The Hemoglobins in Genetics and Evolution

Vernon M. Ingram

The Hemoglobins in Genetics and Evolution

VERNON M. INGRAM

COLUMBIA UNIVERSITY PRESS
New York and London
1963

Library of Congress Catalog Card Number: 63-10416
Printed in the United States of America

PREFACE

This book is based on six Jesup Lectures, delivered at Columbia University in the Department of Zoology during March, 1962. I accepted the invitation to deliver these lectures with some trepidation. The subject which I wanted to discuss is a part of such diverse areas as biochemistry, human genetics, molecular genetics, and evolution, not to mention hematology, and I felt myself to be too much of a specialist to do justice to all these aspects before an audience of biologists. In particular, an early training in classical organic chemistry and zoology followed by a gradual sliding into biochemistry and the other fields had led to the acquisition of only that knowledge which was of immediate interest and to the persistence of large areas of ignorance. Therefore, to attempt a balanced synthesis seemed too daunting. On the other hand, the temptation of the opportunity to present and develop one's own thoughts on a subject so dear to one's heart proved irresistible. The result is this book, which is offered to the reader in the hope that it will interest him and perhaps stimulate him to disagree with some of the ideas put forward.

In such a short book, it has been impossible to include reference to everyone whose work would have been relevant. A rather arbitrary choice was often necessary, with apologies to those who have had to be omitted, but not through lack of an appreciation of their work.

VERNON M. INGRAM

Cambridge, Mass.
July, 1962

ACKNOWLEDGMENTS

In the first instance, I would like to thank most sincerely Dr. Howard Levene and the committee who organize the Jesup Lectures for their kind invitation. My visits to the Department of Zoology and Botany during that period were made most pleasant and interesting through the kindness of my hosts, particularly Drs. Dunn, Moore, Ryan, Sager, and Taylor, as well as many others.

Much of the original work reported here was done either by, or in collaboration with, Drs. Baglioni, Hunt, and Stretton. In particular, I would like to acknowledge the help and stimulation of many critical discussions with Dr. Baglioni and Dr. Stretton concerning some of the more theoretical ideas put forward here, without, however, wishing to evade my responsibility for them.

Many authors and publishers have been most generous in allowing me to reproduce certain drawings and photographs. Proper acknowledgment is made to them in the appropriate places. I am grateful to the staff of the Columbia University Press for their help, efficiency, and patience.

Finally, I would like to thank my secretary, Miss Barbara Garrelts, without whose help this book would not exist.

V. M. I.

CONTENTS

1

INTRODUCTION

The six lectures transcribed in this book were designed to present a personal view of biochemical genetics and evolution. These two topics were treated almost entirely in terms of the biochemistry of the hemoglobins and no attempt was made to give a balanced view of these large areas of scientific endeavor. The same plan was followed when the pleasant task of giving the Jesup lectures (March, 1962) in the congenial environment of the Zoology Department of Columbia University turned into the more arduous job of putting one's thoughts and prejudices onto paper. Not even the more restricted field of the hemoglobins, normal and abnormal, will be extensively reviewed here, since this has been adequately done in several very recent reviews (see later). Only some aspects of particular interest to the author will be discussed.

Most books that are concerned with biochemical genetics open with a reference to Garrod's work early in this century. This is as it should be, since it was he who in 1902 described the first human biochemical inherited disease, alkaptonuria, and who coined the term "inborn errors of metabolism" (see Garrod, 1923). This name, which has persisted as a most useful description of these diseases, underlines the point that we are dealing here with an inherited defect on the biochemical level. In alkaptonuria one can actually point to the defect as the absence of the active enzyme which normally converts homogentisic acid to maleylacetoacetic acid as part of the metabolism of tyrosine. More recently, the term "inborn error of metabolism" has come to mean an "error" or an alteration in the chemical structure of a biochemically important substance. In this last sense, we may talk of the inherited

hemoglobinopathies as "inborn errors of metabolism" of a kind which allows us to study the basic defect directly, in spite of the fact that a hemoglobin defect is a metabolic defect only in the widest sense of the word.

A second important step in the development of this subject came in 1949 when Pauling coined the term "molecular disease." In that year, Pauling, Itano, Singer, and Wells (1949) discovered that patients with sickle-cell anemia carried a hemoglobin which was electrophoretically different from normal hemoglobin. It seemed likely even then, that such a difference was due to a definite biochemical abnormality at the molecular level. Also in 1949 came the demonstration by Neel and also by Beet, that this "inborn error," this "molecular disease," was inherited as a simple Mendelian factor, with the heterozygote possessing the hemoglobins characteristic of both the normal and the mutant alleles.

These are the most important landmarks for the development of the story of the human hemoglobin abnormalities. Since then there has been a very great deal of research in this area, so that we now have a reasonably coherent picture of the biochemical genetics of this protein. Much of this work has been summarized in several recent reviews (Itano, 1957; Lehmann and Ager, 1960; Rucknagel and Neel, 1961; Baglioni, 1962).

A third milestone was the formulation of the "one gene-one enzyme" hypothesis by Beadle and Tatum (1941) which is fundamental to that part of our picture which deals with genetic control of the structure of any particular protein.

GENETIC CONTROL OF CELLULAR ACTIVITIES

It is a commonplace nowadays in molecular biology to state the relationship between a gene and its product as the relationship between genetic DNA which makes template RNA which in turn makes protein (Figure 1-1). The one gene-one enzyme hypothesis leads to the proposition that the chemical structure of the relevant portion of DNA will produce a definite chemical structure in a protein. It is this relationship and its ramifications which we will examine in this book, taking our examples largely from the biochemical work on the hemoglobins. Just

as a normal gene will produce, in this view, a normal protein molecule, so the effects of a gene mutation are felt as a chemical alteration in the structure of that protein, provided that this protein is the first protein product of the gene. One can well imagine that some proteins are one step further removed from the gene than are for example the peptide chains of hemoglobin. Some proteins have carbohydrate constituents

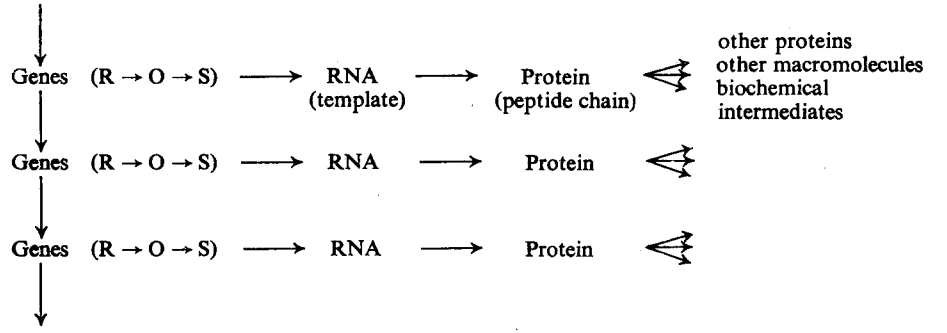

FIGURE 1-1. Scheme of the genetic control of protein synthesis
Three successive generations are shown. R: regulator gene; O: operator gene; S: structural gene. (See Jacob and Monod, 1961.)

attached to them, put there by the action of some enzymes. These enzymes are themselves proteins and are made by their own genes. A mutation in one of these genes may alter or delete the function of such an enzyme and may ultimately show up as an alteration in the carbohydrate portions of one or more other proteins, leaving their peptide chain sequences perfectly normal. This would be a less direct effect of a gene mutation on protein structure.

The complex system of genetic control (Figure 1-1) leaves of course a great deal unsaid, particularly in terms of the relationship between the types of gene (regulator, operator, structural) or indeed the need to postulate them in a mammalian system. Nevertheless, the scheme provides a convenient and useful way for examining the findings of biochemical genetics and for directing the planning of experiments.

This scheme of cellular control is supposed to be applicable in bacteria as well as in mammals. In fact the best studied examples for

the correspondence between genetic and protein structure come from the bacterial systems (Dreyer, 1960; Jacob and Monod, 1961; Garen, Levinthal, and Rothman, 1961; Helinski and Yanofsky, 1962). The same is true for the demonstration of the role of template (messenger) RNA (see for example Brenner *et al.*, 1961; Gros *et al.*, 1961). However, we will confine ourselves in this book to applying the principles derived from the more tractable microbial systems to the control of hemoglobin synthesis and structure.

Before proceeding to the hemoglobins, we must however point out that there are other human proteins which can be used to study the scheme outlined in Figure 1-1, of which only two will be mentioned here, glucose-6-phosphate dehydrogenase (G-6-P-D) and haptoglobin.

G-6-P-D DEFICIENCY

The inherited glucose-6-phosphate dehydrogenase deficiency has been demonstrated in many Negroes (10 percent) in this country who have a severe hemolytic anemia when certain drugs are administered (see Marks, 1960; also Marks *et al.*, 1961; Ramot *et al.*, 1961; Adinolfi *et al.*, 1961). These drugs include sulfanilamide and primaquine. The usual form of the condition is inherited as a sex-linked gene of intermediate dominance. The biochemical reaction which seems to be involved in this disease and which is mediated by the enzyme G-6-P-D is the following: glucose-6-phosphate + 2 TPN $\rightarrow$ 6-phosphogluconic acid + 2 $TPNH^+$. The TPNH which is produced is used in many ways in the cell; for example it is used to keep glutathione reduced.

Siniscalco, Bernini, and Latte (1961) found that in Sardinia G-6-P-D deficiency is linked to color blindness and to hemophilia. There is still some controversy as to whether this disease is really all of one kind in different parts of the world. Although the clinical symptons are similar, this is probably a group of different enzyme defects. Another way in which this deficiency is detected in Italy and elsewhere is after ingestion of the fava bean which in persons with this deficiency will also cause a hemolytic anemia. Siniscalco *et al.* (1961) found that G-6-P-D deficiency is strongly correlated in portions of Sardinia with the presence, either

today or in the past, of malaria. This is an example of balanced polymorphism, where the disadvantage of a gene is balanced by some selective advantage of the heterozygote (usually), leading to an unexpectedly high level of frequency of a deleterious gene. We will come across other examples of this phenomenon in the hemoglobin system. After all, the possession of G-6-P-D deficiency is deleterious—nobody knows how deleterious exactly and yet the frequency is high. Since there is a positive statistical correlation between the distribution of malaria and the distribution of this gene, one is tempted to postulate that partial protection against malaria is the mechanism. There is as yet no actual evidence for such protection in this particular biochemical abnormality. When we turn to sickle cell anemia later, we will find that there is a similar situation of balanced polymorphism, but there is now considerable evidence that malaria really is involved in increasing the frequency of the sickle cell anemia gene. In the case of the G-6-P-D deficiency we have only a correlation at the moment. Sardinia is a favorable place for studying balanced polymorphism, because there are villages high up in the mountains with little malaria and other villages in the lowlands, where there is, or was, a lot of malaria. And just as G-6-P-D deficiency is positively correlated with malaria, so it is negatively correlated with altitude.

The kind of G-6-P-D deficiency which Ramot *et al.* (1961) have studied in Israel amongst the Jews may not be the same condition. She also found that G-6-P-D deficiency is linked to color blindness, but this time the genes are in repulsion, whereas in Sardinia they were in coupling. Ramot also studied the level of the enzymes in tissues other than the red cell (Table 1-1) and found different levels of enzyme activity in affected individuals, indicating that the gene mutation affects the enzyme in a variety of cell types, if not in all.

It is not always clear that the same structural gene is involved in making enzymes of similar function in different tissues. There is some chemical evidence that the enzyme lysozyme in the spleen and the brain of dogs is different chemically; maybe there are different, though related, structural genes involved in producing these enzymes. On the other hand, as far as human G-6-P-D is concerned, it looks as if only

TABLE 1-1

G6PD ACTIVITY IN VARIOUS TISSUES

Tissue	*Number*	*Control*	*Number*	*"Mutants"*
Erythrocytes°	103	12.6 ± 1.18	55	1.08 ± 1.04
Platelets°°	56	0.127 ± 0.038	34	0.024 ± 0.017
Leucocytes°°	26	34.70 ± 11.20	25	7.59 ± 2.72
Saliva**	26	8.00 ± 6.3	14	0.54 ± 0.4
Liver*	18	0.112 ± 0.038	3	0.010–0.030
Skin**	9	2.6–20.9	2	0.167–1.6

° Units = ΔOD/g Hb/min. ± standard deviation.
°° Units = ΔOD/10^9 cells/min.
** Units = ΔOD/mg protein/min.
* Units = ΔOD/100 mg wet Liver Tissue/min.

one structural gene is involved; the mutation, whatever it is, affects the G-6-P-D enzyme similarly in the various tissues. Marks *et al.* (1961, 1962) have reported studies of the biochemical properties of the G-6-P-D enzymes taken from normal individuals and from a variety of mutants, which demonstrate that chemically different enzymes may be found (Table 1-2). For example, in the normal subject and in the affected Negro subject the enzyme has a certain mobility in starch gel electrophoresis; but in one Italian family, the electrophoretic mobility of the G-6-P-D enzyme was increased 35 percent; therefore we would suspect that a different protein is made in these individuals, pointing to at least two types of G-6-P-D deficiency.

In the hemoglobins we are very fortunate because the protein is easy to prepare. Even the mutant hemoglobins are usually easy to obtain. We can therefore settle more easily the question whether we are dealing with a mutant protein or with different rates of production in the hemoglobin abnormalities.

THE HAPTOGLOBINS

Another example of an inherited abnormality in the human system is the haptoglobins. These human proteins maybe a type of antibody, at least they combine specifically with hemoglobin (Jayle and Boussier, 1955; Moretti *et al.*, 1957). It looks as if one molecule of haptoglobin

TABLE 1-2

PROPERTIES OF GLUCOSE-6-P DEHYDROGENASE PURIFIED FROM RED CELLS OF NORMAL SUBJECTS AND PERSONS WITH A DEFICIENCY IN THIS ENZYME

Source	*RBC enzyme activity*	K_{m1} *TPN*	K_m *G6P*	K_m *2-D-G6P*	K_m *nicotinamide*	*pH optima*	*Electrophoresis*
	Units/gmHgb	$M \times 10^{-6}$	$M \times 10^{-5}$	$M \times 10^{-3}$	$M \times 10^{-2}$		*Percent*
Normal subjects (7)	11.5–18.2	3.4–7.1	3.5–5.6	3.0–3.6	2.0	8.5–9.3	100.0
Affected subjects:							
Negro males (4)	1.2–4.1	4.4–8.0	3.7–4.1	4.2–5.0	2.0–3.0	8.5–9.3	100.0
Negro female (1)	8.7	3.6	4.0	4.0	2.0	8.5–9.3	100.0
Barbieri males (2)	6.5–7.2	12–28	5.6–8.0	2.5–5.9	2.0	8.5–9.3	135.0
Barbieri females (1)	8.0	24	6.6	5.0	2.0	8.5–9.3	131.0
Caucasian male (1)	0.9	9.2	3.4–6.7	4.5	—	8.5–9.3	100.0
Caucasian female (1)	0.8	5.0	2.8	—	—	8.5–9.3	100.0
Caucasian-Negro male (1)	1.5	14	7.5	7.2	—	8.6–9.3	—

Abbreviations: TPN, triphosphopyridine nucleotide; G6P, glucose-6 phosphate; 2-D-G6P, 2-desoxy-glucose-6-phosphate. Values indicated in table are range of values where more than one enzyme preparation was studied.

will combine with one molecule of hemoglobin, that is, each haptoglobin molecule has one combining site. This may only be true for the so-called type 1-1 haptoglobin (Smithies and Connell, 1959); type 2-1 or 2-2 may combine with more than one hemoglobin molecule for reasons which will be apparent later. Usually a typical antibody has two combining sites; thereby it is able to form a three-dimensional network with the protein with which it reacts and to precipitate that protein. Haptoglobin does not precipitate hemoglobin, but it does form a strong complex which seems to be a mechanism for removing free hemoglobin from solution in the serum. During a hemolytic crisis, hemoglobin is liberated into serum, the haptoglobin combines with it and the haptoglobin-hemoglobin complex is then eliminated. After a severe hemolytic crisis one does not find any haptoglobin, because it has been used up. There are actually many people who have no haptoglobin at any time; about three or four percent of some of the Negro populations in Africa show this absence of haptoglobin (Allison *et al.*, 1958). This lack of haptoglobin does not seem to be disadvantageous, so that perhaps the haptoglobin is not an essential part of the biochemical mechanism of the body. On the other hand, there may be other functions of this protein of which we know nothing yet.

Smithies, in 1956, introduced a beautiful technique for the separation of serum proteins (reviewed in Smithies, 1959). He performed electrophoresis of these proteins in starch gel, which gives a much better separation and characterization of human proteins than had been possible before. This technique enabled him to see these haptoglobins; on screening populations he discovered that there were three distinct types of individuals with respect to their haptoglobins (Smithies and Connell, 1959). These were called types 1-1, 2-1, and 2-2. The type 1-1 was from a homozygote Hp^1/Hp^1, the 2-2 was the other homozygote (Hp^2/Hp^2) of the allelic system and the 2-1 type was produced by the heterozygote Hp^1/Hp^2. The molecular weight of type 1-1 haptoglobin is about 85,000 (Moretti *et al.*, 1957); it is not entirely a simple protein, but also contains some carbohydrate, which was shown by Jayle and Boussier (1955) some time before Smithies' work. Parker and Bearn (1962) have found recently that neuraminidase, which presumably

removes a carbohydrate constituent, alters the electrophoretic mobility of human haptoglobin 1-1.

Figure 1-2 shows a starch gel electrophoresis taken from Smithies' work, illustrating on the left the three types of haptoglobins which are found. The homozygote, Hp^2/Hp^2, produces a series of bands, some stronger and some weaker, probably due to association of this type of molecule. The heterozygote produces a series of proteins *intermediate* in nature between the two homozygotes. Reduced type 1-1 and 2-2 show single protein bands in different electrophoretic positions. The heterozygote, 2-1, after reduction, has both the bands characteristic of 1-1 and of 2-2. The reduced proteins show phenotypes which correspond to the alleles in the genotypes.

In the abnormal hemoglobins, the heterozygote shows both forms of hemoglobin molecules of the allelic system *without* any previous degradation of the protein. One does not get truly hybrid molecules made up of peptide chains produced by the two allelic genes, as happens in the haptoglobins. This is a very clear distinction between the two systems.

One can recognize two types of peptide chains in the haptoglobins: α chains and β chains. Figure 1-2 shows only the α chains, since in this system of electrophoresis the β chains remain insoluble at the origin. But recently it has been possible (Connell, Dixon, and Smithies, 1962) to move the β chain away from the origin and to show that they are the same in these three types. The mutational differences which distinguish the haptoglobin types reside in the α chain.

It appears that in fact there are three kinds of type 1-1 haptoglobin, the socalled 1^{Fast}-1^{Fast}, 1^{Fast}-1^{Slow} and 1^{Slow}-1^{Slow}, referring to the speed of their α chain components in electrophoresis (Connell, Dixon, and Smithies, 1962).

Figure 1-3 illustrates in schematic form some of the chemical findings of Smithies, Connell, and Dixon (1962), derived from digestion of the peptide chains with chymotrypsin and from isolation and fingerprinting of the fragments in a way similar to the methods discussed in chapter 3. Smithies has shown that the Hp^2 peptide chain differs from $Hp^{1^{Fast}}$ and $Hp^{1^{Slow}}$ α peptide chains by being almost twice as long. By careful

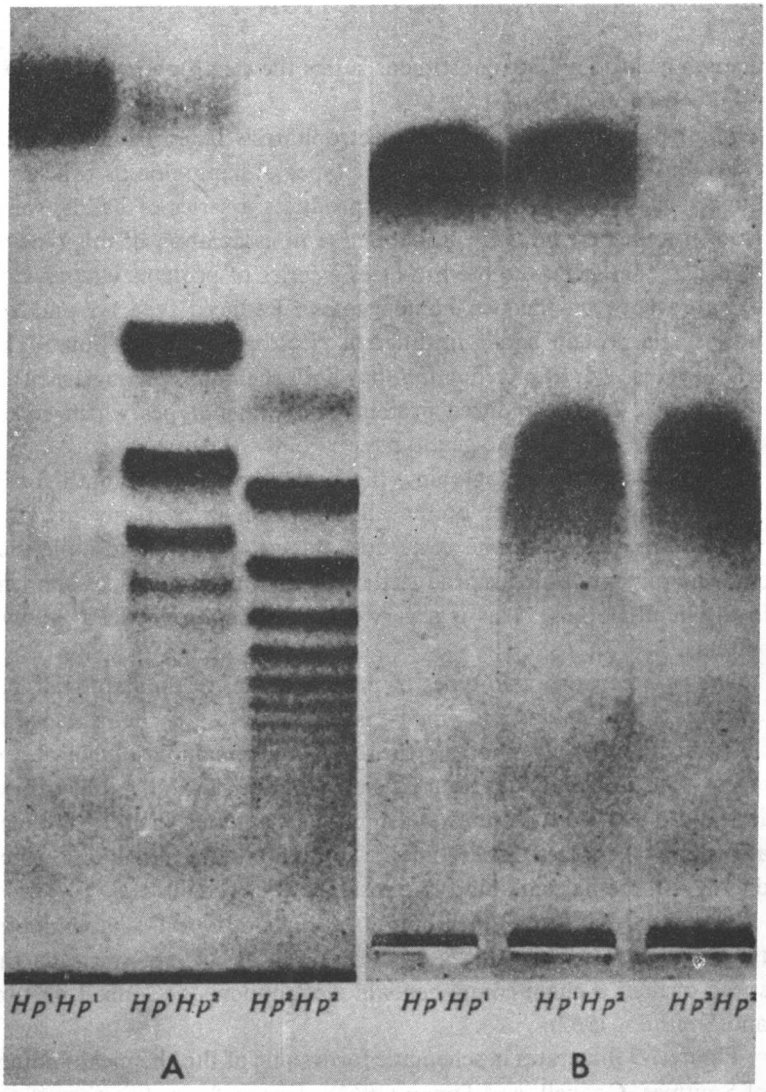

FIGURE 1-2. The haptoglobins of the three genotypes Hp^1/Hp^1, Hp^1/Hp^2, Hp^2/Hp^2

A: purified haptoglobins, B: haptoglobins after reduction of disulfide bonds and alkylation of the liberated —SH groups. Electrophoresis in starch gel, pH 3.2. By permission of Dr. Oliver Smithies.

purification and fingerprinting of enzymatic digests of these chains he deduced (Smithies, Connell, and Dixon, 1962) that most of the peptides derived from the Hp^1 chain are represented in the Hp^2 chain, but that the Hp^2 chains seem to be composed of one Hp^{1Fast} chain and one

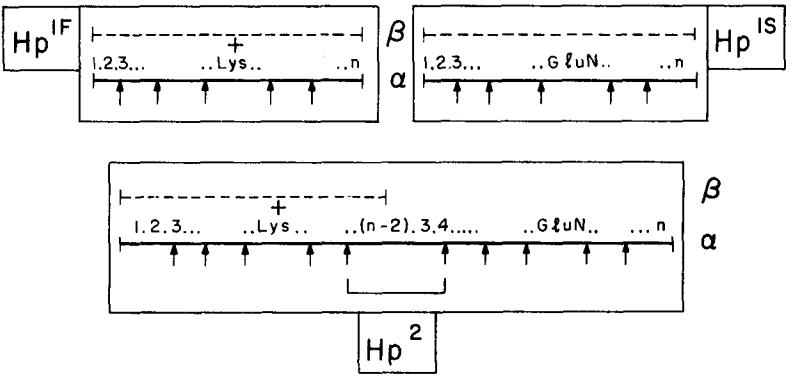

FIGURE 1-3. Schematic representation of the chemical differences in the α peptide chains of haptoglobins Hp^{1F} ($= 1^{Fast}$), Hp^{1S} ($= 1^{Slow}$) and Hp^2 (Based on the results of Smithies *et al.*, 1962.)

Hp^{1Slow} chain joined together with a few amino acids missing at the junction. Hp^{1Fast} differs from Hp^{1Slow} by having a lysine residue instead of a glutamic acid or glutamine residue (see Figure 1-3), thus accounting for the electrophoretic differences. The pronounced difference in electrophoresis between the whole molecules of Hp^1 and Hp^2 is due to the

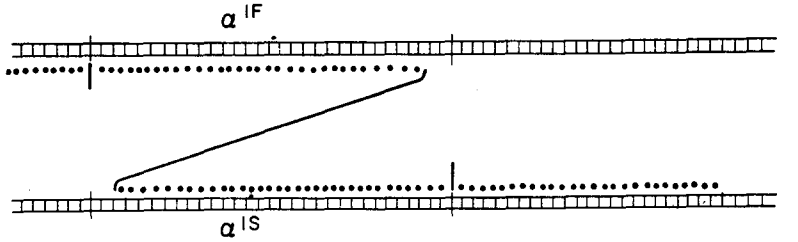

FIGURE 1-4. Schematic representation of an unequal crossing over mechanism to explain the origin of the α^2 chain of Hp^2 from a heterozygote Hp^{1F}/Hp^{1S} which has an α^{1F} and an α^{1S} gene