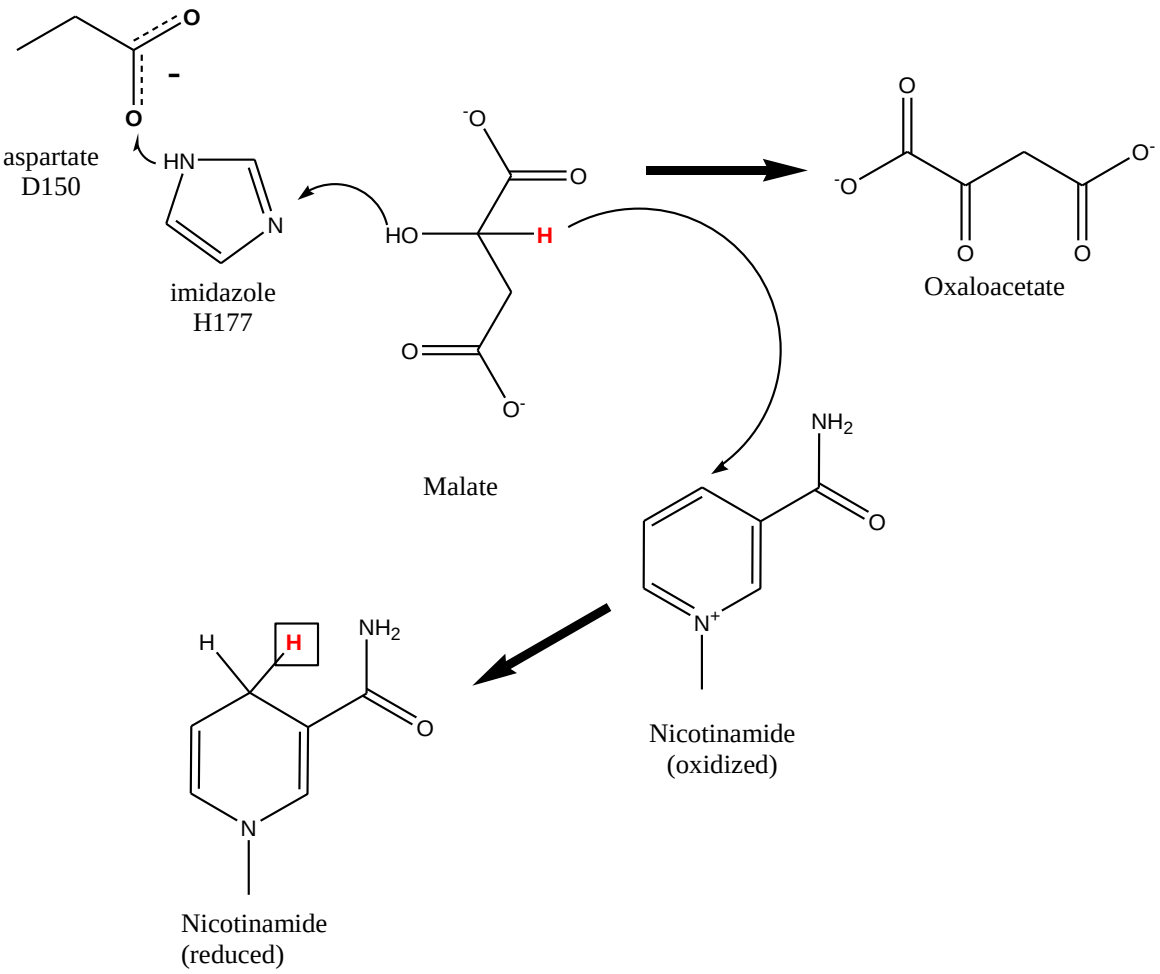
Introduction

During this course you will establish a purification protocol for watermelon glyoxasomal malate dehydrogenase. You also learned how to perform kinetic assays of the enzyme and in the next several weeks will learn the basics of plasmid purification and the design of site directed mutants using the QuikChange protocol. This semester you will utilize some of these basic techniques in addition to more advanced techniques to put together a research project involving an investigation of structure-function relationships of glyoxasomal malate dehydrogenase. The project will culminate in both a research presentation and a research paper that form the final exam in the course.

The enzyme catalyzes the reaction:



The Putative Reaction Mechanism for Malate Dehydrogenase

and in your experiments you have been following the reaction by absorbance measurements at 340nm as a result of the absorbance properties of the reduced nicotinamide ring. Based upon the available crystal structure of watermelon glyoxysomal malate dehydrogenase, H220 plays the role of a base, helping to abstract a proton from the OH of Malate. D193 is thought to play a role in this proton abstraction by increasing the basicity of

H220 [these residue numbers are for the Watermelon crystal structure and will differ in other forms of malate dehydrogenase]. Three arginine residues, R196, R130 and R124 appear to play a critical role in binding the substrate oxaloacetate in the active site of the enzyme.

In the laboratories in this sequence you will investigate the kinetic and structural properties of the native enzyme and one or more mutant form [where a single amino acid side chain has been replaced by site directed mutagenesis] of the enzyme depending upon time. The current sequence of laboratories this semester will take you through the mutagenesis process from design, this week, to isolating the plasmid DNA and constructing site directed mutants to finally placing the mutated plasmid into an expression system and expressing and purifying the protein. These steps play a critical role in the ability to probe enzyme [or protein] structure function relationships.

WRITING A PROPOSAL

The first assignment due in this course is a project proposal: you will spend the rest of the semester during laboratory conducting experiments directed towards answering the questions that you propose and will present the results of your project in two forms: a written thesis and a poster presentation: you will defend your work at the poster presentation to a group of faculty from other institutions including the National Institutes of Health and Virginia Commonwealth University. This presentation and defense constitutes the final exam of the course.

Research Proposal:

The written research proposal must be submitted electronically by 5.30pm Friday February 10th. The proposal should consist of the following sections [with guidelines for length in parentheses] :

Title:

Background Information: describe both the biological and chemical background to the proposed work, reviewing relevant preliminary work and literature. [1-2 pages]

Specific Aims: list a series of specific aims of the proposed work, 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 [1 page]

Methodology: 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]

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.

IDENTIFYING A PROJECT

During an upcoming laboratory period there will be a short presentation on the state of current research on malate dehydrogenase including discussion of current ongoing projects in the laboratory. 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.

Big Picture Questions:

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.

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.

BACKGROUND LITERATURE

On Blackboard are several articles, from the Scientific literature, describing 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 paper 1 on blackboard.

The second way is to use "Pub Med" (http://www.ncbi.nlm.nih.gov/sites/entrez), a data base for all peer reviewed scientific literature in the Life Sciences. For example if you type in the term "Malate Dehydrogenase" you would find all papers that contain the words " malate dehydrogenase "- thisis not necessarily the best way to find a suitable paper but using more descriptive search terms. For example "kinetic mechanism of malate

dehydrogenase" will produce fewer and usually more relevant "hits"- You still need to scroll through the titles to look for ones that will be useful and relevant to you, including some reviews. The articles will usually have free access to the abstract and in some cases you will be able to download the full article.

In class we will discuss how to read a paper and will use the following rubric tom help you get the most out of the papers that you read.

Rubric for Dissecting a Paper

Step	Topic for Student to Identify
1	What is the context of the paper
2	What work by others is critical to the current paper
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
6	Identify preparative experiments
7	What are the critical experiments that test the hypothesis
8	Which is the most important figure in the paper
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

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

As you will see in the next few laboratories, 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.

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 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.

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.

This week's Laboratory

There are two parts to this first week in this sequence: performing a detailed clustalW analysis of Malate Dehydrogenase and examining the three dimensional structure of Glyoxasomal Malate Dehydrognease. Once you have done this you can start designing the mutant that you wish to make.

Design of Site Directed Mutants of Malate Dehydrogenase

ClustalW 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 E Coli MDH and several mammalian forms as well as watermelon glyoxysomal malate dehydrogenase, available in the Protein Data Base. You also have access to many MDH sequences. 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.

To do this you should go to Pub Med and download, in FASTA format the sequences of E Coli MDH and Watermelon glyoxasomal MDH as well as porcine cytoplasmic and porcine mitochondrial [for all of these there are 3 Dimensional structures available]. To maximize the information that you can obtain from a clustalW analysis you should also download and use a wide variety of Malate Dehydrogenase sequences from diverse prokaryotic and eukaryotic sources.

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

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 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.

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.

Appendix Lab
Sequence and Bioinformatics

The nucleotide or amino acid sequence of a protein such as malate dehydrogenase contains a wealth of information and potential information. Bioinformatics is the process of abstracting this information and can be as simple as identifying an open reading frame from a cDNA sequence to comparing the sequence with a data base to abstract information concerning the potential structure and function of the protein. In addition, the alignment of sequences can give significant information both about potentially important amino acid residues in the sequence and the evolutionary relationships that might exist between sequences.

Shown below is the nucleotide sequence of the cDNA clone of the precursor of watermelon glyoxasomal malate dehydrogenase:

```
   1 caacgctaag ttcccaaagg tttctgatct tgaagcggtt ggtttgtttt tctgtttgtc
  61 aaactaatta tgcagccgat tccggatgtt aaccagcgca ttgctcgaat ctctgcgcat
 121 cttcatcctc ccaagtctca gatggaggag agttcagctt tgaggagggc gaattgccgg
 181 gctaaaggcg gagctcccgg gttcaaagtc gcaatacttg gcgctgccgg tggcattggc
 241 cagccccttg cgatgttgat gaagatgaat cctctggttt ctgttctaca tctatatgat
 301 gtagtcaatg cccctggtgt caccgctgat attagccaca tggacacggg tgctgtggtg
 361 cgtggattct tggggcagca gcagctggag gctgcgctta ctggcatgga tcttattata
 421 gtccctgcag gtgttcctcg aaaaccagga atgacgaggg atgatctgtt caaaataaac
 481 gcaggaattg tcaagactct gtgtgaaggg attgcaaagt gttgtccaag agccattgtc
 541 aacctgatca gtaatcctgt gaactccacc gtgcccatcg cagctgaagt tttcaagaag
 601 gctggaactt atgatccaaa gcgacttctg ggagttacaa tgctcgacgt agtcagagcc
 661 aataccttg tggcagaagt attgggtctt gatcctcggg atgttgatgt tccagttgtt
 721 ggcggtcatg ctggtgtaac cattttgccc cttctatctc aggtgaagcc tccaagttct
 781 ttcacacaag aagagattag ttacctgact gataggattc aaaatggtgg aacagaagtt
 841 gtcgaggcca aagcaggagc tggctcagca actctctcaa tggcttatgc tgccgttaag
 901 tttgcagatg catgcctcag gggcttaaga ggagatgctg gtgtcattga atgcgcgttt
 961 gtgtcttctc aggtgactga acttccattc tttgcatcaa aagtacgact tggtcgcaat
1021 ggtatcgaag aagtatactc ccttggcccg ctaaatgagt atgagaggat tggattggag
1081 aaagcgaaga aagagttggc aggaagcatt gagaagggag tttccttcat cagaagctga
1141 agagatgcca attaccatta gttttaatag aaacattcca tctcttatag attacttgtg
1201 ctcaatgttt tcctggagat tgaagttgat tgaaatgata ccacaccacg tatttttata
1261 ctaataaaac tatatcgcca tcatgtcgat atttaatgca caaccaaaag ggttggatta
1321 gagtaccttt tatg
```

Together with the protein sequence:

```
   1 mqpipdvnqr iarisahlhp pksqmeessa lrrancrakg gapgfkvail gaaggigqpl
  61 amlmkmnplv svlhlydvvn apgvtadish mdtgavvrgf lgqqqleaal tgmdliivpa
 121 gvprkpgmtr ddlfkinagi vktlcegiak ccpraivnli snpvnstvpi aaevfkkagt
 181 ydpkrllgvt mldvvrantf vaevlgldpr dvdvpvvggh agvtilplls qvkppssftq
 241 eeisyltdri qnggtevvea kagagsatls mayaavkfad aclrglrgda gviecafvss
 301 qvtelpffas kvrlgrngie evyslgplne yeriglekak kelagsiekg vsfirs
```

How does one go about finding the protein sequence from the nucleotide sequence?

In this case life is simple: the first atg codon is the initial methionine in the precursor sequence. With many cDNA sequences however it is not that simple and to translate the nucleotide sequence often means trying a number of atg codons until the correct reading frame is obtained: often the reading frame set by the first atg codon is not correct and the sequence will not translate completely: you run into stop codons well before the correct stop codon and get a prematurely terminated sequence. Knowing where the sequence stops is another potential problem: when there can be three stop codons, uaa, uag and uga, and one of them, uga sometimes codes for selenocysteine there could be a problem of knowing where the reading frame actually stops. In the case of the watermelon glyoxasomal malate dehydrogenase you will find that the actual stop codon is a uga codon and that the next stop codon is some six residues later (uag). If the potential confusion involves a relatively long stretch of amino acids the molecular weight of the protein can be useful in deciding the actual stop codon.

Once you have the protein sequence what else can you tell about the protein from its amino acid sequence.

First you could run a search of a data base for similar sequences. This usually involves a "Blast" search which aligns the sequence against all those in a data base and relates "homology".

What is homology?

Sequence identity is exact homology. Closely related proteins will have a high degree of sequence identity. More distantly related proteins will have a lesser degree of sequence identity but may also show sequence conservation where residues of similar types [physical properties and function] have been substituted. More distantly related proteins will also often have gaps or inserts where regions of sequence have been deleted or added. The algorithyms that calculate homology usually have some type of penalty for gaps in the alignment since it is of course easier to align two or more sequences if you allow gaps of unlimited size.

Once similar sequences have been identified via blast search sequences can be aligned using a program such as Clustalw. In these types of programs it is usually possible to color code residue types and to mark exact homology and functional homology. In the alignment of 4 malate dehydrogenase sequences [watermelon glyoxasomal, E. Coli, Pig mitochondrial and Pig cytoplasmis forms], the standard four colors used to highlight amino acid residues are shown in the following table:

Color	Amino Acid Side Chains (pK)	Conserved Property	Comments
Purple	**K(10.5), R(12.5)**	**Positive change, length**	
Blue	**E(4.1), D(3.9)**	**Negative Charge**	Q and N are neutral analogs
Green	**G,S,T,Q,N, H(6.0),**	**Hydrophilic**	H can be positive, Y

	Y(10.5), C(8.4)		and C can be negative
Red	**M,L,I,V,A,F,W,P**	**Hydrophobic**	P restricted Φ angle

Note the insertion of two added stretches of sequence:QPII and DSWLKGEFITT in the cytoplasmic malate dehydrogenase and also the fact that when looking at the color coded sequence alignment there is more homology between the pig mitochondrial and the E Coli and glyoxasomal enzymes than there is with the cytoplasmic enzyme.

Second, by comparing the sequence with known "consensus sequences" it is possible to sometimes identify an active site motif or binding domain, or such things as potential sites of post translational modification. For example running "Prosite" [http://us.expasy.org/prosite/] with the glyoxasomal MDH sequence will give an indication of where "possible" sites of asn glycosylation [NSTV], Protein Kinase c phosphorylation [TxR], Casein Kinase 2 phosphorylation sites [S/TxxD/E], Tyrosine Phosphorylation [RxxxExx.Y], myristylation [GGxxGF], RGD cell attachment sequences [RGD] and of course the malate dehydrogenase active site signature sequence [VTMLDxxRAxxxV]. You should note that apart from the malate dehydrogenase active site signature sequence there is no evidence of any of the other sites being actual sites: they are simply sites that have the appropriate consensus sequence.

From the amino acid sequence of a protein [or from the amino acid composition] it is possible to obtain an estimate of the isoelectric point of the protein. For example using the site http://us.expasy.org/tools/pi_tool.html and entering the glyoxasomal MDH sequence one obtains a computed pI of 8.67 for the precursor sequence or 8.25 for the mature watermelon glyoxasomal MDH sequence. These values are calculated as described earlier on the assumption that local environment effects are minimal..

Finally, there are a variety of "prediction" tools available through ExPASy,

http://us.expasy.org/tools/#proteome

which can be used to predict all manner of things such as the potential for a transmembrane helix [you will be surprised to find that this sequence has a quite probably transmembrane sequence near the n terminal region of the protein- it of course does not have a TM helix, just the potential for one! In this region] or various secondary structure prediction programs, which often start by doing a blast search to look for an homologous sequence with known secondary structure. With an "unknown" protein sequence where there are no identified homologies such prediction is of little value.

Appendix:

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 Clustalw 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.

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:

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

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
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
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.

Molecular Visualization Programs.

Protein Data Bank (.pdb) files can be visualized in a number of programs, many of which are freely available from the web. You will use several of these in this course. If you want to download these programs to your own computer they are available through the "software" section of the PDB site. If you go to the "Mage and Kinemages" link you can download two programs, Mage and Prekin which together allow visualization and analysis of pdb files. "Rasmol" and "RasTop", the "Swiss PDB viewer" and VMD. In addition to these programs pdb files are read and used by commercially available programs such as HyperChem and MOE. Each of these programs has tutorials available online or through the commercial web site and you should learn to utilize one or more of these programs to visualize and manipulate three dimensional structures.

The discussion of Malate Dehydrogenase and the analysis of what residues are within the various binding sites or catalytic site are based on analysis using Mage and VMD and are presented to guide you in your analysis of the structure-function relationships in Malate Dehydrogenase as a guide to the design site directed mutagenesis experiments that you will perform in the next laboratory using Quiagen QuikChange Mutagenesis protocols.

Analysis of Malate Dehydrogenase Using Prekin and Mage:

Download the file 1EMD.pdb from the protein data bank or use the gMDH-dimer file from blackboard..

Open Prekin and load the 1EMD.pdb file: this creates a kin file which Mage will utilize. Select "Selection of Built in scripts" and then "aasc + R" and "do all subunits in range" This will create a kin file that shows the backbone and all of the side chains in each subunit in the structure [there is only one subunit in the 1EMD file] After Prekin has created the kin file, close Prekin and open Mage. Open the 1EMD.kin file that you previously created and using the mouse to click on amino acids find the Active Site Histidine residue, H177. If you want you can change the background color to white [from the drop down "display" menu]: some people find it easier to see with a white background and if you want to print the display it looks better with a white background. Once you have found the active site histidine you should use the "measure (distance)" tool from the upper drop down "tools" menu bar and map out other amino acid side chains that are within 5-8A of the active site histidine.

When you have done this you should repeat the process [upto opening the file in Mage] for 1IB6.pdb to create the appropriate kin file. Now go to Mage and open this kin file. This time you will see 4 subunits [there are two dimers in the structure] but you can select (on the upper right hand box) which subunits to show: select either a and b or c and d subunits. This will display a dimer. If you switch off the side chains [right hand panel again] you will see just the backbone connectivity. Again the mouse selects residues and indicates their numbers. Rotate the structure and locate the dimer interface. Examine and make a table of the close contacts across the dimer interface.

In either structure you should map out the regions of the primary sequence that constitute the major helices or beta strands in the structure.

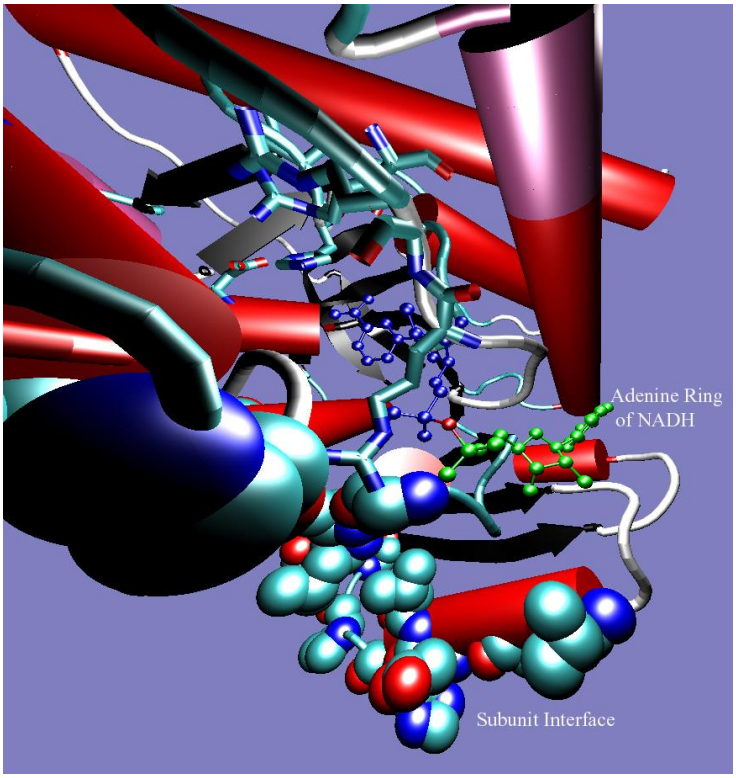
Visualizing the Structures in RasMol or RasTop

Once you have oriented yourself to the general features of the structure using Mage you should open each pdb file using RasMol or RasTop and familiarize yourself with the basic functions of the program.

VMD

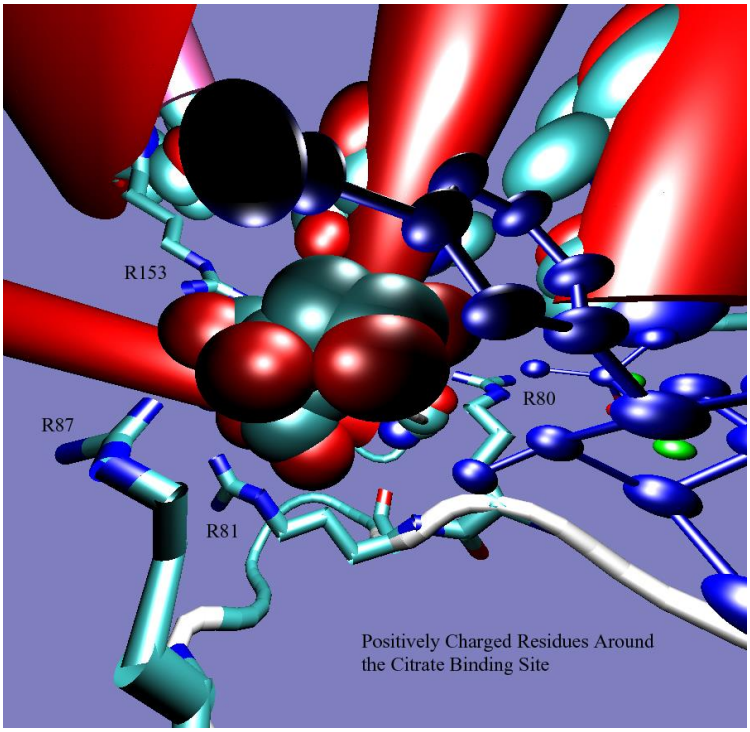
VMD is a powerful graphics program which allows you to display the contents of a pdb file. The program gives you powerful graphics control over the representation of the structure and also allows you to zoom in to a structure to focus on a relatively small area of the structure. Since the program will 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.

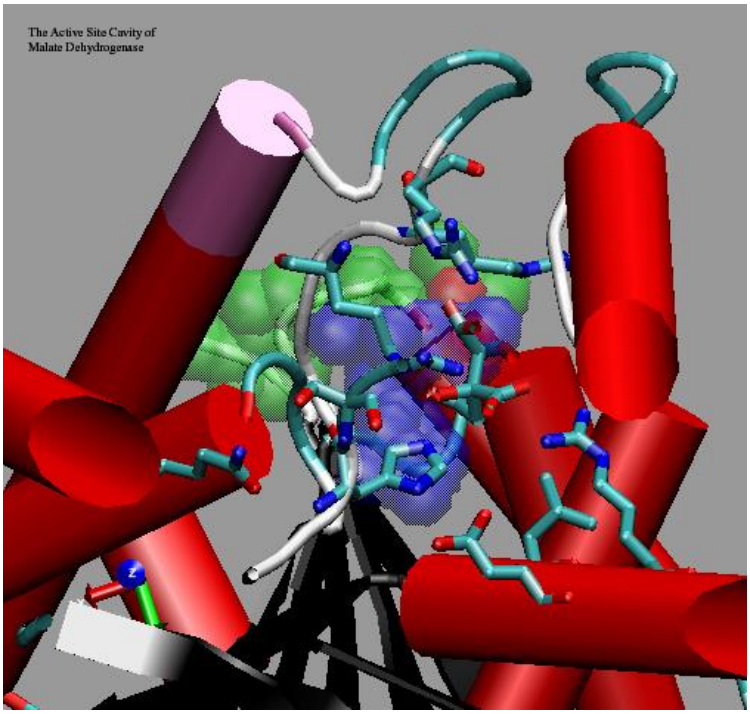


This image was created by loading the 1EMD file twice, representing one data set in cartoon mode and the other designating particular side chains to show with the "resid" command. Finally the cofactor pdb file that was created from the original 1EMD.pdb file was loaded and the cofactor atoms represented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

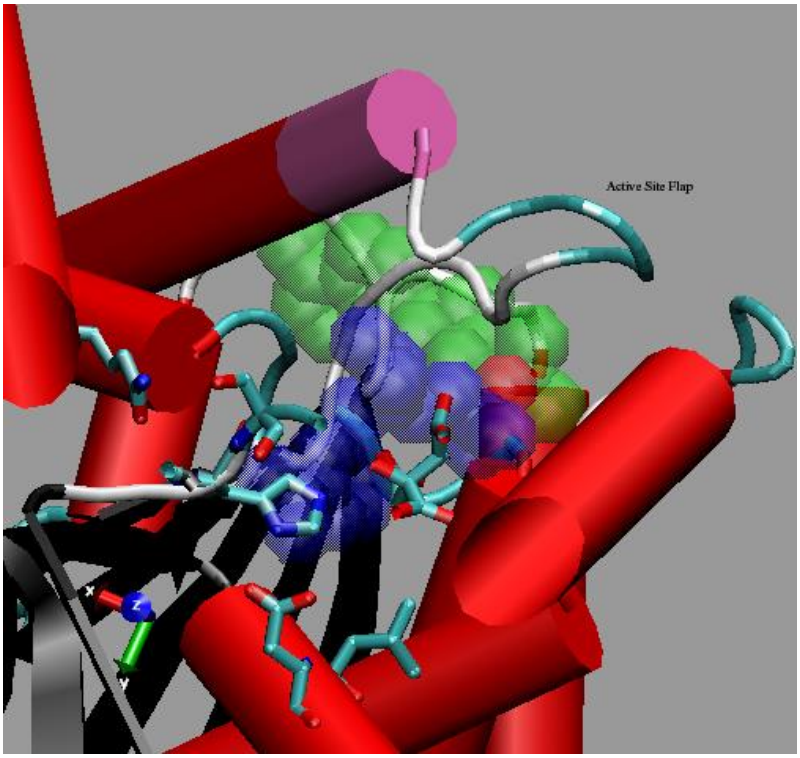


This image was created by again loading two sets of the 1EMD.pdb file, representing one in cartoon fashion and the other in "liquorice". Only certain side chains were designated using the resid command to be shown in the liquorice set. After this had been done the pdb file created for NAD was loaded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

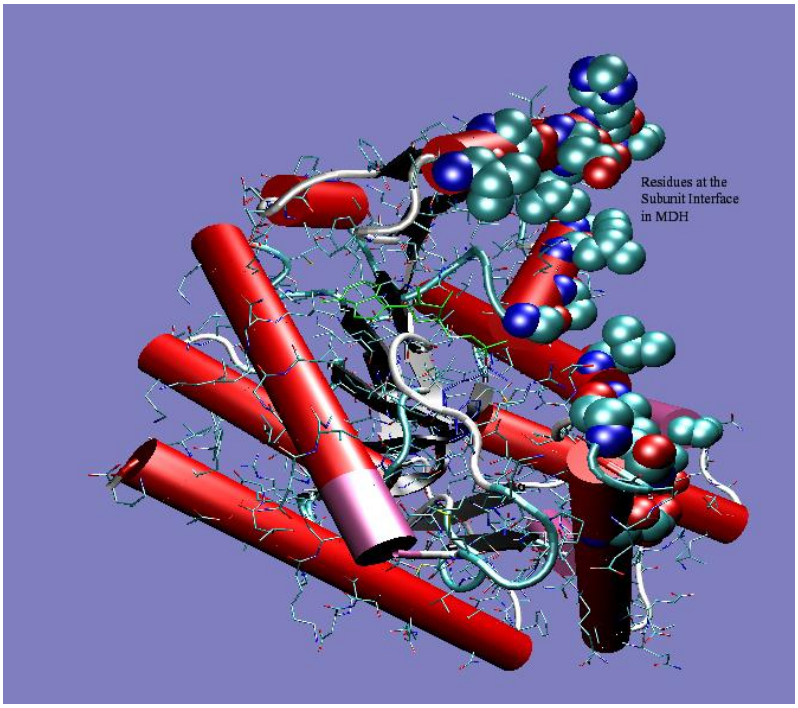


In this image various residues around the active site are shown including an aspartate, a leucine, an asparagine and the arginines shown in the previous picture. You should attempt to define which aspartate, asparagine,and leucine are shown

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 the adenine ring in green..



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.