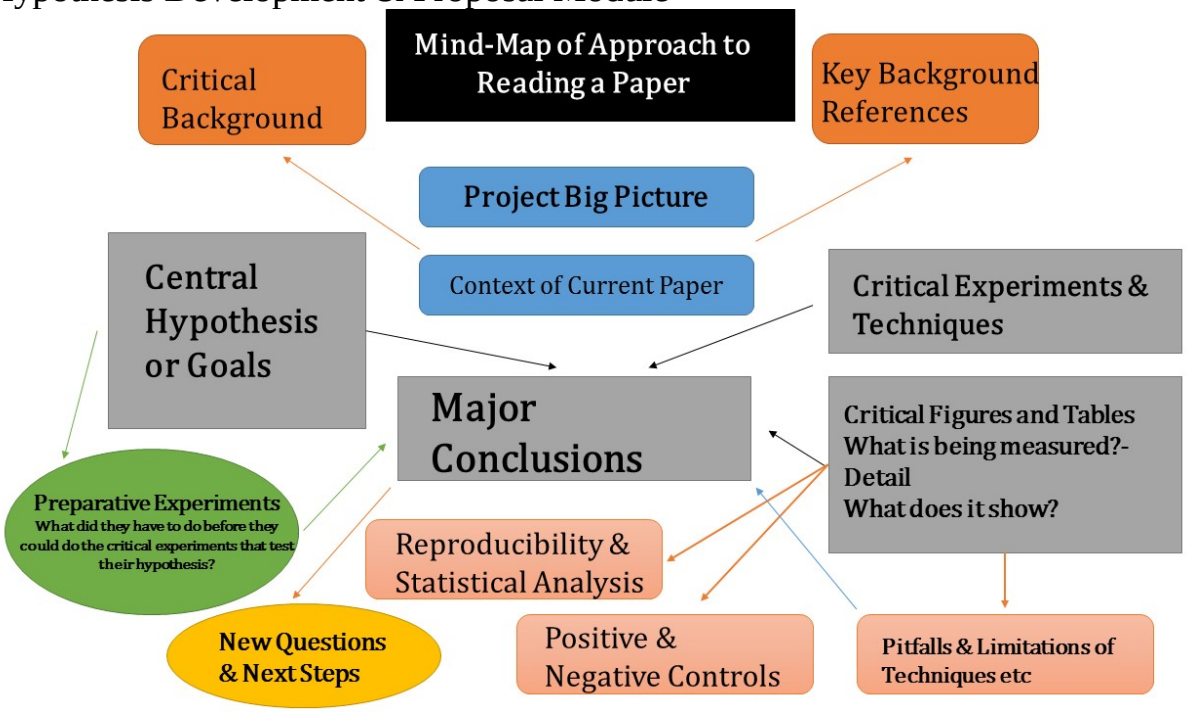
13 What are the potential pitfalls of the techniques used

14 What is the next logical step suggested by the authors

15 What other experiments do these results suggest to you

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module



You can also obtain a wealth of information about both how many malate dehydrogenases, and the types of metabolic pathways and activity that they are involved with by searching PUBMED [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi] using one of its options [the drop down menus on the left of the screen-where it says "search"] to search the protein data base. Enter the search words "Malate Dehydrogenase" and you will come up with a wide variety of Malate Dehydrogenase like proteins. Likewise, PUBMED is a valuable source of literature citations for Malate Dehydrogenase. On the subject of finding background information about a given protein, you can usually find a variety of relevant references by reading the introduction of a paper about Malate Dehydrogenase. The introduction to a well written paper will usually give you background information as to what the protein does, where it is found, often molecular characteristics of the protein and sometimes information about kinetic properties etc. For the papers that you will find useful in this sequence of laboratories, ones about structure or kinetic/regulatory properties are likely to be the most useful.

To get an idea of involvement of Malate Dehydrogenase in human disease you can search OMIN:

"OMIM, Online Mendelian Inheritance in Man. This database is a catalog of human genes and genetic disorders authored and edited by Dr. Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere, and developed for the World Wide Web by NCBI, the National Center for Biotechnology Information. The database contains textual information and references. It also contains copious links to MEDLINE and sequence records in the Entrez system, and links to additional related resources at NCBI and elsewhere."

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Finally, you can easily examine the three dimensional structure of Malate Dehydrogenase [or of course any other protein whose three dimensional structure has been determined] using the tools at the Protein Data Base[http://www.rcsb.org/pdb/]\

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

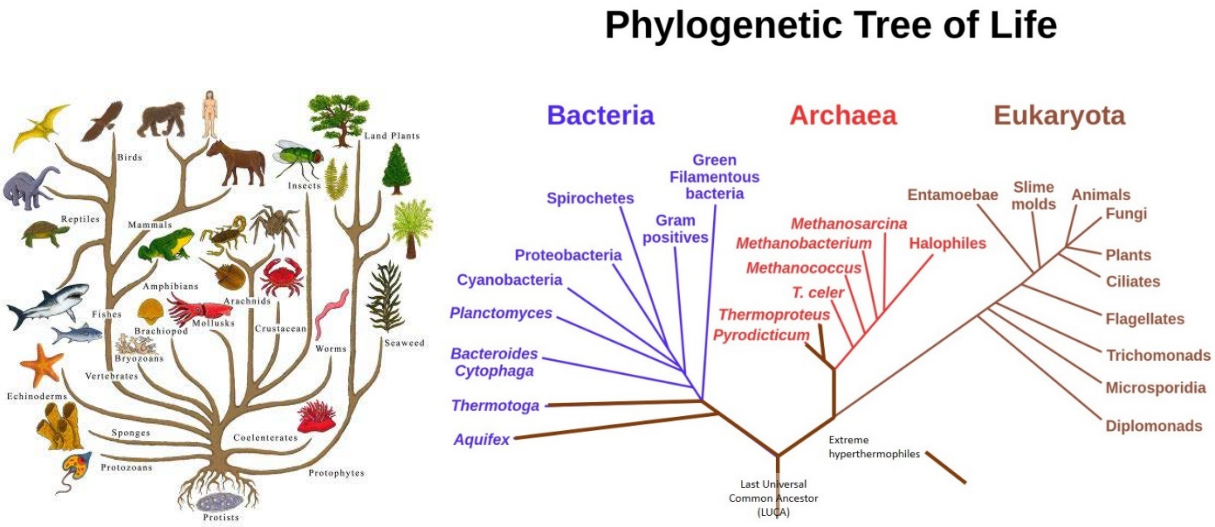
Hypothesis Development & Proposal Module

PART 1B

What is Bioinformatics?

Bioinformatics is a term often applied to the use of computer algorythms to analyze biological information. This information is often sequence information, DNA, RNA, Protein but increasingly "omics" information about levels of biologically important molecules (RNA – transcriptomics-, Proteins – proteomics – lipids – lipidomics- metabolites – metabolomics – and is used to compare metabolic or disease states, or protein structures - (structural proteomics). Bioinformatics hence can range from comparing sequence information, translating sequence information into structural information or comparing contents of cells and tissues etc. In the context of this module it is used in the context of comparing protein sequence information and what you can learn from such comparisons.

These comparisons are based on one of the major learning goals associated with evolution: What is the molecular basis of evolution? – in particular, the learning objective associated with using genetic information to categorize organisms and establish phylogenetic relationships and relates to the so called tree of life or evolutionary tree.



While originally based upon fossil evidence and structural relationships such "trees" can be constructed from DNA, RNA or Protein sequence information.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

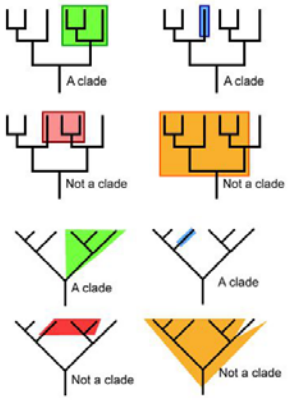
Hypothesis Development & Proposal Module

A clade is a group of organisms that have evolved from a common ancestor.

Evidence for which species are part of a clade can be obtained from the *base sequences of a gene* or the *corresponding amino acid sequence of a protein*.

Sequence differences accumulate gradually so there is a positive correlation between the number of differences between two species and the time since they diverged from a common ancestor.

Cladograms are tree diagrams that show the most probable sequence of divergence in clades.



Cladograms

Tree diagram based on similarities and differences between the species in a clade.

Base sequences/amino acid sequences

Principle of parsimony

- Computer programs calculate how species in a clade could have evolved with the smallest number of changes of base/amino acid sequences
- Does not prove how a clade actually evolved – but shows the most probable sequence

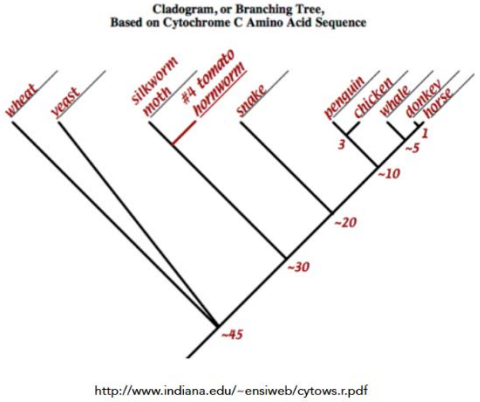
MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Animal	Amino Acid Sequences in Cytochrome-c																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Horse	gln	pro	phe	thr	thr	ala	lys	asn	lys	thr	lys	glu	glu	thr	leu	met	glu	lys	ala	thr	asn	glu
Chicken	gln	glu	phe	ser	thr	asp	lys	asn	lys	thr	gly	glu	asp	thr	leu	met	glu	lys	ala	thr	ser	lys
Frog	gln	ala	phe	ser	thr	asp	lys	asn	lys	thr	gly	glu	asp	thr	leu	met	glu	ser	ala	cys	ser	lys
Human	gln	pro	tyr	ser	thr	ala	lys	asn	lys	ile	gly	glu	asp	thr	leu	met	glu	lys	ala	thr	asn	glu
Shark	gln	gln	phe	ser	thr	asp	lys	ser	lys	thr	gln	gln	glu	thr	leu	arg	ile	lys	thr	ala	ala	ser
Monkey	gln	pro	tyr	ser	thr	ala	lys	asn	lys	thr	gly	glu	asp	thr	leu	met	glu	lys	ala	thr	asn	glu
Rabbit	gln	val	phe	ser	thr	asp	lys	asn	lys	thr	gly	glu	asp	thr	leu	met	glu	lys	ala	thr	asn	glu

	horse	donkey	whale	chicken	penguin	snake	moth	yeast	wheat
horse	0	1	5	11	13	21	29	45	46
donkey		0	4	10	12	20	28	46	45
whale			0	9	10	18	27	45	44
chicken				0	3	18	29	46	46
penguin					0	17	27	45	46
snake						0	29	46	46
moth							0	48	45
yeast								0	47
wheat									0

http://www.indiana.edu/~ensiweb/cytows.r.pdf



http://www.indiana.edu/~ensiweb/cytows.r.pdf

The sequence of porcine Malate Dehydrogenase is easily obtained by going to the protein data base and searching for the three dimensional structure of Malate Dehydrogenase

[http://www.rcsb.org/pdb/]. From the structure file, which you can download to examine later, you can get the amino acid sequence in FASTA format [a format used by most data base search engines].

Copy the sequence and enter it into BLAST [http://www.ncbi.nlm.nih.gov/BLAST/] and see how many similar sequences are found in the non-repetitive protein data base.

To see just a little of the diversity of physiological activities of these Malate Dehydrogenase like proteins randomly investigate the putative roles of some

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

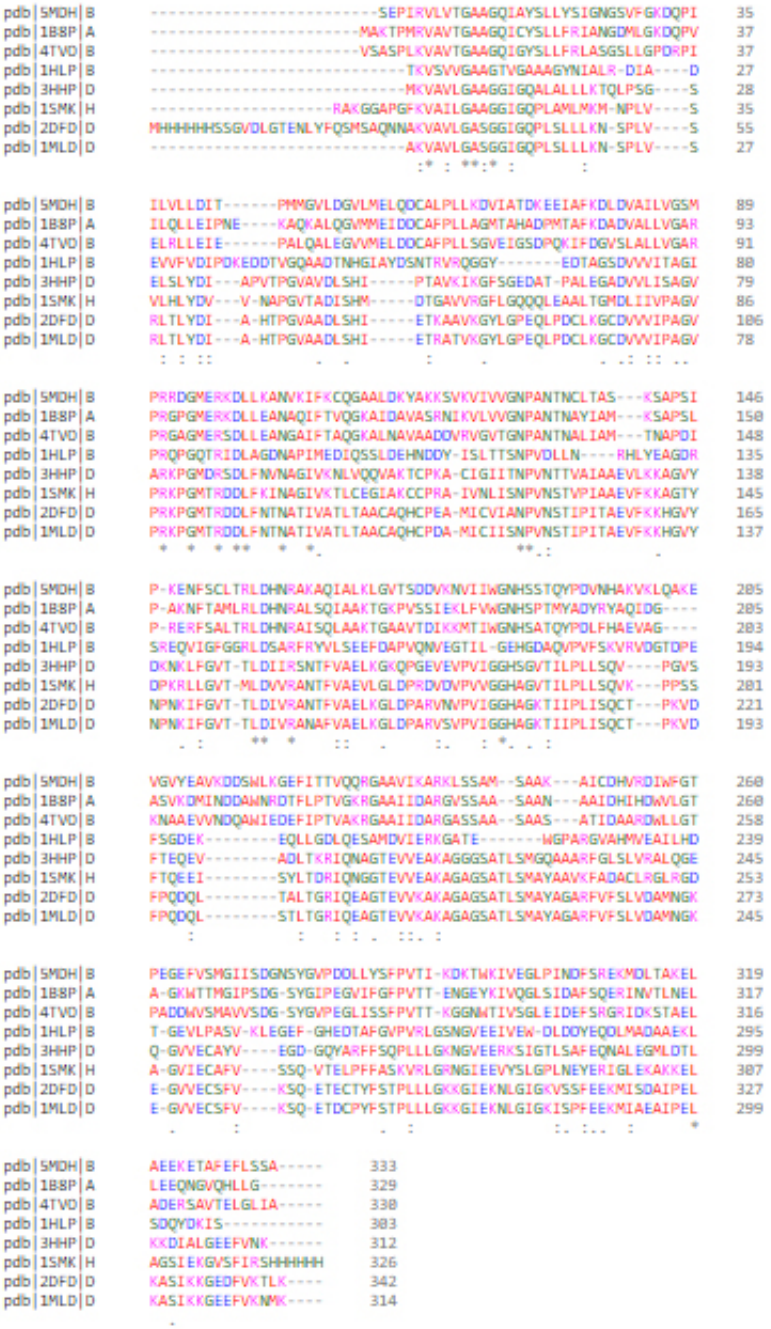
of these proteins using the links provided in the BLAST search results. You will find proteins from almost every type of life form [except viruses] and from a variety of cellular compartments.

Such cladograms can be constructed using Malate Dehydrogenase amino acid sequences: shown below is an example of a clustal alignment obtained using a number of forms of malate dehydrogenase whose 3 dimensional structure is also known., together with its associated "cladogram" showing the evolutionary relationships between the forms of Malate Dehydrogenase. Each form contains the His-Asp pair associated with catalysis, and the three arginines associated with substrate binding.

PDB identifier	Organism	Cellular Location	Catalytic Residues	The three "R"s
5MDH	Pig	cytosol	D158, H185	R91, 97, 161
1B8P	Cold Adapted Aquaspirillum		D217, H244	R95, 101, 220
4TVO	M. tuberculosis		D215, H242	R93, 99, 218
1HLP	Halophilic Archaebacterium		D166, H203	R100, 106, 169
1SMK	Watermelon	Glyoxysome	D193, H220	R124,130,196
3HHP	E. coli		D150, H177	R81, 87,152
2DFD	Human	Mitochondrion	D177, H204	R108, 114, 180
1MLD	Pig	Mitochondrion	D149, H176	R80, 86, 152

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module



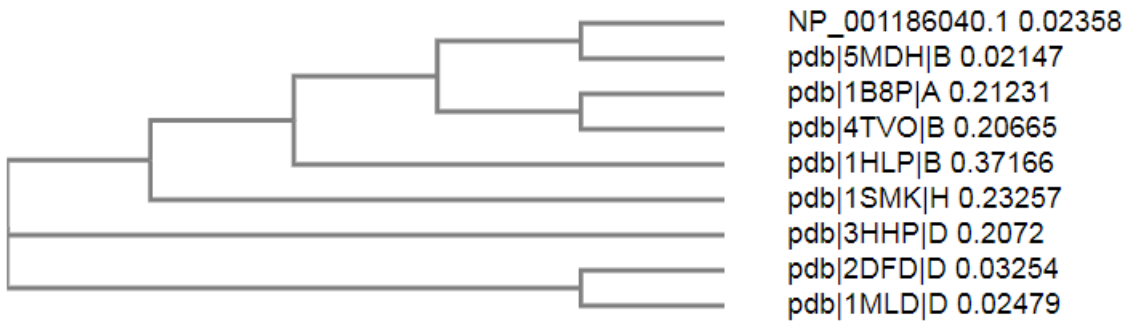
When one examines pairwise comparisons to human mitochondrial Mitochondrial Malate Dehydrogenase and examines identities (*), close homologs (:), and somewhat conserved (.) the following data is obtained:

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Form	Identities	Close Homolog	Somewhat conserved
Human Mitochondrial	325*		
Human Cytosolic	71	75	48
Pig Mitochondrial	296	14	3
Pig Cytosolic	71	54	43
M Tuberculosis	67	55	51
Archibacterium	78	53	37
E. coli	182	52	24
Aquaspirillum	72	61	43
Watermelom Glyoxysomal	175	68	28
All	18	32	23

From this type of data a cladogram can be constructed showing the evolutionary relationships between these Malate Dehydrogenases

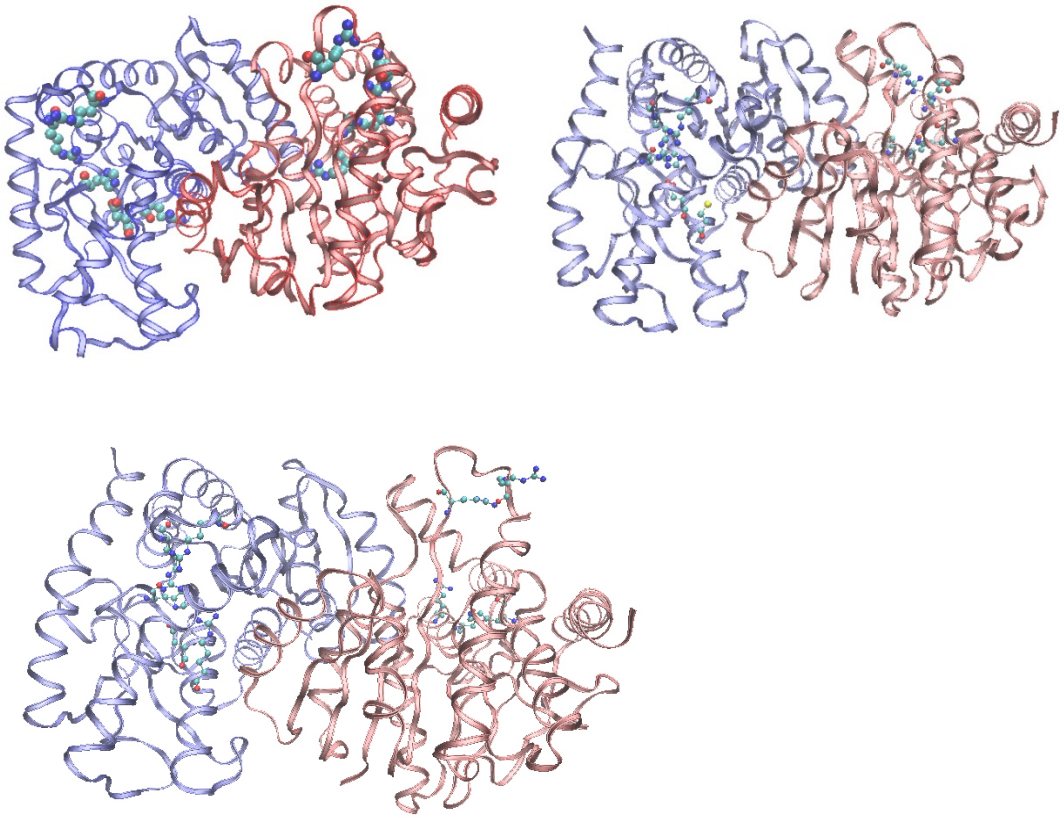


When one examines the cladogram obtained from this limited number of sequences the evolutionary relationships become clear and show that E coli, watermelon Glyoxysomal, pig mitochondrial and human mitochondrial are more related to one another than they are to the cytosolic forms.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Using molecular graphics images of the three dimensional structure of various forms can be constructed from the appropriate protein data base files:



Interestingly, despite quite a wide range of sequence identity, the three dimensional structures of each form are remarkably similar.

Big Picture Questions:

Each project will include the design and construction of a site directed mutant with appropriate characterization of the protein and additional experiments to test a hypothesis that you will construct based upon existing knowledge and a detailed bioinformatics analysis of the protein. As you think about what to propose you should think "big picture" as well as detailed picture.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

As you know malate dehydrogenase exists in every organism and in eukaryotic cells in several sub-cellular compartments- the cytosol and mitochondrion at least and often in other organelles such as the glyoxysome- and often has different biological roles in the different compartments. Likewise different types of organisms may have different needs of their malate dehydrogenase. Despite these potential differences virtually all malate dehydrogenases that have had their structure determined look somewhat similar and have conserved "functional" core amino acids- H220, D193, R196, R130 and R124 noted above that are associated with catalysis or substrate binding. In addition you might expect that any amino acid residue in the sequence that plays a critical role in folding and maintenance of the overall three dimensional structure of the protein would be conserved.

There remain many unanswered fundamental questions about MDH to be investigated. Some examples are i) Folding and oligomeric structure (while most MDHs are dimeric, some are monomers and others form tetramers, ii) Substrate specificity (there exist MDH isoforms with LDH like activity, or NADPH (vs. NADH) affinity), iii) Allosteric regulation (some forms are regulated by citrate inhibition and/or substrate inhibition), iv) Metabolon formation (mammalian

mitochondrial MDH but not cytosolic MDH can interact with other proteins including citrate synthase, both forms may interact with aspartate aminotransferase) and v) Adaptation to extreme conditions (temperature, salinity etc) and evolution (prokaryote vs eukaryote, other different branches of the evolutionary tree of life). Although some forms are well characterized structurally, the relationship of such structural and functional features are not well defined across a range of evolutionarily distinct organisms. Availability of MDH sequences for a wide array of organisms lends itself to bioinformatics approaches to develop hypotheses that can be tested experimentally.

This gives rise to a series of potential 'big picture" questions you can ask such as:

Are certain amino acid side chains responsible for global aspects such as holding the structure in the right shape, allowing the protein to 'fold" properly or contributing to the catalytic mechanism of the enzyme [H220 and D193 would be examples of this later type]

Are there differences between organelle and cytosolic forms of the enzyme that might be responsible for the different biological roles in cytosol versus organelles etc

Are there differences between for example plant and bacterial forms that might yield insight into evolutionary relationships between organisms.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Clustal Ω Analysis

To effectively design a site directed mutant you need to have available a 3-Dimensional structure, not necessarily of the exact form of the protein you wish to mutate but a closely related form. In the case here you have the three dimensional structures of a variety of malate dehydrogenases including E Coli MDH and several mammalian forms as well as watermelon glyoxysomal malate dehydrogenase. (there are many others available in the Protein Data Base and you can always create a "homology" model of a specific form that has not been crystallized- see later). You also have access to many MDH sequences through PubMed. As you start to think about what residue to mutate you should first run a clustalw sequence alignment of a number of MDH sequences including the one you wish to mutate and any that you have crystal structures for.

***Introduction**: This handout, along with the videos and links provided on Blackboard, is designed to guide you through the basics of finding and examining nucleotide and amino acid sequences and their conserved structural and functional domains. You will learn how to align the protein sequences to find important amino acids. You will learn how to render/model a structure to examine the structure and function of any known protein. **Following completion of this handout, you will be able to finish the informatics and PyMOL homework assignment linked on the class webpage.** **Note**: Watching the video tutorial located on Bb prior to this workshop will greatly enhance your ability to complete the assigned tasks in an efficient manner.*

Step 1: Basic introduction to find the function of a gene or gene product using NCBI (~1 hr)

First a simple introduction is required. Go to the NCBI page at www.ncbi.nlm.nih.gov This is the National Center for Biotechnology Information hosted by the National Library of Medicine and the National Institutes of Health. Take a moment or two and click on different options, including PubMed, BLAST, Nucleotide, Protein, and Domains & Structures. Notice under the Resources List a Training & Tutorials tab. There are a number of tutorials to advance your knowledge if you find something of interest or are looking for additional help.

Answer the following questions based on your explorations:

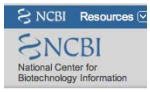
1. Go to PubMed. Select Single Citation Matcher. Enter "NADH" in Title words search box and initiate Search. How many articles were identified? Using the selectors on the left side

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

determine how many articles were published in the last 5 years. How many NADH reviews are available?

2. Go to PubChem. Search NADH in Compound category. Choose a result _____ What is the MW of this compound? Animate the 3D- conformer. What is its toxicity?



After you have seen some of the options and resources at the NCBI website, return to the NCBI home page by clicking on the NCBI icon. Now we will start to use NCBI to find the function of a gene or gene product. There are multiple ways to start. If you know the name of a gene, protein or nucleotide accession number or even part or all of a nucleotide or protein sequence, you can find information about the gene and/or protein and find the related accession numbers. Each time a nucleotide or protein sequence is entered into one of the databases, the gene/protein is given an accession number. Nucleotide accession numbers have 1 letter + 5 numerals OR 2 letters + 6 numerals. Nucleotide sequences determined from mRNA (coding region) will be noted as the mRNA accession numbers and begin with two letters (ie. NM_001282404). Gene sequences will also be listed with a Gene ID number. Protein accession numbers for the GenBank/NCBI database are 3 letters + 5 numerals. For older records, you will find a both a "version" and "gi" as part of the accession number. This is an older version of the nucleotide or protein accession number. If any change in the record of the gene or protein occurs, the version number is increased by one decimal and a new gi number is assigned. Gene and protein records can have multiple identification numbers associated with them. It is always best to keep that in mind when searching. Click here if you want to learn more information about the accession, version and gi number.

Let's begin searching for information on your protein of choice. **START WITH A GENE OR PROTEIN NAME!**

1. From the NCBI home page, click on the Search pull-down menu to select the Gene database, type the Gene Name (malate dehydrogenase) in the text box and click Go. See Gene Help for tips searching Gene.
2. Locate a desired Gene record in the results and click the symbol to open the record.
3. Functional information will be located in the Summary, Bibliography, and General gene info sections. Also, see the Links list for resources such as Conserved Domains and BioSystems.
4. Repeat using the Protein database in the pull-down menu. Explore the BLink tab under Related Information.

Step 2: Searching for DNA vs protein sequences (~1.5 hrs)

How does one find a DNA or protein sequence?

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

- Go to PubMed and search for malate dehydrogenase. Look at links – what kind are they? Think about how many papers you would have to read to find the sequence.
- Go to the pulldown window next to the SEARCH box and find the Nucleotide database option. Enter malate dehydrogenase. Can you find it? Try searching for MDH, MDH1, MDH2 and MDH2 Homo sapiens. Sometimes it is hit and miss with an educated guess. Try the advanced search function. Does this limit your hits to more relevant entries? (Using a specific accession number found in a journal publication can help you narrow the search tremendously).
- Find the human MDH2 gene and protein records. Are there different variants? If so, what are the differences*?*

Be careful when you click on the different links.

You may be taken to a different database such as "gene", "nucleotide" or "protein". The database for each page will be shown at the top and will have very different information.

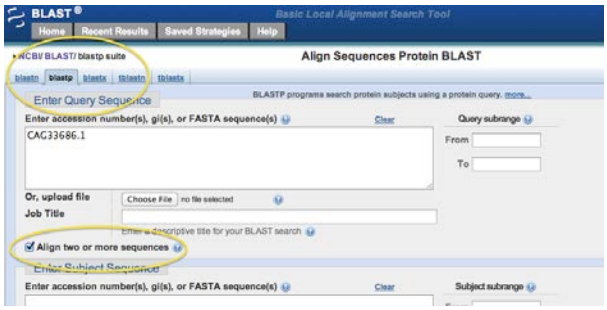
- In the nucleotide window, search for gi: NM_005918.3
- Scan through the record. On the right of the window, you will find options to use the information with the record. Notice there are links to conserved domains, sequence features and articles about the gene. Take a moment to look at those links. Continue to search down the record. You will see information about the size of the gene and/or mRNA, the location of the protein, if it is a variant or not, etc. Scanning further you will find several PubMed references. Read the titles and click on one or two of the links.
- Scan down to find and record the **CDS** number – The CDS is the "coding DNA sequences". This is where the DNA sequence start site is. – A common mistake is to assume that the first nucleotide is the first amino acid codon.
- Continue down to find both the nucleotide sequence and the amino acid translation.
- At the top of the record, click on the FASTA link. This is a format used to compare sequences.
- Click on the Graphics link. Search for hyperlinked sections. Notice the NAD binding site, dimerization interface, and so on. Here you can find the graphical representation of a number of important domains of the protein/gene.
- Search using the nucleotide accession number under the protein database pulldown menu. What happens? Does the window stay in the protein database? Now try searching the protein database using NP_005909.2 What is the difference? Is this the same protein sequence as before?

Step 3: Aligning amino acid sequences (~1.5 hr)

If you need to compare two sequences: say against a protein you have that doesn't have a solved structure, or you read a paper that indicates amino acid 212 is important for something. How would you directly line the sequences up? One way is to manually slug it out by hand. Not fun. Instead, you can use computational alignment tools to examine similarties amongst nucleotide or protein sequences.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module



The Basic Local Alignment Search Tool (BLAST) is a program on the NCBI website that will compare two sequences and match these sequences based on best matches. Very few if any sequences are the same, have the same length or will start with the same sequence. Therefore BLAST helps you align two sequences to find where amino acids or nucleotides are common or unique.

BLAST has different versions for nucleotides and proteins. You must use the correct sequence accession number – a common mistake is using a nucleotide accession number in a protein BLAST search. Click on the BLAST home page and familiarize yourself with the nucleotide and protein BLAST windows.

Click on the NCBI icon in the top left corner, and use the pull down menu to select the Protein: Sequence Database. Then type in human MDH1 and find the *Homo Sapiens* record (right hand side of the screen).

- You will find the "Run BLAST" link to the right. Click on it. Notice your protein record has already been loaded into the BLAST protein search. Convince yourself that you are in the protein BLAST by looking at the tab. It should read "blastp".
- Click on the "Align two or more sequences box and enter the gi, FASTA, or other protein accession number for MDH2.
- Find which amino acid in MDH1 corresponds to residue 58 in MDH2.
- Do the same thing using two nucleotide sequences. Pick any two MDH sequences. You will need to make certain you change the program settings from blastp to blastn (n = nucleotide, p = protein) in the pull-down window.
- Finally, go to the protein record number and find out which triplet bases code for the MDH1 amino acid you queried. Don't forget that the first nucleotide is not the start site. See above for the CDS number! This will be important when comparing critical amino acids and domains between different isoforms of MDH.

A second tool for examining sequence alignments of more than three sequences is Clustal Omega. Multiple sequence alignments (MSAs) are used to examine conservation of an amino acid or nucleotide at a position across several sequences of the same gene or protein. Correlation between conservation and function has been observed suggesting an evolutionary pressure for the organism to preserve a define residue or sequence in order to maintain function.

What sequences should you align? You will use MDH sequences for this exercise but which ones? You should think of a question you would like to pose based upon sequence conservation. For example, malate dehydrogenase occurs in several organelles (mitochondria, glyoxosomes) and the

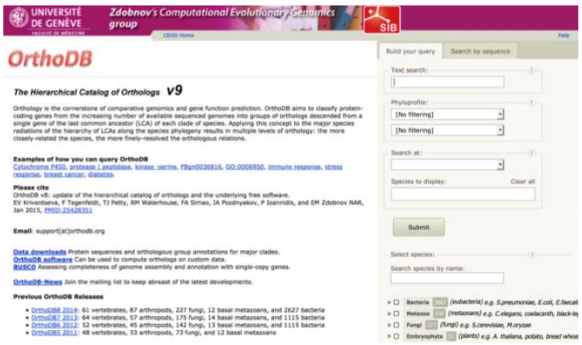
MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

cytosol. Do organelle specific versus cytosolic MDH sequence have distinct patterns of conserved residues? Similarly, MDH is found from archaea, bacteria, plants, through all metazoans. Is the conservation pattern of residues different amongst these kingdoms?

Now that you have a question that forms the basis for the sequences you will align, where can you find these sequences? Several places – NCBI as outlined above but two other databases can serve as a resources: OrthoDB and UniProt. Choose ***either*** OrthoDB or UniProt to complete this exercise.

OrthoDB



Under build your query enter the protein name. Using a full name such as malate dehydrogenase is not as successful as MDH or MDH1 or MDH2. You can limit the search to certain species by using the Select Species function.

Your search will be returned in groups. At the top, click on the Get All FASTA. Copy and paste this list of sequences into a text editor. Select 15-20 sequences that you will align.

Tidying up sequences…

The FASTA sequence format is

>annotation (carriage return)

sequence of 60 characters per line followed by a carriage return

When the sequences are entered into Clustal Omega, the program will use the annotation as the name of the sequence until a space is encountered. For example, human MDH1: "sp|P40925|MDHC_HUMAN Malate dehydrogenase, cytoplasmic OS=Homo sapiens GN=MDH1 PE=1 SV=4" is named "sp|P40925|MDHC_HUMAN." If needed, one can shorten the annotation or rename the sequence (i.e. human) but always keep a record of the original downloaded sequence as it contains an accession number that is universal.

Clustal Omega

Paste FASTA formatted sequences in Step 1. In Step 2, choose Clustal w/ numbers. In Step 3, submit your sequences. If you are aligned >25 sequences you may want to have the results sent to your email account. Download your alignment file. This can be opened in a text editor program but must be formatted with 0.5 inch margins, Courier New font, 10 pt size. Alternatively you should color the alignment- this helps see commonalities of amino acid side chain properties and either print as an appropriately formated pdf file of do a screen capture and save as an image file. In the results summary, examine the percent identity matrix. This is a quantitative measure of the similarity between the entered sequences. Under "Help & Documentation" you

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

will find Clustal Omega FAQ that provides more information on the program including the legend for the conservation symbols and colors used in the alignment.

UniProt

Search malate dehydrogenase, MDH, MDH1 or MDH2. Filter by "Reviewed" on the top left. You can further filter by organism. Select entry by checking in left column. Your results may spread across several pages. If selections are made on one page, "Add to basket" before moving to a new page or selections will be removed. When you have selected your sequences and added them to the basket, click on the basket in the upper right corner. Select the sequences that you will align and then select "Align." This sends the sequences to an integrated Clustal Omega alignment tool. In the left column, you can make several selections to annotate your sequence or highlight amino acid properties. You can download the alignment and open in a text editor program.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Remember, your selection of sequences should be based upon the big picture question you might be interested in asking.

After you have performed a ClustalΩ alignment [use the color coding option for residues] you should map onto this homology alignment the structure of watermelon glyoxysomal malate dehydrogenase, focusing on residues in and around the active site, and on residues at the subunit interfaces

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Learning Objective	*Excellent*	*Acceptable*	*Poor*	*Unacceptable*
Find and use appropriate literature to illustrate the big picture aspects of the work	Identifies big picture of field or component of study background, states focus of the study, provides significance of study.	Funnel approach applied to intro but one or two aspects missing (i.e. broad field or significance, etc.)	Most components of funnel approach present, but not presented in logical order.	Funnel approach not taken nor are the components presented. Lacks enough information for reader to understand current field or
Find and use appropriate literature to document specific background to the enzyme and their hypothesis	Introduces model system, describes relevant previous study(ies) and their findings pertaining to focus of the study. Not a history section.	Introduces model system but missing relevant background to current study or does not relate to current study or is overly long with non-essential background.	Introduces model system but does not provide relevant background to current study.	Does not introduce model system or relevant background.
Use appropriate data bases to obtain sequence information and analyze using Clustal Omega	Rationalizes choice of sequences, justifies number of sequences, illustrates trends in alignment appropriately, discusses evolutionary relationships of sequences	Lacks 1 or 2 points of excellent	Runs Clustal omega but does not select sufficient sequences to draw conclusions	No rationale to selection of sequences, lacks 3-4 points of excellent
Use the protein data base to obtain 3D coordinates for a protein and use Pymol or other visualization tools to illustrate key features of the protein and their hypothesis	Produces several visually effective pictures clearly illustrating appropriate different aspects of structure-function relationships	Pictures show basic ideas but not as effective as excellent	Pictures lack clarity or necessary detail	Minimal image of protein. No detail or insight into structure-function relationships shown
Use appropriate bioinformatics tools to design primers for mutagenesis	Accurately designs primers accounting for codon usage, GC Clamp, & Tm	Designs primers but does not give appropriate detail in excellent answer	Designs primers but does not take into account codon usage etc	Mistakes in primer design, wrong codon, insufficient length etc
Depending upon format of presentation: General flow/organization. Grammar/spelling/general attention to detail	Logical flow from global to particular study point of view. Engaging writing style. Clearly connects ideas.	Solid order & structure. Inviting writing style. Effectively moves the reader through the text.	Organization is functional; some order lacks logical pattern and structure.	Lacks cohesive structure, difficult to follow.
Team Work and Peer Evaluation	Clearly engaged in and contributing to all aspects of the project, discussions and presentations	Contributes to many but not all aspects of the project	Contributes to limited aspects of the projects	Minimally, if at all, contributes.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

PART 2

Learning Goal: Understand the various roles that non-covalent interactions may play in the structure and function of an enzyme

Exploring Structure-Function Relationships in Malate Dehydrogenase II: Using Molecular Visualization tools

After you have performed a ClustalW alignment [use the color coding option for residues] you should map onto this homology alignment the structure of watermelon glyoxysomal malate dehydrogenase, focusing on residues in and around the active site, and on residues at the subunit interfaces

How to Find Residues in and Around the Active Site

From the various pdb files you have looked at you will have found that there is a histidine [H177 using E Coli numbering, H220 in watermelon], and an aspartate [D150-e coli or D193 in watermelon] residue in the active site. Using pkin and mage identify these three residues and then determine what other residues are within say 5-6A of the active site: any of these could be targets for mutagenesis. Carefully examining the substrate binding site using Pymol or VMD visualization reveals the presence of a series of positively charged residues- as you might expect- that attract the negatively charged carboxyl groups of the substrate. The residues too could be targeted for mutagenesis although if you do a literature search of site directed mutagenesis of Malate Dehydrogenase you will find that Bell et al " Structural Analysis of a Malate Dehydrogenase with a Variable Active Site", J. Biol. Chem 276: 31156-31162 have probed the role of Arginine, R153 [e coli, R196 in watermelon] in substrate binding and the stabilization of the hydroxyl/keto group during catalysis. There are other residues in the substrate binding site whose role could be probed by site directed mutagenesis.

If you decide upon one of these residues you should decide upon some hypothesis that you will test by site directed mutagenesis. Remember that the H, R and D have roles defined by the crystal structure which you could also test.

How to Find Residues at the Subunit Interface.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

If you choose one of the pdb files that contains the dimer of MDH, for example 1BDM.pdb and examine it in pkin and mage you can map out the subunit interface. Once you have done this you can then decide what to mutate. Why would you want to mutate a residue at the interface: 1] to increase the strength of the interface, 2] to decrease the strength of the interface, or 3] because you think that the residue might be involved in subunit-subunit communication

Before going on to the Quikchange protocol to design your mutant you should have in mind what residue you wish to change and what you want to change it to. Briefly outline why you have made the choice that you have made and how, after expression of the protein, you would determine whether or not the mutant has had the hypothesized effect.

Structure and Molecular Graphics

Over 20,000 three dimensional structures have been determined, mainly by X ray crystallography but some by NMR. These structures are deposited in th "Protein Data Bank – PDB – which is the single world wide repository for the processing and distribution of 3-D biological macromolecular structure data. As will be discussed below, pdb files are of a uniform type and can be used by many molecular graphics programs. For many proteins therefor there is now a reasonable chance that the three dimensional structure has been experimentally determined and the coordinates available through the PDB.

Frequently with proteins while the exact protein has not been crystallized and its three dimensional structure determined, it is likely that a homologous protein has had its structure determined and a careful examination of the three dimensional structure of a homologous protein together with insight obtained from bioinformatics approaches such as Clustal can be very revealing in terms of structure function relationships in the protein and helpful in the design of appropriate site directed mutagenesis experiments.

In the case of Malate Dehydrogenase while the three dimensional structure of Glyoxasomal MDH has yet to be determined there are a variety of three dimensional structures available from sources such as E Coli [1EMD.pdb] , pig cytoplasmic [4MDH.pdb & 5MDH.pdb] and pig mitochondrial [1MLD.pdb]. The pdb files for these three structures can be downloaded from http://www.rcsb.org/pdb/. Analysis of these structures can be used to show a variety of regions of the protein of potential interest such as the catalytic site and residues potentially involved in catalysis, the substrate binding site, and subunit interface regions.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

What Information Does a pdb File Contain

The pdb file of a protein contains a wealth of information in addition to the actual three dimensional coordinates of the structure. A pdb file is simply a text file and can be opened in any word processing program such as “Word”. You should set the page margins so that you see the files as written rather than with lines wrapped around-this usually is accomplished by setting the right and left hand margins at 0.5 inches.

The pdb file for the E Coli enzyme begins:

```
HEADER    OXIDOREDUCTASE(NAD(A)-CHOH(D))          25-MAR-93   1EMD      1EMD   2
COMPND    MALATE DEHYDROGENASE (E.C.1.1.1.37)                           1EMD   3
SOURCE    (ESCHERICHIA COLI)                                            1EMD   4
AUTHOR    M.D.HALL,L.J.BANASZAK                                         1EMD   5
REVDAT   1   31-OCT-93 1EMD    0                                        1EMD   6
JRNL        AUTH   M.D.HALL,L.J.BANASZAK                               1EMD   7
JRNL        TITL   CRYSTAL STRUCTURE OF A TERNARY COMPLEX OF           1EMD   8
JRNL        TITL 2 ESCHERICHIA $COLI MALATE DEHYDROGENASE, CITRATE     1EMD   9
JRNL        TITL 3 AND /NAD$ AT 1.9 ANGSTROMS RESOLUTION               1EMD  10
JRNL        REF    TO BE PUBLISHED                                     1EMD  11
JRNL        REFN                                            353  1EMD  12
REMARK   1                                                            1EMD  13
REMARK   2                                                            1EMD  14
REMARK   2 RESOLUTION. 1.9  ANGSTROMS.                                1EMD  15
REMARK   3                                                            1EMD  16
REMARK   3 REFINEMENT.                                                1EMD  17
REMARK   3   PROGRAM                   X-PLOR                         1EMD  18
REMARK   3   AUTHORS                   BRUNGER                        1EMD  19
REMARK   3   R VALUE                   0.195                          1EMD  20
REMARK   3   RMSD BOND DISTANCES       0.012  ANGSTROMS               1EMD  21
REMARK   3   RMSD BOND ANGLES          1.65   DEGREES                 1EMD  22
```

The “header” simply gives information about the general class of enzyme, the date of the file and the file name.

The final columns are the file name and line number which runs throughout the file.

The “Compound” line gives the name of the enzyme and the Enzyme Commision number.

“Source” indicates the organism that the protein was obtained from, in this case E Coli.

”Author” is the person or people who published the structure

“RevDat”, for Revision Date is to indicate when revisions to the file were received.

“JRNL” is the citation to the relevant publication

“Remark” lines are for commentary about the structure and usually indicate the resolution, the program used for the refinement of the structure, the R factor, which indicates how good the data is and is defined by:

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

$R = \Sigma|Fo-Fc|/\Sigma Fo$

Where Fo is the actual data point and Fc is the modeled parameter.

and RMSD [root mean square deviations] for the bond distances and bond angles in the structure.

Next in the pdb file comes the 'SEQRES" section which lists the amino acid sequence of the protein with appropriate "FTNOTE" lines-in this case indicating that residue 120 is a cis-proline.

Next comes a listing of lines for "HET" which indicates whether any other molecules are in the structure-this is often the substrate, analog or inhibitor etc. Followed by the Formula of the HET molecules and a line for water molecules in the structure.

This is followed in turn by listings of structure, first "HELIX", then "SHEET" and finally "TURN" lines

The section:

```
SITE     1 ACT  5 ARG    81  ARG    87  ASP   150  ARG   153            1EMD  94
SITE     2 ACT  5 HIS   177                                            1EMD  95
CRYST1  116.800   43.050   83.740  90.00 130.10  90.00 C 2          4  1EMD  96
ORIGX1      1.000000  0.000000  0.000000        0.00000                 1EMD  97
ORIGX2      0.000000  1.000000  0.000000        0.00000                 1EMD  98
ORIGX3      0.000000  0.000000  1.000000        0.00000                 1EMD  99
SCALE1      0.008562  0.000000  0.007210        0.00000                 1EMD 100
SCALE2      0.000000  0.023229  0.000000        0.00000                 1EMD 101
SCALE3      0.000000  0.000000  0.015612        0.00000                 1EMD 102
```

The "Site" lines indicate in this case that 5 residues that are part of the active site have been identified as R81, R87, D150, R153 and H177.

"CRYST!" indicates the unit cell parameters, the space group, in this case c2 and the z score: the number of asymmetric units per unit cell, in this case 4.

Finally the actual three dimensional coordinates begin:

```
ATOM      1  N   MET     1      18.501  11.209  -4.601  1.00 21.76      1EMD 103
ATOM      2  CA  MET     1      18.761  10.071  -3.682  1.00 20.15      1EMD 104
ATOM      3  C   MET     1      18.433   8.774  -4.397  1.00 17.06      1EMD 105
ATOM      4  O   MET     1      17.480   8.716  -5.177  1.00 17.88      1EMD 106
ATOM      5  CB  MET     1      17.893  10.173  -2.409  1.00 24.88      1EMD 107
ATOM      6  CG  MET     1      18.683  10.087  -1.111  1.00 20.12      1EMD 108
ATOM      7  SD  MET     1      19.416  11.699  -0.865  1.00 38.49      1EMD 109
ATOM      8  CE  MET     1      21.095  11.417  -1.207  1.00 35.86      1EMD 110
```

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

```
ATOM      9  N   LYS     2      19.205   7.737  -4.110  1.00 15.56      1EMD 111
```

In this section, each atom is numbered and the element type and some additional information given:

Eg CA is the alpha carbon, CB is the beta carbon etc before the residue type and number given. After the residue number the next three numbers are the three dimensional, Cartesian coordinates of the atom, followed by the "occupancy" of the electron density for that atom: usually 1.0. The next number, the so-called B factor or temperature factor gives an indication of the local motion of the atom: a low number indicates little motion while a high number indicates significant motion of the atom. While it is quite usual for exposed side chains such as the charged or hydrophilic residues to have relatively high temperature factors [upto 30-50] the backbone atoms often have single digit temperature factors unless significant motion is observed.

The final two columns are simply the file name and the file line number.

The end of the protein sequence [remember there may be more than one polypeptide chain: usually indicated by 1A, 1B, 1C etc comes the ter statement:

```
TER    2279     LYS   312                                1EMD2381
```

Indicating the end of the protein.

This is followed by the coordinates of any heteromolecules such as ligands or water: in this case the ter statement is followed by:

```
HETATM 2280  C1  CIT   313       5.426  -7.608  15.720  1.00 25.43      1EMD2382
HETATM 2281  O1  CIT   313       4.760  -7.292  16.696  1.00 22.77      1EMD2383
HETATM 2282  O2  CIT   313       5.710  -8.783  15.470  1.00 23.22      1EMD2384
HETATM 2283  C2  CIT   313       5.220  -6.691  14.572  1.00 22.61      1EMD2385
HETATM 2284  C3  CIT   313       5.865  -5.268  14.934  1.00 23.43      1EMD2386
HETATM 2285  O7  CIT   313       7.150  -5.379  15.668  1.00 20.35      1EMD2387
HETATM 2286  C4  CIT   313       6.220  -4.401  13.666  1.00 23.09      1EMD2388
HETATM 2287  C5  CIT   313       6.872  -5.350  12.811  1.00 31.49      1EMD2389
HETATM 2288  O3  CIT   313       7.741  -6.011  13.344  1.00 34.43      1EMD2390
HETATM 2289  O4  CIT   313       6.506  -5.733  11.536  1.00 32.18      1EMD2391
HETATM 2290  C6  CIT   313       4.833  -4.564  15.697  1.00 21.46      1EMD2392
HETATM 2291  O5  CIT   313       5.262  -3.941  16.722  1.00 17.15      1EMD2393
HETATM 2292  O6  CIT   313       3.669  -4.640  15.174  1.00 19.96      1EMD2394
HETATM 2293 AP   NAD   314       6.405 -11.130   4.954  0.69 26.19      1EMD2395
HETATM 2294 AO1  NAD   314       6.596 -12.584   5.180  0.69 27.85      1EMD2396
HETATM 2295 AO2  NAD   314       5.164 -10.631   4.350  0.69 19.49      1EMD2397
HETATM 2296 AO5* NAD   314       7.698 -10.375   4.274  0.69 28.25      1EMD2398
HETATM 2297 AC5* NAD   314       9.100 -10.752   4.424  0.69 26.99      1EMD2399
```

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

```
HETATM 2298 AC4* NAD   314       9.908 -10.815   3.112  0.69 30.23      1EMD2400
HETATM 2299 AO4* NAD   314      11.245 -11.256   3.417  0.69 29.14      1EMD2401
HETATM 2300 AC3* NAD   314       9.377 -11.852   2.102  0.69 30.36      1EMD2402
HETATM 2301 AO3* NAD   314       8.918 -11.300   0.864  0.69 31.71      1EMD2403
HETATM 2302 AC2* NAD   314      10.516 -12.828   1.856  0.69 30.56      1EMD2404
HETATM 2303 AO2* NAD   314      10.532 -13.467   0.558  0.69 32.13      1EMD2405
HETATM 2304 AC1* NAD   314      11.642 -11.888   2.199  0.69 29.49      1EMD2406
HETATM 2305 AN9  NAD   314      12.946 -12.548   2.294  0.69 30.68      1EMD2407
HETATM 2306 AC8  NAD   314      13.403 -13.455   3.186  0.69 30.38      1EMD2408
HETATM 2307 AN7  NAD   314      14.746 -13.503   3.170  0.69 29.26      1EMD2409
HETATM 2308 AC5  NAD   314      15.076 -12.598   2.249  0.69 30.31      1EMD2410
HETATM 2309 AC6  NAD   314      16.347 -12.115   2.003  0.69 27.98      1EMD2411
HETATM 2310 AN6  NAD   314      17.465 -12.733   2.307  0.69 26.75      1EMD2412
HETATM 2311 AN1  NAD   314      16.406 -11.054   1.195  0.69 30.45      1EMD2413
HETATM 2312 AC2  NAD   314      15.323 -10.531   0.634  0.69 29.90      1EMD2414
HETATM 2313 AN3  NAD   314      14.102 -10.997   0.817  0.69 30.17      1EMD2415
HETATM 2314 AC4  NAD   314      13.972 -12.018   1.668  0.69 30.56      1EMD2416
HETATM 2315  O3  NAD   314       6.470 -10.585   6.442  0.69 25.22      1EMD2417
HETATM 2316 NP   NAD   314       6.574  -9.067   6.899  0.69 29.30      1EMD2418
HETATM 2317 NO1  NAD   314       5.404  -8.704   7.740  0.69 22.95      1EMD2419
HETATM 2318 NO2  NAD   314       6.770  -8.191   5.717  0.69 18.36      1EMD2420
HETATM 2319 NO5* NAD   314       7.869  -9.394   7.848  0.69 23.27      1EMD2421
HETATM 2320 NC5* NAD   314       8.765  -8.341   8.225  0.69 27.18      1EMD2422
HETATM 2321 NC4* NAD   314       9.965  -8.616   9.126  0.69 28.17      1EMD2423
HETATM 2322 NO4* NAD   314      10.320  -7.290   9.589  0.69 33.05      1EMD2424
HETATM 2323 NC3* NAD   314       9.643  -9.373  10.416  0.69 28.95      1EMD2425
HETATM 2324 NO3* NAD   314      10.803  -9.831  11.132  0.69 19.79      1EMD2426
HETATM 2325 NC2* NAD   314       8.983  -8.282  11.258  0.69 32.60      1EMD2427
HETATM 2326 NO2* NAD   314       8.693  -8.760  12.602  0.69 31.87      1EMD2428
HETATM 2327 NC1* NAD   314      10.080  -7.227  11.022  0.69 34.45      1EMD2429
HETATM 2328 NN1  NAD   314       9.747  -5.837  11.353  0.69 40.34      1EMD2430
HETATM 2329 NC2  NAD   314      10.008  -5.147  12.560  0.69 44.29      1EMD2431
HETATM 2330 NC3  NAD   314       9.629  -3.816  12.717  0.69 44.76      1EMD2432
HETATM 2331 NC7  NAD   314       9.879  -3.032  13.977  0.69 44.86      1EMD2433
HETATM 2332 NO7  NAD   314       9.185  -2.066  14.289  0.69 48.45      1EMD2434
HETATM 2333 NN7  NAD   314      10.925  -3.436  14.647  0.69 42.75      1EMD2435
HETATM 2334 NC4  NAD   314       8.913  -3.178  11.754  0.69 45.41      1EMD2436
HETATM 2335 NC5  NAD   314       8.584  -3.871  10.603  0.69 45.19      1EMD2437
HETATM 2336 NC6  NAD   314       9.015  -5.178  10.398  0.69 42.63      1EMD2438
HETATM 2337  O   HOH   315       1.022   8.231  -1.217  1.00 23.42      1EMD2439
HETATM 2338  O   HOH   316       8.525  -7.240   3.696  1.00 16.79      1EMD2440
HETATM 2339  O   HOH   317      20.324 -12.888  -0.305  1.00 30.95      1EMD2441
```

[Note not all of the waters are shown here]

indicating the presence of Citrate and NAD in the structure as well as the presence of crystallographic water molecules.

One of the nice things about pdb files is that you can easily copy sections of the file and use just those sections in a viewer of some type. For example if you have two subunits and several ligands it is often convenient [see why later] to make separate files of each subunit and each bound ligand. You should do this for the 1EMD file and save each set of coordinates in plain text format. Plain text format is read as pdb format by most molecular visualization programs.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

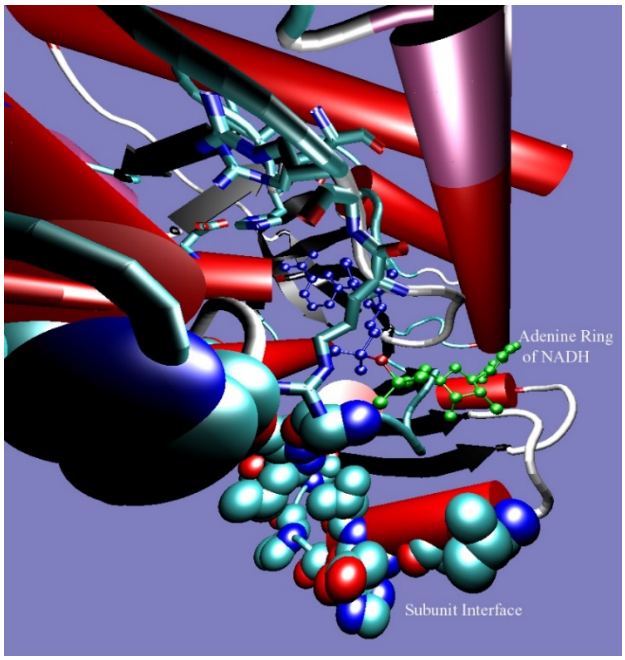
Hypothesis Development & Proposal Module

Molecular Visualization Programs.

VMD

VMD is a powerful graphics program which allows you to display the contents of a pdb file. The program gives you powerful graphics control over the representation of the structure and also allows you to zoom in to a structure to focus on a relatively small area of the structure. Since the program will read any file that is in the format of a pdb file it will read the files that you can create from a pdb file. As mentioned earlier if you have a pdb file that has in addition to the protein, bound substrates etc and you create a separate file for the substrate coordinates then VMD will display that file too. Since one of the advantages of VMD is that you can load the same set of coordinates several times, you can represent a structure in a number of ways and as you will see superimpose side chains onto a schematic representation of a structure. Since the scale of the display is set by the last pdb file you load it is possible to create a represwentation of a protein and then load the pdb file for just the substrate and in effect zoom into the active site automatically. This is extremely useful in terms of examining what is in a particular site.

Earlier with the 1EMDpdb file, the crystal coordinates of E Coli Malate Dehydrogenase, you created separate files for NAD and for Citrate and of course have the original pdb file it is possible to create a series of images that allow you to examine in detail what residues are around the active site, the citrate [substrate] site and of course the cofactor site. Examples of these are shown below.

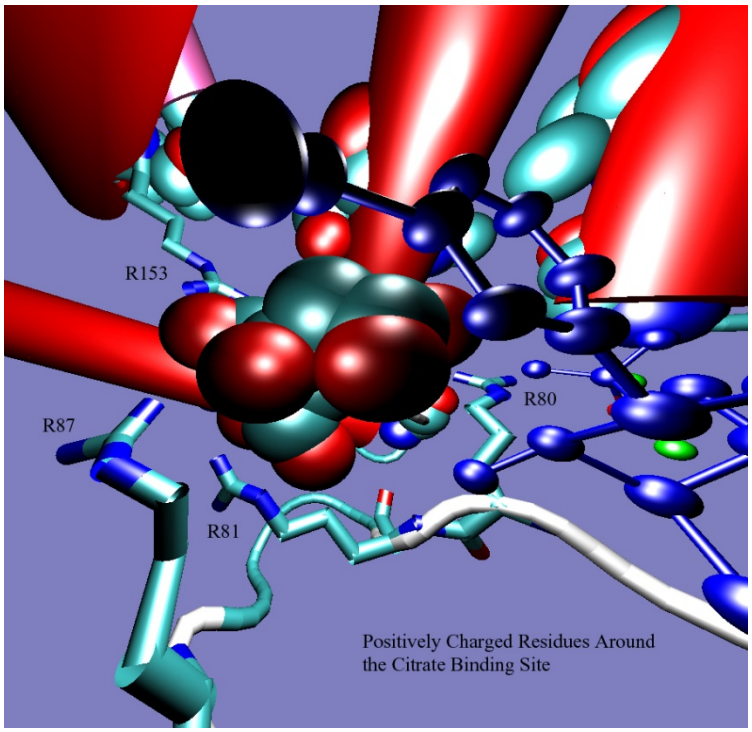


This image was created by loading the 1EMD file twice, representing one data set in cartoon mode and the other designating particular side chains to show with the "resid" command. Finally the cofactor pdb file that was created from the original 1EMD.pdb file was loaded and the cofactor atoms represented in ball and stick fashion.

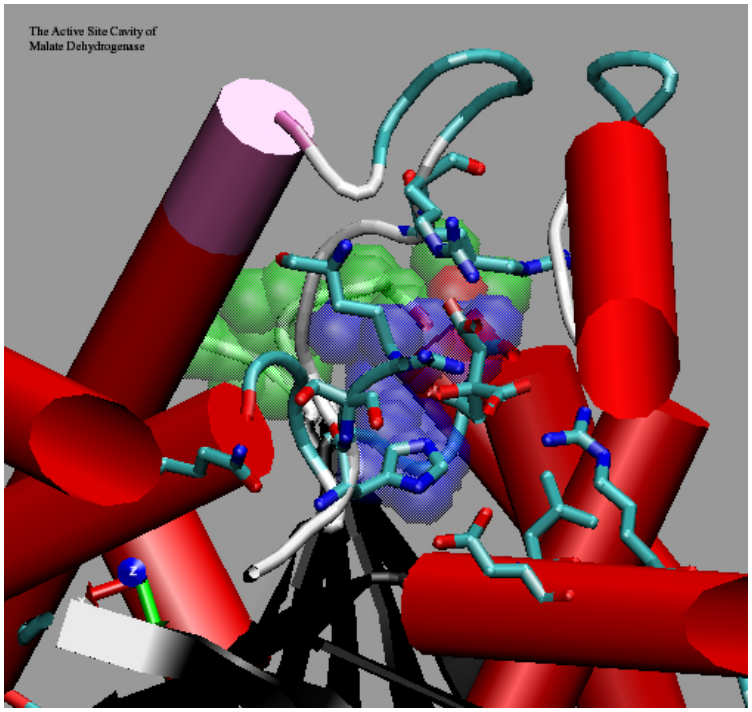
This image shows that the adenine ring of the cofactor is close to some of the residues located at the dimmer interface. Hence adenine ring conformational

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module



This image was created by again loading two sets of the 1EMD.pdb file, representing one in cartoon fashion and the other in "liquorice". Only certain side chains were designated using the resid command to be shown in the liquorice set. After this had been done the pdb file created for NAD was loaded and represented in ball and stick fashion. Finally the pdb file for Citrate was loaded and is shown represented in van der waals

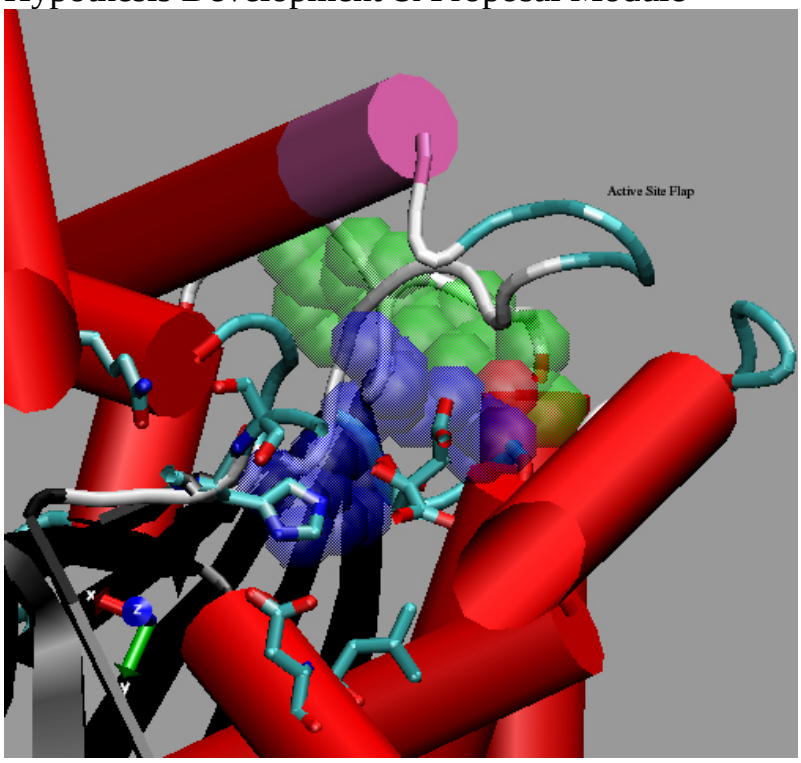


In this image various residues around the active site are shown including an aspartate, a leucine, an asparagine and the arginines shown in the previous picture. You should attempt to define which aspartate, asparagine,and leucine are shown

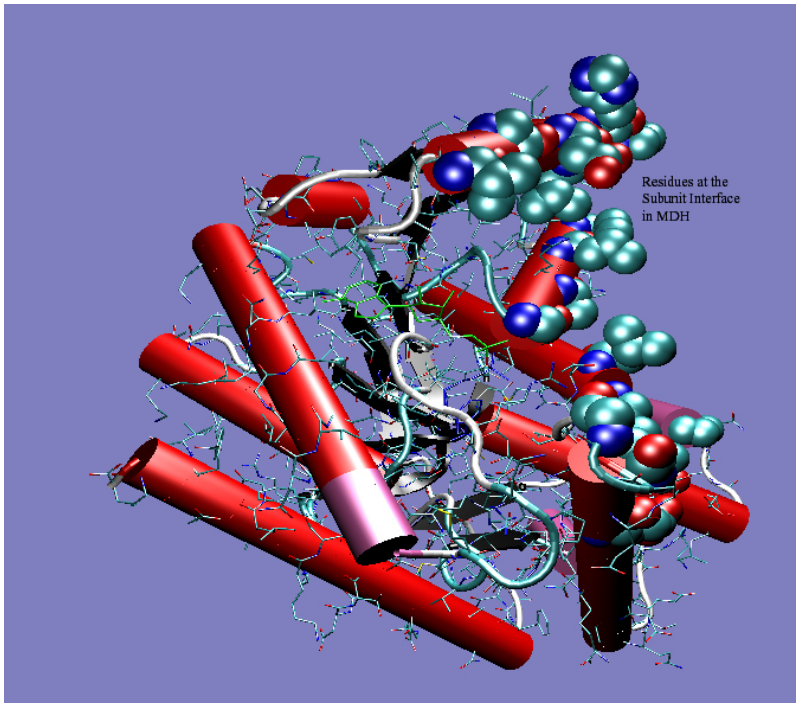
The Citrate in the structure is shown (on the right of the active site cavity and below that in the translucent representation is the cofactor, with the nicotinamide ring in blue and

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module



Same basic area of the protein as above but from a slightly different angle to show the active site loop that closes over the active site during catalysis: an example of induced fit.



In this view there are three copies of the same pdb file superimposed, one shown with cartoon representation, one with bonds and one with just the subunit interface residues shown in van der waals representations.

Pymol Visualization

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

PyMOL Tutorial Interface Part 1 **https://vimeo.com/44836592** 22:27 min

PyMOL Tutorial Animation Part 2 **https://vimeo.com/44801178** 12:07 min

Basics of Pymol Part 1 http://www.youtube.com/watch?v=ai7p9Neguks 13:14 min

(downloading, color coding, saving)

Basics of Pymol Part 2 http://www.youtube.com/watch?v=uxa-9UYnIAw 13:55 min

(measuring tool, polar contacts, mutagenesis)

NOTE: In the following, some of the things that you can do to create and capture protein structure images are illustrated. *To export images from PyMOL, use File → Save Image As → PNG. These .png picture files can be directly inserted into a document. All images should annotated with a figure title and figure legend describing the image composition.*

Using PyMol, you can:

1) Create a complete cartoon image of the protein and ligand as shown in the basic commands and rendering tutorial videos (4 & 5 below). Insert the image into this typed answer word document and annotate to explain your picture.

2) Zoom in on and display the ligand binding site (as in videos 4 & 5 and using the following hint – With the ligand showing, shift and drag a box to select the protein region all around the ligand. For the selected region, then under A (action), choose preset → ligand sites, and then choose how to display). Display to highlight all the binding/interacting/interesting amino acid residues on the screen. The image can then be captured, inserted into a document, and annotated appropriately

3) Measure the distance from 4 or 5 of the binding residues in your protein to the bound ligand/small molecule using the Measurement Wizard. (Video 6) Capture, insert, and annotate the image as before.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

4) Use the Mutagenesis Wizard, (video 6) mutate an interesting or critical amino acid to: a) a conserved amino acid, and b) an amino acid with the opposite chemical characteristics. Describe the changes in structure when you perform each mutation. Capture the more dramatic instance, then insert and annotate the image as before.

5) Use a homologous protein, to create an overlay of both structures as shown in video 8. Capture, insert, and annotate the image as before.

6) Create a publication quality image as described in videos 4 & 5. Use a white background and have some fun with this image. Capture, insert, and annotate the image as before.

7) Create a movie in PyMOL with your protein – see video tutorial 7 below.

As part of the Bioinformatics and Proposal Module there are a series of video tutorials what might be of use to you.

These include:

1. Sequence alignment using EXPASY and UniProt

You Tube URL: http://www.youtube.com/watch?v=xzvUoQHP_W4

2. Downloading a PDB (Protein Databank) File

http://www.youtube.com/watch?v=EWhxjSs5IIQ

3. PyMol Wiki:

http://pymolwiki.org/index.php/Main_Page

4. PyMol Video Tutorial- Basic Commands and Rendering

Youtube URL: http://www.youtube.com/watch?v=sUWwyogc3kc

5. PyMol Video tutorial- Basic Rendering:

Youtube URL: http://www.youtube.com/watch?v=ai7p9Neguks

6. PyMol Video Tutorial- Mutagenesis & Measuring Distance

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

YouTube URL: https://www.youtube.com/watch?v=uxa-9UYnIAw

7. PyMol Video Tutorial- Making a Movie

YouTube URL: http://www.youtube.com/watch?v=I38NZPnEce8

8. PyMol Video Tutorial- Overlaying Two Structures

YouTube URL http://www.youtube.com/watch?v=NhlknnrvlS0

Background to the Roles Non-Covalent Interactions may play in enzyme structure-function relationships

Non Covalent interactions: Using Molecular Structures

While site directed mutagenesis changes the covalent structure of the protein, the impact on function is mediated by the changes in non-covalent interactions that result and a critical component of developing a testable hypothesis is understanding the various types of roles that non-covalent interactions play in enzyme structure and function. There are two conceptual types of interactions that must be considered, attractive and repulsive and these can be further categorized as coulombic and hydrophobic.

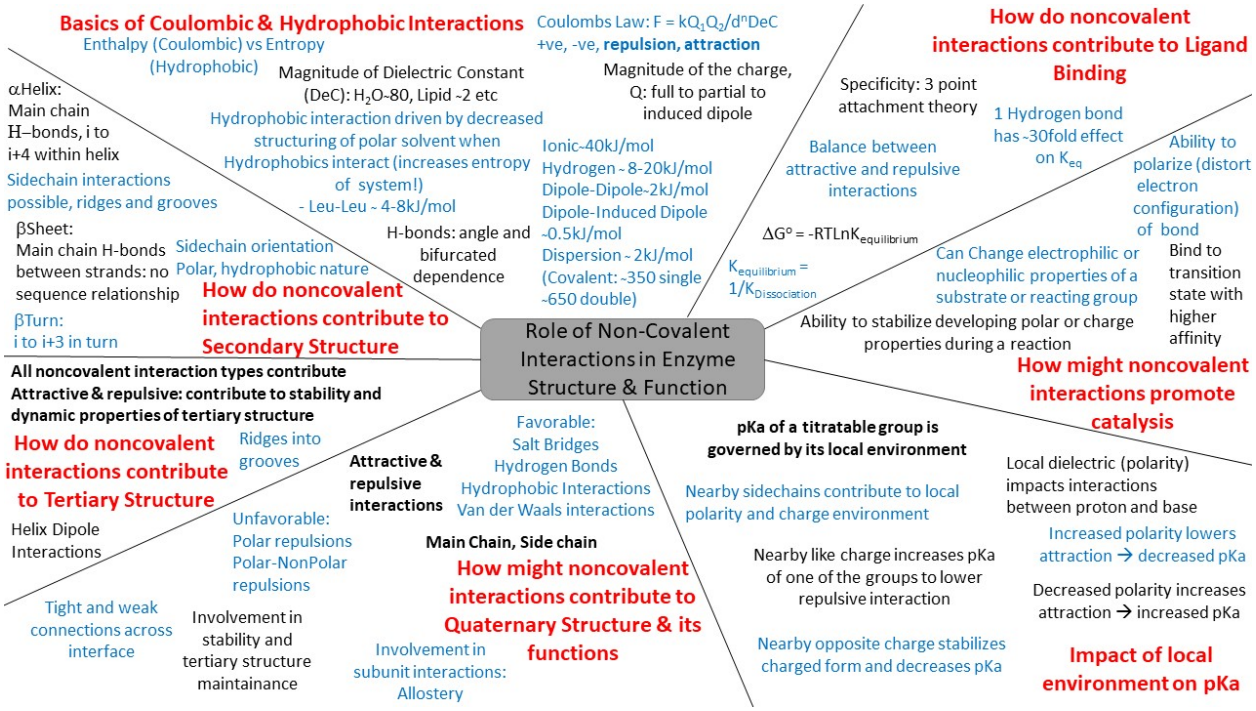
Charges in a molecule can be formal charges, for example at pH 7 a carboxyl group usually carries a formal negative charge, or partial charges- polar parts of a molecule usually have a δ negative, δ positive component. Coulombic interactions involving opposite charges are favorable, while those involving like charges are unfavorable. Hydrophobic interactions require a polar solvent to be present and result from "hydrophobic" moieties clustering to minimize the unfavorable change in solvent entropy that would result in the absence of such clustering. A structure that places a polar group (formal or partial charge) in a "hydrophobic" region is unfavorable because it

Non-covalent interactions play critical roles in a variety of aspects of enzyme structure and function including the folding and overall tertiary structure, the quaternary structure as well as secondary structure, all of which involve both favorable and unfavorable components and are related to the dynamic structure of the protein. Non-covalent interactions clearly play critical roles in binding processes, both specificity and affinity, whether it be substrates, inhibitors and regulators or other proteins, in for example a metabolon or as part of a post-translational modification. When it comes to the involvement of non-covalent interactions in the mechanism of catalysis there are a number of components to consider including stabilization of intermediates/transition states to the polarization of bonds in addition to showing favorable binding to the transition state versus the substrates

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

The mind map below illustrates both the basis of these interactions and summarizes the roles they may play in enzyme structure and function.



The mind map below illustrates both the basis of these interactions and summarizes the roles they may play in enzyme structure and function.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Rubric for Learning Goal 2

Learning Objective	*Excellent*	*Acceptabl*	*Poor*	*Unacceptable*
Compare and contrast the physical basis for coulombic interactions and Hydrophobic interactions	Coulombic interactions: charge - charge interactions, Coulomb's law, attractive, repulsive, depend upon magnitude of charge, full, partial charges, depends upon surrounding dielectric, distance. .Hydrophobic interactions- only favorable, need polar solvent, depend upon entropy of the system, solvent cages around individual hydrophobes decrease entropy, hydrophobic interactions minimize entropy loss	Misses 1-2 points of excellent	Misses 3-4 points of excellent	Minimal or no aspects of answer
Briefly outline the types of non covalent interactions you would expect to stabilize secondary structure in a protein	Hydrogen bonds- in helix between C=O and N-H in i to i+4 relationship- in beta sheets: between strands, C=O to N-H May get charge or hydrophobic stabilization between side chains in helices (helix wheel, ridges and grooves) but not necessary for formation.	Misses 1-2 points of excellent	Misses 3-4 points of excellent	Minimal or no aspects of answer
Briefly outline the types of non covalent interactions you would expect to be involved in maintaining a functional tertiary structure in a protein	Hydrophobic core (solvent exclusion), hydrogen bonds and charge - charge interactions, van der Waals interactions. Attractive and repulsive: importance of dynamic structure, intrinsically disordered regions/proteins. For folded proteins attractive > repulsive, but not by much . Usually greater in extremophile proteins	Misses 1-2 points of excellent	Misses 3-4 points of excellent	Minimal or no aspects of answer
Proteins like malate dehydrogenase have quaternary structure. What types of non covalent interactions might you expect to see across the subunit interface? What functions might these interactions play?	Wide variety of attractive and repulsive interactions- Coulombic, Hydrophobic, van der Waals: govern stability of interface: attractive > repulsive: Functions: govern symmetry, homo and hetero-oligomers, Allosteric regulation, Protein-Protein Interactions	Misses 1-2 points of excellent	Misses 3-4 points of excellent	Minimal or no aspects of answer
What types of non covalent interactions would you anticipate are involved in substrate binding to an enzyme?- compare the relative strengths of the interactions you suggest	Van der Waals (steric) – size, Charge-Charge attractions (full and partial), Hydrophobic. Full charges > partial charges and hydrogen bonds- affected by polarity of local environment. Important for specificity as well as affinity- 3 point attachment theory etc. Any given ligand may have repulsive as well as attractive interactions.	Misses 1-2 points of excellent	Misses 3-4 points of excellent	Minimal or no aspects of answer
How might non covalent interactions in a protein-substrate complex promote catalysis?	Catalysis promoted by orientation (correct orientation increases number of productive collisions) and by strain of bonds involved in reaction. Strain may be physical or electronic (strained bonds more reactive-higher energy, less stable, better nucleophile, electrophile	Misses 1-2 points of excellent	Misses 3-4 points of excellent	Minimal or no aspects of answer

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

PART 3A

Learning Goal: Create/develop and present a testable and falsifiable hypothesis and appropriate experiments to interrogate the hypothesis

Exploring Structure-Function Relationships in Malate Dehydrogenase III

Making Predictions and Testing the Hypothesis

How Do you select what amino acid to mutate?

As you will see in the course of your project, altering specific amino acid residues in the primary sequence of a protein is quite a simple process. However, to effectively use the approach of site directed mutagenesis you need three dimensional structural information. The availability of sequence homology information can also be of significant help in designing appropriate mutants. There are two appendices to this laboratory which detail how you can obtain and analyze both sequence information and structural information. If you are not familiar with these protocols you should read these appendices.

Once you have made the mutants there are still a significant number of problems that you might encounter and while these will be discussed in more detail in more advanced laboratories it is useful to be aware of them at this point.

First, just because you design and make a “mutant” there is no reason to know that the mutant will be expressed and fold correctly- the residue mutated may have played some role in the folding process that you had no reason from the final three dimensional structure to be aware of. In such a case the protein will be expressed, will not fold correctly and is likely to be rapidly degraded.

Predications from the Hypothesis and proposed experiments

Second, to probe the effects of the mutation you must first purify the protein and be able to characterize properties that will likely have been affected by the mutation. Enzyme kinetics studies can give a great deal of information about the functional properties of a protein, and as you design a mutant you should be thinking about what kinetic parameters you might measure to follow the effects of a mutant. For

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

example if you choose to alter a residue that you speculate plays a role in catalysis you would want to measure the maximum rate of the enzyme and know that it is due to catalysis and not due to a product release step. If you change a residue that may play a role in substrate binding you must be able to determine through either initial rate kinetics or direct substrate binding studies that you have an alteration in affinity for the substrate. If you change a residue that may affect the quaternary structure of the protein and affect perhaps subunit cooperativity in an allosteric enzyme you need some parameter that tells you about the cooperativity and the strength of the subunit interactions.

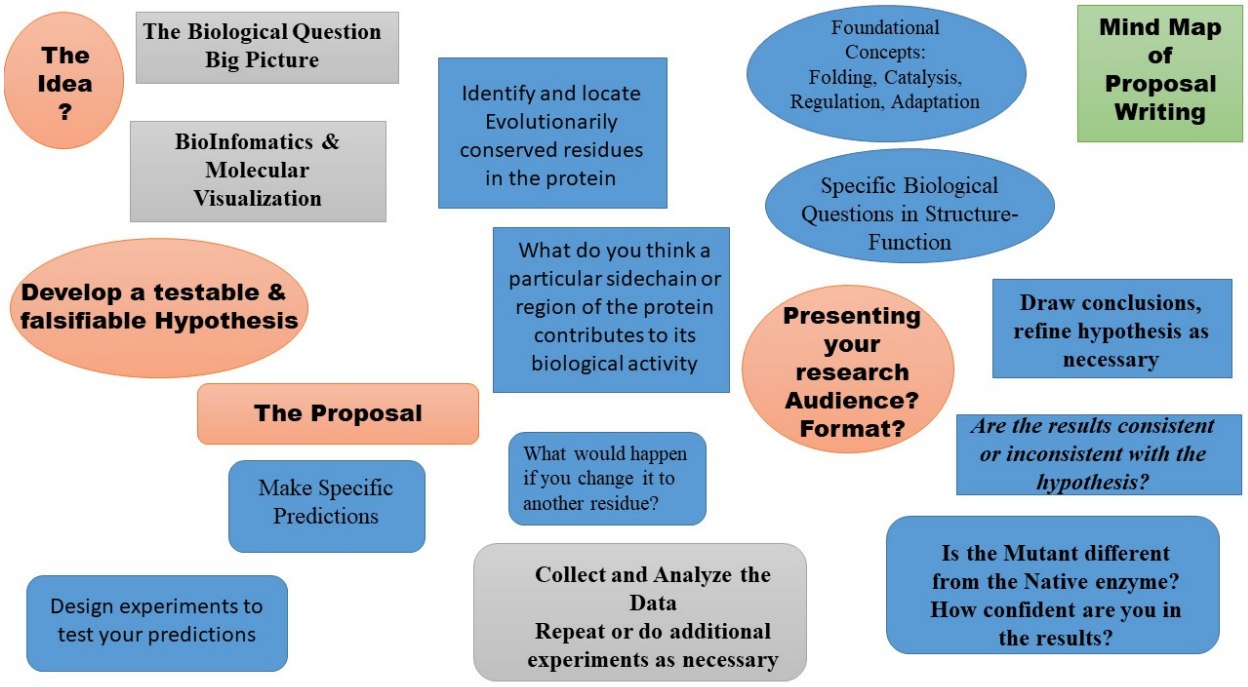
You might have hypothesized that the amino acid plays a role in folding or stability, in which case you need an experimental approach that measures those aspects of the protein. If you think the amino acid plays a role in metabolon formation you need to measure some parameter that tells you about direct interaction with the other protein in the metabolon,that is determine that the mutant no longer effectively interacts with the target protein- this could involve enzyme assays of the coupled reaction, or direct measurement of the physical interaction between the two proteins.

In all of these cases it is important to compare the native protein with the mutant using whatever measurements you decide to make, and to obtain sufficient data to allow you to determine the significance (statistically) of any differences you might find, remembering that when it comes to “testing” your hypothesis, no change can be as important as finding a change in terms of drawing a valid conclusion.

Finally, you should be aware that even if a protein folds and is functional at some level, alterations in the function could have resulted in some overall change in shape triggered by the site directed mutant and not reflect a direct role of the particular side chain in the reaction or substrate binding etc. To guard against such a possibility you need to be able to measure some parameters of the overall shape and conformation of the protein and show that there have been no gross changes in structure. The ultimate proof that the conformation of the protein has not changed is, of course, to crystallize the protein and determine the three dimensional structure of the mutant.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module



PART 3B

Presenting your work.

Writing a Proposal

Research Proposal:

The written research proposal should consist of the following sections [with guidelines for length in parentheses] :

Title: The title should be informative and generate interest in the enzyme/project- you may revise this many times during the project!

Background Information: describe both the biological and chemical background to the proposed work, reviewing relevant preliminary work and literature. [2-3 pages]: you may want to include a picture and/or

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

scheme of MDH that you generate. The following are a list of questions you may want to consider. This section should be appropriately referenced.

1. What is the reaction catalyzed by Malate Dehydrogenase, what is the equilibrium constant for the reaction

2. What are the important biological function(s) of MDH?

3. Briefly describe the reaction mechanism by which MDH catalyzes its reaction. Which amino acids are responsible for substrate specificity and which are involved in the reaction?

4. Is MDH a monomer or an oligomer? Briefly describe.

5. Find some information on the thermal stability of MDH? What do you know?

6. What are the kinetic parameters (Km and/or Vmax) of wild type watermelon glyoxysomal MDH?

7. Does MDH display protein interactions with other proteins? Is MDH allosterically regulated? If yes, what proteins/regulators are known?

8. What are the similarities and differences between MDH and other dehydrogenases?

Specific Aims/Hypothesis: As you develop your hypothesis you will incorporate it, and list a series of specific aims of the proposed work to test your hypothesis, clearly indicating defined landmarks along the way. For each specific aim a brief overview of what experiments will be used to examine the points of the aim should be given in the methodology section [1 page]

Materials and Methods: For each aim and experimental approach to be used there should be a sufficient description of how the experiments are planned and what types of data analysis and interpretation of results will be used. [1-2 pages per experiment/aim]

The materials and methods section should:

Have a logical flow of preparatory experiments with reference to published or commercial protocols

Should describe experiments to determine specific activity and basic molecular properties detailed, with reference to published or commercial protocols

Should describe how is reliability to be assessed

Details of experiments, with relationship of experiment and data to hypothesis: Data analysis including reproducibility outlined, potential outcomes and conclusions discussed

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Literature Cited. You are expected to conduct an extensive literature review for your proposal and to keep up to date on any publications that may appear after the proposal is written but before the presentations are made at the end of the semester.

Summary: Before turning in your proposal you should write a succinct summary of the proposal, highlighting appropriate background, your hypothesis and aims and how each experimental technique you propose can give information relevant to that aim etc. The summary is limited to 300 words and will immediately follow the title of your proposal when turned in for peer review.

Converting a Proposal into a final report

Written Proposal

Presents experimental data in tables/figures, does not interpret data. Logically introduces each group of tables/figures in a separate paragraph where overall trends and data points of interest are noted. Statistics are noted (n, std. dev., etc.). Each table/figure/graphic referenced. Succinct.

Figures, tables, or other graphics *clearly* present results or data with figure legends that add appropriate context

General Guidelines for figures and tables

A nice set of figure guidelines is included on the "Resources for Authors" page of the Journal of Biological Chemistry. http://www.jbc.org/site/misc/itoa.xhtml

Although focused on images for digital submission, many of the tips are generally applicable to any figure preparation. http://art.cadmus.com/da/jbc/guidelines.html. In particular the sections on font usage and multi-panel figures are particularly helpful.

Note: JBC recommends against assembling figures in PowerPoint because of potential loss of color depth and resolution, **however**, this can be a very good way to assemble your figure for our purposes as it can be arranged as desired and then saved as a single file. Once prepared, you could save the image as a png or jpg file for importing into any document.

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Figure Legends: Please note that figure legends can include simple conclusions of the data and function as a mini-method section. Carefully read these examples and other examples from the journal articles you have collected. You are expected to write at the same professional and formal level (including the style and detail) as these legends.

Structure figures:

- Figures must be clearly described and appropriately labeled so that the reader can understand the data you are presenting.

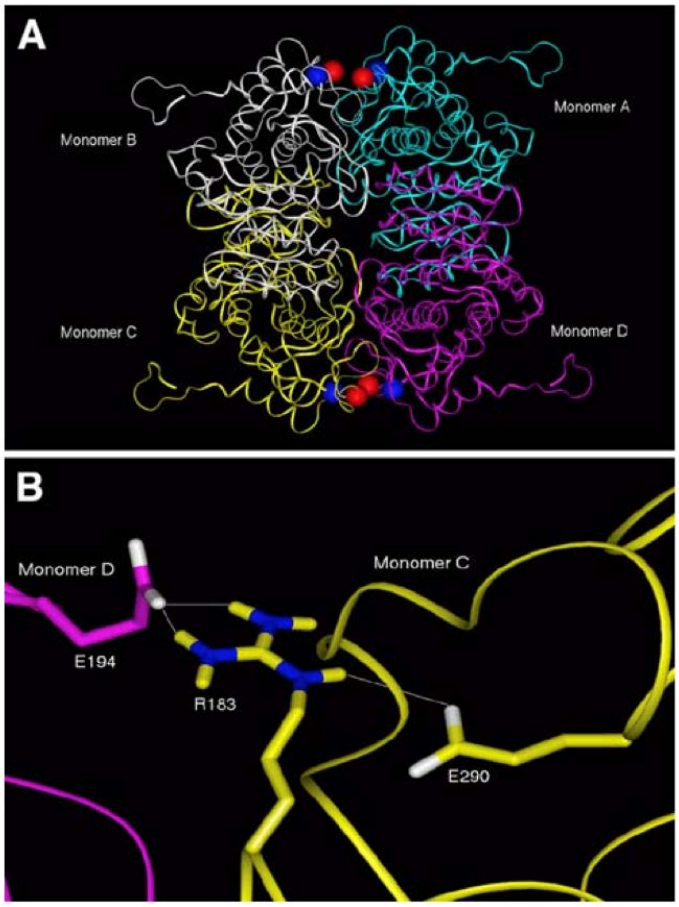


Fig. 4 a) Modeled structure of PfMDH tetramer showing the four monomers A (*cyan*), B (*white*), C (*yellow*), and D (*magenta*). Critical salt-bridge interactions are present at the interfaces of the monomers A/B and C/D. The interactions consist of a pair of salt bridges between arginine 183 (*blue CPK*) and glutamate 194 (*red CPK*) of each monomer. **b)** Intra- and inter-subunit salt-bridge interactions occurring near the interface of the monomers C and D. R183 (monomer C, *yellow*) and E194 (monomer D, *magenta*) are involved in an inter-subunit salt bridge, while R183 (monomer C, *yellow*) undergoes an additional intra-subunit salt-bridge interaction with E290 (monomer C, *yellow*).

In this figure, the part I do not like is the black background! A white background looks clearer and is easier to print if necessary.

Be sure to define any new abbreviations as needed. PfMDH had been previously defined at the beginning of the paper.

Reference: Pradhan, A. et al. (2009) *Mol. Cell. Biochem*., **325,** 141-148.

Graphs:

MCC: Exploring Structure-Function Relationships in Malate Dehydrogenase

Hypothesis Development & Proposal Module

Graphs are to be prepared on the computer. (It doesn't matter which program you use.)

Any graph should follow the following guidelines:

- Each axis must be properly labeled. Do not include a legend within the graph.
- Make the graph fit the data. Limit the axis to the values in the graph.
- When you have multiple data points ($n \geq 1$) for a single reading, do NOT graph all points on one graph. Instead, average them. If the number of replicates is three or greater, then determine the standard deviation and include that information in your graph.
- Remove guidelines from the background of the graph, it should look uncluttered so the data points (and trendline, if necessary) are clearly visible.
- A figure number and legend will be found underneath each graph. The title for each graph is in the figure legend, *not at the top of the graph*. Each legend should be a mini-statement on what you did to get the graph. It is kind of a specific methods section for each graph.
- Do not include the table of data used to generate the graph unless requested (it is in your notebook!).
- The standard default Excel format is NOT acceptable. Lines through the graph will not be accepted. Adjust the size and font of the axis label and numbers accordingly. The default size and font are typically not appropriate.
- Line graphs should be connected ONLY IF there is some sort of numerical / formulaic relationship between points (time, concentration, etc…). Histographs (bar graphs) are for values that are not related (cell lines, types of treatments, etc…).

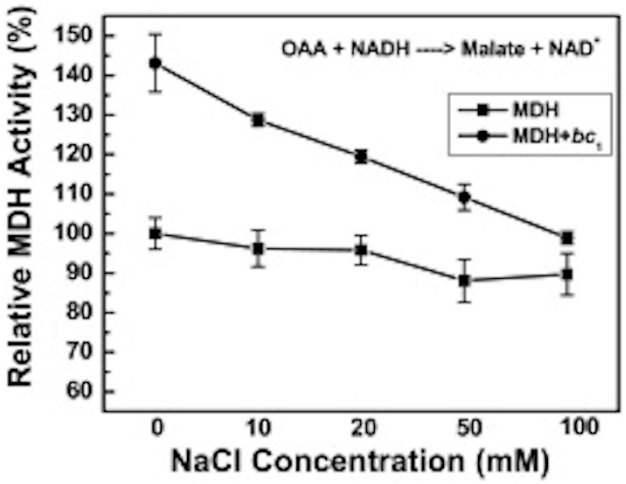


Figure 6. Effect of ionic strength on reverse MDH activity enhancement by bc_1. MDH only or together with bc_1 complex in 100 mM Na^+/K^+ phosphate buffer, pH 7.5, was incubated with the indicated concentrations of NaCl on ice for 30 min before MDH activity (the reverse reaction) was determined. The activity of MDH only was used as 100%. n = 4, data are the means ± S.D.

Tables:

If there are many graphs necessary for a set of data, such compiled data can be presented in a detailed table. Often one representative graph will be included in a figure and sometimes the remaining associated figure(s) will be published as supplementary data.

TABLE 3

Kinetics analysis of the MDH in the present of bc_1 complex

0.1, 0.05, 0.025, 0.016, and 0.0125 mM NADH and 0.2 mM OAA were used in the assay for analyzing the K_m of NADH. 0.1, 0.025, 0.0125, and 0.00625 mM OAA and 0.2 mM NADH were used in the assay for analyzing the K_m of OAA. MDH-only or together with bc_1 complex was used in both of the assays, and V_{max} was also calculated via this Lineweaver-Burk plot.

Enzyme	NADH		OAA	
	K_m	V_{max}	K_m	V_{max}
	μM	mmol/min/mg	μM	mmol/min/mg
MDH	22.2 ± 0.5	1.6 ± 0.2	10.7 ± 0.7	2.2 ± 0.1
MDH + bc_1	15.4 ± 0.5	2.5 ± 0.2	8.9 ± 0.9	3.5 ± 0.1

Reference: Wang, Q., Yu, L., and Yu, C.A. (2010) *J. Biol. Chem.* **285,** 10408-10414

Peer Review and Feedback

Your proposal may be subject to peer review.

Peer review is a critical part of science, particularly of written proposals. Reviewers are usually supplied with guidelines for their review and a rubric to "score" the proposal. When writing your proposal be sure to know what the review guidelines are and the level of detail outlined in the scoring rubric provided to the reviewers.

For example if the reviewers are provided with the following guidelines

Peer Review Form

Proposal Title_____

Reviewer #:

Please state on scale of 1 – 5 your opinion of the Overall Proposal: 1 2 3 4 5

Poor

Excellent

Evaluate the specifics of the proposal using the grading scale and criteria below:

	1	**2**	**3**	**4**	**5**
	No, poorly done				**Yes, very well done**
Q1:Does author state a testable hypothesis with appropriate Aims to test the hypothesis?	1	2	3	4	5
Q2:Does author clearly formulate the research question?	**1**	**2**	**3**	**4**	**5**
Q3:Are experiments clearly explained and justified?	1	2	3	4	5
Q4: Does the proposal use graphics appropriately?	1	2	3	4	5

Q5:Does author clarify or define terms for the reader?	**1**	**2**	**3**	**4**	**5**
Q6:Is the Introduction clear and easy to read?	**1**	**2**	**3**	**4**	**5**
Q7: Is the writing so unclear that you cannot tell what the author is trying to say?	**1**	**2**	**3**	**4**	**5**
Q8:Are there spelling and/or grammar errors?	**1**	**2**	**3**	**4**	**5**
Q9:Are the author's statements supported by sources?	**1**	**2**	**3**	**4**	**5**
Q10:Does author cite a source when presenting previous work or contributions?	**1**	**2**	**3**	**4**	**5**

In addition the review will be asked to make **General Comments which should, at a minimum, address –Q1-4 from guidelines above** , and also **Technical Comments (This is mechanics of writing: layout, flow of writing, grammar, spelling errors)**:

It is a good idea to go through your own proposal with these review criteria in mind and make sure that you meet the criteria for the grade you desire!

Oral Presentations

Several times during the CURE you will have the opportunity to make a ten minute oral presentation- this will occur as part of your development of the proposal, and again at the end of the CURE when you present your project

A good oral presentation will:

Be well organized and clearly presented, its not a bad idea to have an overview slide that outlines your talk

Have slides that are not overly crowded or contain extraneous material- if its on the slide it should be part of the story

Have an appropriate number of slides that can be fully presented in the time allotted- a rule of thumb for a ten minute presentation is to have no more that 8-10 slides

Have slides that are easy for the audience to read, figures and tables are clear and prepared as per the earlier criteria

Uses graphics and schemes

Does not try to say everything- the slide is a guide to what you will say- you are telling the story

Not be read!- as above the slide is a guide to what you will say- embed clues in the slides to remind you of the flow of what you want to say, and practice!!

Keep to the time limit, and use most of the allotted time- if you have ten slides it looks bad if you only get to number 4 during the time allotted! Or have to rush through the last few slides to make the time limit.

Rubric: Learning Goal iii: Create/develop and present a testable and falsifiable hypothesis and appropriate experiments to interrogate the hypothesis.

Learning Objective	*Excellent*	*Acceptable*	*Poor*	*Unacceptable*
Describe how the proposed work fits into the field/fills a gap in knowledge	Relationship of big picture of project to the field clearly indicated and explained	Relationship of big picture of project to the field indicated but explanation unclear	Relationship of big picture of project to the field clearly mentioned but not explained	Relationship of big picture of project to the field not addressed
Clearly states their Hypothesis and the requisite background information that lead to the hypothesis	Hypothesis or goal clearly stated. Gives appropriate background justification for hypothesis	Hypothesis or goal clearly stated but lacking justification	Hypothesis or goal is not clearly stated. Studies may or may not support hypothesis or goal as presented.	Hypothesis or goals lacking
Clearly indicates the testable and falsifiable predictions the hypothesis makes	Clearly states and justifies testable and falsifiable predictions and relates to hypothesis. Indicates controls	Limited predictions and justifications, some indication of controls.	Limited predictions, no justification or controls indicated.	Predictions lacking
Briefly outlines the types of experiments that will be used to interrogate the hypothesis	Summary of experiments is consistent with testing hypothesis or reaching goal. Types of data that will support or falsify hypothesis indicated	Outlines experiments but not how the data will contribute to the interrogation of the hypothesis	Gives detail but some proposed studies are not consistent with hypothesis or goal.	Minimal attention to how experimental data will be obtained or used to interrogate the hypothesis
As appropriate cites necessary References	Additional References added appropriately.	Additional references placed in text but some references missing.	Significant omission of additional references.	Additional references lacking.
General flow/organization	Logical flow from global to particular study point of view. Engaging writing style. Clearly connects ideas. Good use of graphics	Solid order & structure. Inviting writing style. Effectively moves the reader through the text. Graphics present but not well explained	Organization is functional; some order lacks logical pattern and structure. Minimal use of graphics	Lacks cohesive structure, difficult to follow.
Grammar/spelling/general attention to detail	No spelling or grammatical errors; includes all required sections; clearly written in language for reader familiar with biochemistry; well organized and legible.	Minor spelling or grammatical errors; includes all required sections. .	Some spelling and grammatical errors; some sections not complete or less well organized.	Significant spelling and grammatical errors; disorganized, difficult to follow.