Engineering Biotechnology

Gateway Coalition

Drexel University

Faculty Staff: Raj Mutharasan, W.E. Magee, M.A. Wheatley, Y.H. Lee

Multimedia Team: Bhargavi Garapati, Dakai Cheng

Webbooks Programming: Kannan Mutharasan, Adam Rhuberg

Engineering Biotechnology - Foreword

With the advent of significant fundamental advances in biosciences, increasing number of products, particularly pharmaceuticals, are manufactured using biological agents. New applications of genetic engineering in other industrial segments are reported at a seemingly increasing rate. Consequently, a larger fraction of engineers of tomorrow would need to be familiar with fundamental precepts of applications of biosciences, genetic engineering. The central idea the authors propose is to treat in a single course all significant scientific and engineering issues that encompass converting genes, the starting material, to final product that is manufactured for the market place. In our view, it is important to tell the whole story in a single course in sufficient depth so that the relevancy and significance of the emerging area of biotechnology can be communicated.

In order to provide a common application theme (or thread) through all the course topics, the group chose insulin. Insulin was the first product of genetic engineering in the market place and a fair amount of literature on its manufacture is in the open literature. The course is organized into four segments: insulin genetic engineering, manufacture of insulin in bioreactors, delivery of insulin in the human body and biosensors and monitoring devices for biopharmaceuticals. The product insulin served us well in the role of weaving the seemingly different subject matter into one wholesome story. It was not our intention to develop a course on insulin, which would be clear when, sees the details of the course content. The course was team taught by the four instructors whose teaching and research training correspond to the four topic areas to be taught. In the <u>first</u> edition of the course (Winter 93/94), all four instructors were present in the lecture room. The three audience-instructors carefully interacted with lecture-instructor so that appropriate expertise for the question at hand was brought to the fore. Additionally, they assisted with comments and suggestions to improve learning in the classroom. This approach is expected to integrate the expertise of instructor(s) and enhance student learning. In the context of Drexel, this course is expected to be one of the cross-disciplinary electives offered to all engineering pre-juniors and juniors in the Drexel Curriculum. Although the current edition of this integrated "text" appears to be segregated, the intention is to integrate the topics in classroom discussion.

The project of developing the course and the self-standing multimedia was directed by Raj Mutharasan in collaboration with faculty colleagues, Wayne Magee (Bioscience & Biotechnology, Drexel), Margaret A. Wheatley (Chemical Engineering, Drexel), Young Lee (Chemical Engineering, Drexel) and non- Drexel colleagues Ken Foster, David Graves (both of University of Pennsylvania) and Dan Raichel (Cooper Union). The project was funded by NSF supported Gateway Coalition (EEC 9109794).

The authors have taught the integrated course each of the past three years. Evaluation measures through journals, end-of-the course interviews and course evaluations show that the students found the topic interesting and challenging. A very large fraction of the students said that it broadened their technical horizon. In the same breadth, they also said that the biology part was difficult. We discovered through further interviews that the "difficulty" stemmed from their method of learning were through worked-out-examples

rather than reading the subject matter. This pointed to the well-known axiom that "learning by doing" is a very common method used by a large fraction of the engineering students. Such an opportunity is not easily found in the traditional treatment of biology subject matter.

I have heard it said that a piece of work such as this is never completed, but only abandoned. Should you wish to keep up with our evolution through revisions, do drop by at http://www.chemeng.drexel.edu or http://www-gateway.vpr.drexel.edu.

Raj Mutharasan
July 26, 1997
September 10, 2000

Typical Course Outline: **Engineering Biotechnology**

3-0-3   winter 9x/9x

Instructors:
>   Raj Mutharasan, Chemical Engineering Dept., 27A-281, X2236
>   Wayne Magee, Biosciences & Biotechnology, 5-105, X6906
>   Maggie Wheatley, Chemical Engineering, 27A-478, X2232
>   Young Lee, Chemical Engineering, 27A-277, X2230

Instructional Objectives   To introduce underlying biological and engineering principles needed to develop new biopharmaceuticals based on their properties, structure and processes for manufacture. Specifically, the course will be structured to illustrate concepts in genetic engineering of insulin, production and purification of insulin, insulin delivery to human system and glucose sensing.
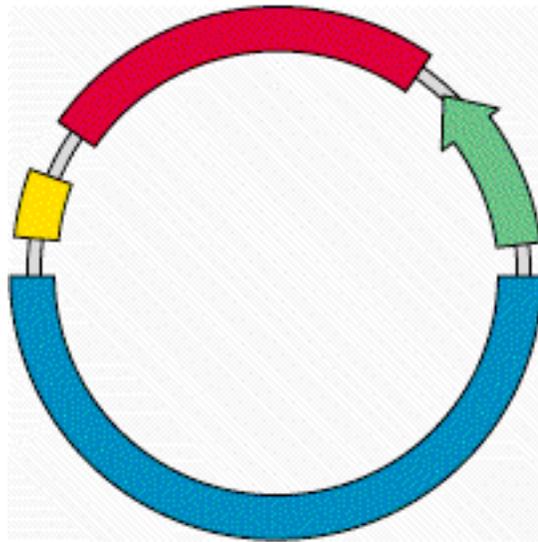
Organization of Topics and Lectures   Meets Tu for one hour and two hours on Th. The first session [labeled as 1] refers to the Tu meeting.

Week of	Instructor	Topic	Assignment
Jan 10	ALL WEM	1. Introduction 2. Insulin and Physiology	Reading on insulin
Jan 17	WEM WEM	1. Cloning 2. Identifying insulin gene & DNA amplification	Reading on cloning
Jan 24	WEM WEM, RM	1. Gene transfer , expression and regulation 2. Stoichiometry and thermodynamics of growth	"gene" calculations
Jan 31	RM RM	1. Design of High cell density reactors 2. Cell separation & disruption. Insulin Recovery	Reactor problem set
Feb 7	ALL RM	1. **[Mid Term 1]** 2. Good Manufacturing Practice & FDA Regulations	Separation problem set
Feb 14	MAW MAW	1. Protein properties relevant to drug delivery 2. Pharmaco kinetics and dynamics: Insulin & others	Manuf. problem set
Feb 21	MAW MAW	1.Insulin delivery requirements - strategies 2. Controlled release mechanisms & methodology	Kinetics problem set
Feb 28	ALL YHL	1. **[Mid term 2]** 2. Glucose sensing: application in insulin delivery	Diffusion/Kinetics set
March 7	YHL YHL	1. Detection methodology for in vitro applications 2. Use of biomolecules for sensing	Electrode problem set
March 14	YHL YHL, ALL	1. Design of biosensors for in vivo applications 2. Design of biosensors for in vitro applications Course Evaluation and student interviews	biosensor calibration & dynamics
March 21		**Finals Week**	

Recommended Text: There is no recommended text for this course. Copies of lecture notes will be provided during the course. Copies of recommended reading material will also be distributed.

MOLECULAR BIOLOGY

**Engineering Biotechnology
Gateway Project**

**Wayne E. Magee
Drexel University**

Contents

Chapter 1 Introduction

This section of the bioengineering course will introduce you to the molecular biology portion of the course. That is, we will examine the properties of the nucleic acids that constitute genes and then proteins which are the end product of gene action and the main focus of genetic engineering. We will investigate our model, the hormone insulin, its structure and the general role it plays in regulating the blood level of glucose. Individuals who suffer from diabetes either are unable to produce insulin, or have some defect in responding to the hormone signals.

There are many analogies between general engineering principles and the manner in which the body regulates glucose levels:

Fluid system
Glucose sensor
    --the beta-cell of the pancreas
Valves to regulate glucose level
    --glucose transporters under control by the hormone insulin
    --glucose release from liver under control by the hormone glucagon
Steady state glucose blood levels
    --analogous to a continuous feed bioreactor
Substrate (glucose) reservoir
    --the liver is the glucose buffer system of the body

Chapter 2 Molecular Biology

This section is intended to provide a brief overview of the properties of the components of genes (nucleic acids and their chemical constituents) and the end products of gene action (proteins and their amino acid sequences). Additional details will be introduced as needed in subsequent sections. The student should supplement the material given here with the suggested readings.

2.1 Properties of nucleic acids

Nucleic acids (deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are extremely long polymers made up of phosphate-sugar-nitrogenous base (nucleotide) units. The bases found in DNA are adenine and guanine (both purines) and cytidine and thymine (both pyrimidines). Thymine is replaced by uracil in RNA. The nucleotides are linked by 3' to 5' phosphodiester bonds. That is, a phosphate group on the 5' position of the sugar residue becomes linked to the 3' hydroxyl group of the preceding sugar group on the chain as the long polymer is synthesized.
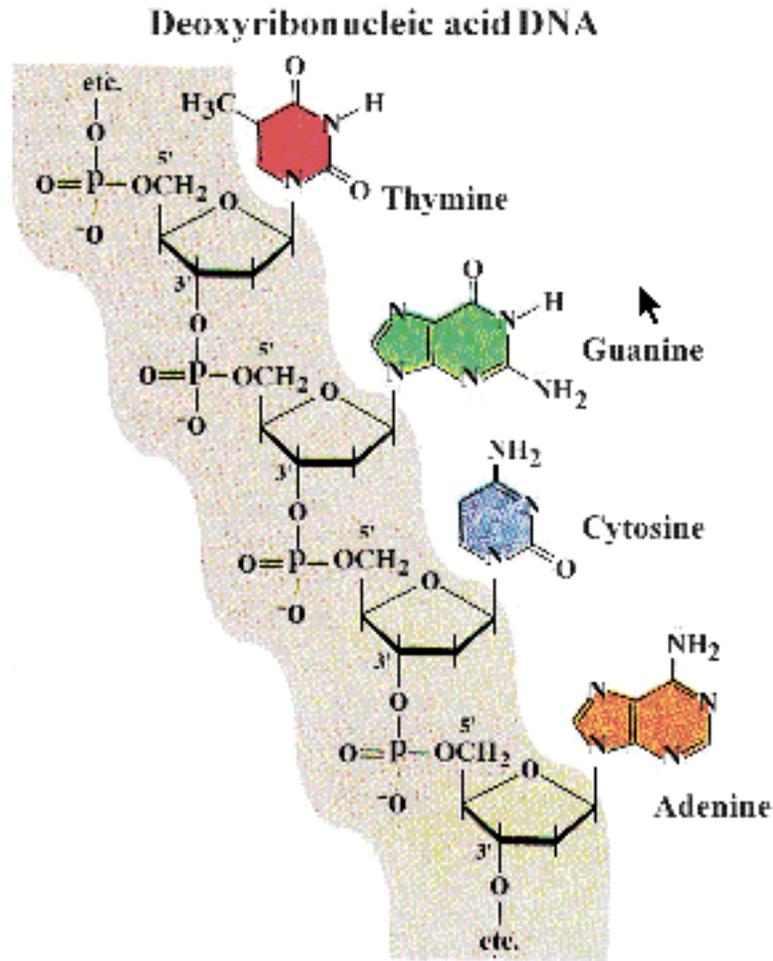
FIGURE 1. The structure of nucleic acids showing the bases found in DNA and how nucleotides are linked together

The double stranded structure of DNA was worked out in 1953 by Watson and Crick . It is absolutely essential to understand the key concepts behind this structure since all cloning strategies are based upon them.

The double stranded (double helix) molecule consists of two strands wound around a central axis with the bases stacked inside. The order of the strands are in opposite directions (5' to 3' in one, 3' to 5' in the other). The bases stack together in the center of the helix because they interact with one another *via* weak hydrogen bonds.
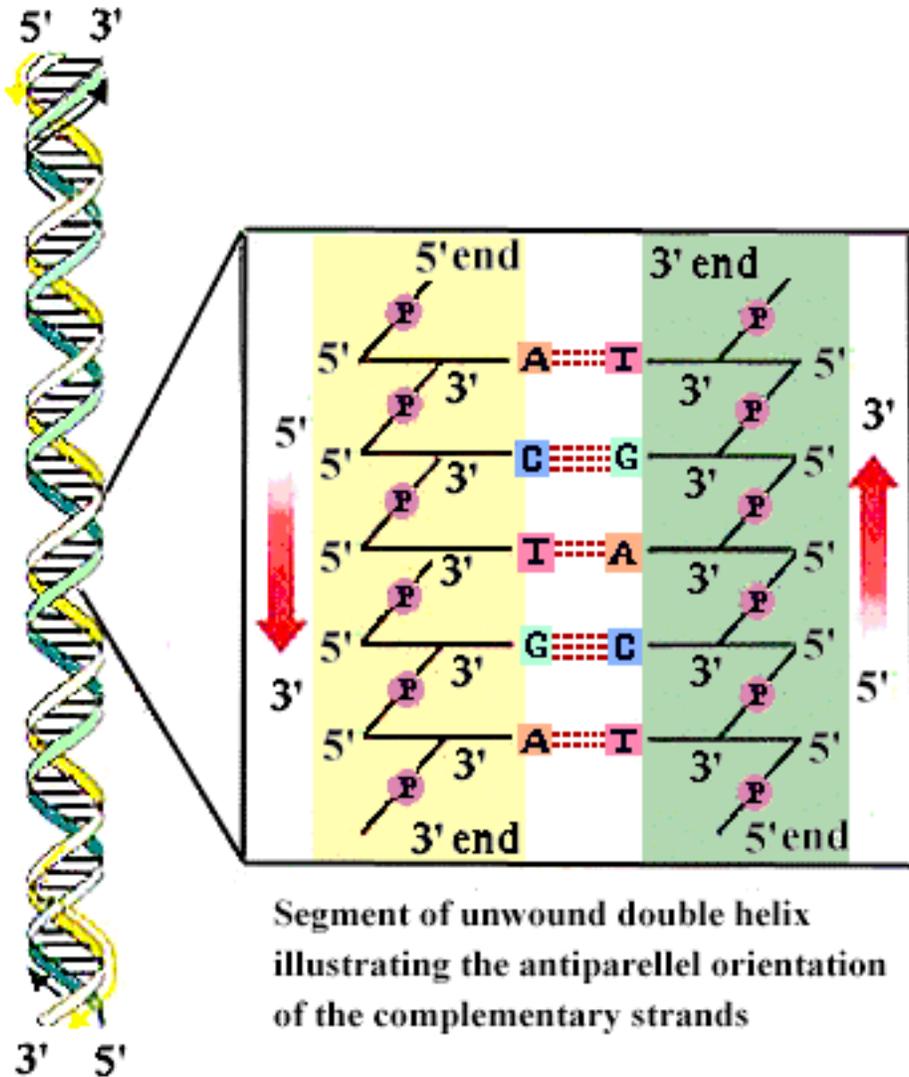
FIGURE 2. Structural features of the DNA double helix.

Hydrogen bonds are much weaker than covalent bonds and are continually forming and disassociating. They form between two electronegative atoms, one of which has a covalent hydrogen. The special properties of water are due to hydrogen bonds of the type:

H-O-H - - - O⟨ H, H (Hydrogen bond)
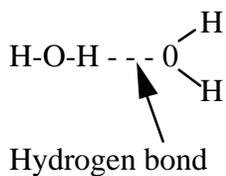
The hydrogen bonds found in nucleic acids are between N and O:

N-H - - - O-- and O-H - - - N--

In double stranded nucleic acids, adenine will only form hydrogen bonds with thymine (or uracil)

and cytidine will only form hydrogen bonds with guanine. Thus, the amount of adenine always equals the amount of thymine and the amount of cytidine always equals that of guanine for the DNA of a given species. The per cent G-C (and thus the per cent A-T) in DNA is surprisingly variable from species to species.

FIGURE 3. Base pairing between A:T and G:C.

The base pairs form a flat plane in the interior of the helix with adenine forming two hydrogen bonds with thymidine and cytidine forming three hydrogen bonds with guanine. It immediately oc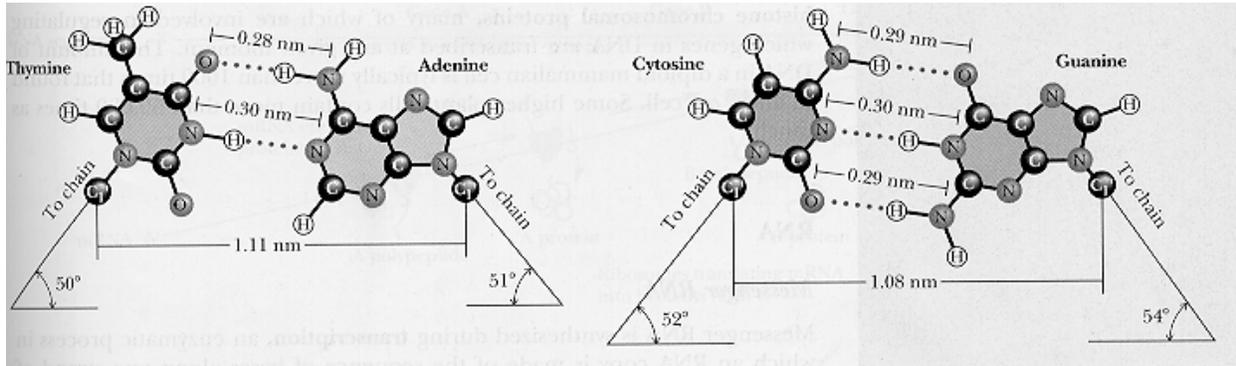curred to Watson and Crick that if you zipped open the two strands and began copying them so that everywhere that there was an adenine in the original strand, the new strand had a thymidine, (a guanine would be matched to a cytidine etc.), then soon you would exactly duplicate both strands using only the concept of base pairing and whatever enzymes and substrates that might be necessary to carry out the synthesis.

At the completion of replication, each original strand (parental strand) is paired with a new (daughter) strand. The replication is said to be semiconservative. Double stranded DNA may be denatured by alkaline conditions or by heat. For example, the two strands will unwind when a solution of DNA is heated to 90° and then will reassociate if the temperature is lowered gradually. Longer DNA molecules take longer to reform their double-stranded structure in perfect alignment.

Even small stretches of DNA (polynucleotides) will anneal with large single-stranded DNA molecules if the base sequence matches a sequence somewhere on the single stranded DNA (by base pairing). A sequence of 10 or more matching base pairs usually is needed to form enough hydrogen bonds for the complex to be stable. Many of techniques used to manipulate DNA make use of this observation.

2.2   Proteins and amino acids

Proteins are made up of long sequences of amino acids. There are 20 common amino acids used for synthesis of the many thousands of proteins found in living organisms. The properties of the proteins depend upon the exact sequence of their amino acids which in turn is determined by the genetic code discussed below. A major goal of genetic engineering is to be able to make useful quantities of peptides and proteins, some of which are as small as a dozen amino acids on up to proteins containing hundreds of amino acids. The proteins we want to make (perhaps using bacteria to produce them) might be human hormones, antibodies and other rare proteins that are difficult or impossible to obtain by other means.

General Formula:

$$R - \overset{\displaystyle NH_2}{\underset{\displaystyle COOH}{CH}}$$

ALIPHATIC AMINO ACIDS

R =

Glycine	H –
Alanine	$CH_3–$
*Valine	$\underset{CH_3}{\overset{CH_3}{>}}CH–$
*Leucine	$\underset{CH_3}{\overset{CH_3}{>}}CH–CH_2–$
*Isoleucine	$\underset{CH_3}{\overset{CH_3–CH_2}{>}}CH–$
Serine	$HO–CH_2–$
*Threonine	$\underset{CH_3}{\overset{HO}{>}}CH–$
Cysteine	$HS–CH_2–$
Cystine	$\overset{S–CH_2–}{\underset{S–CH_2–}{|}}$
*Methionine	$CH_3–S–CH_2–CH_2–$
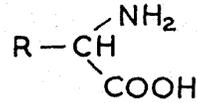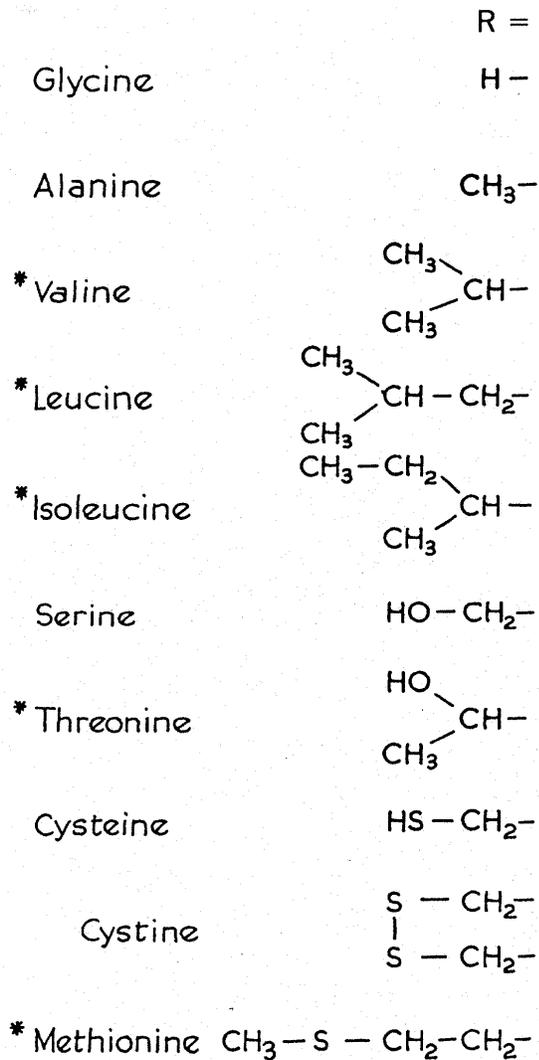
AROMATIC and HETERO-CYCLIC AMINO ACIDS:

*Phenylalanine ⬡$–CH_2–$

Tyrosine HO⬡$–CH_2–$

*Tryptophan (indole)$–CH_2–$

BASIC AMINO ACIDS

*Histidine (imidazole)$–CH_2–$

*Lysine $NH_2–CH_2–CH_2–CH_2–CH_2–$

Arginine $NH_2 \underset{\underset{NH}{C}}{NH}–CH_2–CH_2–CH_2–$
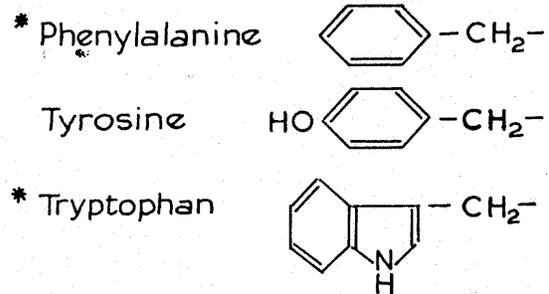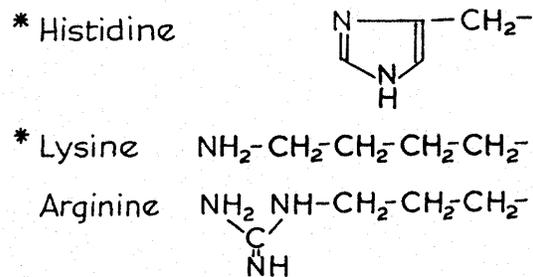
ACIDIC AMINO ACIDS and their AMIDES

Aspartic Acid $\underset{HO}{\overset{O}{>}}C–CH_2–$

Asparagine $\underset{NH_2}{\overset{O}{>}}C–CH_2–$

Glutamic Acid $\underset{HO}{\overset{O}{>}}C–CH_2–CH_2–$

Glutamine $\underset{NH_2}{\overset{O}{>}}C–CH_2–CH_2–$

Proline (pyrrolidine)$–COOH$

FIGURE 4. Structure of the common amino acids

Many proteins found in cells are enzymes that catalyze metabolic reactions. These proteins must have a three dimensional structure that encourages binding of substrates and one or more active sites where the reactions take place. Peptide hormones must be of precise structure in order to bind to their receptors on cells.

The exact amino acid sequence and the precise folding of the protein structure is essential for fulfilling their biological roles . Proteins fold into characteristic structures determined by (guess what) hydrogen bonding, usually into hel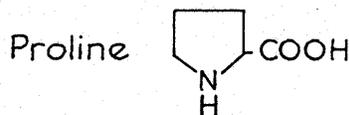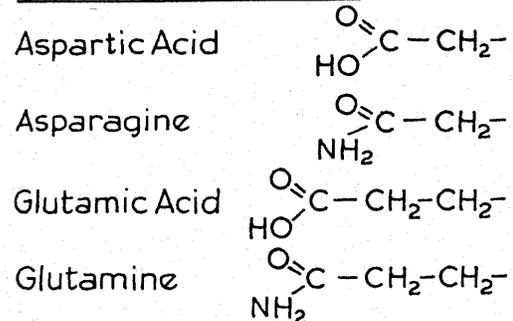ical form (a-helix) or sheet-like structures (b-sheets). A single protein often has several of each of these folding patterns. Also, during protein folding, the amino acids with aliphatic and aromatic side chains tend to locate in the interior of the protein away from water, with charged amino acids at the surface.

The naming of amino acids is trivial (asparagine was isolated from asparagus)

The usual structure is a carboxyl and an amino group covalently bound to the same carbon atom (the a-carbon):

```
NH2--CH--COO--
        |
        R
```
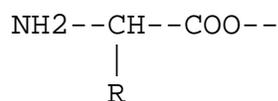
Various side chains are attached to the alpha-carbon (the R-group; see table). There are several amino acids that have only carbon and hydrogen on the R group (alanine, valine, leucine, isoleucine), several that have a positive charge at neutral pH (histidine, lysine, arginine), several that are acidic at neutral pH (aspartic acid, glutamic acid), and several that have aromatic rings (phenyl alanine, tryrosine, tryptophan). The amino acid cysteine ($HS-CH_2 = R$) is of particular interest because disulfide bonds between two cysteine residues are often found that cross-link between two peptide chains or between two parts of a single peptide (to give $R-Ch_2-S-S-CH_2-R'$). The amino acids often are represented by three letter abbreviations (ala, val, leu, iso, etc.) and even one letter abbreviations (A=alanine, R=arginine, D=aspartic acid, etc) to save space when writing out sequences.

α-Helix
only the N-Cα-backbone
is represented. The vertical line
is the helix axis.
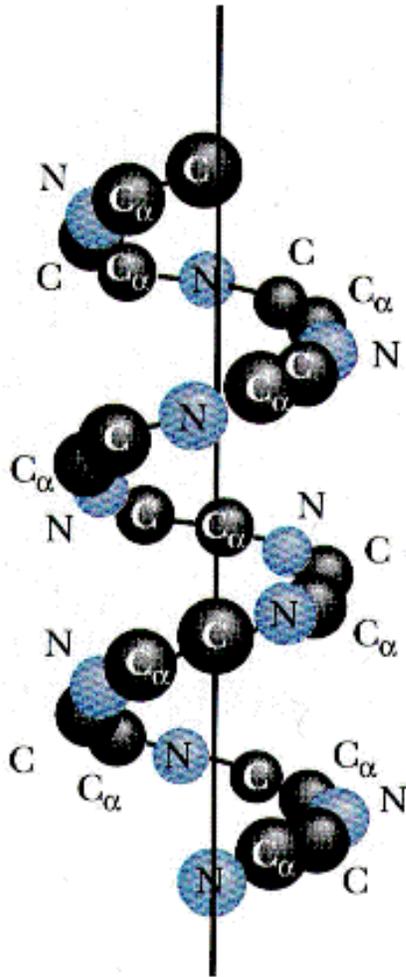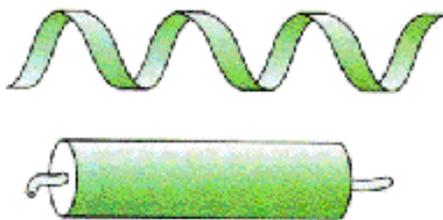
β-Strand
Note that the amide planes
are perpendicular to the page.

"Shorthand" α-helix

"Shorthand" β-strand
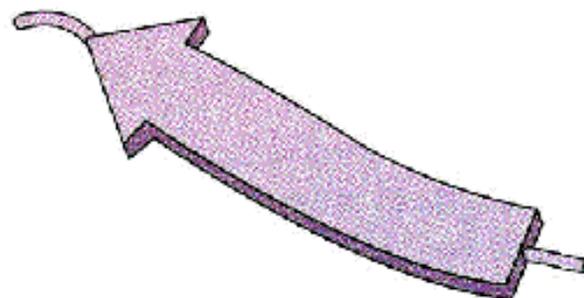
FIGURE 5. Proteins are made up of combinations of helical and sheet structures

During synthesis of proteins, amino acids are added one at a time according to the genetic code using complex structures called ribosomes. Each amino acid is added to the preceding one by a splitting out of water between the carboxyl group of one amino acid and the amino group of the next amino acid. The resulting bond between the two is known as the peptide bond:
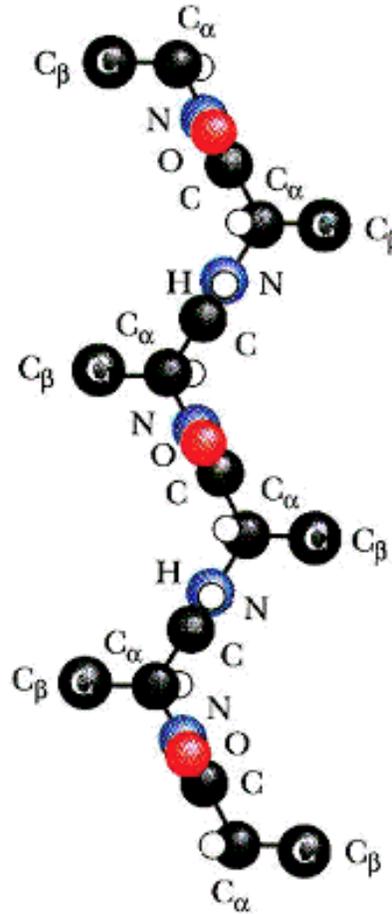
--CH--CO--NH--CH--COO--

It follows that the first amino acid of a protein usually will have a free amino group (the amino end) and the last amino acid will usually have a carboxyl group (the carboxy end). I have said "usually" this is true, because sometimes the ends of proteins are modified after synthesis.

2.3 Replication of Nucleic Acids

Polymerase enzymes catalyze the duplication of nucleic acids. All of the DNA polymerases can only catalyze the addition of additional nucleotides to an existing piece of DNA that has a free 3'-hydroxyl group (they are **unable to start a new strand** without a primer attached). This limitation imposes some special conditions on the way new DNA is made in cells. The reaction itself uses deoxynucleotide triphosphates as substrates. The triphosphate supplies the free energy to drive the reaction (figure). Note that the enzyme only will add the next nucleotide base onto a primer, and the base to be added must be able to hydrogen bond with the base exposed on the strand to be copied (the template strand). During the reaction, inorganic pyrophosphate is liberated and immediately hydrolyzed to inorganic phosphate, insuring that the reaction will be irreversible.

The reaction only proceeds in the 5' to 3' direction (new bases are added to the 3' hydroxyl of the preceding sugar.
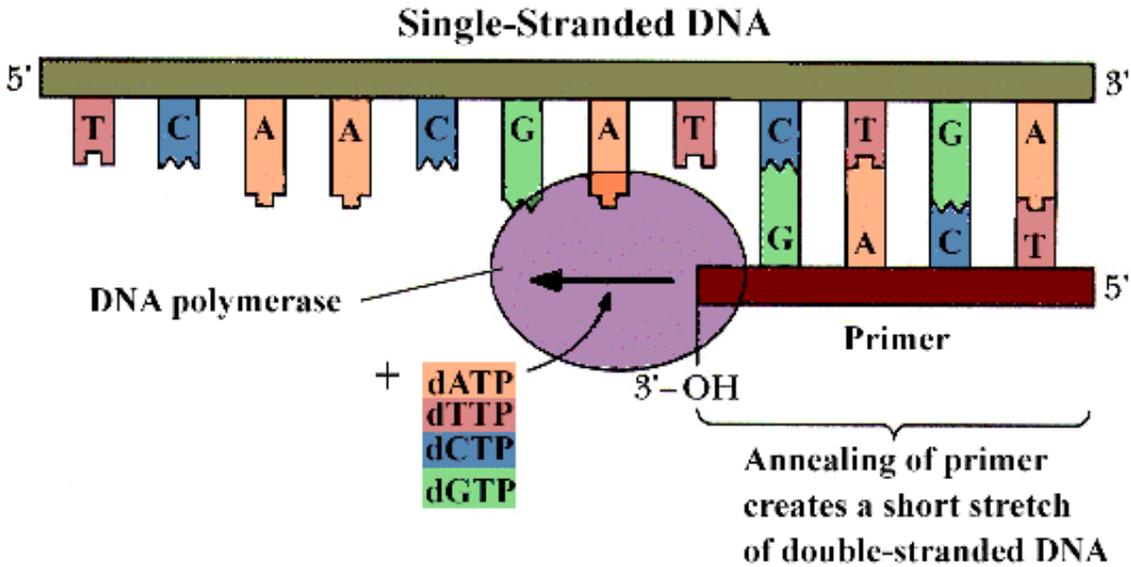
FIGURE 6. DNA polymerase requirses a primer, a template stand and the four deoxy nucleotide triphosphate substrates

What are some of the requirements for DNA replication starting with intact DNA? First of all, replication starts at special sequences of bases that define the origin of replication (indicated as *ori* in our models). Then we must begin to unwind the DNA to provide access for DNA polymerase and other needed enzymes and proteins. As the DNA unwinds, we see two fork shaped areas that are indeed the replication forks. They proceed in both directions down the double stranded DNA, and each must have its own set of enzymes. Remember that the two strands of DNA are running in opposite directions. One strand is ready to serve as a template for the 5' to 3' synthesis, but the other is in the wrong orientation. To solve this problem, RNA polymerase is used to synthesize a small stretch of RNA (Okazaki fragment, named for the scientist who discovered this mechanism) to provide a primer so that DNA polymerase may continue the chain. The RNA stretches are removed, filled in with DNA and sealed together with an enzyme called ligase. A more detailed representation of the replication fork complex is shown in Figure 7. Notice that proteins are needed to help unwind the DNA, which is tightly coiled, before the synthesis may continue. By looping one strand around (the one that needs the RNA start) we may visualize the replication complex as all occurring as one large aggregate at the replication fork.
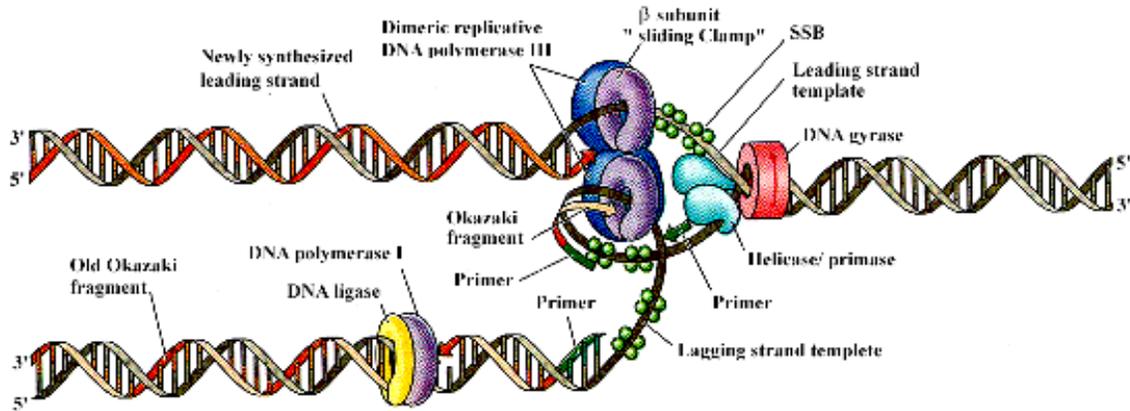
FIGURE 7. A more complete represenation of DNA synthesis at a replication fork

2.4   Making an RNA copy of DNA

You may already know that it is RNA that carries the genetic code to ribosomes with instructions for protein synthesis. Our master DNA tape stays safely away from the action and sends messenger RNA out with the code.

One strand of DNA is copied into RNA using RNA polymerase. In order to copy the genetic code from DNA molecules, we need to copy one of the strands of the DNA into RNA using RNA polymerase and the same base paring principles that we used for DNA synthesis. Again we have the problem of gaining access to a tightly wound coil of DNA. The DNA must be unwound slightly to allow the enzyme to bind and begin copying the sequence.

Control regions are found upstream of the coding region of the gene. The binding only occurs at a region of the gene called the **promoter** region. This base sequence occurs before the 5' end of the gene (upstream) along with many other sequences that serve to control reading of the gene. Synthesis proceeds with the required base paring (remember adenine now pairs with uracil) using the ribonucleotide triphosphates (containing the sugar ribose) as substrates. The next figure shows a more complete model of the polymerization with an indication of how the unwinding of DNA coils may occur.
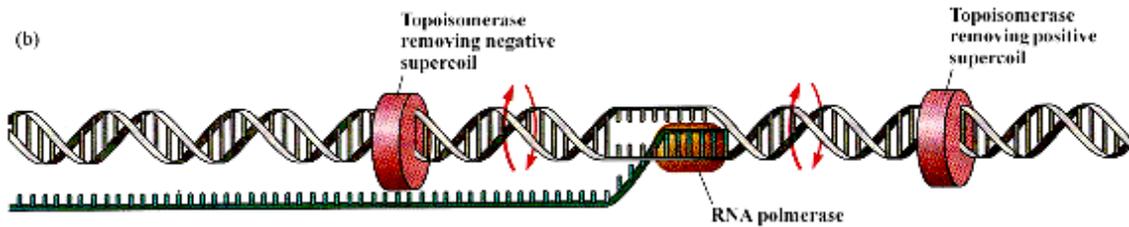
FIGURE 8. Copying the DNA sequence into an RNA is accompanied by local unwinding of the DNA

There are some differences between bacterial messenger RNA (mRNA) and that of higher organisms that have great importance if we are cloning the gene. Bacterial mRNA is used directly without modification. In fact, ribosomes often attach to mRNA of bacteria as it is being synthesized and immediately use the code for new protein synthesis.

The RNA of higher organisms must be processed before it becomes messenger RNA. In higher organisms, there are stretches of RNA that are cleaved out and discarded (called introns). The ends of the remaining RNA that contain the actual code (called exons) are reattached (spliced) to each other to give a shorter mRNA that is now ready to serve as a coding strand for protein synthesis. The mRNA of higher organisms also usually has polyadenylate tail added to the 3' end of each mRNA.
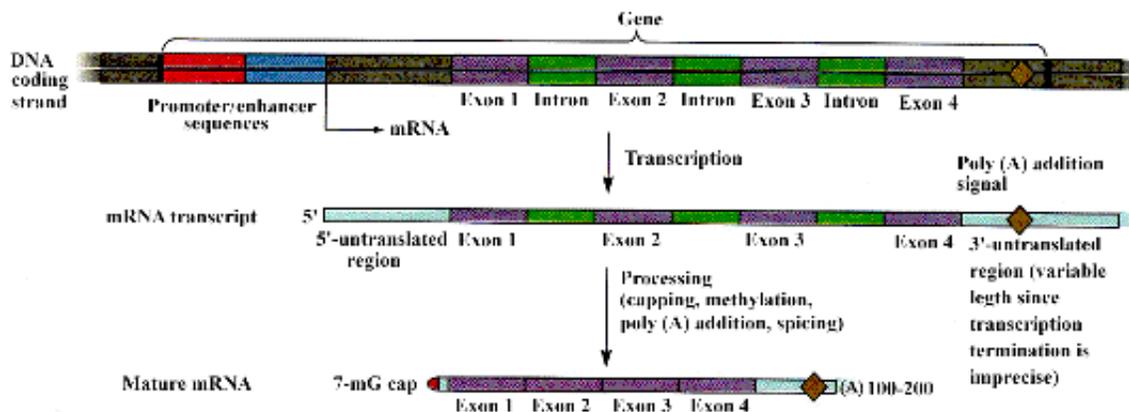
FIGURE 9. The organization of eukaryotic genes

2.5   The Genetic Code for Protein Synthesis

Scientists puzzled for some time on how four nitrogenous bases could provide the code for the 20 amino acids used in protein synthesis. The puzzle was solved once it was realized that three bases were used to code for one amino acid. Now the number of possible combinations could be calculated as $4^3 = 64$, or more than enough!

In general, each amino acid can be coded for by several different three-base codons (the code is redundant; see table). For example, leucine is coded for by CUU, CUC, CUA, CUG, UUA, and UUG. Notice that in the first four examples, only the last base is different. Three codons, TAA, TAG, and TGA are used as underlined stop codons to tell ribosomes to stop reading the code and stop making protein. Methionine has only one codon (AUG) and it is used as the first amino acid to start a new protein (it sometimes is removed later, as all proteins do not have an N-terminal methionine).

The Genetic Code

| First Position (5'-end)	Second Position				Third Position (3'-end)
	U	C	A	G	
U	UUU Phe	UCU Ser	UAU Tyr	UCU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UCC Cys	C
	UUA Leu	UCA Ser	UAA Tyr	UCG Stop	A
	UUG Leu	UCG Ser	UAG Tyr	UCG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	UCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	UCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Ser	A
	AUG Met*	ACG Thr	AAG Lys	AGG Ser	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

*AUG signals translation initiation as well as coding for Met residues.

FIGURE 10. *The genetic code*

We now may think back concerning the necessary sequence of events. One strand of the DNA must be copied into an RNA sequence beginning at the promoter region. Either DNA strand may be used, depending on the location of the promoter region. Each three-base code of the RNA

specifies a specific amino acid that will be polymerized together on ribosomes to give the final amino acid sequence which, in turn, will determine the properties of the new protein.

We will not go into the details of protein synthesis, but the general features are shown in the figure. Ribosomes assemble on the mRNA in the region containing a ribosome binding site and the initiation codon (AUG for methionine). Amino acids are brought to the proper site on the ribosome in the form of complexes with special small RNA molecules known as transfer RNAs. Each transfer RNA is specific for its own amino acid and has an "anti-codon" on one end of the loop shaped molecule and the amino acid at the other. In the case of methionine, the anti-codon would be UAC which will associate with AUG of the bound RNA. After a second transfer RNA with its amino acid enters the site, a peptide bond is synthesized between the two amino acids. The ribosome then shuttles three bases along the mRNA so that the next transfer RNA and its amino acid may enter and undergo polymerization. The process continues until a stop codon is reached and the ribosome falls off the mRNA.

(a) Each tRNA binding site extends over both sub-units and matches a triplet codon in the mRNA

Ribosome movement

P site   A site

5'   3'
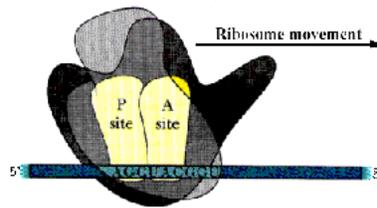
(b) Prior to peptide bond formation, an aminoacyl-tRNA is present in the A site and polypeptidyl tRNA is situated in the P site

Amino acids

5'   3'
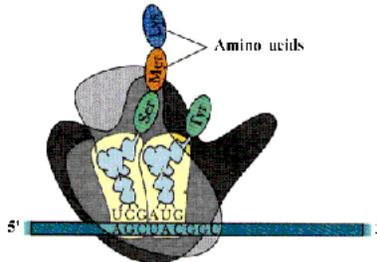
(c) Peptide bond formation involves transfer of the polypeptide to the amino group of the amino acid carried by the tRNA in the A site

5'   3'
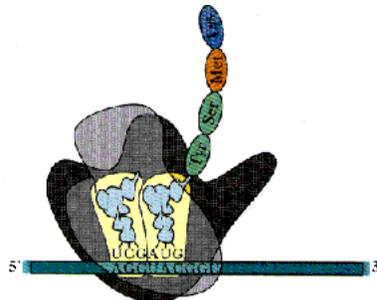
(d) The ribosome then translocates one codon further along the mRNA and the uncharged tRNA is expelled. Translocation places the polypeptidyl-tRNA in the P site and aligns a new codon within the A site, ready to accept the next incoming aminoacyl tRNA.

5'   3'
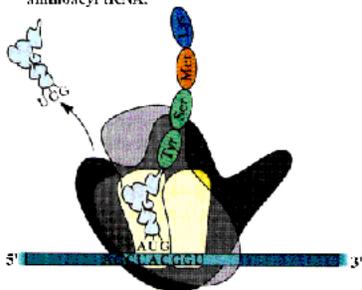
FIGURE 11. The synthesis of proteins on ribosomes

Some proteins have short stretches of amino acids that serve to direct the protein into specific cell compartments as it is being synthesized. These "signal" sequences are then removed by specific enzymes. In this way, proteins may be captured into vesicles for excretion out of the cell (insulin) or, in bacteria, may be directed to the space between the inner and outer membranes or excreted.

2.6   Making a Genetic Probe

Often we may know the amino acid sequence (or part of it) of a protein of interest, and frequently we will wish to synthesize a short nucleic acid sequence that might correspond to the gene sequence used by the cell. Such polynucleotides may be used as molecular probes to help locate the gene and as primers for polymerase reactions.

The redundancy of the genetic code complicates our problem because we have no real way of guessing which codon the cell may be using for a particular amino acid (Figure 12), although a given organism often uses some redundant codons more efficiently that others. In this case, we would choose the sequence starting with glu which uses two codons and proceed through a region where there is a minimal number of possible codons. We still need to consider ten possible sequences in this stretch of 15 nucleotides.

AMINO ACID SEQUENCE   **leu-glu-asn-tyr-cys-asn**
POSSIBLE CODONS: UUA GAA AAU UAU UGU
AAU
                         UUG GAG AAC UAC UGC
AAC
                         CUU
                         CUC
                         CUG
SYNTHETIC PROBES: GA(A)-AA(U)-UA(U)-
UG(U)-AA(U)
                              (G)   (C)   (C)
(C)   (C)

FIGURE 12. A nucleic acid probe should be selected around the least redundant sequece of possible codons. This sequence of amino acids is taken from the insulin molecule. (see FIG 14)

2.7   The Control Regions of Genes

It is essential that the cell be able to regulate the relative amounts of proteins and enzymes that are synthesized according to the needs of cell in its environment.   The process is very well studied in bacteria and an example of one such regulatory region and how it works was worked out by Jacob and Monod and is shown in the figure. *E. coli* makes the proteins and enzymes that are needed to metabolize lactose only if lactose is present in the environment, and actually prefers to use glucose if it is available. Whether or not these genes are transcribed is under control of a repressor protein and several other factors, including cyclic AMP. The gene for the

repressor protein lies upstream from the lactose control region and, in the absence of lactose, the repressor (as a tetramer) binds to a gene segment called the operator and prevents RNA polymerase from binding at the adjacent promoter site. If lactose is present, it binds to repressor and lowers the binding affinity of the repressor for the operator site. An enhanced reading of the lactose genes occurs upon binding of CAP protein (catabolite activator protein) and cyclic AMP leading to an active transcription of the genes and enabling the cell to metabolize lactose. The level of cyclic AMP in the cell is regulated by glucose. Thus, when glucose concentration is high, cyclic AMP concentration declines, and the lactose genes will not be read. We will use parts of this same system later in cloning experiments. The *lac*Z gene is the structural gene for b-galactosidase, the enzyme that hydrolyzes lactose to glucose plus galactose. It can accumulate to equal as much as 10% of total soluble, cellular protein following induction.
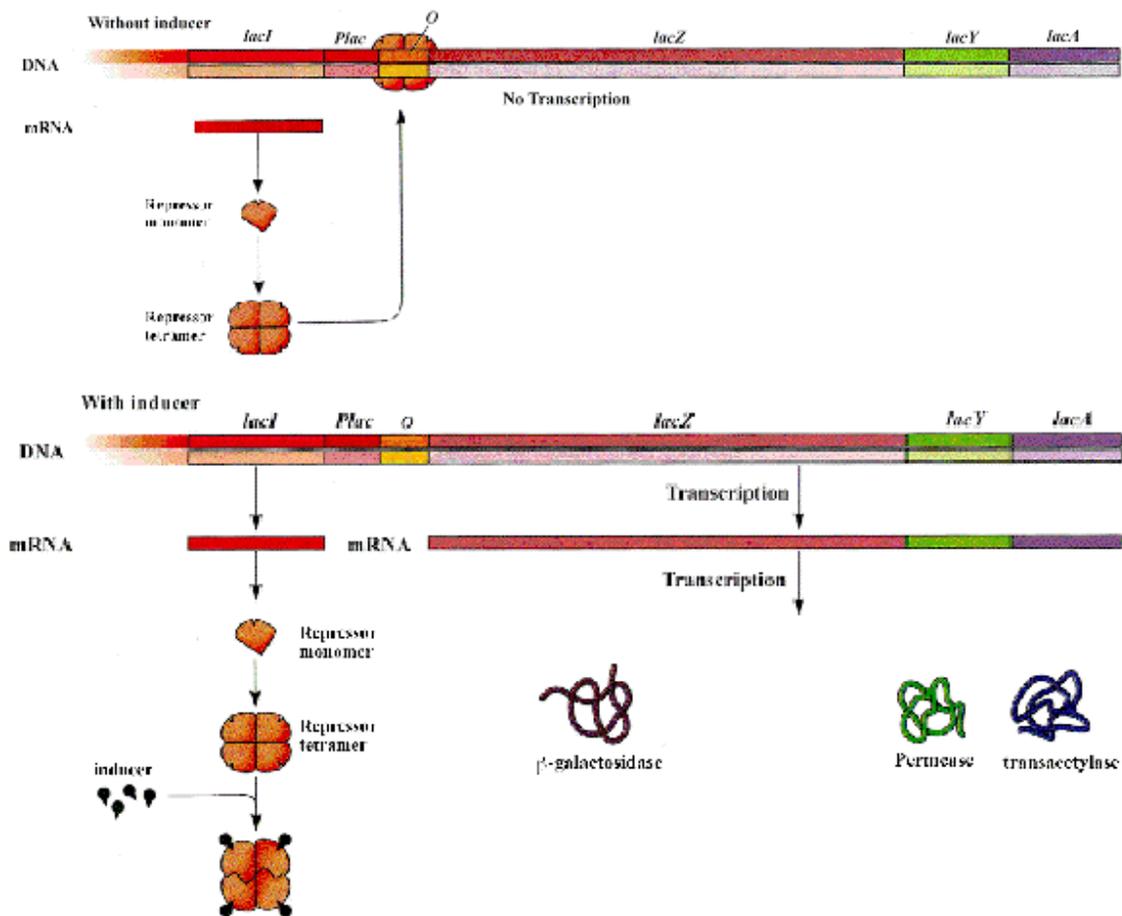
FIGURE 13. Control of transcription of genes involved in lactose metabolism

2.8 Insulin and its Genes

We now may examine the amino acid sequence of insulin and how its gene is organized. Insulin is a small protein made up of two peptide chains held together by disulfide bonds

between cysteine residues. The A chain has 21 amino acids and the B chain 30. The amino acid sequence was determined by Sanger in 1955.
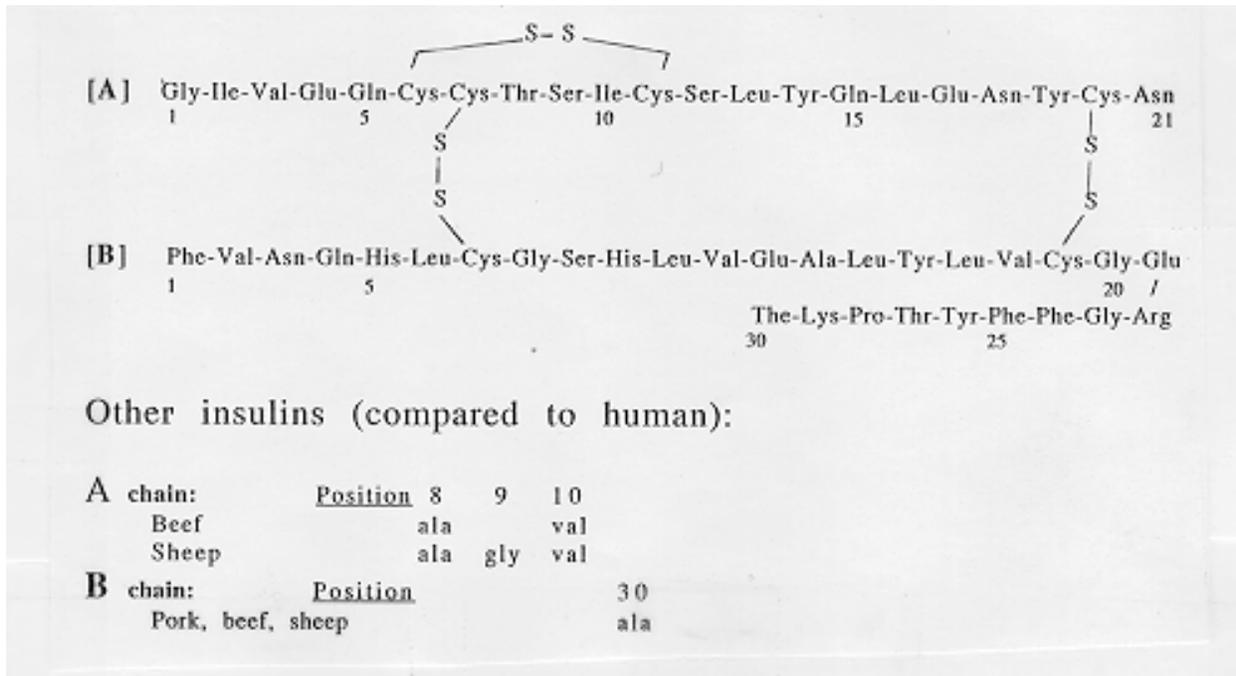
FIGURE 14. The structure of insulins from various sources

Insulin is produced only in the beta cells of the pancreas. Of course, the gene for insulin is in all cells, but insulin is not produced in other tissues because of special control regions that lie in front of the gene and prevent it from being expressed. Proinsulin is the precursor of insulin and contains an extra peptide chain (the C chain).   The C chain that links chains A and B allows the entire molecule to be synthesized in the b-cell as one protein.The C chain is split out after proinsulin is made to give the mature insulin structure of 51 amino acids with a molecular weight of 5,700 .

Insulin is actually produced in the b-cell as **preproinsulin** with an extra 24 amino acids which serves to guide the new protein to vesicles where it is stored. This "signal sequence" is removed shortly after protein synthesis begins.
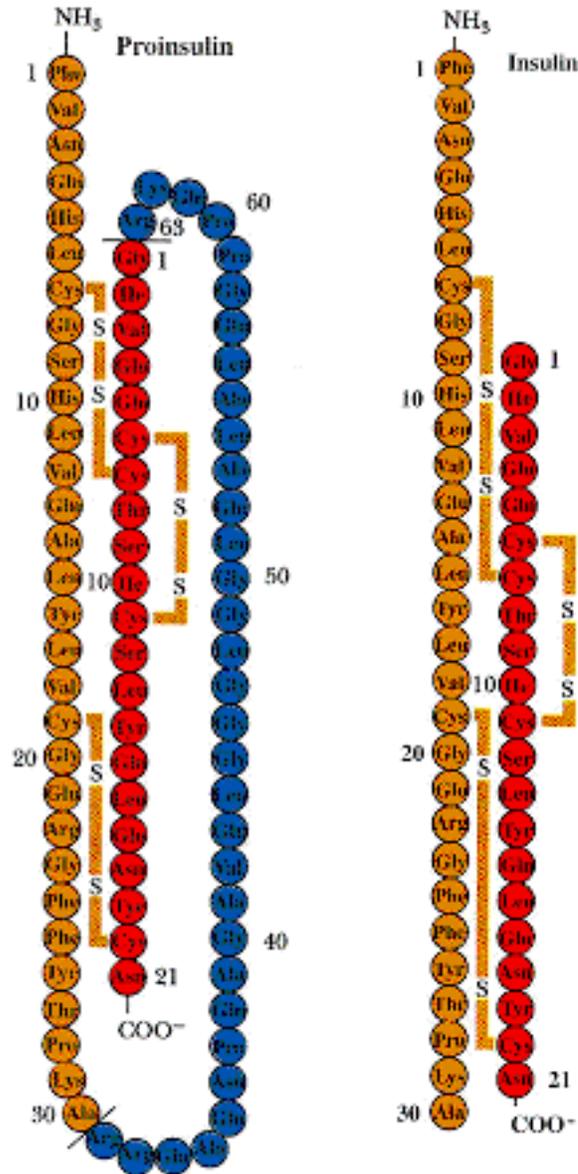
FIGURE 15. Processing of insulin precursors to give mature insulin

Insulin is a member of a large family of related proteins and over 80 of these have been sequenced. Even insects have insulin-like proteins. The disulfide bonds are conserved, and the molecules form three helices in their native form. The three-dimensional shape of the insulin molecule is shown in the figure. Other animal insulins have almost identical amino acid sequences to that of human insulin, which means that we may sometimes use them in place of human insulin.

Pork insulin has only one amino acid change; in chain B at position 30, alanine is found in place of threonine

Beef insulin has two additional changes; in chain A at position 8, alanine is found in place of threonine and at position 10, valine is found in place of isoleucine

Chapter 3 Cloning Genes

3.1   Overview

Now we are ready to examine the technical procedures (and some of the theories behind them) that allow molecular biologists to detect, study and clone individual genes--and even produce the proteins coded for by these genes in commercial amounts. These developments have sparked a revolution in the biological sciences and will continue to have far-reaching implications for every area of biology and medicine in the years ahead.

The practice of genetic engineering, or gene cloning, had to await the development of suitable methods in the early 1970s for cleaving, joining and manipulating nucleic acids. The basic ideas behind the procedures are easy to state: genetic information is encoded in genes which are made up of long polymers of nucleotides in the form of double stranded DNA molecules.

Cleave genomic DNA into large fragments using restriction endonucleases.
Restriction endonucleases are used to cleave genomic DNA into large fragments at defined nucleotide sequences.

Insert the DNA into plasmids. These pieces of DNA are inserted into plasmids, which are self-replicating, extrachromosomal genetic elements originally isolated from the bacterium, *Escherichia coli*. The circular plasmid DNA is opened using the same endonuclease that was used to cleave the genomic DNA.

Join the ends of DNA with the enzyme, DNA ligase. The inserted DNA is joined to the plasmid DNA using another enzyme, DNA ligase, to give a recombinant DNA molecule. The new plasmid vector contains the original genetic information for replication of the plasmid in *E. coli* plus the inserted DNA which may contain, for example, the human gene for insulin.

Introduce the new vector into *E. coli*.   The new vector is inserted back into *E. coli* where many copies of the genetic sequence are made as the bacteria grow and divide with the replicating vector inside.

Isolate the newly-synthesized DNA, or the protein coded for by the inserted gene.
The *E. coli* may even transcribe and translate the gene and obligingly produce insulin, if

that was the gene incorporated into the vector. Alternatively, many copies of the DNA gene itself may be isolated for sequencing the nucleic acid or for other biochemical studies.

Many variations on the basic methodology have been developed. We will consider only three: the use of vectors made from bacterial viruses (or bacteriophages); cosmid vectors; and the polymerase chain reaction (PCR). Additional methodologies may be found in the reference list.

Eukaryotic
chromosome

Cloning vector
(plasmid)

cleavage with restriction
endonuclease

cleavage with restriction endonuclease,
separation of fragments

DNA ligase

Recombinant
vector

transformation (or viral infection)
of bacterial host cell

cell propagation
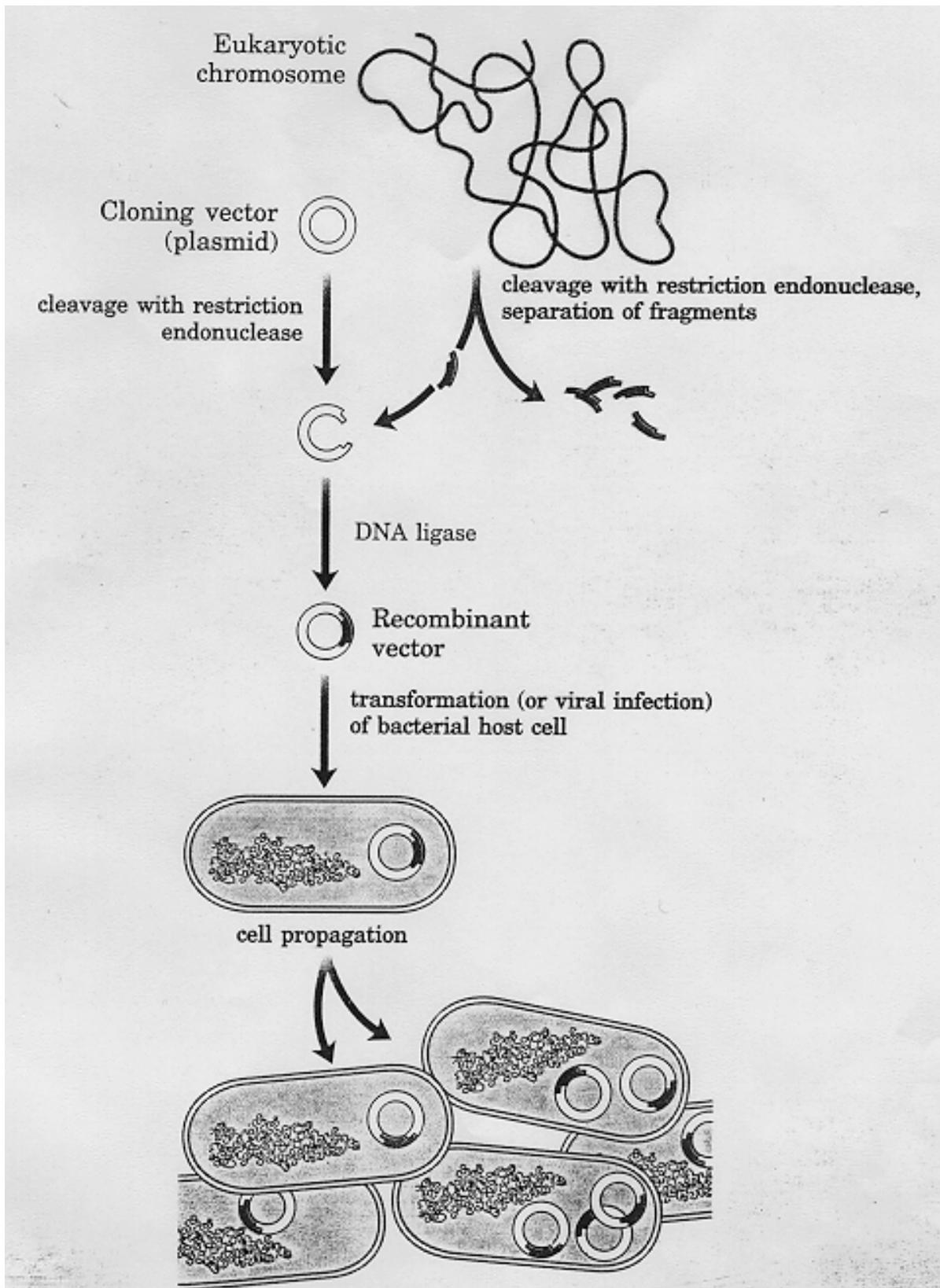
FIGURE 16. The general scheme for genetic engineering

3.2   Cutting DNA

The most important enzymes for gene cloning are the restriction endonucleases. These nucleases make internal cuts in the nucleic acid molecule only where certain sequences of nucleotides are present. Thus, they tend to cleave DNA into rather large pieces rather than small fragments.

These enzymes are found in various bacterial species. They are believed to protect the organisms from any invading foreign DNA, such as might be introduced by bacterial viruses. The DNA of the host is protected by methylation of the DNA bases that are a part of the recognition sequence.

> *Example Escherichia coli K has a restriction endonuclease EcoK which is coded for by the gene hsdR. However, EcoK does not cleave the host DNA because E. coli K has a methylase coded for by the gene hsdM that methylates the bases in the region where the endonuclease would cleave unprotected DNA. Both of the enzymes require the presence of gene hsdS which is needed for DNA recognition.*

Over 800 restriction endonucleases are known. The most useful for cloning work are known as Type II which generally make a staggered cut within the recognition site and across the double stranded DNA.

The enzymes are named by a three letter code indicating the organism from which they were isolated. For example, *Eco*RI was isolated from *E. coli* and *Bam*HI from *Bacillus amyloliquefaciens*.

The action of *Eco*RI is as follows:

5'- - G AATT C - - -
3' - -C TTAA G - - -

5' - -G-3'OH   **AATTC - -**
3' - -CTTAA-   **3'OH-G - -**
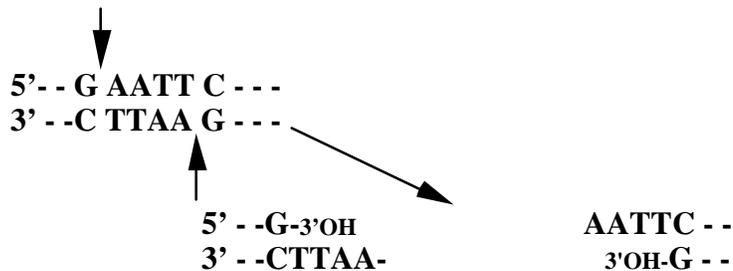
Notice that the enzyme recognizes a six-base pair sequence in the DNA, and the cleavage results in single-stranded ends (sticky ends) that would readily reform hydrogen bonded base pairs again with each other.

Most restriction endonucleases recognize four, five or six base pair sequences. It can be calculated that if the base sequence of DNA were random, then the expected frequency for a given sequence would be 4 (for the number of different bases to the nth power:

Expected frequency of cleavage $= 4^n$ where n is the length of the recognized sequence

Thus, the *Eco*RI site might be expected to occur about every 4096 base pairs. Enzymes that recognize a four base pair sequence (such as *Hae*III)would be expected to produce shorter fragments (around 256 base pairs in length). Other restriction endonucleases are shown in the table.After cleaving with restriction endonucleases, plasmid DNA and the DNA to be inserted are allowed to anneal *via* their sticky ends.

BamHI (5') G G A T C C (3')
    C C T A G G

ClaI (5') A T C G A T (3')
   T A G C T A

EcoRI (5') G A A T T C (3')
    C T T A A G

HaeIII (5') G G C C (3')
    C C G G

HindIII (5') A A G C T T (3')
    T T C G A A

NotI (5') G C G G C C G C (3')
   C G C C G G C G

PstI (5') C T G C A G (3')
   G A C G T C

PvuII (5') C A G C T G (3')
    G T C G A C

SmaI (5') C C C G G G (3')
   G G G C C C

Tth111I (5') G A C N N N G T C (3')
    C T G N N N C A G
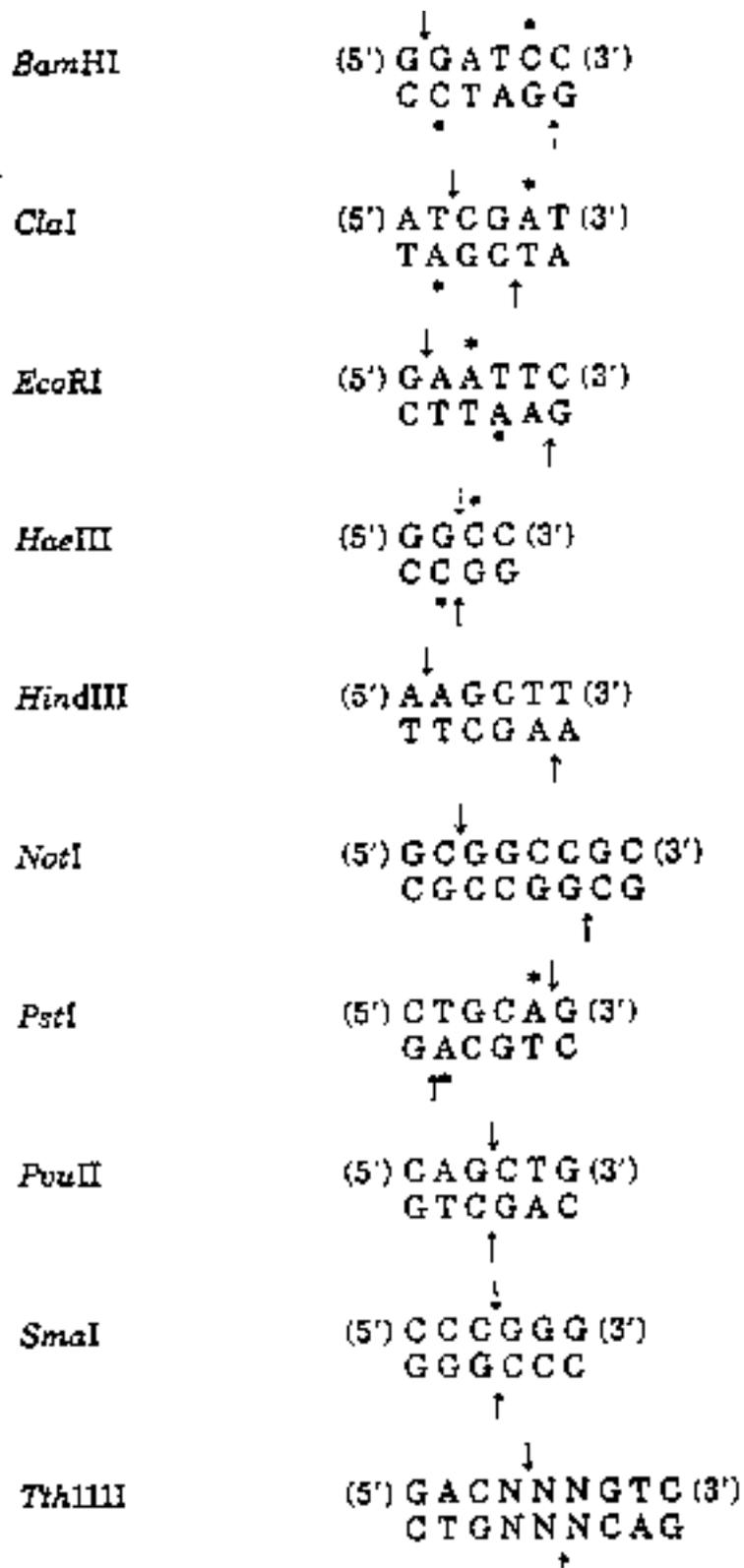
FIGURE 17. Some commonly used type II restriction endonucleases
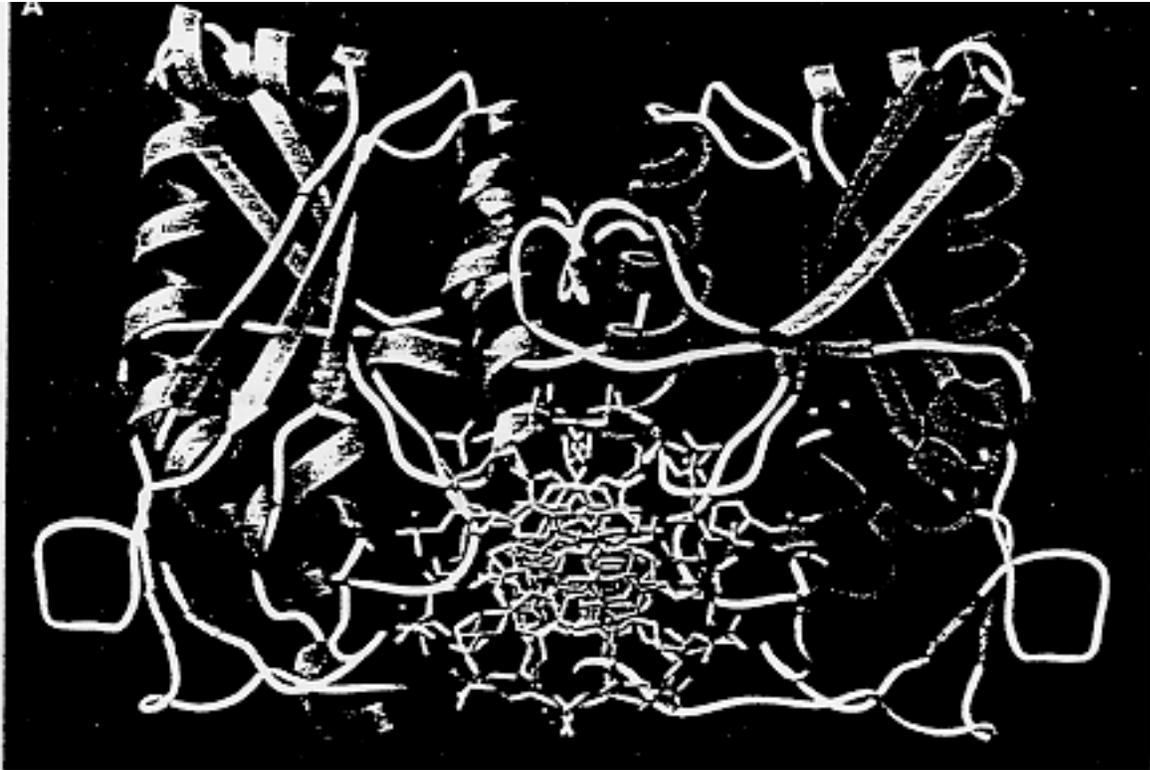
FIGURE 18. The three-dimentional structure of a restriction endonuclease showing the binding of DNA

3.3   Joining DNA: DNA Ligase

The second key enzyme for cloning is DNA ligase. This enzyme will rejoin DNA strands that are close together and that have a free -3'-OH facing an adjacent 5'-phosphate. This will come about naturally if the sticky ends of DNA molecules that have been cleaved by the same nuclease are allowed to anneal with one another. The most commonly used DNA ligase is one coded for by bacteriophage T4. It may be purified from *E. coli* cells that are infected by T4.

```
            3'OH                               5'
    ---------G-                        pApApTpTpC-------
    ---------CpTpTpApAp                      G----
               -5'                          3'-OH
                         |
                         |
                         |          ANNEAL
               OH        |
                         ▼
              3'   5'
    ---------G pApApTpTpC-------
    ---------CpTpTpApAp G------
                      -5'   3'OH
```
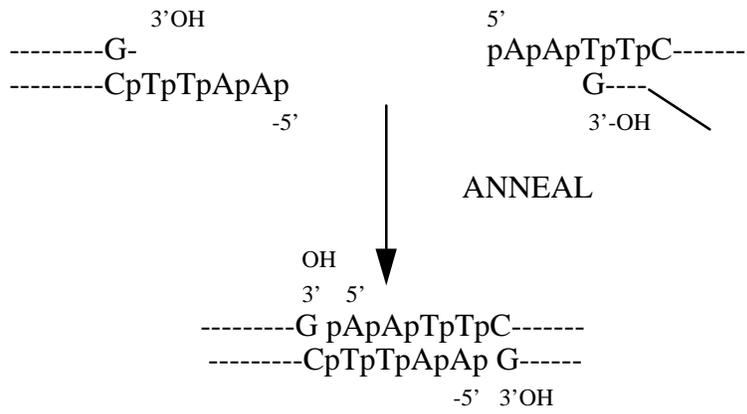
DNA ligase is activated by ATP which adds an adenylate group to the enzyme. The adenylate group is transferred to the 5' terminal phosphate of the DNA:

```
          3'OH   Adenosine Monophosphate
            | |
    ---------G pApApTpTpC-------
    ---------CpTpTpApAp G------
                       | |
                      AMP  3'OH
```
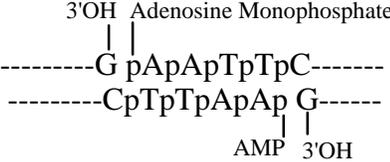
The 3' OH now can perform a nucleophilic attack on the phosphate linkages and form the phosphodiester bond to seal the stands:

```
        5'---------GpApApTpTpC-------3'              + 2 AMP
        3'---------CpTpTpApApG------5'
```
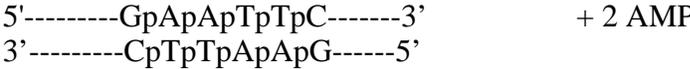
The enzymatic reaction is run at 5-15° (which is below its temperature optimum of 37°) so that the hydrogen bonds which hold the sticky ends together will not dissociate. Sometimes the ligase will simply reseal the original plasmid back together, and often it may seal together the two ends of the DNA gene that we wish to insert instead of inserting it into the plasmid. We will see how to deal with these problems when we screen for transfomants. Since there is a binding site for both strands of the DNA on the ligase enzyme, it is possible to bind and ligate blunt-ended DNA molecules, but with much lower efficiency.
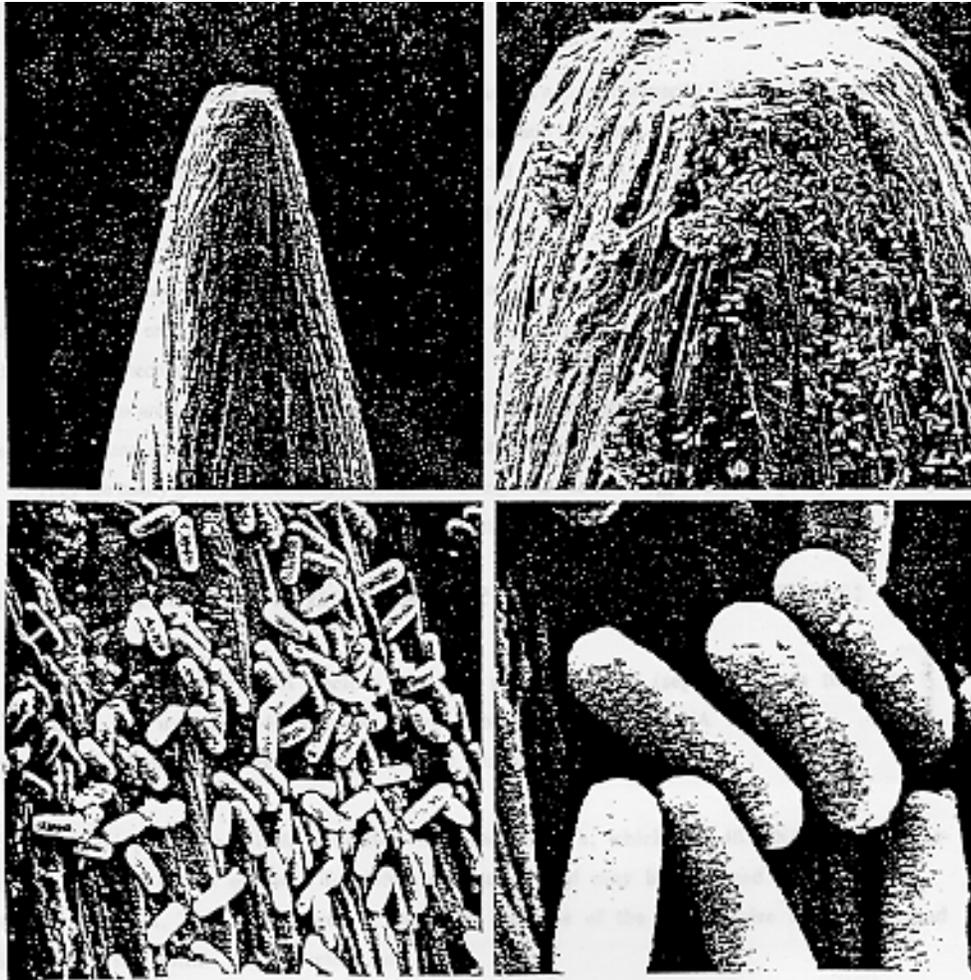
3.4   Host Strains

The most commonly used host strains have been derived from *E. coli* K-12, although other bacteria, yeast cells, plant cells and mammalian cells are used according to the needs of the experiment. For example, if you are producing a human protein that has specific carbohydrate residues needed for activity, it may be essential to produce that

protein in mammalian cells. Also, eukaroytic genes usually have introns and exons that cannot be directly transcribed into useful mRNA by prokaryotes, such as *E. coli* (see molecular biology and cDNA libraries sections). Sometimes a bacterial host may be used for initial experiments and gene isolation, and then the vector is moved to a mammalian host for expression.

Our examples use *E. coli* K-12 strains that have the following properties:
1. These bacteria are easy to handle in the laboratory, and may be used with many different vectors. The fundamentals of genetic engineering were developed using these organisms and are based upon an exhaustive knowledge of their molecular genetics.
2. The strains are selected to contain mutations in host restriction and usually other mutations involving nutritional requirements so that they will not grow outside the laboratory.
3. They are easily grown on minimal medium containing salts and an energy source such as glucose. They also grow on a rich nutrient broth with a doubling time of 20 min.

E. coli is a Gram-negative short rod. The genome consists of a large, compactly folded circle containing 4×10^6 base pairs. *E. coli* proteins are made on ribosomes in the cytoplasm and either remain there or are extruded into the periplasmic space, or even excreted into the medium by special mechanisms. Recombinant proteins often accumulate as insoluble inclusions in the cytoplasm.

E. coli on the head of a pin!

outer lipid bilayer

periplasmic space

inner lipid bilayer

CYTOPLASM

lipopolysaccharide

porin

lipoprotein

peptidoglycan

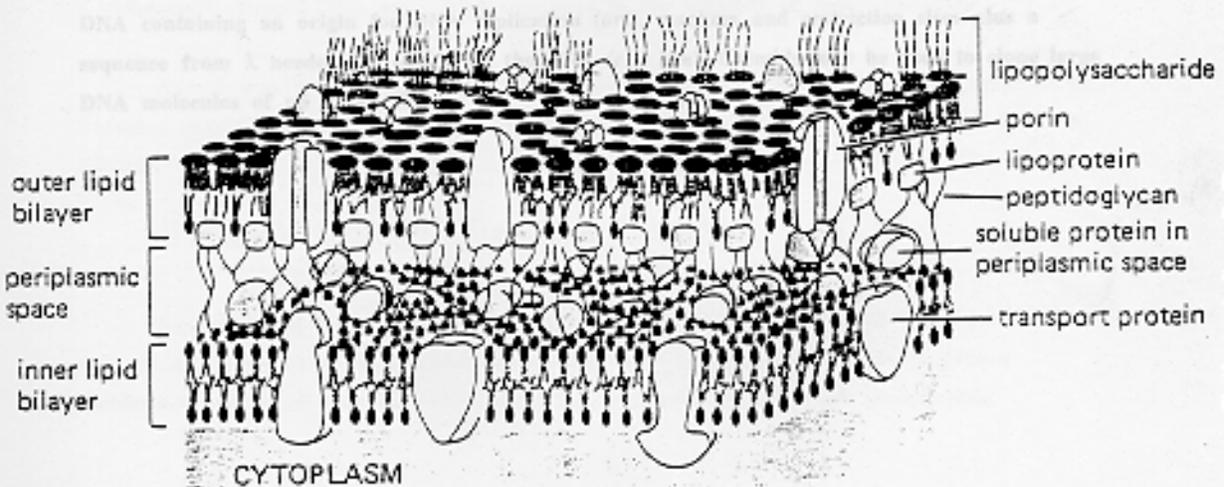soluble protein in periplasmic space

transport protein

Figure 19. E. coli is a commonly used host cell.

3.5 Cloning Vectors for *E. coli* are of three main types

1. Plasmids--circular DNA vectors that are self-replicating (separate from the host DNA). Plasmids are usually 5-40 kilobase pairs (Kbp) in size, and DNA fragments up to 15 Kbp can be inserted into the plasmids

2. Bacteriophages--include modified bacteriophage lambda, which has 48.5 Kbp. About one-third of the genome is not essential for DNA replication and may be replaced with up to 23 Kbp of inserted DNA. The total length of the DNA must be of the correct size to be packaged into virus particles within the infected cells

3. Cosmids--a combination of plasmid and bacteriophage. Small (5-7 Kbp) circular DNA containing an origin for DNA replication (*ori*), markers and restriction sites plus a sequence from lambda needed for packaging the DNA (*cos* site). Cosmids may be used to clone large DNA molecules of up to 45 Kbp.

3.6   Plasmid Cloning Vectors for *E. coli*

There are hundreds of different plasmids that today may be used for cloning experiments with *E. coli*. We will concentrate on describing the plasmid known as pBR322 which was developed by Bolivar and Boyer.

The plasmid known as pBR322 was developed by Bolivar and Boyer in 1977. This plasmid has 4,362 base pairs that includes a special sequence of DNA that is needed as an origin (the *ori* site) for DNA replication and that allows for efficient replication. Usually 10--20 copies of the circular DNA are produced in each bacterial cell that harbors the plasmid.

pBR322. pBR322 is a very commonly used cloning vector. It contains an amplifiable pMB1 replicator and genes encoding resistance to ampicillin and tetracycline. Insertion of DNA into a restriction site in either drug-resistance gene usually inactivates it and allows colonies bearing plasmids with such insertions to be identified by their inability to grow on medium with that antibiotic (Bolivar et al., 1977; sequence in Sutcliffe, 1978).

FIGURE 20. The pBR322 plasmid

Two genes of pBR322 confer resistance to antibiotics to any cell that contains the plasmid. *Amp*R confers resistance to amplicillin and *tet*R confers resistance to tetracycline to cells containing the plasmids. *Amp*R is a gene that codes for the periplasmic enzyme beta-lactamase that cleaves the ring structure found in amphicillin, which is a penicillin antibiotic. *Tet*R is a gene that codes for a protein that modifies the bacterial cell wall and prevents tetracycline from entering the cell.

Multiple restriction endonuclease sites are present where foreign DNA fragments may be inserted.

"Relaxed" plasmid DNA replication continues in the presence of chloramphenicol. An interesting feature of this plasmid is that "relaxed" plasmid DNA replication continues even in the presence of an inhibitor of protein synthesis such as chloramphenicol. This feature allows increased yields of plasmid/cell of up to 100-fold.

3.7   Transforming Cells with Plasmids

As you might imagine, it is not easy to force large DNA molecules into bacterial cells. *E. coli* is protected by two membranes, the outer membrane and an inner plasma membrane, and large, charged molecules do not ordinarily pass through phospholipid membranes. The organism also has protective enzymes; the restriction endonucleases we already have learned about and other nucleases that are ready to attack foreign DNA as it entered the cell. In 1974, Lederberg and Cohen developed a method to partially overcome these natural barriers.

Intact DNA does not ordinarily enter bacterial cells. First, treat *E coli* cells at 0° with a dilute solution of $CaCl_2$ to weaken the cell wall. Next, added DNA forms hydroxy-calcium-phosphate bonds and the complex adsorbs to the cells.   The calcium-phosphate complex not only adsorbs to cells, but the DNA is protected inside the complex.

Next, Raise the temperature to 37-40° to "heat shock" the cells and force DNA inside. During the heat shock, a few of the cells take up complete molecules of DNA. Usually about 1 in 1000 DNA molecules yield a transformant. This mean that 10^5-10^7 transformants may be obtained for each microgram of pBR322 DNA.   Other methods may be used. For example, DNA may be driven into cells by an electric pulse which creates a temporary pore in the membranes, a technique known as electroporation.

Special host strains of *E. coli* are used that lack the nucleases that would otherwise attack and degrade the inserted plasmid DNA (see host strain section)

3.8   Selecting Transformed Cells

Remember that only a small percentage of the bacterial cells may have taken up the plasmid DNA and not all of the plasmids will have the gene we wished to insert. Some transformed cells may contain original plasmid that closed upon itself without

incorporating any of the new DNA--and not all of the DNA that might be inserted in the plasmid may be intact and functional. We will see later that when we are making a gene library, the one gene that interests us may be only one of thousands present in the DNA sample that we have cloned into the plasmids. We clearly need very sensitive and accurate methods for screening for just the transformed cells that we want.

In even the simplest cases, the transformation mixture will contain large numbers of untransformed cells, usually some cells that are transformed with only the original plasmid, and a few cells that are transformed with the complete vector containing the gene of interest.   There are several procedures that may be used to create very efficient selection techniques:

Include in the plasmid a gene that is required for growth of the transformed host cell--usually a gene for resistance to an antibiotic. If the plasmid brings in a gene for antibiotic resistance, then normal host cells will die if the antibiotic is added to the culture and only transformed cells will survive.

Include in the plasmid a gene that will yield colored colonies when a suitable substrate is added. Only those colonies that are transformed by plasmids that have a new gene for an enzyme that will convert a substrate to a colored produce will be colored. Normal cell colonies will appear white and colorless.

A radioactive nucleic acid probe (such as ^{32}P-mRNA) may be used to detect those colonies that have the new DNA gene. For this procedure, we must isolate the mRNA of the gene or chemically synthesize in radioactive form a small section of the DNA sequence which we know to be a part of the gene. Of course, we do not always have this information.

Antibodies (often labeled with radioactivity) that react with the protein product of the gene may be used to determine which colonies contain the new gene. For this procedure to work, the transformed cell must make mRNA copies of the new gene and then translate these messages into new protein products. For example, *E. coli* containing the complete gene for proinsulin will make large quantities of human insulin. These cells could be detected using anti-insulin antibody. The general techniques for these procedures will be described under "replica plating" and other refinements are described in the Nicholl text.

3.9   Replica Plating for Colony Screening

We will now examine how to select for transformed cells that harbor the plasmid of interest if that plasmid containes genes for antibiotic resistance. Think back to the plasmid pBR322. The original plasmid contains genes for resistance to amphicillin and tetracycline. Cells transformed with the original plasmid will be resistant to both of these antibiotics. If we cut the plasmid with *Bam*H1, and insert our new gene at this site we will interrupt the gene for tetracycline resistance. If we transform cells with this new construct, the cells will not be resistant to tetracycline, but they will still be resistant to amphicillin.
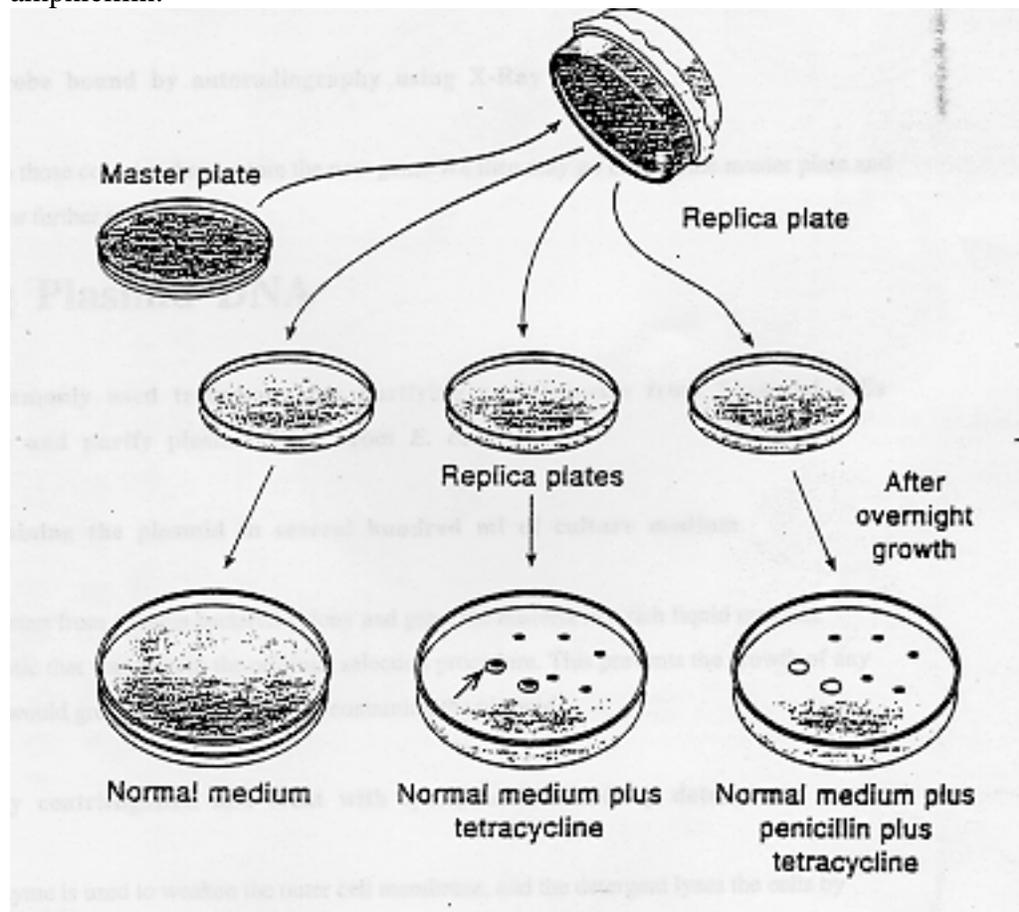

FIGURE 21. Replica platting

Spread about 10^7 bacteria from the transformation reaction mixture on the surface of a
   plate containing nutrient agar . Incubate the plates overnight at 37° C to allow the
   cells to grow as colonies.
Press a piece of velvet against the surface to pick up some of the cells.

Transfer to new plates containing normal media or medium containing antibiotics and grow overnight.

> *Expected Result: The colonies of cells pressed onto normal medium will grow vigorusly as before and they represent the total number of cells that we plated. In our example, cells plated on medium containing amphicillin will not grow unless they were transformed with either the original plasmid or the plasmid containing the inserted gene. The number of colonies tells us about the efficiency of transformation.*

Select colonies of cells containing the desired transformation. Cells growning on medium containing both amphicillin and tetracycline must be those cells that were only transformed with the original plasmid and not the plasmid containing the gene we want. We now must go back to the amphicillin plate and select and save those few colonies that were unable to grow in the presence of tetracyline. These are the cells that we will save and grow and which contain the desired plasmid.

3.10   Nucleic Acid Probes

A similar replica plating procedure may be used if we have a radioactive nucleic acid probe that will aneal with our new gene. The procedure is as follows:

Original dish saved for reference

Disc now has phage DNA at same location as dish

Phage plaques on a lawn of cells

Overlay a nitrocellulose disc

Disc is treated so that DNA unwinds in place

A single strand of radioactive DNA or RNA with a sequence complementary to the DNA (a probe) is added to the solution and allowed to anneal

Radioactive probe

Probe

Remove unbound probe, expose to film

Only a colony that has annealed to the probe will darken the film
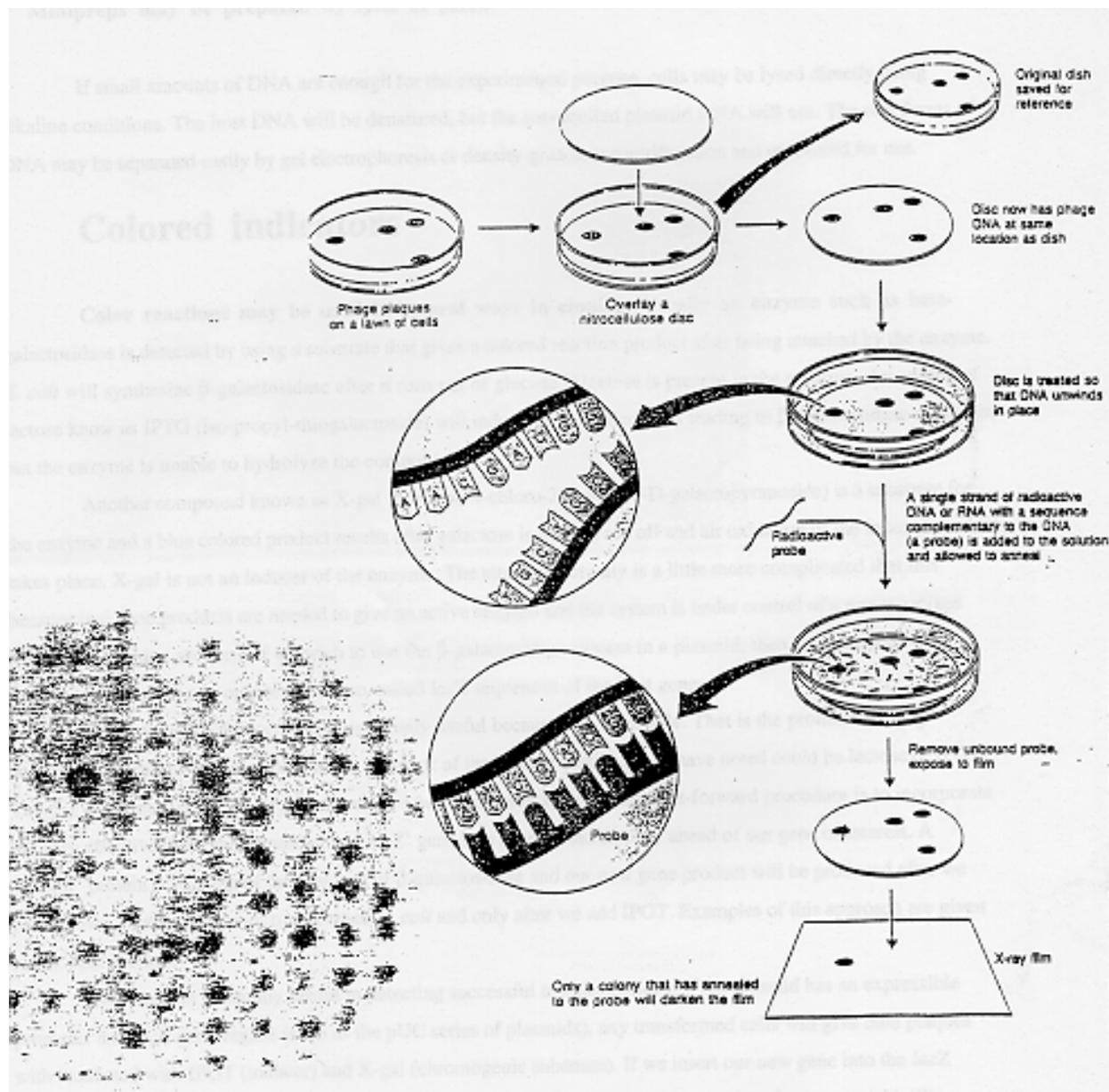
X-ray film

FIGURE 22. Screening with a radioactive probe

Grow the cells on the master plate as before, but substitute a nitrocellulose filter for the velvet cloth to blot the colonies.

Press a nitrocellulose filter onto the colonies; some cells will adhere. When the attached cells are treated with alkali, the cells will break open and the DNA will attach to the nitrocellulose filter. A radioactive probe derived from the new gene may then be incubated with the filters.

Remove the filter with attached cells and treat with alkali.

Incubate with a radioactive nucleic acid probe (RNA or DNA) under anealling conditions.

Detect where the probe bound by autoradiography using X-Ray film.

The probe binds to those colonies that contain the new gene. We then may go back to the master plate and select the desired colonies for further growth

3.11   Purifying Plasmid DNA

Any of the commonly used techniques for purifying nucleic acids from bacterial cells may be used to isolate and purify plasmid DNA from *E. coli*.

Grow bacteria containing the plasmid in several hundred ml of culture medium. It is customary to start from a single bacterial colony and grow the bacteria in a rich liquid medium containing the same antibiotic that was used in the originial selection procedure. This prevents the growth of any non-transformed cells that would grow faster than the cells containing the plasmid.

Harvest the cells by centrifugation and treat with lysozyme followed by detergent. The enzyme lysozyme is used to weaken the outer cell membrane, and the detergent lyses the cells by interaction with the plasma membrane.

Isolate the DNA by gel electrophoresis or density-gradient centrifugation

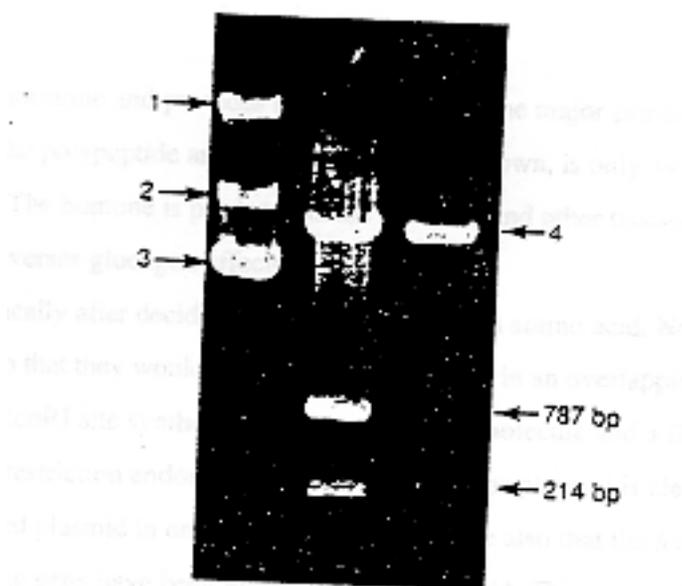*Minipreps may be prepared by lysis at pH12*

FIGURE 23. Separation of DNA fragments by gel electrophoresis

If small amounts of DNA are enough for the experimental purpose, cells may be lysed directly using alkaline conditions. The host DNA will be denatured, but the supercoiled plasmid DNA will not. The two forms of DNA may be separated easily by gel electrophoresis or density-graadient centrifugation and recovered for use.

3.12   Colored indicators

Color reactions may be used in several ways in cloning. Usually an enzyme such as beta-galactosidase is detected by using a substrate that gives a colored reaction product after being attacked by the enzyme. *E. coli* will synthesize b-galactosidase after it runs out of glucose if lactose is present in the medium. An analog of lactose know as IPTG (iso-propyl-thiogalactoside) will induce gene transcription leading to betta-galactosidase synthesis, but the enzyme is unable to hydrolyze the compound.

Another compound known as X-gal (5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside) is a substrate for the enzyme and a blue colored product results after galactose is hydrolyzed off and air oxidation of the indoxyl product takes place. X-gal is not an inducer of the enzyme. The situation actually is a little more complicated that this because two gene products are needed to give an active enzyme and the system is under control of a repressor (see Molecular Biology section). If we wish to use the beta-galactosidase system in a plasmid, then the host cells must contain mutations in one or more of the so-called lacZ sequences of the host genes.
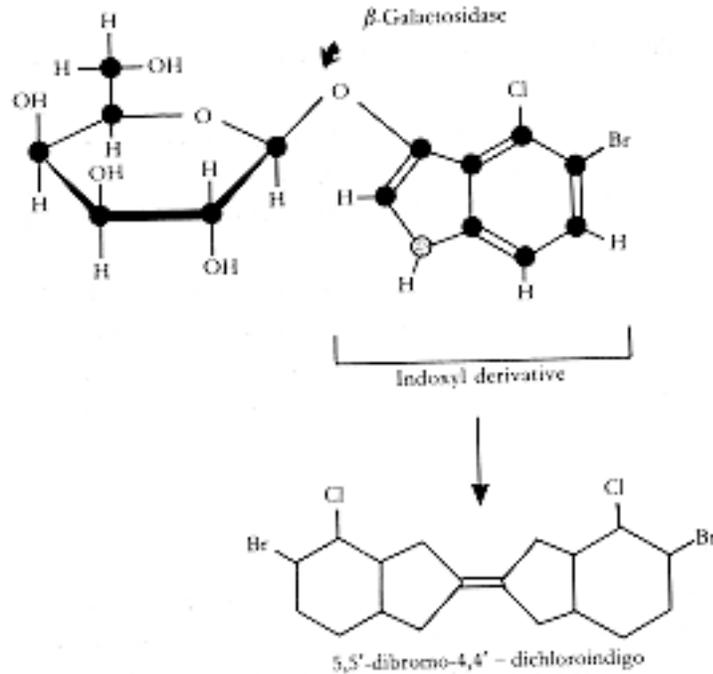
FIGURE 24. Detection with colored indicators

The beta-galactosidase system is especially useful because it is inducable. That is the promoter is only activated (and the gene transcribed) in the presence of the inducer (which as we have noted could be lactose or IPGT). Thus, if we want to produce large amounts of a new protein in *E. coli*, a straight-forward procedure is to incorporate the very effective inducable promoter and *lac*Z' gene into the plasmid directly ahead of our gene of interest. A "fusion" protein consisting of the first part of b-galactosidase and our new gene product will be produced after we have grown a dense culture of transformed *E. coli* and only after we add IPGT. Examples of this approach are given in the next section.

The reaction is equally useful in detecting successful transformations. If the plasmid has an expressible promoter and *lac*Z gene region (such as the pUC series of plasmids), any transformed cells will give blue plaques with incubated with IPGT (inducer) and X-gal (chromogenic substrate). If we insert our new gene into the *lac*Z sequence (interrupting it) then the plaques will be colorless. The pUC plasmid series also have an amphicillin resistance gene for additional selection and a short sequence containing multiple restriction sites to make them highly useful vectors.

3.13 Let's Clone!

We now may use the various procedures that we have learned to examine making a plasmid vector, transforming cells and producing human polypeptides in *E. coli*.

Somatostatin (Itakura, et al., Milestones readings, p. 84)

Somatostatin is a very small peptide hormone and provides a nice example of the major principles to be considered when cloning a gene (see figure 25). The polypeptide amino acid sequence is known, is only 14 amino acids long, and contains one internal disulfide bond. The homone is produced in the pancreas and other tissues and inhibits the secretion of insulin and glucagon. It also reverses glucagon effects in the liver.
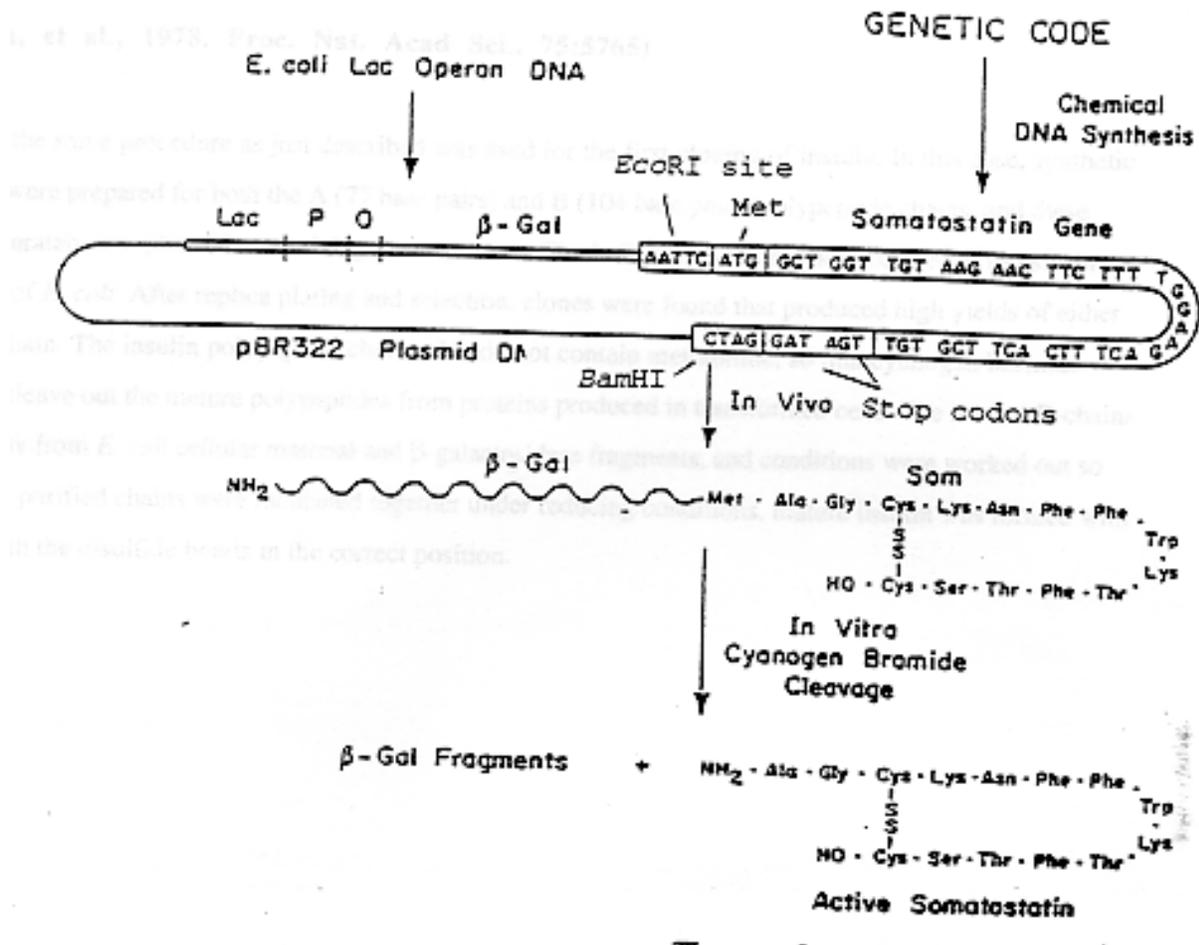

FIGURE 25. Cloning somatostatin

The DNA gene was synthesized chemically after deciding likely codons for each amino acid. Notice that deoxynucleotides A through H were prepared so that they would anneal with one another in an overlapping sequence that could be sealed by ligase. Also notice

the *Eco*RI site synthesized at one end of the molecule and a *Bam*HI at the other. When we open pBR322 using these two restriction endonucleases, a segment of the plasmid is cleaved out and discarded and our new gene will enter the opened plasmid in only one orientation. Notice also that the *lac* promoter and control region and part of the b-galactosidase gene have been inserted into the plasmid. This means that the protein will not be produced until an inducer has been added to the medium and that the protein product will consist of part of the b-galactosidase protein linked to the polypeptide that we want.
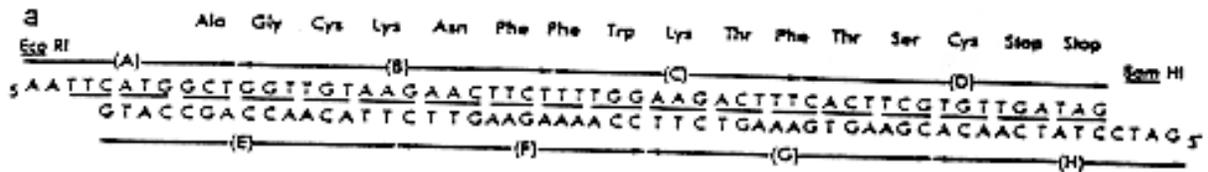

FIGURE 26. Strategy for the chemical synthesis of the gene for stmatostatin

Another trick that was used was to insert the codon for methionine as the first amino acid in front of the code for somatostatin. Somatostatin itself does not contain methionine. Cyanogen bromide may be used to chemically break the polypeptide bond at all methionine residues, thus liberating the polypeptide from the beta-galactosidase-somatostatin protein product. Two stop signals were inserted into the gene at the end of the somatostatin gene sequence to ensure that protein synthesis would not continue beyond this point. Recombinant *E. coli* colonies were selected as usual using antibiotic resistance markers, and clones yielding the highest amounts of somatosatin were further screened with a radioimmune assay. The protein product that contained somatostatin was found to reach over 3% of the total cellular protein in some of the clones. The peptide hormone was successfully purified from the transformed cell.

Insulin (Crea, et al., 1978. Proc. Nat. Acad Sci., 75:5765)

Almost the same procedure as just described was used for the first cloning of insulin. In this case, synthetic DNA sequences were prepared for both the A (77 base pairs) and B (104 base pairs) polypeptide chains, and these were inserted separately into plasmids containing *lac* sequences. The individual vectors could be used to transform separate cultures of *E. coli*. After replica plating and selection, clones were found that produced high yields of either the A- or the B-chain. The insulin polypeptide chains also do not contain methionine, so that cyanogen bromide could be used to cleave out the mature polypeptides from proteins produced in transformed cells. The A- and B-chains were purified away from *E. coli* cellular material and beta-galactosidase fragments, and conditions were worked out so that when the two purified chains were incubated together under reducing conditions, mature insulin was formed with high yield and with the disulfide bonds in the correct position.

GENETIC CODE

E. coli Lac Operon DNA

Chemical
DNA Synthesis

Insulin A or B Gene

Lac   P   O   β-Gal

| AATTC | ATG |

| CTAG | GAT AGT |

pBR322   Plasmid DNA

In Vivo

Insulin A or B

β-Gal

NH2 ⌇⌇⌇⌇⌇⌇⌇⌇⌇⌇⌇⌇ Met

In Vitro
Cyanogen Bromide
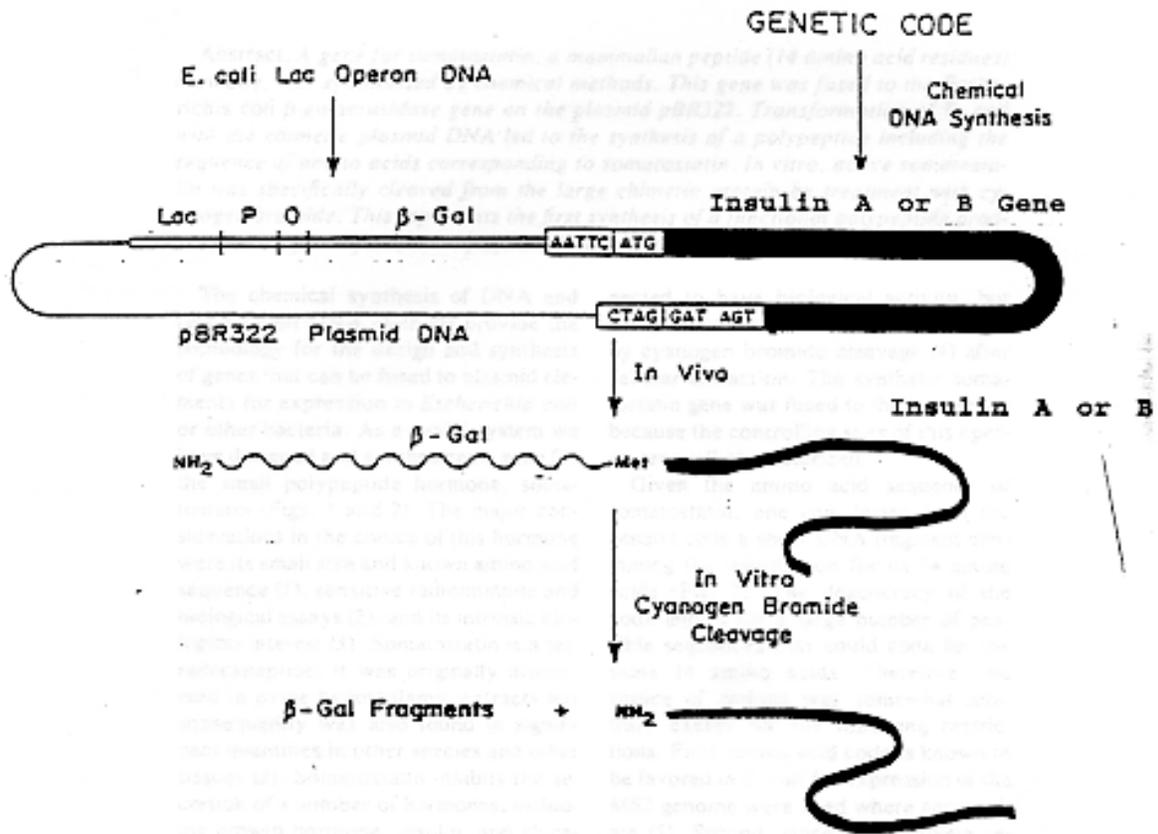Cleavage

β-Gal Fragments   +   NH2

FIGURE 27. Cloning insulin

Chapter 4 More Advanced Cloning Techniques

4.1 Bacteriophage Lambda Vectors

Interestingly enough, even bacteria are plagued by viruses! You may know that when viruses infect cells they manage to transfer their nucleic acid genetic material to the inside of the host cell. There it replicates to produce up to several hundred new viruses. If we could replace one or more of the viral genes with a gene we are interested in amplifying, then we would have another type of vector in addition to the plasmids alreaady discussed. Furthermore, the host cell may obligingly package up the replicated DNA containing our gene into new virus particles that are liberated from the cell.

The lambda (l) genome is 48.5 kbases in length and contains about 46 genes. The bacteriophage lambda is frequently used as a vector. It has a head structure that contains the viral DNA and a long tail which is used in attachment of the virus to *E. coli*. The original virus has a set of genes that allows the viral DNA to insert itself into the bacterial chromosome. These genes may be cut out and discarded to provide a location for our new gene to be inserted.
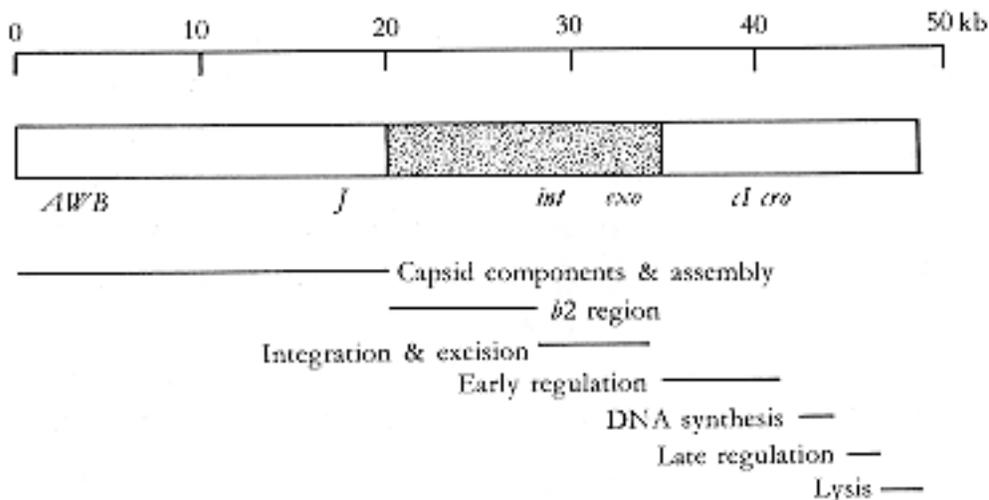

FIGURE 28. The l genome

Following attachment of the virus to *E. coli*, the DNA in the head structure travels down the hollow tail and enters the host cell. The linear, viral DNA has naturally "sticky ends" consisting of a complementary sequence of 12 bases on each end of the molecule. The

sticky ends of the linear DNA (the *cos* site) associate to form circular DNA. The virus codes for its own DNA polymerase which leads to the synthesis of several hundred copies of viral DNA.
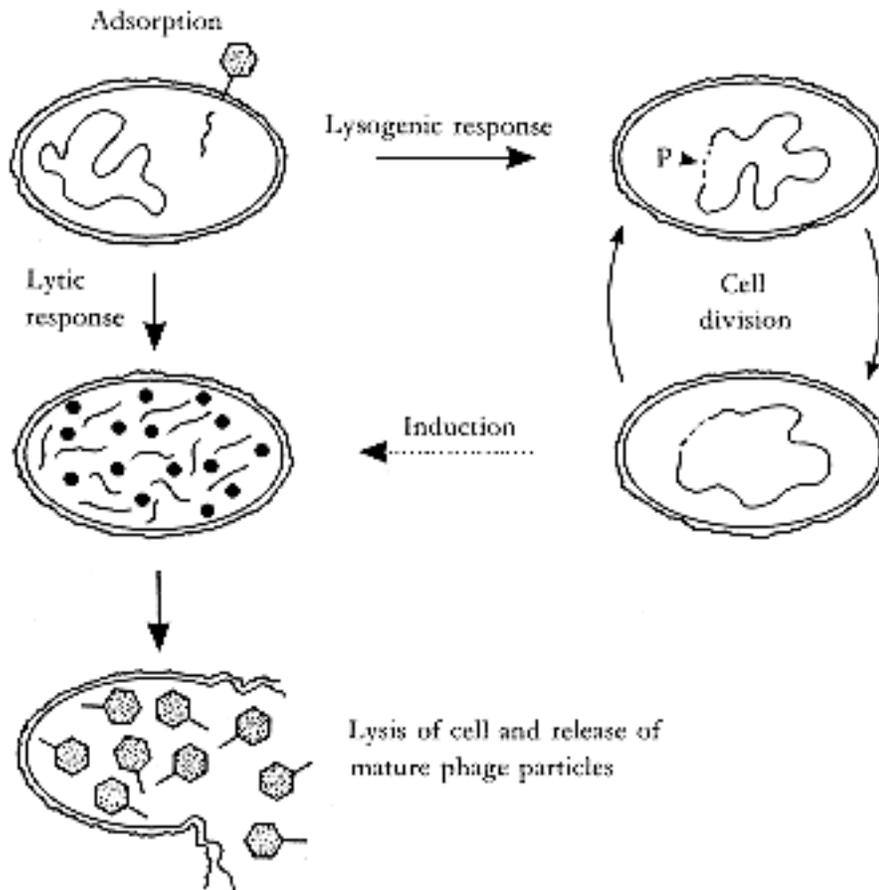
FIGURE 29. The phage life cycle

The DNA is synthesized in very long repeated units which are cleaved at the *cos* sites to give unit length viral DNA for insertion into new particles. The bacteriophage DNA also codes for all of the proteins needed to put together new virus particles which then are assembled inside the host cell.

The new DNA is trapped inside the newly-formed viral coat proteins and the host cell bursts, liberating the new particles.   A most remarkable property of the proteins and the viral DNA is that complete particles will assemble in the test tube when all of the reactants are mixed together.

The infection spreads to adjacent cells and the cycle is repeated . If the infection is carried out with *E. coli* cells that are spread over an agar surface, the infected areas will appear as clear dots (or plaques) surrounded by live cells. Bacteriophage may be recovered from the cear areas. Each plaque originates from a single infected cell.
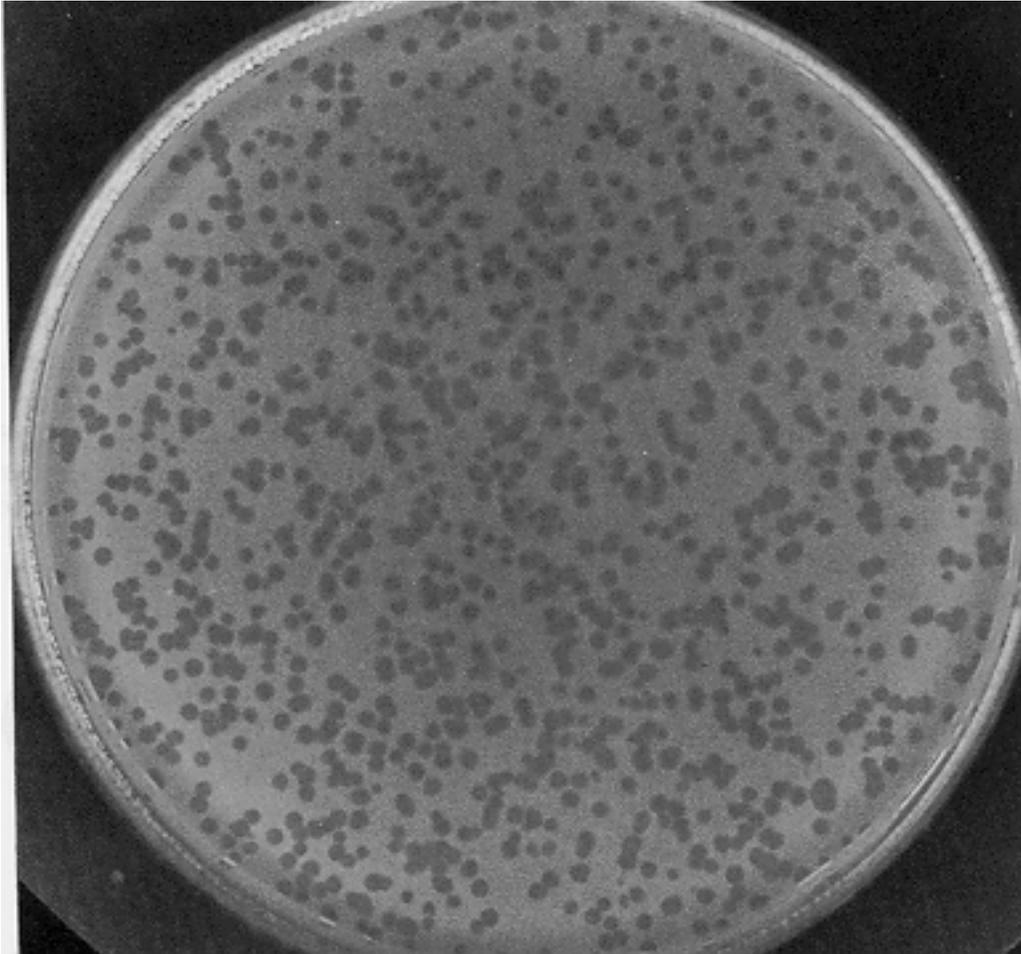
FIGURE 30. Phage l plaques on a lawn of bacteria

Virus particles are separated from cell debris and the DNA isolated. The process may be scaled up by infecting cells growing in liquid culture. Over 100 different lambda vectors have been prepared for use in cloning. The DNA of the cloning vector must be the correct size for it to be packaged into virus particles (between 38 and 51kb).

The vector lgt10 (43.8 kbases) has an *EcoRI* site inserted within the cI gene (lambda repressor) and so could accept a DNA fragment of 7.6 kbases before becoming too large to be packaged into virus particles.

FIGURE 31. The vectors lgt10 and Charon 16A

Other lambda vectors can incorporate DNA fragments of up to 22 kbases (*e.g.*, the Charon series).

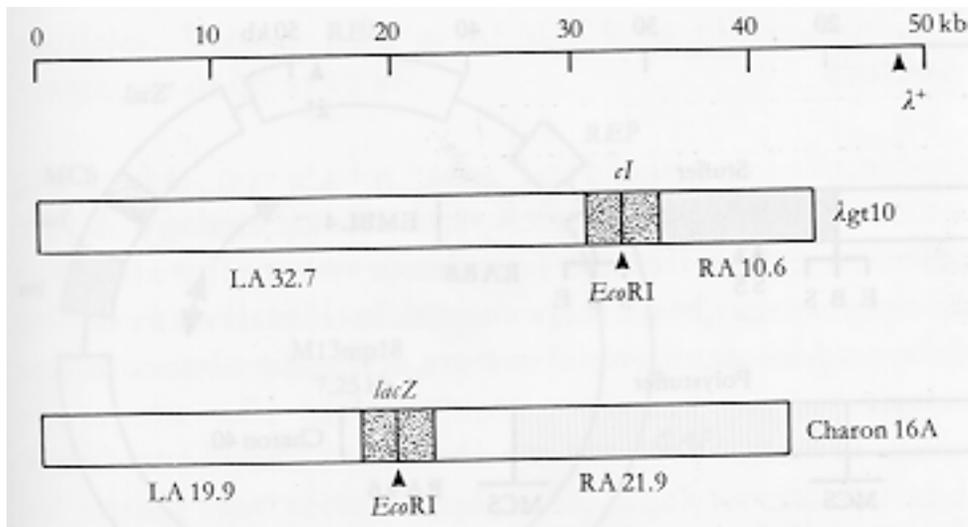To use lgt10, we would cut the vector DNA and the DNA to be inserted with *EcoRI*, incubate the two together, seal with ligase and incorporate the recombinant molecule into particles in the test tube:

Open lgt10 with EcoRI and insert double stranded DNA also cleaved with EcoRI .
Seal with ligase.
Package the DNA into particles with a mix of head and tail precursors.
Infect E. coli with the reconstituted virus particles containing the new gene and recover
  the modified bacteriophage from infected cells as described above.

4.2   Cosmid vectors

All of the lambda DNA sequence may be removed, except for the *cos* sites to allow larger pieces of DNA to be inserted. A site for the origin of replication (*ori*) and perhaps an antibiotic resistance gene also are retained:

Insert DNA from 33 to 47 kbases in length into the cos site fragments using an
appropriate restriction endonuclease.
Seal with ligase.

Package the DNA into particles with the in vitro packaging mixture.

Since no viral genes are included in the vector, they replicate as plasmids in *E coli*. The amplified DNA or its gene product may be isolated from the infected cells.
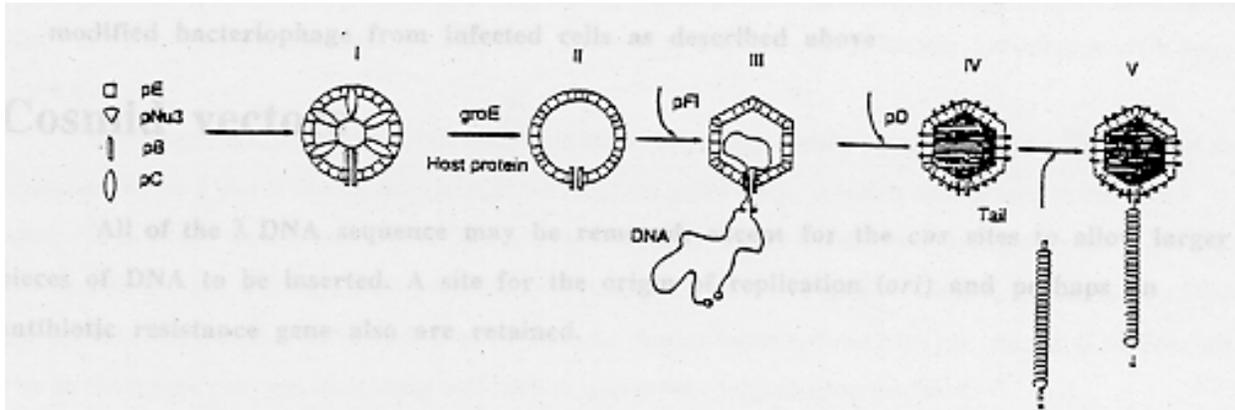
FIGURE 32. Assembly of l phage in the test tube

Infect cells, allow the cosmid to replicate, and isolate DNA. This method allows insertion of larger pieces of DNA and also has the advantage of efficient infection of cells by the bacteriophage particles. Recall that transformation of cells by plasmids alone is relatively inefficient (perhaps 1 in 1000 molecules) and is usually limited to inserted DNA fragments of up to 15 kb.
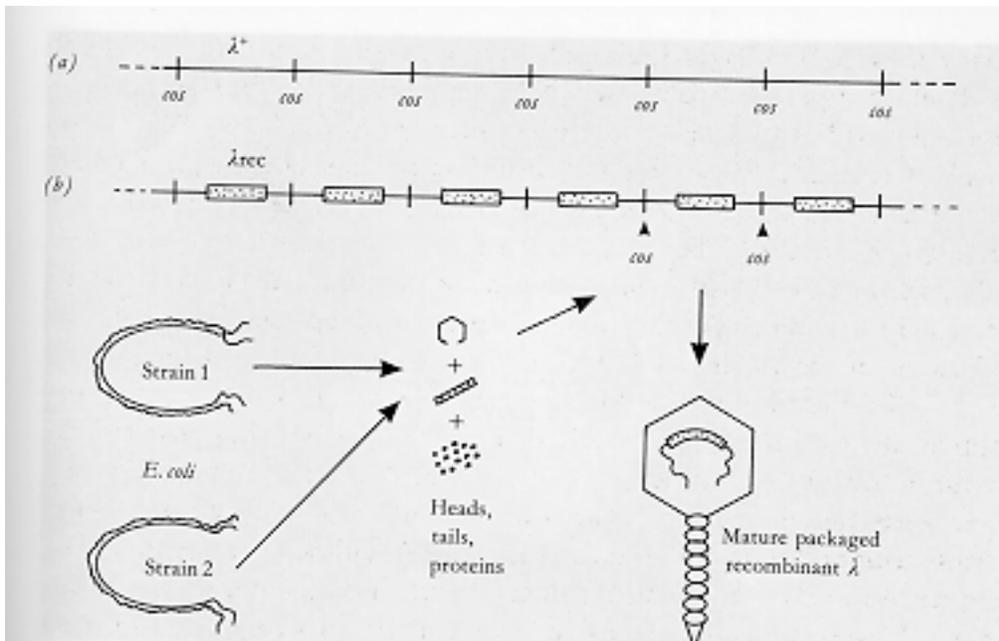
FIGURE 33. Incorporation of DNA into l--derrived particles

4.3   Finding genes: DNA Libraries

We now have all the tools at hand that we need to actually clone a gene. Each new cloning problem is different from from the previous one and our experimental approach must be planned carefully. The first major decision that must be made is whether to clone from the genomic DNA or to start with the messenger RNA that has the final code for a protein of interest.

Why do the two approaches differ? You may have to review this part of the molecular biology section before proceeding.   Recall that the DNA of both bacteria and higher organisms has information at the 5' end of a gene that is related to control of that gene-- turning it on or off. You may or may not want to include this information in your cloned gene.

Most important, the typical DNA gene from higher organisms usually contains introns and exons and the sequences that are a part of introns must be cleaved out of the mRNA copy before it may be used to code for a protein. Thus, if you want to read a gene from a higher organism (such a human) in a bacterium (such as *E. coli)*, to produce a protein you will want to start with the sequences found in mature mRNA. We then need to make a DNA copy (called cDNA) of the mRNA so that we may use the cloning techniques we have just learned. If the host cell is to be an eukaryotic one, you most likely will want to isolate the complete genomic DNA.

cDNA libraries

In our example, we will devise an approach to isolate a set of mRNA molecules from cells that are actively expressing the gene of interest, make DNA copies of the RNA and clone these DNA molecules into *E. coli*.
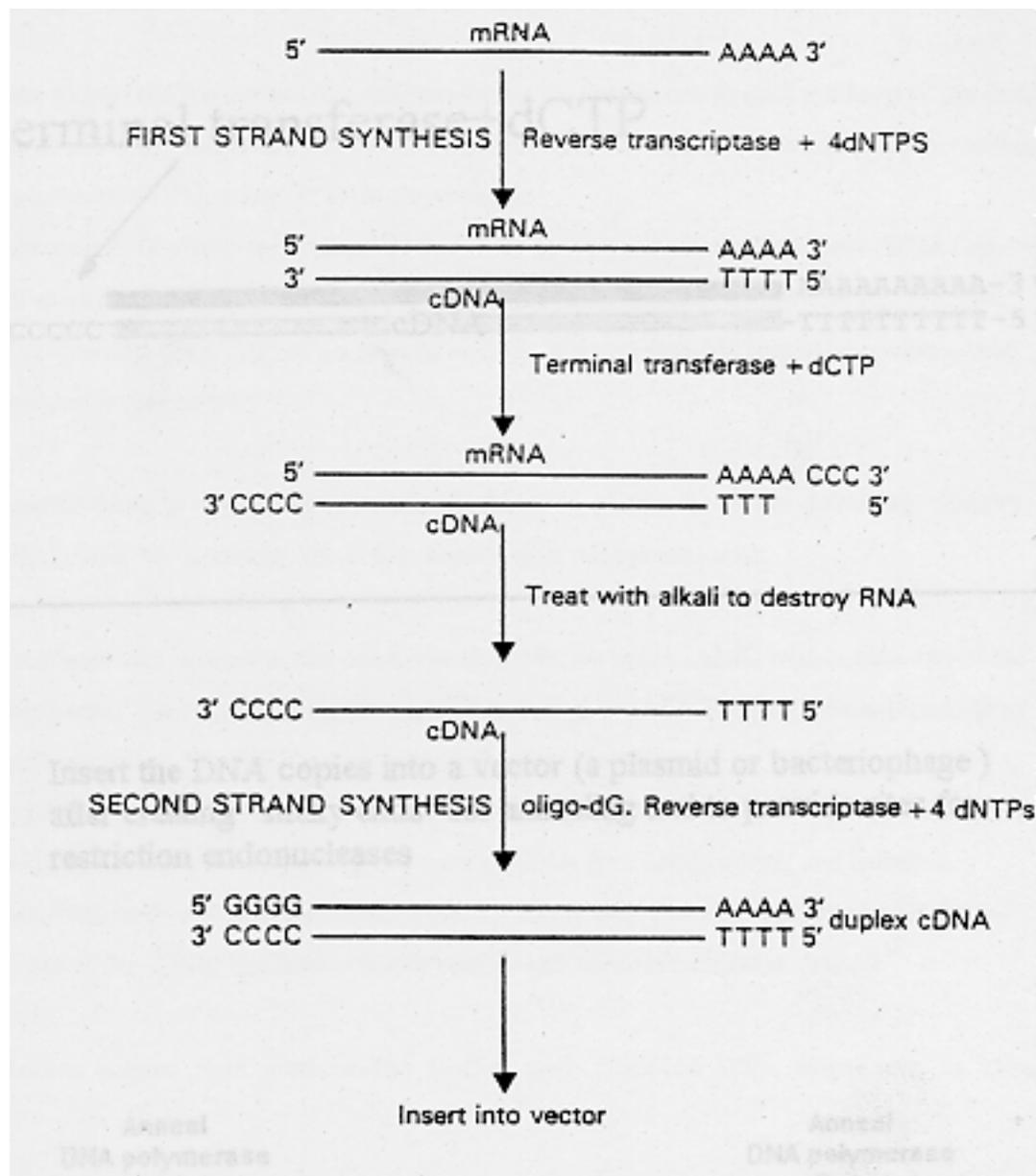
FIGURE 34. Constructing a cDNA library

Isolate mRNA from cells that are actively expressing the gene of interest. Most cells express hundreds of genes at the same time, but often we can take advantage of some specialized tissue that is producing much more of the protein we are interested in than do other cells or tissues. It then follows that these cells will be making increased amounts of the specific mRNA that we want. Examples include liver cells producing albumin, lymphocytes making antibodies, and our favorite: beta-cells of the pancreas producing insulin (the only cells of the body to do so). If we isolate mRNA from these special

tissues, we increase our chances of being able to capture the correct mRNA in our cDNA library.

Use reverse transcriptase to make DNA copies of the RNA. Reverse transcriptase is an enzyme that is found in some viruses, where it makes DNA copies of RNA viral genes. We also can make use of another trick by remembering that most eukaryotic mRNAs have a string of adenine nucleotides at their 3' end so that we can isolate the whole mRNA on a column containing poly dT bound to beads. The poly A tail will aneal to the poly dT and be captured on the column. The column may be washed to remove other nucleic acids and protein and then the mRNA may be eluted. Remember also that polymerases need a primer; now we can use poly dT for this purpose. We might also save only the mRNAs that are of the approximate size that we suspect would be needed to code for our protein.

Use terminal transferase + dCTP to add a poly dC tail which will serve as a primer for second-strand synthesis. We need a primer to start the second strand synthesis, just as we needed one to start synthesis of the first strand. A convenient way of providing this is use an enzyme known as terminal transferase to add cytidine residues to the 3' end of the newly synthesized DNA using dCTP as the substrate.

If we separate the strands in alkali, the original RNA will be hydrolyzed and only our new cDNA copy will be left. We now may add an oligo dG primer and synthesize the second strand using DNA polymerase. We now have created a full length double-stranded DNA copy of the original mRNA. We may also add restriction endonuclease sites to the ends of the molecules (see below)

Insert the DNA copies into a vector (a plasmid or bacteriophage l) after creating "sticky ends" for annealing and to provide sites for restriction endonucleases. One of several methods may be used at this point. For example, we could add dC tails to each end of the cDNA using terminal transferase. Then we might cut the plasmid with *Pst*I and add dG tails to the exposed sticky ends, again using terminal transferase. The DNA now inserts readily into the vector.
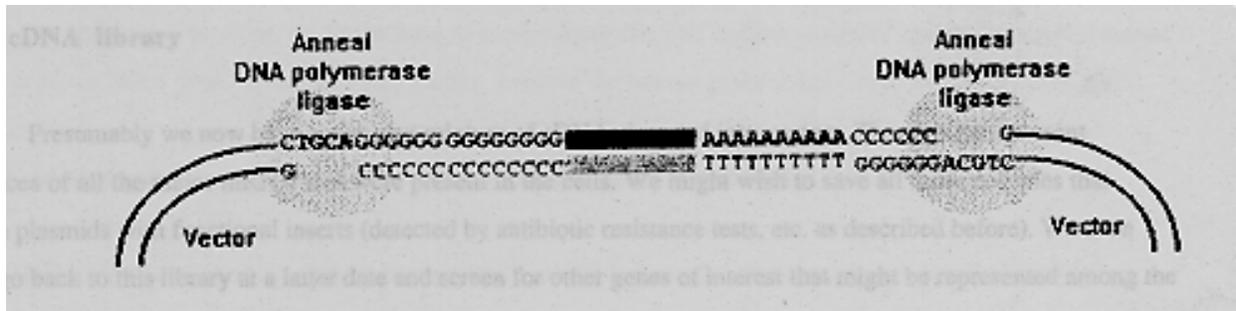
FIGURE 36. Joining and extending

The gaps may be filled in with DNA polymerase after annealing. Note that the *Pst*I site is preserved in the recombinant DNA and will provide an easy way to clip out the new DNA after amplification and isolation. Another method would be to use blunt end ligation (which occurs with low efficiency) to ligate "linker" oligonucleotides to each end of the cDNA. the linkers would contain one or more restriction sites.

Isolate many different clones that contain the vector with inserted DNA sequences to create a cDNA library. Presumably we now have a complex mixture of cDNAs inserted into vectors. The cDNAs represent sequences of all the major mRNA that were present in the cells. We might wish to save all those colonies that contain plasmids with functional inserts (detected by antibiotic resistance tests, etc. as described before). We even could go back to this library at a latter date and screen for other genes of interest that might be represented among the cDNAs.

Screen for the desired clone using any of the techniques discussed in the previous sections. Antibodies specific for the new protein or radioactive nucleic acid probes are particularly useful. We now need to screen our library more specifically for the gene of interest. We need quite a specific test in order to select the correctly transformed cell colonies. If we are looking for the insulin gene, we might test with antibodies for insulin (radioactive antibodies will bind to those colonies producing insulin). Since we know the amino acid sequence for insulin, we could synthesize a radioactive nucleic acid probe that will anneal to the cDNA and perform this test (as described previously).
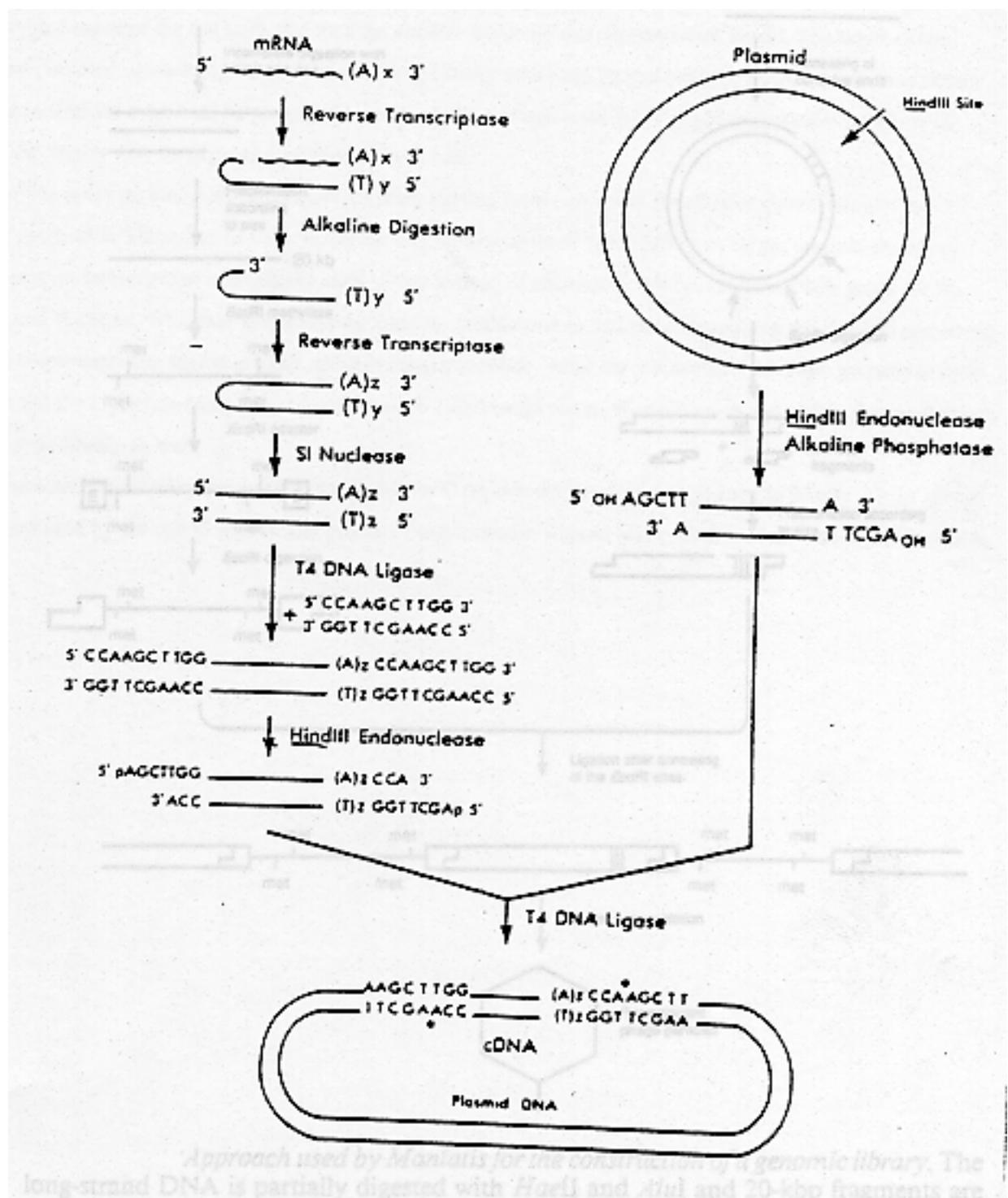
FIGURE 37. A library from mRNA

Genomic Libraries

The goal of creating genomic libraries is to fragment the entire genome of an organism into a series of overlapping fragments and incorporate them all into suitable vectors. The library (or clone bank) may be stored as transformed bacteria (in the case of plasmid vectors) or infected bacteria or bacteriophages (in the case of lambda vectors).

Estimate the number of clones needed to produce a library that will include all of genes of an organism. This estimated number will depend heavily on the size of the genome and will be much lower for the gene of a bacterium than for a human. $N = \ln(1-P)/\ln(1-a/b)$ helps to estimate the number of clones needed where P is the probability that the desired sequence will be represented in the library (usually 0.95-0.99), *a* is the average size of the DNA fragments, and *b* is the size of the genome . For the human genome, we might have to screen about one half million clones of average size of 20 kbases pairs in order to find a gene present as a single copy, because the human genome has a total of 3×10^9 base pairs!

Cut genomic DNA randomly into fragments suitable for cloning. DNA may be sheared (simply be passing DNA solutions through a fine hypodermic needle) but no restriction sites will be generated. We noted that restriction endonucleses that recognize six base pairs give fragments of about 4096 base pairs which is on the small side if we hope to capture complete genes. Also, we know nothing about the actual distribution of restriction sites (such as for *EcoRI*) within a given gene. For these reasons, the usual procedure is to use a restriction enzyme that cuts frequently, on average, (such as *Sau3A*; once every 256 bases) but only allow a partial digestion of the DNA. Under these conditions, cutting is nearly random and restriction endonuclease sticky ends are produced. Either the time of the reaction or the enzyme concentration may be adjusted to obtain the desired size range of fragments.
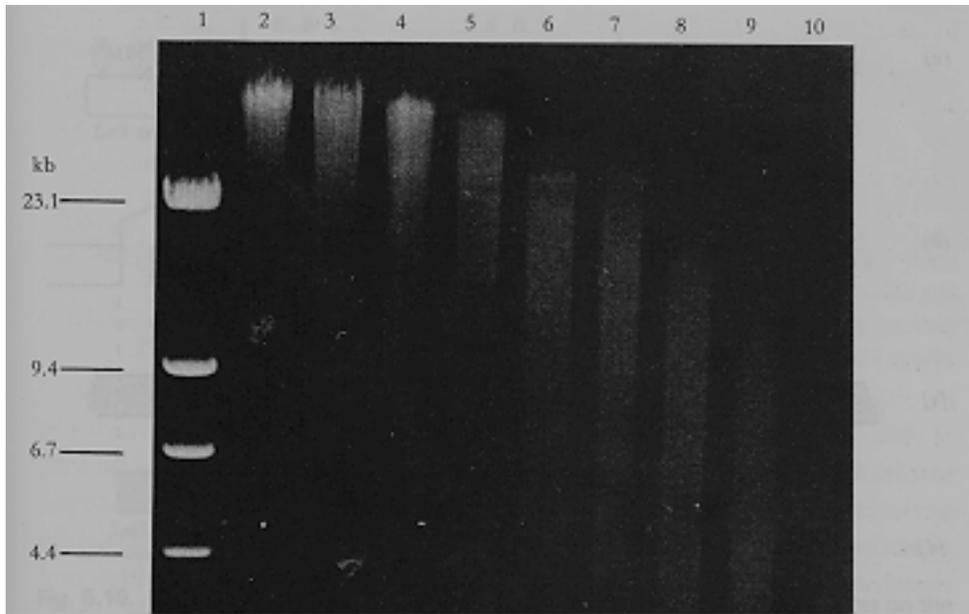
FIGURE 38. Selecting sizes of nucleic acid to clone. Genomic DNA was cleaved with increasing amounts of Sau3A giving progressively smaller fragments. The desired size may be isolated from the gel and used for cloning.

DNA of suitable size for cloning may be purified by gel electrophoresis.
Bacteriophage lambda or cosmids usually are the vectors of choice for DNA libraries because 20 Kb fragments and above can be inserted and cloned.

The vector is opened with a suitable restriction endonuclease, mixed with the DNA fragments, annealed and ligated. If *Sau3A* was used in the fragmentation of DNA, the l vector may be opened with *BamHI* to produce the correct cohesive ends. If lambda is used, the recombinant molecules may be incorporated into particles *in vitro* and the library stored in this way.

Bacteriophage lambda libraries may be amplified by one passage through *E. coli*, stored and used to screen for many different genes. Indeed, libraries based on lambda are available commercially and represent an easy starting point for screening for human genes.

Specific screening techniques (such as antibodies or nucleic acid probes) are used to select transformed clones of interest.

4.4   More Cloning

Several variations on the techniques just presented will illustrate practical cloning solutions (see FIG. 37). In the first, a cDNA library is constructed. Instead of adding a tail of dC residues to provide a binding site for second strand primer, the investigators relied on the ability of reverse transcriptase to make hairpin turn at the 3' end of the new strand. This short hairpin now provides a primer to complete the second strand. The hairpin may be cleaved by S1 nuclease that makes cuts in single stranded regions, such as the extreme end of the hairpin would be. Another feature of this cloning is the attachment of short adapters molecules to each end of the new double stranded DNA using DNA ligase (blunt-ended ligation). These adapters contain one or more restriction sites such as the one shown for *Hin*DIII. The plasmid also was opened with *Hin*DIII for insertion of the DNA.

The next example (FIG. 39) illustrates the construction of a genomic library. The genomic DNA was digested with limiting amounts of *Hae*III and *Alu*I to give DNA fragments in the 10--50 Kbp range. The DNA was fractionated on agarose gels and fragments of about 20 Kbp were selected for the cloning. The fragments were methylated at sites that would protect against *Eco*RI action (can you explain why?), and then an adapter was added to each end with ligase. The completed molecule was digested with *Eco*RI to create sticky ended sites. A Charon 4A was selected and annealed to bind together the *cos* sites, and then the circular molecule was digested with *Eco*RI. The large, central fragment was isolated (containing the *cos* site and *Eco*RI sticky ends) and ligated with the DNA fragments to obtain long, linear molecules that could be packaged *in vitro* into new phage particles. The phage particles constitute the library, or the library may be expanded by growth in *E. coli*.

High-molecular-weight
eukaryotic DNA

Charon 4A

50 kbp

Incomplete digestion with
HaeIII and AluI

Annealing of
cohesive ends

Fractionation
according
to size

20 kb

EcoRI methylase

EcoRI digestion

met   met

EcoRI adaptor

met   met

E ┤   ├ E

EcoRI digestion

met   met

met   met

Central
fragments

+

Fractionation according
to size

III

III

Ligation after annealing
of the EcoRI sites

met   met   met   met

met   met   met   met

in vitro encapsidation

Recombinant
phage particles
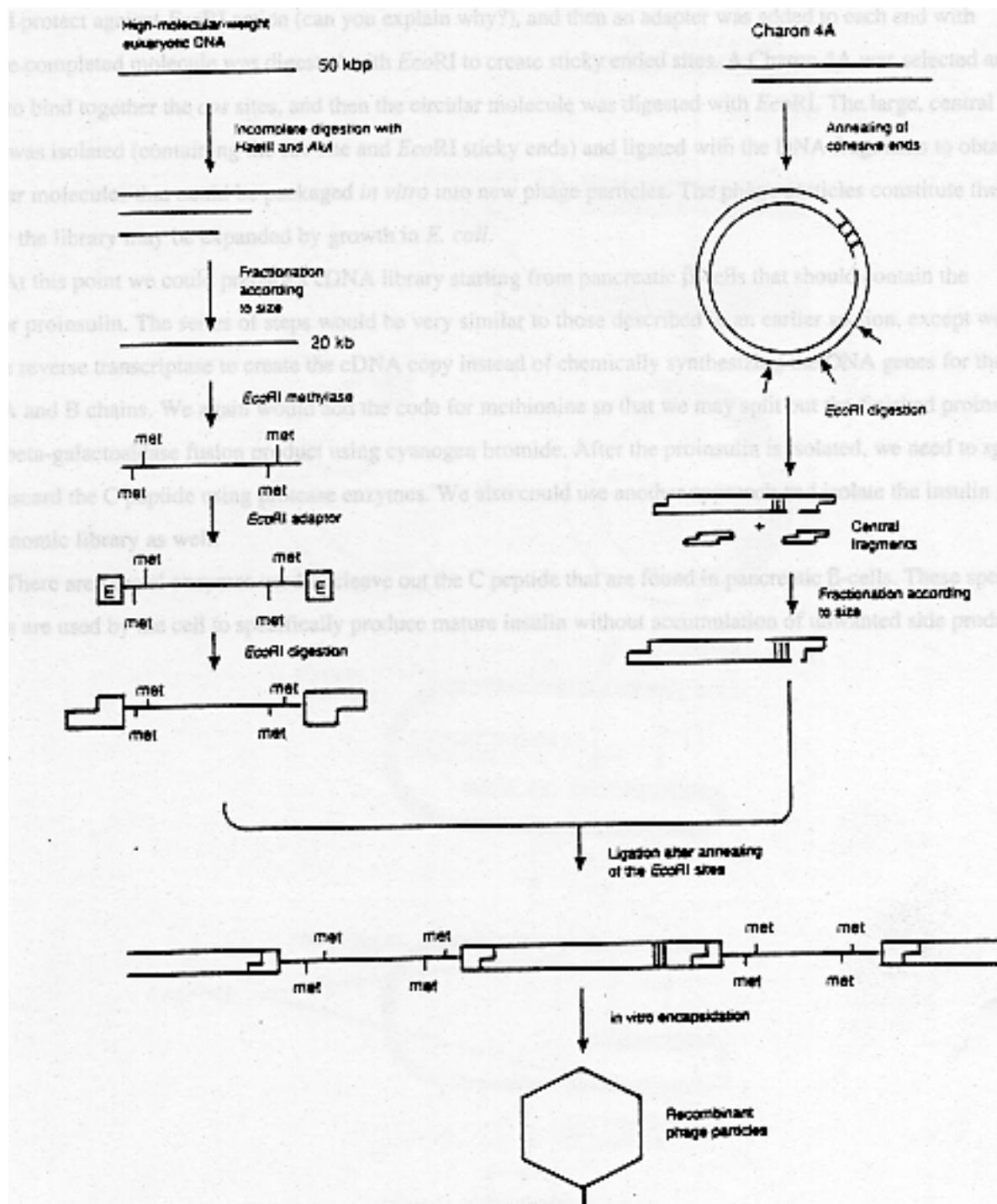
FIGURE 39. Cloning a DNA library

At this point we could prepare a cDNA library starting from pancreatic beta-cells that
should contain the mRNA for proinsulin. The series of steps would be very similar to
those described in an earlier section, except we would use reverse transcriptase to create the
cDNA copy instead of chemically synthesizing the DNA genes for the separate A and B

chains. We again would add the code for methionine so that we may split out the finished proinsulin from the beta-galactosidase fusion product using cyanogen bromide. After the proinsulin is isolated, we need to split out and discard the C peptide using protease enzymes. We also could use another approach and isolate the insulin gene from a genomic library as well.
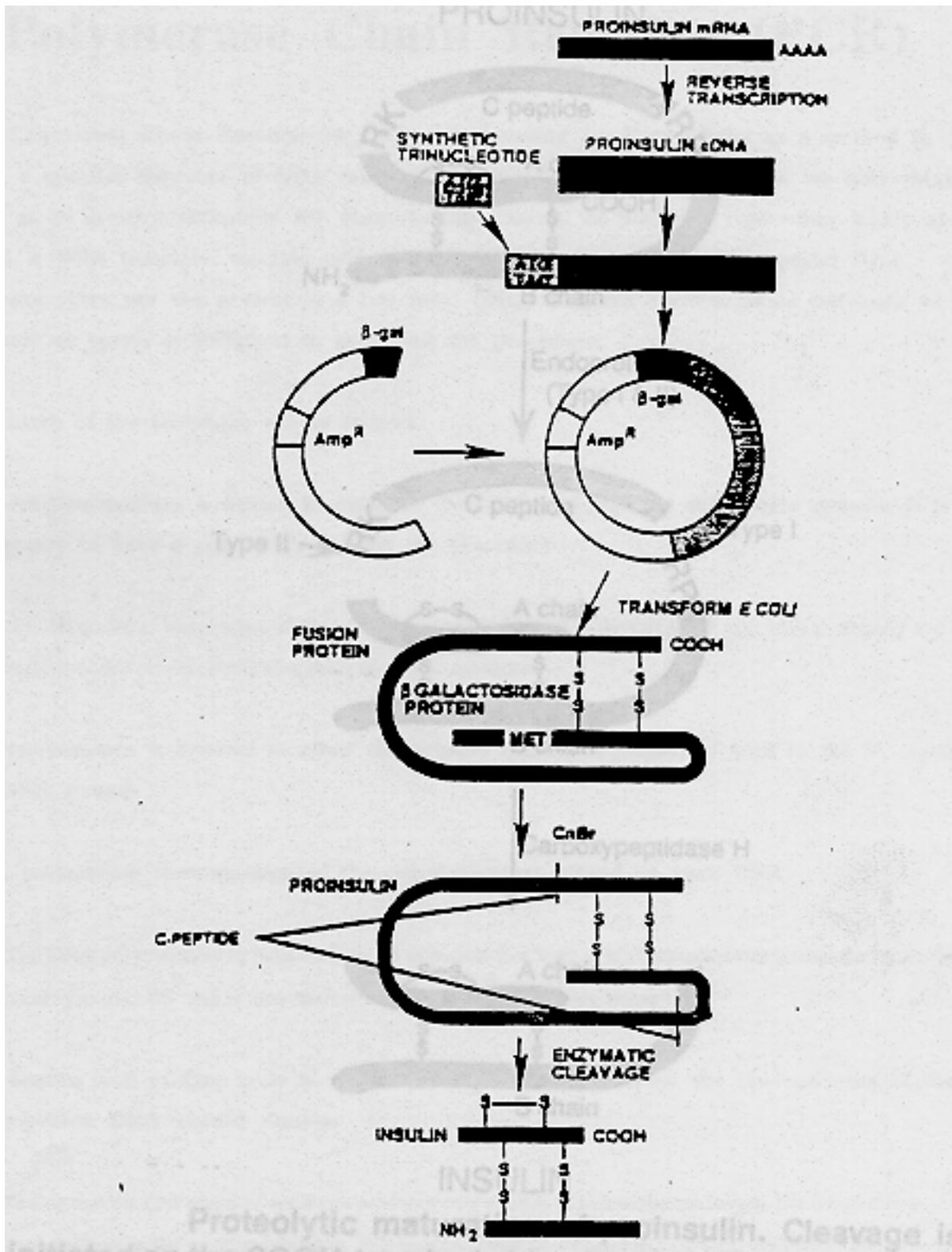
FIGURE 40. Cloning the complete gene for insulin

There are special enzymes used to cleave out the C peptide that are found in pancreatic beta-cells. These special peptidases are used by the cell to specifically produce mature insulin without accumulation of unwanted side products.

Chapter 5 Polymerase Chain Reaction (PCR)

5.1   Introduction

Polymerase Chain Reaction (or PCR) was invented by Kary Mullis as a method to amplify a specific sequence of DNA millions of times in the test tube. Thus, we may think of PCR as an *in vitro* technique for gene cloning. Instead of plasmids replicating inside of bacteria, a DNA template, suitable primers and substrates together with a special DNA polymerase carry out the process in a test tube. Often the DNA from a single cell--such as a single hair or sperm is sufficient to carry out the procedure.
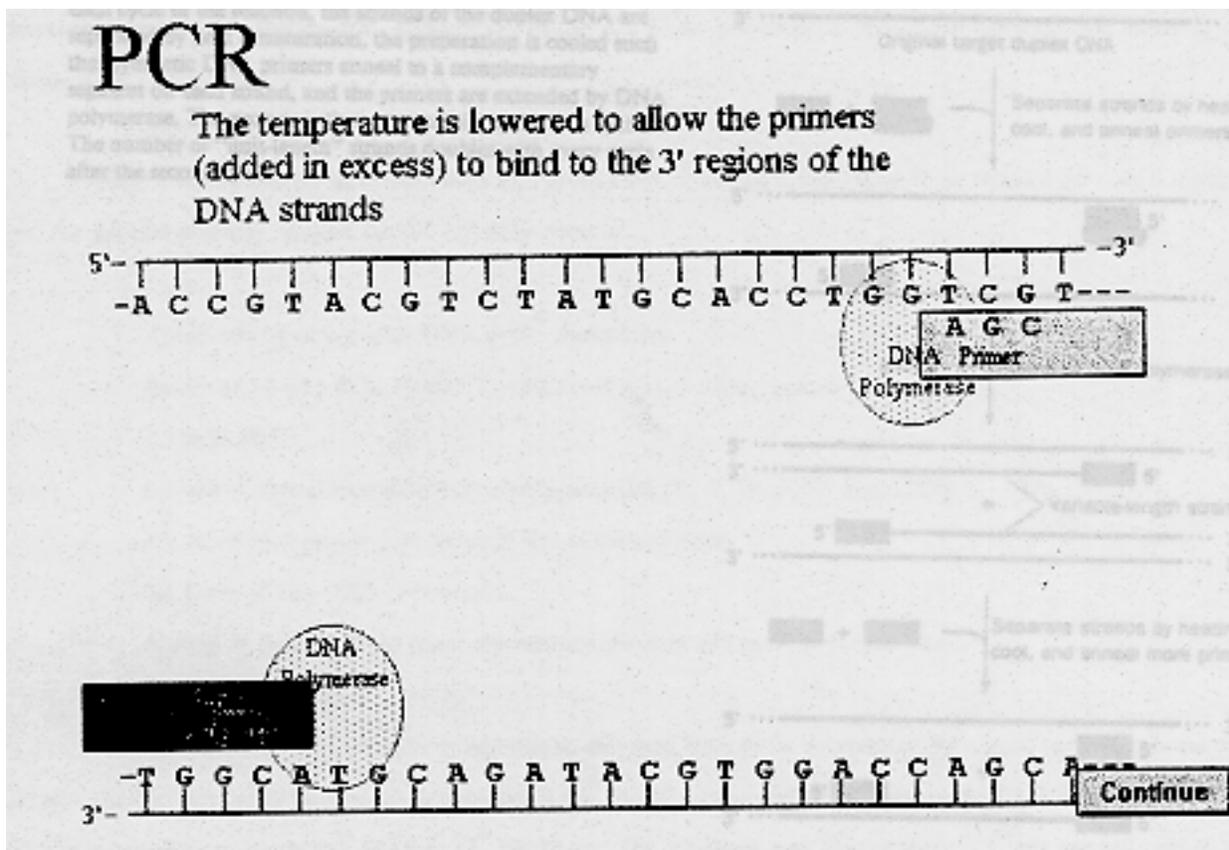

FIGURE 41. The first step in PCR

5.2   The Reaction

The reaction mixture is heated to denature the DNA and separate the double strands. It is not necessary to have a pure DNA to start the procedure. It is necessary to

know some of the base sequence at each end of the DNA that you wish to amplify and to prepare single-stranded primers containing the correct base sequences

The temperature is lowered to allow the primers (added in excess) to bind to the 3' regions of the DNA strands.

DNA polymerase then synthesizes the complementary strand on each DNA. The DNA polymerase was isolated from an organism that lives at high temperatures so that the enzyme is stable to heating to over 90° and is enzymatically active at 70°

The heating and cooling cycle is repeated many times. Each time the concentration of the newly-amplified DNA should double. Notice that the first round of synthesis produces new strands of indeterminate length, but the products become more uniform as the reaction proceeds. This happens because the products are soon defined by the primers at each end of the desired sequence. Commercial thermal cyclers are available that will precisely automate the many heating and cooling cycles.

A particular gene often may be amplified to microgram quantities in as little as four hours.
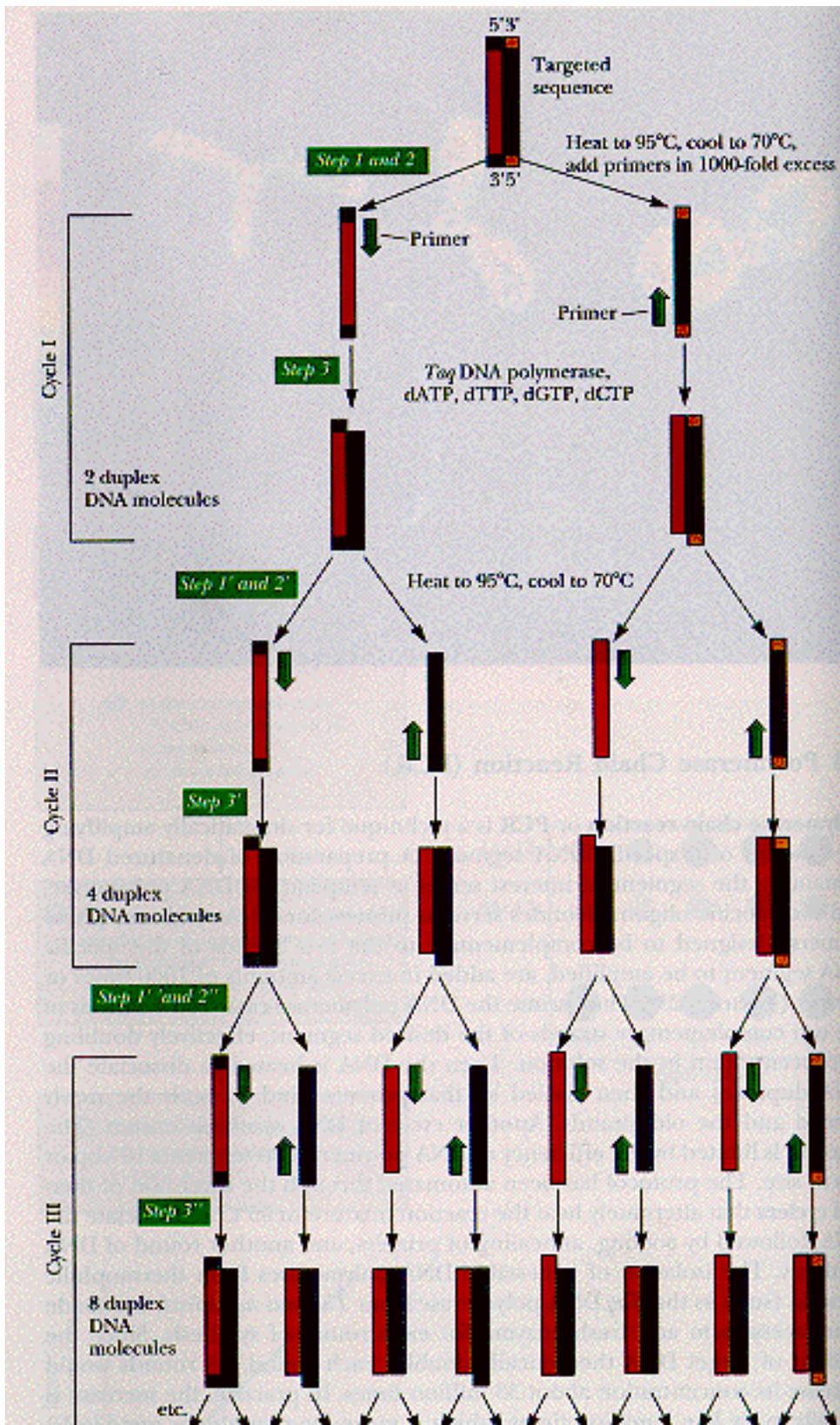
FIGURE 42. Multicycle PCR

5.3   Reaction Particulars

Reaction conditions

Specific conditions for maximal yields of a given DNA sequence usually have to be worked out . As a starting point, the general reaction mixture (in 0.1 ml) may contain:

- About 100 ng of complex DNA (~10^4 molecules)
- Buffer of 50 mM KCl, 10 mM Tris.HCl (pH 8.4), 1 mg/ml gelatin
- 1.5 mM MgCl2
- 0.2mM of each deoxynucleotide triphosphate (dATP, dCTP, dGTP and dTTP)
- 50mM of each primer (one for each end of the sequence)
- 2.5 Units of *Taq* DNA polymerase
- A drop of mineral oil to cover the reaction mixture and prevent evaporation

The concentration of Mg ions is quite critical, and may have to be determined by testing various concentrations.

Typically, the thermal cycler will be set to denature at 94° for 20 sec, anneal the primers to the DNA at 55° for 20 sec, and allow the polymerase to synthesize DNA at 72° for 30 sec. The machines may heat at about 0.3^O per sec and cool at about 1^O per sec, so that a complete cycle takes 3.75 min. The automated machine might run this program for 20 to 30 cycles to amplify the desired sequence.

Reference: Ausubel, F. M., *et al*., Short protocols in molecular biology, 2nd edit., J. Wiley & Sons, N. Y. 1992.

Primer Selection

Select primers (if possible) that do not have repeat stretches of one base, or primers that might fold over on themselves to form internal hairpin structures, or that might hydrogen bond with each other. Primers should be between 20 and 30 nucleotides in length and be of typical G plus C content. If only shorter primers are available (12 to 15 nucleotides) the polymerase reaction temperature will have to be lowered to perhaps $50\text{-}60^O$ so that the primers do not dissociate from the template DNA before polymerase can act. Unrelated nucleotide sequences may be added onto the 5' end of primers to provide other functions in the product. Different sequences could be placed at each end. An example might be the addition of an endonuclease restriction site]

The DNA polymerase

The PCR method first used DNA polymerase I from *E. coli* (the Klenow fragment). This enzyme has an ordinary temperature stability, and so fresh enzyme had to be added after each denaturation step. The *Taq* polymerase was isolated from *Thermus aquaticus* which is was found in a hot spring in Yellowstone National Park. The microorganism grows optimally at 70-75° and the isolated enzyme catalyzes DNA synthesis optimally at 75-80°. It is quite stable at still higher temperatures and retains about 50% of its activity after one hr at 94°. Thus, the enzyme needs to be added only once at the beginning of the procedure. In contrast to the *E. coli* enzyme, *Taq* polymerase has no "editing" function, and may make more errors in synthesizing DNA. Usually this is not a serious problem, but if the product is used to determine the base sequence of a gene, several cloned PCR products should be sequenced to be sure no errors have crept in.

The DNA sample

Often, whole cells may be added to the reaction mixture and the high temperature and alkaline conditions of the denaturation step will lyse the cells and liberate DNA. This procedure works with isolated white blood cells, but not with whole blood. Hematin and other porphyrin breakdown products of hemoglobin were found to inhibit PCR. In some cases, it may be preferable to isolate and quantitate the DNA before using it in PCR.

A procedure for plucked hairs is as follows:
Cut off about 0.5 cm at the root end (containing the cell body). Place in PCR buffer with nonionic detergents (0.45 % NP40 and 0.45% Tween 20) and Proteinase K(0.01 mg in 0.1 ml). Incubate at 50° for one hour to digest cell proteins. Incubate at 95° for 10 min to inactivate Proteinase K. Proceed with PCR or store frozen.

The product

Usually the approximate size of the gene being isolated is known or suspected. Gel electrophoresis should be used to verify that PCR yielded predominately one product in the correct size range. If impurities are detected, often the DNA in the band of the correct size may be eluted from the gel and used. The DNA may be precipitated with ethanol, redissolved and used in DNA sequencing reactions (see above reference for details). PCR also may be used to produce mostly single-stranded product by using one primer in 100-fold excess over the other.

Reaction plateau

Amplification usually slows or stops after 20-30 cycles. The reason for this is believed to be that the *Taq* polymerase cannot keep up with the many strands to be copied after their concentration reaches high levels. Another cause might be that the new strands will anneal with one another before the primers can bind and be extended. If more product is needed, it is simplest to run as many duplicate tubes as required.]

Chapter 6 Insulin and Diabetes

6.1 Insulin

The normal level of glucose in blood is between 80-90 mg/100ml or 4-6 mMolar. When the level of glucose goes up (after eating a meal), the beta (b) cells of the pancreas secrete insulin. When the level of glucose begins to fall, the alpha cells of the pancreas secrete glucagon which acts mostly on the liver to liberate glucose from stored glycogen. Thus, the effects of the two hormones counterbalance each other.
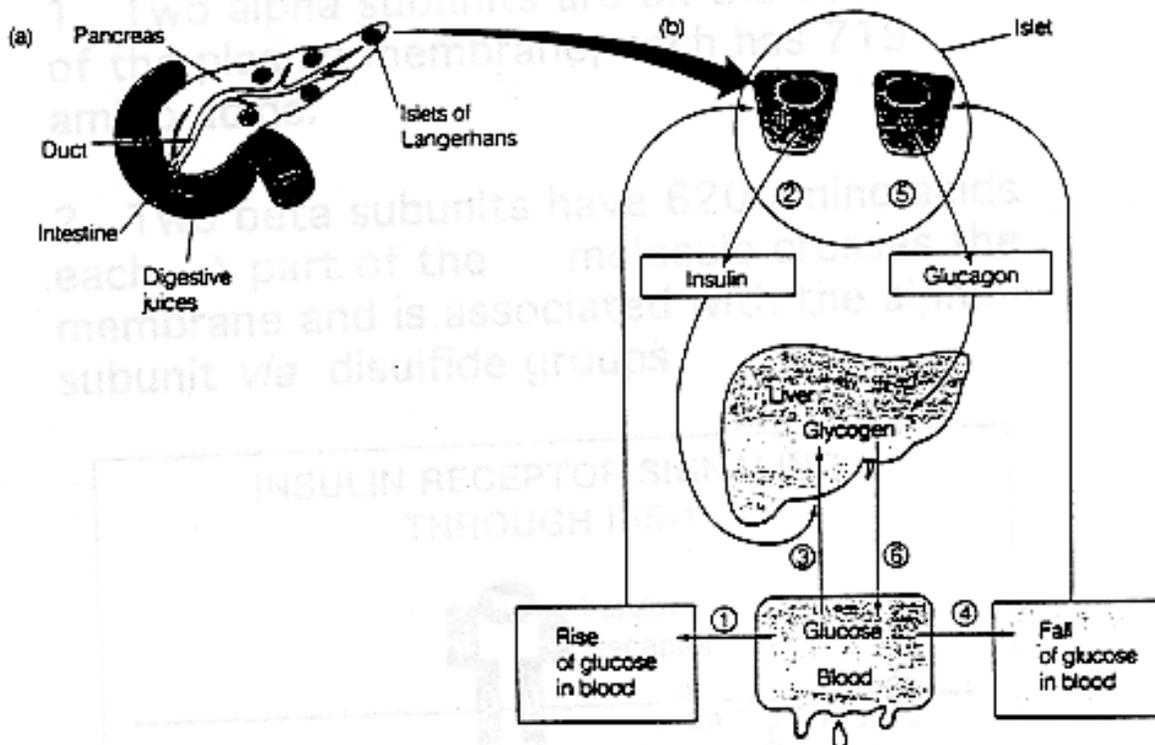

FIGURE 43. Relationships between insulin and glucagon action

For insulin to act, it must bind to insulin receptors . These receptors are membrane proteins found on many tissues (see diagram). The action of insulin is complex and changes with time.

INSULIN RECEPTOR SIGNALING THROUGH IRS-1

Courtesy of Morris F. White, Ph.D.
Joslin Diabetes Center, Boston, MA 02215
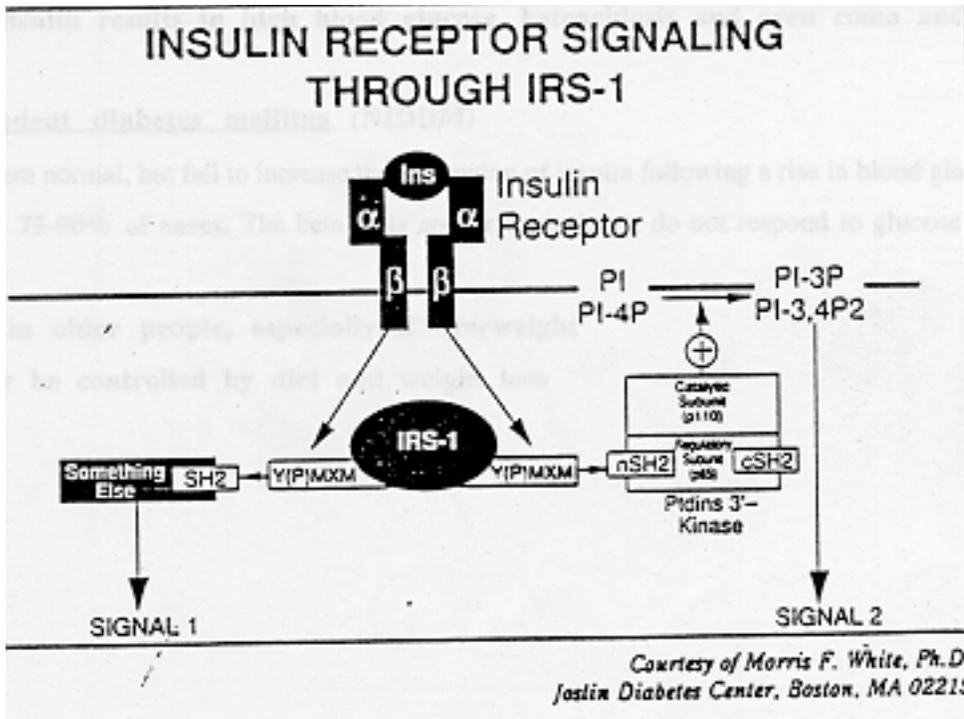
FIGURE 44. The binding of insulin to its receptor leads to complex intercellular reactions.

Short term:

Within seconds after insulin binds to a receptor complex, the receptor proteins become phosphorylated and a receptor tyrosine kinase is activated which may lead to other cellular proteins becoming phosphorylated.

Within minutes there is a large increase in transport of glucose due to a large migration of glucose transporter proteins from the interior of fat and muscle cells to the cell membrane. Liver always has a very large capacity to take up glucose. The excess glucose is converted into glycogen for storage. A change in gene regulation occurs, leading to turning on of new genes. Some insulin receptor in taken into the cell.

Longer term:

There is an increase in glycogen synthesis in liver, fat synthesis increases in fat cells, and growth-promoting factors are synthesized (insulin-like growth factor). As glucose levels fall in blood, insulin secretion also declines. Soon glucagon is released which stimulates glycogen breakdown in liver with release of glucose into the circulation and activation of a hormone-sensitive lipase in fat cells which stimulates release of fatty acids to be used as alternate energy sources to glucose.

Glucose Flux:

Humans use about 10 gm/hr of glucose when resting and 40 gm/hr during exercise. After a meal, about 30 gm/hr may pour into the blood stream. This glucose must be rapidly removed by the liver and muscle and stored as glycogen
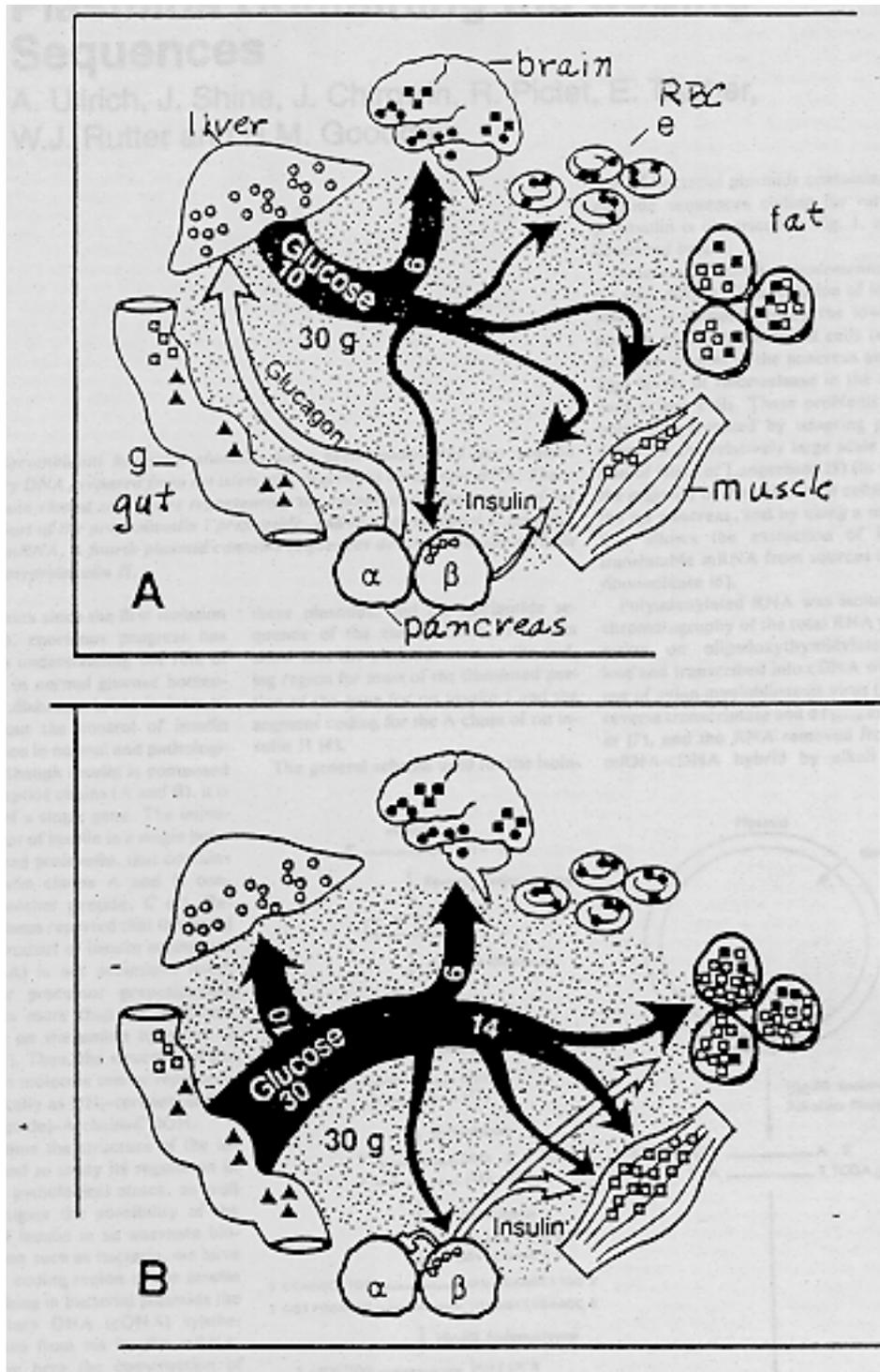
FIGURE 45. Glucose flux in the resting state (A) and after eating (B). There are rapid increases in glucose transporters in muscle (squares) and fat cells (squares) in respond to insulin action

6.2   Diabetes

About 2% of the USA population suffer from diabetes. There are two main types.

Insulin-dependent diabetes mellitus (IDDM).   The pancreas stops making insulin, which then must be supplied (usually by injection). IDDM accounts for 10-25% of cases. Patients do not make enough insulin because of autoimmune destruction of the beta cells. Insulin must be given by injection on a carefully controlled basis. Lack of insulin results in high blood glucose, ketoacidosis and even coma and death.

Non-insulin-dependent diabetes mellitus (NIDDM) .   Is very common. The beta-cells seem normal, but fail to increase their secretion of insulin following a rise in blood glucose. NIDDM accounts for 75-90% of cases. The beta cells appear normal, but do not respond to glucose by producing insulin.   Common in older people, especially if overweight. Often may be controlled by diet and weight loss.

References

Text:

Nicholl, D. S., *An introduction to genetic engineering,* Cambridge University Press, England, 1994.

References for Engineering Biotechnology:

Ausubel, F., *Short protocols in molecular biology*, 2nd edit., Wiley, New York, 1992.

Berg, P., Singer, M. *Dealing with genes. The language of heredity*. University Science Books, Mill Valley, CA. 1992

Davies, J., Reznikoff, W. S. *Milestones in biotechnology. Classic papers on genetic engineering.* .Butterworth-Heinemann, Boston,1992.

Garrett, R. H., Grisham, D. M., *Biochemistry*, Saunders, Philadelphia, 1995.

Lehninger, A. L., Nelson, D. L., Cox., M. M. *Principles of biochemistry*. Worth Pub., New York, 1993.

Lodish, H., Berk, A., Matsudaira, P., Baltimore, D., Zipursky, L., and Darnell, J*., Molecular Cell Biology* and *Molecular Cell Biology* 3.0 CD for Macintosh and PC/Windows. W. H. Freeman and Co., New York, 1996

Lucotte, G., Baneyx, F. *Introduction to molecular cloning techniques.* VCH Pub., New York., 1993

Primrose, S., *Molecular biotechnology*. Blackwell Scientific Pub., Cambridge, 1991.

Recommended articles from Primrose: *Milestones in Biotechnology* :

Itakura, K., et al., Expression in Echerichia coli of a chemically systhesized gene for the hormone somatosatatin (1977) p 84

Southern, E., Detection of specific sequences among DNA fragments separated by gel electrophoresis (1975) p 122

Suggs, S. et al., Use of synthetic oligonucleotides as hybridization probes: Isolation of cloned cDNA sequences for human b2-microglobulin (1981) p 140

Ullrich, S., et al., Rat insulin genes: Construction of plasmids containing the coding sequences (1977) p 243

Villa-Komaroff, L., et al., A bacterial clone synthesizing proinsulin (1978) p 316

Innis, M., et al., DNA sequencing with Thermus acquaticus DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA (1988) p 7

Mullis, K., et al., Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction (1986) p 17

BIOREACTORS

**Engineering Biotechnology
Gateway Project**
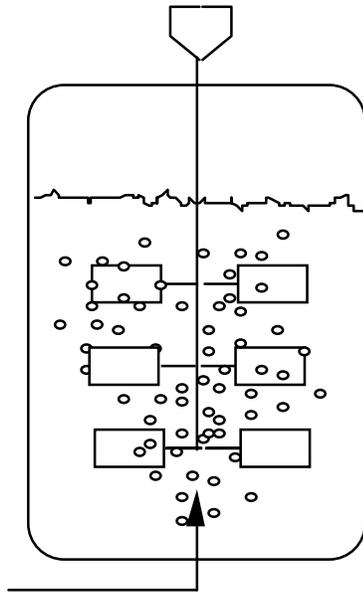
**Raj Mutharasan
Drexel University**

Foreward

The following "text" was written to provide a simple structure for discussion of issues governing manufacture of biopharmaceuticals. The manufacturing section is broken down into two main segments, namely bioreactors and bioseparations. The former deals with all issues from cell to expression of desired protein in a bioreactor while the latter is concerned with engineering issues that relate to purification of the expressed product. In this text, we discuss the bioreactor part. Bioseparations is being developed by Professor Jordan Spencer of Columbia University and will be added as soon as it is available. Most, if not all biochemical engineers employed by biotechnology companies work on problems related to bioreactors and bioseparations.

Although the topic of bioreactors can be discussed in an entire course, the intent here is to provide a brief introduction to it so that the student becomes aware of issues in design of large scale systems. Many exhaustive treatments are available in the literature. To the author's knowledge, the simplified structure provided here is <u>original</u> for bioreactor analysis. It follows the pedagogical structure of building on the principles of stoichiometric calculations, thermodynamic and kinetic analyses an average engineering student learns in freshman courses. The "text" follows the structure: mass balance, then energy balance followed by rate analysis. Such an arrangement has been found to be successful in teaching chemical reactor design.

The "text" was used in a course titled as "Engineering Biotechnology" at Drexel during Winter term of 1996. The material covered herein was discussed in 8 hours of instruction including recitation.

This project, funded by the Gateway Coalition, is concerned with introducing the topic of Engineering Biotechnology to undergraduate engineering students of all majors as an elective. The idea is to provide breadth by integrating biological concepts and ideas with quantitative engineering principles. The challenge is to introduce major ideas from genetic engineering, biomanufacture, drug delivery and biosensors in a course consisting of 30 to 40 hours of instruction for a typical Junior engineering student.

The author welcomes suggestions for improvement. He can be contacted at: Raj.Mutharasan@coe.drexel.edu.

Raj Mutharasan
Philadelphia
March 25, 1996

Table of Contents

Chapter 1 Introduction

Commercial production of products produced by genetically engineered microorganisms requires two distinct body of knowledge, namely,   molecular biology and process engineering. Background in molecular biology will enable us to create effectively expressed <u>genes</u> in microorganisms or cells of animal, insect or plant origin that can be used for industrial production. Background in process engineering principles wil enable us to design and operate large-scale plants for growing genetically-engineered organisms and for the subsequent processing of purification and formulation of product. In the early days, it was thought that scale-up was simply a matter of using larger volumes. That is, conditions that were found to be good at a small-scale would be equally effective on a larger scale and that to achieve this it was merely necessary to use a larger <u>fermentor</u> vessel with a larger medium volume. Such an approach resulted in not only product variability, both in terms of yield and quality, but also expensive operating costs. Hence, a systematic study of process engineering principles is needed for scaling up and operation of biotechnological processes for manufacture.

1.1 What is a Bioreactor?

The heart of a bioprocess used for manufacture of biological, is a bioreactor. A commercial unit is illustrated in Fig 1-1. It is usually a large vessel ranging from 1000

Fig 1-1 Large Scale Fermentor Used for Cultivating Bacteria and Yeast.

liters to 100,000 liters, made of stainless steel equipped with temperature, pH and dissolved oxygen measurement and control systems. The bioreactor is equipped with an agitation system to keep the contents uniformly mixed and to provide oxygen transfer. The design of the bioreactor should ensure sterility and provide for containment of the genetically engineered microorganism. The bioreactor includes sensors that permit monitoring of as many critical process parameters (temperature, pH, dissolved oxygen) as possible so that they can be adjusted to within allowable values.

1.2 Production and Purification

 Generally, large-scale microbial cultivation or cell culture, and product purification steps are carried out in a stepwise manner (Fig. 1-2).
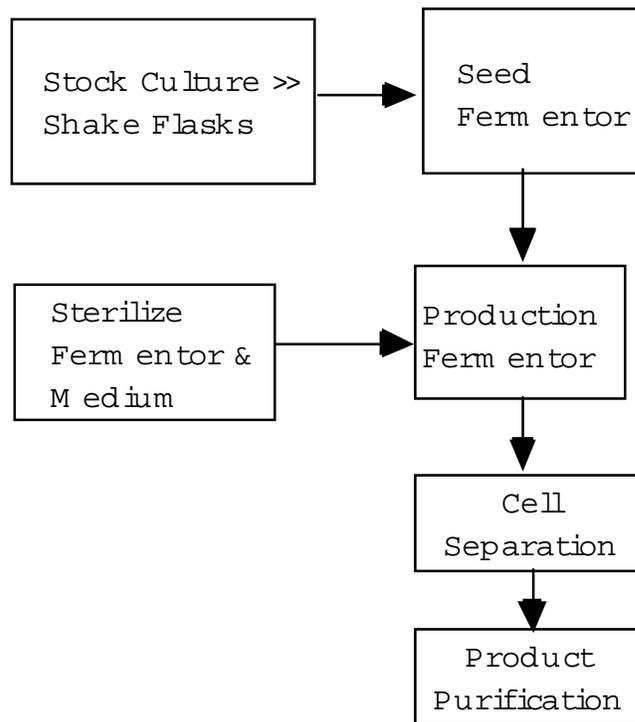
Figure 1-2 Steps in Large Scale Biotechnological Processes

A typical procedure begins with the formulation and sterilization of growth medium and sterilization of the fermentation equipment. The cells are grown first as a stock culture (5 to 10 mL), then in a shake flask (200 to 1,000 mL), and then in a seed fermentor (10 to 100 liters). Finally, the production fermentor (1,000 to 100,000 liters) is inoculated. After the fermentation step is completed, the cells are separated from the culture fluid by either centrifugation or filtration. If the product is intracellular, the cells are disrupted, the cell debris removed, and the

product recovered from the debris-free broth. If the product is extracellular, it is purified from the cell-free culture medium.

Although microorganisms can be grown in a number of different ways (batch, fed-batch, or continuous culture), it is most common to cultivate them in a batch fermentor. In batch fermentation, the sterile growth medium is inoculated with a suitable amount of microorganisms, and the fermentation, i.e cell growth, proceeds without any further addition of fresh growth medium. In some processes the cells themselves will be the product. In others the product is what the cells produce as they grow or as they are induced to produce. For example, in yeast manufacture the product is the biomass (cell) itself while in insulin manufacture, the product is formed as an intracellular product. In this case, the cells are disrupted to harvest the intracellular insulin and the cell debris is discarded.

1.3 Bioreactor Engineering Issues

It is necessary to monitor and control culture parameters such as dissolved oxygen concentration, pH, temperature, and mixing regardless of the process that is used to grow cells. Changes in these parameters can significantly affect the process yield and the stability of product protein.

Optimal growth of *E. coli* cells and many other microorganisms that are used as hosts (see section on Molecular Biology) for recombinant genes usually require large amounts of dissolved oxygen.   Because oxygen is sparingly soluble in water (8.4 mg/L at 25°C), it must be supplied continuously -- generally in the form of sterilized air -- to a growing culture. The air produces bubbles and the stirrer is used to break up the bubbles and mix the content of the reactor. If air flow is inadequate or the air bubbles are too large, the rate of transfer of oxygen to the cells is low and is not sufficient to meet cellular oxygen demand. Thus the fermentors are equipped to monitor dissolved oxygen level of the medium, to transfer oxygen efficiently to the culture medium, and to mix the broth to provide a uniform culture environment.

Temperature is another physiological parameter that is be monitored and controlled. Microorganisms have optimal temperature for growth. If grown at a temperature below the optimum, growth occurs slowly resulting in a reduced rate of cellular production. On the other hand, if the growth temperature is too high, not only will death occur, but in situations where the target protein may be under the control of temperature sensitive promoter, it may be expressed prematurely, lowering product yield.

Most microorganisms grow optimally between pH 5 and 7. As the cells grow, metabolites are released into the medium, a process that can change medium pH. Therefore, the pH of the medium must be monitored and be adjusted by base or acid addition to maintain a constant pH.

Adequate mixing of a microbial culture is essential for ensuring adequate supply of nutrients and prevention of the accumulation of any toxic metabolites within the bioreactor. Although good mixing is easy to achieve at small scales, it is one of the major problems in increasing the scale of bioreactors. Agitation of the broth also affects the rate of transfer of oxygen and heat transfer removal via cooling coils. Excessive agitation can cause mechanical damage to microbial or mammalian cells. Hence a balance must be reached between the need to provide good mixing and the need to avoid cell damage..

The process design should also include factors that make it easy to implement Good Manufacturing Practices. Although most recombinant microorganisms are not hazardous, it is important to design processes that ensure that they are not inadvertently released into the environment. Hence, fail-safe systems should be considered in equipment design and operation to prevent accidental spills of live recombinant organisms and to contain them if a spill does occur. Furthermore, all recombinant microorganisms must be treated by a verified procedure to render them nonviable before they are discharged from the production facility, and the spent culture medium must also be treated to ensure that it does not contain viable organisms and that its disposal does not create an environmental hazard.

Summary

In this chapter you were introduced to main componenets of a biopharmaceutical manufacturing facility, and specifically issues concerning bioreactors. In the chapters following, we will learn how to determine material need of a bioreactor.

Chapter 2 Stoichiometry of Cellular Growth

A good starting point for discussion on cell growth is to examine what the cells are made of, that is its chemical composition. Although there are many different biological species, it turns out that a very large fraction of their mass is made of a few elements - carbon, oxygen, nitrogen and hydrogen. You will note that these are among the most abundantly found elements on earth.

2.1 Cell Composition

Cells primarily contain water! Typically 70% of cell mass is water and the remaining is dry matter. Therefore it is conventional to express cell composition on a dry basis. The microorganism *Eschericia coli* is widely used in genetic engineering. Typical elements found in *Eschericia coli* are given below:

Table 1 Elemental Composition of E. coli
(after Stanier et al)

Element	% Dry Weight
C	50
O	20
N	14
H	8
P	3
S	1
K	1
Na	1
Ca	0.5
Mg	0.5
Cl	0.5
Fe	0.2
others	0.3

Nearly half of the dry matter in cells is carbon and the elements carbon, oxygen, nitrogen and hydrogen total up to about 92% of the total. This observation for E. coli is also found to be generally true for other cellular organisms.

Table 2 Elemental Composition of Microorganisms

Microorganism	Carbon Source	Growth Rate	Composition				Empirical Formula	Molecular Weight
			C	H	N	O		
Klibsiella aerogenes	Glycerol	0.1	50.6	7.3	13.0	29.0	$CH_{1.74}O_{0.43}N_{0.22}$	23.7
Aerobacter aerogenes	Complex		48.7	7.3	13.9	21.1	$CH_{1.78}O_{0.33}N_{0.24}$	22.5
Aerobacter aerogenes	Complex	0.9	50.1	7.3	14.0	28.7	$CH_{1.73}O_{0.24}N_{0.43}$	24.0
Saccharomyces cerevisiae			47.0	6.5	7.5	31.0	$CH_{1.66}O_{0.49}N_{0.13}$	23.5
Sachromyces cervisiae			50.3	7.4	8.8	33.5	$CH_{1.75}O_{0.15}N_{0.5}$	23.9
Candida utilis	Glucose	0.45	46.9	7.2	10.9	35.0	$CH_{1.84}O_{0.56}N_{0.2}$	25.6
Candida utilis	Ethanol	0.43	47.2	7.3	11.0	34.6	$CH_{1.84}O_{0.55}N_{0.2}$	25.5

Table 2 above shows that in different microbes, the carbon content varies from 46 to 50%, hydrogen from 6 to 7%, nitrogen 8 to 14% and oxygen from 29 to 35%. These are small variations and the variations appear to depend on substrate and growth conditions. For many engineering calculations, it is reasonable to consider cell as a chemical species having the formula

$$CH_{1.8}O_{0.5}N_{0.2}$$

This engineering approximation is a good starting point for many quantitative analysis while a more carefully formulated empirical formula based on proximate analysis may be necessary for complete material flow analysis. The cell molecular weight for the above cell formula is 12+1.8 + 0.5(16) +0.2 (14) = 24.6.

Example 2-1

Suppose we want to produce 10 g of cells using glucose as a carbon source. What is the minimum amount of glucose that would be needed?

Solution

Assume cell composition as $CH_{1.8}O_{0.5}N_{0.2}$

Glucose is $C_6H_{12}O_6$

MW of glucose is 180

$$\text{Moles of cells to be grown} = \frac{10}{24.6}$$

Since glucose has 6 moles of carbon per mole of glucose,

$$\text{Moles of glucose needed} = \frac{1}{6} \cdot \frac{10}{24.6}$$

$$\text{Therefore, min glucose needed} = \frac{1}{6} \cdot \frac{10}{24.6} \cdot 180 \approx 12.2 \text{ g}$$

2.2 Growth Reaction

In the above example, we have assumed that all of the carbon found in substrate (glucose) is incorporated into cell mass. This does not happen as the cell needs to "oxidize" or respire some of the carbon to produce energy for biosynthesis and maintenance of cellular metabolic machinery. In addition cells may produce extracellular products that accumulate in the broth. Hence we can represent growth as:
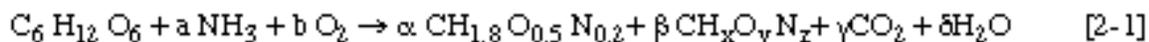
$$\text{Cell} + \text{Medium} + \text{Oxygen} \rightarrow \text{More Cells} + \text{Extracellular Product} + \text{Carbon dioxide} + \text{Water}$$

The medium is the "food" for the cell. It serves as a source for all elements needed by the cell to grow (or biosynthesis) and for product formation. The compounds carbon dioxide and water on the product side of the reaction above result from oxidation of glucose in the medium.

Since the cellular material contains C, N, P, S, K, Na, Ca, etc, the medium must be formulated to supply these elements in the appropriate form. The above growth reaction can be re-stated as

$$\text{Cell} + \{\text{C-source, N-source, others}\} + O_2 \rightarrow \text{More Cells} + \text{Extracellular} + CO_2 + H_2O$$

If we neglect the "others" and assign stoichiometric coefficient for each of the species in the above equation on the basis of one mole of glucose (C-source) consumed, we re-write the above as

$$C_6H_{12}O_6 + a\,NH_3 + b\,O_2 \rightarrow \alpha\,CH_{1.8}O_{0.5}N_{0.2} + \beta\,CH_xO_yN_z + \gamma\,CO_2 + \delta\,H_2O \qquad [2\text{-}1]$$

where ammonia represents the nitrogen source. We will refer to this reaction as **growth reaction**.

Note that whatever nitrogen that is supplied in the medium, it is expressed as equivalent nitrogen in the form of ammonia. Cells require nitrogen in both organic and inorganic form. It is common to supply the inorganic nitrogen as salts of ammonium (e.g. ammonium phosphate) while the organic nitrogen is usually supplied as amino acids or proteinous extracts which are rich in nitrogen. In most production processes using recombinant cells, glucose is used as the carbon source. However, in the production of low value products, less expensive

carbon sources such as molasses ($ 0.10 / lb) or corn meal (about $ 0.12 / lb) are used. Compare this against glucose at $ 1.00 /lb!

The growth reaction derived above is useful in interpreting laboratory data reported in the literature. Because the early work in cell growth were reported by microbiologists, it is necessary for us to learn the terms used by microbiologists to describe growth stoichiometry. We will then relate the above reaction equation to commonly reported cell properties.

2.3 Cell Yield and Stoichiometric Coefficients

Consider the experimental cell (*Pseudomonas lindneri*) growth data shown in Fig 2-1a, originally reported by Bauchop and Elsden. The experiment consisted of inoculating five test tubes containing growth medium with the bacterium. Each of the test tubes contained different concentrations of carbon source - in this case glucose at levels from about 4 mM to 36 mM.   The cultures were incubated anaerobically (i.e. in absence of oxygen) at growth temperature (30 C) for two days or until growth ceases. The resulting cells were filtered, dried and weighed. This mass of bacteria obtained is plotted against the starting glucose concentration. The important observation illustrated by the data is the straight line relationship between carbon source concentration (reactant in chemical parlance) and the cell concentration (product).
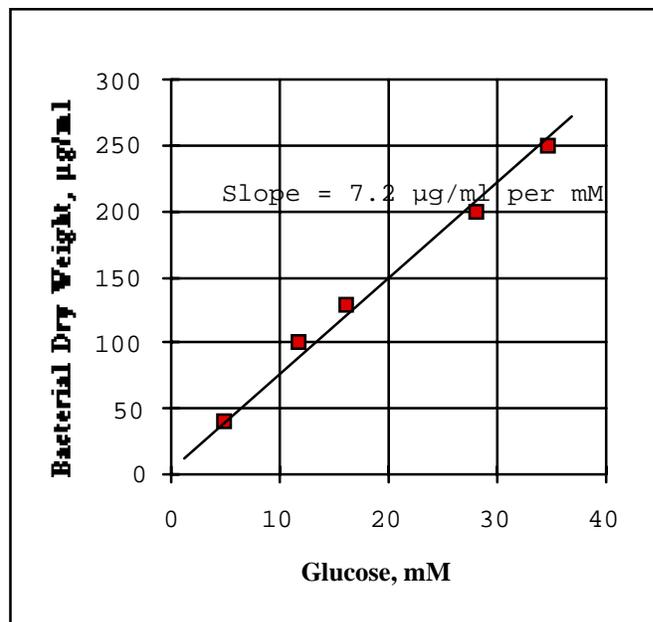

Fig 2-1a Anaerobic growth of Pseudomonas lindneri on glucose .
(Data from Bauschop et al 1960)

The slope of the line represents the amount of cells obtained per unit amount of glucose consumed.

$$\text{Slope} = \frac{270 \ \mu\text{g ml}^{-1}}{33 \ \text{mM}} \implies \frac{270 \ \mu\text{g ml}^{-1}}{33 \ \mu\text{mol ml}^{-1}} \implies 8.2 \ \mu\text{g} \ \mu\text{mol}^{-1}$$

If we convert the above to mass basis,

$$\text{Slope} = 0.046 \ \text{g of cells per g of glucose consumed}$$

The above value is often called **cell yield, growth yield, or yield**. If one examines the growth reaction stated in the previous section, the slope (in mass units) we calculated above can be equated as follows.

$$\text{Cell Yield} = \alpha \left(\frac{\text{MW of Cell}}{\text{MW of Substrate}} \right)$$

In the above the numerator term contains the amount of cell created and the denominator contains the amount of substrate consumed. In other words, the measurements reported by Bauschop and Elsden enable us to calculate the stoichiometric coefficient, α. That is,

$$\alpha = (0.046) \cdot \left(\frac{180}{24.6} \right) \implies 0.33$$

Let us consider another set of data shown in Figure 2-1b. The cell yield depends on growth conditions. You will note that under anaerobic conditions, slope (also yield) is 58.2 g (mol substrate)$^{-1}$ $\implies$ 0.32 g cell (g substrate)$^{-1}$. Similarly under anaerobic conditions, yield is 22 g (mol substrate)$^{-1}$ $\implies$ 0.21 g cell (g substrate)$^{-1}$. Invariably, the yield under anaerobic conditions will be smaller than at aerobic conditions because the cell derives significantly more metabolic energy under aerobic conditions. It is also important to note that not all cells can grow both aerobically and anaerobically.
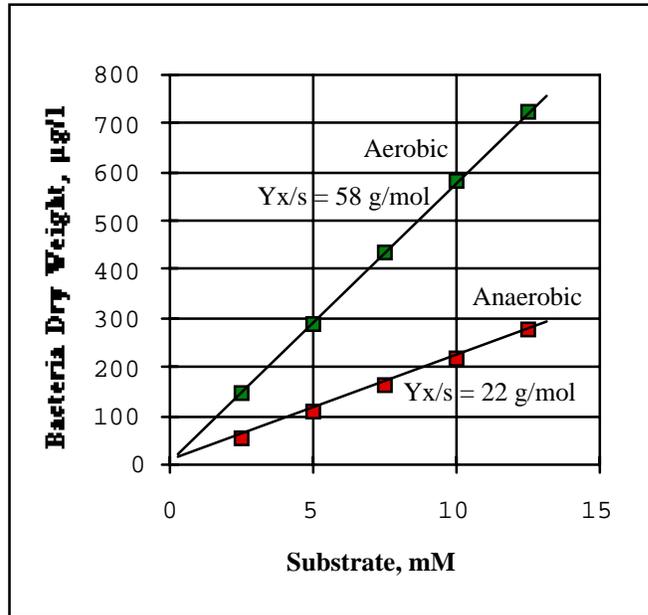
Fig 2-1b Aerobic and anaerobic growth of Streptococcus faecalis
on glucose (Bauschop et al 1960, also Smalley et al , 1968)

From a practical viewpoint, an aerobic organism is preferred. This is because, the amount of product protein produced is proportional to cell amount. Higher biosynthesis is possible with aerobic cultures than with anaerobic ones.

2.4 Mathematical Definition of Yield

Mathematically, cell yield can be defined as

$$Y_{X/S} = \frac{\text{Amount of Cell Produced}}{\text{Amount of Substrate Consumed}} = \frac{\Delta X}{\Delta S} \qquad (2\text{-}2)$$

where ΔX represents change in cell concentration and ΔS represents change in substrate concentration. The subscript X/S indicates the basis of yield - cell on the basis of substrate. This notation comes in handy when we need to calculate yield based on more than one substrate. Examining the above and comparing with **growth reaction**, one notes that the yield defined here corresponds to a mass-based stoichiometric coefficient.
 Taking the limit of Eq(2-2) as ΔS approaches zero,

$$Y_{X/S} = \left| \frac{dX}{dS} \right| \qquad (2\text{-}3)$$

The absolute sign is used to eliminate the negative value of the derivative. Note that dS is negative, because substrate is consumed. Yield is always reported as a positive value.

The above definition of yield can be applied to product, P on the basis of substrate consumed. Thus,

$$Y_{P/S} = \left| \frac{dP}{dS} \right| \qquad (2\text{-}4)$$

Similarly product yield based on cell will be expressed as,

$$Y_{P/X} = \left| \frac{dP}{dX} \right| \qquad (2\text{-}5)$$

In general, yield of the species, i, based on species, j, can be calculated from

$$Y_{i/j} = \left| \frac{di}{dj} \right| \qquad (2\text{-}6)$$

From the above it is clear that we can combine two different yields which have a common species as

$$Y_{i/j} = \frac{Y_{i/k}}{Y_{k/j}} \qquad (2\text{-}7)$$

Example 2-2   Batley (1979) reported aerobic gowth of yeast on ethanol as:

$$C_2 H_5 OH + 0.153\, NH_3 + 1.851\, O_2 \rightarrow 1.03\, CH_{1.704} O_{0.408} N_{0.149} + 0.970\, CO_2 + 2.346\, H_2O$$

Calculate $Y_{X/E}$, $Y_{X/O2}$, $Y_{X/NH3}$ on mass basis.

Solution

MW of cell = 12 +1.704 +(14)•(0.149) + (16)•(0.408) = 22.32
MW of ethanol = (2)•(12) + 5 + 16 + 1 = 46

$$Y_{X/E} = \frac{(1.03) \cdot (MW \ of \ Cell)}{(1) \cdot (MW \ of \ Ethanol)} = \frac{(1.03) \cdot (22.32)}{(46)} \Rightarrow 0.5$$

$$Y_{X/O2} = \frac{(1.03) \cdot (MW \ of \ Cell)}{(1.851) \cdot (MW \ of \ Oxygen)} = \frac{(1.03) \cdot (22.32)}{(1.851) \cdot (32)} \Rightarrow 0.388$$

$$Y_{X/NH3} = \frac{(1.03) \cdot (MW \ of \ Cell)}{(0.153) \cdot (MW \ of \ Ammonia)} = \frac{(1.03) \cdot (22.32)}{(0.153) \cdot (17)} \Rightarrow 8.839$$

Yield of yeast based on ethanol of about 0.5 is consistent with the observation that roughly one half of the substrate is converted to cell mass aerobically. If

yield on a carbohydrate source is significantly less than 0.5, it is likely that medium formulation is inadequate to support good growth.

Example 2-3

Yeast grown on glucose is described by

$$C_6 H_{12} O_6 + 0.48\ NH_3 + 3\ O_2 \rightarrow 0.48\ C_6 H_{10} O_3 N + 3.12 CO_2 + 4.32\ H_2O$$

Calculate the following for a design requiring 50 g/L of yeast in a batch reactor of 100,000 liters.

- Nutirent media concentration for glucose and ammonium sulfate.
- Calculate $Y_{X/S}$ and $Y_{X/O2}$
- Calculate total oxygen required
- Determine oxygen uptake rate (g O_2 L^{-1} h^{-1})when cell concentration increases at a rate of 0.7 g L^{-1} h^{-1},

Solution

Total cell mass to be produced is = $(10^5\ L) \cdot (50\ g\ L^{-1})$ = 5000 kg

$$Y_{X/S} = \frac{(0.48) \cdot (MW\ of\ Cell)}{(1) \cdot (MW\ of\ Glucose)} = \frac{(0.48) \cdot (144)}{(180)}$$

$$\Rightarrow 0.384\ g\ cell\ (g\ substrate)^{-1}$$

$$Y_{X/O2} = \frac{(0.48) \cdot (MW\ of\ Cell)}{(3) \cdot (MW\ of\ O_2)} = \frac{(0.48) \cdot (144)}{(3) \cdot (32)}$$

$$\Rightarrow 0.72\ g\ cell\ (g\ O_2)^{-1}$$

$$Glucose\ needed = \frac{(Cell\ Mass)}{(Y_{X/S})} = \frac{(5000)}{(0.384)} \Rightarrow 13,020\ kg$$

$$Ammonia\ needed\ =\ \frac{(Cell\ Mass)}{(Y_{X/NH3})}\ =\ \frac{(5000)}{(8.471)}\ \Rightarrow\ 590\ kg\ \Rightarrow\ 130\ g\ L^{-1}$$

$$(NH_4)_2SO_4\ needed\ =\ \frac{\frac{1}{2}\bullet\ (MW\ of\ (NH_4)_2SO_4\ )}{MW\ NH_3}\ \bullet\ (Ammonia\ needed)$$

$$=\ \frac{1}{2}\bullet\ \frac{132}{17}\bullet 590\ \Rightarrow\ 2,292\ kg\ \Rightarrow\ 22.9\ g\ L^{-1}$$

$$Total\ Oxygen\ Required\ =\ \frac{(Cell\ Mass)}{(Y_{X/O_2})}\ \Rightarrow\ \frac{5000}{0.72}\ =\ 6,944\ kg$$

$$Oxygen\ Consumption\ Rate\ =\ \frac{Cell\ Mass\ Generation\ Rate}{Y_{X/O_2}}\ =\ \frac{0.7}{0.72}$$

$$\Rightarrow\ 0.972\ g\ L^{-1}\ h$$

2.5 Measurement of Stoichiometric Coefficients

For the growth reaction given in Eq(2-1), the ratio γ/b is called the respiratory quotient, often abbreviated as RQ. It is easily measured in large scale fermentors. In Eq(2-1), if the nature of extracellular product is known (i.e. x,y,z), then it is possible to calculate α,β,γ and δ from experimental measurement of RQ and one other measurement. If no significant amount of extracellular product is formed, as in simply growth processes, then only RQ or one other measurment is needed to compute stoichiometric coefficients. The example given below illustrates this idea.

Example 2-4

For the reaciton equation representing E. coli growth, RQ was measured as 0.85. Calculate α,β,γ and δ.

$$C_6H_{12}O_6 + a\ NH_3 + b\ O_2 \rightarrow \alpha\ CH_{1.8}O_{0.5}N_{0.2}+\ \gamma CO_2 + \delta H_2O$$

Solution

The solution consists of carrying out elemental balances and then solving them. Here, we can write four elemental balances, C, H, O and N. We have five unknowns: α,β,γ and δ, a and b. One additional

relationship is obtained from the given RQ value, thus making the problem solvable!

C balance: $6 = \alpha + \gamma$
H balance: $12 + 3a = 1.8\alpha + 2\delta$
N balance: $a = 0.2\alpha$
O balance: $6 + 2b = 0.5\alpha + 2\gamma + \delta$

$$RQ = \frac{\gamma}{b} = 0.85$$

Rearranging the above system of algebraic equations:

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1.8 & 0 & 2 & -3 & 0 \\ 0.2 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -0.85 \\ 0.5 & 2 & 1 & 0 & -2 \end{bmatrix} \begin{bmatrix} \alpha \\ \gamma \\ \delta \\ a \\ b \end{bmatrix} = \begin{bmatrix} 6 \\ 12 \\ 0 \\ 0 \\ 6 \end{bmatrix}$$

Solution to the above yields

$$\begin{bmatrix} \alpha \\ \gamma \\ \delta \\ a \\ b \end{bmatrix} = \begin{bmatrix} 6 \\ 12 \\ 0 \\ 0 \\ 6 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1.8 & 0 & 2 & -3 & 0 \\ 0.2 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -0.85 \\ 0.5 & 2 & 1 & 0 & -2 \end{bmatrix}^{-1}$$

Summary

In this chapter we introduced the idea of **growth reaction** to characterize material balance associated with cell growth. We defined a chemical formula which represents about 95% of the dry matter of biomass. We also defined yield which enables us to derive usful engineering information from literature articles that report cell yield. To carry out material balances around a fermentor requires, in general, respiratory quotient values and one other measurement.

Chapter 3 Thermodynamics of Cellular Growth

Microbial growth consists of a complex network of metabolic reactions. Coupled catabolic and anabolic reactions take place so that energy released in the former is efficiently used to drive the latter. However, some energy is always lost as heat. The purpose of this chapter is to quantify the heat release due to growth. In large-scale processes it is necessary to remove this heat so that the culture is maintained at physiological temperature. In small reactors metabolic heat is removed quite easily, while in very large fermentors (>10,000 liters) in which rapidly growing cells are cultivated, it is necessary to design adequate heat transfer area for heat removal. Bioreactor temperatures must be maintained within ± 0.5 C to maintain physiologic conditions conducive to optimal growth.

3-1 Heat Release due to Growth

Consider the growth reaction when no significant amount of extracellular product is formed. Under these conditions Eq (2-1) simplifies to

$$C_6 H_{12} O_6 + a\, NH_3 + b\, O_2 \rightarrow \alpha\, CH_{1.8}\, O_{0.5}\, N_{0.2} + \gamma CO_2 + \delta H_2O$$
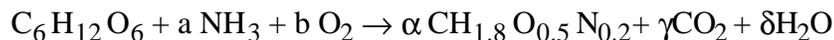
Since nitrogen consumption is usually small compared to the amount of carbon consumed, and that nitrogen does not go through oxidation (while C does !), we can approximate the above as,

$$C_6 H_{12} O_6 + b\, O_2 \rightarrow \alpha\, CH_{1.8}\, O_{0.5}\, N_{0.2} + \gamma CO_2 + \delta H_2O$$
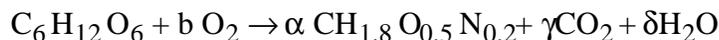
Consider heat balance of this reaction using one mole of glucose consumed as the basis.

Heat released = $(\alpha) \bullet [(MW\ biomass) \bullet (-\Delta H_c)] - (-\Delta H_s) \bullet (MW\ substrate)$ \hspace{1cm} (3-1)

where $^{(-\Delta H_c)}$ and $^{(-\Delta H_s)}$ are heat of combustion per gram of cell and per gram of substrate respectively. Rearranging,

$$\frac{Heat\ \ released}{MW\ \ substrate} = (\alpha) \bullet \left(\frac{MW\ \ biomass}{MW\ \ substrate} \right) \bullet (-\Delta H_c) - (-\Delta H_s) \hspace{1cm} (3-2)$$

The left hand side is the amount of heat released per gram of substrate consumed and the coefficient of the first term on the right is growth yield. That is,

$$Y_{\Delta/S} = (Y_{X/S}) \bullet (-\Delta H_c) - (-\Delta H_s) \hspace{1cm} (3-3)$$

where $Y_{\Delta/S}$ is "heat yield" on the basis of substrate consumed. Dividing the above by $Y_{X/S}$ gives

$$Y_{\Delta/x} = (-\Delta H_c) - \frac{(-\Delta H_s)}{(Y_{X/S})}$$

(3-4)

Both Eq (3-1) and (3-2) are useful in determining heat release due to growth, $Y_{\Delta/X}$ and substrate consumption, $Y_{\Delta/S}$.

3-2 Heat of Combustion Data

Heat of combustion has been reported by a number of authors and a few are given below in Table 3-1. The heat of combustion for a variety of organism falls in a very narrow range of about 22 kJ g^{-1}. Heat of combustion of glucose, commonly used substrate, is 15.6 kJ g^{-1}.

Table 3-1 Heat of Combustion of Biomass

Organism	$(-\Delta H_S)$, kJ g^{-1}
E. coli	23.03
E. cloacae	22.83
B. thuringiensis	22.08
Candida lipolytica	21.34
Candida boidinii	20.14
Kluyveromyces fragilis	21.66

3-3 Experimental Observations

Cooney and co-workers collected heat release experimental data for a number of different organisms by making careful heat balance measurements over a fermentor. The data show a linear relationship between heat released and oxygen consumption rate (Fig 2-1).
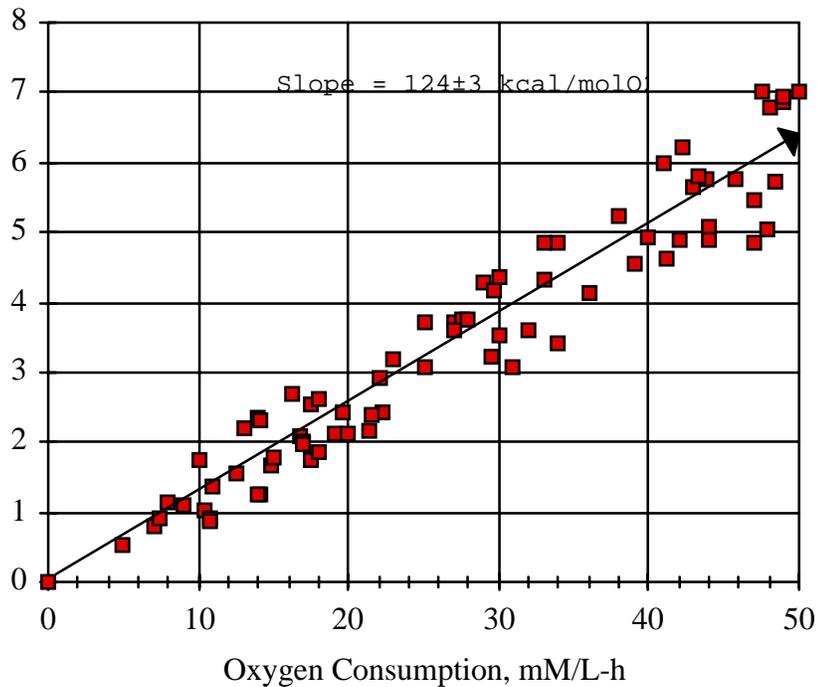
Figure 3-1 Experimental Heat Release Data (after Cooney et al)

The slope of the line above is equal to the amount of heat released per mol of oxygen consumed. That is:

$$Y_{\Delta/O2} = \frac{(124)}{(32)} \text{ kcal } (\text{g } O_2)^{-1} \Rightarrow 3.88 \text{ kcal } (\text{g } O_2)^{-1} \Rightarrow 16.21 \text{ kJ } (\text{g } O_2)^{-1}$$

This result enables one to calculate heat generation rate from oxygen uptake rate information. We will examine oxygen uptake rate data in Chapter 5.

3.4 Heat Release when Extracellular Products are Formed

When significant amount of product is present, Eq 3.1 will be modified to

Heat released = (α) * (MW biomass) * $(-\Delta H_C)$ + (β) * (MW product) * $(-\Delta H_P)$

$$\qquad\qquad\qquad\qquad\qquad - (-\Delta H_S) * (\text{MW substate}) \quad (3\text{-}5)$$

where $(-\Delta H_P)$ is heat of combustion per gram of extracellular product(s). Dividing the above by MW of substate gives

$$Y_{D/S} = (Y_{X/S}) * (-\Delta H_C) + (Y_{P/S}) * (-\Delta H_P) - (-\Delta H_S)$$
$$(3\text{-}6)$$

Dividing the above by $Y_{S/X}$ gives

$$Y_{\Delta/X} = (-\Delta H_C) + (Y_{P/X}) * (-\Delta H_P) - \frac{(-\Delta H_S)}{(Y_{X/S})} \qquad (3\text{-}7)$$

Chapter 4 Kinetics of Growth and Product Formation

4.1 Growth Kinetics

If a viable inoculum is introduced into a medium that contains a carbon source, suitable nitrogen source, other nutrients necessary for growth, and physiologic temperature and pH are maintained, it will grow. The rate of biomass synthesis is proportional to biomass present. That is

$$r_X = \mu\, X \qquad\qquad\qquad (4\text{-}1)$$

where r_X is the amount of cells synthesized in $g\,L^{-1}\,h^{-1}$, X is cell concentration in $g\,L^{-1}$. The parameter μ is called specific growth rate, analogous to the specific rate constant in chemical reaction rate expressions. Recall the treatment of chemical reactions, summarized below for ease of reference.

Reaction: $\qquad\qquad A \rightarrow B$
Rate Expression: $\qquad -r_A = k\, C_A$
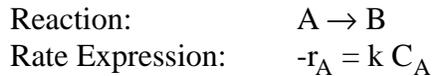
In the above C_A is concentration of A ($mol\,A\,L^{-1}$), $-r_A$ is reaction rate ($mol\,A\,L^{-1}h^{-1}$) and k is rate constant (h^{-1}). The negative sign in front of $-r_A$ is to comply with the definition of r_A, which is the rate of generation of A. In Eq(4-1), the negative sign is not necessary as X increases with time.

Consider cell balance over a batch bioreactor:

Cells in - cells out + Generation of Cells = Accumulation of cells in Bioreactor

$$0\ -\ 0\ +\ (r_X) * (V)\ =\ \frac{d(Vx)}{dt}$$

Substituting for r_X from Eq(4-1) and noting that volume of reactor is constant gives,

$$\frac{dX}{dt}\ =\ \mu\, X \qquad\qquad\qquad (4\text{-}2)$$

The above can be expressed as

$$\mu\ =\ \frac{1}{X}\frac{dX}{dt}\ =\ \left(\frac{\Delta X}{X}\right) \bullet \left(\frac{1}{\Delta t}\right)$$

The term, $\Delta X/X$, represents fractional increase in cell amount and Δt is the time over which the fractional increase was accomplished. That is, μ can be interpreted as fraction of biomass formed per unit time. For example if μ is $0.3\ h^{-1}$, every hour the biomass will approximately increase by 30%. We use

the term "approximately" because we are using finite quantities to describe the rule which applied at infinitesimal scale.

Treating μ as a constant for now, Eq (4-1) can be integrated to give

$$X = X_0 \; Exp(\mu t) \tag{4-3}$$

where X_0 is the initial (inoculum) cell concentration. The time, t, refers to the time since the inoculum emerged from lag phase. Eq(4-3) can be rearranged setting the conditions for doubling of biomass. That is $\left(\dfrac{X}{X_0}\right)$ = 2 and t = the doubling times, t_d .

$$t_d \;=\; \left(\frac{\ln(2)}{\mu}\right) \;=\; \left(\frac{0.693}{\mu}\right) \tag{4-4}$$

The values of doubling time and specific growth rate have been reported by many researchers. Given below is a sample of typical values.

Organism	Growth Rate, μ [h⁻¹]	Doubling time, t_d [h]
E. coli	2.0	0.35
Yeast	0.3	2.3
Hybridoma	0.05	13.9
Insect Cells	0.06	11.6

4.2 What Does μ Depend on?

Specific growth rate (μ) depends on a number of factors such as growth medium composition, temperature, pH and others. Experimental studies have shown that one cannot increase growth rate beyond a certain maximum value, μ_m due to inherent metabolic reaction rate limitations. In general, when substrate, S is limiting growth, Monod (1949) reported that growth rate variations can be expressed as

$$\mu \;=\; \frac{\mu_m \; S}{K_S \;+\; S} \tag{4-5}$$

where K_S is called Monod constant or simply the substrate saturation constant. The significance of K_S is, when substrate concentration is numerically equal to K_S, growth rate is exactly half of maximum growth rate. See Figure 4-1.
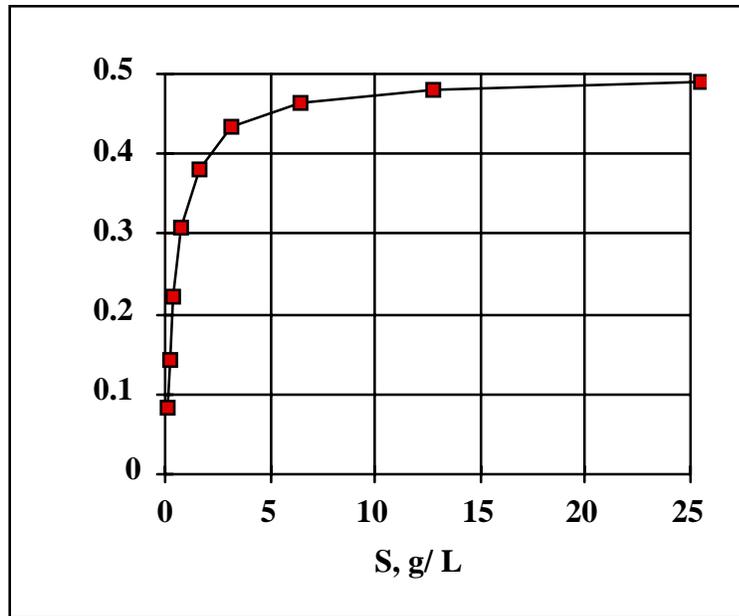
Figure 4-1 Monod Kinetics. Dependence of Growth Rate on Limiting Substrate. Specific growth rate reaches a maximum value of 0.5 h^{-1}. Value of K_S here is 0.5 g L^{-1}. Note that when S = 0.5 g L^{-1}, μ is half of its maximum.

The form of Eq(4-5) can be used to describe dependence of μ on more than one limiting nutrient. In many practical applications availability of oxygen for respiration often limits growth. When both substrate, S, and dissolved oxygen concentration, C_{DO}, are both limiting growth, specific growth rate can be mathematically described as

$$\mu = \frac{\mu_m \, S}{K_S + S} \cdot \frac{C_{DO}}{K_{DO} + C_{DO}} \qquad\qquad (4\text{-}6)$$

Figure 4-2 illustrates the behavior of maximum growth rate when two substrates are limiting. The parameters K_S and K_{DO} are cell specific. K_S is typically in the order of 10 mg/L for glucose and K_{DO} is less than 1 mg/L for oxygen in the case of bacteria and yeast. K_{DO} has been reported to be higher for mammalian and insect cells.
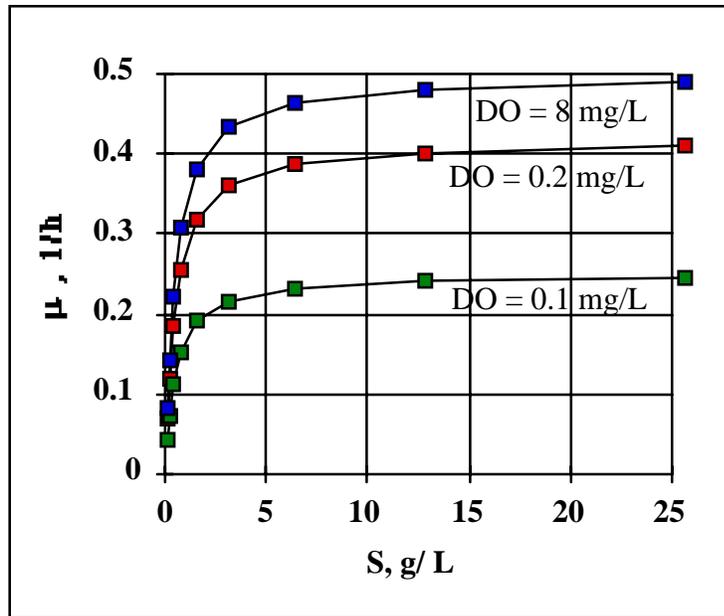
Figure 4-2 Monod Kinetics when two substrates are limiting. Specific Growth Rate reaches a maximum value of 0.5 h^{-1}. Value of K_S here is 0.5 g L^{-1}. Value of K_{DO} is 0.1 mg L^{-1} Note that when $C_{DO} = 0.1$ mg L^{-1}, μ is half of its maximum at values of S >> K_S.

Let us now consider growth under conditions of only substrate limitations in a batch bioreactor. Incorporating the substrate limited condition, bioreactor material balance equation, Eq(4-2), can be modified and we may write,

$$\frac{dX}{dt} = \frac{\mu_m S}{K_S + S} X \tag{4-7}$$

In order to integrate the above, one of the variables, S, needs to be replaced in terms of X. The yield relationship, Eq(2-3), can be integrated as

$$\int_{X_0}^{X} dX = - Y_{X/S} \int_{S_0}^{S} dS$$

which simplifies to

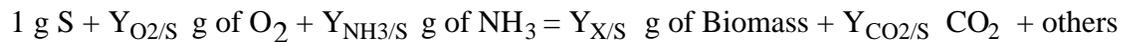$$S = S_0 - \left(\frac{X - X_0}{Y_{X/S}} \right) \tag{4-8}$$

where subscript, 0 refers to initial concentration. Substituting for S from Eq(4-8) in Eq(4-7) and integrating gives,

$$\left(\frac{K_S Y_{X/S} + S_0 Y_{X/S} + X_0}{Y_{X/S} S_0 + X_0} \right) \ln \left(\frac{X}{X_0} \right) - \left(\frac{K_S Y_{X/S}}{Y_{X/S} S_0 + X_0} \right) \ln \left(\frac{Y_{X/S} S_0 + X_0 + X}{Y_{X/S} S_0} \right) = \mu_m t$$

For analyzing batch systems, use the above to calculate cell concentration and then calculate substrate concentration using Eq(4-8).

4.3 Rate Expression and Metabolic Quotient

We have already discussed rate expressions for cell growth, Eq (4-1). Let us now examine rate expressions for other medium components in the **growth reaction**, Eq (2-1). Consider the growth reaction on the basis of <u>one g of substrate</u> consumed. It can be written as ,

$$1 \text{ g S} + Y_{O2/S} \text{ g of } O_2 + Y_{NH3/S} \text{ g of } NH_3 = Y_{X/S} \text{ g of Biomass} + Y_{CO2/S} \ CO_2 + \text{others}$$

The stoichiometric coefficients in growth reaction become yield coefficients on the basis of substrate. See Example 2-2. The general rate expression is then:

$$\frac{r_S}{-1} = \frac{r_{O2}}{-Y_{O2/S}} = \frac{r_{NH3}}{-Y_{NH3/S}} = \frac{r_X}{Y_{X/S}} = \frac{r_{CO2}}{Y_{CO2/S}} \qquad (4\text{-}9)$$

where r_i is expressed in g of i L^{-1} h^{-1}. Since r_x is the most fundamental of the various rates, it is conventional to write the stoichiometric coefficient in terms of it. That is

$$r_S = \frac{r_X}{-Y_{X/S}}$$

$$r_{O2} = r_X \frac{Y_{O2/S}}{Y_{X/S}} \Rightarrow \frac{r_X}{Y_{X/O2}} \qquad (4\text{-}10)$$

Following the examples above, the rate expression for species i can be written as

$$r_i = \frac{r_X}{Y_{X/i}} \qquad (4\text{-}11)$$

Metabolic quotients are rate expressions on the basis of unit mass of biomass. That is

$$q_i = \frac{r_i}{X} \Rightarrow \left(\frac{r_X}{Y_{X/i}} \right) \cdot \left(\frac{1}{X} \right) \Rightarrow \frac{\mu}{Y_{X/i}} \qquad (4\text{-}12)$$

The metabolic quotient for oxygen is of special interest. This single property determines the upper limit of cell concentration that can be achieved in many bacterial fermentation systems. We will see further analysis in the next chapter. Typical values of metabolic coefficients are given below.

Organism	$q_{glucose}$ g g^{-1} h^{-1}	q_{O2} g g^{-1} h^{-1}
E. coli	2.5	0.3
Yeast	0.5	0.2
Hybridoma	0.2	0.02

Example 4-0

If specific growth rate of a bacteria is 0.35 h-1 and cell yield is 0.6, calculate glucose consumption rate.

$$q_G = \frac{\mu}{Y_{X/G}} = \frac{0.35}{Y_{X/G}} = 0.48 \text{ g G (g Cell)}^{-1} \text{ h}^{-1}$$

4.4 Factors Affecting Growth Rate

Nutrients in the medium, pH, temperature, dissolved oxygen concentration and other cultivation environmental conditions all affect growth rate. Temperature and pH dependence are illustrated in Fig 4-3 a and b. In Figure 4-3 a the maximum growth rate is observed at 39 C for E. coli. Product formation kinetics (for example insulin), product yield ($Y_{P/S}$), cell yield ($Y_{X/S}$) are also affected by temperature. In general, cell yield decreases with temperature while similar defining relationships for product has not been reported. It is important to note that the optimum temperature for growth may be different from that for product formation.
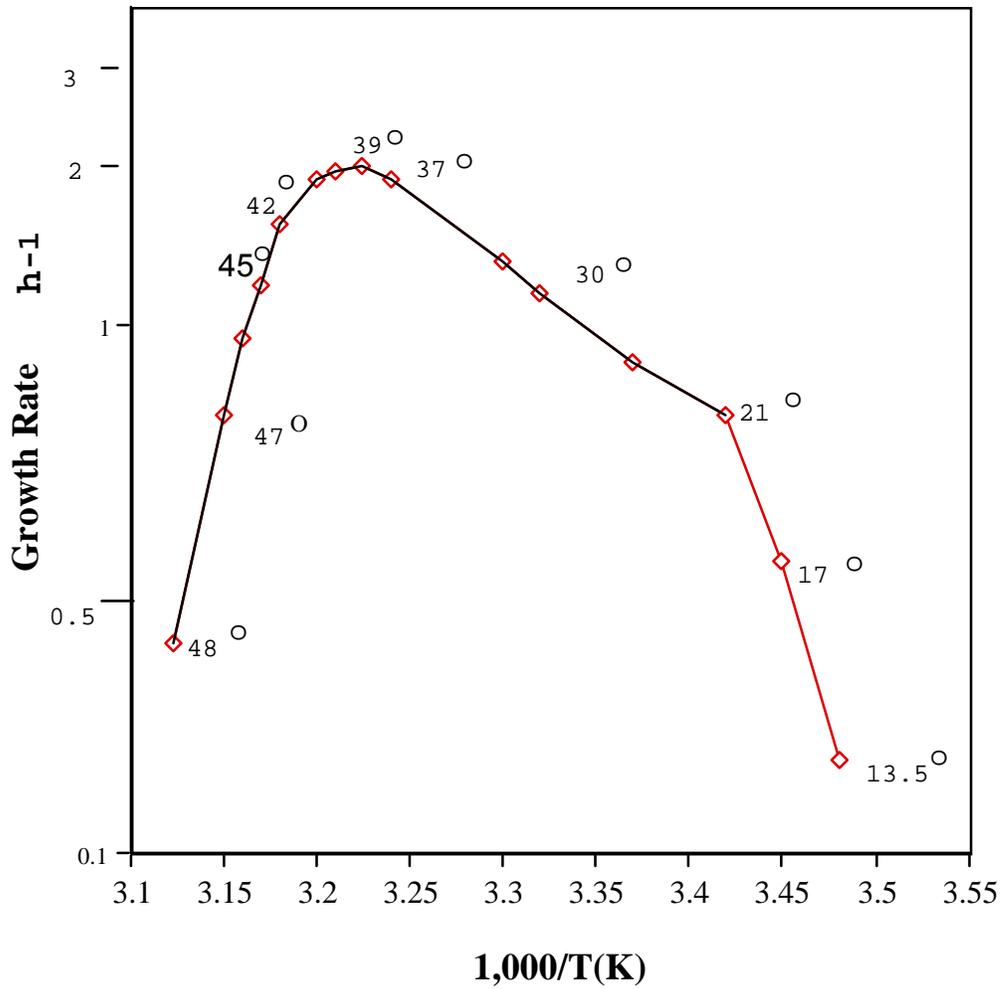
Figure 4-3 a Effect of Temperature on Growth Rate of E. coli. Maximum growth rate is at 39 C. Plot is given as a function of inverse absolute temperature. The declining line from 39 C to 21 and then to 13 C suggest that the growth rate constant behaves somewhat similar to chemical reaction rate constant.
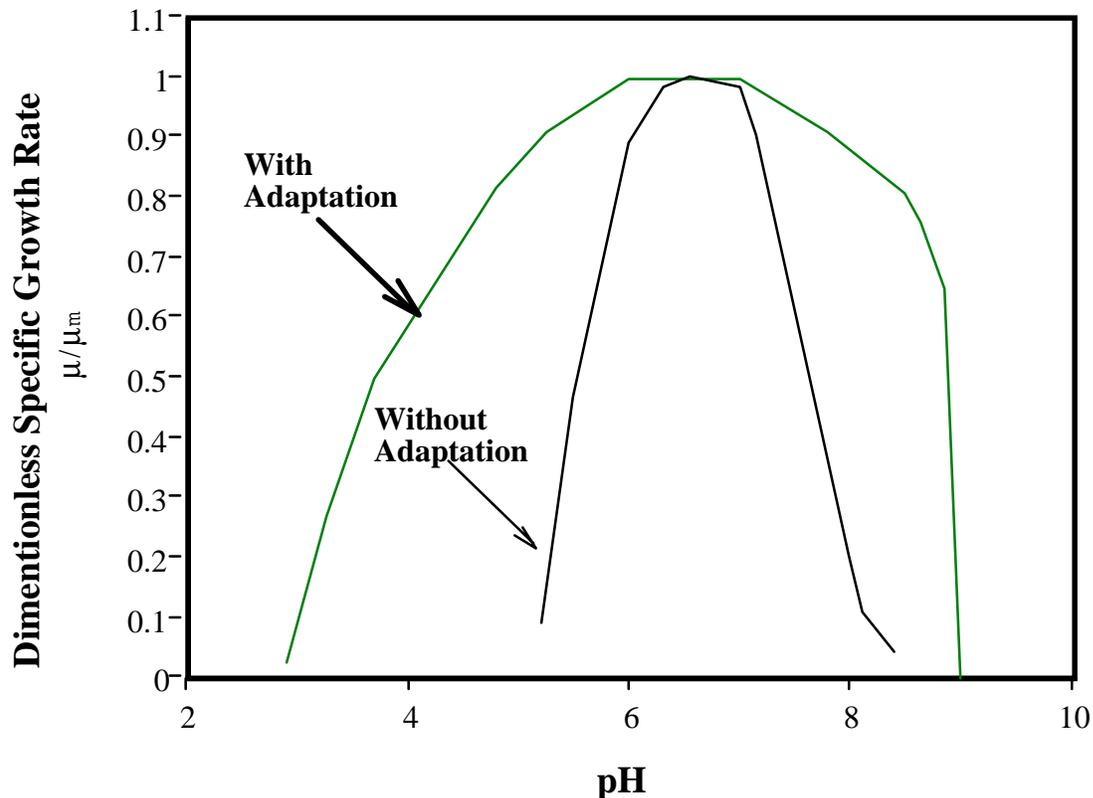
Figure 4-3b Effect of pH on Growth rate. Typical pH ranges over which reasonable growth can be expected is about 1 to 2 units. With adaptation, broader ranges can be achieved.

Optimum pH values for growth range from 4 to 7 for bacteria, from 4 to 7 for yeast and 6.2 to 7.2 for animal cells. Optimum pH for product formation may be different from that for growth. Many bacteria produce a different mix of products when pH is altered. For example, *Clostridium butylicum* produces acetic and butyric acids at near neutral pH while butanol, acetone and ethanol are produced under acidic pH (biological equivalent of Le Chatelier's Principle!). However, in the case of a recombinant cell expressing a recombinant protein, pH usually affects kinetics of recombinant protein generation rather than the product mix. Hybridomas are known to produce antibodies at a higher rate at pH 6.2 than at 7.2. Because of the difference in conditions for growth and product formation, optimization is often necessary.

Oxygen is an important substrate for aerobic organisms. Since metabolic energy production by cells is directly related to oxygen uptake rate (also called respiration rate), oxygen concentration is very strongly coupled to growth rate. As illustrated in Fig 4 - 4 , growth
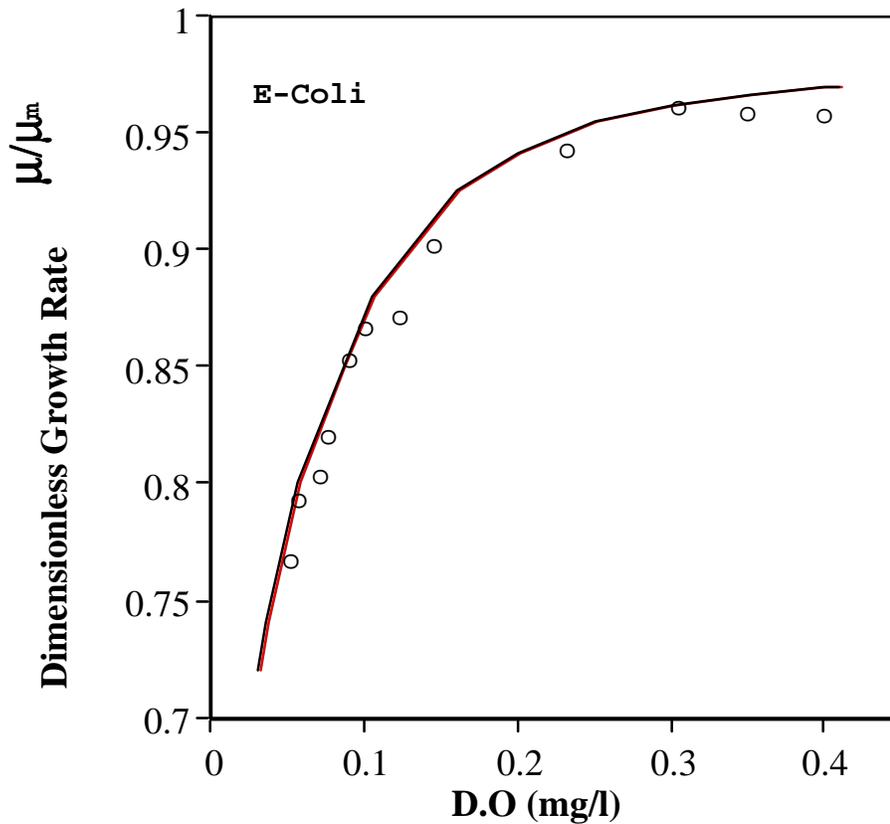
Figure 4-4 Growth Rate depends on dissolved oxygen concentration. The critical dissolved oxygen concentration refers to value of DO below which growth rate is lower than the maximum value.

rate sharply rises to its maximum value with dissolved oxygen concentration. The relationship is similar to the behavior we discussed. See also Figure 4-1. The concentration at which maximum growth rate is attained is often referred to as critical oxygen concentration, C_{O2}^{CRIT}. This value is typically less than 0.5 mg L^{-1} for bacteria and yeast, and about 1 to 2 mg L^{-1} for animal and insect cells. Note that these values are significantly lower than air saturation value of 6.7 mg L^{-1} at 37 C.

4.5 Product Formation Kinetics

Product formation kinetics fall into one of the following three types.
    I. Growth Associated Product Formation
    II. Non-Growth Associated Product Formation
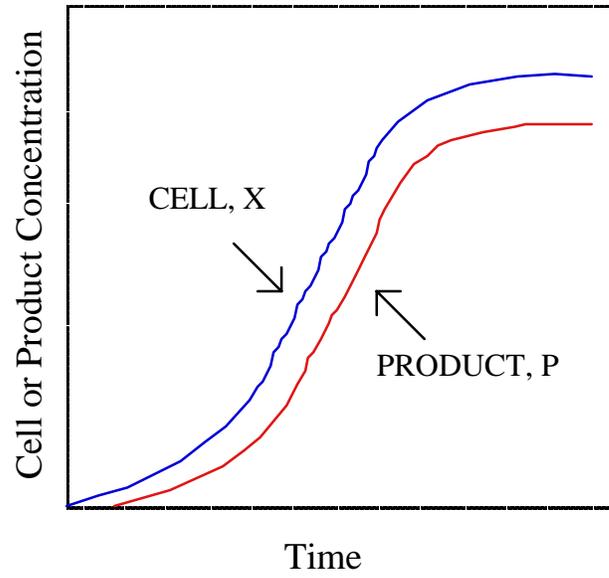   III. Mixed Mode Product Formation

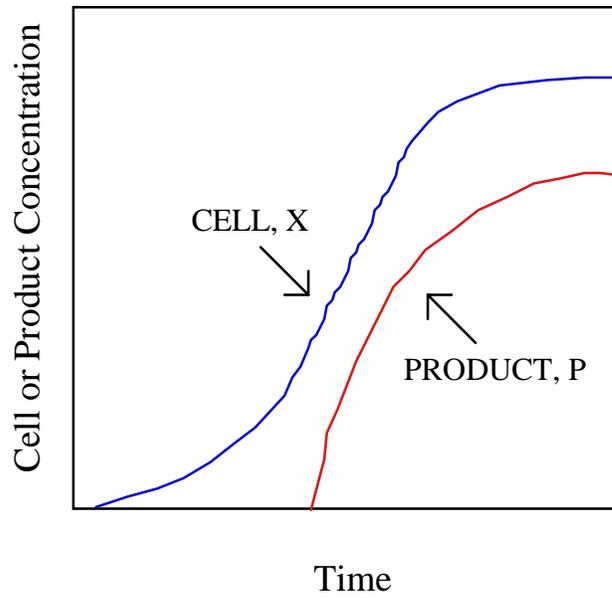**Figure 4-5a** Growth Associated Product Formation

Figure 4-5b Non-Growth Associated Product Formation
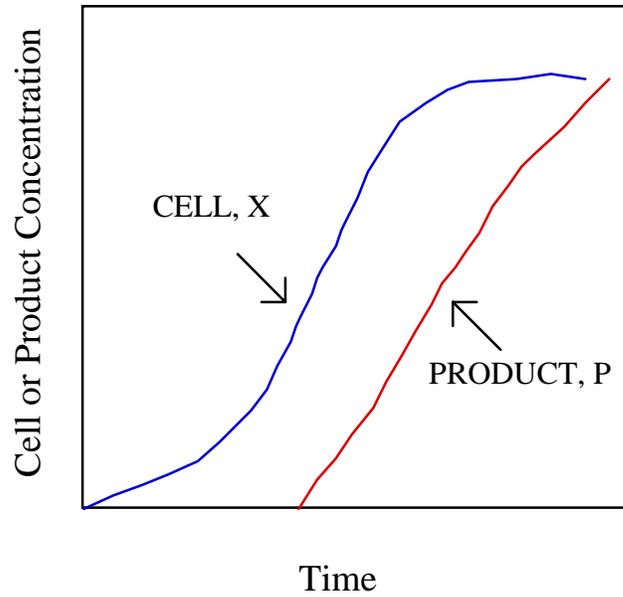
Figure 4-5c Mixed Mode Product Formation

Typical time-profiles of these three cases are illustrated above. In Type I shown in Fig 4-5a, product is formed simultaneously with growth of cells. That is product concentration increases with cell concentration. The metabolic quotient for P can be expressed as a function of μ,

$$r_P = q_P X \Rightarrow \alpha \mu X$$
$$q_P = Y_{P/X} \mu$$

(4-13)

It is clear from the above, the proportionality constant, α is the yield coefficient, $Y_{P/X}$. Anaerobic fermentation of sugars by *Saccharomyces cerevisae* is an example of Type I. Illustrated below are actual data for this bioprocess.

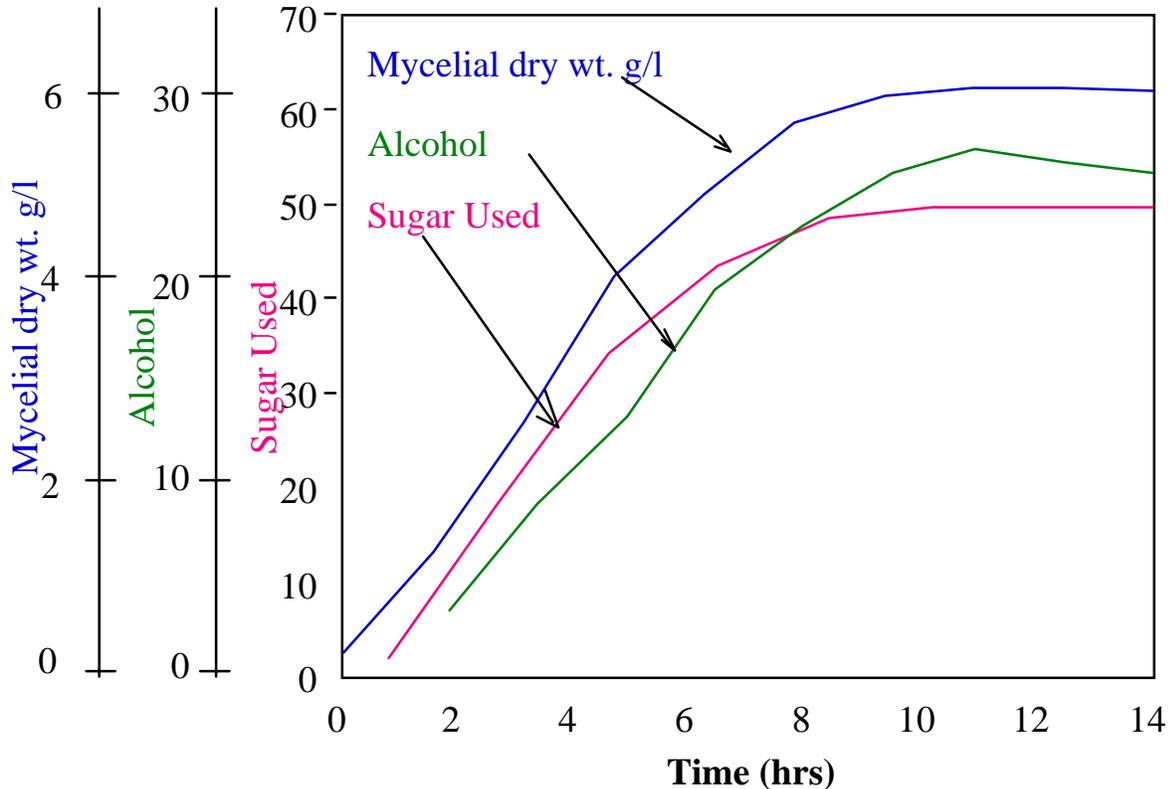**Figure 4-5** Ethanol Fermentation data for yeast illustrates Type I product formation kinetics. Note that formation of alcohol is proportional to cell concentration.

In Type II, product formation is unrelated to growth rate, but is a function of cell concentration. This is expressed as

$$r_P = q_P X \Rightarrow \beta X \qquad (4\text{-}14)$$

Antibody formation by hybridoma, and some antibiotic fermentation exhibit this type of behavior.

In the third category, product formation is a combination of growth rate and cell concentration. That is,

$$r_P = q_P X \Rightarrow (\alpha \mu + \beta) X \qquad (4\text{-}15)$$

Many biochemical processes fall into this category. Note that if β is zero and α is $Y_{P/X}$, this case reduces to Type I. If $\alpha = 0$, it reduces to non-growth associated case. Therefore let us consider this more general case for further analysis.

In a batch reactor, product accumulation can be obtained by carrying out mass balance on the product.

Rate of Product Formation = Accumulation of Product

$$(r_P) \cdot (V) = \frac{d(V \; P)}{dt}$$

For constant V,

$$\frac{dP}{dt} = r_P = (\alpha \mu + \beta) \; X$$

If we consider exponential phase only, $X = X_0 \; Exp(\mu_m t)$. That is, substituting in the above gives

$$\frac{dP}{dt} = (\alpha \mu + \beta) \; X_0 \; Exp \; (\mu_m t)$$

Integrating from $t = 0$, $P = P_0$ we get

$$P - P_0 = \frac{(\alpha \mu + \beta)}{\mu_m} \; X_0 \; (Exp \; (\mu_m t) - 1)$$

The above expression can be used to calculate the amount of product concentration at the end of a growth cycle.

Example 4-1

McCallion reported growth of Thermoanaerobacter ethanolicus under controlled pH of 7.0. Using appropriate graphs calculate $Y_{X/G}$ and $Y_{LA/G}$, where G and LA refer to glucose and lactate. Is lactic acid formation growth associated ? Can you estimate an approximate value for $q_{glucose}$?

Time [h]	Glucose [g/L]	Lactate (LA) g/L	Cell (X) g/L
0	19.50	0.45	0.01
13	16.88	3.88	0.41
14	14.85	4.94	0.54
16	13.11	6.98	0.92
18	10.40	8.98	0.99
19	8.91	10.30	1.05
20	7.75	10.83	1.15
22	5.18	12.57	1.30
24	3.64	14.58	1.35
37	0.25	16.03	0.69

Solution

First plot cell concentration versus time. The slope in the exponential growth phase is approximately 0.24 h^{-1}. Notice that the culture growth slows down shortly after 15 h. One could also analyze the information numerically. Such an approach unfortunately will lack a good overview of the phases of growth.

Now plot LA vs X and S vs X.
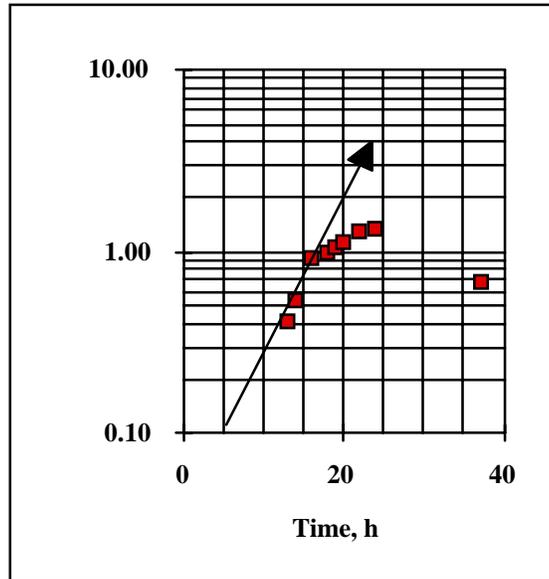
Plot above shows that the exponential behavior deviates at t = 16 h. A straight line drawn during the exponential phase gives growth rate.
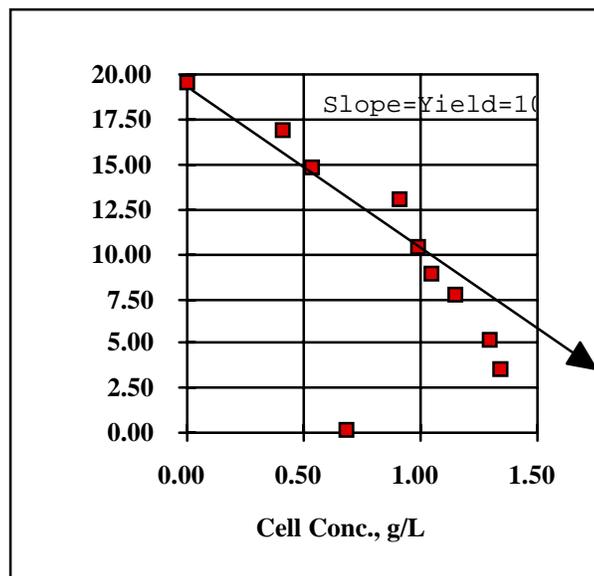
Except for one point, all others lie nearly on a straight line. The data at 0.7 g/L of cell corresponds to the declining phase of growth, which may be ignored for current analysis. The line has a negative slope because glucose decreases as cell concentration increases.
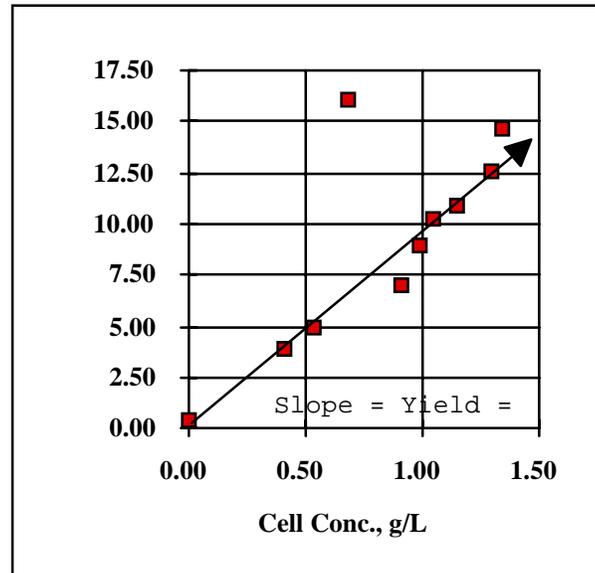
Yields from the above graphs:

$Y_{G/X} = 10$ and $Y_{LA/X} = 9.8$

Required yield of $Y_{X/G} = [Y_{G/X}]^{-1} = 0.1$ g cell/g glucose

$Y_{LA/G} = Y_{LA/X} [Y_{G/X}]^{-1} = (9.8) \cdot (0.1) = 0.98$ g lactate/g glucose

$$q_{glucose} = \frac{\mu}{Y_{X/G}} = \frac{0.24}{0.1} \Rightarrow 2.4 \; \frac{g \; glucose}{(g \; cell) \; (h)}$$

Chapter 5 Oxygen Transfer in Bioreactors

Oxygen is needed by cells for respiration. Oxygen used by cells in suspension must be available as dissolved oxygen. Since oxygen solubility is quite small, about 6 to 7 mg/L under normal cultivation conditions, metabolic oxygen requirement is supplied on a as needed basis by continuous aeration of culture medium. Actively respiring yeast requires about 0.15 g O_2 (g cell)$^{-1}$ h. At a cell concentration of 10 gL^{-1}, medium saturated with air can support less than 30 seconds worth of metabolic oxygen. That is, a continuous supply of oxygen must be maintained in any viable aerobic manufacturing process. In this Chapter, we will first get a quantitative appreciation for metabolic oxygen demand, followed by methods used in calculating rates at which oxygen is transfered from sparged air. We will then examine methods useful in characterizing oxygen mass transfer coefficient. Finally we will evaluate bioreactor operation and design based on oxygen transfer capability.

5.1 Metabolic Oxygen Demand

Metabolic oxygen demand of an organism depends on the biochemical nature of the cell and cultivation conditions. Oxygen need is usually satisfied in most cells if the dissolved oxygen concentraiton in the medium is kept at about 1 mg/L. If the oxygen level is allowed to fall far below this value, oxygen consumption rate decreases with concomitant decrease in biochemical energy production, and as a result cell growth rate also decreases. We described this behavior in Section 4-4.   The value of oxygen concentration above which growth rate is at the maximum was described as the critical oxygen concentration, C_{O2}^{CRIT} . Characteristic values are summarized in Table 5-1.

<div align="center">

Table 5-1 Critical Oxygen Concentration

</div>

Organism	C_{O2}^{CRIT} in mg L^{-1}
E. coli at 37 C	0.26
S. cerevisiae at 30 C	0.13
Penicillium sp at 24 C	0.78

The oxygen requirement for growth is expressed best in the the parameter, yield coefficient, $Y_{X/O2}$. It represents the amount of oxygen required to grow one gram of cells. Typical values summarized in Table 5-2, show that approximately 0.7 to 1 g of oxygen is needed to produce 1 g of cells. In the same table respiration quotient is also included.

Table 5-2 Stoichiometric Oxygen Demand &Respiration Rate

Organism	Substrate	$Y_{X/O2}$ g (g cell)$^{-1}$ h	q_{O2} g O_2 (g cell)$^{-1}$ h
E. coli	Glucose	1.1	0.20
S.cerevisae	Glucose	0.98	0.30
Candida utilis	Glucose	1.32	
Pennicillium sp.	Glucose	1.35	0.18
Hybridoma			
CHO cell line			

5.2 Volumetric Oxygen Mass Transfer Coefficient

In a typical aeration system, oxygen from the air bubble is transferred through the gas-liquid interface followed by liquid phase diffusion/bulk transport to the cells. Although this is a multi-step serial transport, in a well dispersed systems, the major resistance to oxygen transfer is in the liquid film surrounding the gas bubble. Consider the oxygen concentration profiles in the region near the interface illustrated in Figure 5-1.
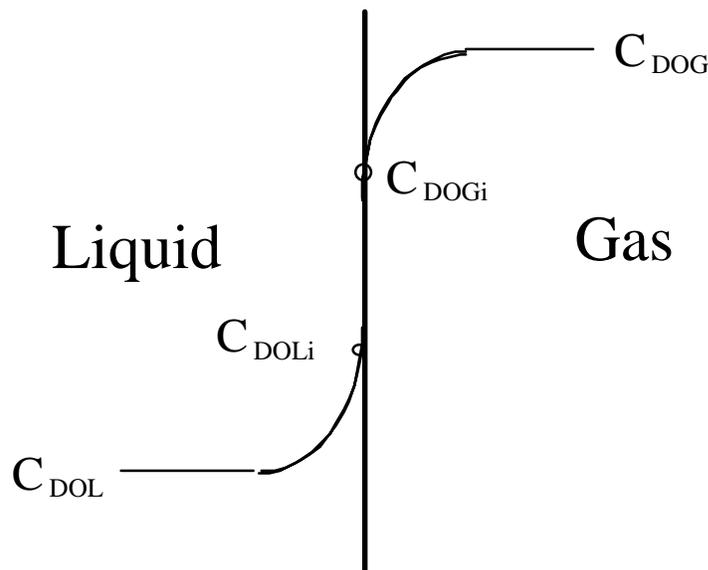

Figure 5-1 Oxygen Concentration Profile at Air Bubble-Medium Interface

The transport of oxygen through the gas and liquid films are equal at steady state. They can be expressed by

$$N_{O2G} = k_G A (C_{DOG} - C_{DOGi}) \qquad (5\text{-}1a)$$

$$N_{O2L} = k_L A (C_{DOLi} - C_{DOL}) \qquad (5\text{-}1b)$$

$$N_{O2G} = N_{O2L} \qquad (5\text{-}1c)$$

where subscript G and L refer to gas and liquid phases respectively. The terms, N_{O2G} and N_{O2L} are oxygen transfer expressed in g O_2 h^{-1}, A is interfacial area and C_{DO} is oxygen concentration expressed in g O_2 per unit volume. At the interface, equilibrium between the liquid and gas phase oxygen is reached. That is

$$C_{DOGi} = m\, C_{DOLi} \qquad\qquad (5\text{-}2)$$

Because of low oxygen solubility and the fact that k_G is much higher than k_L,

$$C_{DOG} \approx C_{DOGi} \qquad\qquad (5\text{-}3)$$

Hence, Eq (5-1a) can be written as

$$N_{O2} = k_L\, A \left(\frac{C_{DOG}}{m} - C_{DOL} \right) \qquad\qquad (5\text{-}4)$$

The subscript L in N_{O2} has been dropped to note that the above represents overall transfer of oxygen. The driving force in the above consists of the difference between bulk oxygen concentrations in the two phases; the first term represents the concentration of oxygen in the liquid which is in equilibrium with the bulk gas phase oxygen. If air is the gas medium, this term will equal to 7 mg/L at 35 C.

When the above oxygen transfer is applied to an entire volume of a bioreactor, A will represent the total interfacial area and k_L will represent an average mass transfer coefficient. The concentrations will be bulk gas and liquid phase oxygen concentrations. If we divide the above equation by volume of liquid phase, V, the resulting term will represent the amount of oxygen transfered per unit volume per unit time --- which is in the same units as the rate expressions we saw in last chapter. Since the rate is due to a physical phenomena, let us distinguish it by the symbol, R_{O2}. That is,

$$R_{O2} = k_L \left(\frac{A}{V} \right) \left(\frac{C_{DOG}}{m} - C_{DOL} \right) \qquad\qquad (5\text{-}5)$$

The term, k_L A represents the product of mass transfer coefficient and interfacial area available for mass treansfer. In a bioreactor, air is sparged and the liquid is agitated to break up the bubbles so that interfacial area can be kept high to enhance rate of oxygen transfer. In such systems, the area, A, is not easily measured or estimated. But, the term consisting of the product - mass transfer coefficient and interfacial area - is more readily measured. Further more, it is convenient to use interfacial area per unit volume, a, rather than total area, A because rate of oxygen transfer is expressed per unit volume of bioreactor, similar to rate of cell growth, which is reported on a volumetric basis. Hence, area per unit volume, a, is combined with the mass transfer coefficient, k_L and is

given by the term, k_{La}. In Eq(5-5) the term, $\dfrac{C_{DOG}}{m}$ can be replaced by oxygen solubility at bioreactor conditions, C^*_{DOL} .

$$R_{O2} = k_{La} \left(C^*_{DOL} - C_{DOL} \right) \tag{5-5}$$

The above will be our working equation for describing transfer of oxygen from gas phase to growth medium. In order for us to calculate oxygen transfer rate (OTR), we need the mass transfer coefficient, k_{La} , solubility of oxygen in the medium, C^*_{DOL} and the dissolved oxygen concentration in the medium, C_{DOL}. In the last chapter we had used the notation, C_{DO} to describe dissolved oxygen concentration. In the discussion above, there was a need to make a distinction between gas and liquid phase concentration. In Eq (5-5), one notes that both concentrations are expressed on the basis of liquid phase. Hence, from here on we will drop the subscript L. In situations where we need to make a distinction between the two phases, we will re-introduce the subscript L and G.

5.3 Bioreactor Oxygen Balance

Let us now consider the case of oxygen balance within a bioreactor in which cells are growing and in the process consuming oxygen. There is a continuous inflow of air at a constant volumetric flow rate. The liquid broth is agitated by a Rushton agitator (flat blade stirrer). Le the metabolic oxygen uptake rate be q_{O2} and cell concentration is X. Let us examine the reactor system over a sufficiently short period that we can treat X as a constant. Consider oxygen balance over the liquid phase of the bioreactor.

O_2 transfered from Gas Phase - O_2 consumed by Cells = Accumulation

$$\left[k_{La} \left(C^*_{DO} - C_{DO} \right) \right] \cdot V - q_{O2} \, X \, V \;=\; \frac{d(V \, C_{DO})}{dt} \tag{5-6}$$

For constant liquid phase volume, the above can be simplified to

$$\frac{d(C_{DO})}{dt} = k_{La} \left(C^*_{DO} - C_{DO} \right) - q_{O2} \, X \tag{5-7}$$

The concentration, C_{DO} is readily measured using an dissolved oxygen electrode. A later segment of the course on Biosensors, will deal with principle of measurement and construction of DO electrodes.

If oxygen being supplied is in exact balance with the oxygen consumed by the cells, we expect the dissolved oxygen concentration to remain constant; that is, the derivative in Eq(5-7) will vanish. That is,

$$q_{O2} X = k_{La} \left(C^*_{DO} - C_{DO} \right) \qquad (5\text{-}8)$$

One useful application of the above is in estimating the maximum cell concentration a particular bioreactor is capable of supporting in terms of oxygen supply. See the example below.

Example 5-1.

A bioreactor has an oxygen mass transfer coefficient capability of 400 h[-1]. What is the maximum concentration of E. coli that can be grown aerobically in this reactor. Respiration rate of E. coli is 0.35 g O_2 (g Cell)[-1] h[-1]. Critical oxygen concentration is 0.2 mg/L. Assume oxygen saturation with air to be 6.7 mg/L.

Solution

From Eq(5-8), we have

$$X = \frac{k_{La} \left(C^*_{DO} - C_{DO} \right)}{q_{O2}}$$

The maximum oxygen concentration driving force that can be expected is

$$= (6.7 - 0.2) = 6.5 \text{ mg/L}.$$

Therefore, maximum cell concentration that can be grown at maximum growth rate is

$$X_{max} = \frac{k_{La} \left(C^*_{DO} - C_{DO} \right)_{max}}{q_{O2}} \Rightarrow \frac{(400 \text{ h}^{-1}) \bullet (6.5 \text{ mgO2 L}^{-1})}{0.35 \text{ gO2 (gCell)}^{-1} \text{h}^{-1}} \Rightarrow 7.4 \text{ gCell L}^{-1}$$

5.4 Factors Affecting K$_{la}$

The mass transfer coefficient is strongly affected by agitation speed and air flow rate. In general,

$$k_{La.} = k \, (Pg/V_R)^{0.4} \, (V_S)^{0.5} \, (N)^{0.5}$$

where k is a constant
Pg is power required for aerated bioreactor

V_R is bioreactor volume

V_S is air flow rate

N is agitator speed

5.5 Measurement of K_{La}

Most common method of measuring k_{La} is to conduct experiments in the bioreactor when cells are absent, or cell concentration is low so that consumption by cells can be neglected. The latter condition is present immediately after inoculating the bioreactor. Consider Eq (5-7) under these conditions:

$$\frac{d(C_{DO})}{dt} = k_{La} \left(C_{DO}^{*} - C_{DO} \right)$$

If we allow steady state to occur, the dissolved oxygen concentration will reach saturation value, C_{DO}^{*} and the concentration-time profile will be flat, as shown in the diagram.
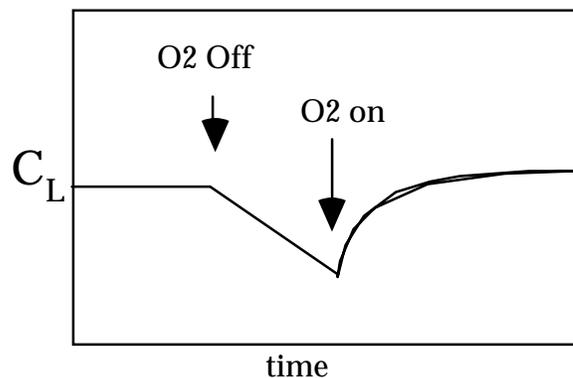

Fig 5-2 Oxygen Profile During a Transient. The responses will be exponential, rather than straight lines.

 If the oxygen source (air) is replaced by nitrogen, the resulting response of the system is described by the above equation with the term, C_{DO}^{*} set to zero. That is,

$$\frac{d(C_{DO})}{dt} = k_{La} \left(0 - C_{DO} \right) \quad \text{and} \quad C_{DO}(t=0) = C_{DO}^{*}$$

The solution to the above is

$$C_{DO} = C_{DO}^{*} \, \text{Exp} \left(- k_{La} t \right)$$

If one plots the response on a semi-log plot, the slope will equal to the negative of mass transfer coefficient. It is relatively a simple experiment and the data analysis is also easy to do. When other type of transient mass transfer experiments are conducted, the above equations should be suitably modified. For example for the case of nitrogen to air switch, we should suitably modify the solution because the initial condition is now different.

5.6 Case Studies

Example 5.2

You are part of a tech service team asked to evaluate if the available 10,000 liter fermentor is adequate to produce 10 kg/day of a recombinant protein using a strain of E. coli that expresses the protein as 20 % of cellular protein. In order to enhance plasmid stability, the nutrients are manipulated to give a low specific growth rate is 0.2 h^{-1}. The oxygen demand is 0.15 g O_2/g cell - h. Assume that the r-protein formation is cell growth associated.

Data: The lag phase is 4 hours. Typical clean-up time following a fermentation batch and preparation for the next batch is 8 hours. The plant runs three shifts. Cell yield on substrate is 0.55 g cell/g substrate. Available support services can supply inoculum of a maximum of 6 kg of cells every 24 hour period. Maximum K_{La} for the available fermentor is 500 h^{-1}. Fermentor accessories are capable of handling cell concentrations of 60 g/L. Assume any other parameters you need to complete the calculation.

Assumption: Critical oxygen conc. is 0.2 mg/L and DO at air saturation is 6.4 mg/L

$$\text{Max. Oxygen Transfer Rate} = K_{La}(C_{DO}^{\bullet} - C_{DO}^{CRIT}) = (500) \cdot (6.4 - 0.2) \cdot 10^{-3}$$
$$\text{g L}^{-1}\text{ h}^{-1}$$

$$\text{Therefore, max. cell conc. sustainable} = \frac{\text{Max OTR}}{q_{O2}} = \frac{3.1}{0.15} = 20.6 \text{ g/L}$$

Solution A: Lag phase and clean-up/ prep time is given as 12 h. If a batch is to completed within each 24 h period, production is limited to 12 h per day. If this is not a limitation, one can optimize production by varying batch time. Let us first evaluate assuming 12 h batch times.

If max. cell concentration of 20.6 g/L is obtained, amount of r-protein produced is = (0.2) (0.5) (20.6) = 2.06 g/L. 50% of cell dry matter was assumed to be protein. Hence in 10,000 liters, we will produce 20.6 kg.

Next to determine the inoculum level. The maximum batch growth phase is 12 h. Substitute in growth eqn, and assuming nutrients are present to support exponential growth during the 12 h period,

$$X = X_0 \ Exp \ (\mu_m \ t)$$

$$(20.6) = X_0 \ Exp(\ (0.2) \ (12)) \qquad or \quad X_0 = 1.87 \ g/L$$

For 10,000 liters, we will need 18.7 kg every 12 h. Since only 6 kg is available, max. protein that can be produced is

$\{(0.2)(0.5)[0.6 \ Exp((0.2)(12)] \cdot 10,000 = 6.61 \ kg$

Solution B: Now let us allow batch times to be longer than 12 h, meaning that there might not be a harvest every day. Since it is advantageous to use the max. inoculum concentration, select X_0 = 0.6 g/L. This value is obtained by diving 6 kg of cells in 10,000 L. Max. cell concentration is fixed due to aeration requirements. Use the batch growth eqn to find the batch growth time of 17.7 h. Hence 20.6 kg or r-protein will be produced every 29.7 h which gives a 24 h production rate of 16.6 kg.

What alternative way of running reactor would you recommend to achieve the production target?

DRUG DELIVERY

**Engineering Biotechnology
Gateway Project**
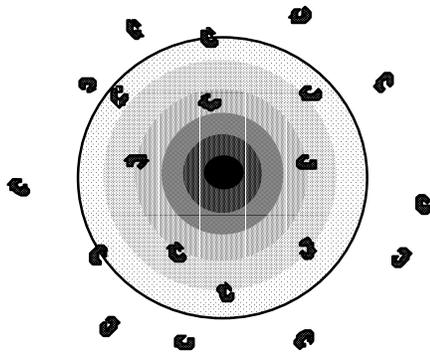
**Margaret A. Wheatley
Drexel University**

Chapter 1.   Overview

In order to address the important issues of drug delivery in the context of insulin, the subject will be divided into three modules:

1.1.   Protein and Hormone Properties

Insulin is both a protein and a hormone.

Before you work out the best way to deliver a protein you will need to know the properties of the substance with which you are dealing. What are the properties of proteins and which of these properties will have an impact on your choice of method and route of delivery? How does a hormone act and how is it usually delivered by the body? (This later topic will be a review of material covered in the insulin and Physiology module).

1.2.   Pharmacodynamics and Pharmacokinetics

You will also need to know how the body takes up and eliminates drugs. A whole discipline deals with the mathematical treatment of these issues. It develops models for the effects of drugs on the body, and how the body handles a drug. The area is too vast to cover in-depth here. We will concentrate on a brief look at Pharmacokinetics. Essentially this is the uptake and elimination kinetics of a drug.

1.3.   Drug Delivery

1.3.1. Drug delivery Strategies and Modeling

Having identified the protein properties of interest and having looked at some models for drug uptake, we look at research being done into alternative drug delivery methods. This also is a huge area which is being studied in universities and industry. Again, mathematical models have been developed for looking at the release of drugs from different situations. There are many similarities to be drawn between the modeling in pharmacokinetics, in drug delivery, and in fermetations that you encounter in the Bioprocess Engineering module.

1.3.2. The Specific Case of insulin

Finally we tackle the central issue of insulin. We will look at the cutting edge research that is being conducted into novel ways of delivering insulin. We will look at everything from pumps to nasal sprays, from transplants to skin patches.

The economic potential for a novel insulin delivery system is huge. The 1993 market for human insulin was 350 million dollars. This is only eleven years after regulatory approval for human insulin to Eli Lilly in 1982. Projected sales for year 2003 are $675 million.

Chapter 2

Protein Properties that Impact Drug Delivery

2.1.   Insulin as a Protein
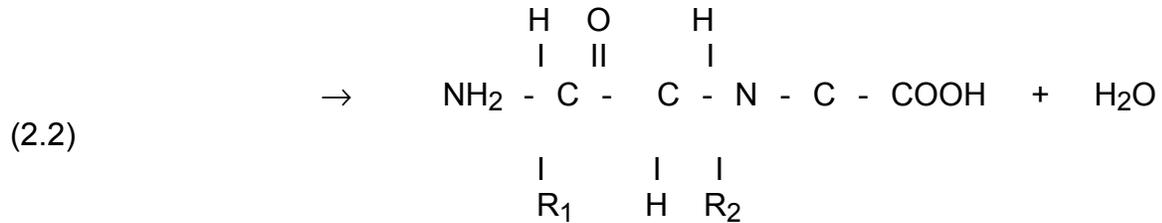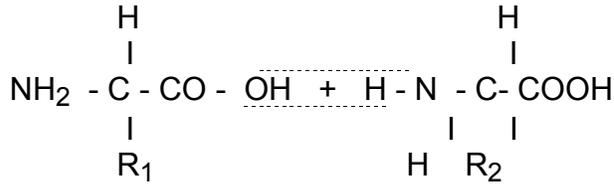
2.1.1 Composition Amino acids

Proteins are made up of molecular units called α **amino acids.** These monomer units consist of a backbone with an amino group at one end and a carboxylic acid group at the other (NH_3-C(R)H-COOH), and a side group, R, on the α carbon. It is the R group that distinguishes one amino acid form the next, gives the acid its name and also defines its properties. These R groups can be of four types: Hydrophobic, Hydrophilic, Basic (+ charged at physiological pH) and acidic (negatively charged at physiological pH). Physiological pH is the pH of the body, which is kept constant at 7.4, except in certain isolated chambers such as the stomach which is between 3 and 4. Because of the four valent bonding structure of the carbon atom, amino acids are optically active, existing in two isomeric forms, the D and L form, but only the L form is found commonly in nature. There are **20 common L-amino acids** that make up proteins. Their structures are given on page three of the class handouts. Since the basic and acidic amino acids have **ionizable groups,** they can be either charged or uncharged depending on the pH of the environment**.** A basic group can be either neutral or positively charged, and an acidic amino acid can be either neutral or negatively charged. It also follows that a protein, which is composed of a string of linked amino acids, can have different charges at different pH, and that there will be a unique pH at which there is no net charge, that is the sum of the negative charges equals the sum of the positive charges. This point is called the **Isoelectric point** (IP), and is a important property of a protein. When the pH is above the IP, the protein will have a net negative charge and when it is below the IP it will have a positive charge. At the isoelectric point a protein has minimal solubility because the lack of a net charge means that the protein molecules no longer repel each other, and they have a tendency to clump together and precipitate. The isoelectric point of beef insulin is 5.7 and porcine insulin is 6.0.

2.1.2. Bonding in proteins

The order in which amino acids are linked together in a protein is called the **primary structure**. Each amino acid can be represented by the first three letters of its name, so that the primary structure can be written out. A complete list of amino acid structures and the three letter code (together with a single letter code that is also used sometimes) is given in the class notes. By convention protein primary sequence is written from the end that contains the NH_2- group. For example the primary sequence of the first seven amino acids of the insulin A chain is :

$$NH_2\text{-Gly} \quad \text{Ile} \quad \text{Val} \quad \text{Glu} \quad \text{Gln} \quad \text{Cys} \quad \text{Cys...}$$

(2.1)

The bond that joins amino acids together is called a **peptide bond**, and is obtained by elimination of H_2O from the acid terminal of one amino acid and the amino terminal of the next.

$$
\begin{array}{c}
\quad\quad\quad H \quad\quad\quad\quad\quad\quad\quad H \\
\quad\quad\quad | \quad\quad\quad\quad\quad\quad\quad\quad | \\
NH_2 - C - CO - \underline{OH} \; + \; \underline{H} - N \; - C - COOH \\
\quad\quad\quad | \quad\quad\quad\quad\quad\quad | \quad\; | \\
\quad\quad\quad R_1 \quad\quad\quad\quad\quad\quad H \quad R_2
\end{array}
$$

(2.2)

$$
\begin{array}{c}
\quad\quad\quad\quad H \quad O \quad\quad H \\
\quad\quad\quad\quad | \quad || \quad\quad | \\
\rightarrow \quad NH_2 - C - \quad C - N - C - COOH \quad + \quad H_2O \\
\quad\quad\quad\quad | \quad\quad\quad\; | \;\; | \\
\quad\quad\quad\quad R_1 \quad\quad H \; R_2
\end{array}
$$

The resulting peptide bond [- $\overset{\overset{\textstyle O}{||}}{C}$ - N -] is a rather rigid bond because of the possible resonance structures, and the peptide unit is planar as a result. The peptide bond is also the major target of **peptidase enzymes**, that is enzymes (natures catalysts) that are specific for breaking the peptide bond. Enzymes themselves are in fact proteins. There are many peptidases in the blood, which normally function to break down waste proteins. This is one of the reasons why insulin and many other protein drugs have a very short half life in the body. Table 2.1.1. lists some half lives of some important peptides in the blood.

Table 2.1.1 Plasma half lives of polypeptide drugs

Polypeptide	Molecular weight	Half-life
ACTH	~4700	< 5 min.
Angiotensin 1	~1200	15 sec.
Bradykinin	1060	30 sec.
Calcitonin	~3600	<40 min.
Enkephalins	~600	2 min.
Gonadotropic hormones	~30,000	0.5-3 hr.
Growth hormone	~22,600	<25 min.
Insulin	~6000	<25 min.
Oxytocin	1007	2 min.
Parathyroid Hormone	9500	<15 min.
Vasopressin	~1200	4 min.

2.1.3. Three Dimensional Protein Structure and Stability

We have already discussed the **Primary Structure** of a protein, that is the sequence in which the constituent amino acids are linked together like beads on a necklace. This primary sequence, which is unique to each different protein, is very important in determining what shape the protein adopts. The different amino acids can interact with each other by weak forces such as hydrogen bonding, Van der Waals forces and hydrophobic interactions. This can give the amino acid strand a structure, like twisting the beads of the necklace back on themselves. This structure is called **Secondary structure**, and it can be in the form of either an α-helix, or a so called β-sheet. These are illustrated on Figure 2.1. Many fibrous proteins such as collagen, or keratin which is in hair, are composed of intertwined α-helices. The β-sheet secondary structure is more stable because of the strong hydrogen bonds between chains.

If we again think of our analogy with a necklace, you can imagine that the necklace, with its sequence of beads, and its local defined α-helix and β-sheet structure, could also be wound up so that beads (amino acids) that are well separated on the necklace can interact. When the amino acid groups are cysteines, this interaction can be covalent, a we saw between the A and B chain on insulin. This longer distance interaction results in what is called the **Tertiary Structure.** The importance of tertiary structure can be seen when we consider that it is this structure that determines the final shape of the protein. In an enzyme the final shape most likely results in a cleft in the protein which is exactly the correct shape to accommodate the substrate on which that enzymes acts. The example that we have already seen is that of a protease that breaks down a protein. The cleft, known as the **active site,** is the correct shape to take in the peptide bond, which is then cleaved. For protein hormones, such as insulin, the shape of the insulin results in an area, called the **binding site,** that will exactly fit into a **receptor site** on a cell surface. For a class of very large proteins know as **immunoglobulins** (antibodies), the final shape results in a Y-shaped molecule, in which the top two prong of the Y posses binding sites for a specific antigen. In all these cases the correct shape of the protein is vital for its activity. If the shape is altered or damaged, the protein will not be able to function. It follows that everything must be done to preserve the protein in its active form. Unfortunately proteins are not robust molecules. As was mentioned above, the forces that hold the secondary and tertiary structure together are all weak, such as hydrogen bonding.

Occasionally proteins have other groups attached to them, such as the heme group in the blood protein hemoglobin. These groups are called **prosthetic groups**. Hemoglobin is also an example of a protein that has yet a fourth level of structure, the **Quaternary Structure**, in which four folded chains get together to from a multi-chain oligomer. Many of these levels of structure are shown in figure 2.1.1.
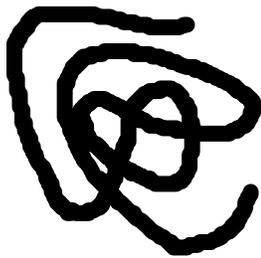
Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
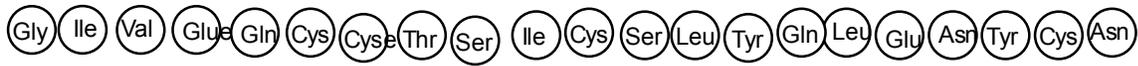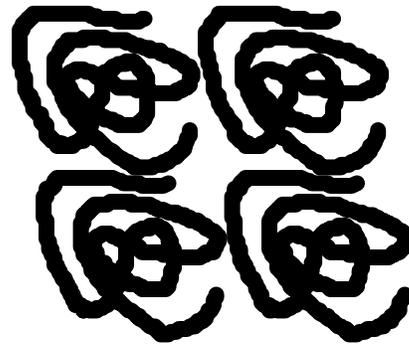
Primary structure of Inslin A chain

α- helical coil

Supercoiling of α-helical coil
in fibrous protein

Tertiary Structure

Quaternary Structure

Figure 2.1.1. Levels of structure found in proteins

One of the most important take-home messages of this section is that proteins
have a very well defined structure, which is vital for their specific activity. The
structure is a result of the interactions of amino acids which make up the
backbone of the protein molecule. The forces that hold these amino acids
together in the secondary through quaternary structures are generally weak
force, such as hydrogen bonds. This means that it is easy to break this all

important structure down. The fragile nature of proteins has important implications on their handling and delivery.

Another consequence of the fact that the side groups of the amino acids in a protein interact is that protein molecules tend to aggregate when they are in solution. Insulin is particularly bad at forming dimers and trimers at the slightest provocation. The solution can be made to form a precipitate simply upon shaking. This is very unfortunate, since insulin is a prime candidate for delivery from a pump.

2.1.4. Molecular Weight

Proteins usually are made up of more than 50 amino acids. Below 50 they are called **polypeptides**, and above 50 proteins. Insulin, with 51 amino acids, has a molecular weight of 6000. This is enormous compared with aspirin, which has the same molecular weight as glucose, 180 Daltons. Since most compounds enter the body by passage through membranes or through minute pores, and travel about the tissues by a process of diffusion, this again has important implication in drug delivery.

2.1.5. Immunologic Response

When a foreign protein enters the blood stream, the body mounts what is called an **immune response**. **Antibodies** are produced which coat the invader and allow the cells of the immune system to destroy it. In fact there is strong evidence that Diabetes itself is caused by the body producing antibodies to its own β cells, the cells in the pancreas which produce insulin. Usually though, antibody production is beneficial to the organism because it helps fight off infection. A vaccine contains proteins of a pathogen, so that the body will produce antibodies, and after a booster shot, the system is primed to mount a very rapid and effective defense against attack by the organism itself. When thinking about proteins as drugs, and how we want to deliver them, we must be aware of the possibility of the body mounting an immune response to the protein and destroying it, or worse giving an allergic reaction after repeated doses. With the development of human insulin, this hazard is greatly reduced for diabetics. Since the gene which expresses the insulin is of human origin, the product of that gene is indistinguishable from that produced by a human. Other researchers have suggested coating foreign proteins with a chemical coat of polyethylene glycol molecules to disguise them from the immune system, or of encapsulating the protein in a tiny capsule called a liposome, that also can evade the immune system.

2.2. Insulin as a Hormone

2.2.1. Activity at the Cellular Level

Insulin binds to a site called the **insulin receptor** on the cell surface of liver and other cells, and in so doing changes the receptor shape and triggers a cascade of events which cause the cells to be more receptive to glucose uptake. The liver stores the glucose as glycogen, a polysaccharide. In situations where the glucose level is below optimum, the glycogen is broken down and released back into the body as glucose. The insulin receptor is itself a protein and it is made up of two alpha subunits on the outside of the cell membrane, each composed of 719 amino acids, and two beta subunits with 620 amino acids apiece, which straddle the cell membrane and are attached to the alpha units by disulfide bridges. Once again we see the importance of being able to deliver the insulin to the cell receptor in the active form so that it can bind properly.
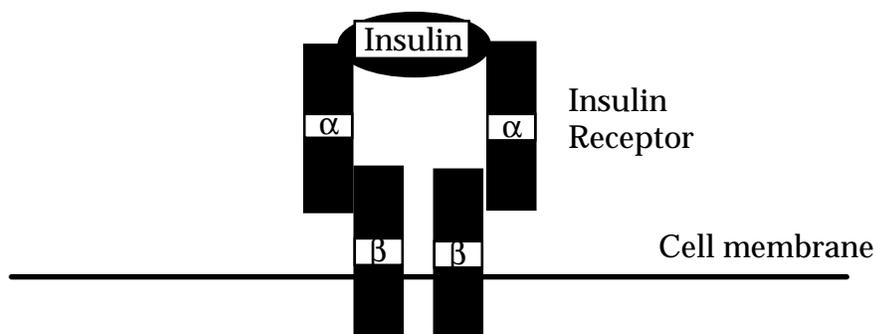
Insulin attached to Insulin receptor   (2.3)

Figure 2.2.1

2.2.2. Natural Delivery Profile

The **control of glucose levels** in the blood of non diabetic people is exquisite. It takes place very rapidly, and on two different levels, one at the gene that expresses insulin in the β cells of the pancreas, and the other at storage vesicles in the same β cells. The net result, which the ideal insulin delivery system should mimic, is that when the blood glucose level rises, insulin is released into the blood, and when it drops back to normal, the insulin supply should be stopped. We would like to produce a **responsive drug delivery system.**

2.3.   Summary

We find then that insulin is a large molecular weight molecule, a protein. The fact that it is a protein and a hormone, means that it must maintain its tertiary

structure to remain active. The ideal drug delivery system will protect this fragile molecule from degradation by physical or enzymatic agents, enable it to get to the receptor where it will act, and deliver it in metered doses in response to elevated blood glucose levels.   The current method of delivery is by intramuscular injection. Diabetics must take and insulin shot three or four times a day, and regulate their diet very strictly. Even with the advent of sophisticated monitoring devices with which the diabetic can monitor blood glucose levels in a small drop of blood, the system still results in wide swings of blood glucose levels, swings that are not found in the non diabetic person, and which have been implicated in the many health problems that face diabetics such as kidney failure, and blindness.

Further Reading Suggestions

Shuler, M.L. and Kargi, F., Bioprocess engineering part 2.2. Prentice Hall 1992
Lehninger, A.L., 3rd ed. Worth publishers New York

Questions you should be able to answer
- What charge will insulin have a physiological pH?
- What are enzymes?
- What famous molecule has an α-helical structure, other than protein?
- Egg whites are composed mostly of the protein ovalbumin. Can you speculate what is happening when you beat an egg into a stiff foam when making a meringue?
- Why does beer have a head of foam on it?
- Would it be a good or bad idea to store insulin in a solution with a pH the same as the isoelectric point of insulin?
- Will the hydrophobic amino acids be found on the inside or outside of a protein molecule in its active form in water?

Questions you should be able to answer with a little library research
- Name an antibiotic that contains a rare D-amino acid
- What is the name given to a molecule that has both positive and negative charges?
- In what area of the body apart from the blood do you find a lot of proteases?

Chapter 3: Pharmacokinetics

3 .1.   Definitions

We will start this chapter with some definitions of terms that are used frequently in pharmacokinetics,

Pharmacokinetics
Pharmacokinetics involves the study of the manner in which the body handles a drug.   It looks at the rate at which drug concentrations change in the body, and is involved with analyzing the kinetics of absorption, distribution, metabolism and excretion of a drug.

Pharmacodynamics
The study of pharmacodynamics involves the study of the effect that a drug has on the body. This will include looking at what effect the amount of drug has, called a dose response, and will also look at the site of action of the drug, such as the kinetics of how a drug interacts with a receptor, known as receptor dynamic.

Pharmacology
Pharmacology is the study of how a drug changes the functioning of an organ.

Toxicology
This is a branch of pharmacology that deals with undesirable effects of molecules on the body. Before a drug company can file an NDA (New Drug Approval) with the FDA, extensive toxicological studies must be performed in at least three species.

All these fields are relate, and play an important role in the observed effects of a drug on the body.

Systemic
The adjective systemic means -involving the entire body. For example one can say that drug delivered by injection will have a systemic effect while one that is applied to a certain area on the skin has a local effect.

Bioavailability
The bioavailability of a drug is the amount of drug that reaches the **systemic** (entire body) circulation in an active form. The process of breaking down a drug starts as soon as it enters the body. Some routes of administration are therefore better than others, depending on the type of drug. Drugs that are broken down by the pH or enzymes in the GI tract have a very low bioavailability via that route.

First pass effect
Another cause for low bioavailability is a phenomenon called the **first pass effect**. All the blood vessel that feed most of the GI tract drain into the portal

vein which goes to the liver. That means that drugs absorbed from the gut go directly to the very organ in the body that is responsible for detoxification. So the first pass of orally administered drugs around the body in the circulation takes them not directly to the area where they are needed, but first to the liver, where they could possibly be broken down.

Biotransformation
Biotransformation is the conversion of a drug molecule to another, usually inactive form, in the body.   Some drugs, however, are administered as a prodrug, which is an inactive form which the body transforms into an active form. For the Parkinson's drug, L-Dopa, the advantage is the inactive form can pass into the brain, whereas the active form can not.

3.2   Absorption of Drugs

3.2.1. Routes of Administration

There are many different routes by which a drug can enter the body, and the route has a large impact on how fast the drug is taken up and how much of it arrives at its destination in an active form.   Table 3.2.1. outlines some of the routes of administration with an indication of some of the advantages and disadvantages of each.

Table 3.2.1. Routes of drug Administration

Route	Advantage	Disadvantage
Oral	Convenient Efficient Absorption (huge surface area)	Drug degradation Requires lipid permeation Gastric irritation Frist pass effect
Inhalation	Large surface area Access nasal or lungs Rapid delivery to blood	Requires special properties (vaporized, atomized) Nasal proteases Irritation
Topical	Convenient Rapid Localized	Only local
Transdermal	Prolonged Release	Lipid permeation Initial time lag Dermal enzymes
Buccal	Self administered No first pass effect	Drug must be lipid soluble and potent

Intravenous	Direct Rapid Programable	Hazardous Requires professional
Intramuscular	Rapid for aqueous Slow depot for oil	Pain, inflamation at delivery site
Others: Intrarteria, Intrathecal, Intraperitoneal		

The **most common route is the oral route**. This route is convenient and very efficient, with a huge surface area and compartments of different pH which can accommodate drugs with different solubilities. Drawbacks include the fact that many compounds, proteins for example, and broken down by the GI enzymes and harsh pH conditions, and the fact that the drug has to have considerable lipid solubility to pass through the fatty cell membranes of the gut wall to reach the blood supply. Another important phenomenon, alluded to earlier, called the **first pass effect** reduces the availability of many drugs. It results form the arrangement of the blood vessels that permeate and drain the GI tract, (with the exception of the lower rectum). Most vessels from the GI drain into the portal vein, which goes to the liver. The liver is the waste management system of the body, breaking down waste and detoxifying compounds. If a drug goes first to the liver it is likely that a large degree of degradation will take place. One route that has become widely studied recently is the transdermal route. Many skin patches have come onto the market, including ones containing the anti angina drug nitroglycerine, patches for hormone replacement therapy, and ones for motion sickness containing scopolamine, which have even found their way on board the space shuttle. The main problem here is that the drug has to be able to permeate the fatty layers of the skin.   Research is ongoing to find ways of encouraging large molecular weight proteins through the skin, opening up the pores in the skin either by chemicals, know as **enhancers**, or by physical means such as applying an electric field (**iontophoresis**) or ultrasound energy (**sonophoresis**).

The traditional route for insulin is intramuscular. This involves injection directly into the muscle of the leg, arm or abdomen. The uptake can be quite rapid if the insulin is in an aqueous solution, but if it is in an oily solution diffusion away from the site of injection is inhibited, and the effect is for the oil suspension to act as a slow release depot. Figures 3.2.1 and 3.2.2 give some idea of how the muscle into which the injection takes place can affect the uptake, as well as the presence or absence of exercise. Figure 3.2.1 is a representation of results from Kiovisto in the English Journal of Medicine and shows the change in plasma glucose levels after insulin is injected into either the leg, arm or abdomen, during one hour of exercise and for a subsequent 6 hours. Since the lowering of blood glucose levels is directly related to the amount of insulin absorbed, it can be seen that significantly more insulin is taken up from the leg muscle, which

presumably has a much greater blood flow during exercise, than from the other two sites

Figure 3.2.2. shows that the amount of **radio labeled** insulin that is taken up following subcutaneous insulin injection was far greater with exercise, presumably again a function of greater blood flow. The use of a drug which has been changed so that it contains an atom which is radioactive, a so-called radio-labeled compound, is frequently used in drug delivery research. It is very easy for the researcher to follow where the drug goes by measuring the amount of radioactivity that is present. When the experiment is performed in volunteers, the levels of radiation are kept very low, so as not to harm the volunteer.
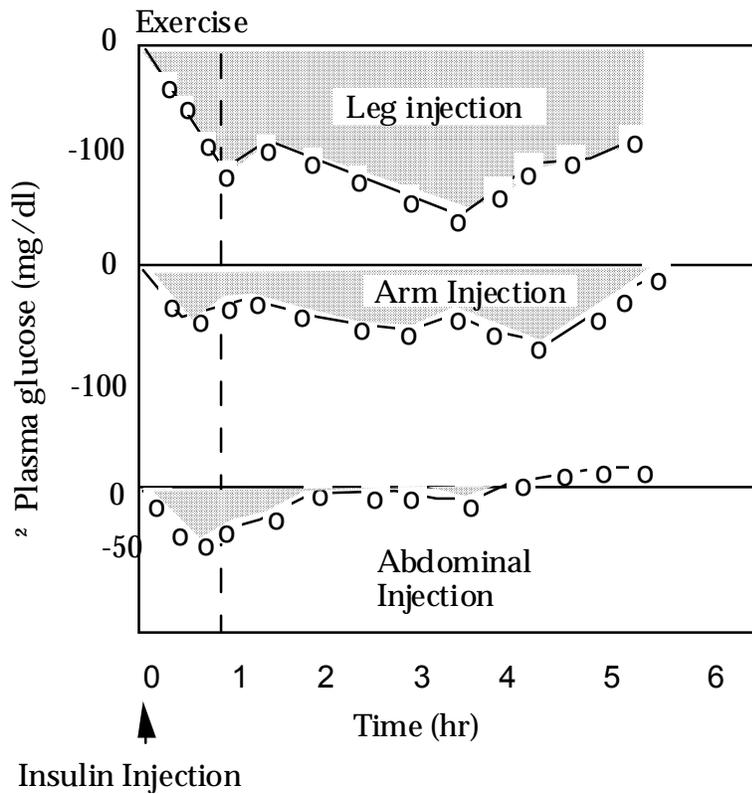
Figure 3.2.1. Influence of injection site on the plasma glucose response to insulin during exercise (leg) and in the absence of exercise (arm and abdomen). (After Kiovisto et al. 1978)
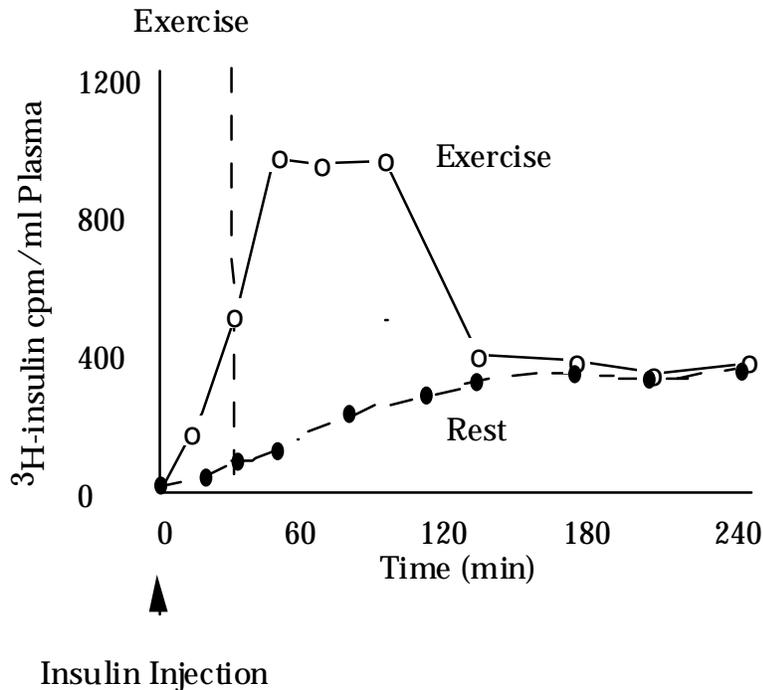
Figure 3.2.2. Effect of exercise on plasma levels of ^{3}H-insulin following subcutaneous injection. (After Berger et al. 1979.)

3.2.2. Factors that affect Drug Absorption

Many factors can effect the rate and extent of drug absorption. A very important factor is the solubility of the drug**.** If the drug is soluble in an aqueous solution it will be absorbed by the surrounding tissue, more rapidly than a drug that is dissolved in an oily solution. However, in the GI tract, the situation is complicated by the need for the drug to have some lipid solubility in order to pass through the cell membranes of the intestine wall and blood vessels and into the blood. So the drug needs to be water soluble and fat soluble at the same time for optimal absorption. In addition a drug that can be ionized will pass through a membrane better in the un-ionized form. This means that weak acids will cross better at a pH lower than the pKa, and weak bases will cross better at pH above their pKa. The **pKa of a substance** is defined as the negative logarithm of the dissociation constant, and is equal to the pH at which the concentrations of the acid and base form are equal. However, the ionized form of a drug is more soluble, and a drug must be in solution to be absorbed.

Also the **rate at which the drug dissolves** at a particular site will have a profound effect on the absorption rate. This can be strongly influenced by the local conditions. Some drugs, for example, and not soluble in the acid of the stomach, but are soluble in the alkaline conditions of the GI tract.

The GI tract is also a good example of how **stability** of a drug can effect its absorption. Many drugs that are of interests today, such as proteins, are not

stable in either the acid environment of the stomach, or the alkaline environment of the intestine, which is rich in protein degrading enzymes .

Not only solubility and stability but also **residence time** at a site can be a factor. Many researchers have investigated the use of substances, called **bioadhesives**, which will hold a drug at a particular site. Drug washout is particularly prevalent in the eye. The tears of the eyewash over 95% of an applied aqueous dose of drug away into the nasolachrymal duct.   Any vehicle that can prevent this from happening will greatly increase the absorption of the drug in the eye. Another example of drug loss is the rapidity of stomach emptying. If you want absorption of your drug from the stomach it is best to take it after a heavy, fatty meal, when the chances of the stomach emptying are diminished.   The best way to get it rapidly into the intestine is to take it on an empty stomach with a lot of fluid, since fluid tends to increase stomach contractions.

Drug **concentration** is an interesting, and contradictory factor in drug absorption. Often high concentrations lead to more rapid absorption, as would be expected since the driving force is greater. However, sometimes in GI absorption this is not the case. Often absorption increases when the drug is either taken as a pill with a large volume of water, or as a dilute solution. The factors that are now becoming important are a more rapid stomach emptying as mentioned above, an increase in surface area contacted by the large volume of the solution, and in the case of a pill, a larger volume of water in which to dissolve more rapidly. This is only true for the GI or **enteral route.** Any route other than by mouth is known as **parenteral** administration.

In   addition to the concentration of drug, the blood flow at the absorption site can greatly influence absorption, which will be increased with increased blood flow. This is an advantage if delivery is to an area that is inflamed due to infection. Also in the GI tract, approximately 1-1.5L of blood pass through the capillaries per minute, compared to around 150 ml/minute in the stomach.

Finally the **area of absorbing surface** is a key factor in drug delivery. Areas such as the lung and the GI tract offer a huge surface over which drug can be absorbed. The intestine, for example, has a surface that is greatly folded, and even the folds posses finger-like projections (page 13 of class notes). The first level of structure is known as the **folds of Kerkring**. The finger-like projections on the surface of these folds are called **villi**, and the cells making up the villi have themselves a serrated edge covered in tiny projections called microvilli. These structures serve to increase the surface area by three then 30 then 600 fold, to end up with a total surface area of around 200m^2.

3.2.3. Barriers to Diffusion

Drugs usually pass into the blood steam by **diffusion**, and then pass to the area of concern by diffusion out of the blood steam into the affected tissue. The drug

has to be in solution to diffuse. Aqueous diffusion can take place through either the spaces between cells, known as **cell junctions**, or through small **pores** in the capillaries. In order to pass **between cells** the molecule must be small, with a molecular weight **less than 100-150 Daltons.**   To pass through most **capillaries** the molecular weight must be **less than 20-30,000 Daltons.**   The exception is in the **brain**, where the capillary pores are a lot smaller, and the compound must be less than **5000 Daltons**.

As mentioned above, drugs can also travel by passing across cell membranes. This is known as lipid diffusion, since the cell membranes are made up of lipophilic substances. This requires the drug to be quite lipid soluble. One way of measuring this parameter is to measure the octanol/ water **partition coefficient** ($K_{o/a}$). A high coefficient indicates a preference for the lipid layer. The $K_{o/a}$ for insulin is very low, 0.0215. A drug such as phenobarbital, which has a pKa of 7.4, exactly physiological pH, is pretty lipid soluble in acid conditions because of its lack of charge, and is reabsorbed by the kidney into the body. However, if the urine is alkaline, phenobarbital has a charge, and is excreted by the kidney.

Some transport appears to be **carrier-mediated**, that is special carrier substances take the compound across the membrane. This type of mechanism uses cellular energy and can be saturated or even inhibited.   Vitamin B_{12} is transported across the gut by a compound called transferin, a process that is sometimes refered to as f**acilitated diffusion** and some amino acids are transported into the brain by active transport.

Finally, very large molecules (M.W. > 1000) can enter the cells by a process called **pinocytosis**, in which the cell membrane forms projections which surround the molecule, which is then engulfed by the cell.

3.3.   Elimination of Drug from the Body

3.3.1. Renal excretion

The most important route of excretion for most substances is the through the kidneys, known as **renal excretion**.   There are three processes, filtration, reabsorption and secretion, and they all take place in different parts of the kidney.   The main structure in the kidney that performs these tasks is the **nephron**, as shown in class notes page 16.

A single kidney contains over a million nephrons. The blood flow into the kidney is around 1.2L/min, or 25% of cardiac output. This contains about 650 ml of liquid, of which 130 ml is filtered out and roughly 128 mL is reabsorbed. In a day roughly 1.5L is excreted. The artery entering the kidney divides up into a ball of fine capillaries in the top of each nephron, forming the **glomerulus**. The high pressure created in these small capillaries causes a large amount of plasma to

be filtered out and pass into the proximal tubule. It is in this tubule that water and salt are reabsorbed, but any drug that has passed along with the plasma is not. There is also active secretion of drugs from the plasma in the capillaries around the tubule. The rest of the nephron is involved with salt, glucose and water reabsorption, and acidification of the urine. In diabetes the active reabsorption mechanism for glucose may get saturated, resulting is glucose excretion. It was in fact the observation that flies settled on the sweet urine samples from diabetics that first alerted physicians to this symptom.

3.3.2. Liver Metabolism

The liver affects drug elimination in two ways. As mentioned previously, some drugs are extensively metabolized, or broken down by the liver in the first pass effect which is the result of the physiological fact that all blood supply from most of the GI tract drains directly into the liver. The part of the liver where most of the drug metabolism takes place is known as the **microsomal fraction.** It is this fraction that is rich in the degradative enzymes. But the liver also excretes some drugs into the **bile**, and this bile is excreted into the **duodenum** by an **active transport mechanism**. If the drug is then reabsorbed from the gut in an active form, it will enter the circulation again. This is called the **enterohepatic circulation** and is responsible for prolonged plasma drug levels for drugs such as the heart drug digoxin.

3.3.3. Degradation in Plasma

Although the liver is probably the main area of drug metabolism, there are also many enzymes circulating in the blood which break down drugs, especially proteins. Often, too, drugs bind with proteins in the blood, and this can effect their metabolism and excretion. **Bilirubin** is an example of a compound that binds to **serum albumin**. In new born babies the liver often has a hard time breaking this plasma conjugate down, and it is deposited in the skin and eyes, giving the yellow appearance that is typical of a jaundiced baby. When a newborn is placed under controlled UV lights, the radiation is sufficient to change the shape of the bilirubin so that it can no longer bind to the protein, and it is much easier to excrete.

3.3.4. Minor Routes

As well as the main areas of kidney and liver, there are many other minor routes of removal such as the **sweat**, in **saliva** and in **tears**.

3.4. Pharmacokinetic Parameters

We will start this section with some more definitions.

Area under the Drug Concentration Curve or $AUC_{0\rightarrow\infty}$

If a plot of the plasma drug concentration versus time is made, then the area under this curve, or $AUC_{0 \to \infty}$ represents the total amount of drug absorbed. This parameter is very important in defining how well the drug is absorbed, and is required by the FDA in the drug approval process.

AUC of (Time x Plasma Drug Concentration) versus Time or $AUMC_{0 \to \infty}$
This is a related parameter to the one above, and represents the concentration of drug in the plasma or body.

Volume of Distribution,Vd
The volume of distribution is a purely kinetic parameter that is used to estimate the degree to which a drug permeates the different fluid compartments of the body. It is defined as the amount of drug in the body divided by the concentration of drug in the blood or plasma. The value can be derived graphically from plasma concentration versus time curves.

$$Vd = \frac{Dose. \quad [AUMC_{0 \to \infty}]}{[AUC_{0 \to \infty}]}$$
(3.1)

Clearance
Clearance of a drug from the body may be expressed as total body clearance or clearance form a specific organ such as the kidney or from the plasma. It is defined as the rate of elimination divided by the concentration:

$$Cl_i = Dose / [AUC_{0 \to \infty}]$$
(3.2)

where i represents plasma (p), an organ (o), or the kidney (r)

Plasma Clearance
The volume of plasma cleared of drug per unit time as a result of all elimination pathways is known as the plasma clearance. If drug is eliminated by a first order rate constant k_{el}, then:

$$Cl_p = k_{el}V_d$$
(3.3)

and

$$Cl_p = \frac{dose}{AUC_{0 \to \infty}}$$
(3.4)

where the dose is given by intravenous injection, i.e. 100% bioavailability.

Renal Clearance

The volume of plasma cleared of drug by the kidneys per unit time, is known as the renal clearance. It can be determined from measurements of the amount of drug voided in the urine and the plasma drug concentration.

$$Cl_r = U V_u / C$$
(3.5)

where U is the drug concentration in the urine, and V_u is the flow rate of urine and C is the mean plasma concentration of drug.

Plasma Half Life

The biological or plasma half life of a drug is the time it takes for the plasma drug concentration to drop to one half. For the situation of equation (3.3) above, the half life can be shown to be related to the k_{el} by:

$$t_{1/2} \quad = \quad \ln2 \ / \ k_{el} \quad = \quad 0.693 \ / \ k_{el}$$
(3.6)

It also follows that:

$$Cl_p = 0.693 V_D / t_{1/2} \tag{3.7}$$

Single organ Elimination rate

Rate at which drug is cleared from an organ is known as the elimination rate.

$$\text{Rate} \quad = \quad Q.C_{in} \quad - \quad Q.C_{out}$$
(3.8)

Where Q = Plasma flow rate through the organ
C = concentration of drug passing in/out of the organ

It follows that organ clearance as defined by flow and concentration is:

$$Cl_{organ} \quad = \quad \frac{Q \ (C_{in} \ - \ C_{out})}{C_{in}} \quad = \quad \frac{dose}{AUC0 \rightarrow \infty}$$
(3.9)

Extraction Ratio

The difference between the inlet concentration and the outlet concentration, divided by the inlet concentration for an organ is known as the extraction ratio, and it defines the degree to which the drug is eliminated or degraded by that organ.

$$E = \frac{(C_{in} - C_{out})}{C_{in}} \quad \text{(3.10)}$$

Oral Availability

The oral availability describes how much of a drug is available after having passed through the liver, and undergone the first pass effect, assuming 100% absorption.

$$OA = 1 - E_H \quad \text{(3.11)}$$

Where E_H is the hepatic extraction ratio.

The extraction ratio, and hence the oral availability of a drug can be calculated by measuring the AUC for the drug administered by IV injection, and knowing the blood flow rate through that organ. By comparing this value with that obtained for oral availability one can get and idea about the efficiency of the absorption process, or about the liver metabolism. An oral availability less than predicted suggests poor absorption, while one greater than predicted indicates that the liver metabolism may be via a mechanism that is saturable at high drug concentrations.

3.5.   Pharmacokinetic Modeling

3.5.1. Modeling approaches

There are three approaches that have been suggested for pharmacokinetic modeling, compartmental, physiological and model-independent.

The first is an empirical approach, which is based on a simple **compartmental model**. These compartments have no strict physiological or anatomic basis. The compartment can represent a body volume, or just as easily it could represent a chemical state, for example a metabolite of the drug. Usually this approach uses one or two compartments. Despite its simplistic nature, many useful quantities can be derived using this approach and by comparing predicted values with actual data.

The **physiological model** identifies the compartments with actual body spaces. The model is a great deal more complex than the compartmental models. Actual transfer and flow rates are employed together with experimentally measured blood/tissue concentration ratios. This model can be used for predictions, and is more adaptable to clinical therapy and to changing situations such as alterations of flow rates due to conditions such as swelling, or fever. Scale up is more easily from one species to another, which is convenient since most of the required measurements must be performed in animals.

The **model independent approach** is the most recent, and is purely mathematical. It avoids recourse to kinetic parameters which may not be valid, and is a lot less complex. It is good for adsorption and elimination rates, and clearances, but gives no specific information about how the drug is distributed.

3.5.2. One Compartment Open Model

This approach models the entire body as a single compartment into which drug is added by a rapid single dose, or **bolus**. It is assumed that the drug concentration is uniform in the body compartment at all times and is eliminated by a first order process that is described by a first order rate constant k_{el}. The model is shown in figure 3.5.1.

Drug Input → Body Compartment / Drug concentration = C(t) / Compartment volume = V → $k_{el}C(t)$
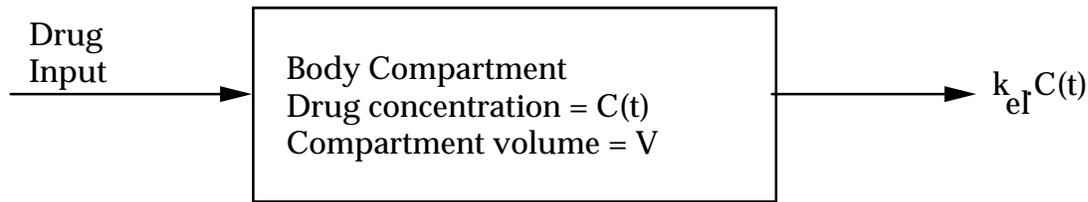
Figure 3.5.1. Basis of the one compartment open model

Analysis of this model relies on a simple mass balance .

$$\frac{dC}{dt} = -k_{el}C(t) \qquad \text{at } t = 0, \qquad C_o = \frac{Dose}{V}$$
(3.12)

$$C = C_o \, e^{-kel\ t} = \frac{D}{V_D}$$
(3.13)

This is a first order rate equation, which can be solved by a plot of ln C against time
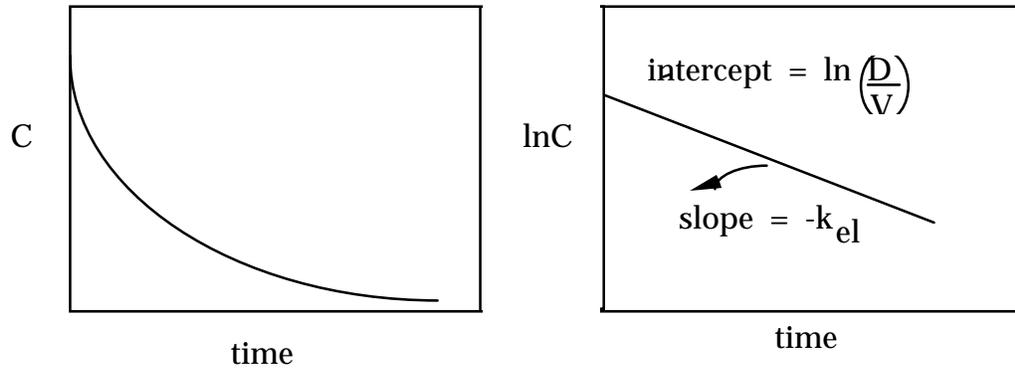
Figure 3.5.2. Graphical treatment of plasma concentration with time to obtain k_{el} and V_D

The one compartment model fails to describe the actual drug disposition when, for example, a particular organ has a small, but strong affinity for a drug, which does not affect the overall plasma concentration, but which leads to toxicity on repeated doses. If this area is the site of drug action, the effect could continue after blood levels had subsided.

3.5.2.1. One Compartment Model: Oral Administration

When a drug is taken orally as a tablet, the drug has to dissolve and be absorbed by the gut. This is often a first order process, and this should be accounted for by a first order absorption term in the kinetic analysis, as shown in figure 3.5.3:
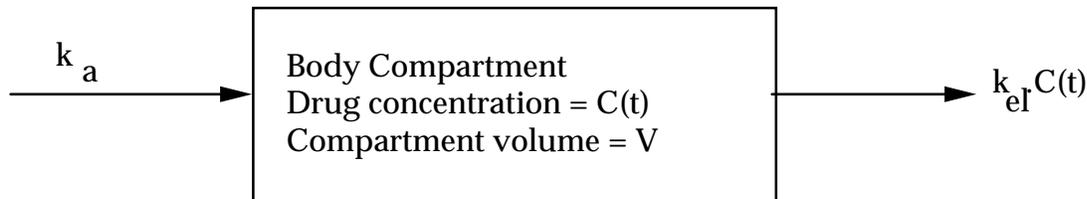
Figure 3.5.3. One compartment model with first order absorption

$$\frac{dC}{dt} = k_a D - k_{el}C \quad (3.14)$$

$$C = C_o e^{-k_a t} \quad (3.15)$$

which yields:

$$C = \frac{k_a C_o}{} [1 - e^{-k_a t}] \quad (3.16)$$

$$V (k_{el} - k_a)$$

3.5.3. Two compartment model

Often drug does not distribute evenly amongst all the organs. To account for this a two compartment model is used in which drug disposition is biexponential. The drug is assumed to distribute into a second compartment but be eliminated from the first compartment only. This is obviously a simplification of the situation in the body, but it can give some data on rates in and out of specific organs. The plasma drug concentration initially declines quite rapidly due to elimination from the plasma and distribution into the second compartment, which can comprise several organs. This phase is called the α phase. Once equilibrium is reached, the plasma drug level declines more slowly due to elimination alone, in what is termed the β phase. This is illustrated in figure 3.5.4.
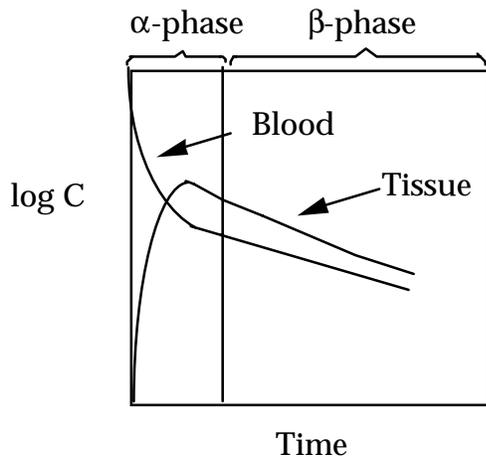
Figure 3.5.4. Plasma-Tissue drug profile that obeys the two compartment model

In the two compartment model, again a rapid bolus of drug is assumed. Also a first order transfer is assumed between compartments and first order elimination from the blood compartment, with no elimination or metabolism in the tissue. The scheme is outlined in figure 3.5.5.
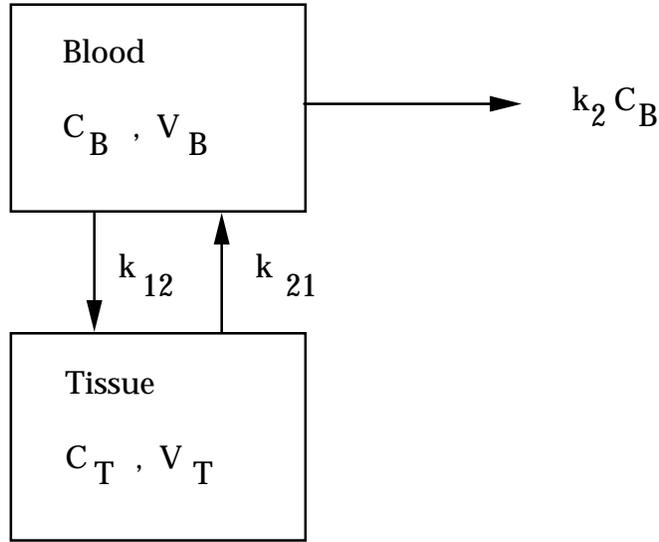
Figure 3.5.5. Two compartment open model

The kinetic equations for this case are as follows:

$$\frac{1}{V_B}\frac{dC_B}{dt} = -\frac{k_{12}\cdot C_B}{V_B} + \frac{k_{21}\cdot C_T}{V_T} - \frac{k_{10}\cdot C_B}{V_B}$$

(3.17)

$$\frac{1}{V_T}\frac{dC_T}{dt} = \frac{-k_{21}\cdot C_T}{V_T} + \frac{k_{12}\cdot C_B}{V_b}$$

(3.18)

Integration with boundary conditions,

$$@ \quad t = 0, \quad C_B = \frac{D}{V_B}, \quad C_T = 0$$

gives:

$$C_B = \frac{D}{V_B}\left[\frac{(k_{21} - \alpha)}{(\beta - \alpha)}e^{-\alpha t} + \frac{(k_{21} - \beta)}{(\alpha - \beta)}e^{-\beta t}\right]$$

(3.19)

$$C_T = \frac{D}{V_T}\frac{k_{12}}{(\alpha - \beta)}\left[e^{-\beta t} - e^{-\alpha t}\right]$$

(3.20)

where:
$$\alpha = \frac{1}{2} [k_{12} + k_{21} + k_2 - ([k_{12} + k_{21} - k_2]^2 - 4k_{21}k_2)^{1/2}]$$

$$\beta = \frac{1}{2} [k_{12} + k_{21} + k_2 + ([k_{12} + k_{21} - k_2]^2 - 4k_{21}k_2)^{1/2}]$$

This represents the biexponential semilog concentration/time plot that is expected with the two compartment model and is shown in figure 3.5.6.

Figure 3.5.6. Method of fitting data to biexponential curve
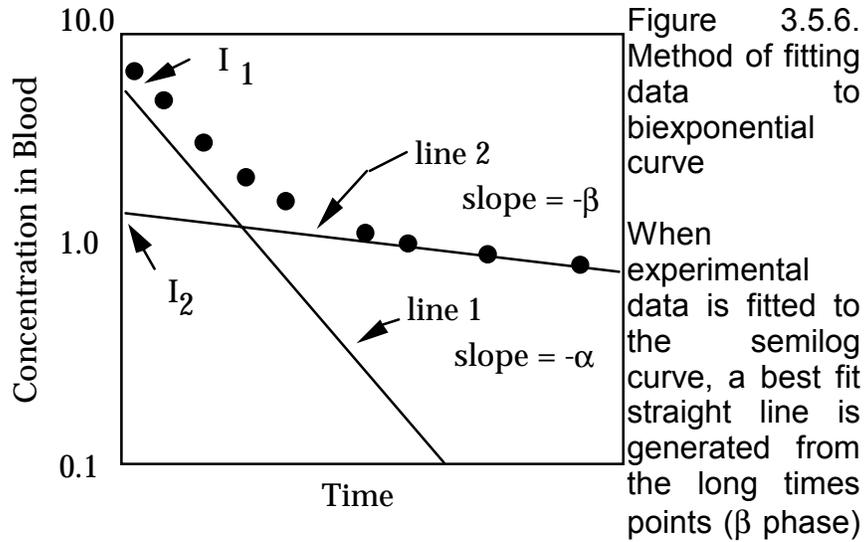
When experimental data is fitted to the semilog curve, a best fit straight line is generated from the long times points (β phase) points, and the resulting slope is - β. When line 2 is subtracted from the remaining short time data points, the slope of the resulting line 1 is -α. This technique is known as curve stripping. From the intercepts I_1 and I_2 of these lines, we can obtain kinetic parameters as:

$$I_1 = \frac{D}{V_B (\beta - \alpha)} (k_{21} - \alpha) \quad (3.21)$$

$$I_2 = \frac{D}{V_B (\alpha - \beta)} (k_{21} - \alpha) \quad (3.22)$$

$$\frac{D}{V_B} = I_1 + I_2 \quad (3.23)$$

$$k_{21} = \frac{D\alpha/V_B + I_2\beta - I_2\alpha}{D/V_B} \quad (3.24)$$

$$k_2 = \frac{\alpha\beta}{k_{21}} \quad (3.25)$$

$$t_{1/2} = \frac{0.693}{\beta} \quad (3.26)$$

Comparing these equations with those for the one compartment model, we can see that drugs with a very short α phase, (rapid distribution) the model approaches that of the one compartment model.

3.5.4. Constant IV Infusion: One Compartment Model

Drugs are not always administered by a rapid bolus, as assumed in these models. A frequent mode of drug administration is by a continuous intravenous infusion, In this case a steady state of plasma drug concentration is achieved. For the one compartment model, we assume a zero order infusion rate constant k_o. The analysis follows:

$$V . \frac{dC}{dt} = k_o - (V.C.k_{el}) \quad (3.27)$$

Integration gives:

$$C = \frac{k_o}{V.k_{el}} - [1\text{-exp} \ (-k_{el}t)] \quad (3.28)$$

$C = 0$ at $T = 0$, $Css =$ steady state concentration as $t \to \infty$

$$Css = \frac{k_o}{V.k_{el}} \quad (3.29)$$

3.5.5. Constant IV Infusion: Two Compartment Model

For the two compartment model and identical expression can be derived for steady state concentration in the body compartment:

$$Css = \frac{k_o}{V.k_2} \quad (3.30)$$

The equivalent expression for the tissue compartment is derived as:

$$Css = \frac{C_B.V_T}{B} \frac{k_{12}}{k_{21}} \quad (3.31)$$

3.5.6. Sustained Release Formulation: One compartment model

Many of the drugs that one can purchase over the counter **(OTC drugs)** are available is **sustained release** formulations. The vehicle in which they are supplied has been modified so that not all of the drug is available for absorption. This means that some fraction of the drug f_i, is available immediately, and some fraction f_r, $[f_r = (1-f_i)]$ is released with first order kinetics, described by a constant k_r as illustrated in figure 3.5.7..
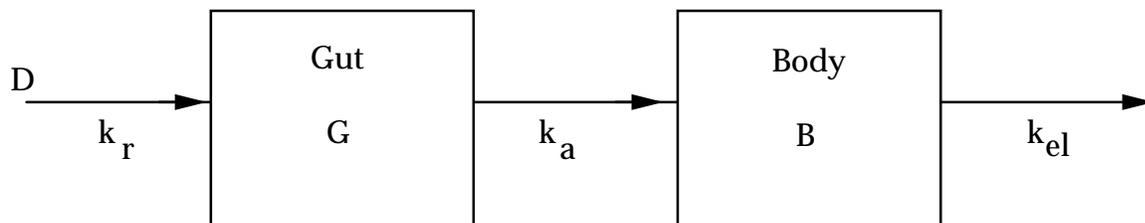
Figure 3.5.7. One Compartment model for a sustained release formulation

$$\frac{dD}{dt} = -k_r D \quad (3.32)$$

@ t = 0   D = $D_0 f_r$

$$D = D_0 \, f_r \, e^{-krt} \quad (3.33)$$

For the concentration of drug in the gut:

$$\frac{dG}{dt} = \text{Drug in - Drug out}$$

$$= k_r D - k_a G \quad (3.34)$$
$$= k_r D_0 \, f_r \, e^{-krt} - k_a G \quad (3.35)$$

@ t = 0   G = $D_0 f_i$

therefore:

$$G = \frac{D_0 \, f_r \, k_r}{k_a - k_r} \left(e^{-krt} - e^{-kat} \right) + D_0 \, f_i \, e^{-kat} \quad (3.36)$$

For the concentration of drug in the body:

$$\frac{dB}{dt} = \text{Drug in - Drug out} \quad (3.37)$$

$$= k_a G - k_{el} B \quad (3.38)$$

@ t = 0 B = 0

$$B = \frac{D_0 \, f_r \, k_a k_r}{(k_a - k_r)(k_{el} - k_r)} \left(e^{-krt} - e^{-kelt} \right) \quad (3.39)$$

$$+ \frac{D_0 \, f_i \, k_a - [D_0 \, f_r \, k_a k_r / (k_a - k_r)]}{(k_{el} - k_a)} \left(e^{-krt} - e^{-kelt} \right) \quad (3.40)$$

This reduces to the expression for oral administered dose with an absorption term, equation (3.16) when $f_r = 0$ and $f_i = 1$, and B = $C_B V_d$. Experimentally one can measure f_r, f_i and k_r *in vitro* for a sustained release formulation and measure C_B (plasma concentration) after administration in animal or human. These data are then fit to find k_{el}, k_a and V_d. An example that appeared in the literature was

given by Wiegand and Taylor (1960) Biochemic. Pharmacol. 3: 256 for the drug 1-(o-methoxy propyl)-4-(g-methoxypropyl) piperazine phosphate.
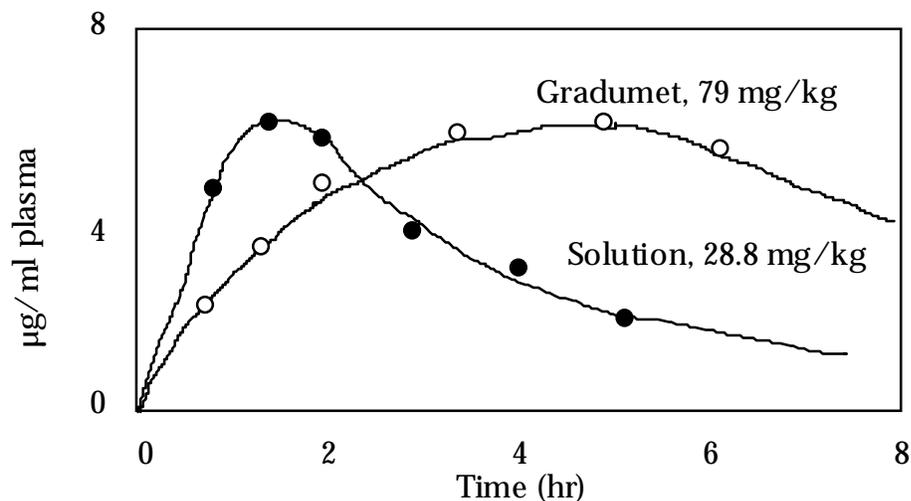
Figure 3.5.8. Comparison of sustained-release tablet and solution kinetics in six dogs. Initial dose is shown on the graph

3.6. Physiological Based Models

The compartmental models do not give any information about the drug concentrations in specific organs, and are not amenable to transfer from one species to another.   To provide some insight into these areas, some investigators have used a physiological model, in which specific organs and tissues of the body are modeled as individual compartments. An example is outlined figure 3.6.1.

The model then uses actual blood flow rates through the specific organ, together with and rates of diffusion of the drug into, and relative affinities of the drug between, the organ and the plasma.   Some examples of blood flow through organs are given in the class notes, and they range form a high of 1500 mL/min in the liver to a low of 250 mL/min in the muscles of the heart. The liver is a highly vascular organ, and since it is the detoxifying site of the body, a high blood flow rate is expected.

Some of the advantages of the physiological model include the more realistic nature of the derived parameters, the ability to change parameters to allow for physiological changes, the ability to predict drug concentrations, and elimination mechanisms for specific organs, and the ability to scale values from one species to another. These advantages are gained at the expense of a far more complicated mathematical treatment, a need for a huge data base for each drug, difficulties of obtaining tissue samples from human subjects, and the possibility of erroneous simplifying assumptions. As a result this approach has only been tried for a few drugs.
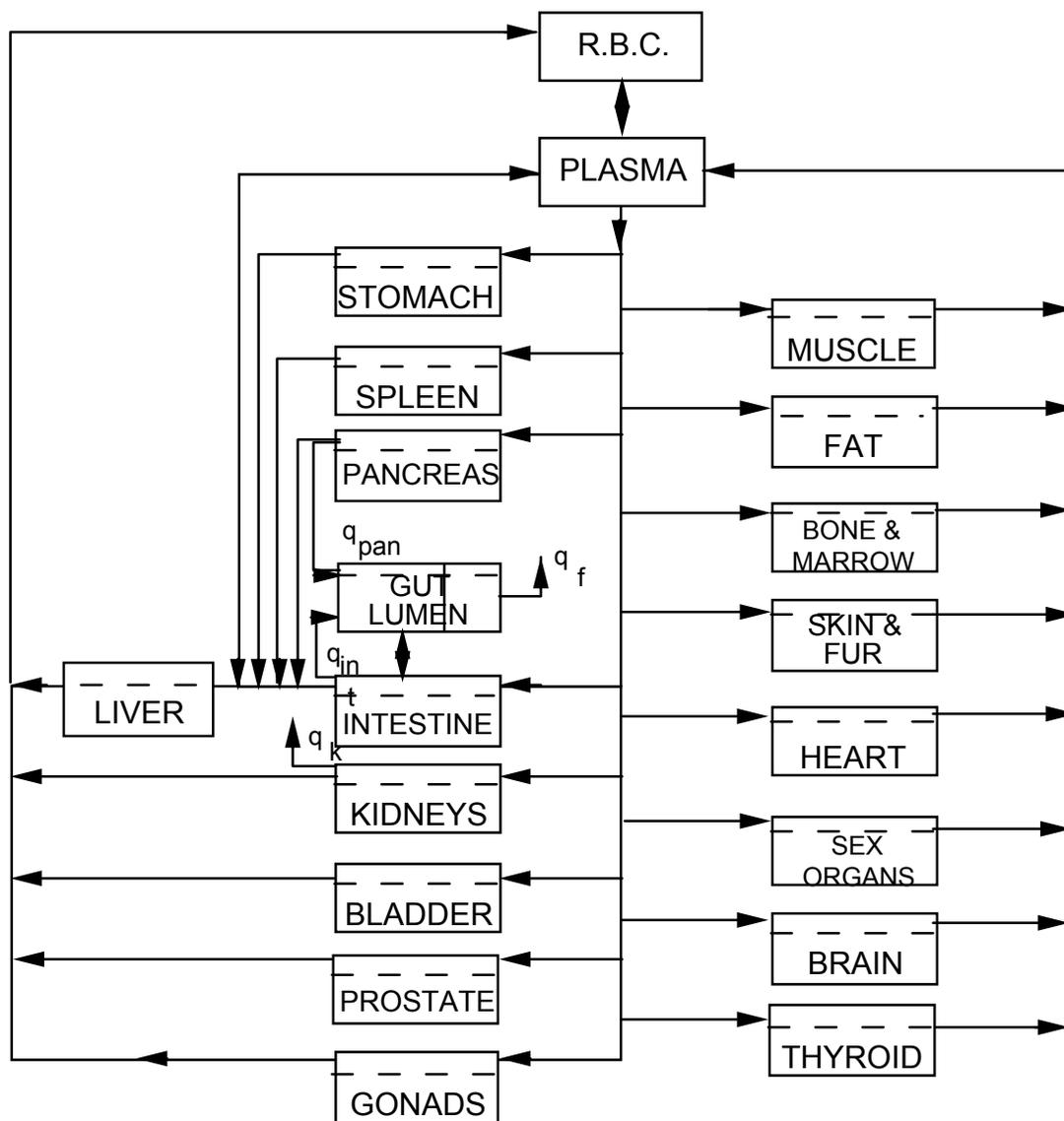
Figure 3.6.1. An example of a comprehensive physiologically based pharmacokinetic model flow scheme after Jain et al. Annals of Biochemical engineering vol. 9 (1982 pp 347-361.

3.7.   Model Independent Approach

The most recent   approach to pharmacokinetic modeling is the Model Independent approach. This is a purely mathematical approach which avoids invoking and kinetic parameters, and thus avoids using any that may be erroneous. This approach gives good rates and plasma clearances, and with some difficulty, overall volumes of distribution, but gives no information about specific rates in individual organs.

3.8.   Non-linear Kinetics

Most of these models assume that the adsorption and/or elimination of the drug is first order, and that a true first order rate constant and half life can be calculated. In reality any drug that is eliminated by other than passive diffusion will involve an energy intensive step, usually enzymatic, which is saturable at high concentrations. It is fortunate that most drugs are given at concentrations which are well below saturation levels, so that pseudo first order kinetics usually describe the situation well.   However some drugs, for example phenytoin, salicylate, theophylline and probenecid, exhibit saturable kinetics in the therapeutic range, and this can have a profound effect on both levels and circulation times.   In these cases Michaelis-Menten type kinetics must be invoked. The implications in adsorption and elimination of regular versus slow release formulations are large. A interesting example of a drug that is consumed at well above saturation levels is alcohol. It is eliminated from the body at an apparent zero-order rate equivalent to 12 oz of beer (equivalent to 8g of pure alcohol), (1 oz of liquor) per hour. At much lower levels a first order rate constant is obtained. This reflects the fact that alcohol is eliminated by enzyme mediated oxidative metabolism in the liver.

Situations in which nonlinear kinetics are observed include; drug absorption involving drugs with low solubility, slow dissolution rates saturable active absorption processes, rates influenced by fluctuating intestinal blood flow rates and changing pH and situations where drug binds to the gut wall, or food interferes with absorption; drug metabolism involving enzymatic reactions, alternate metabolic pathways and drug-drug interactions at the metabolic level; tissue binding by a saturable, or irreversible process, and drug/protein binding with displacement; and finally drug elimination by saturable excretory mechanisms, interference from other drugs or endogenous compounds, and altered renal function or GI motility due to pharmacological of physiological factors.

3.9.   Application of Pharmacokinetic Models

The one compartment model is a good basis for many common drugs. Example include, nortriptyline, pentobarbital, phenytoin, salicylic acid, theophylline, and warfarin. This model fails for some common drugs such as ampicillin, diazepam, lidocane, propranolol and cephalexin.

The two and multi-compartment model has been shown to hold well for amphetamine, chlordiazeperoxide, digoxin, digitoxin, epinephrine, ethchlorvynol, gentamicin, lidocaine, methicillin, ouabain, oxacillin, pentaerythritol, sulfamethazine, sulfisoxazole, theophylline & warfarin.

The physiological model has been used to model anti cancer drugs such as methotrexate and actinomycin D.

Further Reading Suggestions

Welling, P.G. Pharmacokinetics Process and Mathematics. ACS. Washington, DC. (1986)
Notari, R.E., Biopharmaceutics and Clinical Pharmacokinetics. 4th ed.
Marcel Dekker in, New York (1987)
D. Wise in Medical Application of Controlled Release vol II. CRC press Boca Raton, Florida (1984)

Questions you should be able to answer

• Erythromycin has a pKa of 8.8. Which area of the GI tract will it more easily pass through the wall, the stomach or the intestine? How about ampicillin with a pKa of 2.5?
• In cases of drug overdose, the urine of patients who have consumed basic drugs is made alkaline by administration of sodium bicarbonate, or made acid by administration of ammonium chloride for acid drugs. Can you explain the rational behind this treatment?
• Cholestyramine is a nonabsorbable ion exchange resin that can irreversibly bind digoxin, the heart drug. When Cholestyramine is fed to patients who are receiving digoxin, the elimination half life of digoxin is markedly reduced. What is the explanation for this?

Chapter 4.   Controlled Release Technology

4.1.   Rational and Approaches

In the book "Controlled Drug Delivery" the authors define the rational for sustained or controlled drug delivery as 'to alter the pharmaco-kinetics and pharmacodynamics of pharmacologically active moieties by using novel drug delivery systems or by modifying the molecular structure and/or physiological parameters inherent in a selected route of administration.'. In other words, the kinetics of drug uptake and the maintenance of plasma drug concentration is determined not by the properties of the drug, but by the properties of the delivery system. This affords a level of control over the concentrations of drug in the body, that is not available in conventional methods, with the exception of metered intravenous administration.

As we saw in chapter three, the traditional methods and routes of drug administration result in a rate of drug uptake that is controlled by the drug properties (solubility, charge, molecular size, etc.) and the characteristics of the site of administration (pH, surface area, presence of enzymes, active transport mechanisms etc.). The profiles of drug levels in the plasma are also influenced by the rate of the various elimination processes. Figure 4.1.1 shows the plasma drug profile for drug administered by repeated oral doses (traditional) and compares it with the profile from a sustained release device and one delivering drug by zero order kinetics. The traditional method results in peaks and valleys, which are absent in the controlled and sustained formulations.
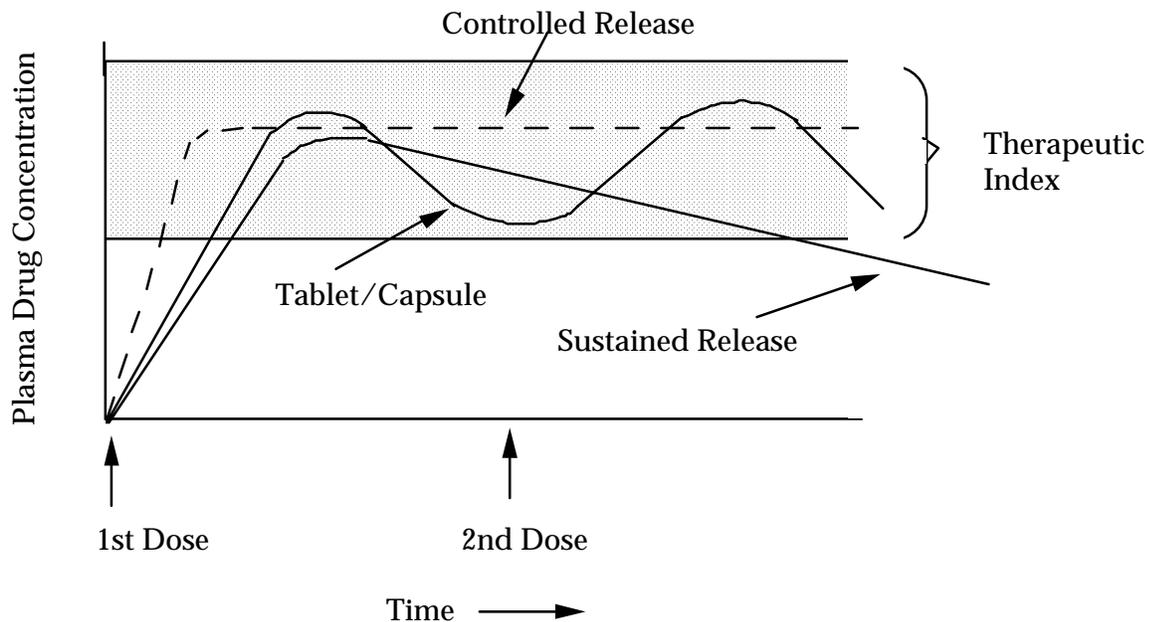
Figure 4.1.1. Concentration profiles for drug delivered by tablet, sustained release device or controlled release device.

The type of delivery vehicle that is chosen for investigation will depend on many factors, including the properties of the drug, its pharmacodynamics and pharmacokinetics, the route of delivery and the nature of the treatment and the disease. Delivery of Insulin is a challenge because it is only required when blood glucose levels exceed normal levels, and so a responsive drug delivery device would be optimal. However, as we have discussed, Insulin is a large molecular weight drug, a protein that needs to maintain its active configuration, and one that is susceptible to the protease enzymes in the body, resulting in a half life of about 25 minutes.

4.2. Classes of Drug Delivery Devices

The types of drug delivery device that have been and are being developed can be classified into four groups, depending on the method used to modify the drug properties.

4.2.1. Polymeric Systems

A great deal of research has been conducted into the use of polymers to develop means of controlling delivery of drugs to the human body. Within this class there are four sub-divisions, classified by the mechanism which controls the rate of drug release.

4.2.1.1. Diffusion Control
In many cases rate of release of a drug from a device is controlled by the rate at which the drug diffuses through a polymer. This class is further subdivided according to how the drug is held in the polymer. If drug is retained in a reservoir it is known as a reservoir device. However, when drug is dispersed throughout the polymer, it is known as a matrix device.

4.2.1.2. Chemical control
Polymeric drug delivery devices have been developed in which the polymer is susceptible to breakdown by agents present in the body (usually water, or enzymes). Two scenarios exist, one in which drug is held inside a biodegradable polymer, and one in which the drug is chemically linked to a polymer by a bond that is biodegradable. The later are known as pendant chain systems, and the polymer is often soluble.

4.2.1.3. Solvent control
Another mechanism for controlling the release of drug from a device involves the use of a solvent, which in the case of the human body is water. In one class the

solvent diffuses into the device and causes swelling, i.e. swelling controlled. In another case the device contains a high concentration of salt to which the solvent is attracted by osmosis through a semi permeable membrane, and the resulting influx of water is harnessed to expel a solution of the drug out of an orifice in the device.

4.2.1.4. Externally activated or modulated
The last class of polymeric device is being developed in order to give some form of control to the release that can be triggered by external means. Two examples include the use of magnetic beads, implanted in a polymer along with the drug, and activated by subjecting the beads to an external, oscillating magnetic field, and the use of externally applied ultrasound energy to increase drug release from polymer-encapsulated drug.

4.2.2. Drug Modification/Protection
An alternative approach is modify the drug so that it is protected form the environment until it is needed, or the environment is protected from it, (as would be the case for drugs that are toxic to certain organs).

4.2.2.1. Produg formation
A prodrug is an in active from of a drug which is converted to the active form in the body. A classic example of a prodrug is the drug L-dopa, used in the management of Parkinson's disease.

4.2.2.2. Conjugation to a homing molecule
A great deal of interest has been generated by the prospect of developing a magic bullet that will home in on the disease area and deliver drug. This is particularly attractive in the case of cancer drugs, which are often highly toxic to many normal cells in the body, as well as the cancer cells. In this way the drug can be made to avoid areas in which it would cause damage, at the same time delivering a large payload to the target site.

4.2.2.3. Liposomal entrapment
It is possible to form micro-vesicles that contain a drug. These bilayered micro-vesicles, known as liposomes, are formed spontaneously when phospholipids contact an aqueous environment. Phospholipids are naturally occurring amphipathic molecules (contain both a hydrophilic and hydrophobic region). If they contact a solution of drug, they will form a vesicle entrapping some of the drug solution in the interior.

4.2.3. Pumps

One can imagine that if drug is in solution it would be very convenient to pump it into the body through a thin catheter. This is in fact the method of choice in hospitals, and currently there are sophisticated syringe pumps which are available to be programmed to pump a drug solution through an intravenous line. This, however, requires that the recipient be under constant supervision by trained personnel, and also that the protective barrier of the skin is broken to allow insertion of a needle. This means that intravenous injection of a drug is an

expensive proposition, which can lead to complications such as infections. One of the first attempts to regulate insulin in diabetes centered on the development of a implantable pump containing insulin solution which could be delivered in a metered doses into the blood stream. For the system to exactly mimic the delivery system of the body requires a sensor which can measure blood glucose level and feed this information back to the pump to administer insulin only when blood glucose levels are high. This is a closed-loop, feedback mechanism. Early versions were simpler, and merely pumped a continuous low, or basal dose of insulin, and the patient was required to administer external insulin by the usual injection routes after a meal.

4.2.4. Modification of the site of delivery

It is also possible to control how the body takes a drug up by changing either the properties or the location of the site of administration.

4.2.4.1. Use of enhancers

Many drugs can not pass through barriers such as the skin, or the nasal mucosa, so one approach to deliver them by these route is to change the skin permeable either by the use of chemicals known as enhancers, or by physical methods such as the use of electrical current (iontophoresis) or ultrasound (sonophoresis).

4.2.4.2. Shift in absorption zone

It is also possible to change the site of delivery by protecting the drug so that it will pass the first site, and arrive at a site where it is more likely to be absorbed. This is the case with drugs that are absorbed through the intestine, but not throughout the stomach, where they arrive from the mouth. If the drug is destroyed in the stomach, it must be protected, and polymeric coatings have been developed that remain in tact in acid conditions, but break down in the alkaline environment to the upper intestine. These coatings are known as enteric coatings.

4.3.   Modeling of the Release Kinetics of the Devices

Detailed mathematical models have been developed for most of the drug profiles obtained from the devices described in section 4.2.

4.3.1 General Models

We are all familiar with the phenomenon that is observed when a drop of dye is added to a beaker and in time the color in the beaker becomes homogeneous, when the dye molecules have distributed themselves throughout the entire volume. This phenomenon is called diffusion, and it is due to the random movement of molecules in the solution. The tendency is for molecules to move from an area of high concentration to one of low concentration, that is they move against the concentration gradient. This is because initially there are far more molecules in the area of high concentration, and therefore the probability that

they will move away is far greater than the probability that a molecule will move into that area of high concentration.

4.3.1.1. Fick's first law

The flux (amount of solute i crossing a plane of unit surface area, normal to the direction of transport, in unit time) J, can be described by Fick's law which in the differential form is:

$$J_i = - D_i \left(\frac{dc}{dz}\right) \quad (4.1)$$

where D_i is the diffusion coefficient,
   c is the concentration of solute mol/cm³,
   z is the position in the device.

In the integral form Fick's Law becomes:

$$J_i = - D_i \frac{K\Delta c}{\delta} \quad (4.2)$$

where K is the partition coefficient which is the ratio of concentration of drug in the polymer, to that in the solution
δ is the device thickness.

The term $D_i K/\delta$ is referred to as the permeability coefficient, P. It is a relatively easy parameter to measure, since from (4.2) it can be seen to be equal to the flux divided by the concentration difference across the membrane. D_i, K and P all relate to the solute properties such as molecular size, polarity, and solubility in the polymer phase and to the structure of the polymer.

$$P = \frac{D_i K}{\delta} \quad (4.3)$$

4.3.1.2. Fick's second law

For unsteady state situations, where concentration is changing with time, we use Fick's second law, which assumes constant D_i and constant boundaries (i.e. no polymer swelling):

$$\frac{\partial c_i}{\partial t} = D_i \frac{\partial^2 c_i}{\partial z^2} \quad (4.4)$$

4.3.1.3. Stokes Einstein equation

The diffusion coefficient D is a very important parameter in drug delivery. It is related to the size of the molecule (radius r) and the temperature (T absolute) and viscosity (η) of the liquid through which the species is diffusing. The relationship is known as the Stokes-Einstein equation:

$$D = \frac{k_B T}{6\pi r \eta} \quad (4.5)$$

Where k_B is the Boltzmann constant.

The radius is related to the molecular weight of the molecule, and the diffusion coefficient is inversely proportional to the cube root of the molecular weight.

For Insulin the **diffusion coefficient is 8.2 x 10^{-7} cm^{2}sec^{-1}**.

4.3.2. Diffusion through polymers

4.3.2.1. Effective diffusion coefficient

When we are considering diffusion through polymers, it is usual to use an effective diffusion coefficient D_{eff} which is related to the diffusion coefficient in water-filled pores D_{sw}, and is given by equation (4.6):

$$D_{eff} = D_{sw} \frac{\varepsilon \quad K_r}{\tau} \quad (4.6)$$

where ε is the void fraction (porosity),
τ is the tortuosity and
K_r is the restriction coefficient which relates
r_p the peptide radius to
r_s, the average pore radius, by
$\lambda = r_s/r_p$.

4.3.2.2. Semi crystalline polymers

For diffusion in semi-crystalline polymers, a modified diffusion coefficient must be used. The crystalline areas are a barrier to diffusion, and can even block the passage of large solutes.

$$D_{ic} = \frac{D_{ia} \, \underline{v}_a}{\psi} \qquad\qquad (4.7)$$

Where D_{ic} is the diffusion coefficient of solute in the crystalline regions

D_{ia} is the diffusion coefficients in amorphous regions, and

v_a is known as the blocking factor and

ψ the detour factor.

These effects can be visualized in diagrammatic form in figure 4.3.1.
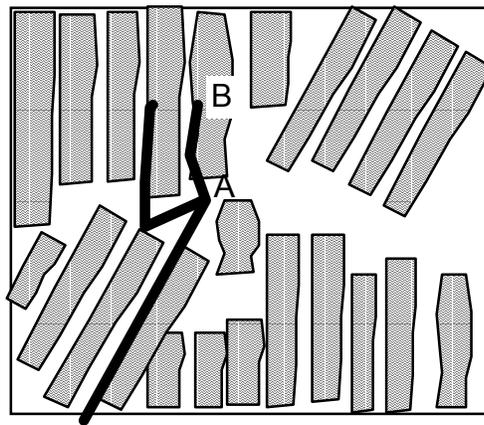
crystalline area

Path of diffusion

A   Detour point
B   Blocked point

Figure 4.3.1. Blocking and detour effect of crystalline regions in a polymer.

4.3.3. In vitro measurements

The mechanisms and rate of diffusion can be studied in the laboratory (in vitro) using special equipment known as a diffusion cell. When a small molecule or peptide is involved the free volume theories of molecular motion can be invoked. For larger peptides (above 1000 molecular weight), and proteins the molecules diffuse through the polymer structure that is unique to the polymer in question. Proteins are said to diffuse by reptation, which as the name implies, involves a reptile-like movement through the available intermolecular space.

Measurements of the rate of diffusion are made in diffusion cells. Figure 4.3.2 shows the basic setup. The polymer of interest is cast into a membrane which is clamped between two chambers, which are filled with buffer. The solute of interest is introduced into the donor compartment, and samples of the buffer from the acceptor side are taken at subsequent time points. The compartments are kept well stirred, and sample size is kept small so that the reduction in acceptor chamber buffer volume is insignificant.
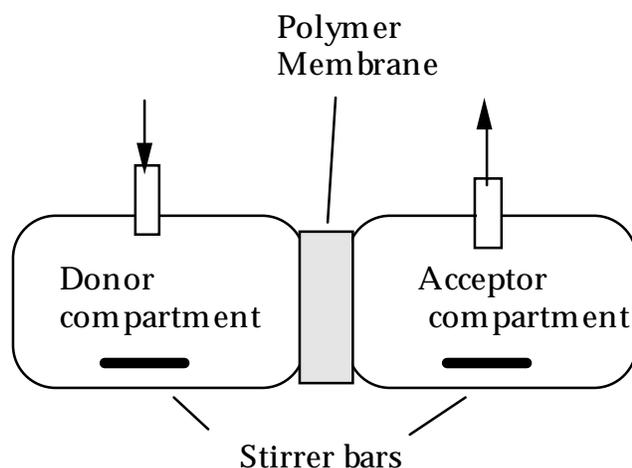
Figure 4.3.2. A typical diffusion cell

4.4.   Specific Systems

Of the four main approaches to controlled release, polymeric devices, drug modification, pumps and site modification, the area that has received most attention is the use of polymers to regulate the delivery profile. In describing the systems, we will go into the greatest detail with these polymeric systems since they have received the lions share of attention.

4.4.1. Reservoir Systems

One of the most widely used polymeric system is called the reservoir system, and it has found commercial success as skin patches. Reservoir systems could also be in the form of microcapsules or hollow fibers. Essentially the drug is retained in a central compartment and surround by a polymeric membrane, through which it must diffuse, and which therefore controls the rate of delivery. Figure 4.4.1 describes the system.
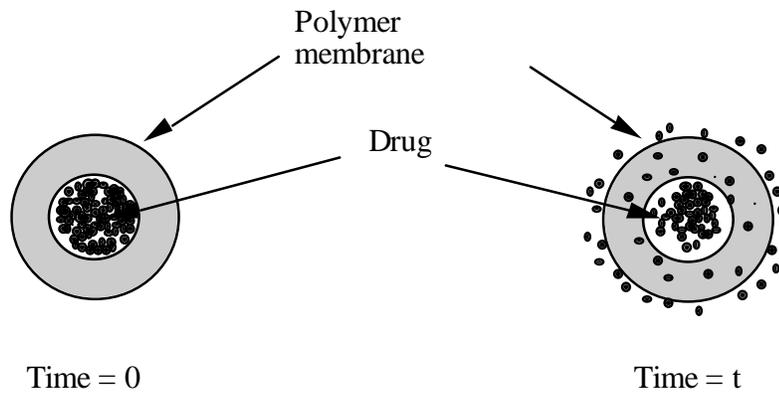
Figure 4.4.1. The reservoir system

The most common polymers that are used are silicone, ethylene vinyl acetate and hydrogels. One disadvantage of these polymers is that if they are implanted in the body, they must be removed once the drug is exhausted. In addition, reservoir devices are not suitable for delivery of high molecular weight drugs, because the drug will not diffuse through the polymer. However, the major disadvantage is that if they contain a potent drug, and the device were to rupture, a potentially fatal dose of drug would be administered.

Drug is released from a reservoir system by first dissolving in the polymer membrane, diffusing across it., and dissolving in the fluid on the other side. The diffusion is down a gradient from the highly concentrated core to the very low concentration of the body. Since body fluid is constantly being exchanged, the conditions which most mimic the in vivo situation are what are called, infinite sink conditions. This means that drug concentration at the body side of the membrane is essentially zero.
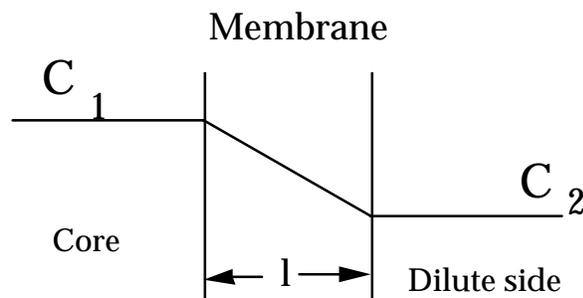
Figure 4.4.2. Representation of the concentration gradient across a membrane.

The flux (g/cm^{2}sec), can be described by Fick's first law:

$$J \qquad\qquad = \qquad\qquad -D \qquad \frac{dC}{}$$

(4.8)

$$\frac{dz} {} = \frac{-D}{l}(C_2 - C_1) \quad (4.9)$$

where l is the membrane thickness

C_2 is the concentration of drug in the core

C_1 is the concentration of drug at the membrane/body interface.

Frequently the drug will preferentially partition into the membrane, in which case the situation is as depicted in Figure 4.4.3.
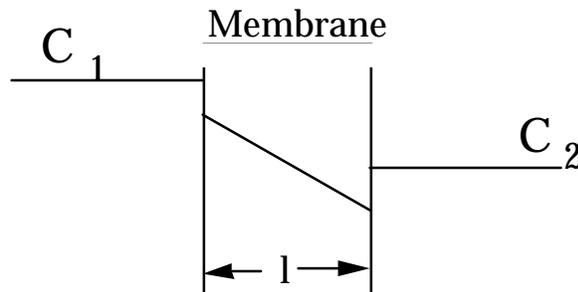
Figure 4.4.3. Preferential partitioning into a membrane.

In this case flux is modified to include a partition coefficient , K, and can be written as:

$$J = -\frac{DK}{l}\Delta C \quad (4.10)$$

Here ΔC is the difference in concentration between the concentrated and dilute sides of the membrane. As before the term DK/l is referred to as the permeability coefficient, which is equal to $J/\Delta C$, which is a straightforward parameter to measure.

If we are dealing with a slab with a surface area A, with a constant concentration of drug in the reservoir, and infinite sink conditions in the body, we can write:

$$\frac{dMt}{dt} = \frac{ADK\Delta C}{l} \quad (4.11)$$

Where M_t is the total amount of drug released at time t.

Similar rate equations can be derived for a cylinder and a sphere. For a cylinder of height h and outside and inside radius r_o and r_i respectively:

$$\frac{dM_t}{dt} = \frac{2\pi hDK\Delta C}{\ln(r_o - r_1)} \quad (4.12)$$

For a sphere of outside and inside radius r_o and r_i respectively:

$$\frac{dM_t}{dt} = \frac{4DK\Delta C}{(r_o - r_i)} \quad r_o r_i \quad (4.13)$$

The rate of release from any of these shapes can be affected if the solute does not get removed from the membrane/body interface rapidly. This can happen if the drug is poorly soluble. This buildup of drug will lower the concentration gradient and the flux will drop. In extreme cases if buildup is so bad that the drug reaches saturation concentration at the membrane surface, then release will stop

Another situation that affects the release rate is the case of a freshly prepared membrane. Just after manufacture the membrane contains no drug, and a steady state situation has not developed. The release at time t during this initial phase can be described by:

$$M_t = \frac{ADK\Delta C}{l} \left(t - \frac{l^2}{6D} \right) \quad (4.14)$$

On the other hand when a membrane has been stored for a long period of time, drug will accumulate in the membrane, and this will result in a burst of release at the initial time points:

$$M_t = \frac{ADKC}{l} \left(t + \frac{l^2}{3D} \right) \quad (4.15)$$

If the membrane dimensions remains the same, i.e. if it does not swell, and if the core concentration of drug remains saturated, and if infinite sink conditions are maintained at the membrane surface, then the steady state rate of drug release will be a function of (and can be controlled by), the area, the diffusion coefficient of the drug, the partition coefficient the concentration gradient and the thickness often membrane. The overall rate will be constant, that is it will be zero order with respect to drug.

It is possible to derive similar equations for various different geometries of reservoir systems. Some common examples are given below.

Equation for a Slab

$$\frac{dM_t}{dt} = \frac{ADK\Delta C}{l} \quad (4\text{-}16)$$

Where: M_t is the total amount of drug released at time t

Equation for a cylinder

$$\frac{dM_t}{dt} = \frac{2\pi hDK\Delta C}{\ln (r_o - r_i)} \quad (4\text{-}17)$$

where: h is the cylinder of height
r_o is the outside and
r_i is the inside radius

Equation for a sphere

$$\frac{dM_t}{dt} = 4DK\Delta C \frac{r_o r_i}{(r_o - r_i)} \tag{4-18}$$

where: r_o is the inside radius and
r_i is the outside radius

Table 4.4.1. gives a list of some of the reservoir devices that are commercially available for drug delivery.

Table 4.4.1. Commercial Pharmaceutical Controlled Release Devices

Product and Manufacturer	Application	Type of Device
Norplant® Population Council	Subdermal implant to deliver levonorgestrel for contraception	Reservoir,hollow fiber/silicone rubber membranes
Progestersert® Alza	IUD to deliver progesterone Delivers 65 µg/day for 1 year	EVA membrane
Ocusert® Alza	Ocular insert delivering pilocarpine for treating glaucoma. Delivers 20 or 40 µg/h, 7 days	EVA membrane
Transderm®-Scop Alza/Ciba-Geigy	Transdermal Scopolamine delivers 10 µg/h for 72 h for motion sickness.	Microporous membrane
Transderm®-Nitro Alza/Ciba-Geigy	Transdermal Nitroglycerine delivery for angina	Microporous membrane

4.4.2. Matrix Systems

4.4.2.1. Description

Matrix systems consist of a polymer throughout which drug is dispersed. They are easy to make. Any leaks in the system do not create severe problems, as would leakage from a ruptured reservoir, and they can be made to release high molecular weight drugs. The main disadvantage of a matrix system is that is must be removed from the body after the drug has been released.
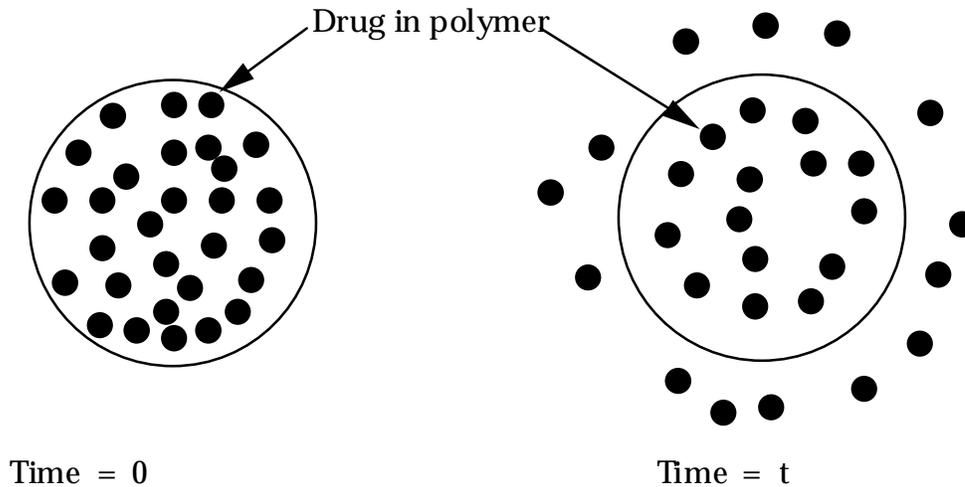
Figure 4.4.4. Matrix System (after Langer)

4.4.2.2. Modeling Matrix Systems

For modeling purposes four cases are considered. The drug can be either dissolved or dispersed in the polymer, and it can travel out of the device by either diffusion through the polymer, or diffusion through channels. The modeling in each case is somewhat different. In the case of dissolved drug diffusing through the polymer two time periods are considered an early and a late time. In the early time approximation (which usually involves about 60% of release) release is inversely proportional to the square route of time:

Dissolved drug, diffusion through polymer

Early time approximation:

$$\frac{dM_t}{dt} = 2M_\infty \left\{ \frac{D}{\pi l^2 t} \right\}^{1/2} \tag{4-19}$$

where M_t is the amount of drug released at time t
M^∞ is the total amount of drug released
l   is the length of the diffusion path
D   is the diffusion coefficient of drug in the polymer

For the late time approximation the equation is as follows

$$\frac{dM_t}{dt} = 8DM_\infty \ \exp \left\{ \frac{-\pi^2 D t}{} \right\} \tag{4-20}$$

$$\frac{d}{dt} \qquad l^2 \qquad\qquad l^2$$

Matrix devices can be made by loading drug into a polymer mixture, physical blending followed by compression or injection molding, extrusion, calendering or solvent casting. In this case drug is dispersed in the polymer, and if it leaves the polymer by diffusion through the polymer, then the model for release becomes:

Dispersed drug, diffusion through polymer

$$\frac{dM_t}{dt} = \frac{A}{2} \left\{ \frac{2 \; DC_s \; C_L}{t} \right\}^{1/2}$$

(4-21)

where A   is the exposed area of the device
       D   is the diffusion coefficient of drug in the polymer
       Cs   is the solution solubility of the drug in the matrix
       C_L is the total loading of drug in the matrix

This assumes that $C_L >>> C_S$. The device maintains a $t^{1/2}$ dependency until the active agent remaining in the device falls below saturation levels. This is a well known relationship known as the Higuchi equation after the scientist who first published this modeling in the Journal of Pharmaceutical Sciences volume 50 p 874-875 in 1961. If a device is following this type of release kinetics a plot of cumulative release against the square root of time will give a straight line.

4.4.2.3. Bioerodible systems

A specific type of matrix system is one in which the polymer is broken down by the body. This breakdown can be via hydrolysis or it can be enzymatic. They are called bioerodible systems, and they have all the advantages of matrix systems, and in addition they do not require removal from the body when release is completed.

Drug disperse in polymer

Time = 0   Time = t
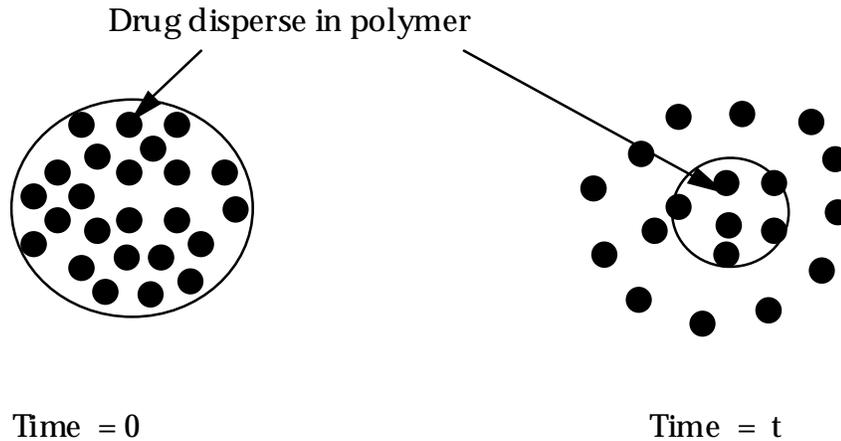
Figure 4.4.5. Bioerodible polymeric system

If the method of degradation is by surface erosion, that is similar to peeling off the skins of an onion, then release can be shown to be zero order if the surface area does not change substantially with time. However if, as often happens with hydrophilic polymers, erosion takes place by solvent infiltrating the polymer and causing chunks of polymer to break off, similar to the weathering of rock seen in the Grand Canyon, then release can be in bursts and it can be hard to control. Some of the most widely studied polymer are those used as biodegradable sutures in surgery. These polymers, known as poly lactides and poly glycolides, have been passed for us in humans by the FDA, but they have the disadvantage of eroding by bulk erosion. Poly anhydrides and polyorthoesters can be made to erode by surface erosion. In all these cases care has to be taken to ensure that the breakdown products of the polymer are non toxic, and that they do not interact with the drug.

4.4.3.   Polymers with Pendent Drugs

Another system that is being investigated, especially for cancer drugs, is that of chemically attaching a drug to a polymer backbone, usually a soluble polymer, by means of a bond that is susceptible to enzymatic or hydrolytic breakdown in the body. In many cases this changes the way that the body processes the drug, and it has even been suggested that this is a way of overcoming drug resistance in some cancer cells. In this system very large drug loadings are possible.
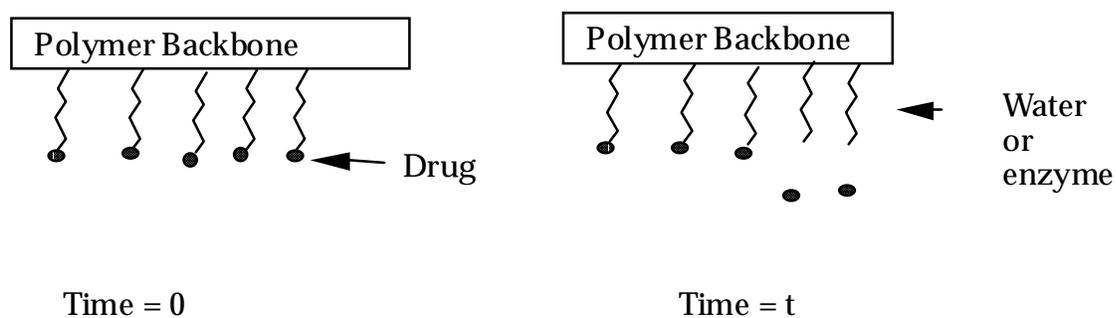
Figure 4.4.6. Polymer with pendent drug attached by susceptible linkage.

4.4.4.   Solvent Control
4.4.4.1. Swelling Controlled Systems

One way to control release of drug that is dispersed in a polymer, is to use a polymer that is normally glassy, but when solvent penetrates the matrix swelling takes place and the polymer chains become relaxed and allow the drug to diffuse out. As the solvent front advances, the area that is penetrated swells, and the glassy core area begins to shrink.
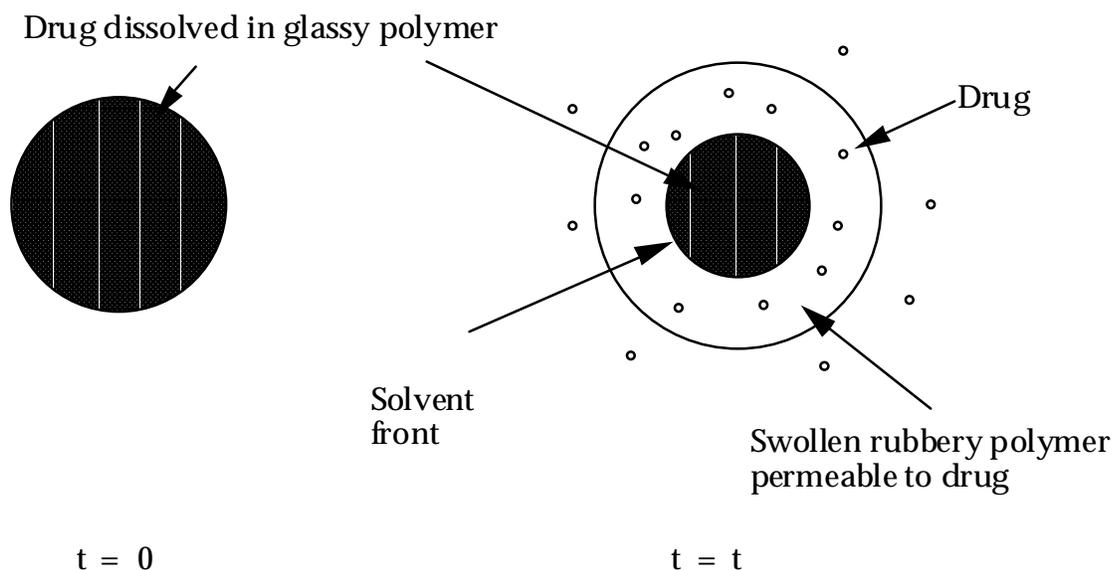

Figure 4.4.7. Swelling controlled system

Some of the advantages of these systems include the fact that the releases is a function of the polymer swelling, not of the particular drug, so that one can be used for different drugs without the need to reformulate and as a matrix system, there is no potential for catastrophic rupture with release of large quantities of drug. Also there is no burst effect. However, the number of available polymers is not large, and although for linear polymers dissolution follows the swelling process, cross linked or partly crystalline matrixes can give problems with areas that do not swell, and the resulting introduction of mechanical weakness.

4.4.4.2. Osmotic control
4.4.4.2.1. Description
Another use of solvent to control drug release involves the use of an osmotic core surrounding a flexible reservoir that contains a solution of drug. If the core is itself surrounded with a semipermeable membrane, and if the drug chamber has access to the outside via a small laser-drilled orifice, drug solution will be squeezed out when water enters through the semipermeable membrane, is imbibed by the osmotic core and the core swells.
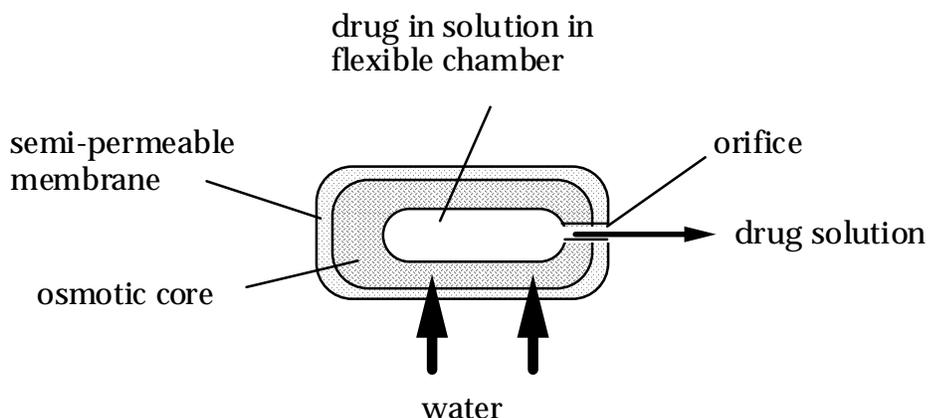
Figure 4.4.7. Schematic of an osmotic device

The advantages of these systems include the fact that osmosis is a constant driving force which results in zero order release which is independent of the environment; they can be designed to deliver any solution, including solutions of macromolecules such as proteins; and the rate can be designed to be much higher than that with diffusion alone. The best method of achieving zero order release is in the pump type of arrangement described above, although it is also possible to has a system where drug is dispersed in a polymer, together with an osmotic agent, and the water that enters forms pores in the polymer, through which drug is released. These systems tend not to be zero order. Several different patented designs have been developed by the pharmaceutical industry, some for patient use, and others for use in research.

4.4.4.2.2. Modeling of Osmotic Systems
Osmotic systems are modeled by considering the osmotic driving force developed by different substances. Table 4.4.2. lists the osmotic pressure of saturated solutions of some common examples.

Table 4.4.2. Osmotic Pressure of saturated solutions

Compound	Pressure (ATM)
Sodium chloride	356
Fructose	355

Sucrose	150
Potassium sulfate	28

Osmotic pressure $\Delta\pi$, is related to concentration (N moles per volume V or N/V) by:

$$\Delta\pi = \frac{N}{V} = CRT \quad \text{(4-22)}$$

The rate of change of the volume V with time t is related to the area A, the thickness δ and the difference between the osmotic pressure and the hydrostatic pressure by:

$$\frac{dV}{dt} = \frac{A}{\delta}(\sigma\Delta\pi - \Delta P) \quad \text{(4-23)}$$

where σ = the reflection coefficient

σ = 0 for a course filter and

= 1 when the membrane is impermeable to solute

For a large orifice $\Delta\pi >>> \Delta P$

$$\frac{dV}{dt} = \frac{A}{\delta}L_p\sigma\Delta\pi = \frac{AK\Delta\pi}{\delta} \quad \text{(4-24)}$$

where L_p = membrane permeability
   K = permeability coefficient

To find the amount released with time:

$$\frac{dM}{dt} = \frac{dV}{dt}C = \frac{AK\Delta\pi C}{\delta} \quad \text{(4-25)}$$

Note: C = saturation concentration for a saturated solution

Further Reading Suggestion

Robinson, J.R. and Lee, V. Eds. Controlled Drug Delivery, Fundamentals and Applications, Drugs and Pharmaceutical Sciences Vol 29. Marcel Dekker Inc. New York. 1987

Chapter 5.   Systems for Insulin

5.1. Current Delivery options

The current preferred method of administering insulin is via subcutaneous injection. Multiple doses are required, up to three times daily.   Different formulations are available, for example oily suspensions which have a depot effect, or formulations which consists of insulin complexed with protamine, which is a strongly basic protein from fish testes. In addition the insulin can be in either the crystalline or amorphous form, and this influences the rate of dissolution, and hence the rate of uptake. These many different formulations are necessary, because diabetics do not all respond equally to a given dose. Some diabetics have a very slow reaction time, and require a rather fast acting formulation, others react rapidly, requiring a slow acting formulation while others still are in between these two extremes.

The standard Insulin injection, a solution called USP standing for the United States Pharmacopoeia, has a rapid onset of activity of between one half and one hour, which peaks in two to three hours. After six hours the activity has subsided. Other forms with different activity are shown in Table 5.1.1. The amorphous form, or Semilente, has a similar onset to the USP, of 0.5 to one hour, but it peaks in 5-7 hours and lasts up to 16 hours and the Ultralenet form, which is the crystalline form has an onset of between 4 to 8 hours, a peak in 16 - 18 hours, and a duration of about a day, and finally the Lent form, which is a mixture of the two other forms, has an onset and duration between the amorphous and the crystalline form.   With careful titration of the patient, is possible to limit the number of daily injections for the patient.   Another long acting form is the complex with globulin from beef blood, and zinc. This form has a two hour onset peaks between 8 - 16 hours and lasts for about a day. The longest acting formulation is the protamine zinc suspension, which has a slow onset between 4 to 8 hour after injection, and a peak between 14 - 20 hours and a 36 hour duration. In the table, Normglycemic activity refers to the time at which insulin activity is evident.

Table 5.1.1. Timing of different insulin formulations

Insulin Preparations	Normglycemic activity		
	Onset (hrs)	Peak (hrs)	Duration (hrs)
Insulin injection USP	0.5 - 1.0	2 - 3	6
Semilente	0.5 - 1.0	5 - 7	12 - 16
Lente	1.0 - 1.5	8 - 12	24

Ultralente	4 - 8	16 - 18	>36
Globin-Zn	~2	8 - 16	~24
Isophane suspension	1 - 1.5	8 - 12	~24
Protamine suspension	4 - 8	14 - 20	~36

5.2. Novel Approaches

5.2.1. Alternate routes

Many new approaches are under investigation to improve the administration of insulin. One involves the search for an alternate route of administration. The usual method of administration by subcutaneous injection results in 80% of the dose being absorbed form the injection site. Table 5.2. compares the efficiency of different sites for administering insulin. When a protein is administered by mouth only 0.5% of the dose is absorbed intact and of that none is available to the systemic circulation as active drug. When a protein is delivered through the nasal membranes, thirty percent of the dose is absorbed, but very little is active. If an enhancer chemical is used to open up the membranes, then 10 - 15 % is available in the active form, and if a substance is used which causes the drug to stick to the nasal membranes, a so called bioadhesive, then 30% is bioavailable. This contrasts strongly with the situation in the mouth, where buccal bioavailability is a mere 0.4%, or the eye where it is only between 5 and 15%.

Table 5.2.1. Efficiency of different novel routes of administration

Route	% or dose absorbed	% of dose bioavailable
Buccal	-	0 - 0.4
Nasal (no bioadhesive)	30	0 - 2
Nasal (with bioadhesive)	-	30
Nasal (with enhancer)	-	10 -15
Oral	0.5	0
Vaginal	18	-
Subcutaneous	80	-
Ocular	-	5 - 15

When comparing different sites of administration it is useful to talk about a permeability coefficient P, defined as the mass of drug penetrating per unit area, time and concentration. This can be measured for insulin across various different membranes. In the mouth, the buccal permeability coefficient is 2 times 10^{-9} which translates into a total of 2.4×10^{-5} mg penetrating the membranes per minute. In the nose the permeability coefficient is between one and two times 10^{-6} which translates into a total of 0.0 24 mg penetrating the nasal membranes per minute. The G.I. tract has a permeability of 10^{-7} which results on a penetration of 6 times 10^{-4} mg per minute. If one were to instill insulin into the eye, the ocular permeability coefficient is similar to the GI permeability coefficient .

Permeability Coefficient:

$$P = \frac{M}{A\,t\,C}$$

(5-1)

where: M = mass penetrating
A = Area
t = time
C = concentration

The values for permeability in different sites are summarized in table 5.3.

Table 5.2.2.   Permeability and total amount of drug penetrating different sites

Route	Permeability coefficient	Mass penetrating (mg min^{-1})
Buccal	2×10^{-9}	2.4×10^{-5}
Nasal (rat)	$1\text{-}2 \times 10^{-6}$	2.4×10^{-2}
G.I.	10^{-7}	6×10^{-4}
Ocular	10^{-7}	3×10^{-3}

5.2.2.   Transplantation

Since it is the cells in the pancreas that are damaged in diabetes one could envision that pancreas transplantation might be a treatment option for diabetics. In 1966 the first whole pancreas transplantation was performed, but it was not successful. Since then there have been other attempts, with limited success. One of the main problems, apart from the fact that the pancreas is a very fragile organ, is that, as with any organ transplantation, there is a low availability of donor organs. Also there is a very high cost and the necessity for immune suppression to prevent rejection. In addition, diabetics often have circulating antibodies against their own islet cells, and this can lead to destruction of implanted islets.

In some cases there have been attempts to transplant only the islet cells, but again there are problems of rejection and supply. A far more promising line is to microencapsulate the islets before transplantation so that they are protected from immune attack. This will be discussed later.

5.3.   Experimental devices

Research into development of new and improved insulin delivery systems is quite extensive, with many different approaches reported in the literature. In this section we will briefly look at some of the ideas, although only a few have been investigate in humans. This will give you a flavor for the type of research that is being conducted in the drug delivery field as it relates to diabetes management.

5.3.1. Compressed tablets

One attempt to prolong the time range of an insulin dose was reported by Wang et al. in 1987. They made tablets by compressing insulin and cholesterol, and the resulting tablets were implanted under the skin of rats which had been made diabetic by injection of a substance which destroys the islet cells. They showed that the blood glucose levels dropped from a high of about 25 mMolar to below 10 mMolar after implantation. The implant lasted for about 15 - 20 days before a further implant became necessary.   The data are shown in figure 5.3.1. The advantage of this type of device is that it is composed of naturally occurring substance, although the advisability of increasing anyone's cholesterol intake might be questioned.
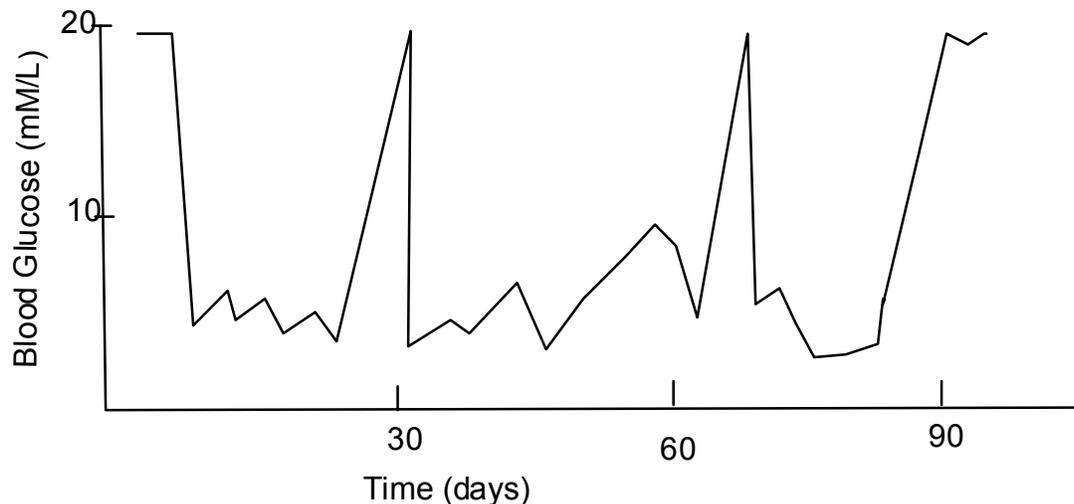
Figure 5.3.1.   Blood glucose levels in diabetic rats after implantation of compressed

cholesterol/ insulin pellets. Fresh pellets were inserted every 30 days.

(after Wang, Implant Diabetes 36:Sept. '87)

5.3.2. Sodium alginate/poly-L-lysine microcapsules

Another approach was taken by Leung et al. in which they microencapsulated insulin in an alginate capsule which was then coated with a rate controlling membrane of poly-L-lysine. They observed that with the highest loading of 10 mg of insulin in the capsules, release was 60% in the first day, and essentially complete by day 4. Decreasing the loading did not have a substantial influence on the release pattern. These experiments were performed in vitro, that is in a test tube in the lab as opposed to in vivo, which refers to experimentation in live animals. Figure 5.3.2. shows the results.
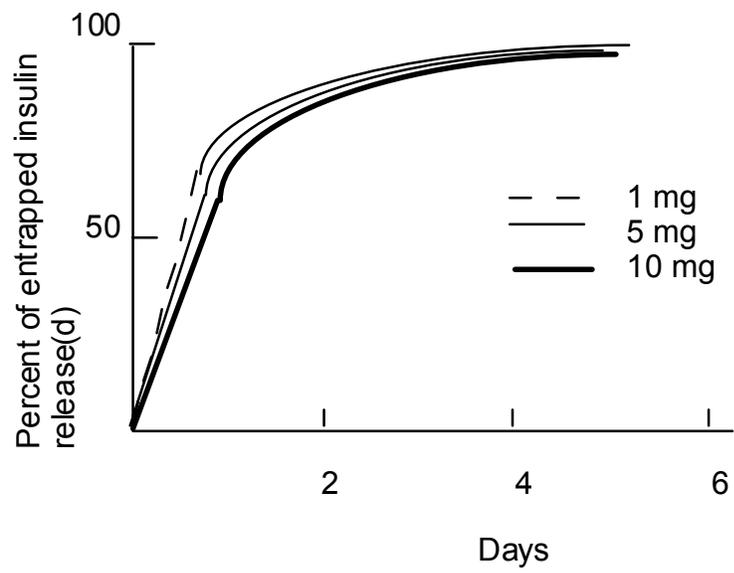
After Leung et al. Artif.Orans 7(2) 1983

Figure 5.3.2. Release of insulin from sodium alginate capsules.

5.3.3.   Encapsulated Islet cells
A great deal of research has been conducted into the possibility of entrapping viable islet cells (the cells in the pancreas that actually make insulin) and coating the capsule with a membrane that will allow glucose and insulin to pass, but will exclude large molecular weight species such as immunoglobulins which are involved in the rejection process.

5.3.3.1. Alginate microcapsules
The largest body of work is with the use of the polymer sodium alginate. This is a water soluble polymer extracted from sea weed, in fact you have probably all eaten it because it is a widely used thickener in the food industry and finds its way into many products from instant pudding to ice cream. The polymer strands are made up of repeating units of two different sugar molecules called **guluronic** and **manuronic** acid. The acid refers to the carboxylic acid group on each sugar molecule. At pH 7.4, which is physiological pH (the body pH) this acid group is negatively charged, and an ion with more than one positive charge is added, such as the calcium ion which has two positive charges, the carboxylates from adjacent strands will ionically bond with the calcium, forming a cross link between the strands. This happens all along the polymer and the result is a fairly rigid gel.

Islet cells are extracted from the pancreas by a somewhat elaborate process involving collection of the pancreas, enzymatic digestion of the collagen that binds the islet cells together and hand picking.   They are then suspended in sodium alginate solution which is made up in a buffer which is at physiological pH and contains all the salts that are found in the body fluid. This is called phosphate buffered saline, or PBS. The entire mixture is then is taken up in a syringe which is fitted with a needle that is adapted so that a coaxial stream of sterile air an pass down its shaft. The entire assembly is placed in a syringe pump, and solution containing the islets is sprayed into a beaker containing 1.2% calcium chloride solution at pH 7.4. This is shown in figure 5.3.3.
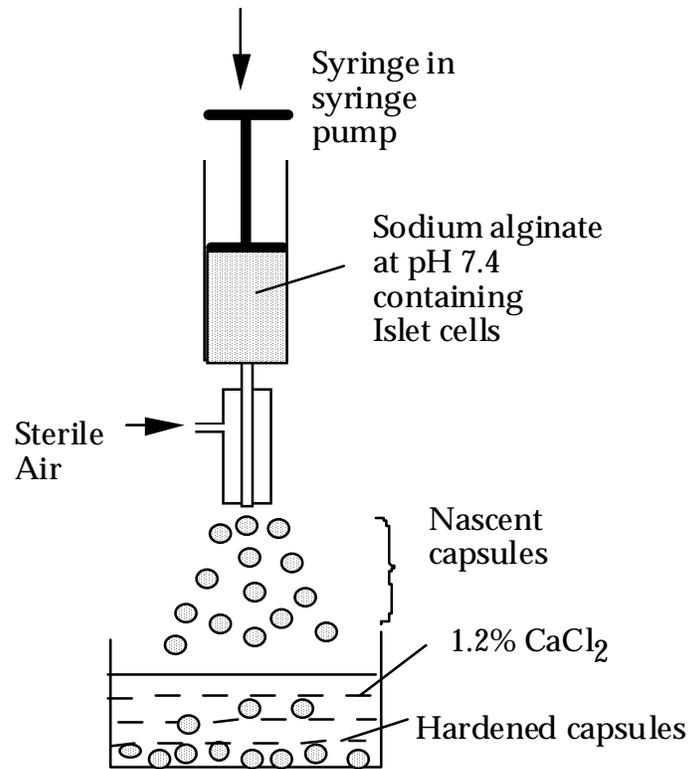
Figure 5.3.3. Apparatus for encapsulation of islet cells in alginate

When the small drops of alginate containing one or two islets contact the chloride solution they gel. These gelled beads are then coated with a polycationic skin by swirling them in a solution of the poly amino acid poly-L-lysine. The resulting capsule is represented in figure 5.3.4.
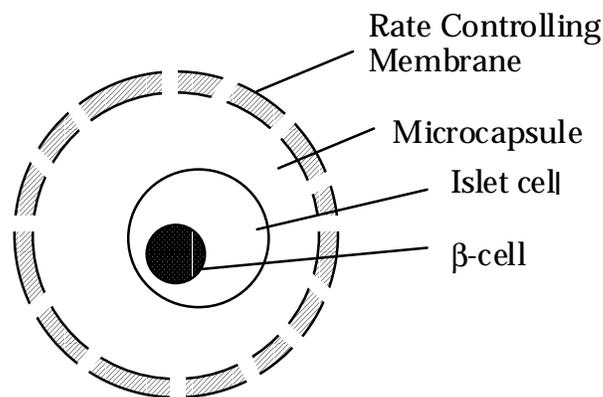
Figure 5.3.4. Schematic of an encapsulated islet cell

Before implanting these encapsulated islets into rats, researchers investigated how they respond in the test tube, that is in vitro, to a glucose challenge. The graph in figure 5.3.5. shows that when glucose was added to the solution, insulin

was released. In vivo, diabetics rats have been kept normal for over a year with implanted microencapsulated islet cells
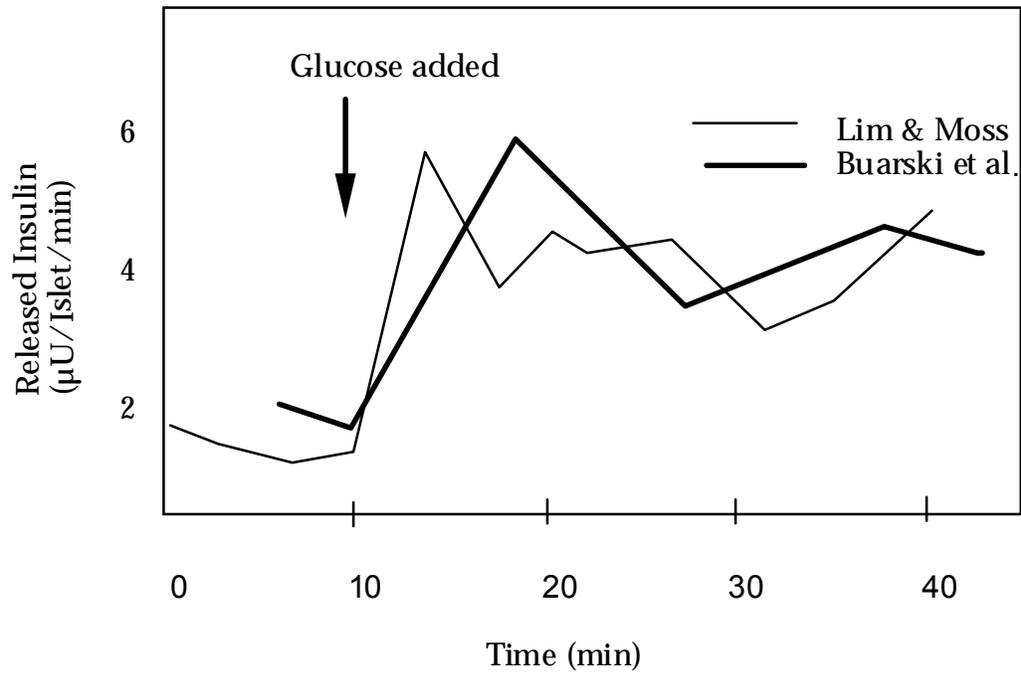
Figure 5.3.5.   Insulin release from encapsulated islets when stimulated by glucose.

5.3.3.2. Acrylic-methacrylic acid microcapsules

Researchers in Canada have used a commercially available copolymer of acrylic and methacrylic acid, known as Eudragit RL to encapsulate islets by a dual barrel syringe method. Islets encapsulated in this way continued to excrete insulin for over three weeks at a steadily decreasing rate. They then used a milder encapsulation process and the copolymer HEMA-MMA. These capsules were soft and elastic and more water compatible. The capsules secreted insulin for 24- 30 days at one third the rate of unencapsulated controls, as shown in figure 5.3.6..
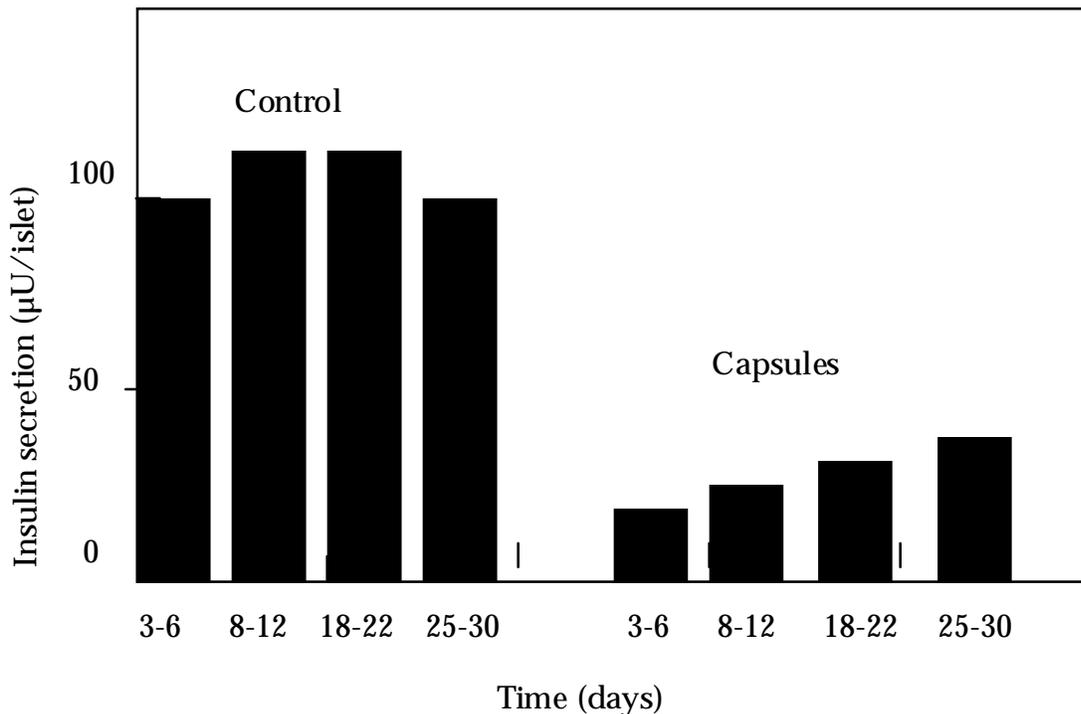
Insulin secretion from microencpsulated islet cells.
After Sefton et al. Proc. 4th Conf. Polymers in Medicine

Figure 5.3.6. Response of islet cells encapsulated in acrylic-methacrylic acid copolymer.

5.3.3.3. Agarose microencapsulation

A company, Iwata and associates, has explored the encapsulation of islets in agarose gel capsules. Agarose, like alginate, is a sugar polymer, and encapsulation can   under mild, aqueous conditions. These capsules were shown to be viable for over 100 days, secreting insulin at a level of 250 - 400 U per islet per day. One hamster remained normoglycemic for 53 days when implanted with 100 such islets.

5.3.3.4. Encapsulation in Hollow Fibers

A alternative method of encapsulating and isolating islets is to entrap them in hollow fiber membranes (figure 5.3.7.). This has been the subject of intensive research for a number of years. One group used the fibers to exclude interleukin 1, which has a molecular weight close to insulin and has been implicated in the destruction of islets. They used a hollow fiber with a cutoff of 50,000 D made form polyetheretherketone.
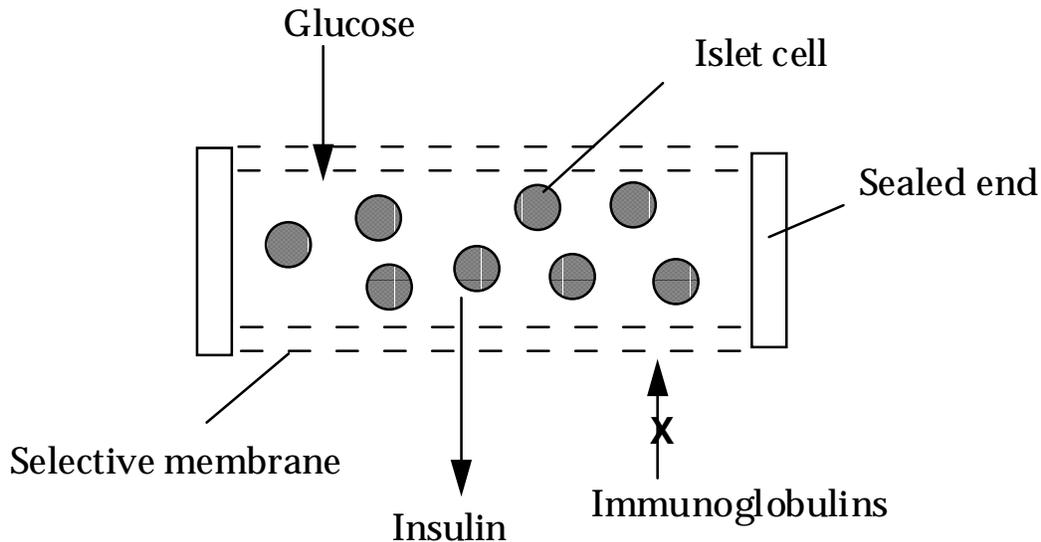
Figure 5.3.7   Schematic of a hollow fiber containing islet cells

5.3.4. Implantable Matrix systems

5.3.4.1. Implanted slabs

One of the early attempts to improve diabetes management with a controlled release device studied the use of an implantable polymeric system which could release a basal dose of insulin over an extended period of time. Langer's group at MIT used the polymer Ethylene vinyl acetate, and devised a method of casting the polymer in different shapes using methylene chloride as a solvent, and suspending the crystalline insulin in the polymer solution before casting it at -80°C, followed by extensive drying. A schematic of a slab is shown in figure 5.3.8. Insulin released is initiated when solvent, in this case body fluid, penetrates the matrix by dissolving the insulin that is exposed at the matrix surface. With high loadings of insulin the crystals are in such tight packing that each one is usually in contact with at least one other crystal further in the matrix. As a result solvent slowly penetrates the entire matrix, and dissolved insulin diffuses out through the fluid filled, interconnecting channels that are formed. Recall in our modeling section we considered modeling of systems which consist of drug dispersed in a matrix. There is a t1/2 dependence for release for the first 60% of release.

Interconnecting channel

EVA matrix

Crystals of insulin
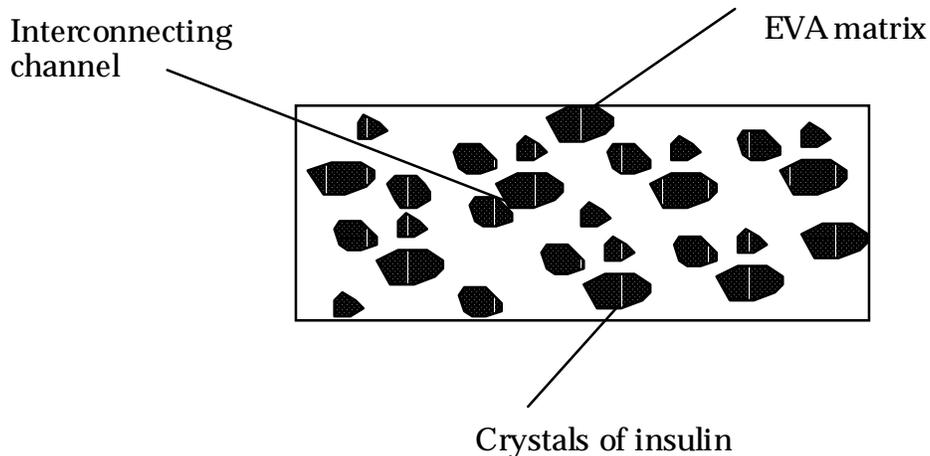
Figure 5.3.8. Schematic of drug crystals dispersed in an EVA matrix

When these matrices were implanted subcutaneously in diabetic rats, the blood glucose levels returned to normal for 8 days, until the implant was removed. Glucose levels then shot up and remained high until the EVA disc was re-implanted at day 14, as shown in figure 5.3.9.
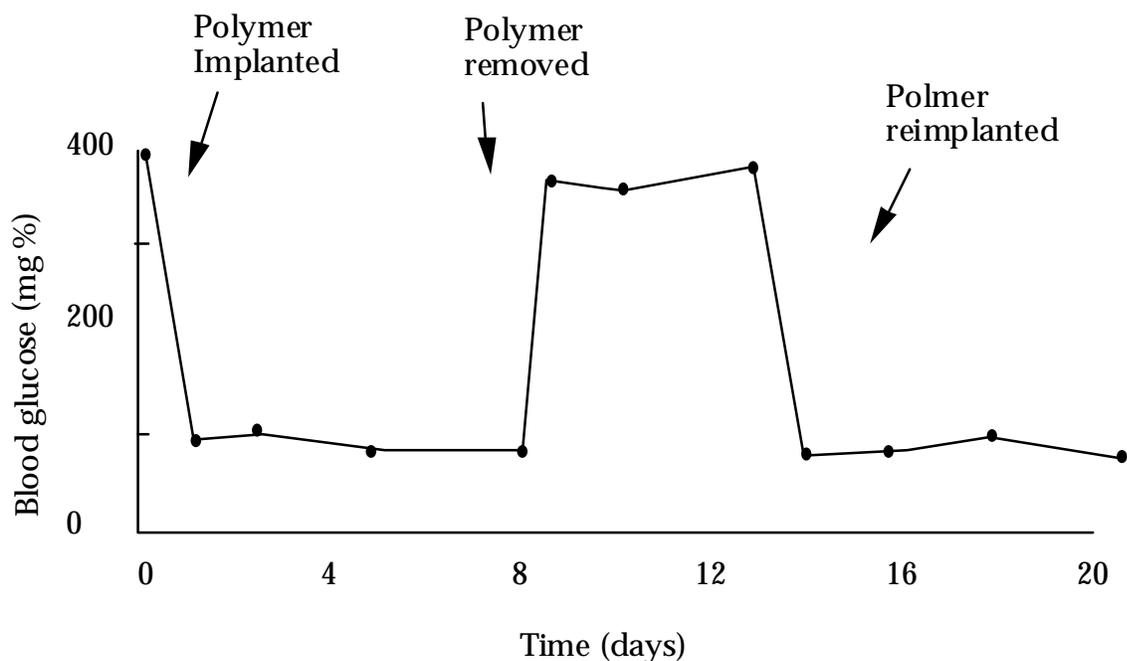

Figure 5.3.9.   Blood glucose levels in diabetic rats in response to periodic implantation and removal of insulin-releasing EVA discs. After Langer et. al.

5.3.4.2. External Control
Attempts have been made to allow for external control of insulin release from these matrixes, to supplement the basal level of insulin when required such as at meal times. Langer approached this by casting magnetic rings into an EVA

device. When an alternating magnetic field was imposed over the matrix, insulin release was enhanced. They report an experiment in which a hemisphere of insulin-containing EVA with an embedded magnetic bead was implanted subcutaneously in diabetic rats. The blood glucose levels after triggering were significantly lowered, as can be seen in figure 5.3.10. The hemispherical shape is unique in that it allows for zero order release at the basal level.
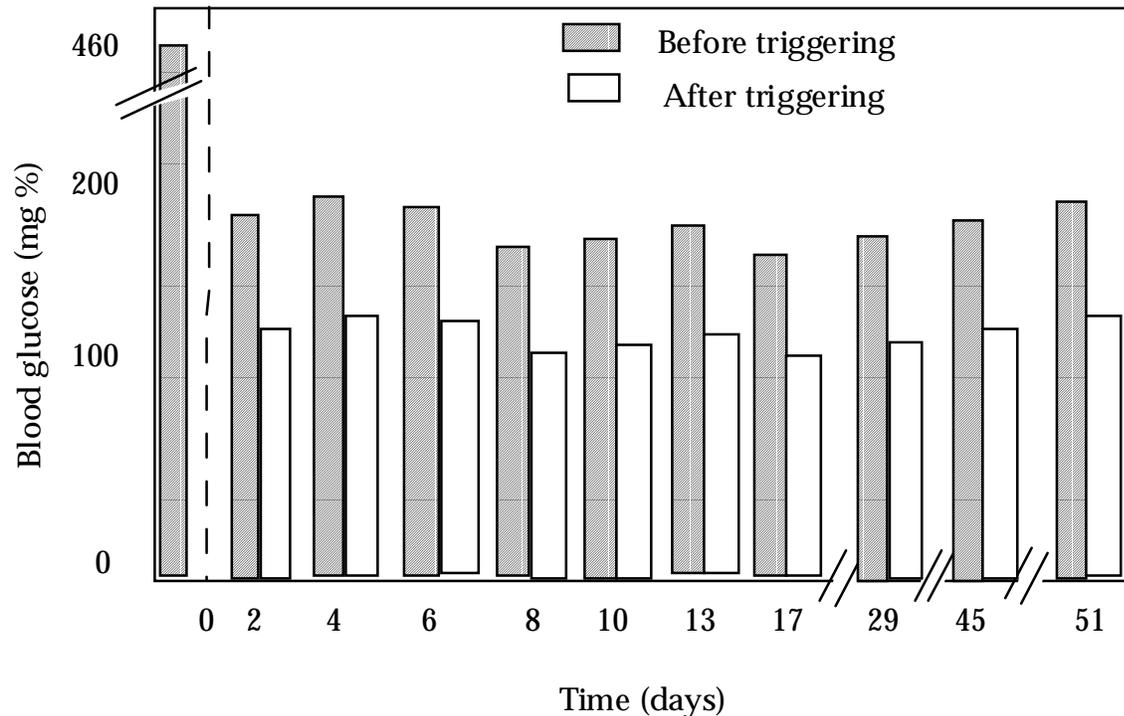
Figure 5.3.10.   Triggering experiment for rats having an EVA-insulin implant containing a

          magnetic bead.

You may want to see if you can do the mathematical modeling of release from this system. The hemisphere is coated with an impermeable layer of EVA, and a small hole is drilled in the upper flat surface by laser. As drug is released with time the surface area that is exposed to the advancing solvent front increases at the same rate as the distance that the drug has to travel increases. With a very small hole in the device, the result is a zero order release, unlike the first order release usually encountered with matrix devices. The device has the appearance of half a cantaloupe melon as shown in figure 5.3.11.
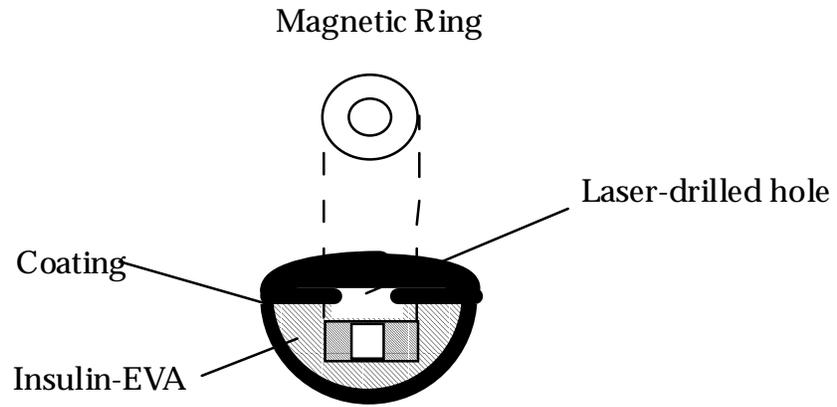
Figure 5.3.11. Hemispherical EVA-insulin device with embedded magnet for triggering
experiment.

5.3.5. Competitive Desorption

Researchers in Utah approached the problem in a different way. They wanted to develop a device that would be responsive to blood glucose levels. For this they synthesized a special form insulin which was glycosylated, that is glucose groups were chemically bonded to it. This glycosylated insulin was found to bind reversibly to a substance called concanavalin A. Glucose also binds to Con A, in fact it binds more strongly, so if it is present under conditions where there are limited binding sites, it will competitively replace the glycosylated insulin. Kim et al devised an implantable chamber (figure 5.3.12.) in which was placed Con A that had been covalently bound to a small sepharose bead. This prevented the Con A from diffusing out of the chamber.
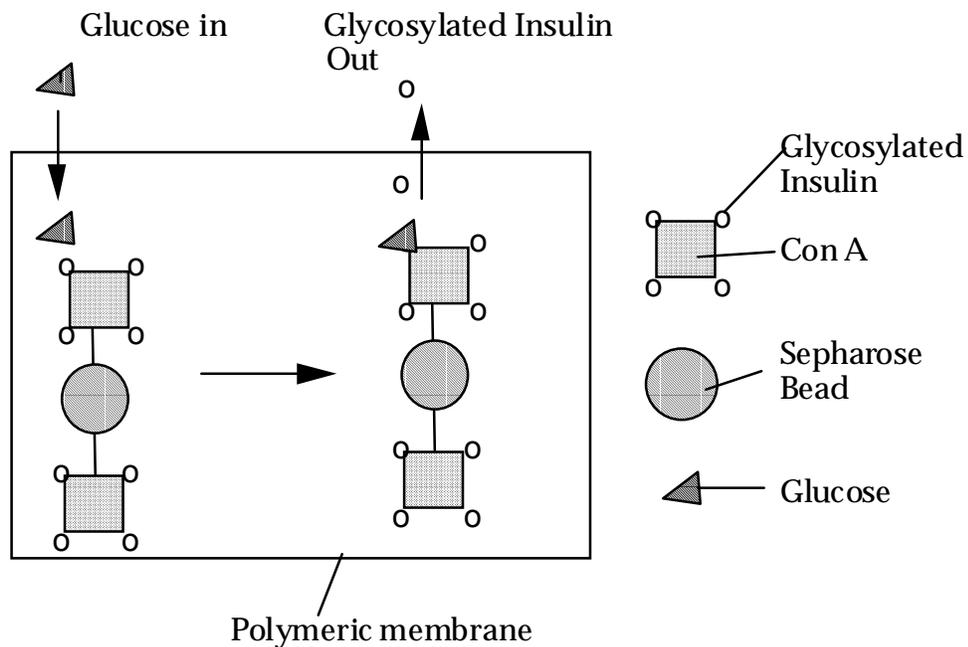
Figure 5.3.12. Schematic of chamber designed for competitive desorption of glycosylated

insulin (after Kim).

Before being placed into the chamber, the immobilized Con A was loaded with glycosylated insulin. When the chamber containing the con A-insulin beads was tested in vitro with a glucose challenge, insulin release was measured for over 70 hours. A glucose pulse was given at roughly 20 hour intervals, as shown in figure 5.3.13..

Insulin (µg/hr)

Time (hrs)

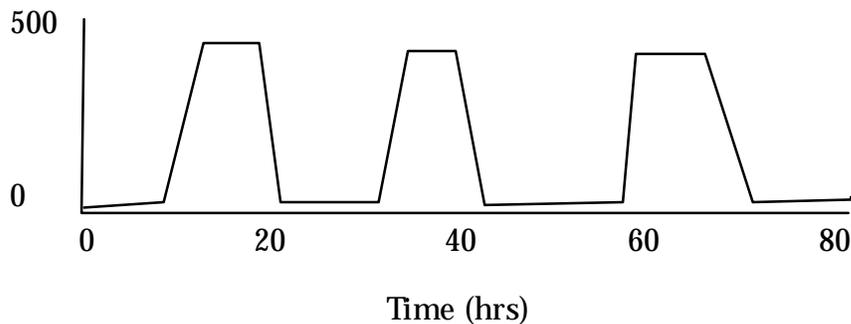Glucose (mg/dl)

Time (hrs)
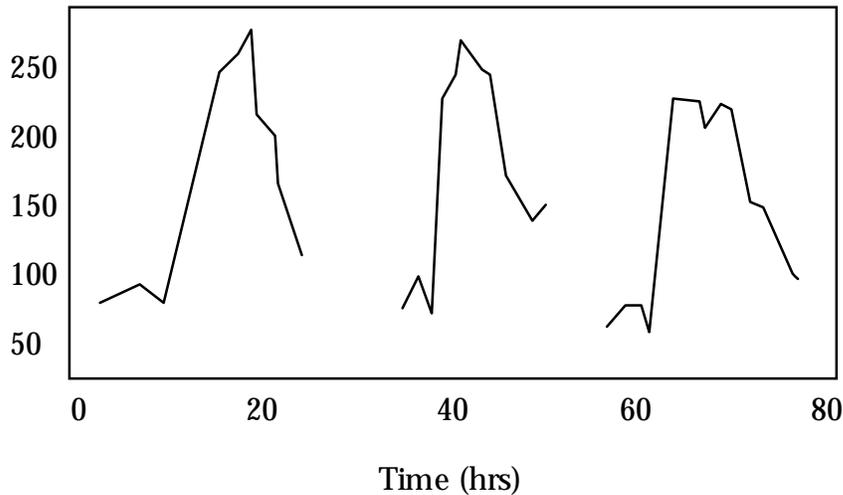
Figure 5.3.13.   Response of competitive desorption device to a glucose challenge. (After
   Kim)

5.3.6.   Glucose sensitive membranes

Another approach to developing a glucose-sensitive system is to synthesize a membrane which will keep insulin in a reservoir when there is no glucose present and allow insulin to diffuse across when it is present. One example of this is a polymer that swells in the presence of glucose. Such a polymer has been developed in the laboratory of Horbett and coworkers at the University of Washington. He embedded the enzyme glucose oxides (GOD) in a polymer containing tertiary amino groups (-NR $_2$). When glucose diffuses into the polymer it reacts at the GOD enzyme to form gluconic acid. This lowers the pH and the resulting proton protonates the tertiary amino group. These groups are close together on the polymer network and the positive charge that they now carry causes repulsion between them. This repulsion results in a swelling of the polymer.   If an insulin solution is held in a reservoir behind the hydrogel membrane, it will be able to diffuse out through the swollen polymer.   When glucose is no longer present the gel will re-equilibrate with the ions of the body and return to physiological pH of 7.4. At this point the amino groups will

deprotonate and the membrane will return to its close packed structure. Insulin will not be able to diffuse across. The device is shown in figure 5.3.14.
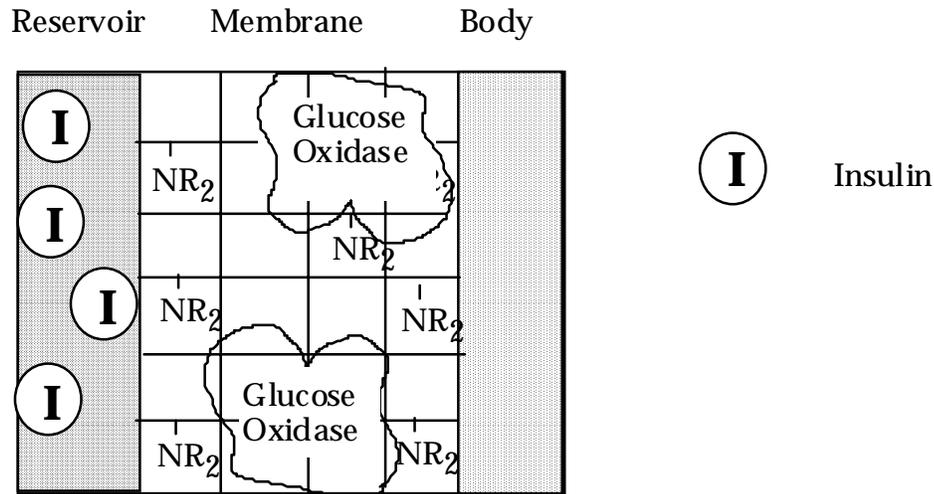

Figure 5.3.14. pH sensitive membrane (after Horbett)

5.3.7. pH Dependent solubility

An approach which has features of both of the previous two systems is one that attempts to produce a glucose sensitive device by linking the drop in pH caused by the reaction of glucose on glucose oxidase, with the fact that protein solubility if highly pH dependent. Here a modified insulin is again used, this time trilysyl insulin, that is insulin with three lysine groups attached to it. Lysine is a basic amino acid and it renders the insulin high at around its new isoelectric point between pH 7 and 8. When insulin is at physiological pH, that is pH 7.4, it is insoluble. When the pH is lowered when glucose reacts at the catalyst glucose oxidase, to form gluconic acid, then it becomes soluble and can diffuse out of the devise into the body. The solubility curves with respect to pH are shown in figure 5.3.15.
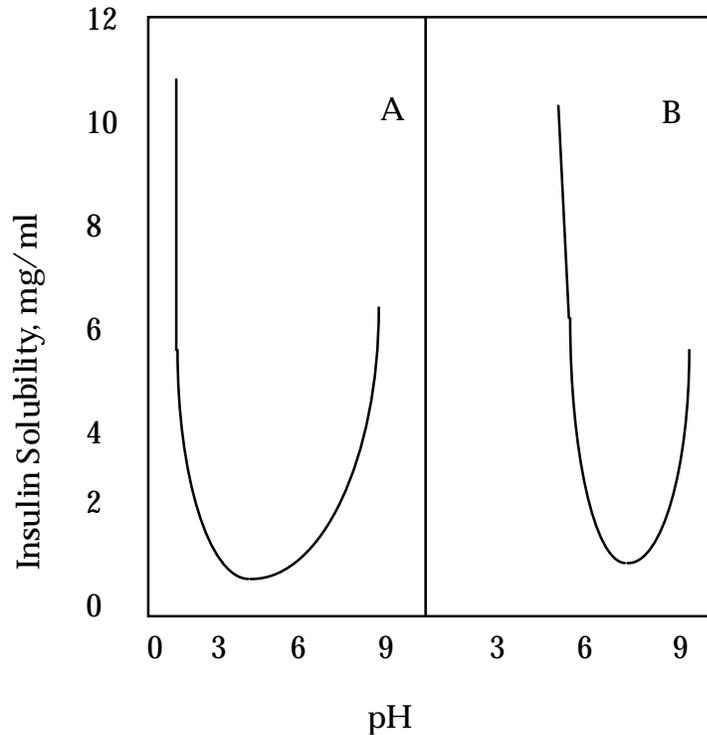
Figure 5.3.15. Solubility dependence of (A) Regular insulin. and (B) Trilysyl. insulin

(after Godshial)

5.3.8. Polymer Erosion on pH Change

Heller and coworkers investigated a glucose sensitive device that was composed of insulin dispersed in a polymer that was eroded by low pH. A hydrogel which contains glucose oxidase enzyme surrounds the insulin-containing polymer. When glucose interacts with the glucose oxidase, the resulting low pH erodes the polymer, and insulin is released. A schematic is given in figure 5.3.16.
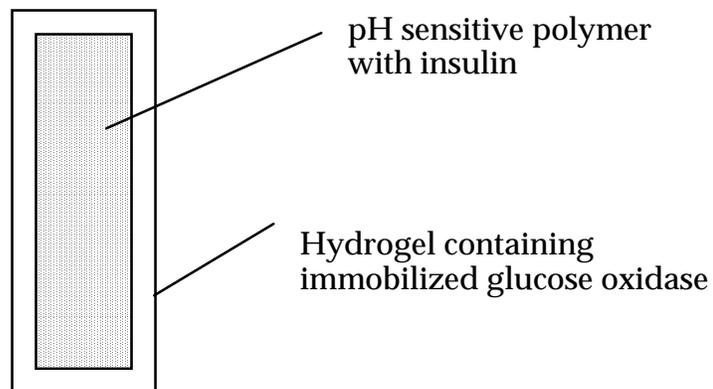
Figure 5.3.16. Schematic of pH sensitive polymeric device (after Heller)

This device was tested in vitro, and dramatic release of insulin was observed at pH 5.5 and below, and at pH 5.6 no release was observed (figure 5.3.17).
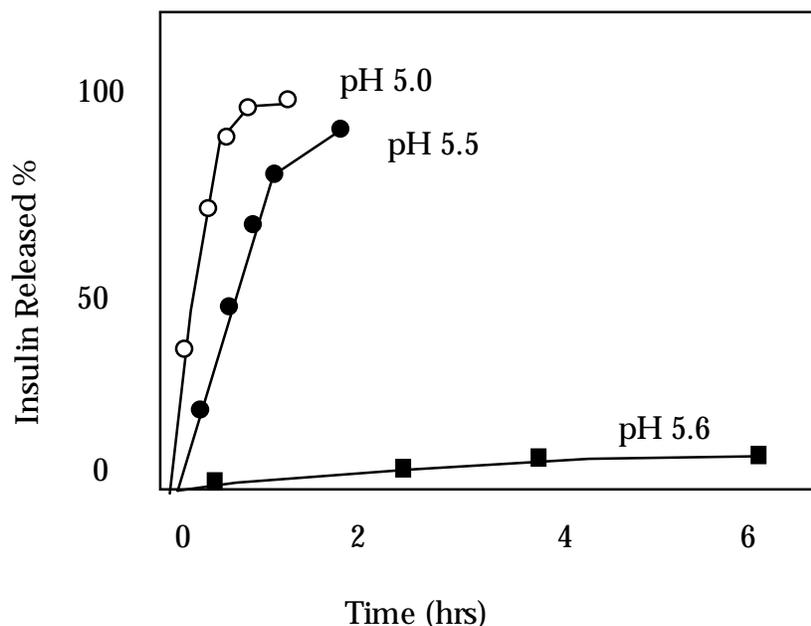
Figure 5.3. 17.   Effect of pH on rate of insulin release from a cross-linked polymer
prepared   from   3,9-bis(ethyenediene   2.4.8.10-tetraethanolamine). (After
Heller)

5.4.   Gene Therapy

The very latest attempts to manage diabetes have focused on gene therapy. Every cell in the body carries the insulin gene, but it is only switched on in the β-cells in the islets in the pancreas.   In order to produce human insulin in fermentation, genetic engineers have inserted an active insulin gene into bacteria so that they can produce it in vast quantities. In theory in should also be possible to implant a cell line in to the body which also produces insulin. It would have to be encapsulated to protect it from the immune system of the body, and it should also function in a similar way to islet cells, that is only produce insulin when blood sugar levels are high.   Several interesting cells lines are being investigated. Some have been engineered and others have been found naturally in tumors.   Table 5.4.1. lists some of the most actively investigated lines, with some of their characteristics. AtT-20 has been encapsulated in alginate. There is great hope that one day in the not too distant future, diabetes will be managed not by daily insulin injections, but with a single implantation of a set of cells that will secrete insulin in the same way as normal islet cells, and that diabetics will not longer be troubled by the many side effects that come when blood glucose
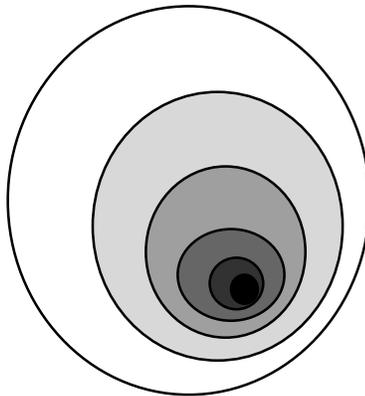
levels fluctuate widely about the norm. If a cure does come, bioengineers will have played a large and vital role in its realization.

Table 5.4.1. Insulin secreting cell lines

Insulinoma cell line	Characteristics
RIN 1046-38	Glut-2 & Glucokinase expressed, but gradually lose that & the ability for glucose stimulated insulin release (GSIR)
INS-1	Retains high GSIR. Cell growth slow
MIN-7	Expresses GLUT-1 not GLUT-2 as major glucose transporter isoform
MIN-6	Expresses GLUT-2, and normal GSIR
βTC-7	Expresses GLUT-2 and normal GSIR, and Glucokinase: Hexokinase activity ratio similar to normal islets
AtT-20	Derived from non-islet cells. Insulin secretion stimulated by cAMP. Secretes mature insulin.

BIOSENSORS

Engineering Biotechnology Gateway Project

Young H. Lee
Drexel University

Table of Contents

Chapter 4. Biosensor Fundamentals

Chapter 1. Overview of Biosensors

1. Overview of Biosensor

1.1. What Is a Biosensor?

Biosensor = bioreceptor + transducer. A biosensor consists of two components: a bioreceptor and a trans-ducer. The bioreceptor is a biomolecule that recognizes the target analyte whereas the transducer converts the recognition event into a measurable signal. The uniqueness of a biosensor is that the two components are integrated into one single sensor (Fig. 1.1). This combination enables one to measure the target analyte without using reagents. For example, the glucose concentration in a blood sample can be measured directly by a biosensor (which is made specifically for glucose measurement) by simply dipping the sensor in the sample. This is in contrast to the conventional assay in which many steps are used and each step may require a reagent to treat the sample. The simplicity and the speed of measurement is the main advantages of a biosensor.

Enzyme is a Bioreceptor When we eat food such as a hamburger and French fries, they are broken down into small molecules in our body via many reaction steps (these breakdown reactions are called *catabolism*). These small molecules are then used to make building blocks of our body such as proteins (these synthesis reactions are called *analbolism*). Each of these catabolism and anabolism reactions (the combination is called *metabolism*) are catalyzed by a specific enzyme. Therefore, **an enzyme is capable of recognizing a specific target molecule** (Fig. 1.2). This biorecognition capability of the enzyme is used in biosensors. Other biorecognizing molecules (= bioreceptors) include antibodies, nucleic acids, and receptors.

Immobilization of Bioreceptor One major requirement for a biosensor is that the bioreceptor molecule has to be immobilized in the vicinity of the transducer. The immobilization is done either by physical entrapment or chemical attachment. Note that only minute quantities of bioreceptor molecules are needed, and they are used repeatedly for measurements.

Transducer A transducer should be capable of converting the biorecognition event into a measurable signal (Fig. 1.3). Typically, this is done by measuring the change that occur in the bioreceptor reaction. For example, the enzyme glucose oxidase (used as a bioreceptor in a glucose biosensor) catalyzes the following reaction:
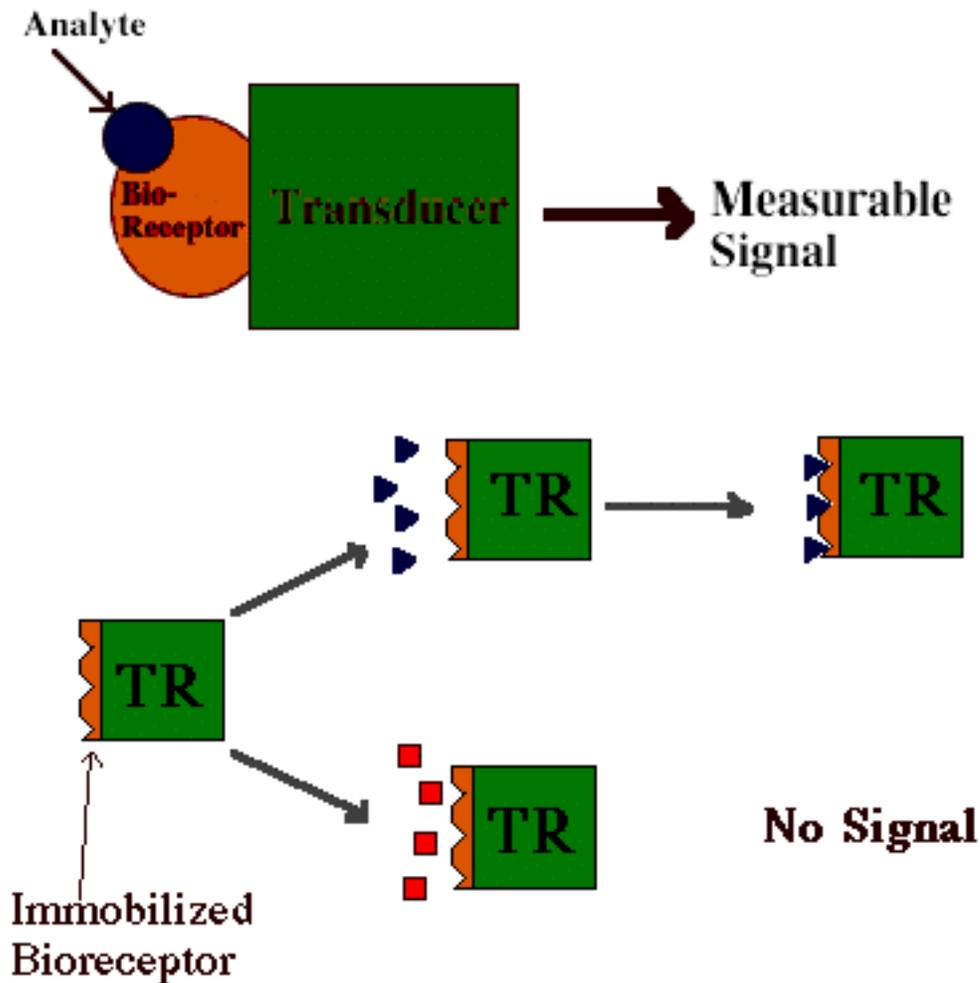
$$\text{Glucose} + O_2 \xrightarrow{\text{Glucose Oxidase}} \text{Gluconic acid} + H_2O_2$$
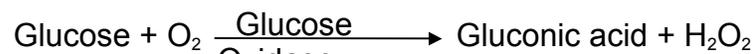
Fig. 1.2. Specificity of biosensor (TR: transducer).

To measure the glucose concentration, three different transducers can be used:

     1. An oxygen sensor that measures oxygen concentration
     2.. A pH sensor that measures the acid (gluconic acid) production
     3. A peroxide sensor that measures H_2O_2 concentration.

Note that an oxygen sensor is a transducer that converts oxygen concentration into electrical current. A pH sensor is a transducer that converts pH change into voltage change. Similarly, a peroxidase sensor is a transducer that converts peroxidase concentration into an electrical current.

Considerations in Biosensor Development Once a target analyte has been identified, the major tasks in developing a biosensor involves:

       1. Selection of a suitable bioreceptor molecule
       2. Selection of a suitable immobilization method

3. Selection of a suitable transducer
4. Designing of biosensor considering measurement range, linearity, and minimization of interference
5. Packaging of biosensor

The item 1 requires knowledge in biochemistry and biology, the item 2 requires knowledge in chemistry, the item 3 requires knowledge in electrochemistry and physics, and the item 4 requires knowledge in kinetics and mass transfer. Once a biosensor has been designed, it has to be put into a package for convenience manufacturing and use. The current trend is miniaturization and mass production. Modern IC (integrated circuit) fabrication technology and micromachining technology are used increasingly in fabricating biosensors. Therefore, interdisciplinary cooperation is essential for a successful development of a biosensor.

Requirements for Sensors   To be commercially successful, a biosensor has to meet the general requirements of commercial sensors (Table 1.2). These are:

1. Relevance of output signal to measurement environment
2. Accuracy and repeatability
3. Sensitivity and resolution
4. Dynamic range
5. Speed of response
6. Insensitivity to temperature (or temperature compensation)
7. Insensitive to electrical and other environmental interference
8. Amenable to testing and calibration
9. Reliability and self-checking capability
10. Physical robustness
11. Service requirements
12. Capital cost
13. Running costs and life
14. Acceptability by user
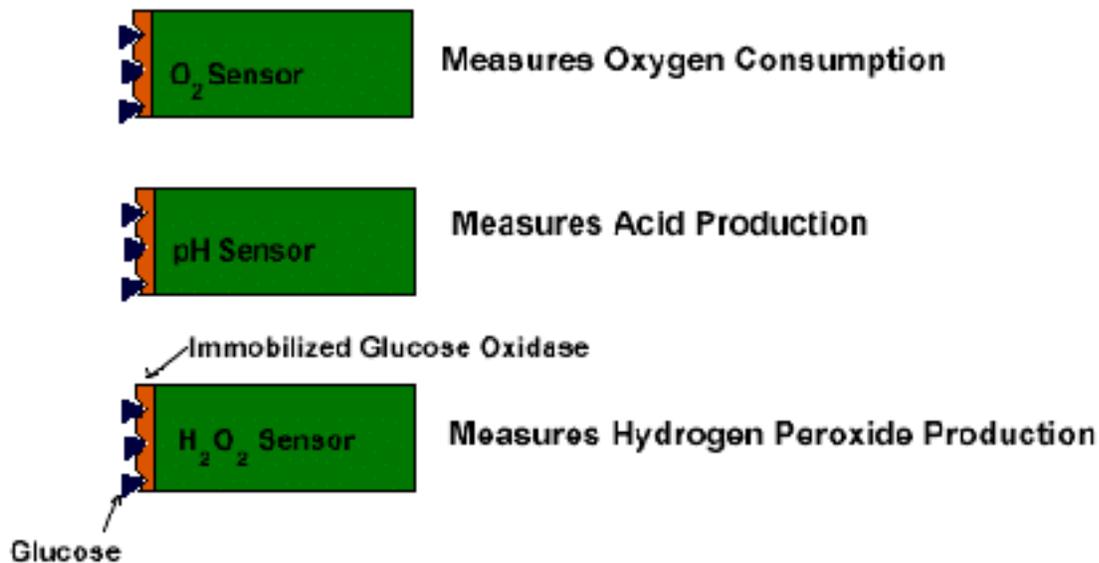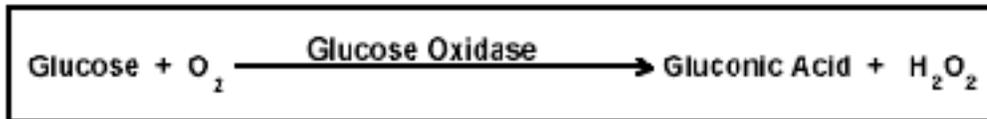15. Product safety—sample host system must not be contaminated by sensor

$$Glucose + O_2 \xrightarrow{\text{Glucose Oxidase}} Gluconic\ Acid + H_2O_2$$

O₂ Sensor — Measures Oxygen Consumption

pH Sensor — Measures Acid Production

Immobilized Glucose Oxidase

H₂O₂ Sensor — Measures Hydrogen Peroxide Production

Glucose

Fig. 1.3. Three possible transducers for glucose measurement.

Table 1.1. Considerations for biosensor development
_____

- **Selection of a suitable bioreceptor molecule**
- **Selection of a suitable immobilization method**
- **Selection of a suitable transducer**
- **Designing of biosensor considering measurement range, linearity, and minimization of interference**
- **Packaging of biosensor**

_____

1.2. Need for Biosensor

Diagnostic Market.   Diagnostics represent a very large and well established market that is continually expanding. Particularly in the current climate of prevention rather than remedy, the need for detection at increasingly lower limits

is increasing in many diverse areas. Estimates of the market size and future projections tend to be difficult and inaccurate.

Clinical Testing   However, undoubtedly *clinical testing* one of the biggest diagnostic markets. A study of the European market suggests a clinical testing products market in excess of 4000 million US$ in the 1990s (Biomedical Business International). In the US, the current biosensor market is already reported to be 12 million dollars, and future prospects vary from 100 to 10,000 million dollars by the turn of the century. This compares with a world market in 1985 of 1.5 billion dollars with an estimated growth rate of 9.5%, achieving a world market of 2 billion in 1990 and then expanding upwards and outwards.

Other Markets   Among the market shares, nearly 50% belongs to the medical arena (Technical Insights Inc.) with veterinary and agricultural applications amounting to a figure of half the size (Table I. 2). Although the actual figures may differ between surveys, the relative proportions for different applications seem to be in agreement for the 1990s. Figure 1.5 (Business Communications Co.) plots the US market projection into the next century and identifies a changing emphasis in applications with the agricultural and environmental component becoming increasingly significant.

Table 1.2. Fifteen characteristics required in a commercial sensor
_____
- **Relevance of output signal to measurement environment**
- **Accuracy and repeatability**
- **Sensitivity and resolution**
- **Dynamic range**
- **Speed of response**
- **Insensitivity to temperature (or temperature compensation)**
- **Insensitive to electrical and other environmental interference**
- **Amenable to testing and calibration**
- **Reliability and self-checking capability**
- **Physical robustness**
- **Service requirements**
- **Capital cost**
- **Running costs and life**
- **Acceptability by user**
- **Product safety—sample host system must not be contaminated by sensor**

_____

Table 1.3. Estimated markets in the 1990s

Market	Value (million dollars)
Medical and surgical	220
Veterinary and agricultural	105

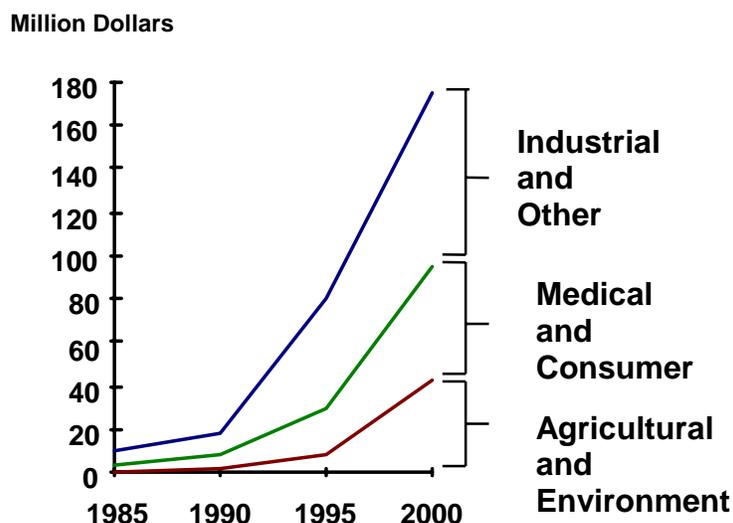|	
Environmental and monitoring and safety	**67**
Industrial process monitoring	**59**

Million Dollars

_____

Fig. 1.5. USA market projection for biosensors.

1.3. Applications of Biosensor

1.3.1. Health Care

Measurement of Metabolites The initial impetus for advancing sensor technology came from health care area, where it is now generally recognized that measurements of blood gases, ions and metabolites are often essential and allow a better estimation of the metabolic state of a patient. In intensive care units for example, patients frequently show rapid variations in biochemical levels that require an urgent remedial action. Also, in less severe patient handling, more successful treatment can be achieved by obtaining _instant_ assays. At present, the list of the most commonly required _instant_ analyses is not extensive. In practice, these assays are performed by analytical laboratories, where discrete samples are analyzed, frequently using the more traditional analytical techniques.

Market Potential.   There is an increasing demand for inexpensive and reliable sensors to allow not only routine monitoring in the central or satellite laboratory, but also analysis with greater patient contact, such as in the hospital ward, emergency rooms, and operating rooms. Ultimately, patients themselves should be able to use biosensors in the monitoring and control of some treatable condition, such as diabetes. It is probably true to say that **the major biosensor market may be found where an immediate assay is required.** If the cost of laboratory maintenance are counted with the direct analytical costs, then low-

cost biosensor devices can be desirable in the whole spectrum of analytical applications from hospital to home.

Diabetes.   The 'classic' and most widely explored example of closed-loop drug control is probably to be found in the development of an artificial pancreas. Diabetic patients have a relative or absolute lack of insulin, a polypeptide hormone produced by the beta-cells of the pancreas, which is essential to the metabolism of a number of carbon sources. This deficiency causes various metabolic abnormalities, including higher than normal blood glucose levels, and where patients have suffered a complete destruction of the insulin-secreting islets of Langerhans, then insulin must be supplied. This has usually been achieved by subcutaneous injection, but fine control is difficult and hyperglycaemia cannot be totally avoided, or even hypoglycaemia sometimes induced, causing impaired consciousness and the serious long-term complications to tissue associated with this intermittent low glucose condition.

Insulin Therapy.   Better methods for the treatment of insulin-dependent diabetes have been sought and infusion systems for continuous insulin delivery have been developed. However, regardless of the method of insulin therapy, its induction must be made in response to information on the current blood glucose levels in the patient. Three schemes are possible (Fig. 1.6), the first two dependent on discrete manual glucose measurement and the third a 'closed-loop' system, where insulin delivery is controlled by the output of a glucose sensor which is integrated with the insulin infuser. In the former case, glucose has been estimated on 'finger-prick' blood samples with a colorimetric test strip or more recently with an amperometric 'pen'-size biosensor device by the patient themselves. Obviously these diagnostic kits must be easily portable, very simple to use and require the minimum of expert interpretation. However, even with the ability to monitor current glucose levels, intensive conventional insulin therapy requires multiple daily injections and is unable to anticipate future states between each application, where diet and exercise may require modification of the insulin dose. For example, it was shown that administration of glucose by subcutaneous injection, 60 min before a meal provides the best glucose/insulin management.

Artificial Pancreas.   The introduction of a closed-loop system, where integrated glucose measurements provide feedback control on a pre-programmed insulin administration based on habitual requirement, would therefore relieve the patient of frequent assay requirements and perhaps more desirably frequent injections. Ultimately, the closed-loop system becomes an artificial pancreas, where the glycaemic control is achieved through an **implantable glucose sensor**. Obviously, the requirements for this sensor are very different to those for the discrete measurement kits. As summarized in Table 1.4, the prolonged life-time and biocompatibility represent the major requirements.

1.3.2. Industrial Process Control

Bioreactor Control. Real-time monitoring of carbon sources, dissolved gases, etc., in fermentation processes (Fig. 1.7a) could lead to optimization of the procedure giving increased yields at decreased materials cost. While real-time monitoring with feedback control involving automated systems does exist, currently only a few common variables are measured on-line (e.g. pH, temperature, CO_2, O_2)) which are often only indirectly related with the process under control.

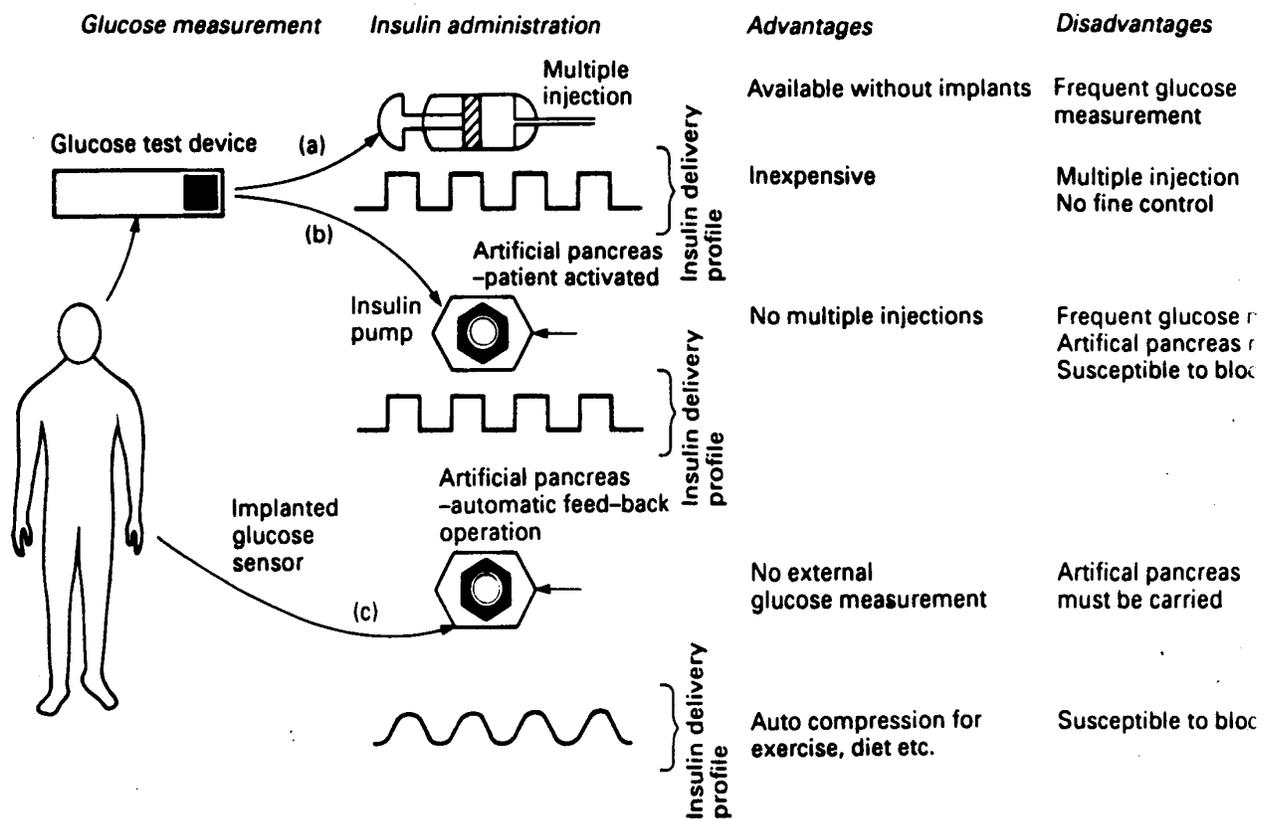| Glucose measurement	Insulin administration	Advantages	Disadvantages
Glucose test device (a)	Multiple injection	Available without implants	Frequent glucose measurement
	Insulin delivery profile	Inexpensive	Multiple injection No fine control
(b)	Artificial pancreas –patient activated		
	Insulin pump	No multiple injections	Frequent glucose r Artifical pancreas r Susceptible to bloc
	Insulin delivery profile		
	Artificial pancreas –automatic feed–back operation		
Implanted glucose sensor (c)		No external glucose measurement	Artifical pancreas must be carried
	Insulin delivery profile	Auto compression for exercise, diet etc.	Susceptible to bloc

Fig. 1.6 Schemes for insulin therapy.

Table 1.4. Seven requirements for an implantable glucose sensor.

● **Linear in 0 - 20 mM range with 1 mM resolution**

- **Specific for glucose; not affected by changes in metabolite concentrations and ambient conditions**
- **Biocompatible**
- **Small---causes minimal tissue damage during insertion and there is better patient**
  acceptability for a small device
- **External calibration and < 10% drift in 24h**
- **Response time < 10 min**
- **Prolonged lifetime—at least several days, preferably weeks in use**

_____

–

Three different methods of controlling a bioreactor are:

1. Off-line distant: central laboratory coarse control with significant time lapse
2. Off-line local: fine control with short time lapse
3. On-line: real-time monitoring and control

On-Line Control.   Method 3 is most desirable, which allows the process to follow an ideal pre-programmed fermentation profile to give maximum output. However, many problems exist with on-line measurements including in situ sterilization, sensor life-time, sensor fouling, etc. Some of the problems can be overcome if the sensor is situated so that the sample is run to waste, but this causes a volume loss, which can be particularly critical with small volume fermentations.

Off-Line Control.   Although Method 3 may be the ultimate aim, considerable advantage can be gained in moving from Method 1 to Method 2 giving a rapid analysis and thus enabling finer control of the fermentation. The demands of the sensor are perhaps not as stringent in Method 2 as in Method 3.

Benefits of Control.   The benefits which are achievable with process-control technology are considerable:

• Improved product quality; reduction in rejection rate following manufacture
• Increased product yield; process tuned in real time to maintain optimum conditions throughout and not just for limited periods
• Increased tolerance in quality variation of some raw materials. These variations can be compensated in the process-control management
• Reduced reliance on human 'seventh sense' to control process
• Improved plant performance—processing rate and line speed automated, so no unnecessary dead-time allocated to plant
• Optimized energy efficiency

The use of biosensors in industrial process in general could facilitate plant automation, cut analysis costs and improve quality control of the product.

1.3.3. Military Applications

Dip Stick Test   The requirement for rapid analysis can also be anticipated in military applications. The US army, for example, have looked at dipstick tests
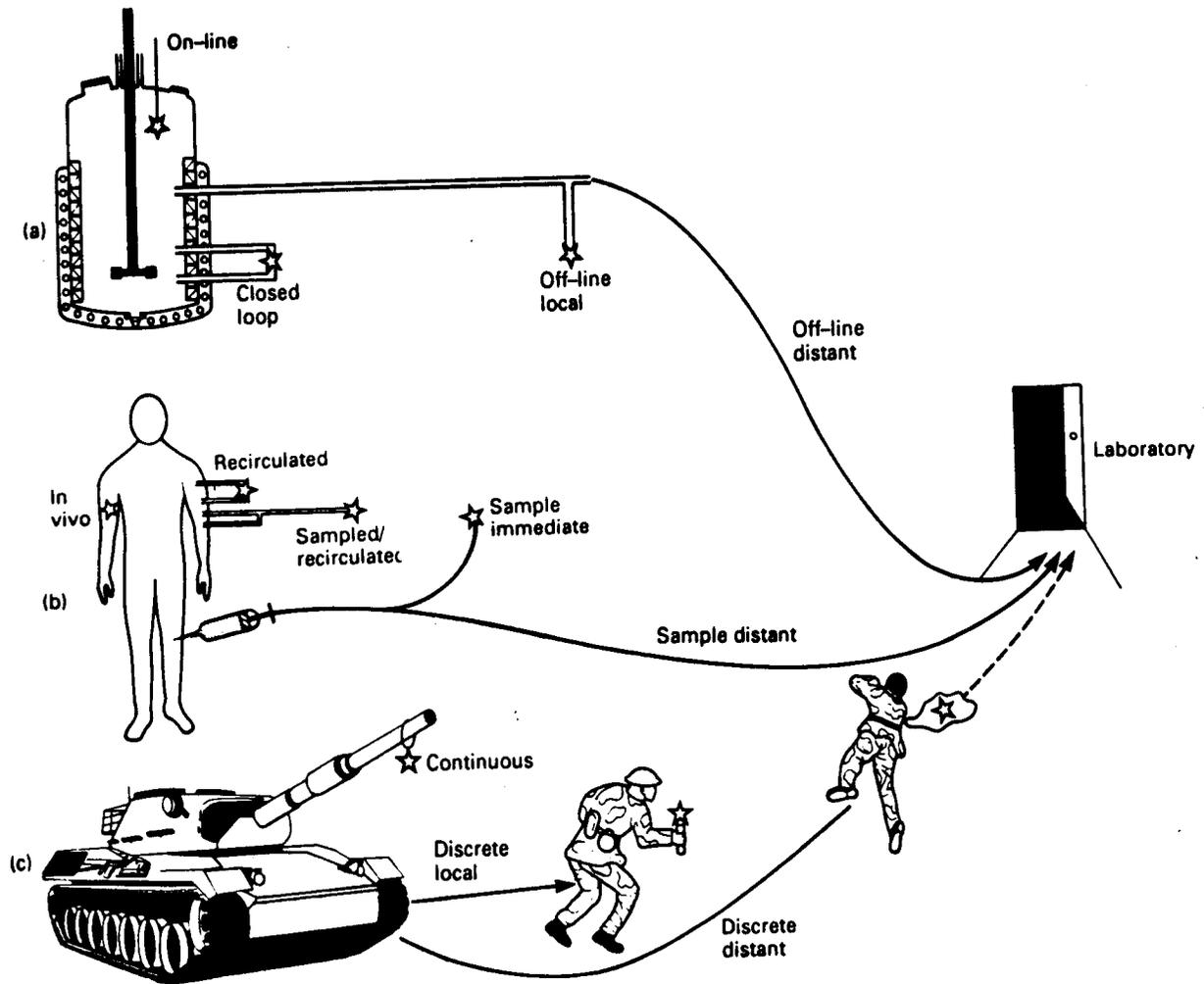
Fig. 1.7. Comparison of sensing modes: (a) bioreactor; (b) clinical applications; (c) military or environmental monitoring.

Table 1.4 Summary of potential applications for biosensors

• **Clinical diagnosis and biomedicine**
• **Farm, garden and veterinary analysis**
• **Process control: fermentation control and analysis food and drink production and analysis**
• **Microbiology: bacterial and viral analysis**
• **Pharmaceutical and drug analysis**
• **Industrial effluent control**
• **Pollution control and monitoring o Mining, industrial and toxic gases**
• **Military applications**

based on monoclonal antibodies. While these dipsticks are stable and highly specific (to Q-fever, nerve agents, yellow rain fungus, soman, etc.) they are

frequently two-step analyses taking up to 20 min to run. Such a time lapse is not always suited to battlefield diagnostics; the resulting consequences are suggested in Fig. 1.7(c).

A particularly promising approach to this unknown hazard detection seems to be via acetylcholine receptor systems. It has been calculated that with this biorecognition system, a matrix of 13-20 proteins are required to give 95% certainty of all toxin detection.

1.3.4. Environmental Monitoring

Air and Water Monitoring.   Another assay situation which may involve a considerable degree of the unknown is that of environmental monitoring. The primary measurement media here will be water or air, but the variety of target analytes is vast. At sites of potential pollution, such as in factory effluent, it would be desirable to install on-line real-time monitoring and alarm, targeted at specific analytes, but in many cases random or discrete monitoring of both target species or general hazardous compounds would be sufficient. The possible analytes include biological oxygen demand (BOD) which provides a good indication of pollution, atmospheric acidity, and river water pH, detergent, herbicides, and fertilizers (organophosphates, nitrates, etc.). The survey of market potential has identified the increasing significance of this area and this is now substantiated by a strong interest from industry. The potential applications of biosensors are summarized in Table 1.4.

Tuning to Application.   The potential for biosensor technology is enormous and is likely to revolutionize analysis and control of biological systems. It is possible therefore to identify very different analytical requirements and biosensor developments must be viewed under this constraint. It is often tempting to expect a single sensor targeted at a particular analyte, to be equally applicable to on-line closed-loop operation in a fermenter and pin-prick blood samples. In practice, however, the parallel development of several types of sensor, frequently employing very different measurement parameters is a more realistic.

1.4. Origin of Biosensor

Enzyme Electrode.   The biosensor was first described by Clark and Lyons in 1962, when the term *enzyme-electrode* was adopted. In this first enzyme electrode, an oxido-reductase enzyme, glucose oxidase, was held next to a platinum electrode in a membrane sandwich (Fig. 1.8) . The platinum anode polarized at + 0.6 V responded to the peroxide produced by the enzyme reaction with substrate. The primary target substrate for this system was glucose:

$$\text{Glucose} + O_2 \xrightarrow{\text{Glucose Oxidase}} \text{Gluconic Acid} + H_2O_2$$
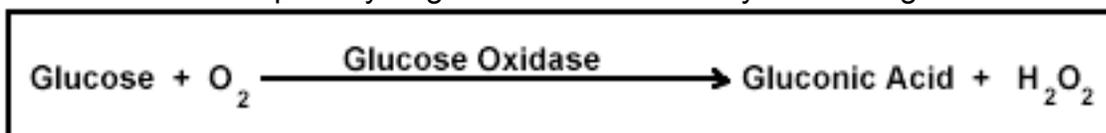
and led to the development of the first glucose analyser for the measurement of glucose in whole blood. This Yellow Springs Instrument (Model 23 YSI) appeared on the market in 1974, and the same technique as employed here has been applied to many other oxygen mediated oxido-reductase enzyme systems.

Use of Membrane for Selectivity.   A key development in the YSI sensor was   the employment of membrane technology in order to eliminate interference by other electro-active substances. Polarized at +0.6V, the major interference to the peroxide measurement is ascorbic acid. Various combinations of membrane-enzyme sandwich have been developed, all satisfying the following criteria:

• the membrane between electrode and enzyme layer should allow the passage of H_2O_2, but prevent the passage of ascorbate or other interferents
•   the   membrane   between   enzyme   layer   and   sample   should   allow substrate/analyte to enter
  the enzyme layer

This was accomplished in the YSI, for example, with an enzyme layer sandwiched between a cellulose acetate membrane and a Nucleopore polycarbonate membrane.

1.5. Bioreceptor Molecules

Enzymes have been the most widely used bioreceptor molecules in biosensor applications. Recently, antibodies and protein receptor molecules are increasingly incorporated in biosensors. The specificity of a biosensor comes from the specificity of the bioreceptor molecule used. An enzyme is a good example. It has a three dimensional structure that fits only a particular substrate (Fig. 1.9a). An enzyme is a protein synthesized in the cell from amino acids according to the codings written in DNA. Enzymes act as catalysts for biochemical reactions occurring in the cell. To maintain high enzyme activity, the temperature and pH of the environment have to be maintained at proper levels.
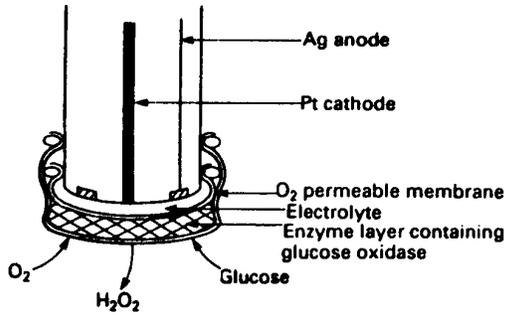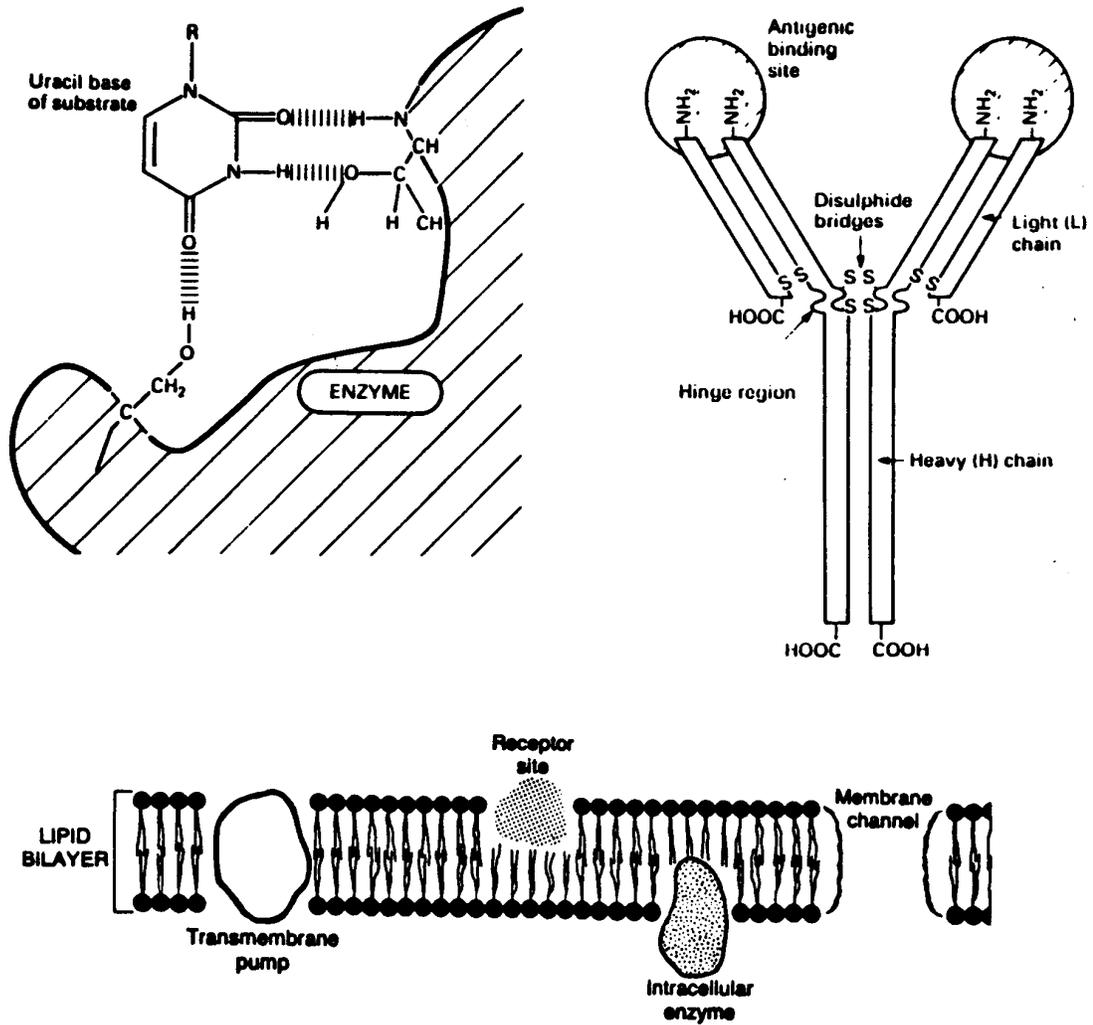
Fig. 1.8. The Clark enzyme electrode.

**Fig. 1.9. Bioreceptor molecules used for biosensor applications:
(a) enzyme; (b) antibody; (c) protein receptor**

Antibody   Antibodies represent one of the major class of protein; they constitute about 20% of the total plasma protein and are collectively called

immunoglobulins (Ig). The simplest antibodies are usually described as Y-shaped molecules with two identical binding sites for antigen. An antigen can be almost any macromolecule that is capable of inducing an immune response. The antibody has a basic structural unit consisting of four polypeptide chains - two light chains and two heavy chains (Fig. 1.9b). The antibody binds reversibly with a specific antigen. Unlike the enzyme proteins, the antibody do not act as catalysts. Their purpose is to bind foreign substances - antigens - so as to remove them from the system.

Receptor Protein.   Receptor proteins have specific affinity for hormones, antibodies, enzymes,and other biologically active compound. These proteins are mostly bound to membrane (Fig. 1.9c). There are hormone receptors, taste receptors, olfactory receptors for smelling, photoreceptors for eyes, etc. Receptor proteins are responsible for opening and closing of membrane channels for transport of specific metabolites.

Other Possibilities.   In principle, any biomolecules and molecular assembly that have the capability of recognizing a target substrate (= the analyte) can be used as a bioreceptor. In fact, membrane slices or whole cells have been used in biosensors. Fig. 1.10 summarizes possible bioreceptors that can be utilized in a biosensor. Note that the bioreceptors require suitable environment for maintaining their structural integrity and biorecognition activity. These requirements are described in Fig. 1.10 along with the type of signal generated as a result of the biorecognition activity. The transducer in a biosensor has to be capable of measuring this signal.

1.6. Transducers Used

Conventional Transducers.   Majority of biosensors existing today use three types of transducers for converting the action of the bioreceptor molecule into measurable signal. These are: (1) amperometry based on H_2O_2 or O_2 measurement; (2) potentiometry based on pH or pIon measurement; (3) photometry utilizing optical fibers (see Fig. 11).
Biorecognition reactions often generates chemical species that can be measured by electrochemical methods. In amperometry (Fig. 1.11a), typically the reaction product is H_2O_2 (or the reactant is O_2) which can be measured by a pair of electrodes (Fig. 1.11a). When a suitable voltage is impressed on one of the electrode against a reference electrode (typically Ag/AgCl or Calomel), the target species (H_2O_2 or O_2) is reduced at the electrode and this generates electrical current (hence the name 'amperometry'). In potentiometry, a glass membrane or other membrane electrode is used for measuring the membrane potential (hence the name potentiometry) resulting from the difference in the concentrations of H^+ or other positive ions across the membrane. In photometry (Fig. 11c), the light from an indicator molecule is the measured signal. For this method to work, one of the reactants or products of the biorecognition reaction has to linked to colorimetric, fluorescent or luminescent indicator molecules. Usually, an optical

Biolayer Type	Main needs for structural integrity	Typical signal generated
End of Organ (eg. Olfactory)	Intact Tissue Architecture	Action potential
Tissue	Nutrient / Oxygen Supply	Metabolic end product
Whole Cell		
Cell Organelle (eg. Mitochondrion)	Osmotic / pH Stability	Product of electron chain
Biomembrane (eg. Receptor)	Mechanic protection	Released contents
Enzyme	pH / electrolyte stability	Reaction product
Antibody	pH stability	Antigen uptake mass change
Ionophore	Adequate retention	EMF/absorbance change (chromionphore)

Complexity hierarchy

Fig. 10. Possible bioreceptor molecules and molecular assemblies for biosensor applications; their requirements for structural integrity and signals generated.

OPTICAL METHOD

ELECTROCHEMICAL METHOD

(a)

(b)

(c)

Electrode holder

Ag/Ag Cl internal reference

Ag/AgCl anode/ reference

Electrolyte

Internal filling soln (0.1M HCl)

20 μ Pt Cathode

pH sensitive glass membrane

Teflon membrane

Typical pH glass electrode

Radiometer Clark-type electrode
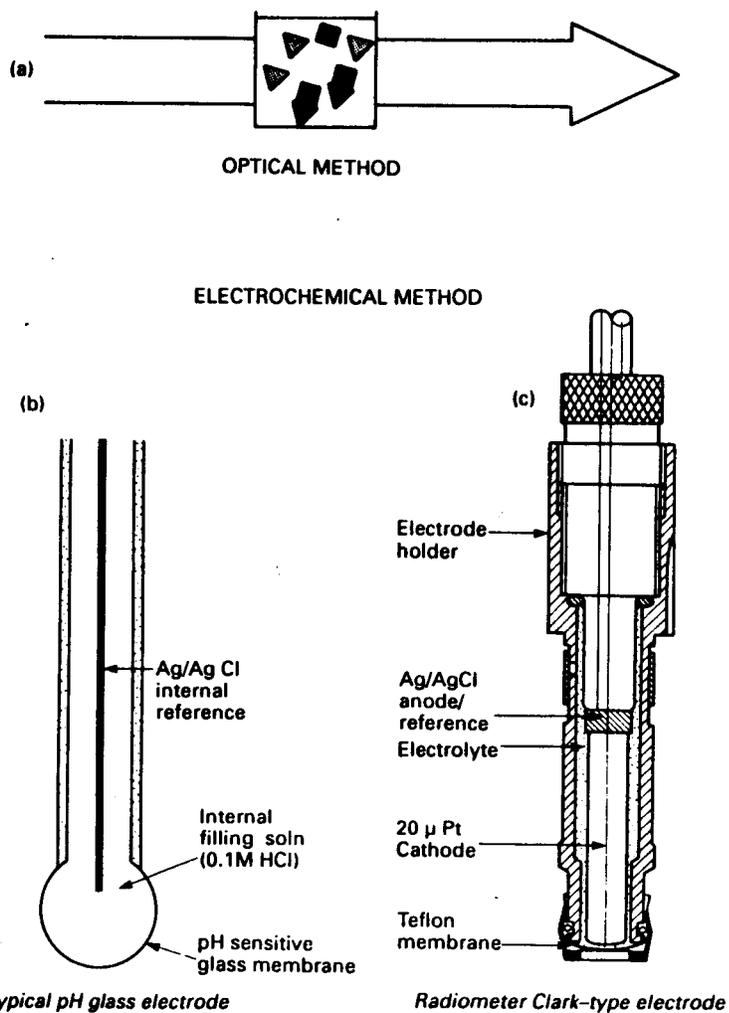
Fig. 1.11. Three conventional transducers used for biosensor development: (a) photometric; (b) potentiometric (based on pH sensor); (c) amperometric (based on Clark oxygen sensor).

Table 1.5. Other transducers used in biosensors.

Category	Measures what?	Examples
Piezoelectric	change in mass	microbalance based sensors
		SAW device based sensors
Conductive	conductivity change	
Capacitive	dielectric constant	antibody sensors
Thermometric	temperature	enzyme thermistor

used for guiding the light signals from the source to the detector. Adaptation and exploitation of these three routes (**photometric, potentiometric and amperometric)** where user acceptability is already established, has been an obvious approach to the development of reagentless biosensor devices with a high specificity and selectivity.

Piezoelectric Transducers. The transducer of a biosensor is not restricted to the three described above. In principle, any variable which is affected by the biorecognition reaction can be used to generate the transduced signal.The **piezoelectric materials** and surface acoustic wave devices offers a surface which is sensitive to **changes in mass.** These transducers have been used where the biorecognition reaction causes a change in mass.

Conductimetry Monitoring **solution conductance** was originally applied as a method of determining reaction rates. The technique involves the measurement of changes in conductance due to the migration of ions. Many enzyme-linked reactions result in a change in total ion concentration and this would imply that they are suitable for conductimetric biosensors.

Capacitance Measurement. When the biorecognition reaction causes a **change in the dielectric constant** of the medium in the vicinity of the bioreceptor, capacitance measurement method can be used as a transducer. Antigen-antibody reaction is a good example. Suppose antibody molecules are immobilized between two metal electrodes of known area. When antigen is added and binds with the antibody, the dielectric constant of the medium between the two electrodes is expected to change significantly. This change translates into a change in capacitance.

Thermometry All chemical reactions are accompanied by the absorption (endothermic) or evolution (exothermic) of heat. Measurements of ΔH, the enthalpy of reaction at different temperatures allows one to calculate ΔS (entropy) and ΔG (Gibbs free energy) for a reaction and therefore collect basic thermodynamic data. The hydrolysis of ATP for example is exothermic:

$$ATP^{4-} + H_2O \sim ADP^{3-} + HPO^{4-} + H^+; \quad \Delta H298 = -22.2 \text{ kJ (pH 7)}$$
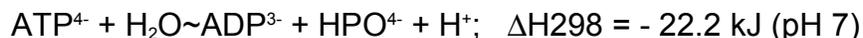
or the immunoreaction between anti-HSA and its antigen HSA yields -30.5 kJ/mol. For this latter reaction, the total increase in temperature for 1 μmol of antibody is of the order of 10^{-5} K, but many enzyme-catalysed reactions have greater ΔH, and produce more easily measurable changes in temperature.

Enzyme Thermistor. For a biosensor device, the biorecognition compound must be immobilized on a temperature-sensing element capable of detecting very small temperature changes. The major initiative in this area has come from the Mosbach group at the University of Lund. Initially, they immobilized glucose oxidase or penicillinase in a small column, so that temperature changes in the column effluent were monitored by thermistors to give an *enzyme thermistor* sensitive to glucose and penicillin, respectively. They have also applied the

technique to other substrates and to immunoassay using an enzyme-labeled antigen.

FET as a Transducer.   As advances are made in biosensors, there was a need for iniaturi-zation and mass production. Field effect transistors (FET) used extensively in semiconductor industry in memory chips and logic chips respond to change in electric field (in front of the 'gate' of FET). An FET is thus capable of detecting changes in ion concentration when the gate is expose to a solution that contains ions. Therefore, pH and ion concentration can be measured with an FET. The advantage of this transducer is that it can be incorporated directly to the electronic signal processing circuitry. In fact, pen-size FET based pH sensor is being marketed commercially.

1.7. Growth of Biosensor

Current Status   Since the development of Clark's glucose sensor, many enzyme electrodes have been developed based on amperometry, potentiometry, and photometry. Some of these biosensors are summarized in Tables 1.6, 1.7, and 1.8. The term 'optode' (see Table 1.8) is used for sensors utilizing optical fiber for light signal transmission. Note that the bioreceptors used are all enzymes (see the 'Bioreceptor' column; enzymes end with suffix '-ase') except the antibody sensor. This represents the current state of the art in biosensor development - that bioreceptors other than enzymes are not explored extensively.

Common Products.   The column 'Product detected' represent the type of transducer used. Note that common products (of biorecognition reactions) are used for measurement. For amperometry, the majority is H_2O_2 (with the exception of NADH and quinone) which is the common product for oxido-reductase enzymes. For potentiometric biosensors, the majority is acid which can be detected by a pH sensor (CO_2 and NH_3 are indirectly detected by measuring the change in pH).

Biosensor Configurations.   When bioreceptor molecules are combined with a suitable transducer, a biosensor is made. Fig. 1.12 shows various biosensor configurations. Note that the bioreceptor molecules are immobilized in a suitable matrix to form a biolayer which is then placed in the immediate vicinity of a transducer. The transducers ion-selective electrode and FET belong to potentiometric transducer category; the coated wire belong to amperometric sensor category; the surface plasma detector and the surface acoustic wave detector belong to piezo transducer category. The materials of constructions for the transducers are also given in the figure.

Discriminative Membranes.   Membranes are one of the essential component of a biosensor. They are used for (1) preventing fouling; (2) eliminating interference; and (3) controlling the operating regime of the biosensor. When a small molecule is the analyte, macromolecules such as proteins can be prevented to enter the active sensing

Table 1.6. Amperometric biosensors

Substrate	Biorecptor	Product detected	Range, mM
choline	choline oxidase	H_2O_2	500
ethanol	alcohol oxidase	H_2O_2	0 - 10
formaldehyde	f. dehydrogenase	NADH	10-3
glucose	glucose oxidase	H_2O_2, O_2	0-7 g/L
glutamine	glutamine oxidase	H_2O_2	0-25
glycerol	g. dehydrogenase	NADH, O_2	
hypoxanthine	x. oxidase	H_2O_2	4-180
lactate	lactate oxidase	H_2O_2	1-40
oligosaccharides	glucoamylase, glucose oxidase	H_2O_2	0.1-2.5
phenol	polyphenol oxidase	quinone	

Table 1.7. Potentiometric Biosensors

Substrate	Bioreceptor	Product detected	Range, mM
aspartam	L-aspartase	NH_3	0.1-0.6
fats	lipase	fatty acids	0.005-0.05
glucose	glucose oxidase	gluconic acid	0.12-2 g/L
urea	urease	NH_4, CO_2	0.01-10
nitrite	nitrite reductase	NH_4	1
penicillin	penicillinase	H^+	0.2-70
sulfate	sulfate oxidase	HS	
antigen or antibody	partner of couple	complex	0-100 ppm

Table 1.8. Enzyme sensors based on optodes

Substrate	Bioreceptor	Product detected	Range, mM
ethanol	alcohol dehydrogenase	NADH	0-1
glucose	glucose oxidase	O_2	0.1-20
urease	urease	ammonia	0.3-3
lactate	lactate monooxygenase	pyruvate	0.5-1
penicillin	penicillinase	penicillinic acid	0.25-10

Table 1.9. Biosensors based on FET (pH)

Substrate	Bioreceptor	Product detected	Range, mM
glucose	glucose oxidase	gluconic acid	0-20
urea	urease	CO_2, NH_3	0-6
penicillin	penicillinase	penicillic acid	0.2-20
triolein	lipase	fatty acids	0.6-3

zone by using a membrane that has small pore size. Note that proteins are notorious for causing fouling of the sensor. The transport of charged molecules can be controlled by using ion exchange type membranes. A combination of different discriminative membrane can also be used for blocking the passage of different interferents. A summary is given in Table 1.10.

Sensitivity Requirements.   The range and type of analytes are also varied and cannot be considered under a single umbrella. The particular application imposes a final concentration range requirement, but initially the concentration level that must be achieved can be estimated by the type of analyte of interest. Metabolites, for example, are commonly found at a level >10^{-6} mol/L, whereas hormones may be in the 10^{-10}-10^{-5} mol/L range and levels as low as 10^{-20} mol/L would be desirable. For virus, 10^{-12} mol/L is desirable. This vast range of concentrations is summarized in Fig. 1.13. It is obvious from this figure that, just based on detection limits, very different approaches should be for an antigen sensor compared with those measuring ion concentration.
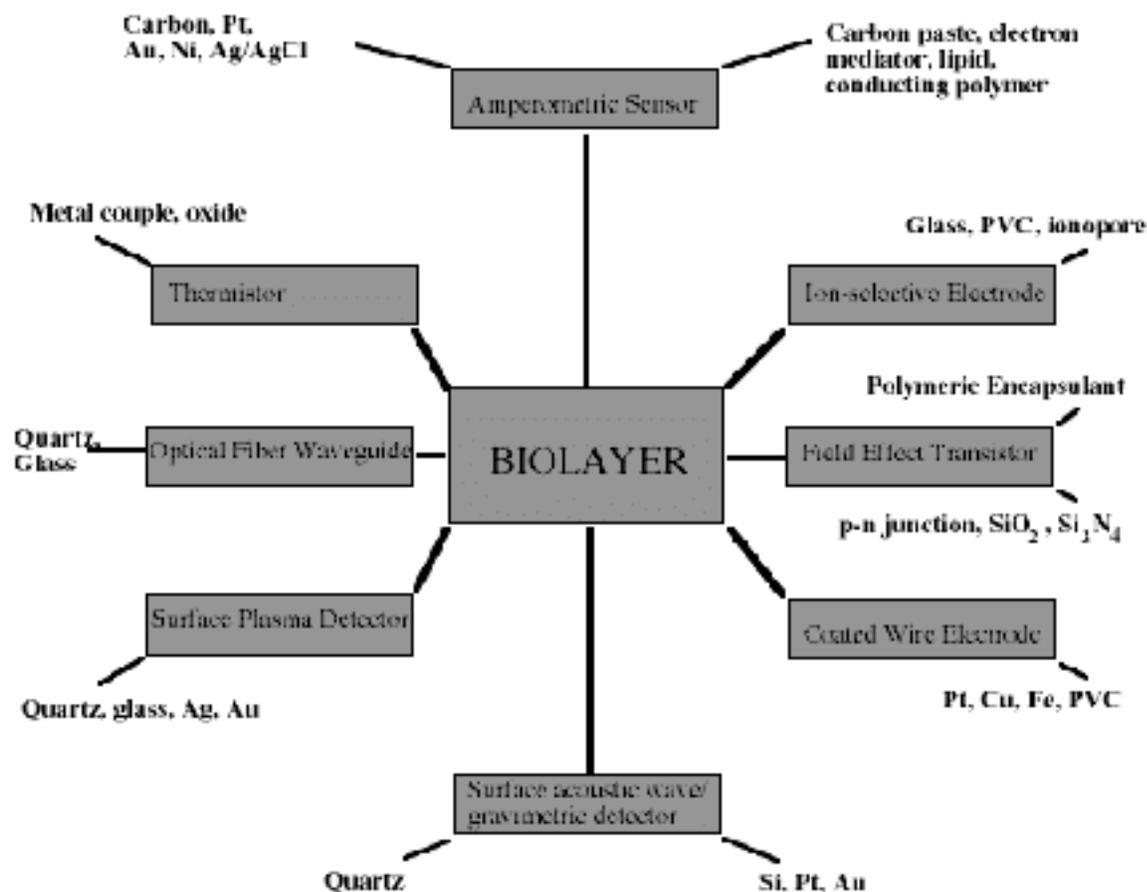
Fig. 1.12. Various biosensor configurations.

Table 1.10. Discriminative coatings for amperometric biosensors.

Transport mechanism	Permeselective film
Size exclusion	Cellulose acetate
	Base-hydrolyzed cellulose acetate
	Phase-inversion cellulose acetate
	Polyaniline, Poly pyrrole
	Polyphenol
	Gamma radiated poly(acrylonitrile)
Charge exclusion	Nafion
	Poly(vinylpyridine)
	Poly(ester-sulfonic acid)
Polarity	Phospholipid
Mixed control	Cellulose acetate - Nafion
	Cellulose acetate - Poly(vinylpyridine)
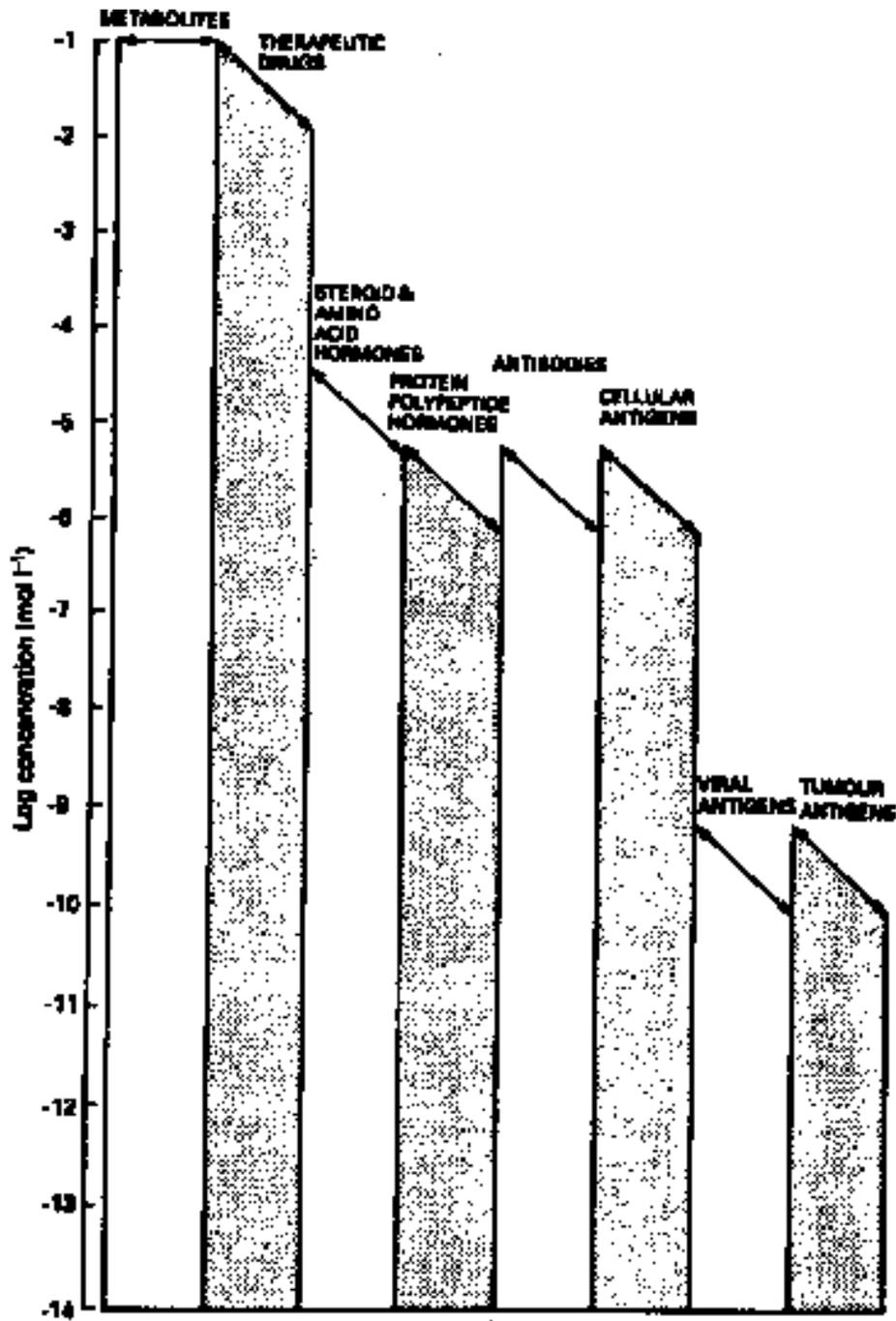
Fig. 1.13. Detection ranges required for some clinically important analytes.

Fig. 1.13. Detection ranges required for some clinically important analytes.

Immunoassay   An important stream of analytical developments, which has been widely applied, is the immunoassay techniques and the DNA probes. In immunoassay, the binding of antibody and antigen results in an increase in

molecular mass and volume. Although current biosensor research is investigating the transduction of this phenomenon, the event is usually followed with a photometric, radioactive or even enzyme marker.

DNA Probe   In DNA probe assay, hybridization of strands of DNA antigen results in an increase in molecular mass and volume. The detection of this event is the same as those of immunoassay. There are many major current reasons for replacing the radioisotopic labels with non-radioactive ones, but the direct use of photometric indicators have rarely provided the same degree of sensitivity, so that enzymes have to date frequently been proven to be the most promising form of labeling. The principle of these label-linked assays is similar for both immunoassays and DNA probes. Both these techniques are heterogeneous assays - so that they are already developed along the lines of the biosensor concept.

Evolution of Biosensors.   Biosensors can be classified into three generations according to the degree of integration of the separate components, i.e. the method of attachment of the biorecognition molecule (= bioreceptor) to the base indicator (transducer) element. In the first generation, the bioreceptor is retained in the vicinity of the base sensor behind a dialysis membrane, while in subsequent generations immobilization is achieved via cross-linking reagents or bifunctional reagents at a suitably modified transducer interface or by incorporation into a polymer matrix at the transduction surface. In the second generation, the individual components remain essentially distinct (e.g. control electronics—electrode—biomolecule), while in the third generation the bioreceptor molecule becomes an integral part of the base sensing element (Fig. 1.14). While these definitions were probably intended for enzyme electrode systems, similar classifications appropriate to biosensors in general can be made. It is in the second and third generations of these families that the major development effort can now be seen.
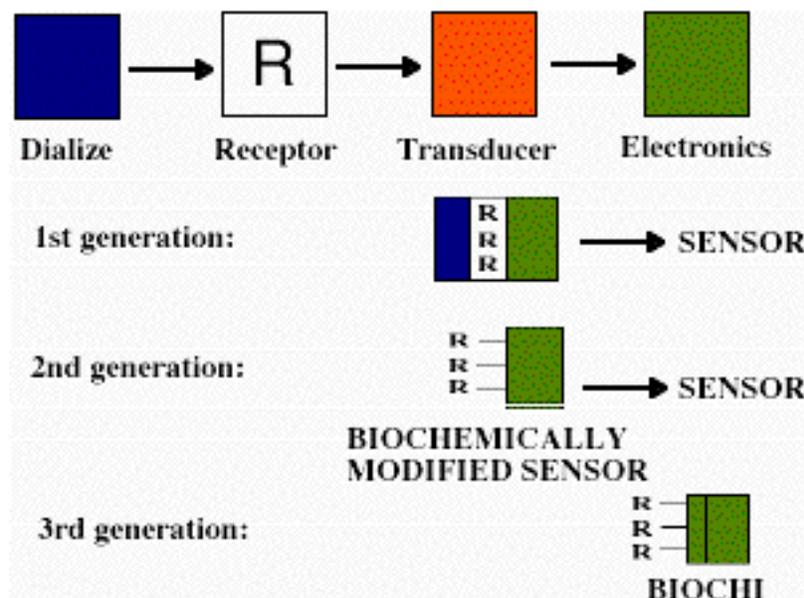
Fig. 1.14. Three biosensor generations (R: Bioreceptor molecule)
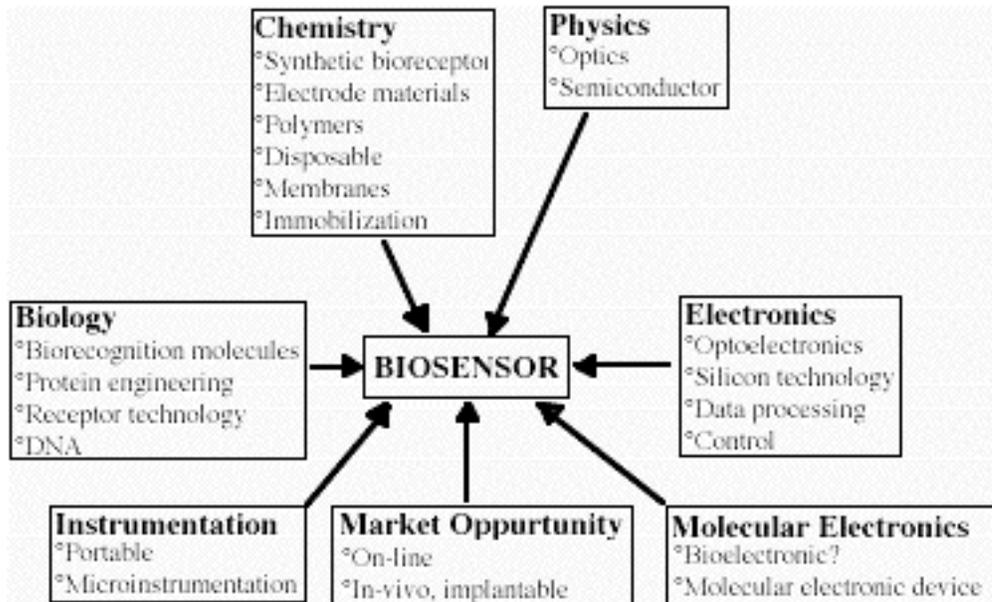
Fig. 1.15. Multidisciplinary nature of bisensor development.


1.8. Future Prospects

Data Processing and Pattern Recognition. If we compare present biosensors with the natural ones (for example, the nose or the eye), they are very crude and simplistic. The recognition molecules in the 'natural sensors' are not necessarily highly specific but the signal transduction via the biomolecules is sophisticated. The specificity often comes from processing of the data collected and recognizing the pattern via a continuous learning process. This mode of operation using the data collected from multiple biosensors is expected to be exploited in the future because the ever increasing capability of microprocessors will provide fast computation.

Micro Instrument. As shown in Fig 1.14, the third generation biosensors have built-in signal processing circuitry. When such sensors are combined with the micro scale valves and actuators currently under development (utilizing micromachining technology), a whole analytical instrument can be built on a silicon wafer. Such an instrument can be mass produced and used in a variety of applications including homes, hospitals, automobiles, toxic dump sites, etc.

Molecular Electronics. The effort to continuously increase the density of electronic components to obtain ever smaller 'packages' will be limited eventually, not by the microlithographic technique employed but by the minimum size allowable for a transistor (note that 'transistor' is the building block of microprocessors and memory chips). Many biological molecules are able to synthesize complex self-organizing molecules with apparently just the required

electronic properties. This suggests that the solution to this problem may be found in replacing silicon with biomolecular components. This idea has led to the proposition of many *molecular electronic* systems. In the past, materials and processing methods developed for microelectronic applications have been exploited in sensor developments. Therefore, any future developments in molecular electronics are expected to be imported into biosensor technology.

Multi-Disciplinary Nature.   The arena of expertise required for biosensor development can be sustained by collaboration from many areas of academia and industry (as illustrated in Fig. 1.15). The resulting output of this collaboration is likely in many cases to be a slow process, but is probably the only realistic route to successful future advances.
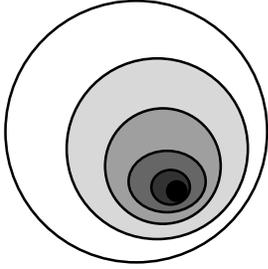
Chapter 2. Transducer Fundamentals

2.1. Dissolved Oxygen Electrode and Amperometry

2.1.1. Principle of Amperometry (Polarography)

Polarogram   When an electrode of noble metal such as platinum or gold is made 0.6 to 0.8 V negative with respect to a suitable reference electrode such as Ag/AgCl or an calomel electrode in a neutral KCl solution (see Fig. 2.1), the oxygen dissolved in the liquid is reduce at the surface of the noble metal. This phenomenon can be observed from a current-voltage diagram - called a *polarogram* - of the electrode. As shown in Fig. 2.2a, the negative voltage applied to the noble metal electrode (called the cathode) is increased, the current increases initially but soon it becomes saturated. In this plateau region of the polarogram, the reaction of oxygen at the cathode is so fast that the rate of reaction is limited by the diffusion of oxygen to the cathode surface. When the negative bias voltage is further increased, the current output of the electrode increases rapidly due to other reactions, mainly, the reduction of water to hydrogen). If a fixed voltage in the plateau region (for example, - 0.6V) is applied to the cathode**,** the current output of the electrode can be linearly calibrated to the dissolved oxygen (Fig. 2.2b). It has to be noted that the current is proportional not to the actual concentration but to the activity or equivalent partial pressure of dissolved oxygen, which is often referred to as *oxygen tension. A* fixed voltage between -0.6 and -0.8 V is usually selected as the polarization voltage when using Ag/AgCl as the reference electrode.

DO Sensor   When the cathode, anode, and the electrolyte are separated from the measurement medium by a polymer membrane, which is permeable to the dissolved gas but not to most of the ions and other species, and when most of the mass transfer resistance is confined in the membrane, the electrode system can measure oxygen tension in various liquids. This is the basic operating principle of the *membrane covered polarographic DO probe* (Fig 2.3).

Signal Conditioning.   To read the output from the sensor, the current from the sensor is converted to voltage by the circuit shown in Fig. 2.4 (the first 1/2 of an operational amplifier LF412). This circuit has a gain of 10,000,000:

$$V_1 = - I * 10{,}000{,}000$$

Therefore, 0.1 μA sensor current will produce an output of - 1V at pin 1 of LF412 (note that R1 can be changed to obtain other amplifier
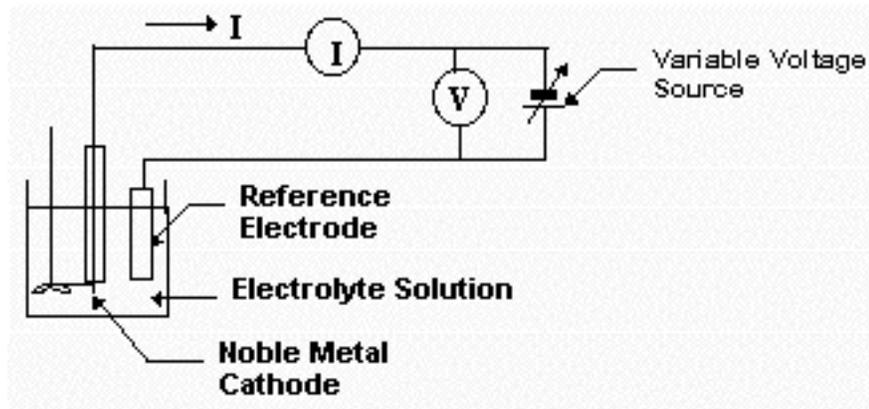
Fig. 2.1. Setup for polarography.
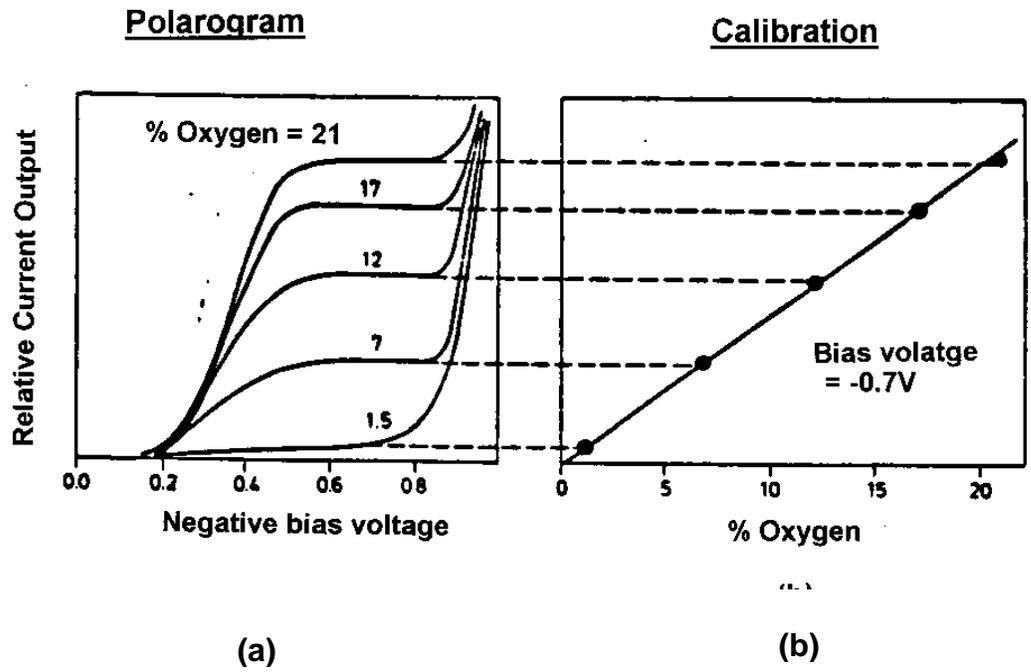

Polarogram

Calibration

(a)

(b)

Fig. 2.2. (a) Current-to-voltage diagram at different oxygen tensions; (b) calibration obtained at a fixed polarization voltage of -0.6V.
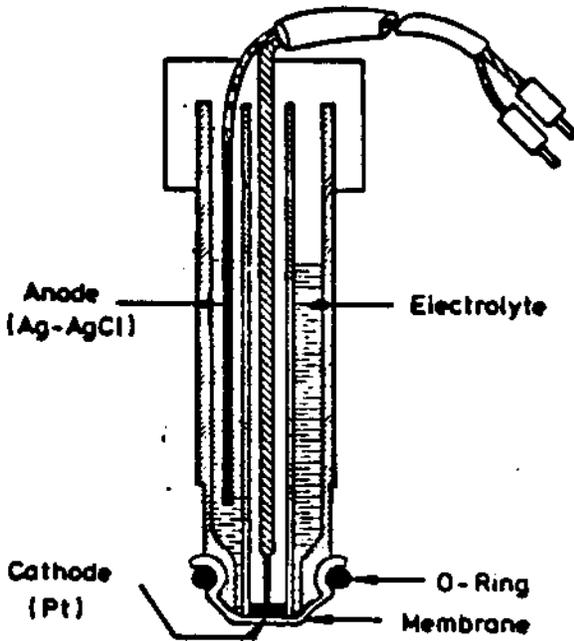
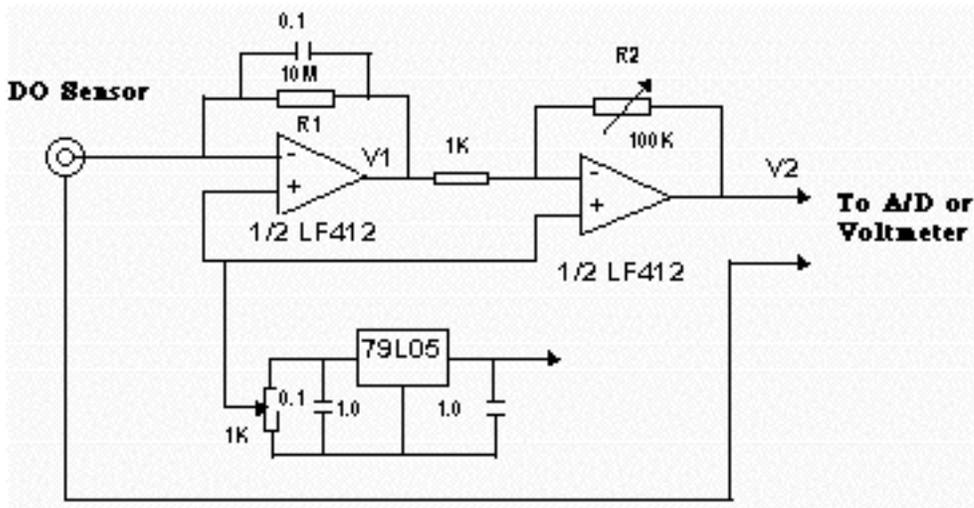Fig. 2.3. Membrane covered polarographic oxygen sensor.

Fig. 2.4. Circuit for current to voltage conversion and application of polarization voltage.
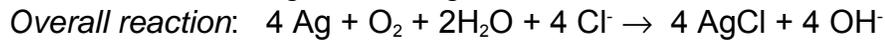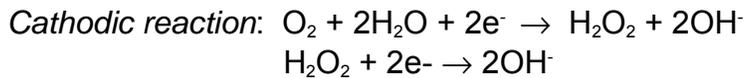
gains). The next stage is an inverting amplifier with gain. The output from this stage is:

$$V_2 = - V_1 \times (R2/1000) = 10,000 \times R2$$

where R2 is the resistance in the feedback loop which can be adjusted. The application of the polarization voltage is done by a 79L05 voltage regulator that converts its input voltage of -12V to -5V. At the output of 79L05, a voltage divider (R3) to convert -5V to -0.7V, which is then applied to the + input of LF412. The voltage output V2 can be read either by a voltmeter or by a computer equipped with an analog-to-digital converter.

Electrode   For polarographic electrodes, the reaction proceeds as follows:
Reactions

Cathodic reaction: $O_2 + 2H_2O + 2e^- \rightarrow H_2O_2 + 2OH^-$

$H_2O_2 + 2e^- \rightarrow 2OH^-$

Anodic reaction: $Ag + Cl^- \rightarrow AgCl + e^-$

Overall reaction: $4\ Ag + O_2 + 2H_2O + 4\ Cl^- \rightarrow 4\ AgCl + 4\ OH^-$

the reaction tends to produce alkalinity in the medium together with a small amount of hydrogen peroxide.

Number of Electrons Involved.   Two principal pathways was proposed for reduction of oxygen at the noble metal surface. One is a *4 electron pathway* where the oxygen in the bulk diffuses to the surface of the cathode and is converted to H_2O via H_2O_2 (path a in Fig. 2.5). The other is a *2 electron pathway* where the intermediate H_2O_2 diffuses directly out of the cathode surface into the bulk liquid (path b in Fig. 2.5). The oxygen reduction path may change depending on surface condition of the noble metal. This is probably the cause for time-dependent current drift of polarographic probes. Since the hydroxyl ions are constantly being substituted for chloride ions as the reaction starts, KCl or NaCl has to be used as the electrolyte. When the electrolyte is depleted of Cl^-, it has to be replenished.

2.1.2. Relationship between Sensor Output and Design Variables

One Layer Model   The current output of the sensor can be related to its design parameters by using a simplified electrode model. The assumptions used for developing the model:
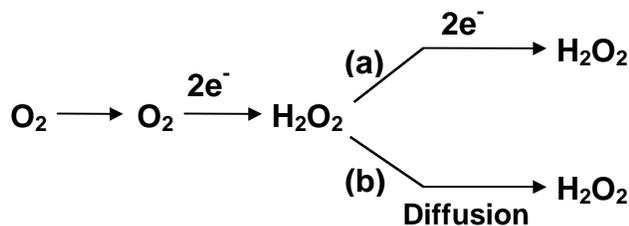

_____

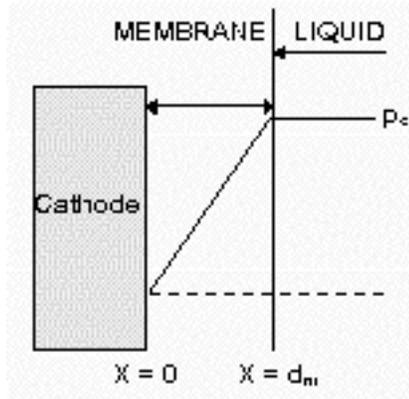**Fig. 2.5. Alternative pathways of oxygen reduction at cathode surface.**

Fig. 2.6. One layer electrode model.

Current Output:	$I_s = NFA\dfrac{P_m}{d_m}p_o$
Response Time:	$\tau = \dfrac{d_m{}^2}{D_m}$
Design Variables:	P_m, d_m, D_m, A

Fig. 2.7. Design variables of DO sensor.

Assumptions Used
1. The cathode is well polished and the membrane is tightly fit over the cathode surface such that the thickness of the electrolyte layer between the membrane and the cathode is negligible.
2. The liquid around the sensor is well agitated that the partial pressure of oxygen at the membrane surface is the same as that of the bulk liquid.
3. Oxygen diffusion occurs only in one direction, perpendicular to the cathode surface.

1-D Diffusion Equation   This so-called one layer model can be extended to include the effects of other layers (such as the liquid layer) as will be shown later. According to Fick's 2nd law, the unsteady-state diffusion in the membrane is described by (using the coordinate system shown in Fig. 2.6):

$$\frac{\partial p}{\partial t} = D_m \frac{\partial^2 p}{\partial x^2} \qquad (1)$$

where D_m is the oxygen diffusivity in the membrane and x is the distance from the cathode surface.

Boundary Conditions   The initial and boundary conditions are:

$$p = 0 \qquad \text{at } t = 0 \qquad\qquad (2)$$
$$p = 0 \qquad \text{at } x = 0 \qquad\qquad (3)$$
$$p = p_o \qquad \text{at } x = d_m \qquad\qquad (4)$$

where d_m is the membrane thickness and p_o is the partial pressure of oxygen in the bulk liquid. The second boundary condition (Eq. (3)) assumes a very fast reaction at the cathode surface, which is generally achieved when the cathode is properly polarized (with the right voltage).

Unsteady Oxygen Profile   The solution of Eq. (1) with the boundary conditions is:

$$\frac{p}{p_o} = \frac{x}{d_m} + \sum_{n=1}^{\infty} \frac{2}{n\pi}(-1)^n \sin\frac{n\pi x}{d_m}\exp(-n^2\pi^2 D_m t / d_m^2) \qquad (5)$$

The current output I of the electrode is proportional to the oxygen flux at the cathode surface:

$$I = NFAD_m(\partial C/\partial x)_{x=0} \qquad\qquad (6)$$

$$= NFAP_m(\partial p/\partial x)_{x=0}$$

where N, F, A, and P_m are the number of electrons per mole of oxygen reduced, Faraday constant (= 96,500 coul/mol), surface area of the cathode, and oxygen permeability of the membrane, respectively. The permeability P_m is related to diffusivity by:

$$P_m = D_m S_m \qquad\qquad (7)$$

where S_m is the oxygen solubility of the membrane.

Unsteady Current Output   From Eqs. (5) and (6), the current output I_t of the electrode is:

$$I_t = NFA\frac{P_m}{d_m}p_o\left[1 + 2\sum_{n=1}^{\infty}(-1)^n \exp(-n^2\pi^2 D_m t / d_m^2)\right] \qquad (8)$$

Steady Current Output   The pressure profile within the membrane and the current output under steady-state conditions can be obtained from Eqs. (5) and (8), respectively:

$$\frac{p}{p_o} = \frac{d}{d_m} \qquad\qquad (9)$$

and

$$I_s = NFA\frac{P_m}{d_m}p_o \qquad (10)$$

At steady state, the pressure profile is linear and the current output is proportional to the oxygen partial pressure in the bulk liquid. Eq. (10) forms the basis for DO measurement by the sensor.

Response depends on Time   Eq. (8) shows that the rapidness of the sensor response depends on the following term:

$$\tau = \frac{d_m{}^2}{D_m} \qquad (11)$$

When τ is large (a thin membrane and/or high D_m), the sensor responds fast. Note that these conditions tend to weaken the assumption of membrane-controlled diffusion. Therefore, a compromise has to be made for an optimum sensor performance. Note that adjusting d_m (rather than D_m) is more effective in adjusting τ (because τ depends on the square of d_m). Eqs. (10) and (11) indicate that the design variables for a DO sensor is P_m, d_m, D_m, and A (Fig. 2.7).

2.1.3. Effect of Liquid Velocity and Properties on Sensor Performance

Real Situation   In reality, the assumption 2 made earlier is not entirely satisfactory. A stagnant liquid film always exists outside the membrane even at high liquid velocity. Actually, when a DO sensor is used to measure DO concentration in two different liquids at the same partial pressure of oxygen, the readings are not the same. This indicates that the sensor output depends, to a certain extent, on the properties of the liquid. One layer model cannot explain this behavior. A more realistic model has to consider both the membrane and the liquid film as shown in Fig. 2.8.

Two Layer Model   The effect of liquid layer on sensor current output can be estimated
by expanding the "one layer" model. At steady state, the oxygen flux J through each layer in Fig. 2.8 should be the same:

$$\begin{aligned}
J &= K p_o \\
&= k_L(p_o - p_m) \\
&= k_m p_m
\end{aligned} \qquad (12)$$

where K is the overall mass transfer coefficient and k_L and k_m are mass transfer coefficient for liquid film and membrane, respectively.

Ohm's Law Analogy   The inverse of the mass transfer coefficient can be termed as the mass transfer resistance. From Eq. (12), it can be shown that:

$$\frac{1}{K} = \frac{1}{k_L} + \frac{1}{k_m} \tag{13}$$

Equation (13) says that the overall mass transfer resistance, $1/K$, is the sum of the liquid phase mass transfer resistance, $1/k$, and the membrane phase mass transfer resistance, $1/k_m$. The derivation is based on Ohm's law analogy - J is considered the current and p as the voltage. The individual resistances can be replaced by:

$$\frac{1}{K} = \frac{d_L}{P_L} + \frac{d_m}{P_m} \tag{14}$$

where d_L and P_L are liquid film thickness and the oxygen permeability of the liquid film, respectively. Note that a stagnant liquid film was assumed here, although it is more accurate to use the convective mass transfer coefficient.
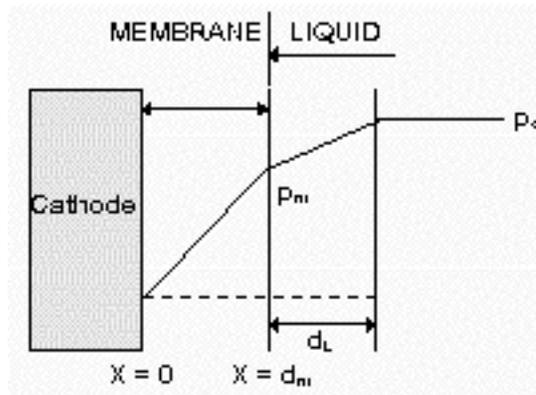
Fig. 2.8. Two layer model for DO sensor.

(a) Sensor output:

$$I = NFA \frac{p_o}{d_m / P_m + d_L / P_L}$$

(b) Response Time

$$\tau = d^2/D_m$$

where

$$d = d_m + (D_m/D_L)^{1/2} \, d_L$$

τ

LIQUID VELOCITY
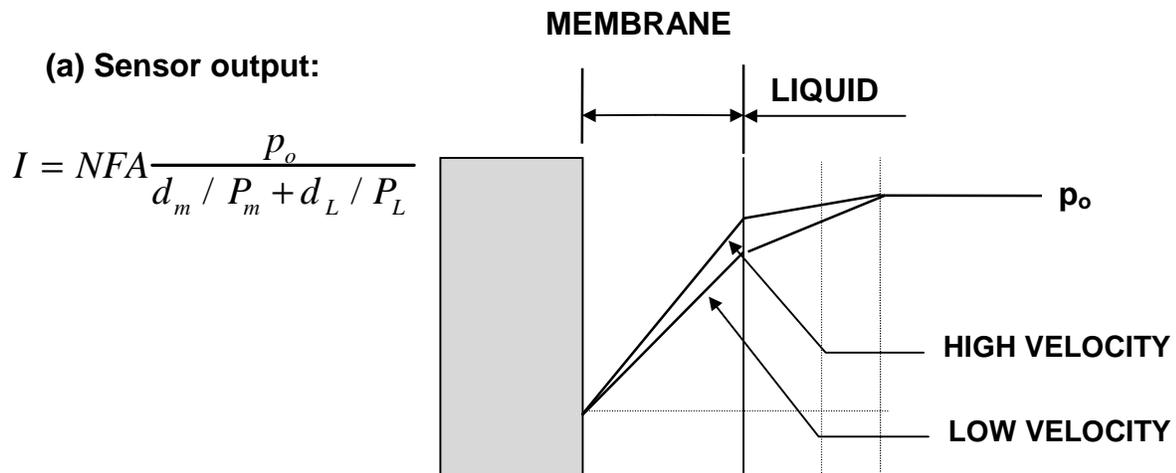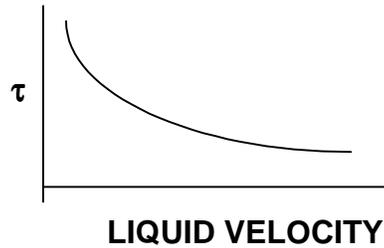
Fig. 2.9. Flow sensitivity of DO sensor.

Two Layer Model   When the individual mass transfer resistances are considered the steady state sensor output becomes:

$$I_s = NFA\frac{P_m}{\overline{d}}\, p_o \tag{15}$$

where $\overline{d}$ is defined by:

$$\overline{d} = d_m + P_m \frac{d_L}{P_L} \tag{16}$$

Alternatively, I_s can be written as:

$$I_s = NFA\frac{p_o}{d_m/P_m + d_L/P_L} \tag{17}$$

The time constant τ of Eq. (11) can be modified to:

$$\tau \quad = \frac{d^2}{D_m} \tag{18}$$

where d is defined by:

$$d = d_m + \sqrt{\frac{D_m}{D_L}} d_L \qquad (19)$$

Flow Sensitivity   The DO sensor, when placed in a stagnant liquid, produces a
diffusion gradient extending outside the membrane and farther into the liquid..
When the liquid is stirred, the diffusion gradient can no longer be extended
beyond the liquid film around the membrane. Since the diffusion gradient
becomes steeper with decreasing liquid film thickness, the current output of the
sensor increases with increase in liquid velocity (Fig. 2.9a). Note also that the
response time of the sensor increases as the liquid velocity decreases (Fig.
2.9b). This so-called "flow sensitivity" is greater for a sensor with a larger
cathode because the size of the stagnant diffusion field is proportionally greater
with a larger cathode.

Condition for Membrane Control of Diffusion   From Eq. (17), the
condition for a membrane-controlled diffusion becomes:

$$d_m/P_m >> d_L/P_L \qquad (20)$$

To achieve this condition, a relatively thick membrane with a low oxygen
permeability have to be used. Note that this contradicts the requirement for a fast
sensor response. When this condition is achieved, the oxygen sensor output
depends only on membrane properties as given by Eq. (10) and the sensor
calibrated in one liquid can be used in other liquids without recalibration. In
reality, however, there is always a liquid film (however thin it may be) and this
causes variations in calibration in different liquids.

2.1.4. Effect of Temperature on Sensor Performance

Temperature Dependency   It has been observed that there is 1 to 5%
increase in sensor output current per $^{\circ}C$ increase in temperature (Fig. 2.10a).
The temperature effect comes from P_m and P_L in Eq. (17) because they are
functions of temperature. When the sensor is operated under membrane-
diffusion control, the temperature dependency of I_s should come entirely from P_m
whereas when the liquid film resistance is not negligible, both P_m and P_L
contribute to the temperature dependency. Generally, P_m is expressed as:

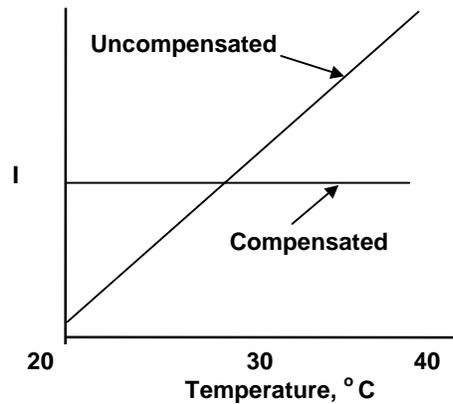$$P_m = P_m^* \exp(-E/RT) \qquad (21)$$

where E is the activation energy for the permeation and amounts to be 8.8
kcal/g-mol for polyethylene and 7.8-9.6 kcal/g-mol for polypropylene membranes.
The response time of the sensor also depends on temperature but the
temperature dependency originates from diffusivity (see Eqs. (18) and (19)). As
the temperature increases, the diffusivity increases and this makes the sensor
respond faster (Fig. 2.10b). In commercial instruments, thermistors are used for
compensating temperature variation in sensor output (see Fig. 2.10a).

2.1.5. Effect of Cathode Diameter on Sensor Performance

Effect of Sensor Size on Flow Dependency   When a DO sensor is placed in a stagnant liquid a diffusion field is generated due to the consumption of oxygen at the cathode surface. The size of steady-state diffusion field is proportional to
the size of the cathode. When the liquid is stirred, the diffusion gradient can no longer be extended beyond the liquid film around the membrane. Since

 (a) Sensor output:

$$I = NFA \frac{p_o}{d_m / P_m + d_L / P_L}$$

Uncompensated

Compensated

I

20   30   40

Temperature, $°C$

(b) Response Time

$$\tau = d^2/D_m$$

where

$$d = d_m + (D_m/D_L)^{1/2} d_L$$

τ
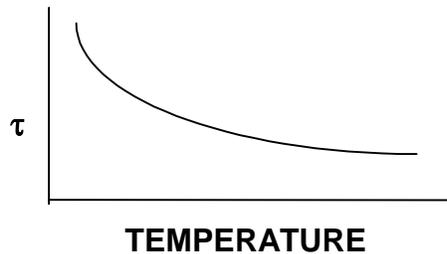
TEMPERATURE

Fig. 2.10. Temperature dependency of DO sensor.

Small Cathode

Large Cathode

I

V_c V_c
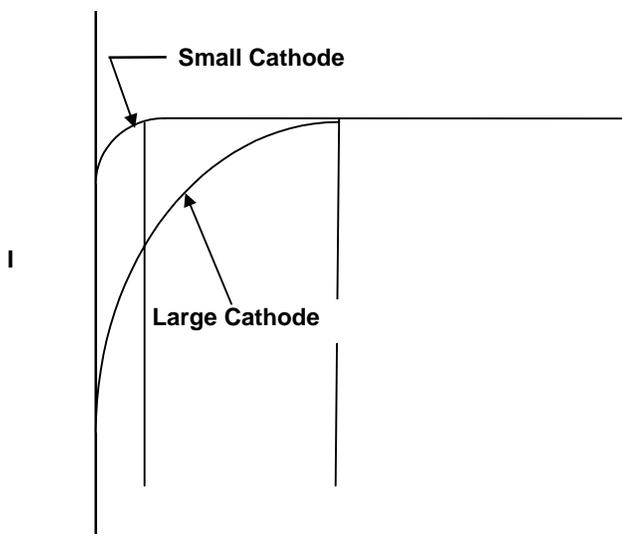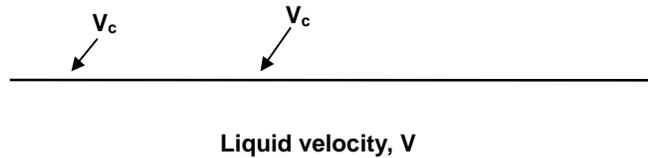
Liquid velocity, V

Fig. 2.11. Effect of cathode diameter on flow dependency of DO sensor.

the diffusion gradient becomes steeper with decreasing liquid film thickness, the sensor output current increases with increase in liquid velocity as shown in Fig. 2.11. This flow dependency is larger for larger cathode area because the size of the stagnant diffusion field is proportionally larger. For proper operation of the sensor, the liquid has to be stirred beyond a certain level to maintain membrane control of oxygen diffusion. The critical velocity, V_c, of the liquid is the velocity where the probe output reaches 95% (95-99% depending on the definition) of the steady state value. For a given liquid, V_c is smaller for smaller cathodes. For example, with a 25 μm Teflon membrane, a cathode of 5 mm diameter requires V_c of 70 cm/s in water, whereas only 5 cm/s is required for 25 μm cathode. When the cathode diameter is less than 1 μ m, the sensor becomes insensitive to liquid flow even without the membrane. In this case, the diffusion field of the cathode is so small that it is always contained inside the minimum liquid boundary layer around the cathode.

Effect on Current Output   One obvious effect of the cathode size is the area effect. Eq. (17)shows that the current output is directly proportional to the sensor area A. When the current produced by the sensor is too small (when the cathode area is small), the sensor signal tend to be more susceptible to noise. The current output should be greater than 10^{-6} to 10^{-7} range for ease of signal amplification.

2.2. Construction of a Steam-Sterilizable Dissolved Oxygen Electrode

2.2.1. Fabrication of DO Sensor

DO Measurement   Oxygen must be supplied to growing microorganisms just like we need it for maintaining our lives. Oxygen dissolves very sparingly in water - about 8 ppm (parts per million by weight) at 20°C. Therefore, one of the major design variables for a bioreactor is the oxygen transfer rate (OTR) because the maximum growth rate of an aerobic microorganism depends on it. The dissolved oxygen in bioreactors are usually measured by a dissolved oxygen (DO) sensor. In this Section, we will construct a steam-sterilizable DO sensor suitable for bioreactor applications. The idea is to appreciate the relationship between the sensor design/fabrication methods and the sensor performance.

Materials of Construction   The following materials are needed:

Electrode material
• 75 µm (or 25 to 125 µm) diameter **Pt wire**, 3 cm length
• a short length (20 cm) of **Ag wire**, 0.25 mm OD
• 1 mil (1"x1") thickness Teflon membrane
• 1M KCl solution in ethylene glycol

Construction aids
• lead glass capillary tube (2 mm OD, 0.5 mm ID), 15 cm long
• glass tube (10 mm OD, 8 mm ID), 12 cm long
• a short length (20 cm) of copper wire, 0.25 mm OD
• a short length (1 cm) of 7 mm ID silicon tubing
• a short length (10 cm) of 2 mm ID rubber tubing
   (for use as a vent tube)
• a BNC female connector (obtainable from Radio Shack)
• epoxy (2 part, ribbon type)

Chemicals for cleaning and plating
• 1M nitric acid solution
• Concentrated nitric/sulfuric (1:1) solution
• Methyl chloroform solution
• 0.1 M HCl solution

Fabrication
**Step 1.** Cut a short length of copper wire (20 cm) and attach 75 µm Pt wire by soldering
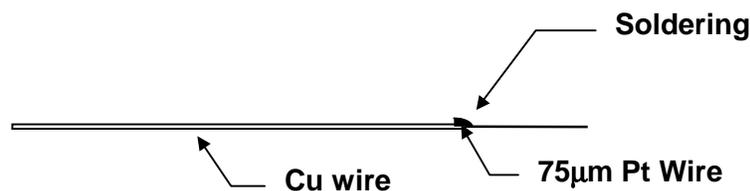
Step 2. Clean the Pt wire (just the Pt wire portion, not the Cu wire) by dipping in a 1:1 concentrated nitric and sulfuric acid solution. Wash with DI water.

Step 3. Place it in the glass capillary tube (cleaned) as shown. Flame the end until the glass melts and fuses around the Pt wire. A good wetting of the Pt wire with molten glass is necessary. Lead glass is best for this purpose because, it wets the metal well and the thermal expansion coefficient is closest to that of Pt.

Pt wire

Glass capillary
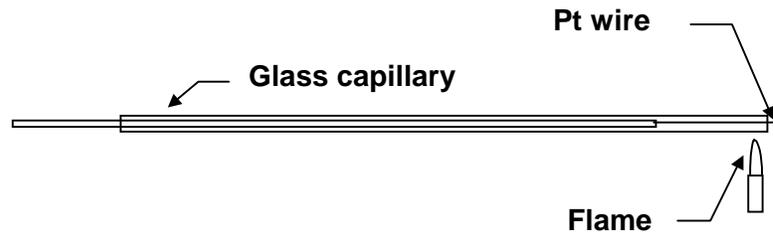
Flame

Step 4. Grind the flamed end flat (use a sand paper: start with a coarse one and then gradually use finer grit sand paper). The end should looks like:
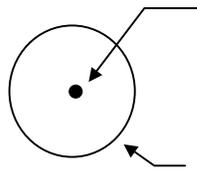

It is important that the glass seals the metal tightly around the metal wire. The goodness of the seal may be observed under a microscope but a better way is to measure the conductivity in an electrolyte solution (by using Ag/AgCl as the counter electrode). The conductivity should remain more or less constant. If it increases with time, the seal is not good and a DO sensor made with such a leaky seal tend to be unstable.

Step 5. Grind flat both ends of the 10 mm OD glass tube. Grind the edge of one end round.


Grind this end flat

Round edges

Step 6. Make a Ag/AgCl reference electrode by electrolytic chlorination of Ag wire.

- Clean Ag wire by dipping in 1 M nitric acid for 10 s.
- Anodize the wire in 0.1M HCl at current density of 0.4 mA/cm² for 30 min.
- Store in 1M KCl solution overnight before use.

1.5V

Ag

Ag

0.1 M HCl
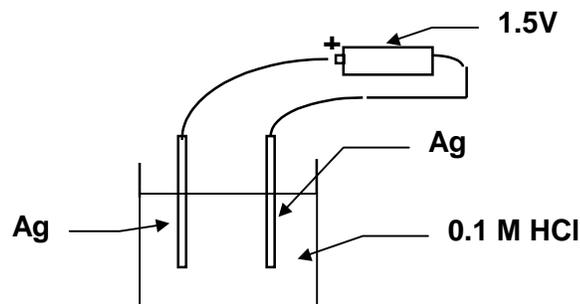
When the chlorination is done properly, the wire will have a brownish coat on it.

Step 7. Put the glass capillary tube in the large glass tube, and insert the chlorinated silver wire and the vent tube.
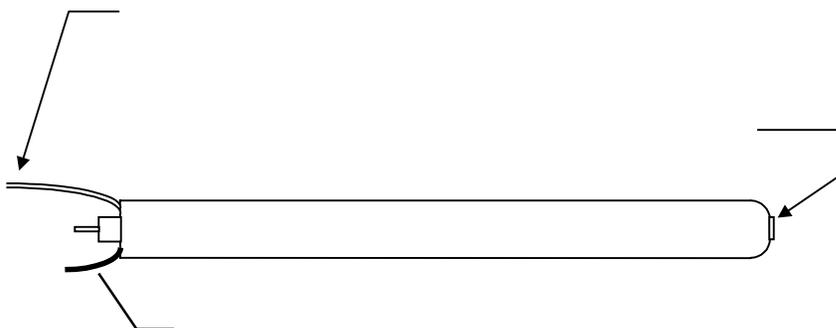

Step 8. Connect the copper wire to the center pin of BNC connector; connect the Ag/AgCl wire to the outer body of BNC by soldering.


Step 9. Fix the BNC connector and the vent tube by molding with epoxy.
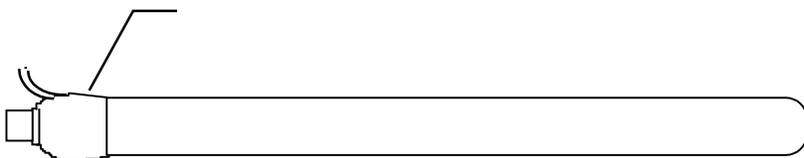
Step 10. Cut a 3/4" length of silicone tubing and place it in methyl chloroform for 2 min. The tubing swells. Cover the free end of the sensor assembly with a Teflon membrane and fix it tightly in place with the swelled silicone tubing. Make sure the membrane is tightly fit. Let the solvent evaporate.

Silicone tubing

Teflon membrane
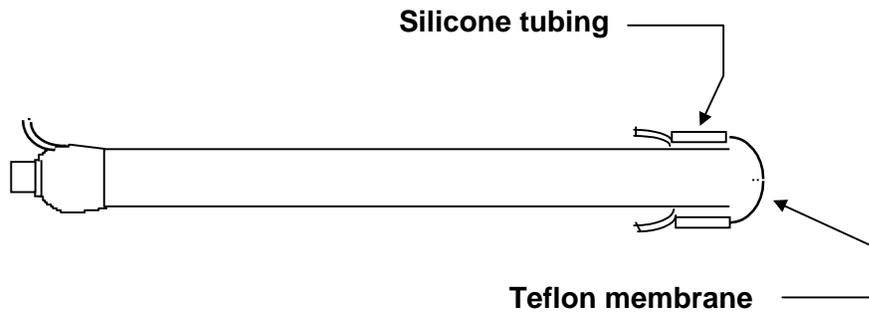
Step 11. Fill the electrolyte solution with a syringe via the vent hole. After the filling, clamp the vent tube.

Syringe filled with electrolyte
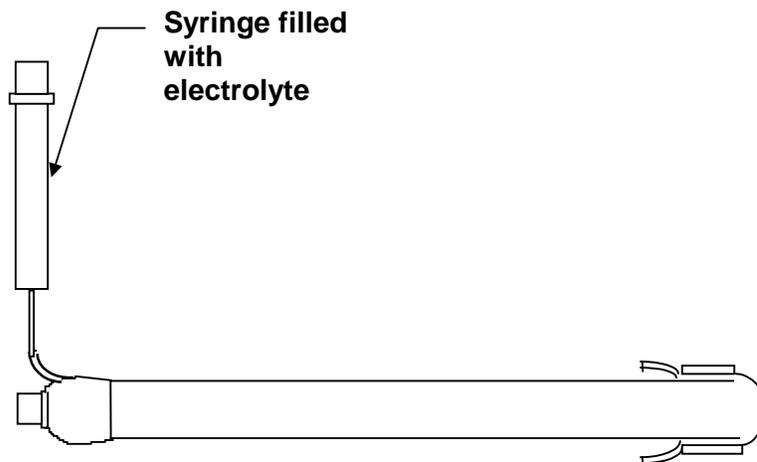
2.2.2. Calibration and k_La Measurement

Liquid Phase Calibration   Prepare an air saturated water by passing air bubbles into a small volume (100 mL) of water. Prepare a nitrogen saturated water in the same way. Connect the fabricated DO sensor to signal amplifying circuit of Fig. 2.3b, and then measure the voltage output for both water solutions. The liquids have to be agitated at high speed to obtain proper calibration. This is so-called a 'two point' calibration.

Gas Phase Calibration   Perform the calibration in gas phase by exposing the sensor to air. Do the same using nitrogen as the gas phase. Compare the two calibrations (between liquid and gas). Should they be the same? If not, why not?

Measurement of Response Time   The other important parameter of the sensor is the Its response time can be measured by making a step change in

oxygen partial pressure in the measurement medium and measuring the sensor response. The sensor can be approximated as a first order system:

$$c - c_p = \tau_p \, (dc_p/dt) \tag{22}$$

where c is the oxygen concentration in the measurement sample, c_p is the oxygen concentration measured by the sensor, and τ_p is the sensor time constant. When a step change is made in c (by transferring the sensor from air into a nitrogen saturated, stirred water), the sensor output decreases roughly exponentially (not exactly exponentially because the sensor may not be a true first order system). The time constant τ_p is the time when the sensor response reaches 63.7% of the ultimate response (Fig. 2.12a). The solution to Eq. (22) with the following boundary condition is an exponential function.

$$c = 1 \qquad \text{at} \quad t = 0 \tag{23}$$

Note that a normalized concentration is used: c of 1 means 100% air saturation and 0 means nitrogen saturation. The solution is:

$$c/c_p = 1 - \exp(-t/\tau_p) \tag{24}$$

Eq. (24) indicates that when $t = \tau_p$, c/c_p will be 0.64. The time constant τ_p can also be determined conveniently by using an integral method - the area above the response curve is equal to τ_p (see Fig. 2.12b). This method is especially useful when there is a lot of noise in the measured signal. The integration can be carried graphically using either trapezoidal rule of Simpson's rule.

Measurement of $k_L a$ The oxygen absorption capability of a bioreactor is represented by $k_L a$, the liquid phase overall volumetric mass transfer coefficient. DO sensor is used frequently to measure $k_L a$. Typically, the reactor is first sparged with nitrogen and at time zero, the nitrogen is switched to air. The oxygen mass balance in the reactor yields:

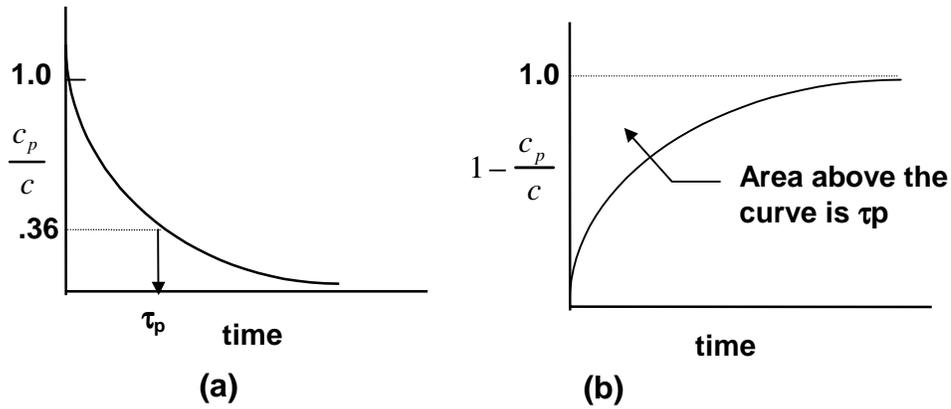$$dc/dt = k_L a (c^* - c) \tag{25}$$

Fig. 2.12. (a) Sensor response time measurement;
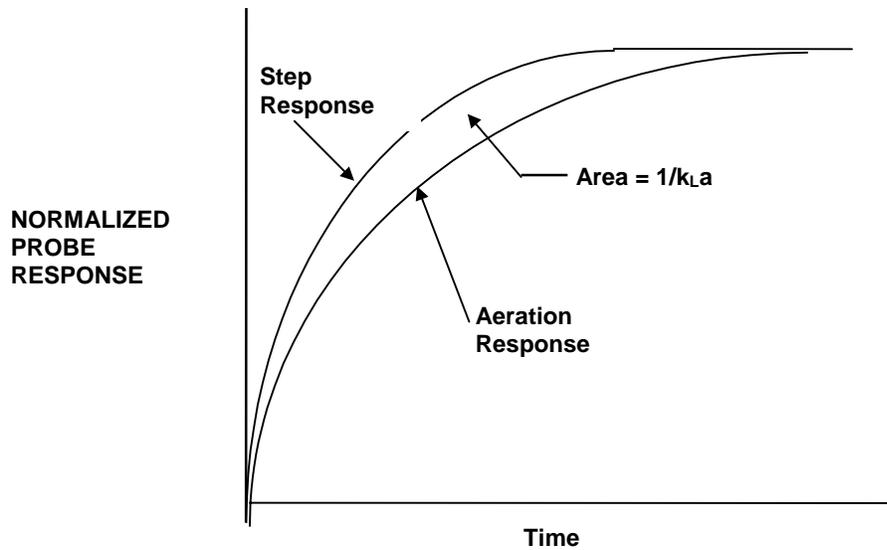(b) integral method for measuring the sensor time constant.

Fig. 2.13. Measurement of $k_L a$ by integral method.

where c is the oxygen concentration in the reactor and c^* is the oxygen concentration at the gas-liquid interface. This equation can be rearranged to:

$$c^* - c = \tau_k (dc/dt) \qquad (26)$$

where

$$\tau_k = 1/k_L a \qquad (27)$$

Eqs. (22) and (26) can be solved simultaneously to obtain an expression for $k_L a$. However, τ_k can be obtained graphically as shown in Fig. 2.13 when τ_p is known.

Caution in $k_L a$ Measurement   Note that the magnitude of τ_p depends on the liquid velocity in the vicinity of the sensor. Therefore, if a τ_p measured at one agitation rate is used for measuring $k_L a$ for different agitation rates, the results will be in error. A safe way is to used the same agitation rate for both τ_p and τ_k measurements. However, if τ_k is much greater than τ_p, such a precaution is not necessary.

2.3. Peroxide Sensor

Many oxidoreductase enzymes produce H_2O_2 as the reaction product. An example is glucose oxidase enzyme that produces H_2O_2 when glucose reacts with oxygen. H_2O_2 can be detected by polarography just as with DO sensor. The only difference is that the polarization voltage has to be + 0.7 V instead of - 0.7 V. The circuit given in Fig. 2.14 has to be used instead of the one given in Fig. 2.4 (note that +0.7 V is applied using 78L05 voltage regulator).
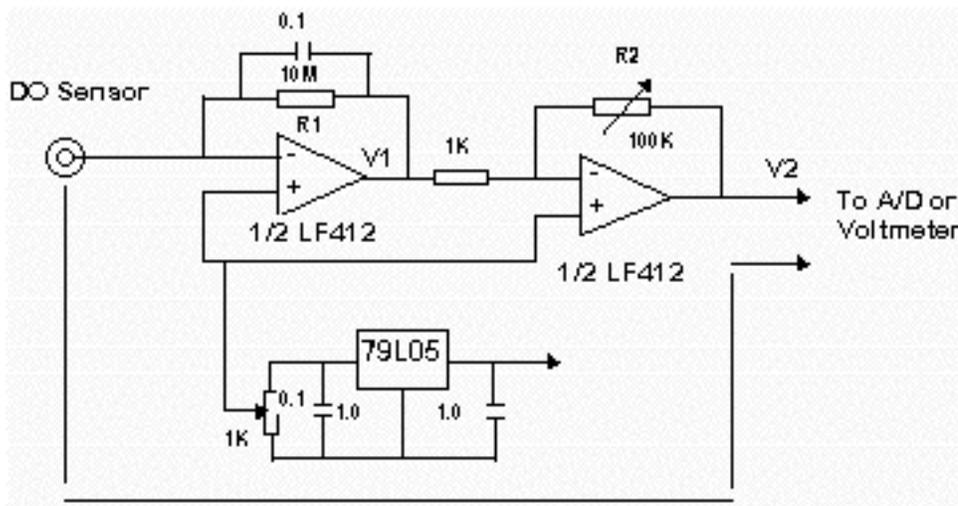

Fig. 2.11. Signal amplifier circuit for H_2O_2 measurement.

2.4. Ion Selective Electrodes and Potentiometry

2.4.1. Definition of pH

Definition of pH   Acids may be defined as proton donors and bases as proton acceptors:

$$HA = H^+ + A^- \qquad\qquad (28)$$
$$\text{(acid) (proton) (base)}$$

The dissociation of weak acids and bases are equilibrium processes, and the equilibrium law can be applied to them:

$$K = \frac{[H^+][A^-]}{[HA]} \qquad\qquad (29)$$

where K is the dissociation constant. For water,

$$H_2O \rightleftharpoons H^+ + OH^-$$

and

$$K = \frac{[H^+][OH^-]}{[H_2O]} \qquad\qquad (30)$$

K$_w$ for Water   Since only a small fraction of water is dissociated, [H$_2$0] may be considered constant:

$$[H^+][OH^-] = K_w \qquad\qquad (31)$$

Conductivity measurements have shown that K_w is 1×10^{-14} mol^2/L^2. 'The acidity or alkalinity of a solution can be measured by its hydrogen ion concentration but it is more convenient to use pH defined by:

$$pH = -\log[H+] \qquad\qquad (32)$$

Therefore, from Eq. (31),

$$pH + pOH = 14 \qquad\qquad (33)$$

When pH = pOH = 7, the solution is said to be in 'neutral' pH.

K$_a$ of Weak Acid   From Eq. (29), pH of a weak acid can be described as:

$$pH = pK_a + \log([A^-]/[HA]) \qquad\qquad (34)$$

Therefore, if K_a of an acid is known and pH is measured, the concentration of A$^-$ ion can be calculated when the concentration of HA is known.

2.4.2. Measurement of pH

Glass type pH Electrode   Usually, pH is measured by a glass electrode type pH sensor. The glass electrode consists of a glass membrane which is selectively permeable only to H+ and an internal Ag/AgCl electrode immersed in 0.1 M KCl solution (Fig. 2.15). When this glass electrode is immersed in a test solution of unknown pH and the voltage is measured against a standard electrode (such as a Calomel electrode; Fig. 2.16), the measured voltage follows Nernst equation:

$$E_m = \varphi_o - \varphi_i = \frac{RT}{zF} \ln \frac{C_o}{C_i} \qquad (35)$$

where E_m is the difference in potential between the inside (φ_i) and outside (φ_o) of the membrane measured by the sensor, R is the gas law constant (8.31 J/mol-K), T is the absolute temperature, z is the valance of the ion, F is Faraday constant (96,500 coul/mol; note that 1 coul = 1 J/V), and C_i and C_o are the concentrations of H^+ ions inside and outside of the membrane, respectively. The C's should be activities but for dilute solutions, molar concentration can be used. In recent pH sensors, the glass electrode and the reference electrode are combined in one body (Fig. 2.16b). Note that the voltage output (E_m) is proportional to the log of the concentration (log C_o), not the concentration itself (C_o) (see Fig. 2.17).

Measurement Circuit   From Eq. (35), a decade change in C_o will give E_m of 58.1 mV for H^+ at 20°C. This voltage output has to be amplified if one desires to take data, for example, with a computer. A circuit shown in Fig. 2.18 can be used for such an amplification. It uses one operational amplifier in a non-inverting mode. The operational amplifier has to be a FET-input type that has an input resistance on the order of 10^{12} ohms. This is because the resistance of the glass pH electrode is close to 5×10^9 ohms. To make voltage measurement from such a high resistance source, the input impedance of the amplifier has to be much greater. The other point is that only the non-inverting configuration (of the operational amplifier)

Ag/AgCl
internal
reference

Internal
filling solution
(0.1 M HCl)
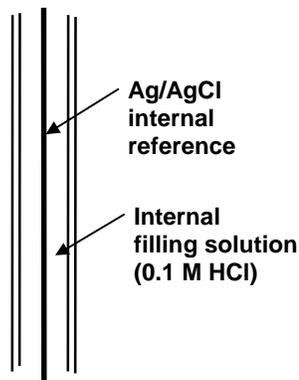
$$E_m = \varphi_o - \varphi_i = \frac{RT}{zF} \ln \frac{C_o}{C_i} \; (Nernst \; Eq.)$$

where R = 8.31 J/mole-K
F = 96,500 coul/mol
1 coul = 1 J/V
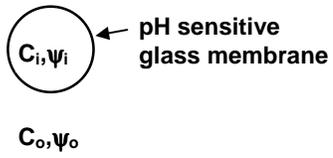z = valence of ion

For H^+ at 20°C, 58.1 mV/decade of C^o

pH sensitive glass membrane

C_i, ψ_i

C_o, ψ_o

Fig. 2.15. Glass membrane electrode for pH measurement.
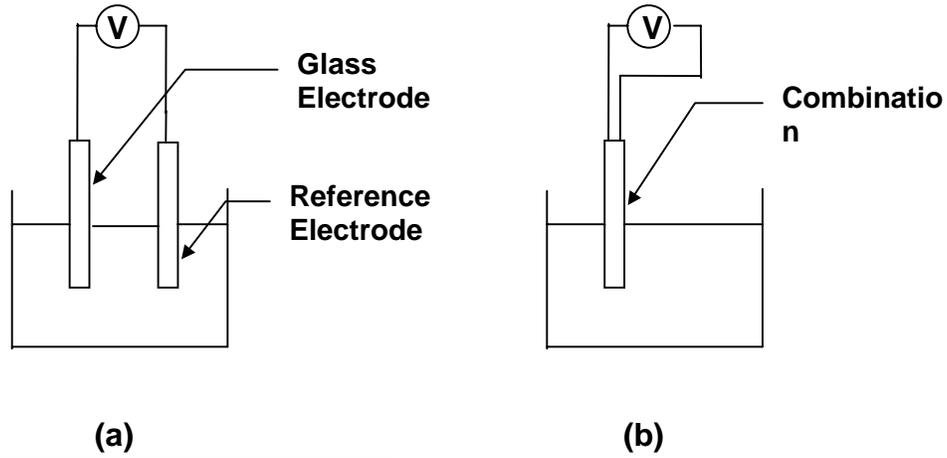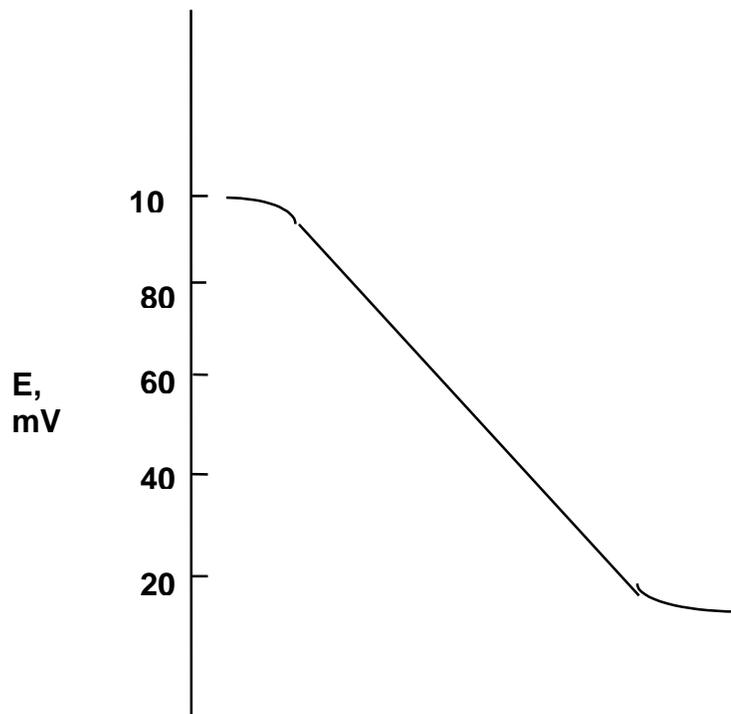

Glass Electrode

Reference Electrode

Combination

(a)

(b)

Fig. 2.16. Measurement setup for pH: (a) separate glass electrode and reference electrode; (b) a combination electrode.

E, mV

10

80

60

40

20

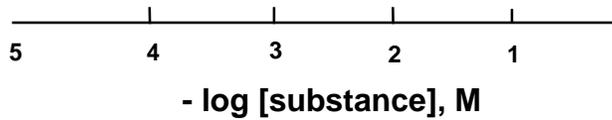|				
5	4	3	2	1

- log [substance], M

Fig. 2.17. Response of potentiometric sensor to variations in ion concentration.
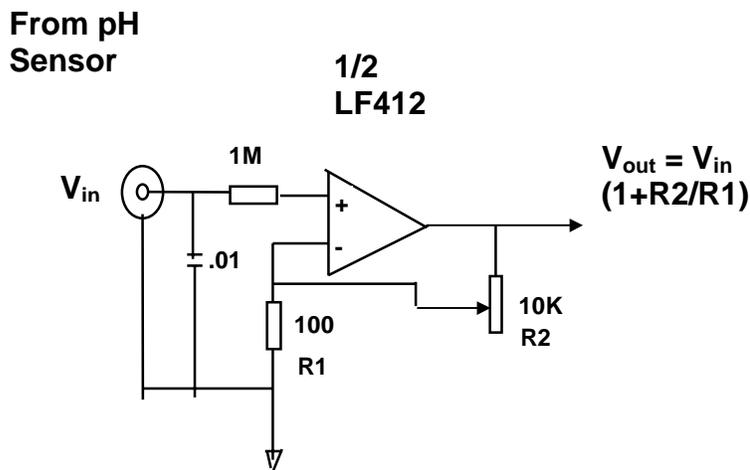

Fig. 2.18. Signal conditioning circuit for potentiometry.
has to be used to take full advantage of the input resistance of the operational amplifier.

2.4.3. Other Ion-Selective Electrodes

pH measuring glass electrode is just one example of ion-selective electrodes (ISE). There are many different ISE's but it would be convenient to classify them in terms of the membrane used for partitioning.

Glass Membrane These originate with the hydrogen-ion selective electrode having well-behaving glass membrane that has high mobility for H^+ ion. Subsequently, electrodes have been developed for other cations such as Na^+, K^+, and NH_4^+

Inorganic Salt Membrane These electrodes are based on inorganic halides and sulphides, for example, silver salts, lanthanum fluoride and heavy-metal sulphides. These membranes have been produced from preparations ranging from whole crystals to dispersions in an inert matrix, such as polythene or silicon rubber. They are targeted at ions such as halides, CN^-, S^-, Ag^+, Cu^{++}, and Pb^{++}.

Organic Membrane The electrodes based on neutral carriers generally have the highest selectivity in this class. However, cation exchangers or complexing agents or anion exchangers have been successfully employed in electrodes with liquid or solid membranes, selective to cations or anions, respectively.

Gas Sensing Electrode These electrodes are an extension of ion-selective measurements to detection of gaseous analytes. Gas-sensing probes are complete electrochemical cells, incorporating both the ion-selective electrode and a reference electrode within the sensor. Assay of the target gaseous sample is not performed directly, but is related to a changing parameter (usually pH) which can be monitored by an ion-selective electrode.

Problems

1. Calculate the steady state current output from the DO sensor that we have made in the lecture when it is placed in air saturated water. Assume that we used 25 μm thick Teflon FEP membrane. Use the table given below for the permeability data.

Properties of Teflon membrane

Membrane	P_m	D_m
Teflon FEP	4.4 x 10^{-10} cc/s-cm-cmHg	1.07x10^{-7} cm^2/s
Silicone	480 x 10^{-10} cc/s-cm-cmHg	-

2. Suppose we use two 25 μm Teflon membranes to cover the cathode.
(a) What will be the output current under the same condition as in Problem 1?
(b) Comment on the response time. Will this sensor become slower? By how much? Take diffusivity of oxygen in water as 2x10^{-5} cm^2/s.

3. Suppose you place the DO sensor in a stagnant liquid which has a magnetic stirrer in it. Qualitatively show the sensor output current as you increase the stirring rate.

4. Flow dependency of DO sensor is a big problem in actual measurement. The flow dependency can be reduced by placing a silicone membrane over the Teflon membrane. Explain why this reduces the flow sensitivity.

5. Suppose we measure DO concentration in (a) dense aerobic culture, and (b) low cell density aerobic culture. In which case the accuracy will be better. Explain why. What can you do to improve the accuracy?
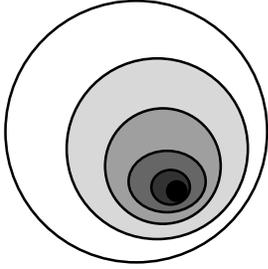
Chapter 3. Bioreceptor Molecules

3.1. Enzymes

3.1.1. Enzyme in Biomolecular Ladder

Biomolecular Ladder   Most of the chemical components of living organisms are organic compounds of carbon, many also containing oxygen and nitrogen. Although each living species contains various combination of these biomolecules, the diversity can be reduced to a few building blocks of common structure. It is possible to organize these building blocks on a hierarchical ladder, according to molecular weight (Fig. 3.1). The bottom of' the ladder is occupied by the low-molecular-weight gases (oxygen, nitrogen and carbon dioxide) and by water. These molecules together with the monoatomic ions, in particular Na^+, K^+, Mg^{2+}, Ca^{2+}, Cl^- and the elements P, S, Mn, Fe, Co, Cu and Zn are most common in further involvement as progress is made up each rung of the molecular-weight ladder. The monitoring of these ions is frequently used to follow the metabolic state of a patient under care and is commonly achieved using ion selective electrode sensors, where the recognition surface of the sensor is provided by a membrane, containing a molecule with a selectivity for the target ion (see Chapter 2) . The oxygen and carbon dioxide electrodes f'orm two of the well-established base sensors to which analyte specific reactions have been linked via a biorecognition macromolecule. Note where 'protein', 'antibody', and 'enzyme complex' lie in the ladder. These are the major molecules with *molecular-recognition capabilities* that are utilized in biosensor development.

Metabolism Point of View   The food that we eat consists mainly of proteins, carboydrates (or polysaccharides), and lipids. These nutrients are broken down in our body with the aid of oxygen that we breathe. This breakdown process called 'catabolism' produces the basic building blocks of the body - amino acids, simple sugars, and fatty acids. Also, energy is produced (and stored in the form of ATP) during the catabolism which is essentially a partial oxidation process. These building blocks and energy are then used to make the main components of our body - proteins, polysaccharides, and lipids. These components are used to make supramolecular assemblies such as enzyme complexes, ribosomes (this is where proteins are made), etc. The synthesis processes are collectively called the 'anabolism' which is dictated by the genetic codes written on the genes of the cell (the DNA helix).

Protein Structure   Proteins are polymers of amino acids. All amino acids have a common structure - a carbon molecule with a carboxyl group (-COOH), an amino group   ($-NH_2$), and R group (Fig. 3.2a). It is the R group which
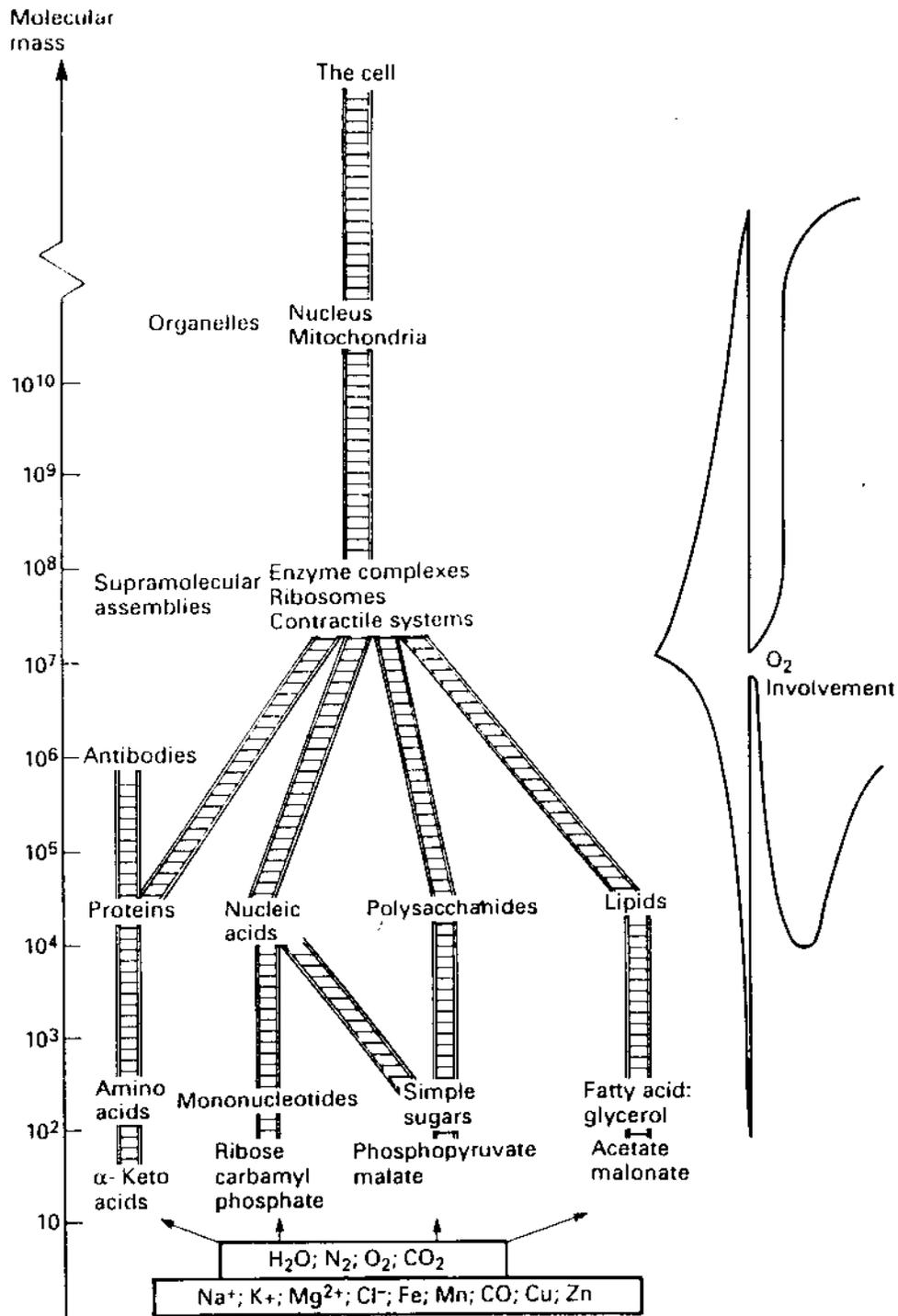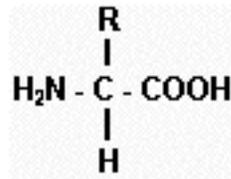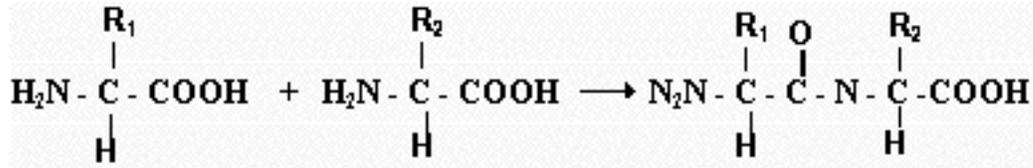
Fig. 3.1. Biomolecular ladder.

(a) Basic structure of amino acid

$$\begin{array}{c} R \\ | \\ H_2N - C - COOH \\ | \\ H \end{array}$$

(b) Formation of peptide bond

$$H_2N - \underset{\underset{H}{|}}{\overset{\overset{R_1}{|}}{C}} - COOH \;+\; H_2N - \underset{\underset{H}{|}}{\overset{\overset{R_2}{|}}{C}} - COOH \;\longrightarrow\; N_2N - \underset{\underset{H}{|}}{\overset{\overset{R_1}{|}}{C}} - \overset{\overset{O}{\|}}{C} - N - \underset{\underset{H}{|}}{\overset{\overset{R_2}{|}}{C}} - COOH$$

(c) Limited rotation in peptide chain

Peptide bond | Overlap between lone pairs on oxygen and nitrogen | Planar amide group | Free rotation about these bonds

(d) non-covalent inter- and intra-molecular bonds in peptide strands.

Interpeptide   Side chain   IONIC BOND   HYDROPHOBIC BOND
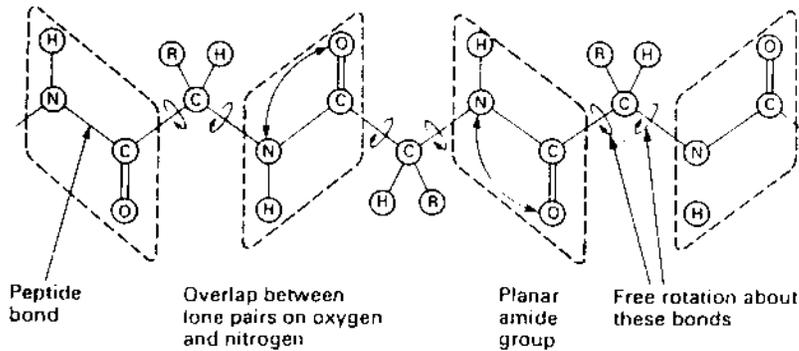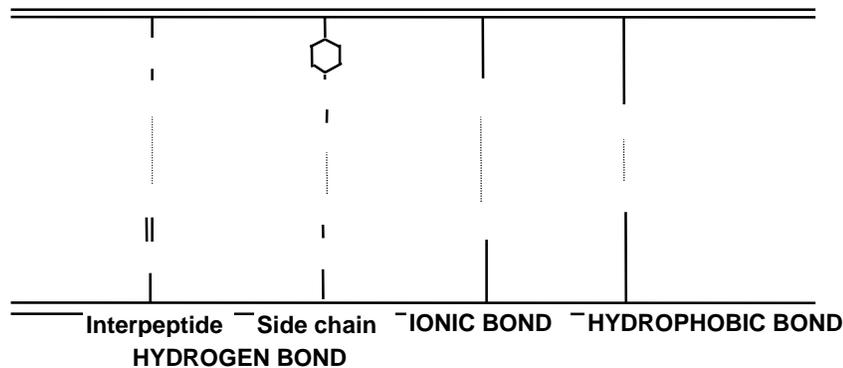HYDROGEN BOND

Fig. 3.2. (a) Basic structure of amino acid; (2) formation of peptide bond; (3) limited rotation in the peptide strand; (4) non-covalent inter- and intra-molecular bonds in peptide strands.

makes each amino acid unique. For example when R = CH_3, it is called alanine. The amino acids are connected together by a peptide bond, which is the bond between -COOH of one amino acid and $-NH_2$ of the other (Fig.3.2b). The proteins are made in the riobosomes (an organel made of ribosomal RNAs) of the cells. The strucutre and the function of the proteins depend not only on the

amino acicd sequence (called the primary structure), but also on the conformation (the secondary, tertiary, and quaternary structures). The peptide bond and the disulphide bridges impart certain restrictions on the structure (Fig. 3.2c). The peptide strands are further organized by interactions between residue side chains. The nature of the bonds include hydrogen bond, ionic bond, and hydrophobic bond (Fig. 3.2d). An example of an enzyme is shown in Fig. 3.3.

Ionic Property of Proteins   The net result of all the interactions between the amino acids is that there is a spontaneous folding of a protein to give a unique structure. All the amillo acids have at least two groups capable of existing in ionic form. The α-(carboxyl groul), -COOH, can lose H^+ to become COO^- . The reaction is pH dependent, and is characterized by a pK_a typically in the range 2 to 3. Similarly the α-amino group, NH_2, can be protonated to give NH_3^+ and has a pK_a value of about 10. Therefore, between about pH 4 and 9 the amino acid exists as a dipolar ion *zwitterion* with little net charge (Fig. 3.3d). At the isoelectric point, pI, the protein has no net charge, and it will not move in an electric field.

Ioselectrical Point   Where R contains no ionizable groups,

$$pI = \frac{(pK_a^{NH2} + pK_a^{COOH})}{2} \tag{1}$$

The movement of the amino acids under the influence of an electric field allows their separation and identification; it can be used as a powerful assay technique. The dual polarity feature accounts for many of the properties of amino acids, e.g. the large dipole moments, the high solubility in water and low solubility in organic solvents

Ionic Behavior of Proteins   By analogy, it would be expected that each peptide strand would also contain at least two ionizable groups, but since the α-carboxyl or α-amino groups are now involved in peptide-bond formation, they are not available for ionization. The zwitterionic behaviour is therefore more restricted to the terminal amino group and the terminal carboxyl group. These groups are considerably further away from one another than they would be in free amino acids and so the electrostatic interactions between them are diminished and their pK_a values are lower than in the α-amino acid. It follows that the groups in proteins

Fig. 3.3. An examples of protein molecule: hen-egg lysozyme (the abbreviations are for amino acids).

Uracil base of substrate

ENZYME

Fig. 3.4. Eamples of the interaction between enzyme surface and incoming substrate.

that are principally involved in acid-base equilibria are the side-chain R groups. Ordinarily there may be 50-60 titratable groups per 100,000 molecular weight of protein. Therefore, the titration curves for proteins are complex and difticult to interpret.

Molecular Recognition by Enzyme   *The biorecognition properties of protein molecules will depend almost entirely on the amino acids of the exposed surfaces.* Weak non-covalent interactions can occur between the residues on the exposed surfaces of the protein and other non-protein molecules (Fig. 3.4). If a sufficient number of these weak bonds are formed simultaneously with the incoming molecule, then the molecule can bind tightly to the protein. Obviously for this to occur the molecule must fit precisely into the *binding site* on the protein surface. This feature is analogous to the recognition surface of the model biosensor.

Enzyme Complexes   One of the most important functions of proteins is to act as catalysts or enzymes for chemical reactions. These enzymes are able to stabilize the transition state between a substrate and its products by interactions at the binding site (of the substrate). The activation of most biochemical reactions fall in the range of 40-80 kcal/mol without the enzyme. The enzymes lower this activation energy. For example, the splitting of H_2O_2 takes 75.4 kJ/mol without the enzyme, whereas a catalase enzyme lowers the activation energy to 23 kJ/mol. Substrate specificity by the enzyme is provided by the surface interactions and this characterisitcs is exploited in the development of enzylne-based biosellsors. The none-covalent binding of the enzyme substrate transition state lowers the activation energy for the reaction and thus catalyzes the reaction.
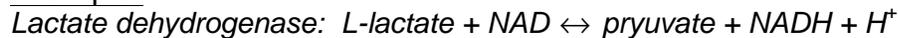
3.1.2. Enzyme Classification

There are approximately 3000 known enzymes. These enzymes are classified into six categories based on the type of reaction they catalyze.

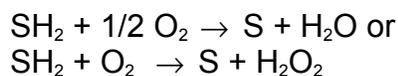**1. Oxido- reductase:**   Oxidizes or reduces by transfer of hydrogen or electrons.
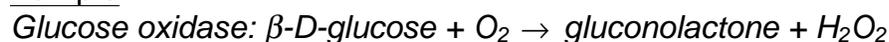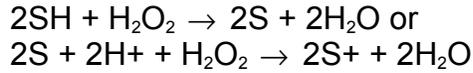
(a) dehydrogenases:

$SH_2 + A \leftrightarrow S + AH_2$ (S: Substrate, A: acceptor)

Example:
Lactate dehydrogenase: L-lactate + NAD $\leftrightarrow$ pryuvate + NADH + H$^+$

(b) oxidases:

$SH_2 + 1/2\ O_2 \rightarrow S + H_2O$ or
$SH_2 + O_2\ \rightarrow S + H_2O_2$

Eample
Glucose oxidase: β-D-glucose + O$_2$ → gluconolactone + H$_2$O$_2$

(c) peroxidases:

$2SH + H_2O_2 \rightarrow 2S + 2H_2O$ or
$2S + 2H+ + H_2O_2 \rightarrow 2S+ + 2H_2O$

Example
Horseradish peroxidase:
$2[Fe(CN)_6]_4- + 2H^+ \rightarrow 2[Fe(CN)_6]_3- + 2H_2O$

(d) oxygenases:

$SH + DH + O_2 \rightarrow S\text{-}OH + D + H_2O$
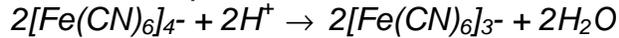
Example
Lactate 2-monooxygenase:
L-lactate + O_2 $\rightarrow$ acetate + CO_2 + H_2O
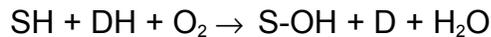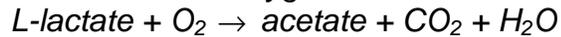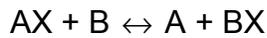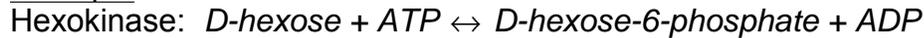
2. Transferase: Transfers C-, N-, P-, or S-containing functional groups such as aldehydes and ketones, glycosils, acyls, phosphates, and sulfur   containing groups.

$AX + B \leftrightarrow A + BX$

Example
Hexokinase: *D-hexose + ATP $\leftrightarrow$ D-hexose-6-phosphate + ADP*

3. Hydrolase: Hydrolyses esters, anhydrides, peptide bonds, other C-N bonds, glycosides

Example
Cholesterol esterase:
cholesterol ester + H_2O $\rightarrow$ cholesterol + fatty acid
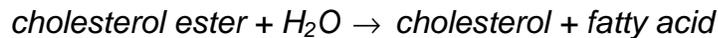
Glucoamylase:
amylose + n H_2O $\rightarrow$ n β-D-glucose
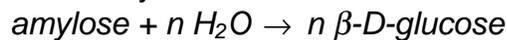
4. Lyase: Adds to double bonds:

$> C = C <$
$> C = O$
$> C = N$
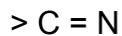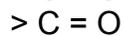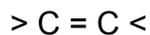
5. Isomerase: Isomerizes optical iomers

Example
Glucose iomerase: *D-glucose $\leftrightarrow$ D-fructose*

6. Ligase: Splits C-C, C-O, C-N, C-S and C-halogen bonds without hydrolysis or oxidation, mostly with ATP

Example
Pyruvate carboxylase:
$pyruvate + HCO_3^- + ATP \leftrightarrow oxaloacetate + ADP + P_i$

3.1.3. Coenzymes, Prosthetic group, Effectors

Sometimes the surface cavity does not act as a catalytic site until it is modified by a second incoming molecule. These participants known as the *coenzymes* are non-peptide molecules capable of completing the binding site for the transition state. Other molecules that do the similar function are prothetic group, and effectors.

Coenzyme Coenzyme is a non-peptide molecule capable of completing the binding site for the transition state. Examples include many vitamin derivative such as coenzyme A, thiamine, pyrophosphate, vitamin B12

Prosthetic Group Prosthetic group is the same as the coenzyme but are tightly bound to the enzyme. When they are split off, the enzyme is mostly denatured. Examples include flavin nucleotides and heme.

Effectors Effectors accelerate (activators) or block (inhibitors) enzyme reaction Examples of activators include Mg^{++}, Ca^{++}, Zn^{++}, K^+, and Na^+, while the examples for the inhibitors include Hg, and substrate analogs.Table 3.1. lists functions of some of the important coenzymes and prostshetic groups.

Table 3.1. Function of some important coenzymes and prosthetic groups.

Compound	Function
(a) Oxidoreduction	
Nicotinamide adenine dinucleotide (NAD)	hydrogen transfer
Nicotinamide adenine dinucleotide phosphate (NADP)	hydrogen transfer
Flavin monocucleotide	hydrogen transfer
Flavin adenine dinucleotide (FAD)	electron transfer
Heme (cytochromes)	
Ferredoxins	electron transfer
(b) Group Transfer	
Pyridoxalphosphate	transamination, decarboxylation
Adenosine triphosphate	phosphate group donor
Tetrahydrofolic acid	C1 group transfer
Biotin	carboxylation, decarboxylation

Coenzyme A	transacylation
Thiamine pyrophosphate (Vitamin B1)	C2-group transfer
Riboflavin	hydrogen transfer
5'-deoxyadenosyl-cobalamine	transfer of methyl group, isomerization

3.2. Enzyme Kinetics

3.2.1. Kinetics of Enzyme-Catalyzed Reactions

Reaction Velocity   The time course of an enzymatic reaction permits one to deduce the substrate affinity, the catalytic mechanism in the active center, and the efficiency of the enzyme (maximum rate, turnover number). The rate of an enzyme-catalyzed single reactant reaction depends on the concentration of substrate and product, respectively. The velocity of the reaction V is:

$$V = - dS/dt = dP/dt \qquad\qquad (2)$$

where first term is the rate of dissapearance of substrate S and second term is the rate of appearance of product P (both S and P are in concentration).

Behavior of Initial Rates   The initial rate (V_o) is determined by extrapolating the slope of the time course of the substrate or product concentration to time zero (Fig. 3.5). The dependence of V_o on the substrate concentration, S (at constant enzyme concentration), is shown in Fig. 3.6. It reflects the typical substrate saturation. At first, V_o increases proportionally to the amount of substrate. Upon further enhancement of substrate concentration V_o levels off. The curve asymptotically approaches a maximum value, V_{max}. When this plateau is reached, a change of S does not lead to a measurable change of V_o: the enzyme is saturated by substrate and has thus reached the limit of its efficiency.

Micahaelis- Menten Kinetics These kinetics result from the fast and reversible formation of an enzyme-substrate complex, ES, which dissociates in a second,
slower reaction under liberation of the product, P (Fig. 3.7):

$$E + S \qquad\qquad ES \qquad\qquad E + P \qquad (3)$$

Because the second reaction is rate-limiting, at very high substrate concentration almost all enzyme is present as enzyme-substrate com-plex. Under these conditions a steady state is reached in which the enzyme is steadily saturated by substrate and the initial rate is at a maximum (V_{max}). This relation between

substrate concentration and reaction rate may be described by the **Michaelis-Menten equation**:

$$V_o = \frac{V_{max}S}{K_M + S} \tag{4}$$

where K_M is the Michaelis constant of the enzyme for the given sub-strate. K_M may also be described by:

$$K_M = \frac{k_{-1} + k_{+2}}{k_{+1}} \tag{5}$$

Meaning of K_M   The relevance of K_M becomes evident at $S = K_M$. Then $V_o = V_{max}/2$, i.e., K_M is the substrate concentration at which the reaction rate is half maximum (Fig. 3.6). The K_M value characterizes the affinity between the substrate and the enzyme. At known K_M and V_{max}, V_o can be calculated for each value of substrate concentration. A low K_M value reflects high affinity. At substrate concentrations $S \ll K_M$, the reaction rate is directly proportional to the substrate concentration (first order reaction); at high substrate concentration ($S \gg K_M$) the reaction is zero order and is no longer dependent on the substrate concentration but only on the enzyme activity.
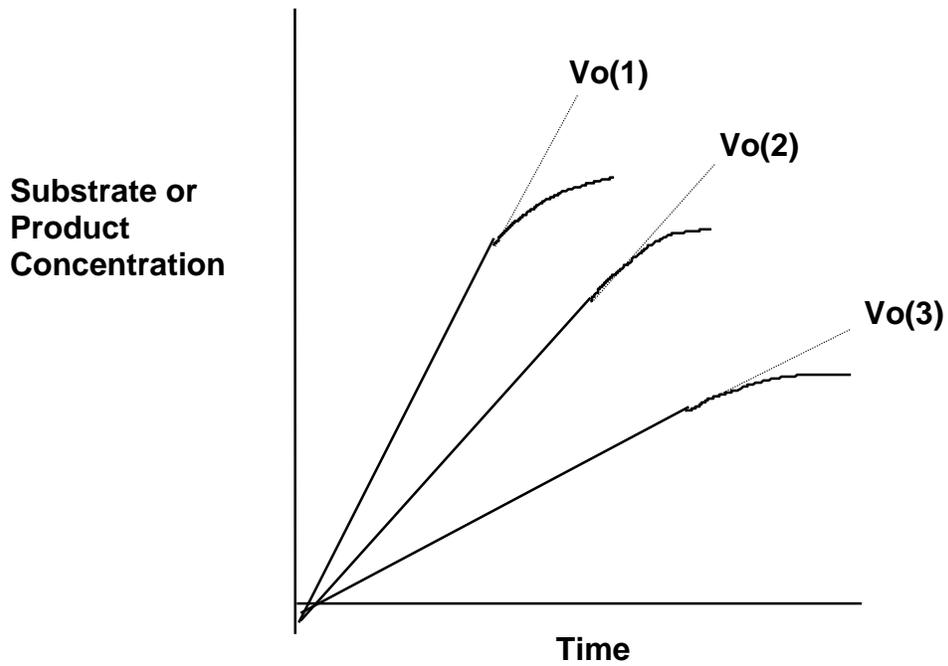
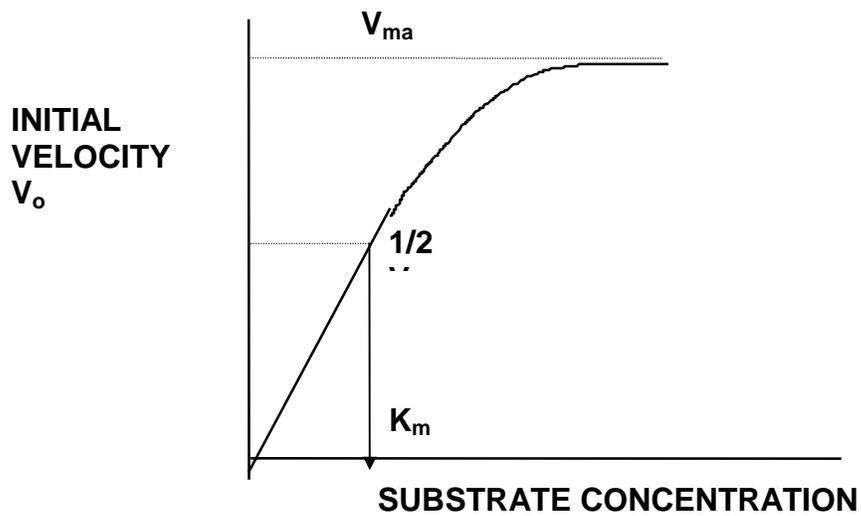**Fig. 3.5. Determination of initial rates at different substrate concentrations.**

Fig. 3.6. A plot of V_o vs. substrate concentration S.

Enzyme Kinetics

$$k_{+1}$$
$$\text{(Step 1)} \quad E + S \leftrightarrow E\text{-}S$$
$$k_{-1}$$

$$k_2$$
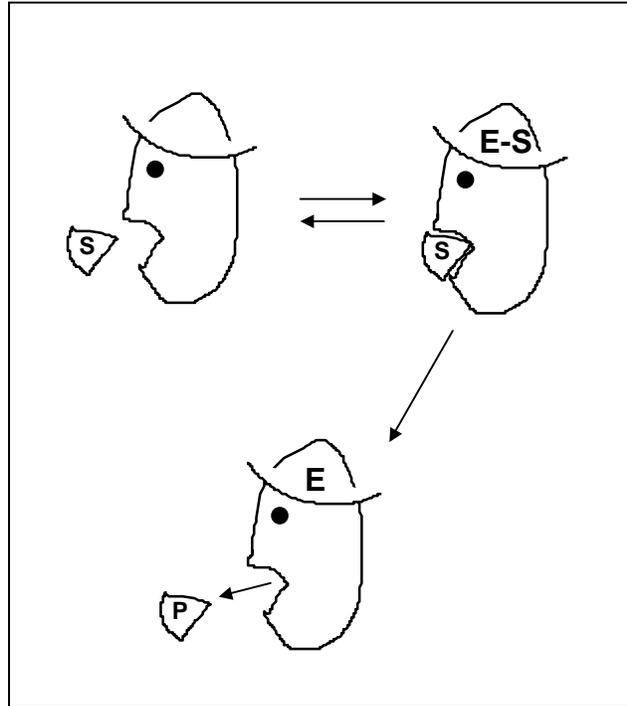$$\text{(Step 2)} \quad E\text{-}S \rightarrow P$$

Rate of reaction:

$$V = -\frac{d[S]}{dt} = \frac{d[P]}{dt} = k_2[ES]$$

Net formation rate of ES:

$$\frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_2[ES]$$

During reaction, the # of active sites occupied by S is constant:

$$\frac{d[ES]}{dt} = 0$$

Total enzyme sites = occupied sites + free sites:

$$[E_o] = [ES] + [E] \quad \text{or} \quad [E] = [E_o] - [ES]$$

Therefore,

$$[ES] = \frac{[E_o][S]}{(k_{-1} + k_2)/k_1 + [S]} = \frac{[E_o][S]}{K_M + [S]}$$

$$\boxed{V = -\frac{d[S]}{dt} = k_2[ES] = \frac{k_2[E_o][S]}{K_M + [S]} = \frac{V_{max}[S]}{K_M + [S]}}$$

Fig. 3.17. Enzyme kinetics.

Line-Weaver- Burk Plot   To calculate K_M and V_{max} (and inhibitor constants) it is advantageous to transform the Michaelis-Menten relation so as to obtain linear relationships between S and V_o that can be evaluated graphically. An

example is the Line-Weaver-Burk equation, containing the reciprocal values of V_o and S:

$$\frac{1}{V_o} = \frac{1}{V_{max}}\left(1 + \frac{K_M}{S}\right) \qquad (6)$$

An example of Line-Weaber-Burk plot is shown in Fig. 3.8.

3.2.2. Enzyme Activity and Enzyme Concentration

Turnover Number   In addition to K_M, and V_{max}, the *turnover number* (molar activity) and the *specific activity* are important parameters for the characterization of enzyme reactions. Both are determined under substrate saturation. With highly purified enzymes the **turnover number** reflects the number of substrate molecules converted in unit time by a single enzyme molecule (or a single active center). Catalase, one of the most potent enzymes, has a turnover number of 2×10^5/s.

Specific Activity   The specific activity of enzymes is given in units. One international unit (IU) is the amount of enzyme consuming or forming 1 μmol substrate or 1 μmol product per minute under standard conditions. The base unit is 1 katal, corresponding to the amount of enzyme converting 1 mol substrate per second:

$$1 \text{ kat} = 6 \times 10^7 \text{ IU,}$$

$$1 \text{ IU} = 16.67 \text{ nkat.}$$

Usually U is used instead of IU. For the quantitative determination of enzyme activity, initial rates are measured at different enzyme concentrations and near substrate saturation, in a suitable temperature range (25-37°C) and at optimal pH. In a certain range the enzyme activity is proportional to the enzyme concentration. The enzyme activity of a sample can be estimated from the linear part of the plot.
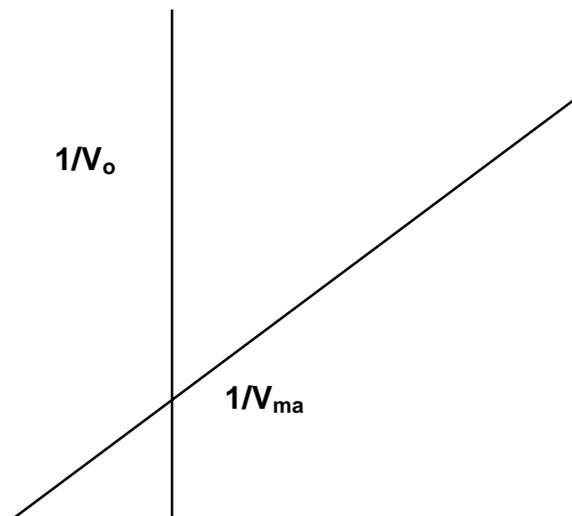
←

—————————————————————————

-1/K$_M$ **1/S**

—————————————————————————

Fig. 3.8. Line-Weaver-Burk plot.

(a)
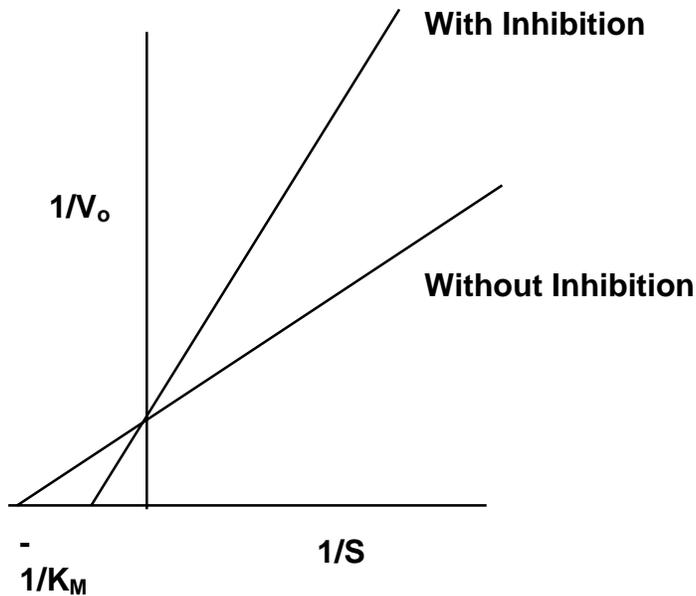
(b)
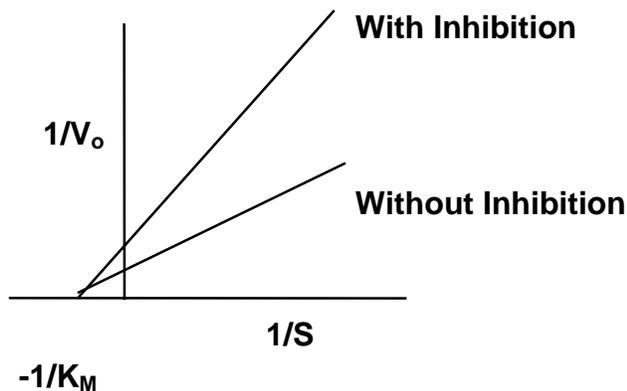
Fig. 3.9. (a) Competitive inhibition; (b) non-competitive inhibition

3.2.3. pH and Temperature Dependence

pH Effect   Each enzyme has a characteristic pH optimum at which its activity is at a maximum. In the range of this optimum essential proton-donating or proton-accepting groups in the active center of the enzyme are in the ionized state required for the enzyme to function. Outside this range, substrate binding is no longer possible, and at extreme pH values the enzyme may be irreversibly denatured. The pH optimum depends on the composition of the medium, the temperature, and the enzyme's stability in acid and alkaline environments. The

pH stability does not necessarily coincide with the pH optimum of the reaction rate.

Temperature Effect As with all chemical reaction rates, those of enzyme reactions increase with increasing temperature (by a factor of 1.4 - 2.0 per 10 K), a limit being set by the thermal stability of the protein. The optimum temperature may be in a wide range, roughly between 30 and 80°C.

3.2.4. Inhibition of Enzyme Reactions

Types of Inhibition   The function of enzyme-based biosensors may be severely restricted by inhibitors. The inhibition is either reversible or results in an irre-versible inactivation of the enzyme.

Competitive Inhibition   Inhibitors structurally related to the substrate may be bound to the enzyme active center and compete with the substrate (competitive inhibition).

Non- competitive Inhibition   If the inhibitor is not only bound to the enzyme but also to the enzyme-substrate complex, the active center is usually deformed and
its function is thus impaired; in this case the substrate and the inhibitor do not compete with each other (noncompetitive inhibition). Competitive and noncompetitive inhibition effect the enzyme kinetics differently. A competitive inhibitor does not change V_{max} but increases K_M (Fig. 3.9a) in contrast, noncompetitive inhibition results in an unchanged K_M and an increased V_{max} (Fig. 3.9b). Some enzymes, e.g. invertase, are inhibited by high product concentration (product inhibition).

3.3. Antibodies (= Immunoglobulins)

3.3.1. Antibody Is Immunoglubulin

Immuno- globulin   Antibodies (Ab) are high-molecular weight (140,000 - 1000,000) soluble proteins (immunoglobulins) produced by organisms in response to foreign substances, antigens (Ag), with whom they form immunochemical complexes:

$$Ab \ + \ Ag \ \rightleftharpoons \ Ab\text{-}Ag$$
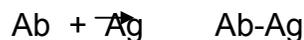
Production of Antibody   The immune system consists of B cells and T cells in the blood serum. The B cells produces antibodies and when antibody graps onto antigen (the foreign invader), T cells destroy the antigen. Each antibody-producing B cell is programmed to make just one antibody, which is placed on its surface as an antigen receptor. Each B cell has a different antigen binding specificity (1-n in Fig. 3.10). Antigen binds to only those B cells with the appropriate surface receptor. These cells are stimulated to profilereate and mature into antibody-producing cells and the longer-lived memory cells, all with the same antigen binding specificity (2 in Fig. 3.10).

Structure of Antibody   The basic structure of all immunoglobulin molecules is a unit consisting of two identical light polypeptide chains and two identical heavy polypeptide chains linked together by disulphide bonds (Fig. 3.11). The class and

subclass of an immunoglobulin molecule is determined by its heavy chain type. Thus the four human IgG subclasses (IgG1, IgG2, IgG3 and IgG4) have heavy chains called y1, y2, y3, and y4 which differ only slightly although all are recognizably y heavy chains. The differences between the various subclasses within an immunoglobulin class are less than the differences between the different classes; thus IgG1 is more closely related to IgG2, 3, or 4 than to IgA, IgM, IgD or IgE.

Subclasses of Antibodies The four subclasses of human IgG occur in the approxi-mate proportions of 66, 23, 7 and 4 per cent respectively. There are also known to be subclasses of human IgA (IgA1 and IgA2) but none have been unambiguously described for the other classes. Immunoglobulin subclasses appear to have arisen after speciation and the human subclasses cannot be compared with, for example, the four known subclasses of IgG which have been identified in the mouse.
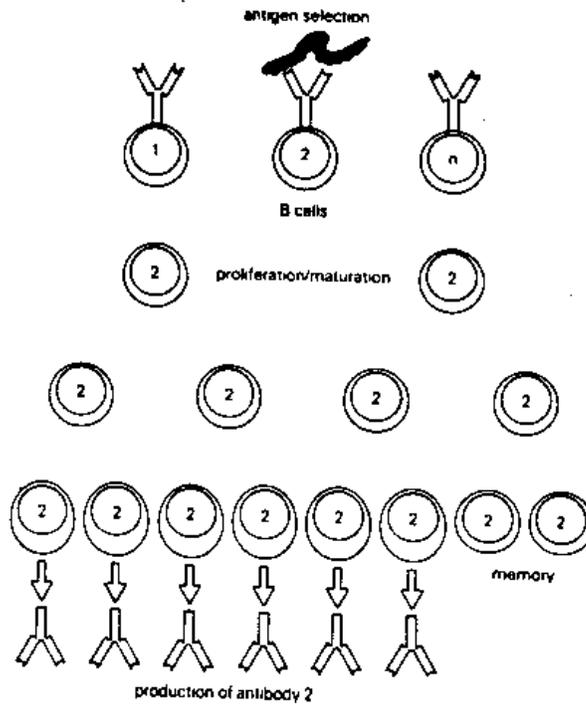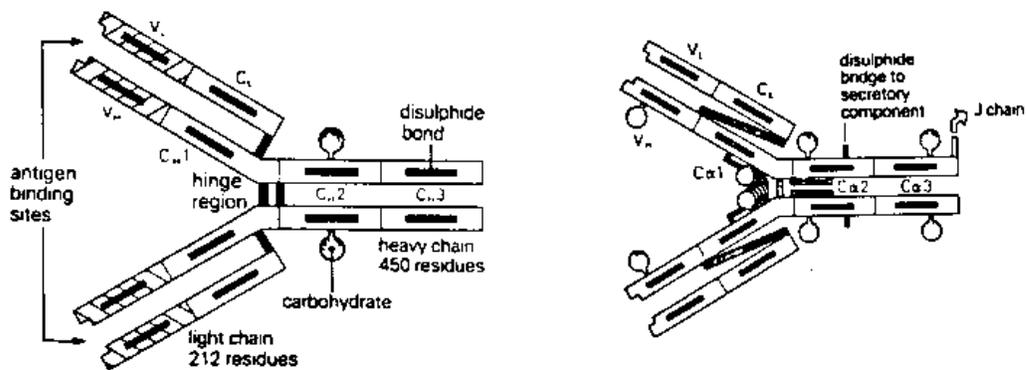
Fig. 3.10. Production of antibodies by B cells.

Fig. 3.11. Structures of antibodies: (a) IgG; (b) IgA1

3.2.2. Properties of Immunoglobulins

General   All immunoglobulins appear to be glycoproteins but the carbohydrate content ranges from 2-3% for IgG to 12-14% for IgM, IgD and IgE. The physicochemical proper-ties of the immunoglobulins are summarized in Table 3.2. Each class possesses a characteristic type of heavy chain. Thus IgG possesses γ chains; IgM, μ chains; IgA, α chains, igD, δ chains and IgE, ε chains. Variation in heavy chain structure within a class gives rise to immunoglobulin subclasses. For example the human IgG pool consists of four subclasses reflecting four distinct types of γ heavy chain. The physicochemical properties of the immunoglobulins vary between the different classes. Note that IgA occurs in a dimeric form {sigA) in association with a protein chain termed the secretory piece. The diversity of structure of the different classes suggests that they perform different functions, in addition to their primary function of antigen binding. In spite of this diver-sity all antibodies have a common basic structure. The structures of IgG, and IgA1 are shown in Fig. 3.11.

IgG   IgG is the major immunoglobulin in normal human serum accounting for 70-75% of the total immuno-globulin pool. IgG is a monomeric protein with a sedi-mentation coefficient of 7S and a molecular weight of 146,000. However, studies of IgG subclasses have indicated that IgG3 proteins are slightly larger than the other subclasses and this increase is due to the slightly heavier y3 chain. The IgG class is distributed evenly between the intra- and extravascular pools, is the major antibody of secondary immune responses and the exclusive anti-toxin class.

IgM   IgM accounts for about 10% of the immunoglobulin pool. The molecule has a pentameric structure in which individual heavy chains have a molecular weight of approximately 65,000 and the whole molecule has a molecular weight of 970,000. This protein is largely confined to the intravascular pool and is the predominant 'early' antibody frequently directed against antig,enically complex infectious organisms.

IgA   IgA represents 15-20% of the human serum immuno-globulin pool . In man more than 80% of IgA occurs as the basic four chain monomer but in most mammals the IgA in serum is mainly polymeric, occurring mostly as a dimer. IgA is the predominant immunoglobulin in sero-mucous secretions such as saliva, tracheobronchial secretions, colostrum, milk and genito-urinary secretions. Secretory IgA (sigA) which may be of either subclass, exists mainly in the 1 15, dimeric form and has a molecu-lar weight of 385,000. sigA is abundant in

Table 3.2. Physicochemical properties of human immunoglobulins.

Immuno-globulin	IgG1	IgG2	IgG3	IgG4	IgM	IgA1	IgA2	sigA	IgD	IgE
name of heavy chain	γ_1	γ_1	γ_1	γ_1	μ	α_1	α_1	α_1 or α_2	δ	ϵ
concentrati-on, mg/mL	9	3	1	0.5	1.5	3	0.5	0.05	0.03	0.00005
sediment-aton constant	7S	7S	7S	7S	19S	7S	7S	11S	7S	8S
molecular weight, x1000	146	146	170	146	970	160	160	385	184	188
MW of heavy chain, x1000	51	51	60	51	65	56	52	52-56	69.7	72.5
# of heavy chain domains	4	4	4	4	5	4	4	4	4	5
carbohyd-rates, %	2-3	2-3	2-3	2-3	12	7-11	7-11	7-11	9-14	12

seromucous secretions and is protected from proteolysis by combina-tion with another protein - the secretory component.

IgD   IgD accounts for less than 1% of the total plasma immunoglobulin but it is known to be present in large quantities on the membrane of many circulating B lymphocytes. The precise biological function of this class is not known but it may play a role-in antigen-triggered lymphocyte differentiation.

IgE   IgE though a trace serum protein, is found on the surface membrane of basophils and mast cells in all individuals. This class may play a role in active immunity to helminthic parasites but in Western countries is more commonly associated with immediate hypersensitivity diseases such as asthma and hayfever.

3.3.3. Antigen-Antibody Interaction

Lock and Key Type Binding   Foreign molecules that generate antibodies are called antigens. Antigen molecules each have a set of antigenic determinants also called **epitopes** (Fig. 3.12). The epitopes on one antigen (Ag1 in Fig. 3.12) are usually different from those on another (Ag2 in Fig. 3.12)). Some antigens (Ag3 in Fig. 3.12) have repeated epitopes. Epitopes are molecular shapes

recognized by the antibodies and cells of the adaptive immune system. Each cell recognizes one epitope rather than the whole antigen. Even simple microorganisms have many different antigens.

Nature of Binding Forces   There are several types of intermolecular attractive forces binding antigen to antibody. These forces require the close approach of
the interacting groups. **Hydrogen bonding** results from the formation of hydrogen bridges between appropriate atoms; **electrostatic forces** are due to the attraction of oppositely charged groups located on two protein side chains. **Van der Waals bonds** are generated by the interaction between electron clouds (here represented as induced oscillating dipoles) and **hydrophobic bonds** (which may contribute up to half the total strength of the antigen-antibody bond) rely upon the association of non-polar, hydrophobic groups so that contact with water molecules is minimized. The distance of separation between the interacting groups which produces optimum binding varies for the different types of bond (see Fig. 3.13).

3.2.4. Antibody Affinity

Nature of Affinity   The affinity with which antibody binds antigen results from a balance between the attractive and repulsive forces. A high affinity antibody impies a good fit and conversely, a low affinity antibody implies a poor fit (see Fig. 3.14).

Affinity Constant, K_a   All antigen-antibody reactions are reversible and the   equilibrium constant K is called, the affinity constant (Fig. 3.15):
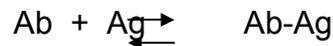
$$Ab \ + \ Ag \ \rightleftharpoons \ \ Ab\text{-}Ag$$

$$K_a = \frac{[Ag - Ab]}{[Ag][Ab]} \tag{7}$$

K_a ranges from 10^4 - 10^{12} liters/mole. Immunoglobulins with $K_a < 10^4$ for a particular antigen is ineffective.
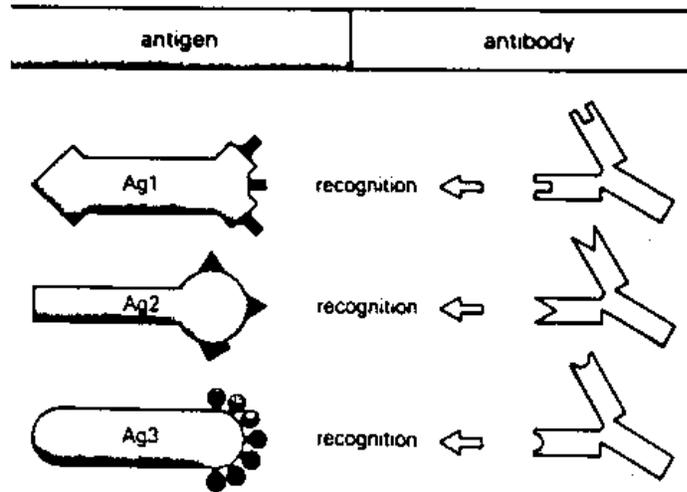
Fig. 3.12. Nature of antibody-antigen binding.
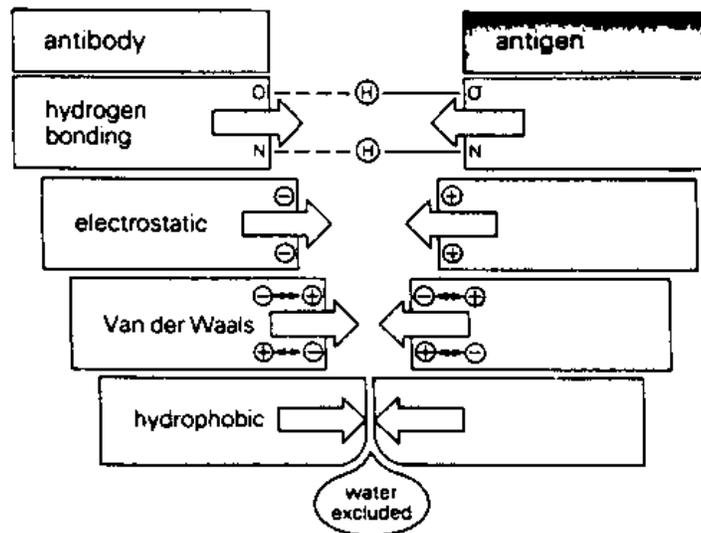
Fig. 3.13. Types of intermolecular attractive forces between antibody and antigen.

Affinity = Σ attractive and repulsive forces

| high affinity – high attraction | low affinity – low attraction |
| GOOD FIT | POOR FIT |
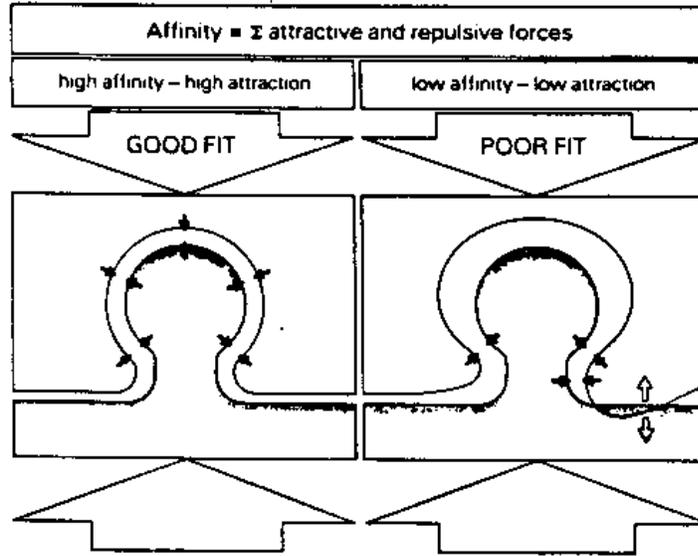
Fig. 3.14. Antibody affinity.

Definition of Antibody Affinity, K_a

antigen–antibody reactions are reversible

equilibrium constant or affinity, K, is given by

$$Ab + Ag \rightleftharpoons AbAg$$

$$K = \frac{[AbAg]}{[Ab][Ag]}$$

Fig. 3.15. Definition of antibody affinity, K_a.

3.3.5. Experimental Determination of K_a (Homogeneous Case)

Suppose we start with an antibody concentration of $[Ab]_o$ and vary antigen concentration $[Ag]$. At equilibrium, the antigen concentration will be:

$$[Ab]_e = [Ab]_o - [AbAg]_e \qquad (8)$$

Let K_d be reciprocal of K_a:

$$K_d = \frac{1}{K_a} = \frac{[Ab]_e [Ag]_e}{[AbAg]_e} \qquad (9)$$

Substituting Eq. (8) in Eq. (9):

$$[AbAg]_e = \left(\frac{[Ag]_e}{K_d + [Ag]_e} \right)[Ab]_o \qquad (10)$$

In reciprocal form,
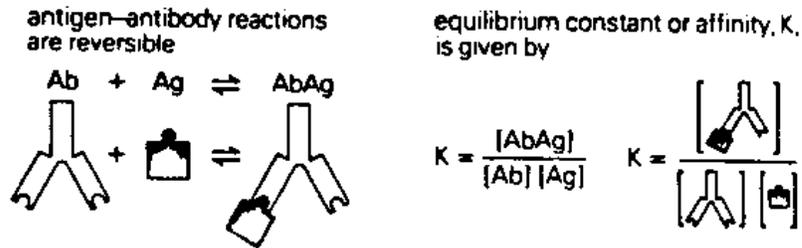
$$\frac{1}{[AbAg]_e} = \frac{1}{[Ab]_o} + \left(\frac{K_d}{[Ab]_o} \right)\frac{1}{[Ag]_e} \qquad (11)$$

Therefore, from a plot of $1/[AbAg]_e$ vs. $1/[Ag]_e$, K_d can be obtained experimentally, and $K_a = 1/K_d$. This process is shown graphically in Fig. 3.16b. K_d is equal to the reciprocal concentration of free antigen necessary to occupy half of the antigen binding sites of the antibody (Fig. 3.16a).
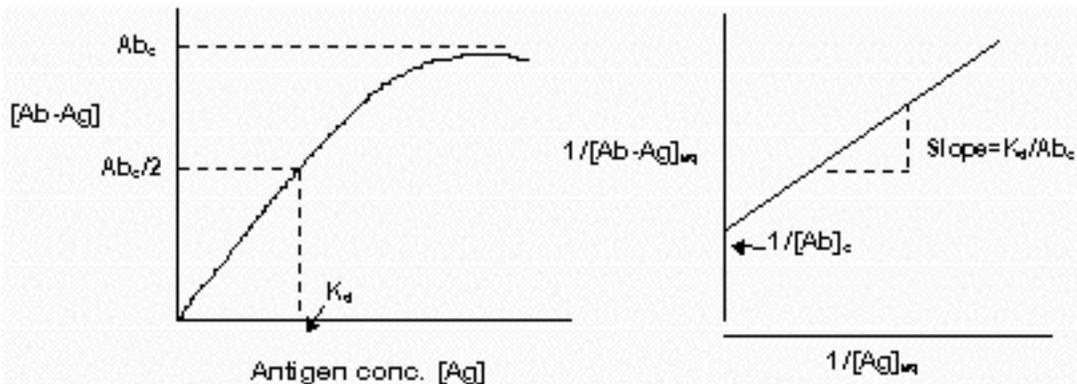
Fig. 3.16. (a) Definition of K_d; (b) graphical determination of K_a.

Avidity   Multivalent binding between antibody and antigen (avidity or functional affinity) results in a considerable increase in stability as measured by the equilibrium constant, compared to simple monovalent binding (affinity or intrinsic affinity, in the example of Fig. 3.17, an arbitrarily assigned value of 104 L/mole) is used). This is sometimes referred to as the 'bonus effect' of multivalency. Thus there may be a 10^3 fold increase in binding energy of IgG when both valencies (combining sites) are utilized, and 10^7 fold increase when IgM binds antigen in a multivalent manner. Monovalent antigen combines with

multivalent antibody with no greater afinity than it does with monovalent antibody.

3.3.6. Antibodies in Assay

Monoclonal Antibody When antibodies are produced from B cells, some of them are polyclonal - i.e., binds with more than one type of antigen, whereas others are monoclonal - i.e., binds with only one type of antigen. Usually, antibodies are produced by immunizing an animal (such as a mouse). The antibodies produced are collected and the monoclonal antibodies are separated. The monoclonal antibodies (Mab) is the key for immunoassays.

ELISA Although the antibody-antigen binding is highly specific, the direct detection of this measurement has been difficult. Therefore, labels are used for measuring the complexation product. Enzymes, fluorescent molecules, and radioactive labels are used for the labeling of the complexation. An example is **ELISA (Enzyme-Linked Immuno Sorbent Assay).** In this method, a known amount of enzyme-labeled antigen is added to a sample with unknown antigen concentration. When this mixture reacts with antibody, wnzyme-labeled and non-labeled antigens compete for the binding sites of the antibody. The more antigen there is in the sample, the lower is the fraction of enzyme-labeled antigen in the antigen-antibody complex. After removal of unbound antigen the amount of bound enzyme-labeled antigen is determined via the enzyme-catalyzed reaction (Fig. 3. 18). Widely used indicator enzymes are horseradish peroxidase and alkaline phosphatase. Urease and deaminases have also been used in immunosensors

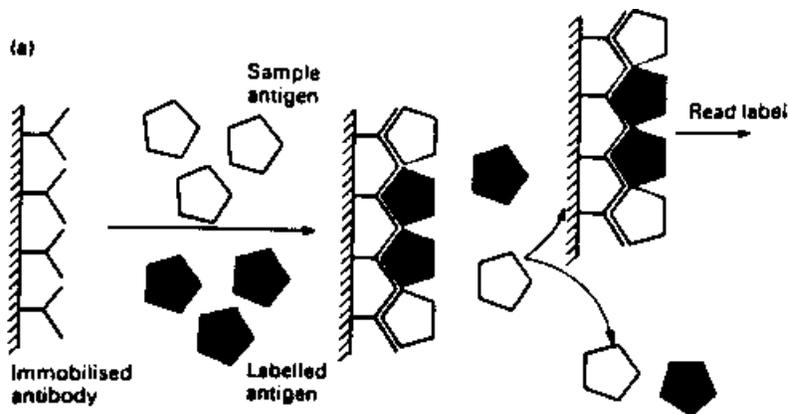| antibody	Fab	IgG	IgG	IgM
effective antibody valence	1	1	2	up to 10
antigen valence	1	1	n	n
equilibrium constant	10^4	10^4	10^7	10^{11}
advantage of multivalence	-		10^3 fold	10^7 fold
definition of bining	affinity	affinity	avidity	avidity

Fig. 3.17. Antibody avidity.

Fig. 3.18. Competitive binding immunoassay (ELISA).

3.4. Receptors

Receptors at Membrane　Biological receptors are protein molecules with a specific affinity for hormones, antibodies, enzymes, and other biologically active com-pounds; most of them are bound to the cell membrane. A receptor-ligand interaction is transmitted to other molecules inside the cell, where consecutive reactions are triggered.

Hormone Receptors　The currently best-known **receptors are those for hormones**. Many hormones released into the blood, e.g. insulin, glucagon, or adrenaline, do not penetrate the cell membrane but react with specific receptors at the cell surface. These are present in enormously high amounts, e.g., *a single fat cell of about 50 µm diameter carries 160, 000 insulin receptors*, which corresponds to about 20 receptors per μm^2.

How Receptor Works　The receptor molecules mostly penetrate the cell membrane (Fig. 3.19) and many of them are coupled inside to an enzyme system. A conformational change of the receptor molecule by hormone binding may be directly mediated to the enzyme and result in its activation. Thus, for example the **adrenaline receptor at the surface of the liver cell** reacts with adrenaline formed in the adrenocortex and released into the blood under stress. The resulting conformational change of the receptor molecule activates associated adenylate cyclase reaching into the internal space of the cell and converting ATP to cyclic adenosine monophosphate (cAMP). cAMP initiates the phosphate transfer from ATP to other enzymes by protein kinase. In this way a number of other enzyme reactions are started, leading to a cascade of activated enzymes. Finally, a single adrenaline molecule will have stimulated several thousands of enzyme molecules, which will themselves liberate about three million glucose molecules from glycogen within a few seconds. An extremely weak chemical signal is thus immediately enzymatically amplified a millionfold and the sugar reserve of the body is mobilized.

Smell Receptor   In addition to hormone receptors, taste and olfactory receptors are typical examples of this biospecific recognition process. Presumably there are about 20 to 30 primary smells. After being bound to the appropriate receptor, their molecules cause conformational changes in the receptor molecule leading to a depolarization of a part of the nerve cell membranes and initiating an action potential.

Light Receptor   Another receptor type is that of light receptors. The retina of the human eye contains about 108 tightly packed receptor cells. Here, biochemical reactions, namely of the rhodopsin molecule, are not in-itiated by the binding of chemical substances but by light quanta. Thereactions are enzymatically amplified and transformed into electrical impulses via membrane potential changes. Because of its light-absorbing chemical group the protein bacteriorhodopsin from salt-tolerant halobacteria has been studied in detail as a photoreceptor model. A single photon is sufficient to give rise to a conformational change of the protein and to transport two protons outside the cell. This 'proton pump' forms a proton and voltage gradient across the cell membrane driving the production of energy-rich ATP.

Current Status   Compared to the investigation of enzymes, that of the structure and function of membrane-bound receptors and their biotechnological application is only at the very beginning. An analogous classification, e.g., according to recognition mechanism or specificity has not yet been attempted. However, any progress in this field will provide impetus to the development of new biosensors based on receptors.
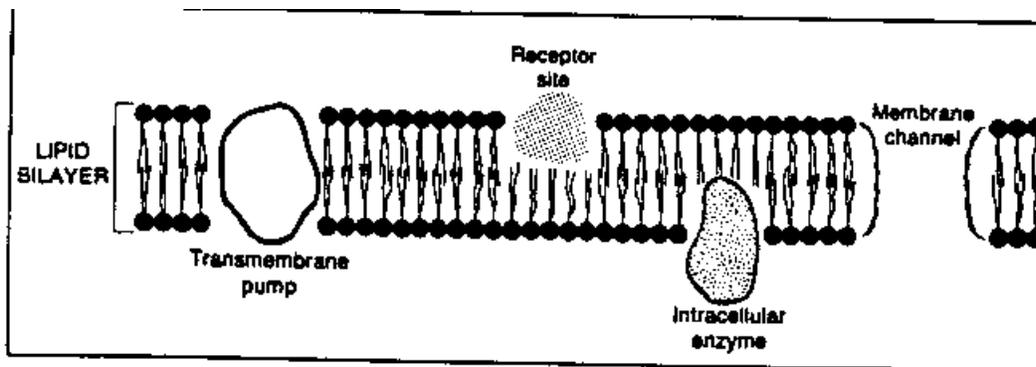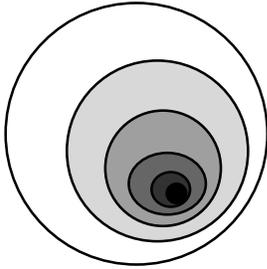

Fig. 3.19. Location of receptor.

Chapter 4. Biosensor Fundamentals

4.1. Fabrication of Glucose Biosensor

4.1.1. Glucose Biosensor

Use H$_2$O$_2$ Sensor for Transducer   Glucose sensor can be made in many different ways by using the enzyme glucose oxidase. This enzyme glucose oxidase catalizes the following reaction:

$$\text{Glucose} + O_2 \xrightarrow{\text{Glucose Oxidase}} \text{Gluconic acid} + H_2O_2$$

To measure the glucose concentration, three different methods can be used:

1. Measure oxygen consumption by an oxygen sensor
2. Measure acid (gluconic acid) production by a pH sensor
3. Measure production of H$_2$O$_2$ by a peroxide sensor.

Note that an oxygen sensor is a transducer that converts oxygen concentration into electrical current. A pH sensor is a transducer that converts pH change into voltage change. Similarly, a peroxidase sensor is a transducer that converts peroxidase concentration into an electrical current.

Structure of Biosensor   The structure of the biosensor to be made is shown in Fig. 4.1. The base transducer consists of H$_2$O$_2$ sensor which is essentially the same as the oxygen sensor fabricated in Section 2.2 (the signal conditioning circuit is shown in Fig. 2.11). The enzyme glucose oxidase is immobilized in front of the H$_2$O$_2$ sensor between two membranes. The **inner membrane** is a permeselective membrane that allows passage of H$_2$O$_2$ where as the **outer membrane** separates the biosensor from measurement medium.

Enzyme Immobilization   The enzyme can be immobilized by: (1) physical entrapment; (2) cross-linking; and (3) covalent attachment. All three methods are to be used. The enzyme used is glucose oxidase in powder form (Sigma Type II, specific activity 25 IU/mg - see Section 3.2.2 for the definition of specific activity).

4.1.2. Glucose Biosensor by Physical Adsorption

Step 1. Prepare a mixture of 24 g of cyclohexanone, 24 g of acetone and 1 g of cellulose acetate (39.8% acetyl content, available from Aldrich). This is for casting the permeselective membrane.
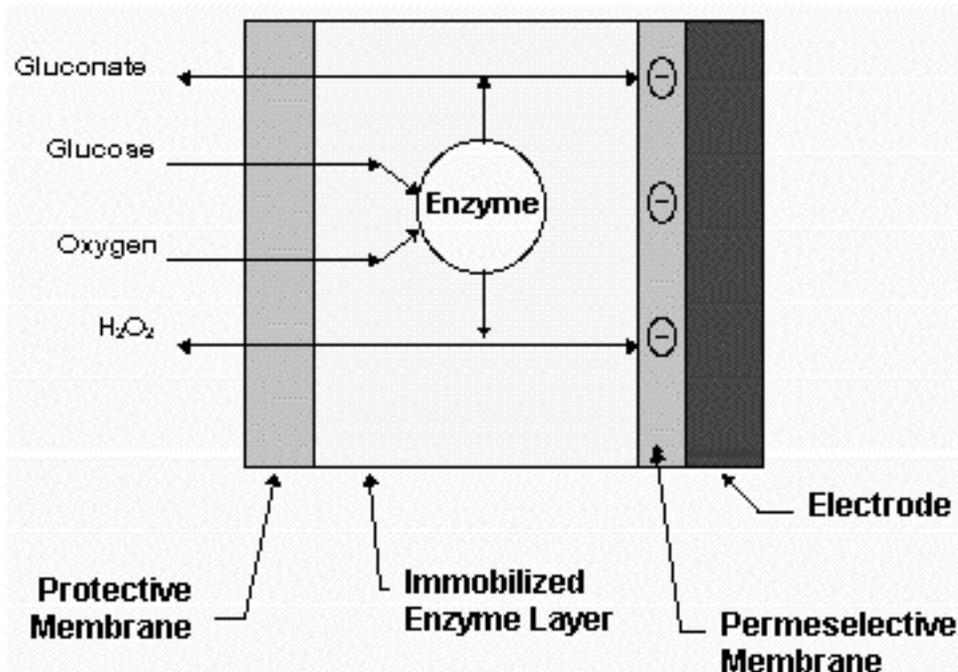
Fig. 4.1. Overall structure of glucose biosensor.
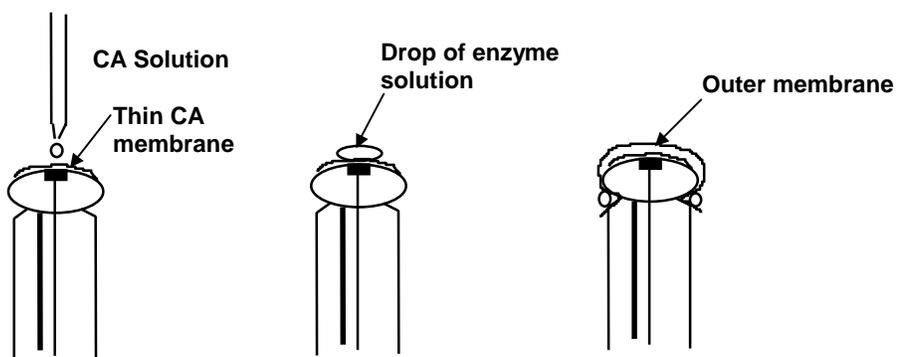
Fig. 4.2. Glucose biosensor by physical adsorption of enzyme.

Step 2. Stir the mixture at room temperature until the cellulose acetate has dissolved and then cast a thin film on to the surface of the sensor probe. Allow the solvent to evaporate.

Step 3. Dissolve glucose oxidase in 0.1 M phosphate buffer, pH 7.4 to a final concentration of 25 mg/mL. Place 20 μL of the enzyme solution on top of the cellulose acetate membrane and allow the water to evaporate (5-10 min).

Step 4. Cover the dried enzyme layer with a 4 cm^2 membrane of polycarbonate membrane (0.05 μm pore size, 10 μm thick; Nucleopore) or general purpose dialysis tubing (MW cut off of 12,000-14,000). Fix the membrane with an O-ring (or silicone tubing).

Step 5. Trim off the excess membrane and place the probe in a 0.1 M phosphate buffer solution, pH 7.4 for 2 hr before use.

Useful life of the sensor will be several months if the probe is stored in room temperature in phosphate buffer. The steps involved are shown in Fig. 4.2.

4.1.3. Method 2: Crosslinking with glutaraldehyde

Generally higher loading of active enzyme can be obtained if the adsorbed enzyme is cross-linked with glutaraldehyde.

The method is the same as in Method 1 except one more step is added right after Step 3.

Step 3a. After the enzyme solution has dried, add 10 μL of a 1% solution of glutaraldehyde (Sigma Type I, supplied as a 25% solution which should be stored frozen and diluted in water immediately before use). Allow the glutaraldehyde solution to evaporate before proceeding to Step 4.

Variation: crosslinking with BSA

Mutual crosslinking of the enzyme with another protein such as bovine serum albumin (BSA; Sigma, Fraction V powder). This procedure can lead to higher enzyme activity and greater stability. Step 3 of Method 1 is modified:

Step 3. Prepare the glucose oxidase solution as in Step 3 of Method 1. Also, prepare 50 mg/mL BSA solution in the same phosphate buffer. Mix 10 μL of each solutions and place the resulting 20 μL on the cellulose acetate membrane. After 1-2 min. add 10 uL of 2.5% glutaraldehyde solution.The liquid layer should harden rapidly. After 1-2 hr, go to Step 4 (Method 1).

4.1.4. Method 3: Covalent attachment to membrane

Covalent attachment is most complicated but is useful in cases when the sensor is so small that the membrane must be fabricated directly on the sensing element. Covalent attachment gives more stable and reproducible enzyme activity.

Step 1. Dissolve 1.8 mg of cellulose acetate in a mixture of 20 mL of acetone and 3 mL of water.

Step 2. Cast 1 mL of this solution onto a clean dry glass plate using a spreader* (Touzart) and allow the solvent to evaporate for 1 min at room temperature. *Note: The spreader has four channeled surfaces which yield films of 5, 10, 15, and 30 μm thickness. A 15 μm thickness is chosen.

Step 3. Remove the membrane by immersing the glass plate in distilled water and floating it off. The resulting membrane is cut in to smaller pieces and stored at room temperature in water.

Step 4. Suspend four membranes (each 2.5 cm square) in 100 mL of 0.1 M sodium periodate for 20 min at room temperature.

Step 5. Wash the membranes in distilled water for 5 min then immerse them in 10 mL of a 10 mg/mL solution of BSA in 0.1 M borate buffer pH 9 for 2 h.

Step 6. Remove 9 mL of the BSA solution and add 4 mg of sodium cyanoborohydride (Aldrich). Incubate at room temperature for 2 h.

Step 7. Wash the membranes in distilled water for 5 min and then store in phosphate-buffered saline at room temperature.

Step 8. Recrystallize *p*-benzoquinone (Merck) from petroleum ether and prepare a solution of 15 mg/mL in ethanol.

Step 9. Add 100 μL of the freshly prepared *p*-benzoguinone to 0.5 mL of a 20 mg/mL solution of glucose oxidase in 0.1 M phosphate buffer pH 7.4 in a tube covered by aluminum foil.

Step 10. Incubate the mixture for 30 min at 37°C and then remove the excess *p*-benzoquinone by gel filtration through a Saphadex C-25 column (1 x 10 cm) equilibrated with 0.15 M sodium chloride and operating at a flow rate of 20 mL/h. Collect the pink-brown band that elutes in the void volume (2-3 mL).

Step 11. Suspend the BSA-cellulose acetate membranes in 2-3 mL of the activated glucose oxidase solution after adjusting the pH of the latter to 8 -9 with 0.25 mL of 1 M sodium carbonate. Incubate at room temperature for 38 h.

Step 12. Remove the membranes, wash them by stirring in 0. 15 S M potassium chloride solution for 24 h and then store them in phosphate-buffered saline pH 7.4 containing 1.5 mM sodium azide.

4.2. Design Variables

4.2.1. Immobilization Methods

Four methods are used for immobilizing enzyme for use in a biosensor: (1) adsorption; (2) entrapment; (3) covalent coupling; and (4) cross-linking. These

four methods are compared in Table 4.1. Among various methods, the cross-linking method is most frequently used because it has the advantage of the covalent bonding yet the cost is inexpensive.

Effects of Immobilization   With *immobilized enzymes* the measured reaction rate depends not only on the substrate concentration and the kinetic constants K_M and V_{max} but also on so-called immobilization effects. These effects are due to the following alterations of the enzyme by the immobilization process.

1. Change in Conformation   Conformational changes of the enzyme caused by immobilization usually decrease the affinity to the substrate (increase of K_M). Furthermore, a partial inactivation of all, or the complete inactivation of a part of the enzyme molecules may occur (decrease of V_{max}). These two cases of a conformation-induced drop of V_{max} may be distinguished by measuring the activity of the resolubilized enzyme or by titration of the active center with an irreversible inhibitor.

2. Change in Microenvironment   Ionic, hydrophobic, or other interactions between the enzyme and the matrix (microenvironmental effects) may also result in changed K_M and V_{max} values These essentially reversible effects are caused by variations in the dissociation equilibria of charged groups of the active center.

3. Non-Uniform Distribution   A non-uniform distribution of substrate and/or product between the enzyme matrix and the surrounding solution affects the measured (apparent) kinetic constants.

4. Reaction and Diffusion   In biosensors the biocatalyst and the signal transducer are spatially combined, i.e., the enzyme reaction proceeds in a layer separated from the measuring solution. The substrates reach the membrane system of the biosensor by convective diffusion from the solution. The rate of this external transport process depends essentially on the degree of mixing. In the multilayer system in front of the sensor the substrates and products are transferred by diffusion. Slow mass transfer to and within the enzyme matrix leads to different concentrations of the reaction partners in the measuring solution and in the matrix. Diffusion and the enzyme reaction do not proceed independently of one another; they are coupled in a complex manner.

4.2.2. Outer Membrane

Requirements
1. The outer membrane has to be compatible with the medium into which it will be placed. Therefore, the requirement will be different depending on the nature of the measurement medium. For example, the outer membrane for biosensor used in liquid samples should be different from that intended for implantation application. For the latter application, bio-compatibility becomes an issue (the rejection of the sensor by body may occur).

2. The outer membrane should offer low diffusional resistance to analytes   while the resistance should be high for macromolecules.

3. For long-term continuous use applications, the fouling of the membrane must be minimal. The fouling causes an increase in the diffusional resistance of the analyte and thus the signal of the sensor changes as the fouling progresses. If microorganism grows on the surface of the outer membrane, the passage of oxygen to the enzyme layer is hindered which makes the sensor to behave erroneously.

Table 4.1. Comparison of four enzyme immobilization methods.
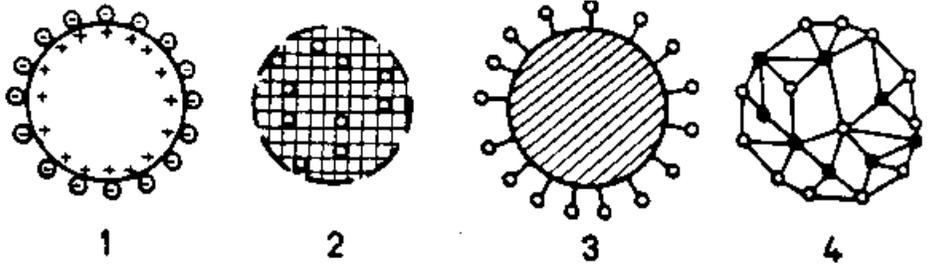
	1. Adsorption	2. Entrapment	3. Covalent coupling	4.Crosslinking
Matrix material	ion exchange resins, active charcoal, silica gel, clay, aluminum oxide, porous glass	alginate, carageenan, collagen, polyacrylamide, gelatine, silicon rubber, polyurethens	agarose, cellulose, PVC, ion exchange resins, porous glass	Crosslinking agents: glutaraldehyde, bisisocyanate, bisdiazobenzidine
Nature of bonding	reversible; changes in pH, ionic strength may detach the enzyme	physical entrapment	chemical bonding	entrapment; functionally inert proteins are often used together (BSA, gelatin)
Enzyme loading	low	low	high	high
Enzyme leakage	some	some	very low	low
Loss of activity	negligible	negligible	significant	small
Cost	inexpensive	inexpensive	expensive	inexpensive

Optimization of Biosensor   Variations of the diffusion resistance of the semipermeable membrane are being used to optimize the sensor performance. A semipermeable membrane with a molecular cutoff of 10, 000 and a thickness of 10 μm only slightly influences the response time and sensitivity. In contrast, thicker membranes such as polyurethane or charged material, significantly increase the measuring time, but may also lead to an extension of the linear measuring range.   Table 4.2. lists some of the commercially available membranes that can be used as the outer membrane.

Table 4.2. Available pre-cast membranes

	Collagen	Polycarbonate (Nucleopore)	Cellulose acetate
Characteristics	a hydroxylic natural protein	uniform pore size	slightly negative due to -COO$^-$
Derivatizability	easy		easy
Temperature stability	O.K. at room T unstable at 37°C	Stable	stable
Permeselectivity	exclude protein	exclude protein	exclude protein, retard transport of anionic species
Source	Sigma	Nucleopore	Amicon

4.2.3. Inner Membrane

Requirements: The inner membrane should be permeselective to target species only (for example H_2O_2 only for the current glucose sensor). Also, it should be as thin as possible and stable for long-term use. Some of the solution castable membranes and their characteristics are compared in Table 4.3.

Table 4.3. Solution castable membranes

	Cellulose acetate	Nafion	Polyurethane*
Characteristics		perfluorosulphonic acid ionomer, negatively charged	widely varying MW dep. on source
Solvent	acetone, cyclohexanone	low MW alcohols	dissolves in 98% tetrahydrofuran and 2% dimethylform-amide
Coating method	dip coating, in-situ formation	dip coating as thin as 1000A (with 5% solution)	dip coating
Other characteristics		tend to adsorb proteins & cations, not useful as an outer membrane	biocompatible, retard glucose access to enzyme layer, passes O_2 well

* More useful as an outer membrane.

4.2.4. Effect of Enzyme Loading

Internal Diffusion Is Important   Usually in the operation of biosensors the flow conditions are adjusted to provide a mass transfer rate from the solution
to the membrane system faster than that of in the enzyme layer (the internal mass transfer). In the immobilized enzyme layer, reaction and diffusion occur simultaneously. Therefore, rigorous modeling is required to fully characterize the behavior of a biosensor. The key question in designing a biosensor is: (1) how thick should the enzyme layer be? and (2) how much enzyme has to be placed in the layer? Although rigorous modeling is required to fully characterize the behavior of biosensors, the design can be carried out by considering limiting cases.

Enzyme Loading Factor, f_E   The key variable for determining limiting cases is the enzyme loading factor defined by f_E:

$$f_E = \frac{(V_{max} / K_M)}{D_S / d^2} \qquad (1)$$

where D_s and d are diffusivity of the substrate and the enzyme immobilization layer thickness. The enzyme loading factor is essentially the ratio of the reaction rate (for the case when $K_M \gg S$; or first order reaction) to the diffusion rate (when $k = D_s/d$ and the mass flux J is expressed as $k(dS/dx)$).

Diffusion Control: $f_E > 25$ When the rate of reaction is much faster than the rate of diffusion, bottle neck of the transport process is the diffusion. In such a case, we say, the biosensor is operated in a diffusion-controlled regime. The condition is:

$$f_E > 25 \qquad (2)$$

When this condition is met, the sensor signal depends on the diffusion process. This means the sensor output signal is linear with the analyte concentration, and is independent of the reaction rate of the enzyme layer.

Reaction Control: When the rate of reaction is much slower than the rate of diffusion, the bottle neck of the transport process is the reaction. In such a case, we say, the biosensor is operated in a reaction-controlled regime. The condition is:

$$f_E > 25 \qquad (3)$$

When this condition is met, the sensor signal depends on the reaction rate and independent of the diffusion rate. This means that the sensor output signal depends on the reaction rate expression - i.e. the Michaelis-Menten equation. Therefore, the sensor output signal will be non-linear with the analyte concentration.

Advantages of Diffusion Control There are many advantages of diffusion-controlled biosensor. These include:

1. Sensitivity is independent of enzyme content and activity
2. Sensitivity is independent of inhibitors and pH variations.
3. There is small temperature effect. Because the activation energy of diffusion is much less than that of reaction.
4. Extended linear range. Extendible beyond $[S] = K_M$.
For kinetics controlled case, linear range is achieved only for $[S] < K_M$
5. Slower response time. The response time is determined by the diffusion time of the reaction product (sensed): d^2/D_p
This is the major disadvantage of the diffusion controlled biosensor.
6. Greater functional stability. Due to the excess of enzyme in the membrane, the sensor will have higher functional stability. With diffusion controlled sensor, 2000-10,000 measurement is possible, while kinetically controlled sensors permit only 200-500 measurement.

Apparent K_M When an enzyme is immobilized, the K_M value changes because of the changes in enzyme conformation and the microenvironment. Table 4.4. lists the K_M values for some of the immobilized enzyme systems.

Table 4.4. Apparent enzyme activity and KM values of adsorbed layers and enzyme membranes.

Enzyme	Immobilization	Apparent enzyme activity, mU/cm^2	K$_M$ value
GOD*	gelatine entrapment	110	7.5
GOD	collagen, covalent	60-80	3.0
GOD	cellulose acetate	340	
GOD	cellulose acetate	>1000	
GOD	PVA entrapment	160-700	
GOD	spectral carbon, adsorbed	150-200	
GOD	carbon, covalent	50-170	3.1-19.1
β-galactosidase	gelatin entrapment	1000	
Urease	cellulose triacetate, entrapment	3-30	2.4
Cholesterol oxidase	collagen, crosslinked	3	
Creatinine amido-hydrolase	cellulose acetate, covalent	1140	278
Creatine amido-hydrolase	cellulose acetate, covalent	110	64.9
Sarcosine oxidase	cellulose acetate, covalent	13	2.4

* **GOD stands for glucose oxidase.**

Homework Problems 2

1. For the glucose sensor that we fabricated in class, design an experiment to obtain K_m value. State what kind of measurements are necessary under what conditions.

2. How can one make a diffusion-controlled biosensor?

3. What would be the flow dependency (i.e., the effect of agitation of the measurement medium) of the biosensor output for
   (a) kinetically controlled biosensor
   (b) diffusion-controlled biosensor

4. How can you determine whether a biosensor is diffusion-controlled or kinetics-controlled without the information on V_{max}, d, K_M, and D_s.

5. State your criterion on selecting the outer and the inner membrane for a glucose sensor intended for implantation in a human body. Consider long term stability, biocompatibility, and fast response time.
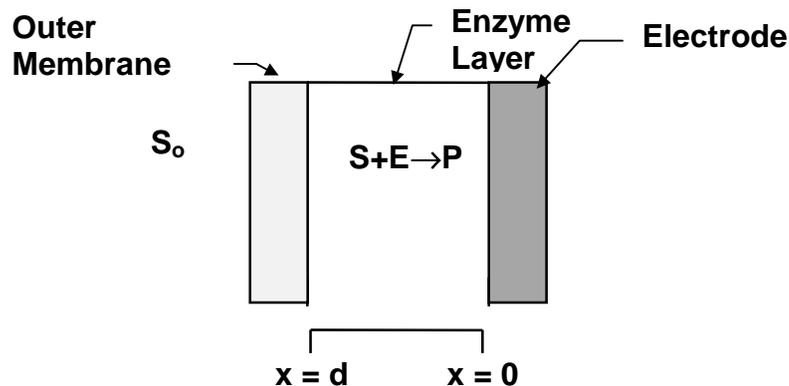
4.3. Modeling of Biosensors

Modeling allows establishing the relationship between design parameters and the sensor signal output. In modeling, simplified and limiting cases are often useful to gain intuition and understanding of the sensing principle.

4.3.1. Amperometric Biosensor

Glucose Biosensor   Glucose biosensor using glucose oxidase is used as an example. The reaction and the transducer used are:

> Reaction: Glucose + $O_2 \rightarrow$ Gluconic acid + H_2O_2
> Transducer: polarographic H_2O_2 electrode
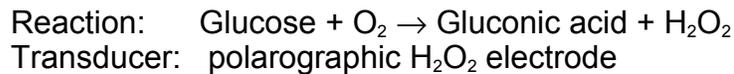
Model geometry

Assumptions   To write mass balance equations for the reaction and to set up boundary conditions, simplifying assumptions are necessary. Here, we will consider a very simple case for illustration purpose.

1. The outer membrane offer no mass transfer resistance to the substrate (glucose) or the product (H_2O_2).
2. No oxygen limitation in the enzyme layer.
3. The substrate concentration at x=d is the same as that of bulk (no mass transfer resistance of the liquid film)
4. The enzymes are homogeneously distributed in the immobilized layer.
5. The inner membrane is extremely thin and offers no mass transfer resistance to the product or the substrate.
6. The pores of the outer membrane is sufficiently large that it offers no mass transfer resistance to the substrate.
7. One dimensional transport.

Differential Mass Balance   Unsteady state, differential mass balance is written for both the substrate glucose and the product (H_2O_2).

For the substrate (glucose):

$$\frac{\partial [S]}{\partial t} = D_s \frac{\partial^2 [S]}{\partial x^2} - \frac{k_2 [E_o][S]}{K_M + [S]} \tag{4}$$

For the product (H_2O_2):

$$\frac{\partial [P]}{\partial t} = D_p \frac{\partial^2 [P]}{\partial x^2} + \frac{k_2 [E_o][S]}{K_M + [S]} \tag{5}$$

Boundary Conditions   To solve these equations, boundary conditions are necessary.

For the substrate:

    @ t = 0,   [S] = 0   (substrate diffusion has not started)   (6a)

    @ x = 0, $\dfrac{\partial [S]}{\partial x} = 0$ (no transport of substrate)   (6b)

    @ x = d,   [S] = [S]$_o$ (negligible liquid film resistance)*   (6c)

For the product:

    @ t = 0,   [P] = 0   (no product formation)   (7a)
    @ x = 0,   [P] = 0   (rapid reaction of H_2O_2 by polarography) (7b)
    @ x = d,   [P] = 0   (well-mixing assumed)   (7c)

---

* When liquid film resistance is considered, replace Eqs. (6c) and (7c) by:

    @ x = d, $D_S \dfrac{\partial [S]}{\partial x} = k_L ([S]_\infty - [S]_o)$ (6c')

    @ x = d, $D_P \dfrac{\partial [P]}{\partial x} = k_L ([P]_o - 0)$ (7c')

Steady State Case When [S] << K_M   Since the solution process is complicated for the full unsteady state case, we will only solve the steady case when [S] << K_M

For the substrate (glucose):

$$D_S \frac{d^2 [S]}{dx^2} = k[S] \tag{8}$$

where
$$k = \frac{k_2 [E_o]}{K_M} \quad \left(= \frac{V_{max}}{K_M} \right) \qquad (9)$$

B.C. 1: @ x = 0, d[S]/dx = 0   (10a)
B.C. 2: @ x = d, [S] = [S]ₒ   (10b)

For the product (H₂O₂):

$$D_P \frac{d^2[P]}{dx^2} = -k[S] \qquad (11)$$

B.C. 1: @ x = 0, [P] = 0 (due to rapid consumption)   (12a)
B.C. 2: @ x = d, [P] = 0 (due to rapid mixing)   (12b)

Solution of Eq. (5) with BCs (7a) and (7b):

$$[S] = [S]_o \left(\frac{\cosh(x\sqrt{k/D_S})}{\cosh(d\sqrt{k/D_S})} \right) \qquad (13)$$

Substituting Eq. (13) in Eq. (11), and solving with BCs (12a) and (12b):

$$[P] = [S]_o \frac{D_S}{D_P} \left(\frac{x}{d} \left\{ 1 - \frac{1}{\cosh(d\sqrt{k/D_S})} \right\} + \frac{1 - \cosh(x\sqrt{k/D_S})}{\cosh(d\sqrt{k/D_S})} \right) \qquad (14)$$

Sensor signal output will become:

$$I = nFAD_P \left(\frac{d[P]}{dx} \right)_{x=0} = nFAD_S \frac{S_o}{d} \left(1 - \frac{1}{\cosh(d\sqrt{k/D_S})} \right) \qquad (15)$$

4.3.3. Effect of Design Parameters on Sensor Performance

Solution in terms of f_E   The solution of Eq. (15) can be rewritten in terms of the enzyme loading factor f_E:

$$d\sqrt{k/D_S} = d\sqrt{\frac{V_{max}}{K_M D_s}}$$

$$= \sqrt{\frac{V_{max} d^2}{K_M D_S}} = \sqrt{f_E}$$

Therefore,

$$I = nFAD_S \frac{S_o}{d} \left(1 - \frac{1}{\cosh(\sqrt{f_E})} \right) \qquad (16)$$

Sensor Output for Diffusion Control   When $f_E > 25$ (criterion for diffusion control), i.e.,

$$f_E = \frac{V_{max} d^2}{K_M D_S} = \frac{k_2 [E_o] d^2}{K_M D_S} > 25$$

$$I \cong nFAD_S \frac{S_o}{d} \qquad (17)$$

Key Design Parameter   Note that the key design parameter is f_E:

$$f_E = \frac{V_{max} d^2}{K_M D_S} = \frac{k_2 [E_o] d^2}{K_M D_S}$$

To obtain **diffusion control**, how does one adjust V_{max}, d, and D_s ? Which variable is most effective?

Response Time   The response time of the sensor will be proportional to the diffusion time of the product τ_d:

$$\tau_d = \frac{d^2}{D_p} \qquad (18)$$

A. Effect of enzyme loading factor, f_E, on sensor output

Enzyme Reserve   The variation of the enzyme loading is a means of determining the minimum amount of enzyme required for maximum sensitivity. Furthermore, this test should reveal the magnitude of the enzyme reserve of diffusion controlled sensors.

Loading Test   Fig. 4.4. shows the results of a loading test of GOD entrapped in a gelatin layer of 30 µm thickness between two dialysis membranes of 15 µm thickness each. The stationary currents for 0.14 mmol/L glucose (lower part of the linear measuring range) and for 5 mmol/L glucose (saturation) increase linearly with enzyme loading from 46 mU/cm² to 1 U/cm². At higher

GOD loading a saturation value is attained. To calculate the enzyme loading factor, f_E, the following values have been used:

Parameters used
d = 30 μm
K_M = 10 mM/L
D_S = 1.63x10^{-6} cm^2
A = 0.22 mm2
Polarization: +0.6V

As is evident from Fig. 4.4, the transient from the linear region to saturation occurs at f_E values between 7 and 20. This agrees with the theoretically predicted value and indicates that above 1 U/cm^2 the function of the GOD electrode is controlled by internal diffusion.

B. Concentration dependence of signal output

Linear Range   The linear measuring range of biosensors extends over 2-5 decades of concentration The lower detection limit of simple amperometric enzyme electrodes is about 100 nmol/L whereas potentiometric sensors may only be applied down to 100 μmol/L. This shows that the sensitivity is affected not only by the enzyme reaction but also by the transducer.

Oxygen Effect
The linear range extends to 2 mmol/L glucose in the measuring cell. In this region, saturation of the measuring solution by oxygen increases

$$I = nFAD_s \frac{S_o}{d}\left(1 - \frac{1}{\cosh(\sqrt{f_E})}\right)$$

Parameters used

$d = 30$ μm

$K_M = 10$ mM/L

$D_S = 1.63 \times 10^{-6}$ cm^2
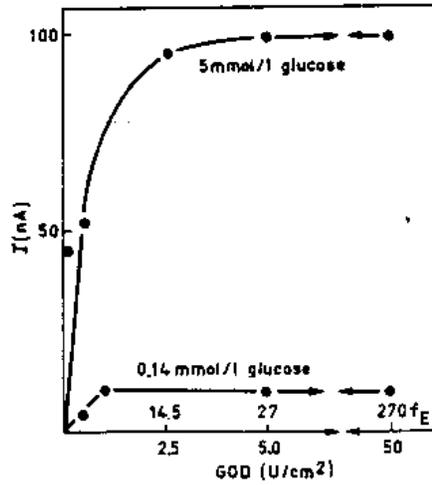
$A = 0.22$ mm2

Polarization: +0.6V

Fig. 4.4. Effect of enzyme loading factor on sensor output.

$$I = nFAD_s \frac{S_o}{d}\left(1 - \frac{1}{\cosh(\sqrt{f_E})}\right)$$

1,2, and 3: under O$_2$ saturation

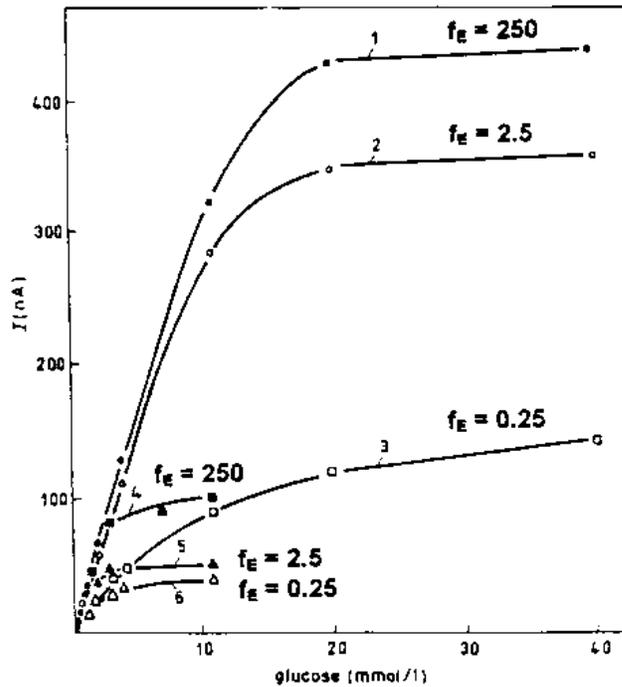4,5, and 6: under air saturation

Fig. 4.5. Sensor output as function of glucose concentration.

the measuring signal by only 10%. At low glucose concentration the cosubstrate concentration (ca. 200 μmol/L at air saturation) influences the enzyme reaction only slightly. By contrast, in the saturation region above 2 mmol/L glucose the

current rises by a factor of 4.5. At the same time the linear range is extended by oxygen saturation (see Fig. 4.5).

C. Effect of pH on sensor output

With a high enzyme excess in the membrane, pH variations should have only a minor influence on the measuring process. Therefore the pH profiles in the linear measuring range and under diffusion control should be substantially less sharp than those of the respective enzyme in solution. The results obtained with a GOD-gelatin membrane (Fig. 4.6) agree with this assumption. With 0.14 mmol/L glucose the curve is almost as flat as that of the H_2O_2 signal. On the other hand, with 10 mmol/L a pronounced maximum is found. At this saturating concentration, the signal depends on the enzyme activity and therefore distinctly on pH. The pH optimum of immobilized GOD is about 0.9 pH units more alkaline than that of the soluble enzyme. Obviously the formation of gluconic acid within the enzyme membrane causes a local pH decrease, shifting the optimum to higher pH in the solution.

D. Effect of temperature on sensor output

The rate of enzyme reactions rises with temperature up to a certain optimum. Above that, the effect of thermal inactivation dominates over that of the increase of the collision frequency. Enzyme stabilization by immobilization is frequently reflected by an increase of the temperature optimum for substrate conversion. If kinetic and diffusion control are superimposed, the higher activation energy results in a predominant acceleration of the enzyme reaction with rising temperature. Thus, *the slower enhancement of the diffusion rate makes mass transfer the limiting factor.* Therefore, the activation energy determined at lower temperatures is ascribed to the enzyme reaction, and that at higher temperatures to diffusion. Besides this, the temperature profile is affected by temperature-dependent conformational changes of the enzyme and decreasing solubility of the cosubstrate. The glucose sensor with the GOD-gelatin membrane exhibits a temperature optimum of about 40°C . Below the optimum the Arrhenius plot (Fig. 4.7) gives parallel straight lines for different glucose concentrations and enzyme loading. The difference between the activation energy of H_2O_2 diffusion, 33.5 kJ/mol, and that of GOD-catalyzed glucose oxidation, 25.5 kJ/mol, is probably too small to give rise to two separate linear regions. That is why purely diffusion controlled GOD

For $[S] \ll K_M$

$$I = nFAD_s \frac{S_o}{d}\left(1 - \frac{1}{\cosh(\sqrt{f_E})}\right)$$

where

$$f_E = \frac{V_{max} d^2}{K_M D_S} = \frac{k_2 [E_o] d^2}{K_M D_S}$$

Parameters used

d = 30 μm
K_M = 10 mM/L
D_S = 1.63x10^{-6} cm^2
A = 0.22 mm2
Polarization: +0.6V
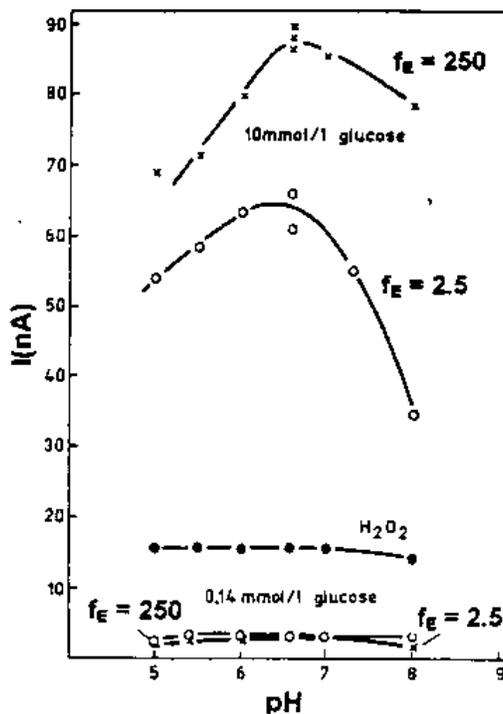

Fig. 4.6. Effect of pH on sensor output.

$$I = nFAD_s \frac{S_o}{d}\left(1 - \frac{1}{\cosh(\sqrt{f_E})}\right)$$

Parameters used

d = 30 μm
K_M = 10 mM/L
D_S = 1.63x10^{-6} cm^2
A = 0.22 mm2
Polarization: +0.6V

For H_2O_2 diffusion: E_a = 33.5 kJ/mol
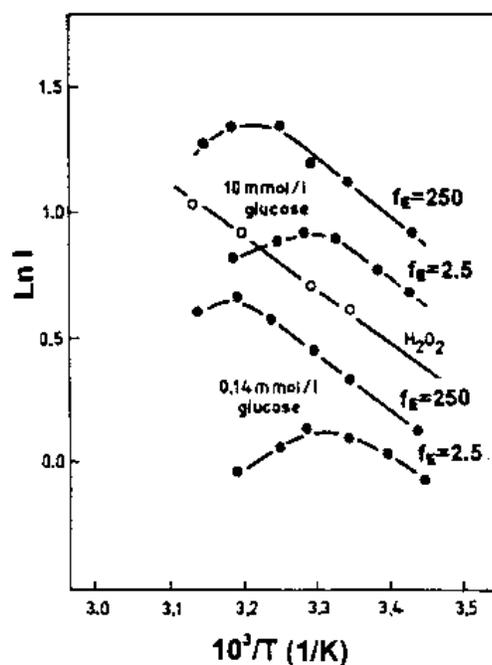For GOD reaction: E_a = 25.5 kJ/mol

Fig. 4.7. Effect of temperature on sensor output.

electrodes are not significantly different from kinetically controlled ones with regard to activation energy.
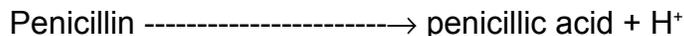
4.3.4. Potentiometric Biosensor

Characteristics of Potentiometric Sensor   The potentiometric sensor (the base sensor) has the characteristic that are different from amperometric sensors. These are:

1. The measured species (such as H^+, NH_4^+, etc.) is not consumed.
2. The sensor measures the activity (for dilute solutions, molar concentration can be used), of a specific ion (C_o) in reference to its internal standard (C_i).
3. The output is in voltage (E_m).
4. The ion specificity comes entirely from the membrane. For example, there is a membranes specific to H^+, one for NH_4^+, one for Ca^{++}, etc.
5. The output of the base sensor is in voltage; independent of the sensor size; and proportional to natural log of C_o.
(Note: for symbols refer to Fig. 2.15).

Examples of Potentiometric Biosensor   many biosensors have been developed using pH sensor as the base transducer. Some of the examples and their performances are illustrated in the following.
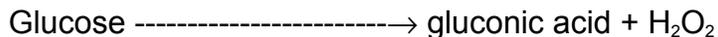
1. Penicillin biosensor
$$\text{Penicillin} \xrightarrow{\textbf{Penicillinase}} \text{penicillic acid} + H^+$$

    Response: 52 mV/decade over 5×10^{-2} to 10^{-4} M.

2. Glucose biosensor

$$\text{Glucose} \xrightarrow{\textbf{Glucose oxidase}} \text{gluconic acid} + H_2O_2$$

    Response: log linear response from 0.1 mM to 1 mM

3. Urea biosensor
$$\text{Urea} \xrightarrow{\textbf{Urease}} 2\ NH_4^+ + NCO_3^-$$

    Response: log linear response from 5×10^{-5} to 5×10^{-3}

Configuration of Biosensor   For potentiometric biosensors, the outer membrane is often not used.Often, the inner membrane is not used either. In such a case, the enzyme is immobilized directly on the surface of the potentiometric sensor.

Model geometry and assumptions used for the modeling

Same as those of the amperometric biosensor.

Steady State Solutions for the Case When [S] << K$_M$

Equation and Bcs The same equations (Eq. (8) through Eq. (12b)) are used to obtain the steady state solutions. The only change that has to be made is B.C.1 for the product. B.C. 1 has to be replaced by:

B.C. 1: @ x = 0, d[P]/dx = 0 (no transport; P is not consumed)

The solutions are:

$$[S] = [S]_o \left(\frac{\cosh(x\sqrt{k/D_S})}{\cosh(d\sqrt{k/D_S})} \right) \qquad (19)$$

$$[P] = [S]_o \frac{D_S}{D_P} \left(1 - \frac{\cosh(x\sqrt{k/D_S})}{\cosh(d\sqrt{k/D_S})} \right) \qquad (20)$$

Solution Sensor signal output will be proportional to [P] at x = 0:

$$[P]_{X=0} = [S]_o \frac{D_S}{D_P} \left(1 - \frac{1}{\cosh(\sqrt{f_E})} \right) \qquad (21)$$

$$E_{out} = \frac{RT}{zF} \ln \frac{[P]_{X=0}}{[P]_i} = \frac{RT}{zF} \ln[P]_{X=0} + const \qquad (22)$$

Sensor Output For sensors under diffusion control (f_E > 25):

$$E_{out} = \frac{RT}{zF} \ln \frac{[P]_{X=0}}{[P]_i} = \frac{RT}{zF} \ln \left([S]_o \frac{D_S}{D_P} \right) + const \qquad (23)$$

It has to be noted that under this condition, the sensor output E$_{out}$ is proportional to the natural log the analyte concentration *S*, while the temperature dependence is linear:

$$E_{out} \propto \ln [S]_o \qquad (25)$$

$$E_{out} \propto T \qquad (26)$$

Effect of enzyme loading factor on sensor output

$$E_{out} = \frac{RT}{zF} \ln \frac{[P]_{X=0}}{[P]_i} = \frac{RT}{zF} \ln \left([S]_o \frac{D_S}{D_P} \left(1 - \frac{1}{\cosh \sqrt{f_E}} \right) \right) + const \qquad (27)$$

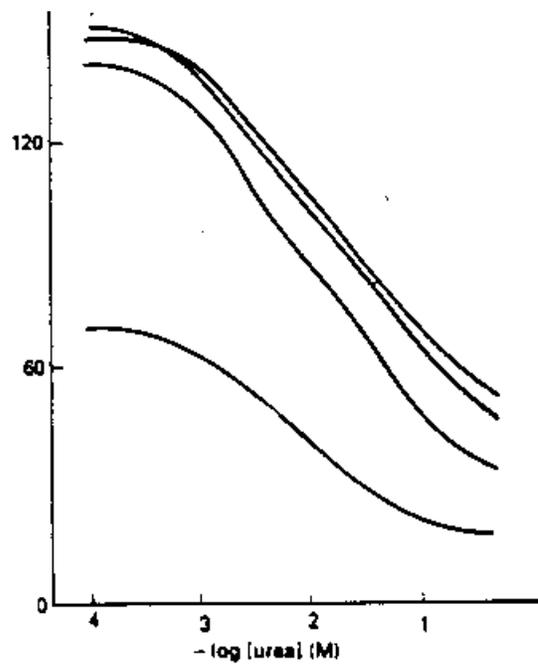Fig. 4. 8 shows the effect of enzyme loading on sensor output.

Fig. 4.8. Enzyme loading effect on potentiometric biosensor.

Homework Problems

In this section, we derived expressions for I and E_{out} (Eq. (15) for the amperometric sensor, and Eqs. (23) for the potentiometric sensor, respectively) for case when $[S] << K_M$. Derive equivalent expressions for both type sensors under 'kinetics control'. What would be the effect of temperature and enzyme loading for the sensor ouputs?

(Hint) Under 'kinetics control', the reaction is the slow step and [S] in the enzyme layer is the same as $[S]_o$. Therefore, Eq. (8) will become:

$$D_s \frac{\partial^2 [S]}{\partial x^2} = \frac{k_2 [E_o][S]_o}{K_M + [S]_o}$$

The same modification has to be made with Eq. (11). The solution procedure will be simpler because the right hand side of the equation is a constant.

Problem Set Biosensor Part

1. Between amperometric sensor and potentiometric sensor, which one tends to have mass transfer limitation? Why?

2. Enzyme-based glucose sensor has two membranes - one outer and one inner. What are the roles of each membranes?

3. What is the advantage of using optical fiber as the transducer in making biosensors over the amperometry.

4. In class, we discussed amperometry, potentiometry, optical method, thermistors, and piezo crystal for use as a base transducer for making biosensors. Illustrate other possible forms of transducers and discuss their advantages and disadvantages.

5. What is the advantages and disadvantages of using biomolecules such as enzymes and antibodies for sensor applications?

6. Calculate the steady state current output from the DO sensor that we have made in the lecture when it is placed in air saturated water. Assume that we used 25 μm thick Teflon FEP membrane. Use the data given on pp 2-27.

7. Suppose we use two 25 μm Teflon membranes to cover the cathode. (a) What will be the output current under the same condition as in Problem 1?
(b) Comment on the response time. Will this sensor become slower? By how much? Take diffusivity of oxygen in water as $2x10^{-5}$ cm^2/s.

8. Suppose you place the DO sensor in a stagnant liquid which has a magnetic stirrer in it. Qualitatively show the sensor output current as you increase the stirring rate.

9. Flow dependency of DO sensor is a big problem in actual measurements. The flow dependency can be reduced by placing a silicone membrane over the Teflon membrane. Explain why this reduces the flow sensitivity.

10. Suppose we measure DO concentration in (a) dense aerobic culture, and (b) low cell density aerobic culture. In which case the accuracy will be better. Explain why. What can you do to improve the accuracy?

11. For the glucose sensor that we discussed in class, design an experiment to obtain K_m value. State what kind of measurements are necessary under what conditions.

12. How can one make a diffusion-controlled biosensor?

13. What would be the flow dependency (i.e., the effect of agitation of the measurement medium) of the biosensor output for (a) a kinetically controlled biosensor (b) a diffusion-controlled biosensor

14. How can you determine whether a biosensor is diff usion-controlled or kinetics-controlled without the information on V_{max}, d, KM, and Ds.

15. State your criterion on selecting the outer and the inner membrane for a glucose sensor intended for implantation in a human body. Consider long term stability, biocompatibility, and fast response time.

16. What is the disadvantage of using antibodies instead of enzymes for sensing? Suggest
    possible solutions.

17. What is the effect of enzyme loading in terms of sensor performance?

18. Discuss the effect of temperature on 'kinetics limited' biosensor and 'diffusion limited' biosensor.

19. Discuss the effect of response time on 'kinetics limited' biosensor and 'diffusion limited' biosensor.