

Lecture 1

This course is concerned with **proteins**, polymeric molecules made up, at least initially, of 20 natural L-amino acids, coded for by DNA, which message is transcribed into mRNA and translated into protein on the ribosomes; thus cyclic peptide antibiotics assembled by bacteria in various ways and including D- and other unnatural amino acids are not included. Once incorporated into proteins, the basic amino acids may be changed into other things; this we will discuss later. Proteins may also have **prosthetic groups**, tightly attached non-amino acid groups, for special purposes; these may be either covalently or non-covalently attached - the operative definition is that they not be dialyzable away. Sometimes in enzymatic reactions these have to undergo a second reaction to return to the original chemical form; but this is really metabolism and outside our concern. Many proteins also bind metal ions tightly. And there are a lot of other natural modifications of proteins whose significance for their function may or may not be clear. I have for the past several years spent a fair amount of time on them - one could teach a whole course.

We assume that the properties of a protein can be described *as a first approximation* by the properties of the constituent amino acids, and it is essential to know these properties; but the properties will be modified considerably by their surroundings, by the three dimensional structure of the protein, and it is the continuing challenge of study of proteins and enzymes to learn how the special properties of a protein result from its unique structure. Sometimes the function of a protein involves a prosthetic group or a metal ion, or a modified amino acid, but not all that often. In some cases, as I mentioned, proteins are found to have non-protein modifications, and the problem is to understand their significance to the function of the protein. This is particularly difficult when the *chemical* function of the protein is itself unclear, as in oncogene proteins.

We will be considering some of the major ways of investigating protein properties, primarily chemical ways and those involving measurement of enzyme activity.

I don't spend time on three-dimensional structure of proteins; but I have asked Dr. Kahn to give one lecture on modeling protein structures in the computer, which is a real show. If you want more about it, I recommend the books by Creighton, *Proteins Structures and Molecular Properties*, and by Brandén and Tooze, *Introduction to Protein Structure*, which are mentioned in the syllabus.

If your interest is primarily in protein purification, I recommend Dr. Ward's course Biochemical Separations, which provides a much more basic and comprehensive look at ways of purifying proteins than I have time for, though as you see from the syllabus I do quite a bit. I should also mention a two-credit course on receptors, occasionally given at the Medical School by Drs. Langer and Rashidbaigi. I will talk some about the general topic of binding of ligands by proteins, and how it is measured, but I have no experience with receptors as such.

The course schedule might be summed up as follows: after today, nine lectures on the assay and purification of proteins, how to get them to the point where they can be studied in isolation, as they need to be before you work back toward studying them *in vivo*, or in as complex a system as you can handle. Then five lectures on enzyme kinetics. I probably won't be able to cover both allosteric enzymes and organized enzyme systems - there were previously listed in separate lectures, and I rarely was able to speak on the latter, because I was behind and sacrificed the topic to get back on schedule. Then five lectures on protein modification, natural and chemical. Natural post-translational modifications are a growth area in studying the specifics of protein action, and artificial modification is a powerful tool in the investigation of the relationships between structure and function in proteins. This includes the currently popular topic of site-

specific mutation, which is indeed powerful, but requires that a lot already be known about the protein. I am also going to try to fit in something on disulfide bond formation, since there is a recent review on this. Then I'm going to try to give a lecture on proteomics, which the recent review defines as the large-scale analysis of proteins, and I perceive as two topics – large-scale analysis of what proteins are expressed under a particular condition, and trying to figure out what the functions of these proteins are. This is also being addressed by my senior seminar. In order to do this I'm dropping Dr. Poretz's lecture on glycoproteins. Then two or three lectures on the general theory of enzyme catalysis, which is probably Fersht's favorite subject. Then three or two on mechanisms of specific enzymes, including tyrosyl-tRNA synthetase which comes almost entirely from Fersht's lab and is covered in detail in his text, and one on catalytic antibodies, another hot topic nowadays. We would like eventually to be able to design enzymes to order, to carry out whatever reaction we want, and this is currently the most promising, if still empirical, initiative in this direction. Dr. Kahn will talk about using computers to change known protein structures into the unknown structures of other proteins - what is called homology modeling.

I would like to give a lecture on the use of fluorescence in studying proteins, but simply have not been able to find room. New topics keep appearing, while old ones do not become unimportant.

In the past I have made copies available for purchase, at a nominal price of 8¢ a page, in the departmental office, Lipman 207. This is particularly important if you miss a lecture. For the past couple of years they have instead been posted on the departmental web site under my name.

I give three 'take-home' examinations, of increasing importance to your grade, as you get used to the idea. These are very different from the type of exams normally given to undergraduates, where you are tested to find out how many right answers you can identify in an hour or so. I give you a journal paper, which you are to read, and a series of questions on the paper, the answers to which you should be able to figure out if you have understood the material in the lectures up to that point. This is much more like what you do as graduate students, in reading and understanding papers from journals - 'the literature' as we call it. I shall try to give you other papers to read, to be discussed briefly in the following lecture, but this depends on how able I am to get an appropriate paper at the right moment in the course. Reading and understanding journal papers, at the level of understanding the techniques as well as the conclusions, is one of the aims of this course, even though I have not in the past worked on this as much as I should.

The final item is the course paper, an original research proposal (though without the detail of an NIH grant proposal, which is sometimes assigned in graduate courses). You are to describe a protein and some aspect of it, or of its interaction with other molecules, which is not yet understood, and propose one or more experiments designed to understand it better. For graduate students, this should **not** be your thesis proposal, though it may be related to it and arising out of reading you have done for it - perhaps something that intrigues you, but which your professor isn't interested in having you work on; and it should not be straightforward purification, which is largely empirical, i.e. you try things until they work, though defining an affinity chromatography method is a reasonable proposal. Nor should it be straight sequencing, or isolating the gene, or anything else which is a standard method whose application to this protein doesn't require much thought. In any case, DNA sequencing and gene isolation are problems in molecular biology, not in protein biochemistry. The paper must focus on a **protein** problem, not a molecular biology problem. In 1996 I had to reject something like five papers because they simply didn't use the material of the course. Molecular biology can be used as a **tool** in approaching a protein problem, as for instance site-directed mutagenesis, or the yeast two-hybrid system for finding a protein which binds to another protein, but the question to be answered must be a **protein** question. Chemical modification, active-site-directed modification, site-specific mutagenesis, enzyme kinetics, use of physical techniques to investigate conformational changes, are some techniques

you might use. I suggest you make a tentative selection of a project early and then come to see me, both for me to pass on whether I think it is an appropriate project and to get any help I can give you about how to approach it. If you don't already know how to use the library I can make suggestions, but you can get fuller help from the reference librarians. The point of the paper is to select an appropriate problem **and** an appropriate technique to solve it, understanding the limitations of the technique as well as its powers.

I should perhaps comment here that this course, though not a laboratory course, does expect some lab experience - some years ago a junior undergraduate took it without having taken biochem lab, and she often didn't understand what was going on. When I talk about different types of column chromatography, I don't want to have to explain what column chromatography is to someone who has never done any sort of it.

I should probably say that in a course like this it is hardly possible to make you immediately able to do research in a given area, rather only to make you aware that the area exists, so that you may go on to learn more specifics in an area when you need to. The Rosenberg book does give you a number of experimental protocols to do certain things. What I **can** do is give you an idea how the various areas fit together, and some idea of strategy, how you use them to answer questions; I suppose one could say that practically everything in the course is methods, except for the few lectures at the end about the mechanisms of specific enzymes. But the problems you will be trying to answer in your work are many, and I can give only a selection of current **problems** in the area of proteins and enzymes; rather, I try to cover **ways of approaching their answers**. When you read a journal paper, always make sure you understand the techniques used, how they get their answers; you may find the techniques useful for your own research. The course culminates in a research proposal paper, in which you are to propose some question about an enzyme or other protein, and how to answer it. You may need to know more details of a technique in order to write the paper, but at least I should get you to the point where you can learn more by reading.

About the reading listed in the syllabus: the specific textbooks are alternate sources of information at the level of the lectures - probably the best thing to do would be to read the text assignment **before** the lecture, to help you understand it, but others may find reading afterward more useful. They are also useful references for the future, particularly Scopes if you are going to purify proteins. The references to *Annual Reviews of Biochemistry* and to review papers generally indicate what I have used in preparing the lecture. They would be good sources for more information, especially for references on original research papers. I will give more references as we go along in the course.

I have been teaching this course for over thirty years, and sometimes worry whether I am getting out of date, particularly in light of the mushrooming growth of 'molecular biology'. There is no general agreement on distinguishable differences in definition of biochemistry and molecular biology. A biochemist and a molecular biologist will generally each try to define the other's discipline as a subdiscipline of theirs; but it is at least interesting that, whatever the definition, they know what *they* are. I would define molecular biology perhaps as synonymous with molecular genetics, as having at its core a genetic way of thinking, based on number (as, number of recombinants) which is very different from a chemical way of thinking based on weight (amount of a substance); and as concerned with the central dogma, the flow of information from DNA to RNA to protein synthesis, the control of each step of this, and the controls within this process of cell function up to differentiation. The manipulation of nucleic acids is central to molecular biology. The term molecular biology arose as a contrast to organismic and physiological biology, as describing understanding of biological processes at the molecular level; but I would say that biochemistry has always been at that level. Biochemistry has its origin what was called physiological chemistry, identification and measurement of all the chemical

compounds of living systems, and in how they are transformed. It came to be realized that these transformations were carried out by enzymes, which are virtually always proteins. Virtually all cell functions are carried out by enzymes and other proteins. Hence there was much interest in how enzymes catalyze reactions, and more generally in how even non-enzyme proteins carry out their functions.

I want to emphasize, now and especially when you write your course papers, **quantitation**, measurement of **how much of** something you have, of a specific protein or of its biological and chemical activity. Molecular biology has its roots in genetics, and often is content simply to know whether a gene or its product is present or absent in an individual or tissue. You can understand more about a biological function if you can measure **how much** of the function is there, and especially if you can compare amount of function, such as but not limited to enzymatic activity, to the amount of the protein responsible for the function. I shall try to stress this often in the course.

We are realizing more and more things that proteins do in addition to catalyzing reactions of other molecules; the problem is that the functions become very difficult to measure outside the complete biological system in which they occur. This is a practical problem, not a philosophical one; once you have a system for measurement - what we call an assay - most of the approaches we have for studying enzymes are also applicable to the study of other proteins, with the principal exception of enzyme kinetics, the study of the rate of enzyme-catalyzed reactions as influenced by conditions such as but not limited to substrate concentrations. I try to give papers to read on assay and characterization of **non-enzyme** proteins. And one of the strong points of molecular biology is that it can enable you to express a protein, or repress the normal expression, **inside** a cell, so that you can at least qualitatively look at its role in a function only observable in the intact cell.

Scopes, in the introduction to the third edition of his book, makes the point that increasingly often proteins are expressed in and purified from a non-natural host cell, the gene having been introduced into the cell by recombinant DNA techniques. One can thus get a great deal more of the protein - in some cases as much as 50% of the host cell protein - and assay by function may be unnecessary, you just purify the most abundant protein present. Or, knowing the DNA sequence and thus that of the protein, you synthesize peptides of that sequence, have a rabbit make antibodies to the peptides, and use them for assay. This is OK for assay during purification, but of course tells you nothing about the protein's function - particularly important when you have changed the structure in some way. Purification may however be important in itself, if the protein is to be used for therapeutic purposes; the FDA then insists that you get it *very* pure, purer than you would need it for experimental purposes.

Several people have remarked to me that proteins are much harder to work with than nucleic acids, primarily because each protein is a separate problem in purification - one must determine in each individual case how to purify it while retaining its biological activity, as well as how to assay that activity. There are many ways to lose that activity, and in general milder conditions must be used - for instance, nucleic acids are typically precipitated at 70% ethanol, and often heated to at least 65°; neither of these procedures are generally applicable to proteins if you want to maintain their biological activity. Beyond that, one should remember that with rare, if Nobel Prize-worthy, exceptions nucleic acids are *only* informational molecules, while proteins **do** virtually everything in cells, including catalyzing the vast number of chemical reactions which continually occur in cells and whose sum is the life process. From the point of view of the enzymologist and proteinologist, the techniques of modern molecular biology are just added ways to study proteins, how do they do what they do. You can learn about the molecular biological methods in other courses; in this course we shall cover the kinds of **results** available from one of them, site-specific mutagenesis.

One matter I always promise at this time but rarely get to: enzymes are classified into six groups, with specific numbers which allow you to recognize an enzyme even if the paper is in Serbo-Croatian, and which are assigned by the Enzyme Commission:

1. Oxidoreductases, i.e. enzymes which catalyze the oxidation of one substrate and the reduction of another. The prototype, EC 1.1.1.1, is alcohol dehydrogenase, which in the physiological direction usually reduces acetaldehyde and oxidizes NADH; note that the name may not describe the prevailing physiological direction of the reaction - pyruvate kinase, which is irreversible in the other direction, is an even better example.

2. Transferases - those transferring a group from one molecule to another, as for example phosphotransacetylase, $\text{CH}_3\text{COSCoA} + \text{Pi} \rightarrow \text{CH}_3\text{COOPO}_3^- + \text{CoASH}$.

3. Hydrolases - those hydrolyzing a bond, such as proteases, esterases, glycosidases. Of course this is transfer to water, and since some may use alternate substrates, water or an alcohol, the distinction from transferases can be blurry. Similarly there are enzymes which are both oxidoreductases and transferases, such as aldehyde dehydrogenases which make acetylCoA rather than acetate.

4. Lyases - which break bonds, usually from carbon to carbon, nitrogen, oxygen or sulfur, **without** adding water, often leaving double bonds in the product; for instance phenylalanine ammonia-lyase, whose product is cinnamate (3-phenylprop-2-enoate).

5. Isomerases, which leave the same atoms in the product, but in different arrangement. Racemases, sugar epimerases. The reactions may seem simple, but often are complicated in mechanism.

6. Ligases - all the enzymes which **build up** more complex molecules, usually with hydrolysis of ATP or other nucleotide triphosphate - these are termed **synthetases**, while those which use energy from another type, of high-energy bond such as a thioester, are called **synthases** citrate synthase, for instance.