Submitted to
Angewandte Chemie

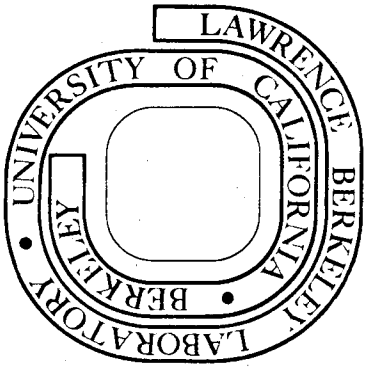
LBL-1559
Preprint

BIOPOLYMERS: ORIGIN, CHEMISTRY AND BIOLOGY

Melvin Calvin

January 11, 1973

Prepared for the U. S. Atomic Energy
Commission under Contract W-7405-ENG-48



DISCLAIMER

BIOPOLYMERS: ORIGIN, CHEMISTRY AND BIOLOGY

Melvin Calvin*

Laboratory of Chemical Biodynamics, University of California, Berkeley, California

ABSTRACT

A brief statement concerning the way in which biopolymers may have originated in the nonbiological world is made, including experimental evidence. This also includes a discussion of such matters as the way in which the code might have originated, that is, the relationship between polypeptides and polynucleotides as well as the secondary and tertiary structure resulting from the primary structure determination. The importance of the interaction of biopolymers with lipids for the formation of limiting membranes is discussed, leading to the formation of cells and other selforganizing cellular type organelles. Thus, the second critical physical-chemical problem for cellular organization, namely, the biopolymer-lipid interaction, is now coming under scrutiny, both in terms of synthetic systems as well as natural ones.

---

* The work described in this paper was sponsored by the U.S. Atomic Energy Commission and by the Elsa U. Pardee Foundation for Cancer Research. This is a transcription of an address presented at the dedication of the Midland Macromolecular Institute, Midland, Michigan, September 29, 1972.

It seems that Dr. Elias has a particular talent for asking his guest speakers critical questions. Just as his question to Professor Overberger induced Professor Overberger to reorganize his thoughts in a new framework, so his question to me has done the same thing. I constructed this collection of ideas over a period of several months, and came out with an outline which I will place on the board, so as I go through it you will see the way in which the relationships which I will try to describe are related to each other and to what Professor Overberger has just said and to what the speakers later today will continue. The first section of this discussion will be the origin of biopolymers, particularly of proteins, including the chemistry which may be involved. The second part will describe the secondary-, tertiary-, quaternary-structure of biopolymers and the interaction between proteins and nucleic acids. Following this is the third section on the origin of the code; the fourth section deals with the interaction between the proteins and lipids.

Why have I chosen this particular method of organization? Many of the phenomena (reactions) which I will discuss are actually extrapolations and modifications of reactions with which polymer chemists are familiar, but, as Professor Flory mentioned yesterday, there is a kind of a curtain which keeps the polymer chemist, whose background is oriented to the structure and synthesis of macromolecules of any kind and the determination of the basic principles involved, and the group that has grown up in biology and biochemistry and which has learned about the physical chemistry and the chemistry of the behavior of the kinds of

materials they deal with from quite a different point of view. The Biochemists have felt that these things are almost magic, and they believe new rules, or laws, or chemistry, may be required to understand in detail how the naturally-occurring biopolymers (proteins, nucleic acids, etc.) perform their function. There appears to be some "mystique" about this, and I hope that when I finish this discussion you will see some of the reasons for the "mystique" which is generally in the minds of those who grew up from the side of biology. I have tried to interpret the behavior of biopolymers in terms which are thoroughly familiar to you, as polymer chemists, and may even be naive in your terminology. You must remember that most of this work is being done by people who have not come from your background, and that is why some of the material may seem obvious to you. Some of it, however, may not be obvious to you and I hope will give you suggestions of what to do next.

FORMATION OF BIOPOLYMERS, PARTICULARLY PROTEINS

Let us begin with a consideration of the general scheme by which the two major biopolymers are constructed by living organisms. These are the proteins and the nucleic acids, and a sequence is shown schematically in Figure 1, beginning with the DNA double helix of the cell which can in some way (which is still not clarified) reproduce itself. That is, it can copy a particular sequence of nucleotide monomers in some particular order, and make a complementary copy, giving rise to a double set. However, in addition to that, there is another type of reaction which occurs

inside the cell. Instead of copying the DNA into another DNA strand, part of the DNA (which is an enormously long polymer, made up of essentially four bases hooked together by ribose phosphate linkages) can be copied into a ribose phosphate polymer, with a series of slightly different bases but related to the initial sequence. That part which is so copied contains in it the "message" for a particular protein. That is why this type of mateial is called "messenger RNA". The messenger RNA then comes out of the nucleus and thus becomes the message for constructing a particular protein. The message is contained in the sequence of bases in the polymer. As you know, it takes three of these bases to designate a particular amino acid. The particular "messages" are then hooked onto the catalysts, which are themselves made up partly of protein and partly of nucleic acid and some lipid substances, and the situation is now ripe to put together the amino acids into a peptide. In order to do that, the amino acids must be suitably prepared. Each one of these different length straiht lines in Figure 1 represents a different amino acid. The wiggle line represents the amino acid, acylated with a phosphate residue (amino acyl phosphate or amino acyl adenylate). This, then, is handed over to a small piece of polymer (RNA) constructed of about 70-100 monomeric units, containing in it somewhere a certain triplet (three bases in a row) which is characteristic of a particular amino acid; this distinction is designated by the different symbols in Figure 1, representing a different sequence of the three bases (circles, lines, crosses, squares). Thus, each activated amino acid is loaded onto the terminus of a specific small oligomer (small polymer of the order of 70-100 units). We now have the amino

acid loaded onto a particular transfer RNA (tRNA) with a particular triplet of bases characteristic of that amino acid. These now are free to attach themselves to the messenger RNA, which will be done at points in the messenger which are complementary to the three bases of which they are composed, thus providing the order of the amino acids which go into the polymer to make the polypeptide. The polypeptide, then, is "zipped up" to construct the protein. The protein is, in general, itself a characteristic polypeptide, having a characteristic shape and structure, which will be discussed later.

All of these reactions -- the copying reaction (replication), the transcription into RNA, and the activation, the loading reaction, and the zipping up on the ribosomes -- are dependent for their specificity in part at least on the proteins which are themselves thus made.This is a "chicken and an egg" problem today, and it sets the stage for the next part of the discussion.

How did this reflexive system, which is represented diagramatically in Figure 1, arise? As chemists, we would like to see if chemistry could have designed such a system and put it together in just this way. This is, of course, what I mean by the nonbiological origin of the polymers and what I mean here by the origin of the code. How did this relationship between particular triplets and particular amino acids come into being?

I believe this relationship is the result of chemistry and not a "frozen accident" as some of the biologists believe,. The evidence for this is yet, however, to be forthcoming in the laboratory. I think that most of you as polymer chemists will find this a rather amenable point

of view, and you will, I hope, turn to trying to devise experiments which will show there is some reason for this triplet corresponding to this particular amino acid and not to some other one. That is something which remains yet to be definitely shown in the laboratory and is a subject of considerable controversy.

Returning then to possible ways in which the proteins and nucleic acids might have come into being in a nonliving world: The world with which we deal is a world in water; it is not a world in methylene chloride, dioxane, DMF, or other solvents. And all of these reactions which I have described to you in Figure 1 occur in water. I feel that whatever reactions we use, whatever reactions we call upon, to do the originating of the protein and the nucleic acid should be essentially in an aqueous medium to begin with. This can be modified somewhat by the resulting polymer itself in terms of the protein-lipid interaction. The reactions do, however, occur primarily in an aqueous environment.

We set out to try and generate polypeptides and other polypolymers in an aqueous environment.[1] Chemically and thermodynamically this is rather a difficult operation. The formation of the polymer from the monomer of each one of the biopolymers I have mentioned (proteins, nucleic acids, carbohydrates, lipids) is the result of a dehydration condensation. To make a dehydration condensation occur in water is a tricky operation. Living organisms have long ago learned how to do this, but the question is could we, as chemists, do something similar. The reactions which must be accomplished in this aqueous environment are shown in Figure 2. You can see that the formation of proteins and peptides makes use of a bifunctional molecule, and from that molecule a water molecule can be

eliminated, leaving a bifunctional dimer which, in turn, can grow at either end. Formally, the same kind of system occurs in each of the four cases -- proteins, polysaccharides, lipids, nucleic acids. In each case, the water molecule is eliminated to result in another bifunctional molecule. This type of operation can go on continuously. With the lipids, however, this same type of reaction is not subject to as extensive extension as with the proteins or polysaccharides; with the elimination of water the ester bond is formed, which can go on. This is, however, not quite in the same category but it is possible to make fairly large molecules.I discuss this now because the lipids play an important role in the whole evolutionary scheme.

Figure 3 shows the methods for the formation of the nucleic acids by dehydration condensation reactions. Here, actually, there are three points of dehydration involved. The first involves the formation of the amino glycoside on a base, using the NH group of the base with the glycosidic semiacetal hydroxyl of the polysaccharide. The second is the formation of the phosphate ester with the primary alcohol, forming the terminal 5'-phosphate. In this case the ribose sugar is shown which forms RNA. (The DNA, of course, is the material in which the 2'-carbon lacks the hydroxyl, i.e., carries two hydrogen atoms). The third dehydration condensation occurs between the monophosphate ester and the secondary alcohol on the 3'-carbon of another nucleotide. The result is a bifunctional molecule, with both functions available; at one end the phosphate ester is available and at the other end the 3'-alcohol is usable. The nucleic acids are built on the same principles as the two other major biopolymers shown in Figure 2.

We asked the question: Is there any way to generate the polypeptides? There are many reactions for the construction of polypeptide linkages which would work in an aqueous environment. This required only a small modification of existing knowledge of reactions. One of the important reagents that has been used for the dehydration condensation of a variety of functional groups is the carbodiimide (R-N=C=N-R). The carbodiimide structure contains in it the ability to remove water, in stages, by more or less specific reactions with acid groups such as carboxyl, phosphate, and certain alcoholds. This leads to condensation reactions.[2,3] This particular reagent (carbodiimide) has been used widely for nucleotide and polypeptide synthesis, but generally not in aqueous environments. We wanted to change the R groups, to make them useful in aqueous environments. The R groups can be made hydrophilic so that the carbodiimide can be dissolved in water. This is useful since the carbodiimide does nct hydrolyze very rapidly directly but does so by virtue of its reaction with the carboxyl group and then with the phosphate.

The next question arose: How could this type of reaction occur in a natural environment? It turned out that there was a very easy way in which this particular structure could have evolved in a natural environment. As you may recall, ammonia was one of the primary molecules on the earth's surface, and HCN as well. With reactions among these two materials it is possible to make cyanamide, probably through cyanic acid or hydroxylamine. This is a tautomeric form of a carbodiimide. The carbodiimide in that tautomeric form is not stable; it tends to dimerize, making the dicyandiamide (DCDA) which is the common form for cyanamide. It also contains

the same type of pi-bonds as carbodiimide. We used the DCDA as one of the starting materials to demonstrate this type of dehydration condensation reaction for biopolymeric materials. We took an amino acid, adjusted the pH to slightly acid conditions, and added the water-soluble carbodiimide material (dicyandiamide) to see if we could produce polypeptides in the homogeneous aqueous environment. The results of some of these experiments are shown in Figure 4.[4] We started with glycine; the reaction was performed a out pH 3.5, the glycine disappears, and the dimer (diglycine), trimer (triglycine) and tetramer (tetraglycine)come out. This is a very slow reaction, and it does not go very far. You can hardly call the tetraglycine a "polymer", but at least the reaction does occur in aqueous conditions.

The next question was: Is there any selectivity of this reaction? Do the amino acids self-select each other when they condense? This is a natural question to ask in the ordinary course of evolutionary studies. About five or six years ago, Steinman, in trying to answer this question put amino acids acid (B) on polymer beads, through the carboxyl group and used an amino protected amino acid (A) to couple to it. He measured the relative rates of coupling of amino acid (A) in the homogeneous environment to amino acid (B) attached to the bead; this is a model, of sorts, of polypeptide formation. Table 1 shows the relative rates of coupling in this series, and you can see that the rates vary over a factor of ten.[5] This is the measured coupling efficiency, and the calculation is based upon the frequency of a particular pair in existing proteins, (the frequency of phenylalylglycine, glycylphenylalanine, or isoleucylglycine, etc.) as they exist in today's proteins with a statistical analysis of

the situation, normalizing the frequencies to glycylglycine as unity. There seems to be slight relationship between what was found for the efficiency of coupling in the heterogeneous system and the occurrence in nature of these particular pairs of peptides. This idea has been developed still further, and there are now better ways of studying the possible self-selection of amino acids in peptide formation.

There are two other methods by which polypeptides have been made, one of which is in a nonaqueous environment. Table 2 shows the results of the experiments of Sidney Fox in which he took molten glutamic acid, containing a small amount of pyrophosphoric acid, as a solvent (hardly an aqueous environment) and put in an equimolar mixture of all the other amino acids, other than aspartic and glutamic.[6] Here, again, you can see that the incorporation into the resulting polypeptides is not statistical. The point of these data is to show the variabilities of occurrence, ranging from 0.5% to 5%, and the selectivity is evident.

The final study to which I want to call your attention is, perhaps, the most elegant, and I don't know whether it will continue. This is the work of Aharon Katchalsky using amino acyl adenylates as his starting point and using montmorillonite clay as the catalyst, both of which are easily available in nature. The amino acyl adenylate is available in nature because the adenylic acid can be generated by ordinary chemical reactions and the coupling of the amino acid to the adenylic acid, linking a phosphate and a carboxyl anhydride, can be easily achieved using the carbodiimide dehydration condensation reaction in aqueous environment. So the system with which Katchalsky began is achievable biologically. I do want to describe Katchalsky's discovery because it is the beginning of a very interesting idea.

The amino acyl adenylate used by Katchalsky is the activated amino acid in the first stage of normal biological amino acid activation; it is usually a part of an enzyme complex but, nevertheless, it is a good beginning for this type of study. He first tried to polymerize this material in an aqueous environment in a homogeneous solution. The amino acyl adenylate has in it the mixed anhydride of the carboxyl group and the phosphoric acid. It also has the free amino group of the amino acid itself. Thus, the free amino group can act as a nucleophile on the carboxyl group of the carboxy anhydride of the amino acyl adenylate.

Figure 5 encompasses the basic idea that Katchalsky introduced, but it shows more as well. If you look at the amino acyl adenylate skeleton on the polynucleotide chain you can see that the amino group can act as a nucleophile on the acyl group of the anhydride (reaction a) which would leac to the formation of a polypeptide (poly)adenylate and free adenylic acid.The alternative reaction, which Katchalsky considered but had not published on with any experimental information, is that in which the secondary hydroxyl of the sugar wi l attack, again, as a nucleophile on the phosphate of the phosphate anhydride (reaction b), resulting in the reverse type of polymerization, the products being a free amino acid and the (poly)amino acyl polyadenylate. There are, of course, alternative reactions which can occur (rearrangements, etc.) to stop or slow down the polymerization reactions.

What Katchalsky found in his report of 1970[7] and which he reported in more detail last May in Gottingen just before he died is that in homogeneous solution the reaction is slow and the polymers which are formed are not very large (8-10 units). He found, however, that if he performed

this reaction in the presence of a properly prepared montmorillonite, the whole system changed. The reaction became extremely rapid and the polypeptides went to fairly high degrees of polymerization -- high for this kind of a reaction, that is.[8]

We have attempted this same type of reaction, and indeed, he and I were moving in this same direction, making both polypeptide and polynucleotide from the same starting material. We can now begin to see the real relationship between the polynucleotide and polypeptide, or the nucleic acid and the protein, and how they can arise as a result of the stereochemistry of the reaction.

Katchalsky found that there were a number of discrete polypeptides formed. The result is given in Figure 6, which is a chromatogram of a second fraction. Table 3 also gives more information on the two principal components which were in the initial extract of this reaction. Fraction I is mostly the adenylic acid and three groups of polypeptides. This was a one gram reaction which resulted in 450 mg of adenylic acid, 20 mg of a 1000 MW substance, and another 20 mg of a 2000 MW substance. Fraction II, however, contained mostly polypeptides with some adenylic acid. The gel permeation chromatography of fraction II is shown in Figure 6.[7] You can see the presence of some eight components, with the largest one being adenylic acid, and components 1,2,3 and 4 are polypeptides; there are two peaks which were not identified which were inorganic salts and two more peaks, not identified, were organic in nature. When I spoke to Aharon about it he said he thought they might be polynucleotides, and I think they are; however, they have not yet been identified.

Perhaps in the future someone else will carry this work forward: I think it represents an extremely important development in the nature of the polymerization reaction. Why is it that you do not get a distribution; why do you get discrete groups of polymers? I think these discrete groups occur because it is not a single addition reaction. Some amino acids are being added, and some polynucleotides are being constructed to make a polypeptide-polyadenylate, and these materials are regulating the size of the polypeptides which are found. It will probably turn out that the polynucleotide is also discrete in some fashion.

THE ORIGIN OF THE CODE

As perhaps some of you are aware, in today's living organisms the specification of the linear array of amino acids in a polypeptide is contained in a corresponding linear array of bases in a polynucleotide. A series of three bases in the polynucleotide (DNA) specifies a specific amino acid in a polypeptide, or protein. This correspondence of a triplet of bases with a particular amino acid is universal in all living organisms. One form of expressing this is shown in Table 4. Here you can see there is a certain amount of redundancy in the code, but the universality of it throughout the living world is now fairly well established.

Thus, in addition to the question of how these biopolymers may actually have been formed abiogenically we must address ourselves to the question of how the linear relationship between the two major biopolymers, that is, the polypeptide (protein) and the polynucleotide (DNA), evolved. That is: What is the possible origin of the code?

There are those who believe that the particular specificity we now find in living organisms between a particular triplet and a particular amino acid as assembled in Table 4 is the result of an accident in some early catalytic reaction which, by virtue of a selective advantage in producing autocatalytic systems, has now been frozen into all of biology.[9] It is selected because a single autocatalytic system eventually dominated all the others. Another way of expressing this "frozen accident" notion is to make the statement that given exactly the same starting conditions and allowing chemical evolution to occur once again, that a different codeal relationship might very well have appeared.[10,11] It is my personal belief, however, that this would not be the case. I believe that the codal relationship reflects some characteristic molecular interactions between amino acids (polynucleotides) and nucleotides (polynucleotides).

A modicum of evidence for this already exists in the form of some experiments done on the rate of coupling of amino acids to nucleotides in a specific instance. Some years ago[12] we attached a nucleotide (in this case adenylic acid) to a synthetic polymer (in this case polystyrene). The polystyrene was in the form of microspheres. The next step was to measure the rate of coupling of two different amino acids to each of these two polystyrene-nucleotide preparations. The coupling was performed using an N-protected amino acid adenylate. The reaction is shown in Figure 7 and the results are given in Table 5.[13] Here you can see that glycine reacts more rapidly with both nucleotides than does phenylalanine and adenine reacts more rapidly with both amino acids than does cytosine. The overall range of reactivity is roughly a factor of three, the slowest reaction being phenylalanine with cytosine and the most rapid the reaction of glycine with adenine.

Here the beginning of a kind of selectivity in a reaction rate is apparent.[14] However, I believe that this kind of selectivity will be very much enhanced if both the amino acid and the nucleotides are each, respectively, part of a polymer. For example, I would expect that the alternative reactions (a) and (b) indicated in Figure 6 will be very dependent upon the nature of the amino acid and the nature of the nucleotide involved. Furthermore, I would expect that selectivity would be increased with the length and character of the polynucleotide, on the one hand, and possibly even with the polypeptide on the other. This last experiment is yet to be done, but is underway.

SECONDARY AND TERTIARY STRUCTURE OF PROTEIN AND INTERACTION WITH NUCLEIC ACIDS

This part of our discussion will describe what is known about the secondary, tertiary and quaternary structures, mostly of proteins. There are several principles which I want to emphasize here: One is the way in which proteins fold up, in their secondary and tertiary structures, another is the way enzymes interact with substrates, and a third is the way in which the proteins interact with the nucleic acids. (Finally, I wish to bring up the protein-lipid interaction, Part III.)

Myoglobin, with its tertiary structure, is shown in Figure 8. The left-hand side of the figure shows the alpha helical structure and the drawing on the right-hand gives a more three-dimensional view of the same material, to show that the myoglobin is a folded protein, with the heme stuck in the center. This secondary and tertiary structure is built into the molecule as a result of the amino acid sequence in the polypeptide which takes up this structure, given the opportunity in

water.[15] The structure of cytochrome c, showing the hydrophobic internal arrangement and hydrophilic external arrangement, is shown in Figure 9. The colors on this figure are classified according to the hydrophilic and hydrophobic characters of the amino acid residues. By observation, you can see that the hydrophobic parts are more or less in the middle and the hydrophilic (the cool ones) parts are on the outside of the molecule. This is the only characteristic which has so far approached a generalization on the structure of soluble proteins. In general, the structure of soluble proteins is such that when they fold up in their secondary and tertiary structures they do so in a way which places the hydrophobic chains on the inside (away from the water environment) and the hydrophilic chains on the outside, near the water environment.

It is for this reason that I had the reservation, which I mentioned earlier, about the generation of polypeptide and polynucleotide structures in a purely aqueous environment. You see that there can be modifications during the course of their construction which will remove their functions from the water environment into a nonaqueous environment, by virtue of their own structure, so to speak.

The structure of lysozyme, showing the substrate cleft, is given in Figure 10. I believe it is easy to distinguish between the lysozyme itself and the shaded area which is the synthetic disaccharide substrate which fits right into the active site cleft. The same molecule, lysozyme, is shown again in Figure 11, illuminating the mechanism of the polysaccharide hydrolysis, and showing the actual function of the active site. You can see the polysaccharide lying in the active site of the lysozyme.

The bond which is hydrolyzed is the glycosidic linkage, by means of transfer to the polymer (Asp 52) of one part of the substrate, thus letting the other part go free. The Asp glycosidic ester, in turn, is very rapidly hydrolyzed. This is the way in wh ich the hydrolytic enzymes and many of the transferring enzymes function. Thus, the tertiary structure is evolved to give the particular tertiary architecture which will take hold of the proper substrate.

The next figure (which is an old one) reaches into the area of visibility. This is collagen (Figure 12), showing that we can refold molecules into tertiary structure but, even further than that, the molecules can actually reassemble into more complex quaternary structures which are dependent upon their primary sequence.[16] The upper part shows the separated collagen fibrils, and in the bottom sequence are the reaggregated fibrils which appear similar to the natural ones. The fibrils are highly ordered arrays and this is a protein-protein aggregation. The next few figures show the dissociation and re-aggregation of the tobacco mosaic virus (a relatively simple virus particle) which consists of a set of identical protein molecules and one nucleic acid molecule. These two can be separated and then reassembled again, into a complete virus particle.[17] Since that time, more complex virus particles have been reconstituted as well. Figure 13a shows the native TMV virus, with the uniform particles, and Figure 13b shows the reassembly of the TMV protein itself alone. The protein is separated from the nucleic acid, redissolved and by changing the salt concentration so the material will reaggregate, with the results shown here. You can see that the material has reaggregated with the correct dimensions in one way, but not in the other; particles have all different lengths, in contrast to the particles of the native TMV virus; because there is no nucleic

acid "information" to give the correct dimensions to the particle. In Figure 13c you can see the reconstituted TMV virus particle, which occurs when the nucleic acid component is put back into the protein solution. This same type of experiment has been done with much more compleparticles, T4 bacteriophage, MS2 phage protein, etc. Progress in the reassembly of very complex structures is being made.[18] The TMV particle itself is shown diagramatically in Figure 14. The tertiary structure of the subunit protein material in the TMV virus particle has not yet been established, but we know that there is a sequence of 120 amino acids and we know which ones they are, but, as yet, we do not know the detailed structure of this subunit.

INTERACTION OF PROTEIN AND LIPIDS

Protein-lipid interaction is one of the areas of very high scientific activity, in the technological polymer area and more selectively in the laboratory than biochemists have done in the past. Lipids have been used for handling biochemical proteins for some time, The present concept of membrane structure, showing the protein embedded in the bilipid membrane, is given in Figure 15. The phospholipid consists of hydrophilic ends (the phosphoethanolamine end represented by spheres), and the lipid chains. The globules which are membrane proteins in various stages of insertion into the bilipid membrane make the total actual biological membrane itself. These proteins may be structural or they may be transport proteins , i.e., those carrying material into the cell through the biological membrane. The fact that many of the proteins are directly associated with lipids implies that there must be hydrophobic connections on the outside of the proteins in order to be able to perform this type of insertion. This has not yet been established in the laboratory,

and I will give you more evidence for that guess later on. There are clearly hydrophobic parts on the proteins, but whether they are on the outside or the inside, or both, is a subject of some concern.

A synthetic "biological cell" is shown in Figure 16. It is made by shaking up a phospholipid (lecithin) with a protein cytochrome to form the liposomes (soft spheres). The multiple and single bilipid layers are visible in this electron micrograph, and these layers are filled with protein.[19] This work was done about eight years ago in a laboratory in GreatBritain. I have called the spheres "Bangasomes" after the man who constructed them first.[20] The spheres are called generally liposomes, and their osmotic properties, transport properties, and a variety of other properties resemble those of living cells. How do materials get in and out through that membrane? What is the mode of their transport. Can specific proteins be associated with the membranes to produce these specific results? These are questions still unanswered. Studies are now in progress, and this field is really just blossoming out as a new area and a new era. Only now have the polymer physical chemists gotten into the act, and Midland Macromolecular Institute is one of the leading sites of this type of research.

As you know, it is quite common to use detergents, anionic, cationic and non-ionic, to try and help in the isolation of a particular enzyme from a living cell. The procedures for the soluble proteins (soluble enzymes) are now well estsablished. The isoluble enzymes, however, are a different story. The membrane-attached enzymes, as the membrane of the cell, or the membrane of the mitochondrion,

or the membrane of other internal cellular structures, present more difficulties. It is necessary to use detergents of various types for enzyme extraction. A few years ago, we naively (and this quality has some merits as well as demerits) began this type of work. We had to learn that in order to extract the enzyme in which we were interested (which happened to be the RNA-instructed DNA polymerase, RIDP) detergent, was required. I wondered why it was necessary. We also learned that when the detergents were eliminated, the enzyme itself seemed to be eliminated, or at least the activity of the enzyme was eliminated. This gave rise to the idea that perhaps the activity of the enzyme was dependent upon the presence of detergent molecules, i.e., the lipid component in some way was changing the conformation of the enzyme and inducing its activity, or at least raising it. Therefore, the residual enzyme activities which are observed when synthetic detergents were <u>not</u> used for extraction were simply due to the natural phospholipid which came out when the solution was sonicated.

We found that when we added small amounts of certain types of non-ionic detergents it was possible to get the enzyme activity back again. This constitutes a new development. No one really believed that the intrinsic activity of the enzyme was dependent upon its association with a detergent molecule. Up until now it has been considered that the detergent molecules weresolubilizing entities to take the enzyme out of the lipid membrane. The detergent does, in fact, take the enzyme out of the lipid membrane, but large amounts of the detergent are not needed to make the enzyme active.

The activation of the RIDP enzyme by detergents is shown in Figure 17. RIDP is the enzyme which copies RNA into DNA, which is popularly known as "reverse transcriptase activity". Work on this

enzyme is one of the major breakthroughs of the last two or three years, particularly in connection with tumor viruses. The activity of the enzyme is increased by addition of suitable amounts of various types of non-ionic detergemts.[21,22] If you translate the effect into a molecular basis, this is a positive molecular interaction. This result occurred as a technical development of how to study the enzyme, but we did all the work on the detergents more or less as a side issue. However, this turned out to be the central theme, which is not uncommon in this type of work. We were actually studying the inhibition of the RIDP enzyme activity by certain drugs with a view toward using those drugs possibly to inhibit the transformation of normal cells into cancer cells by the viruses which carry this enzyme. The RIDP enzyme copies the RNA from DNA in the virus, which copy is then insert ed into the DNA of the cell, thus making a transformed cell. The rationale was to find a drug which would inhibit the RIDP enzyme and thus block the transformation into cancer cells. In order to study the inhibition of the enzyme, we had to study inhibition of the enzyme activity under many conditions. This gave rise to our studies on detergent effect on enzyme activity. We found that we had some excellent enzyme inhibitors in the form of the drug itself. This particular drug, rifampicin and its derivatives is a lipophilic substance. Here, it turns out that if there is too much detergent the drug is ineffective in its inhibition of RIDP enzyme activity. The structure of the non-ionic detergents which we used in this study is shown in Figure 18. Some of the detergents (the Tritons) are aromatic and others are not. This fact is important. The effect of three different detergents on the ability of a certain drug, which has a hydrophobic tail and a hydrophilic end, to prevent

the enzyme from working. The enzyme is shown in Figure 19. We have tried to plot the critical micelle concentration for each of the detergents, and you will note that the removal of the drug from its ability to inhibit the enzyme is dependent upon the formation of the micelles.[22] We interpret this to mean that if there is too much detergent present there is micelle formation, and the micelles dissolve the drug, taking it away from the hydrophobic portion of the enzyme. The drug no longer can act on the enzyme, so the enzyme returns to its full activity. In Figure 20 are two magnifications of a tissue culture which has been transformed into malignancy. You can see the growth of the cells to confluence here, but wherever the cells have been transformed into malignancy they overgrow each other, and they make little "piles" called foci. We have found that by suitable adjustment of the drug concentration it is possible to prevent the virus from transforming cells as determined by this focus formation inhibition.[23] We have gone even further with the drugs to show we can actually prevent the formation of tumors in whole animals.[24] Again, this requires a suitable combination of detergent and drug to prevent the detergent from spoiling the activity of the drug.

I hope that some of the things described above will give you some concept of the kind of a china shop you could get into, if you are willing to do it. I want to give you who are not in the biochemistry-molecular biology business some idea of the way in which a polymer chemist could effect the development of our fundamental concepts of the nature of life, and how it came about, and its application in the problems of the day.

References

1. M. Calvin, "Chemical Evolution", Oxford University Press, Oxford England (and New York), 1969, pp. 152-155.

2. ibid, pp. 161-169.

3. D. H. Kenyon and G. Steinman, "Biochemical Predestination", McGraw-Hill Publishing Co., New York (1969), pp. 182-192.

4. D. H. Kenyon, G. Steinman and M. Calvin, Biochim. Biophys. Acta, 124, 339 (1966).

5. G. Steinman and M. N. Cole, Proc. Nat. Acad. Sci. US, 58, 735 (1967).

6. K. Harada and S. W. Fox in "Origins of Prebiological Systems and Their Molecular Matrices", S. W. Fox, ed., Academic Press (1965), p. 296.

7. M. Paecht-Horowitz, J. D. Breger and A. Katchalsky, Nature, 228, 636 (1970).

8. M. Paecht-Horowitz and A. Katchalsky, J. Molec. Evol., in press.

9. M. Eigen, Die Naturwissenschaft., (1971).

10. F. H. C. Crick, J. Mol. Biol., 38, 267 (1968).

11. L. E. Orgel, J. Mol. Biol., 38, 381 (1968).

12. M. A. Harpold and M. Calvin, Nature, 219, 486 (1968).

13. M. A. Harpold and M. Calvin, submitted to Biochim. Biophys. Acta.

14. M. Calvin, Proc. Roy. Soc. Edinburgh, 70, 273 (1969).

15. For a discussion of the principles involved in three-dimensional structure and self-assembly, see M. Calvin, "Chemical Evolution", pp. 187-193.

16. For a discussion of the principles involved in quaternary structure, see M. Calvin, "Chemical Evolution", p. 196.

17. For a discussion of reassembly, see M. Calvin, "Chemical Evolution", pp. 216-218.

18. D. J. Kushner, Bacteriol. Rev., 33, 302 (1969).

19. D. Papahadjopoulos and N. Miller, Biochim. Biophys. Acta, 135, 624 (1967); D. Papahadjopoulos and W. Watkins, Biochim. Biophys. Acta, 135, 639 (1967).

20. A. D. Bangham, J. DeGier and G. D. Grevitch, Chem. Phys. Lipids, 1, 225 (1967).

21. F. M. Thompson, L. J. Libertini, U. R. Joss and M. Calvin, Science, 178, 505 (1972).

22. F. M. Thompson, L. J. Libertini, U. R. Joss and M. Calvin, submitted to Biochemistry.

23. M. Calvin, Radiation Res. 50, 105 (1972).

24. U. R. Joss, A. M. Hughes and M. Calvin, submitted to Nature.

Table 1

Comparison of Experimentally Determined Dipeptide Yields and Frequencies Calculated from Known Protein Sequences

	Values (relative to Gly-Gly)	
Dipeptide*	Experimental	Calculated
Gly-Gly	1.0	1.0
Gly-Ala	0.8	0.7
Ala-Gly	0.8	0.6
Ala-Ala	0.7	0.6
Gly-Val	0.5	0.2
Val-Gly	0.5	0.3
Gly-Leu	0.5	0.3
Leu-Gly	0.5	0.2
Gly-Ile	0.3	0.1
Ile-Gly	0.3	0.1
Gly-Phe	0.1	0.1
Phe-Gly	0.1	0.1

* The dipeptides are listed in terms of increasing volume of the side chains of the constituent residues. Gly, glycine; Ala, alanine; Val, valine; Leu, leucine; Ile, isoleucine; Phe, Phenylala ine. Example: Gly-Ala: glycylalanine.[5]

Table 2

Amino Acid Composition of Proteinoid Prepared with 200 ppa[6]

Amino Acid	%	Amino Acid	%
Thr	0.55	Lys	2.79
Ser	0.63	His	2.53
Gly	2.93	Arg	1.83
Ala	1.31	Total	7.15
Val	1.33		
Met	0.86	Asp	51.9
Iso	0.71	Glu	13.3
Leu	3.44	Total	65.2
Tyr	3.87		
Phe	5.87		
NH_2	5.02		
Total	27.6		

Asp:Glu: equimolar mixture of basic and neutral amino acid; ratio - 2:1:3; i.e., 33%, 16%, 50%; 100°C for 150 hrs. ppa = polyphosphoric acid in ppm.

DATA ON PEPTIDES I

Peptide Mol. Wt.	Degree of Polymerization	Wt. (mg)
640	9	10.3
1,120	16	35.2
1,900	27	20.0
Adenylic acid	1	470.5

PEPTIDES II OBTAINED ON POLYCONDENSATION OF ALANYL ADENYLATE

Peak No.	Mol. Wt. of Peptide Adenylate	Degree of Polymerization	Wt. (mg)
4	2,130	30	8
2	2,310	32	5.2
1	3,020	42	10.8
3	4,000	56	17.1
6	Adenylic acid	1	260.1

XBL 729-4742

TABLE 3

SECOND LETTER	FIRST LETTER A		FIRST LETTER C		FIRST LETTER G		FIRST LETTER U	
A	AAA	Lys	CAA	Gln	GAA	Glu	UAA	Termi-nate
	AAG		CAG		GAG		UAG	
	AAC	Asn	CAC	His	GAC	Asp	UAC	Tyr
	AAU		CAU		GAU		UAU	
C	ACA	Thr	CCA	Pro	GCA	Ala	UCA	Ser
	ACG		CCG		GCG		UCG	
	ACC		CCC		GCC		UCC	
	ACU		CCU		GCU		UCU	
G	AGA	Arg	CGA	Arg	GGA	Gly	UGA	Term.
	AGG		CGG		GGG		UGG	Trp
	AGC	Ser	CGC		GGC		UGC	Cys
	AGU		CGU		GGU		UGU	
U	AUA	Ilu	CUA	Leu	GUA	Val	UUA	Leu
	AUG	Met	CUG		GUG		UUG	
	AUC	Ilu	CUC		GUC		UUC	Phe
	AUU		CUU		GUU		UUU	

XBL729-4743

TABLE 4

PERCENT OF BOUND NUCLEOTIDE REACTED

Base / Amino acid	Adenine	Cytosine
Phenylalanine	6.7	2.9
Glycine	10.0	6.5

XBL 729-4754

TABLE 5

Figure Captions

PART I: FORMATION OF BIOPOLYMERS, PARTICULARLY PROTEINS

Figure 1 Mechanism of protein biosynthesis

Figure 2 Dehydration condensation of polypeptides, carbohydrates, fats

Figure 3 Dehydration condensation of polynucleotides

Figure 4 Homogeneous polypeptide formation by dicyandiamide

Figure 5 Polymerization of amino acyl adenylate, alternative reactions (Katchalsky)

Figure 6 Polypeptides resulting from alanyl adenylate on montmorillonite (Katchalsky)

PART II: ORIGIN OF THE CODE

Figure 6 The coupling of the polymer-AMP complex with the anhydride form of an N-protected amino acid

PART III: SECONDARY AND TERTIARY STRUCTURE OF PROTEIN AND INTERACTION WITH NUCLEIC ACIDS

Figure 8 Tertiary structure of myoglobin

Figure 9 Structure of cytochrome c showing hydrophobic internal arrangement and hydrophilic external arrangement

Figure 10 Structure of lysozyme showing substrate cleft

Figure 11 Structure of lysozyme showing mechanism of polysaccharide hydrolysis

Figure 12 Structure of collagen

Figure 13 a. Tobacco mosaic virus (TMV), native.
b. TMV protein, reconstituted
c. TMV, reconstituted.

Figure 14 Diagrammatic structure of TMV

PART IV: INTERACTION OF PROTEIN AND LIPIDS

Figure 15 Present concept of membrane structure showing protein embedded in bilipid membrane.

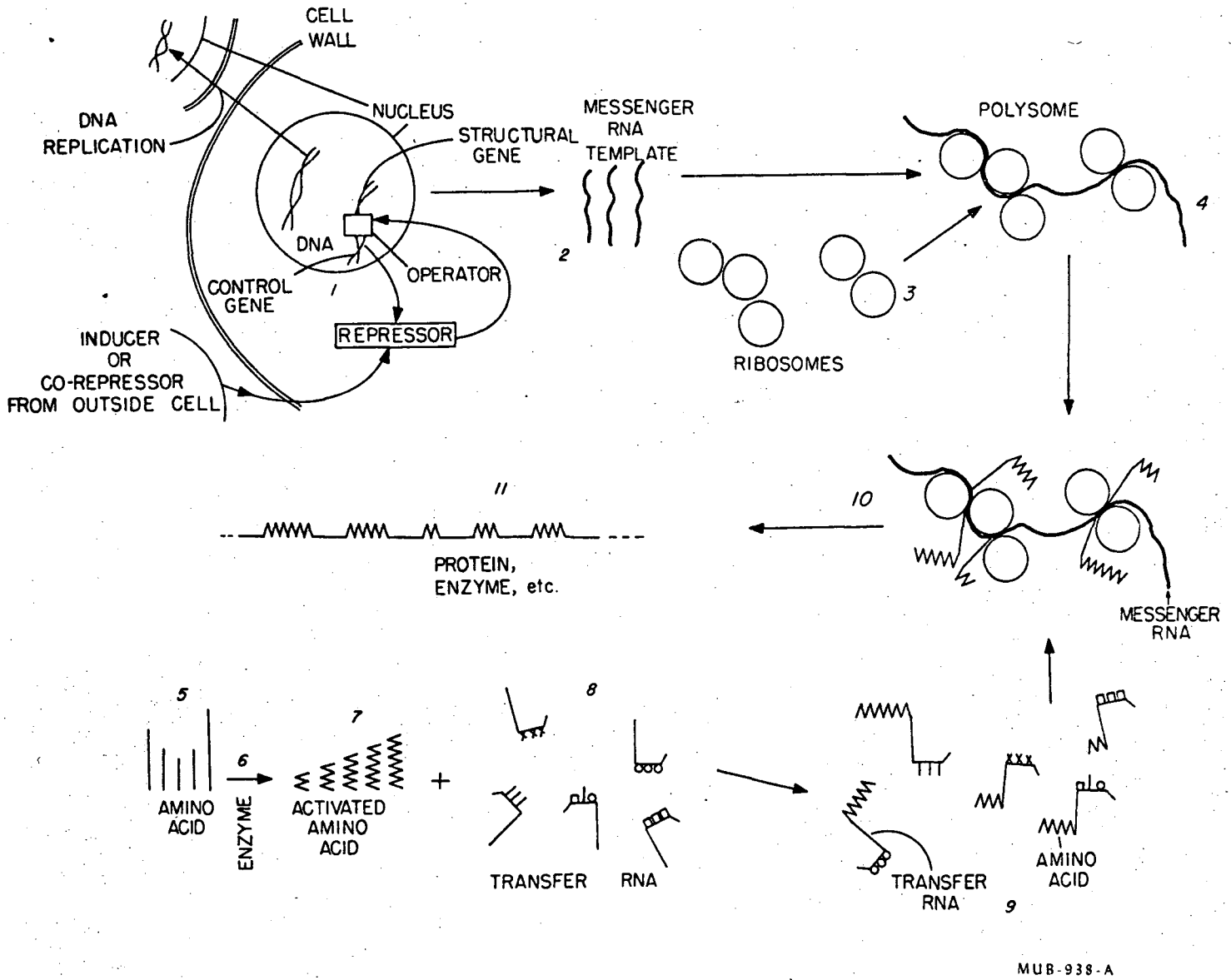
Figure 16 Liposomes of phospholipid with cytochrome (Horne & Watkins)

Figure 17 Activation of RNA-instructed DNA polymerase by detergent

Figure 18 Structure of detergents used in RNA-instructed DNA polymerase activation

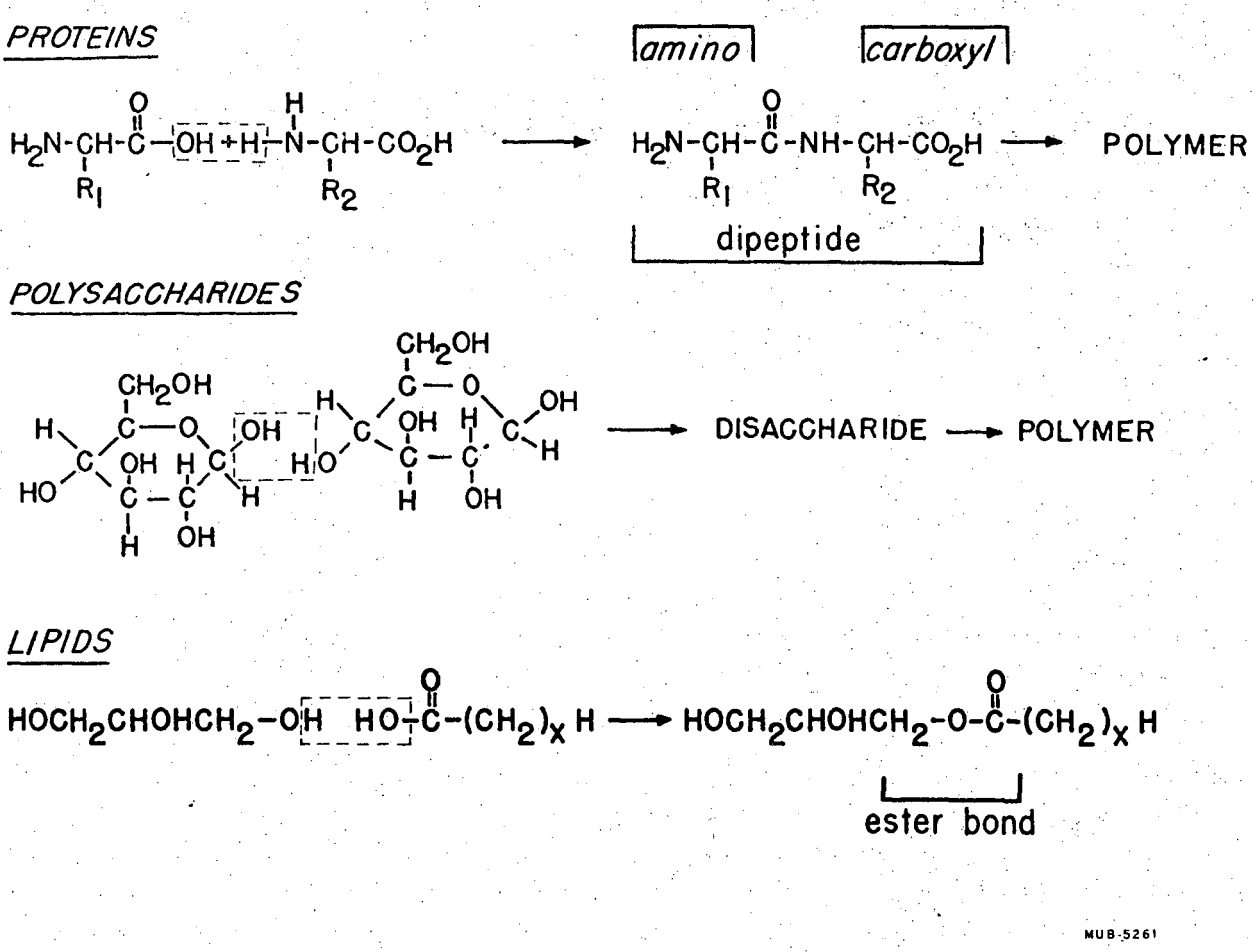
Figure 19 Suppression of drug effect on RNA-instructed DNA polymerase by detergents

Figure 20 MSV focus formation on Balb/3T3 cells



Calvin

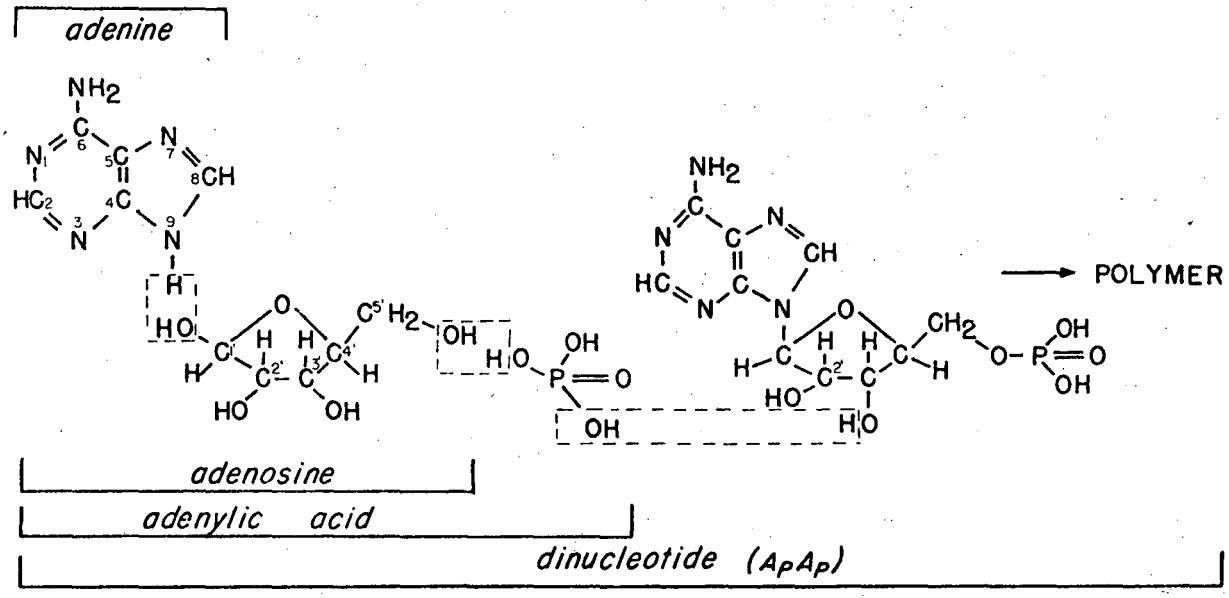
Figure 1



Calvin

Figure 2

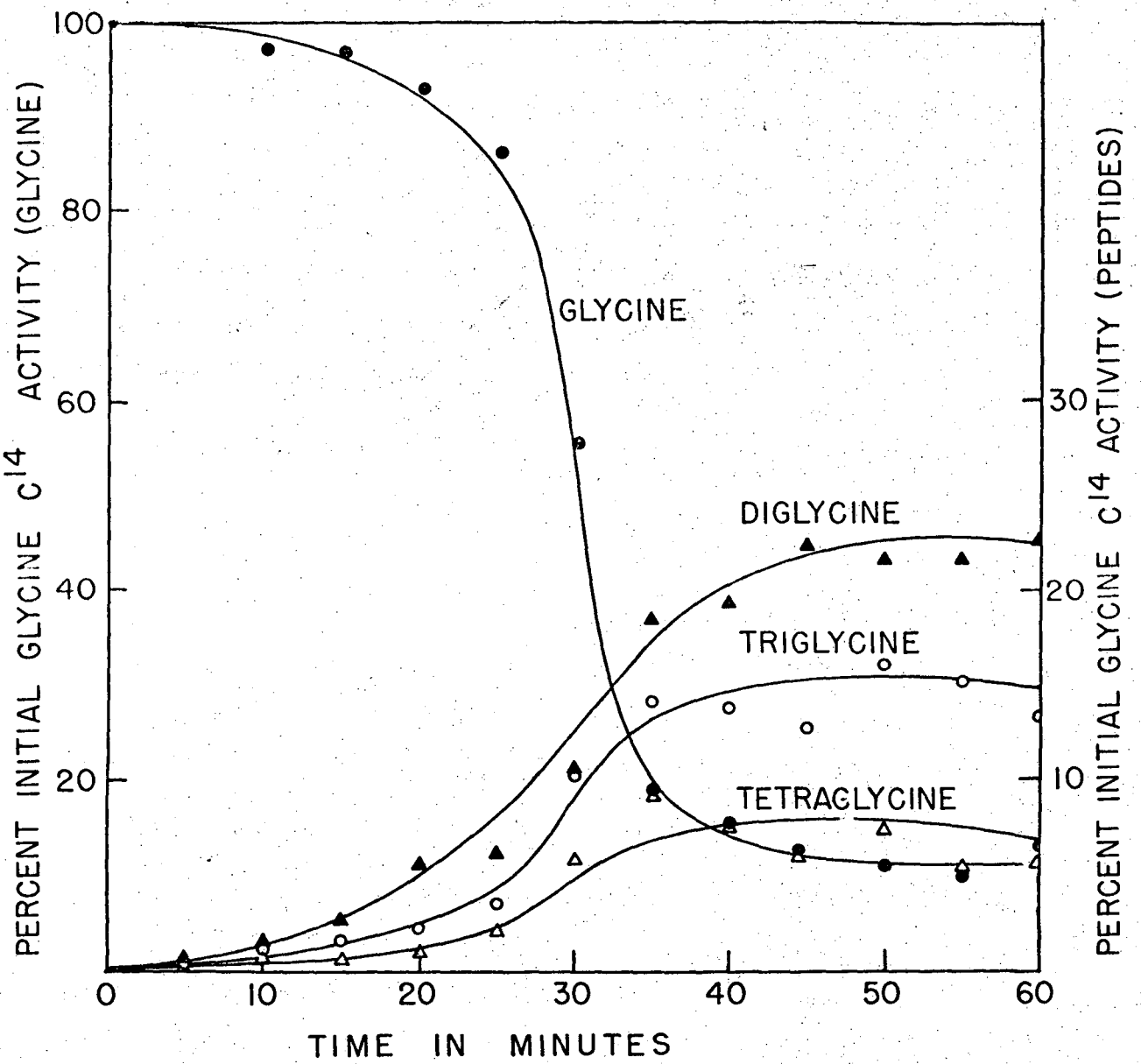
NUCLEIC ACIDS (3 STAGES) RNA SHOWN - DNA LACKS OH ON 2' POSITION



MUB-5262

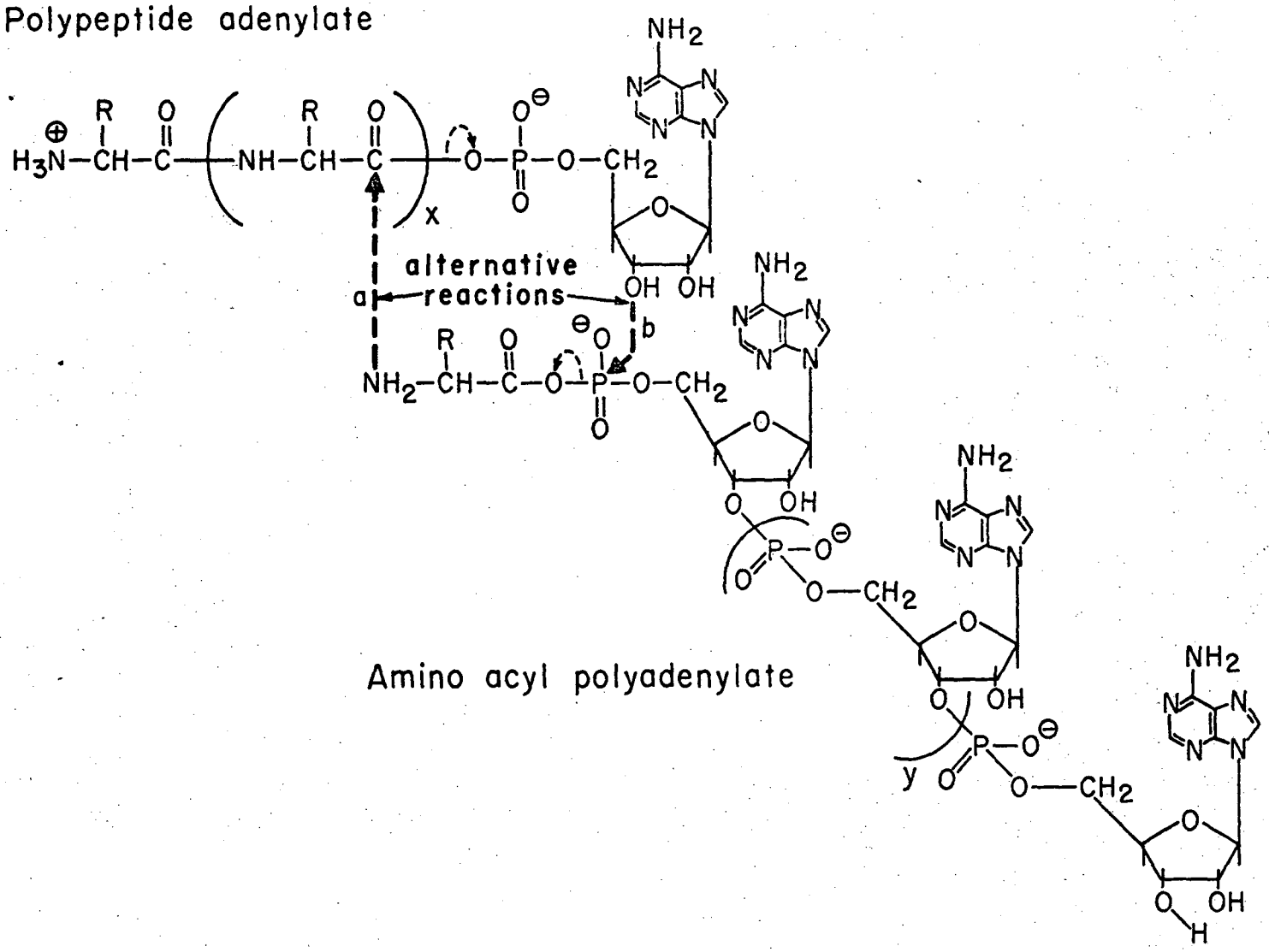
Calvin

Figure 3



Calvin

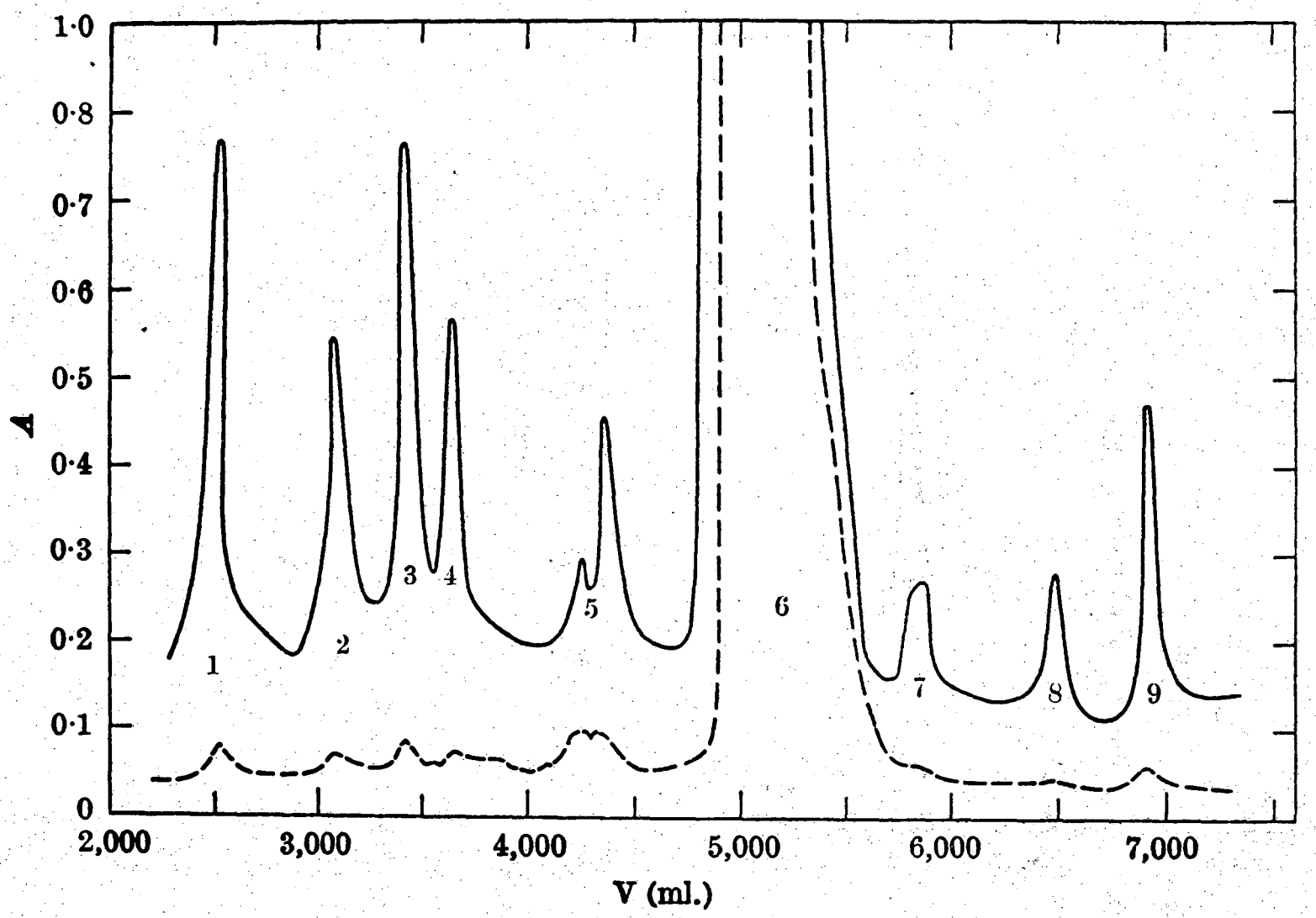
Figure 4



XBL729-4756

Calvin

Figure 5

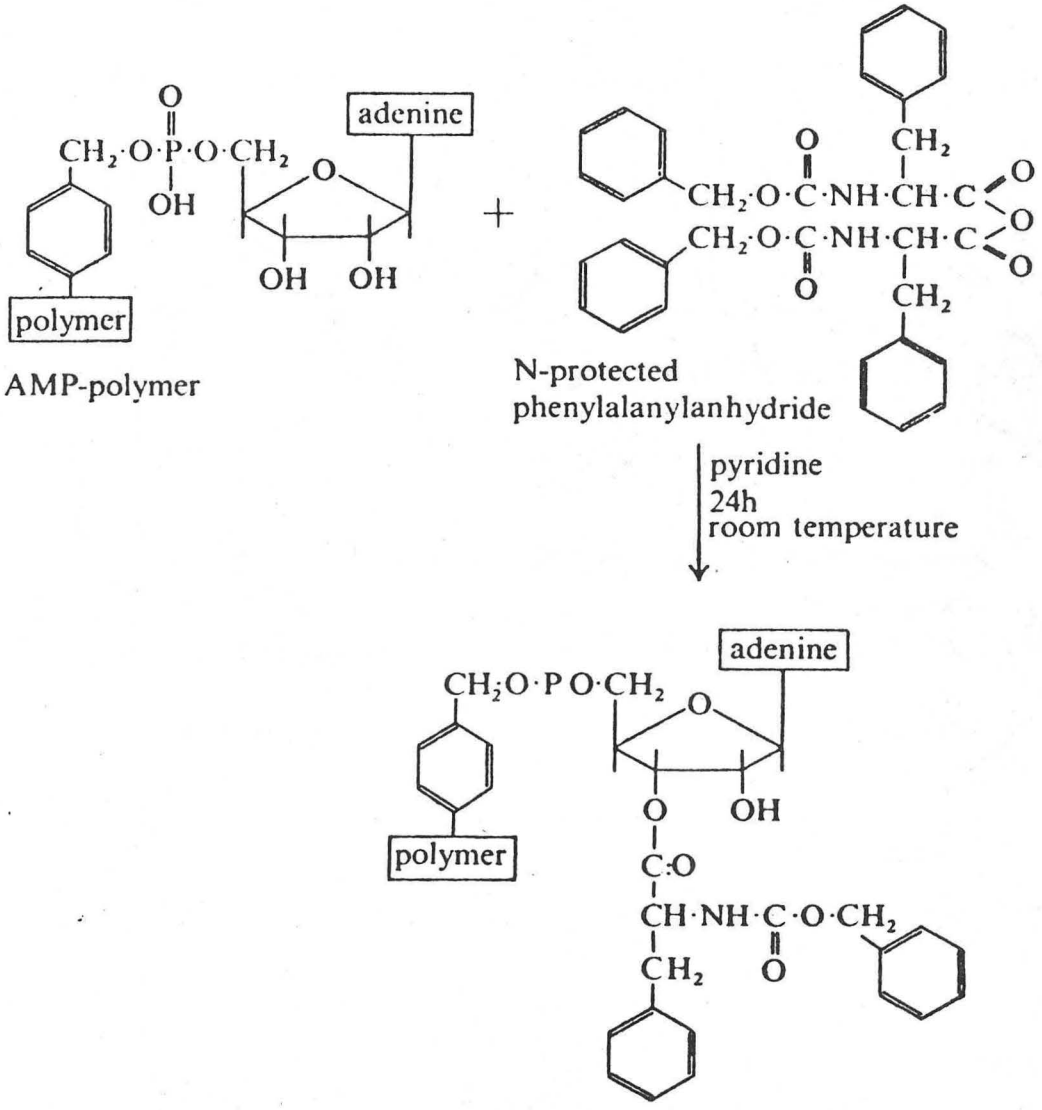


Chromatography of fraction II on 'Sephadex G-25', with gel bed 3 × 4 × 150 cm.

XBL 729-4745

Calvin

Figure 6



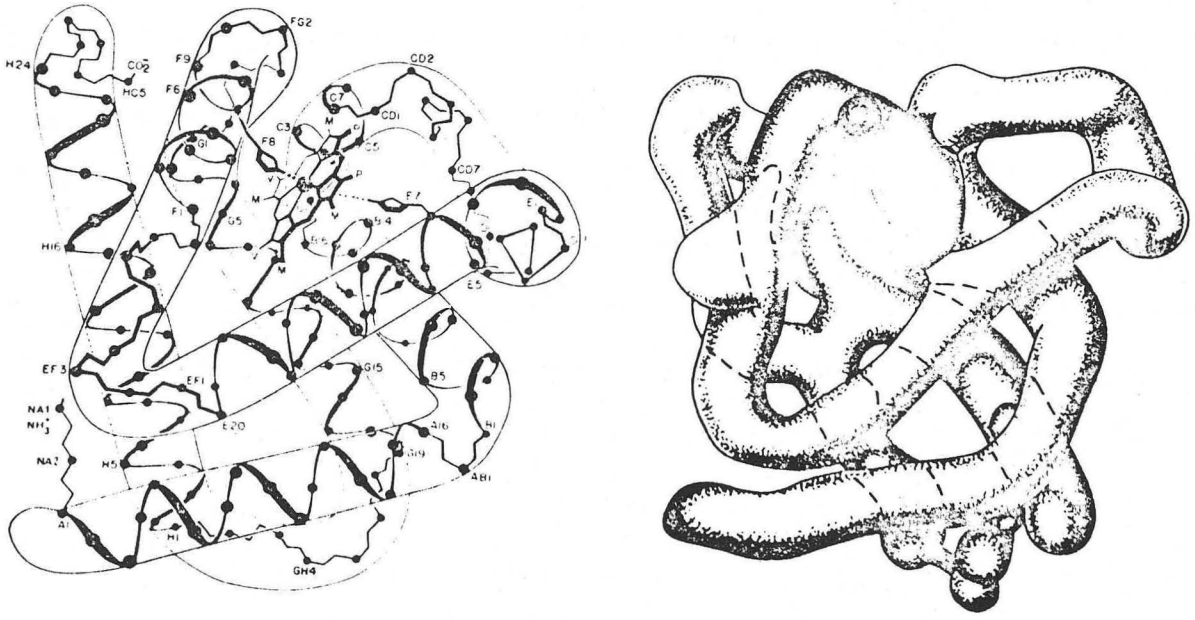
N-protected phenylalanyl -AMP- Polymer

The coupling of the polymer-AMP complex with the anhydride form of an N-protected amino acid.

XBL 729-4750

Calvin

Figure 7



XBL 727-4704

Calvin

Figure 8

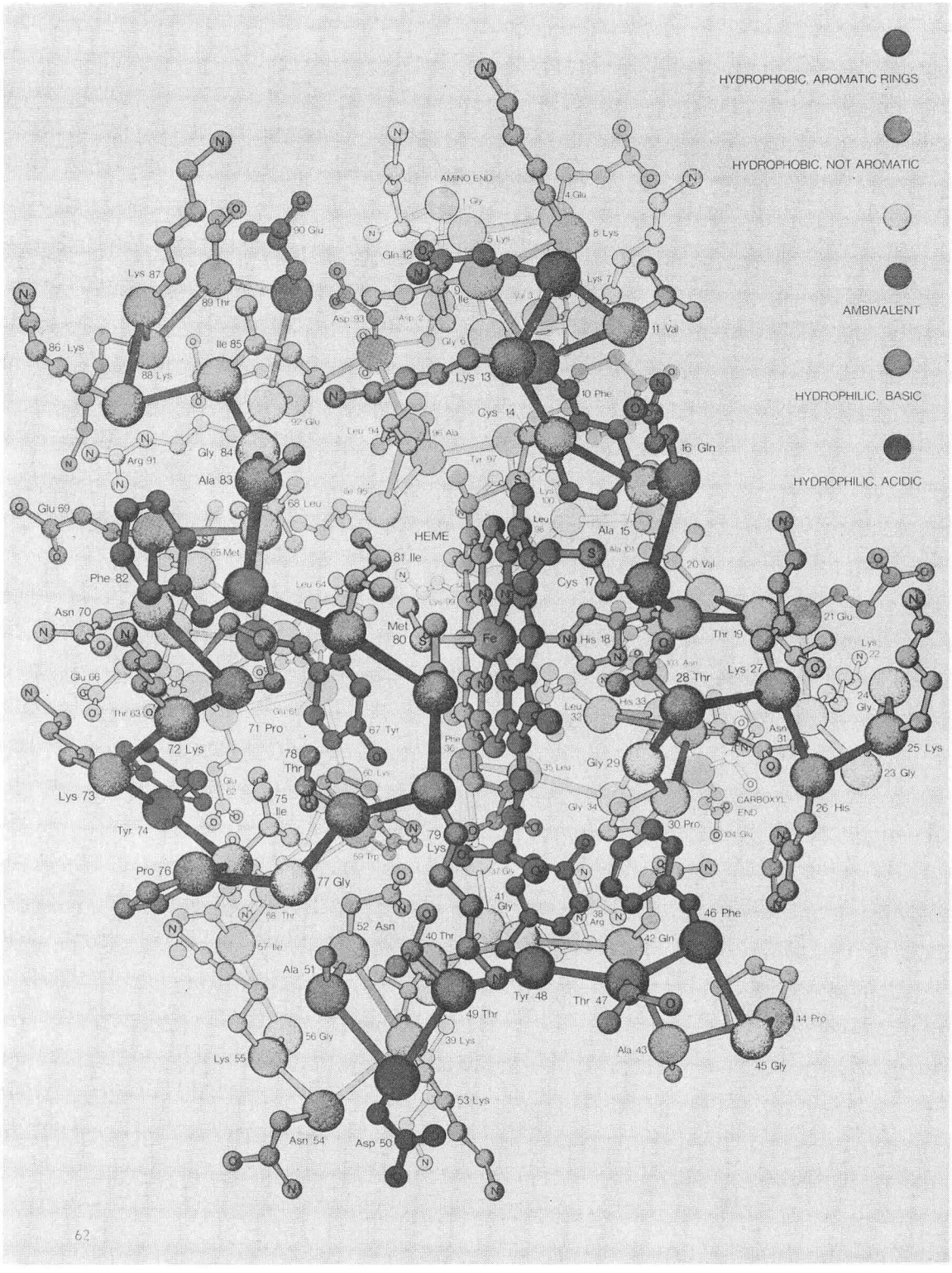


Fig. 9 CBB 728-4287

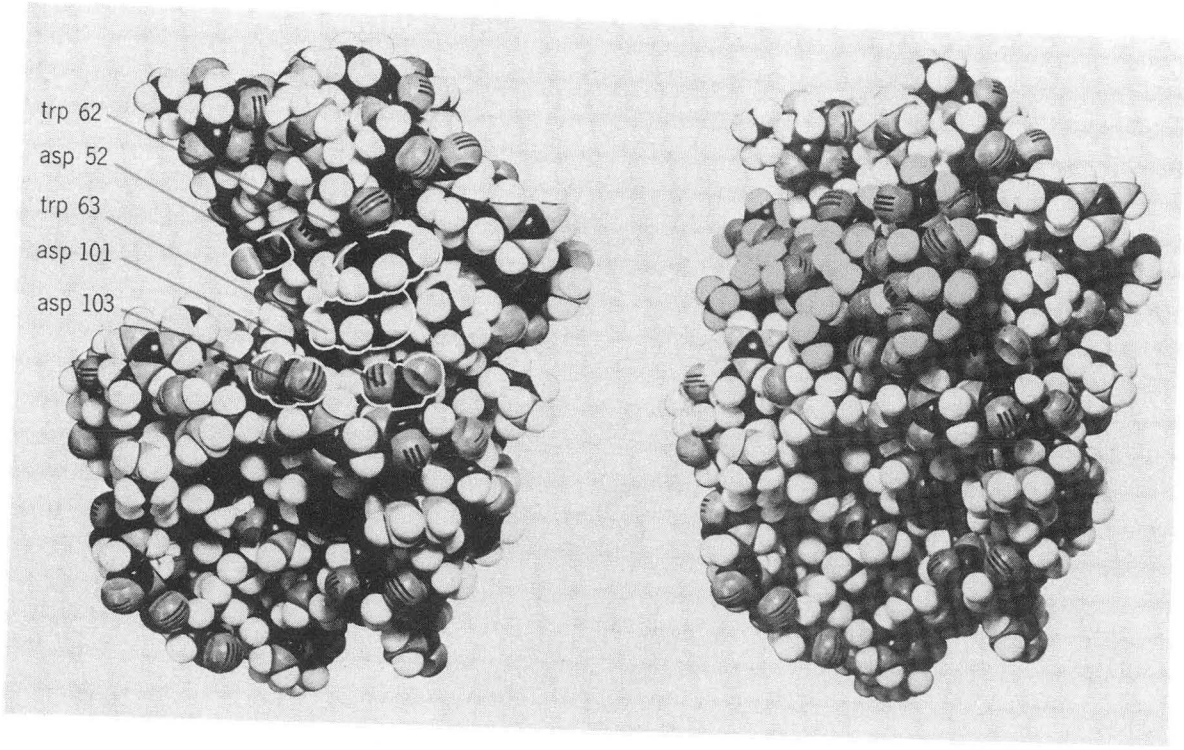
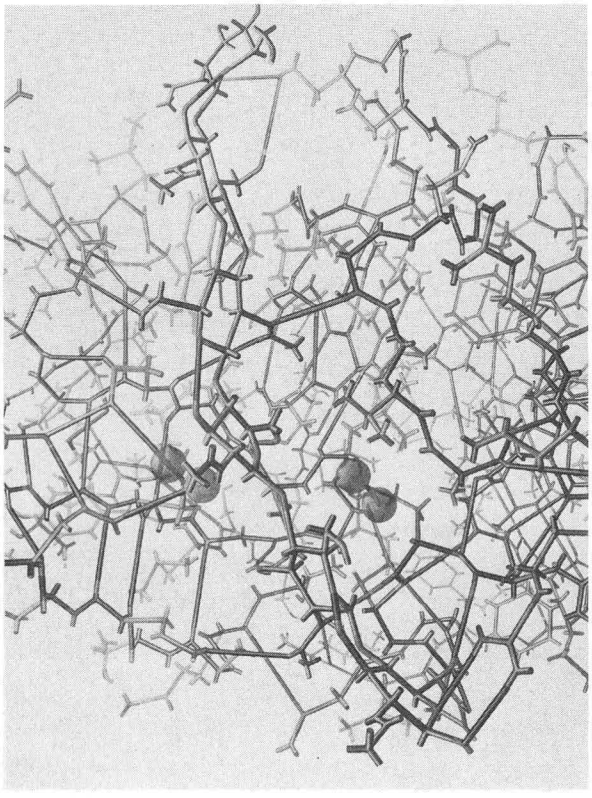


Fig. 10 XBB 729-4365



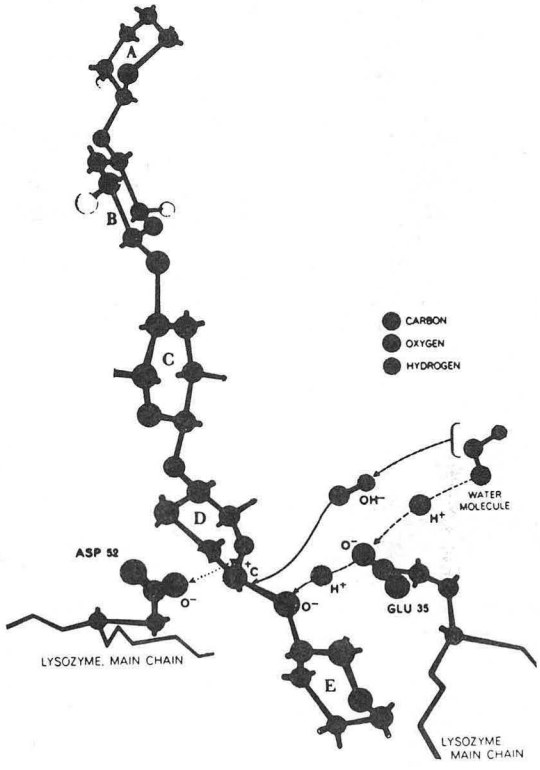
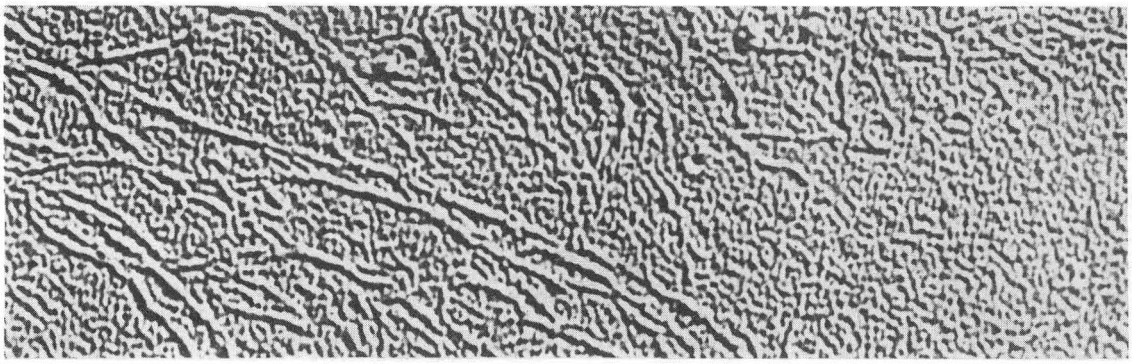
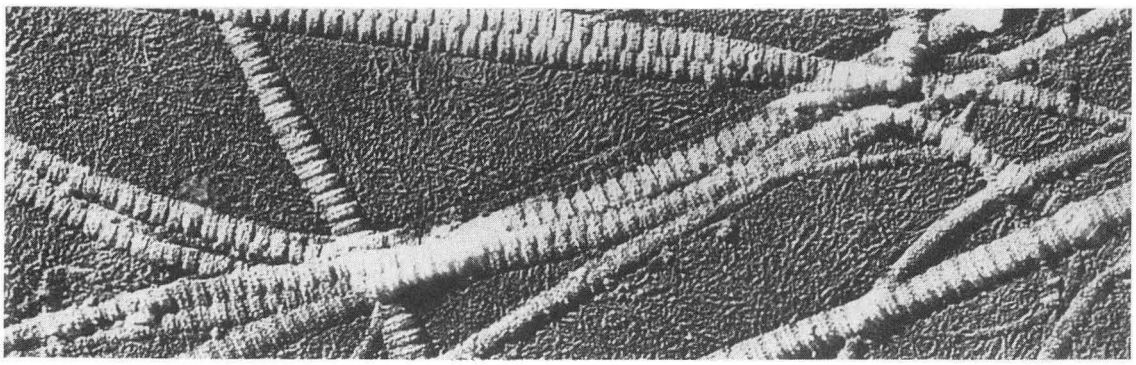


Fig. 11

CBB 728-4289



FILAMENTS OF COLLAGEN, a protein which is usually found in long fibrils, were dispersed by placing them in dilute acetic acid. This electron micrograph, which enlarges the filaments 75,000 times, was made by Jerome Gross of the Harvard Medical School.



FIBRILS OF COLLAGEN formed spontaneously out of filaments such as those shown above when 1 per cent of sodium chloride was added to the dilute acetic acid. These long fibrils are identical in appearance with those of collagen before dispersion.

Fig. 12 BC-272

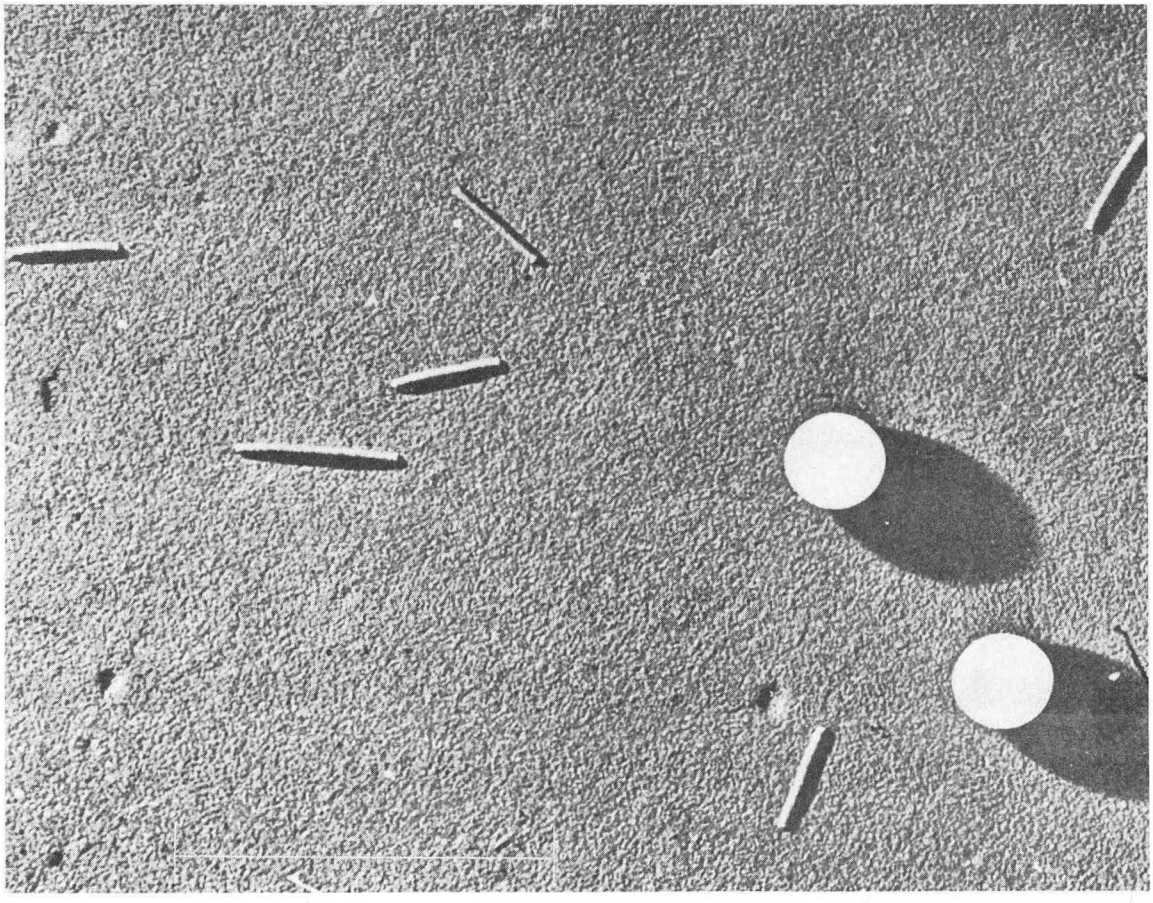


Fig. 13a BC-542

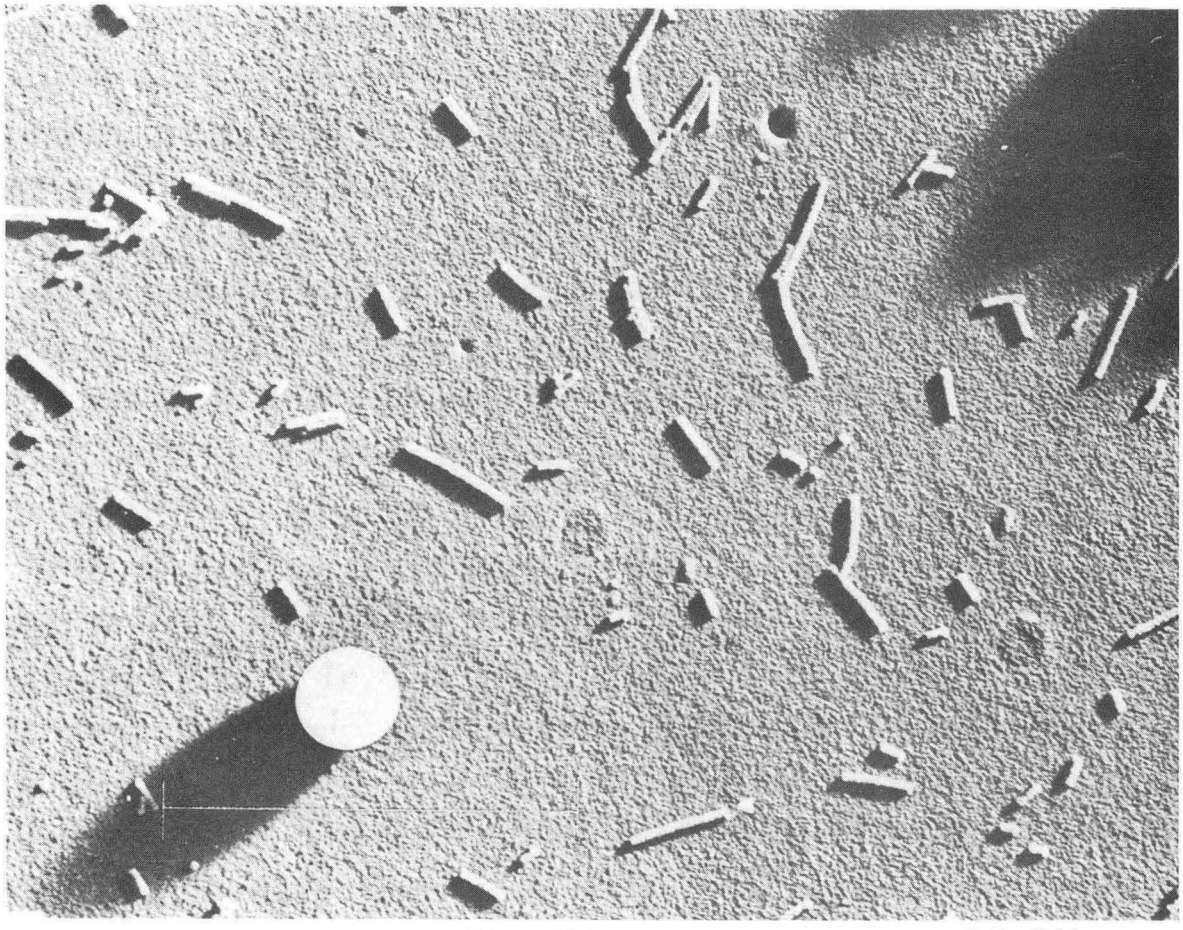


Fig. 13b BC-541

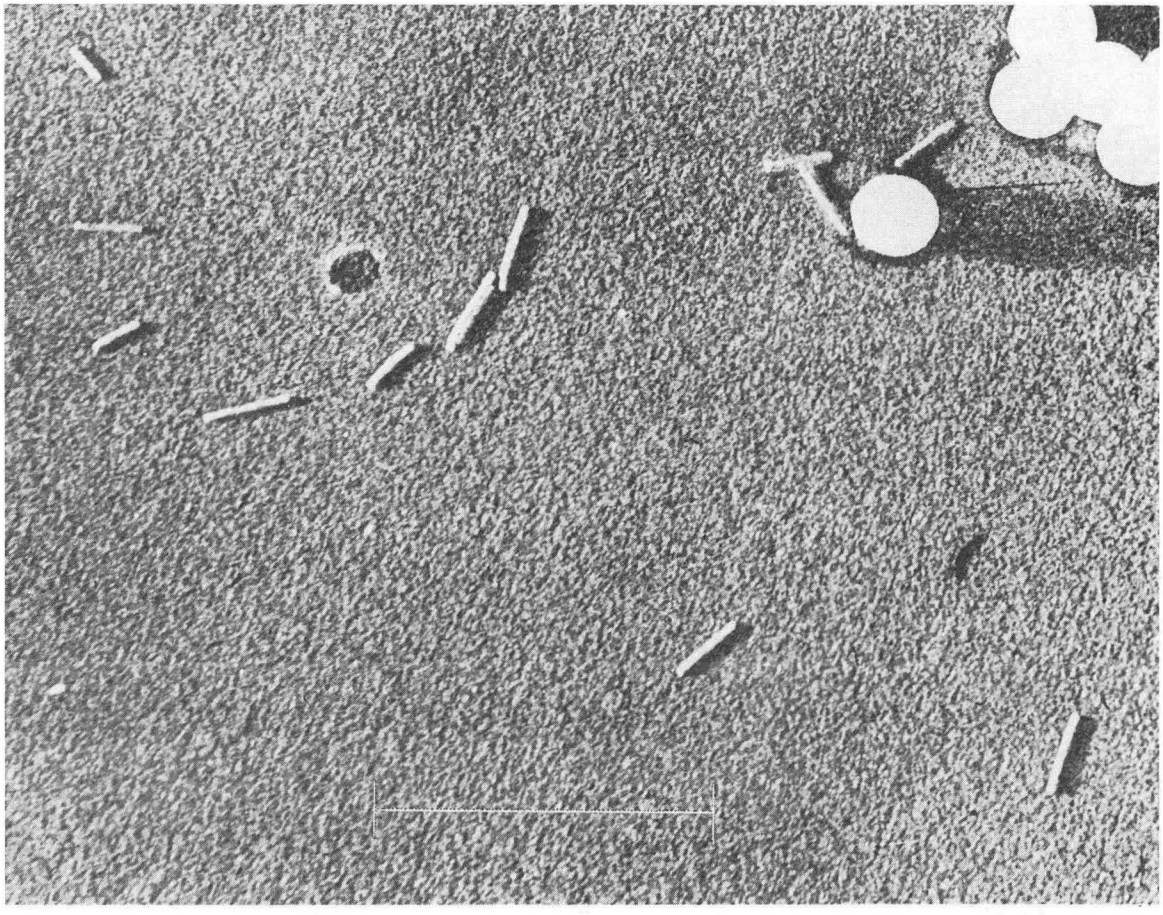


Fig. 13c BC-540

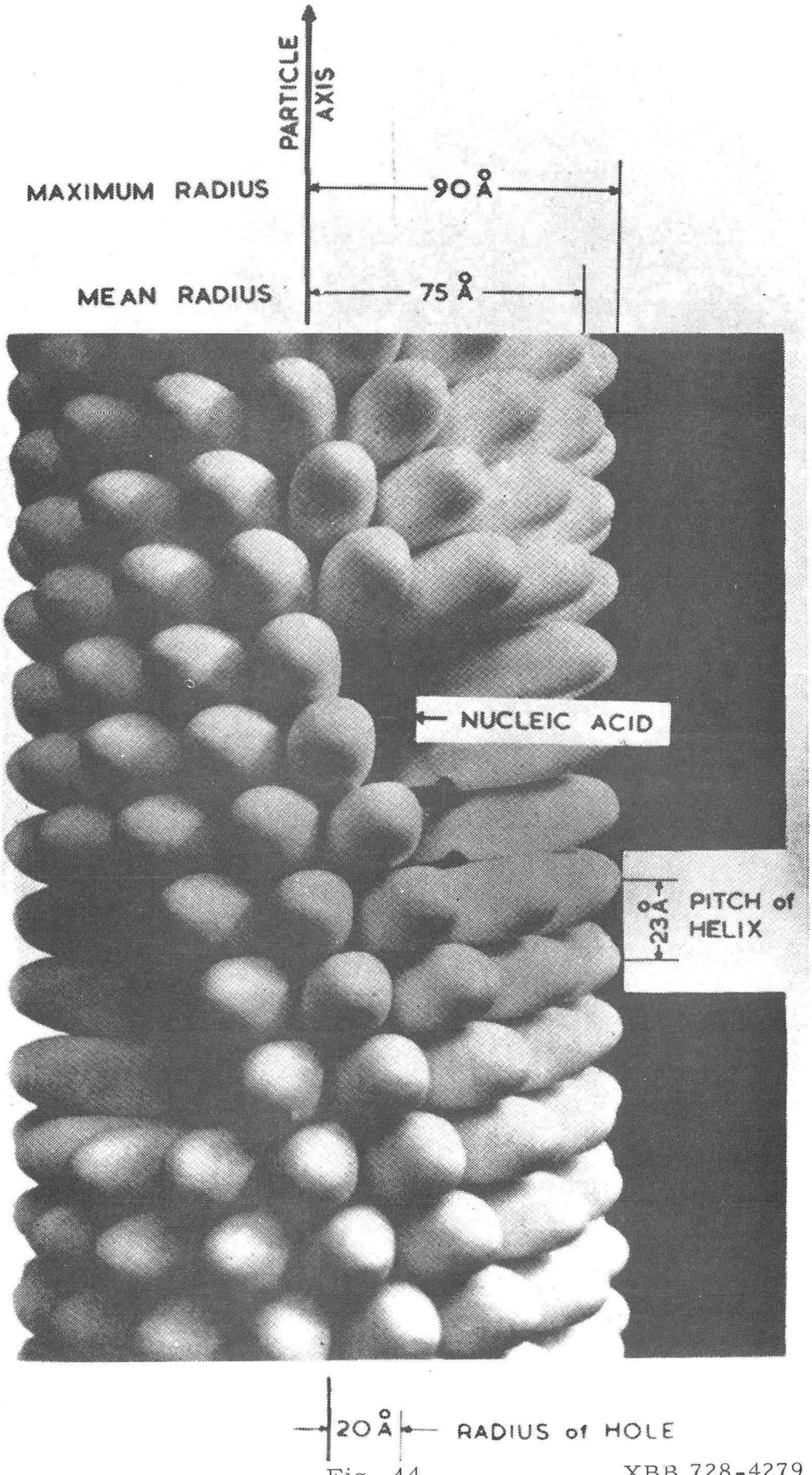
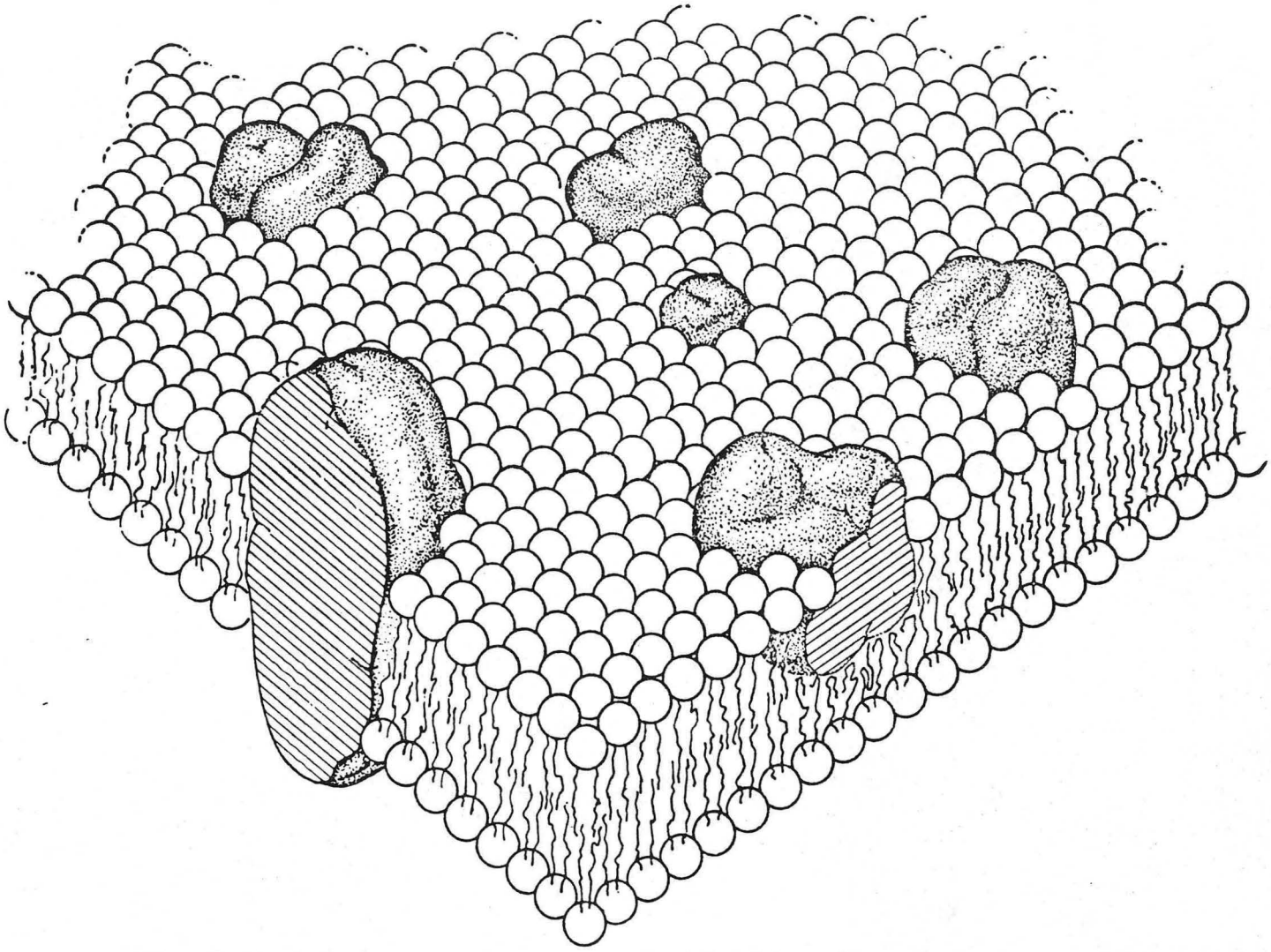


Fig. 14 XBB 728-4279



The lipid-globular protein mosaic model with a lipid matrix (the fluid mosaic model); schematic three-dimensional and cross-sectional views.

XBL 728-1625

Calvin

Figure 15

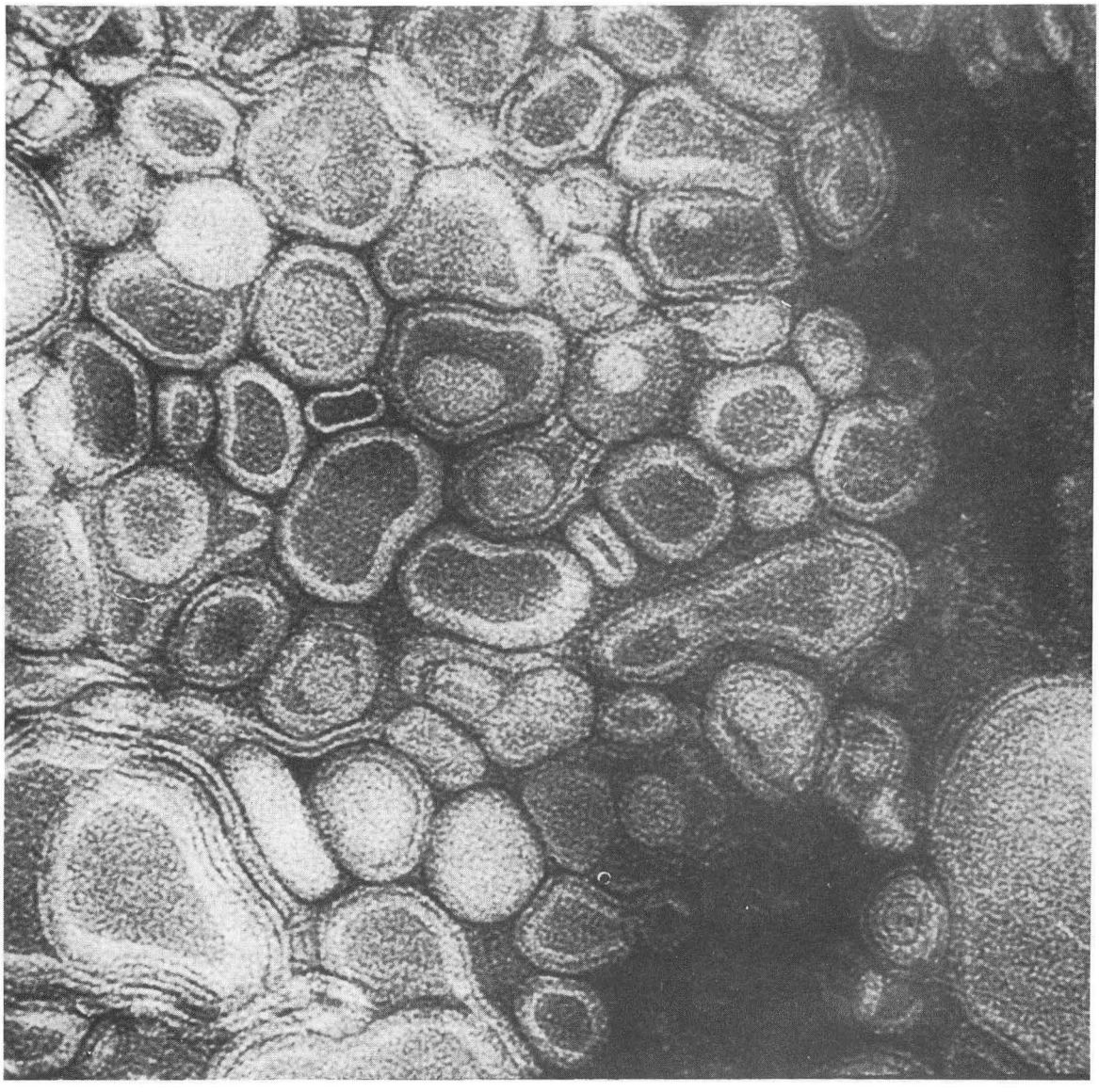
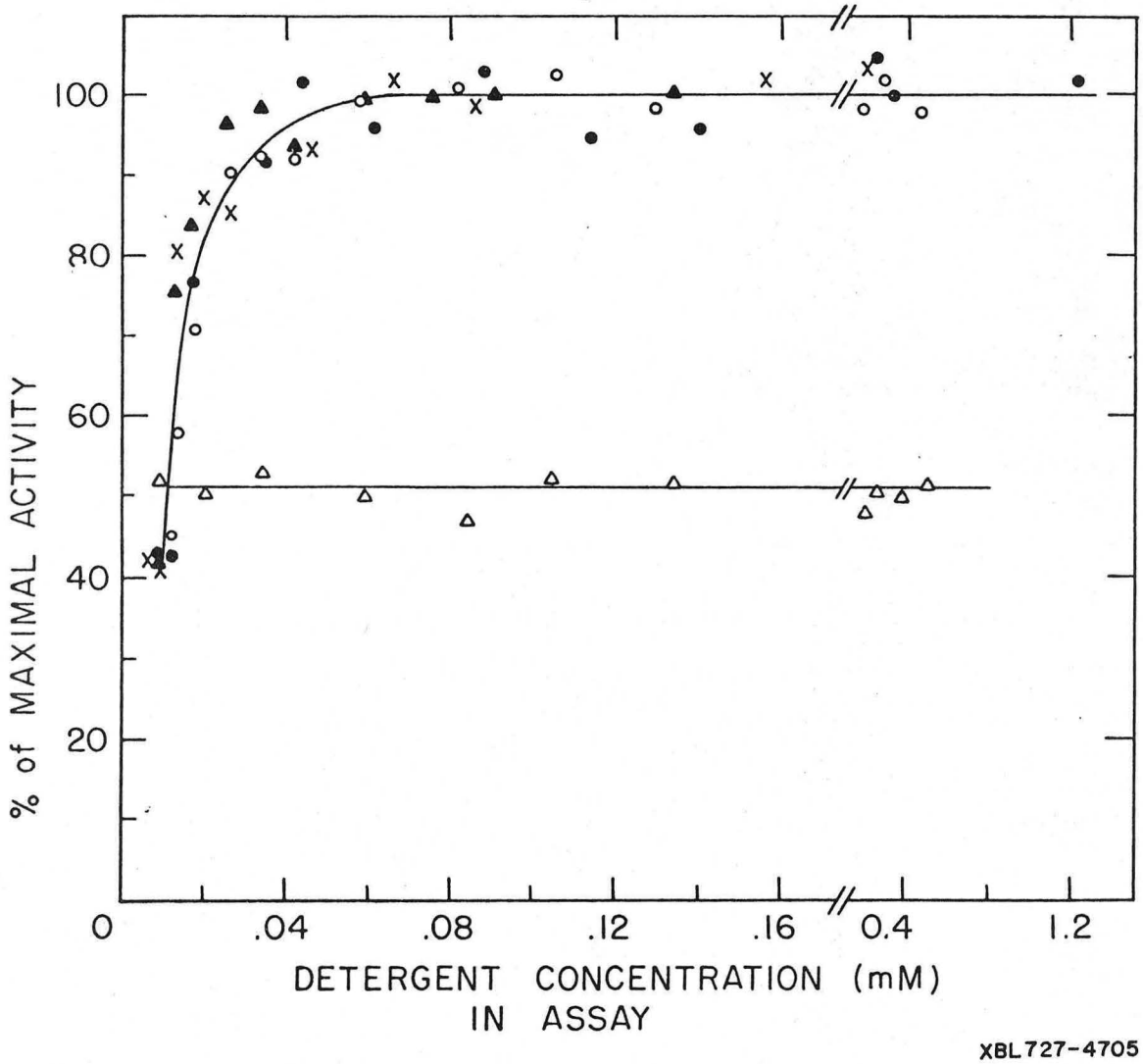


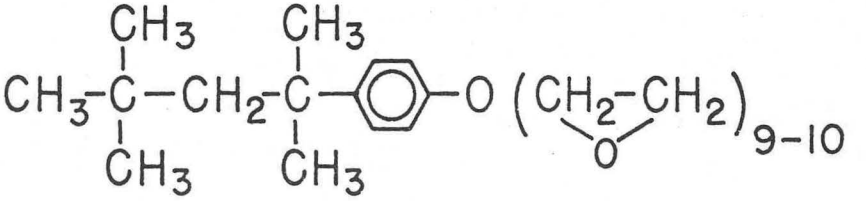
Fig. 16 XBB 729-4501



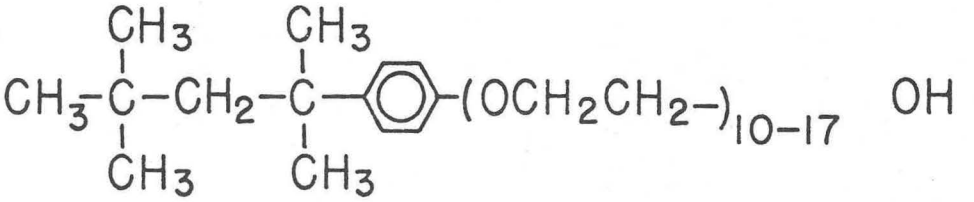
Calvin

Figure 17

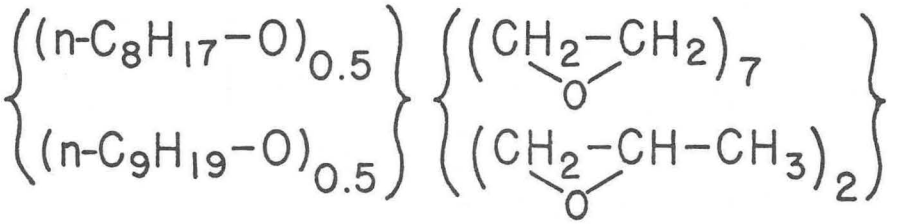
Triton X-100
(Rohm and Haas)



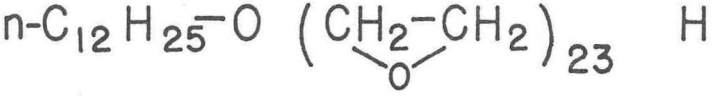
Triton X-1017



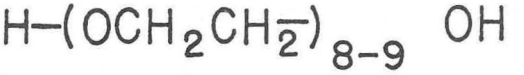
Triton DN-65
(Rohm and Haas)



Brij 35
(Sigma)



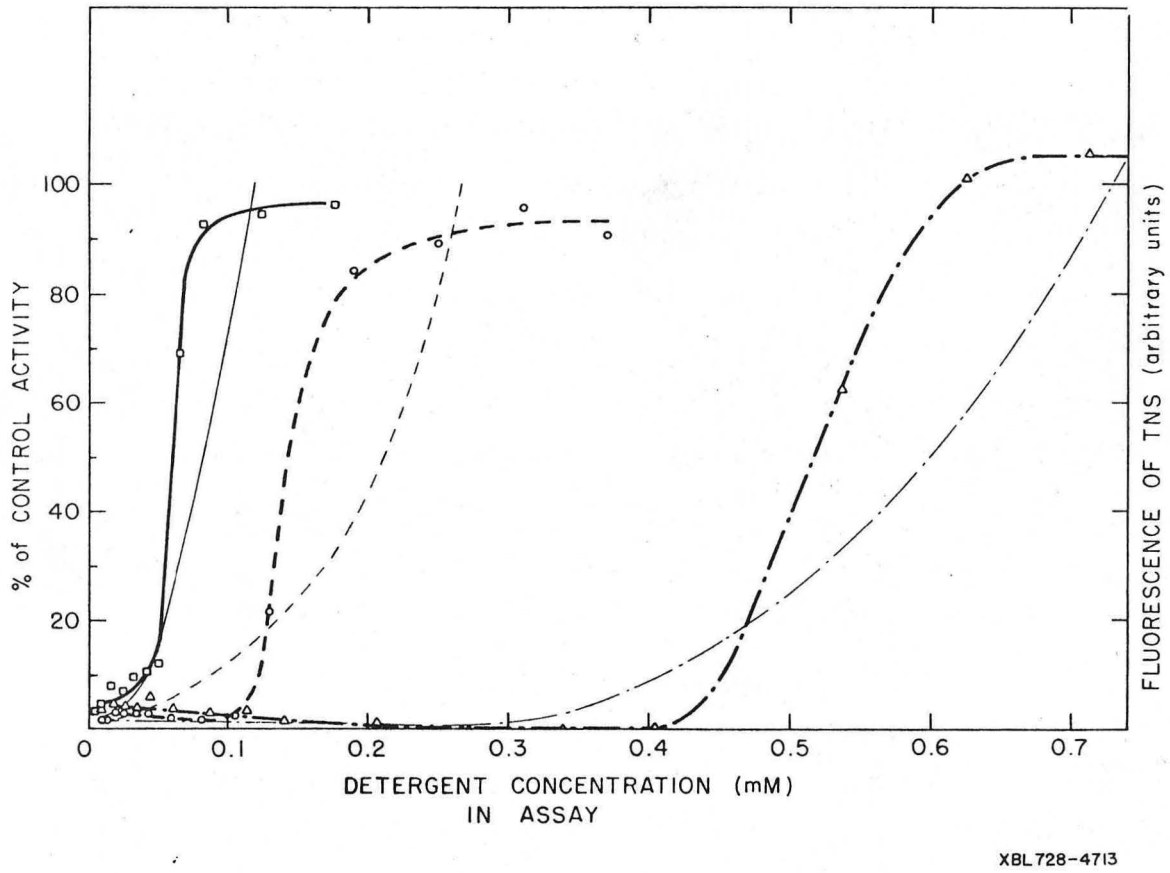
Polyethylene
Glycol-400



XBL729-4755

Calvin

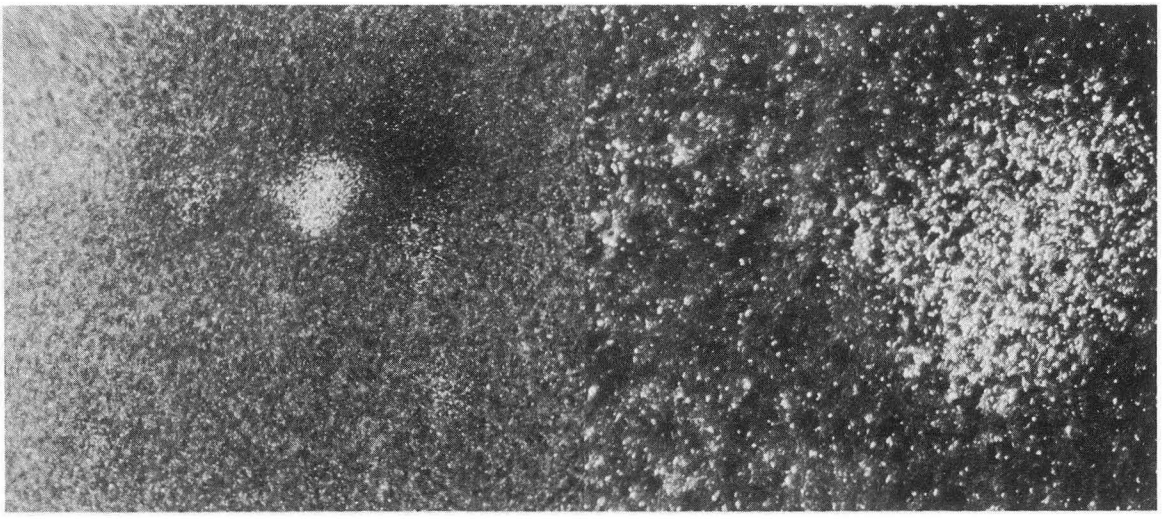
Figure 18



Calvin

Figure 19

MSV Foci on BALB / 3T3 Culture



10x Dark field illumination 40x

Fig. 20 XBB 713-842

TECHNICAL INFORMATION DIVISION
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720