A MODEL FOR THE SPECIFIC PAIRING OF HOMOLOGOUS DOUBLE-STRANDED NUCLEIC ACID MOLECULES DURING GENETIC RECOMBINATION

STEWART McGAVIN

Department of Chemistry, University of Dundee, Dundee, Scotland

SUMMARY

A model involving a specific pairing of homologous double-stranded nucleic acid molecules is applied to some parts of genetic recombination.

The most original features of this application of the model relate to the initiation process, a part of which can be described by a highly compact and symmetrical structure. The model also provides a simple view of the formation of hybrid nucleic acid.

The possibilities of detecting four-strand structures are briefly discussed.

1. INTRODUCTION

IN this paper I attempt to apply a nucleic acid model (McGavin, 1971*a*, *b*) to some parts of genetic recombination. In the model two double helices of the Watson-Crick (Watson and Crick, 1953) kind, with identical base sequences, are paired together specifically about a common long molecular axis. The model was arrived at partly from reading in genetics which suggested to me that the specific pairing of homologous double-stranded nucleic acid was likely.

While there are a large number of models for genetic recombination, or at least for parts of it, there are not many which treat the process at a detailed molecular level and as far as I know there are no others which provide a detailed mechanism for the specific pairing of double strands. Models for genetic recombination usually postulate some kind of approach between homologous double strands then assume splitting down into single strands followed by the recognition of complementary single strands by each other. All such models seem to involve at some stage rather complicated intermediate structures although they can lead to simple structures such as the crossed strand exchange described by Holliday (1964) and Sigal and Alberts (1972).

Whatever the mechanism, there must be precise matching between homologous double strands and apart from models which involve breaking down into single strands, the model discussed here is the only one I know of which gives this necessary precision of matching.

It may be that apart from a description of the initial specific association, nothing is proposed which other models do not already include. It does seem though to be an interesting exercise to follow up, in as much detail as possible, the consequences of applying the model. In what follows, therefore, no attempt is made to take existing models into account such as, for instance, those of Holliday (1964), Whitehouse and Hastings (1965), Sobell (1972) and others which may be correct, and some of which might be

reformulated to include the kind of specific association which I suggest. (See also the books by Fincham and Day (1971) and by Kushev (1975)). Similarly the Gierer structures (Gierer, 1966) which are included in some recombination models might be rearranged to form four-stranded structures.

In what follows, genetic terminology has been avoided, the aim being to place emphasis on the association and possible consequences of the association between homologous double-stranded nucleic acid molecules to form a four-stranded structure. I have tried also, as much as possible, to avoid introducing enzymes into the discussion although several genes are well known to be necessary in recombination.

2. The model and its specificity

The model, as described earlier, is simply two homologous double helices of the Watson-Crick kind related by a dyad axis coincident with a common long molecular axis. The double helices are in contact (paired) along their " deep " grooves and it follows that if pairing is over more than a few bases that they are plectonemically wound together.
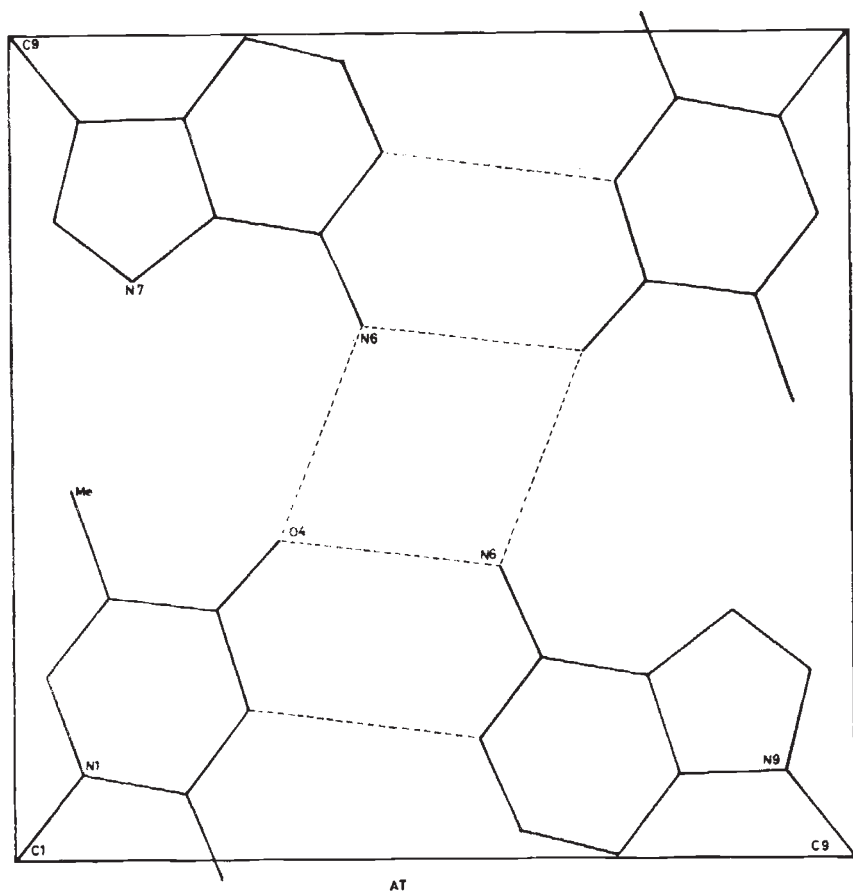

Fig. 1.—Pairing of AT with AT.

It must be stressed that this is only a model and that the structure has not been shown to occur naturally. There is, however, experimental evidence for parts of the structure in the observation of schemes, or of schemes close to the ones used for the pairing of pairs of base pairs, in crystal structures (O'Brien, 1967; Simundza, Sakore and Sobell, 1970).
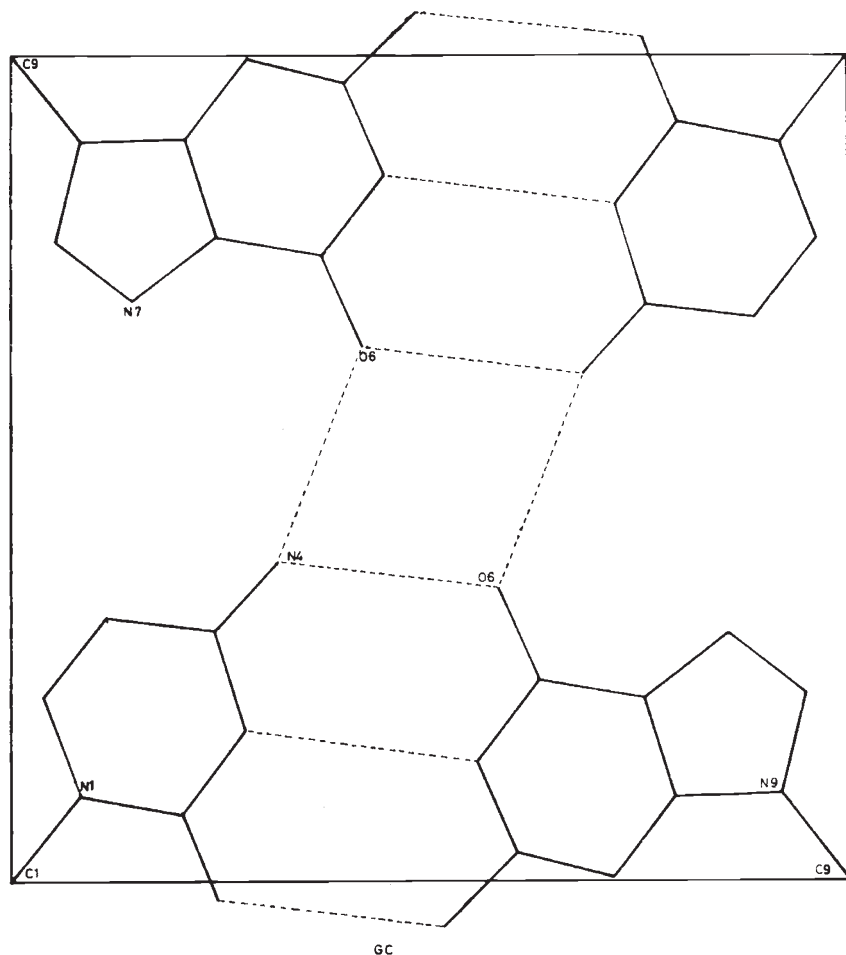

FIG. 2.—Pairing of GC with GC.

The pairing schemes used in the model are reproduced in figs. 1 and 2. It should be noted that they, or at least ones close to them, were first discussed by Löwdin (1964) and also by Kubitschek and Henderson (1966).

Much of the interest in this model is in its apparent specificity and for this reason I have studied these pairing schemes in more detail since publishing the papers mentioned above and have also considered the possibility of pairing of unlike pairs of base pairs. This work is described elsewhere. It supports the use of the pairing schemes for pairs of identical base pairs and seems to exclude the possibility of the frequent pairing of unlike pairs of base

pairs in a regular four-strand structure (see fig. 3). In the paper just men-
tioned (McGavin, 1977) discussion of the model is brought more up to date
in relation to recent work on the principles involved in nucleic acid structures
and in relation to recent work on polynucleotides (see, for example, a recent
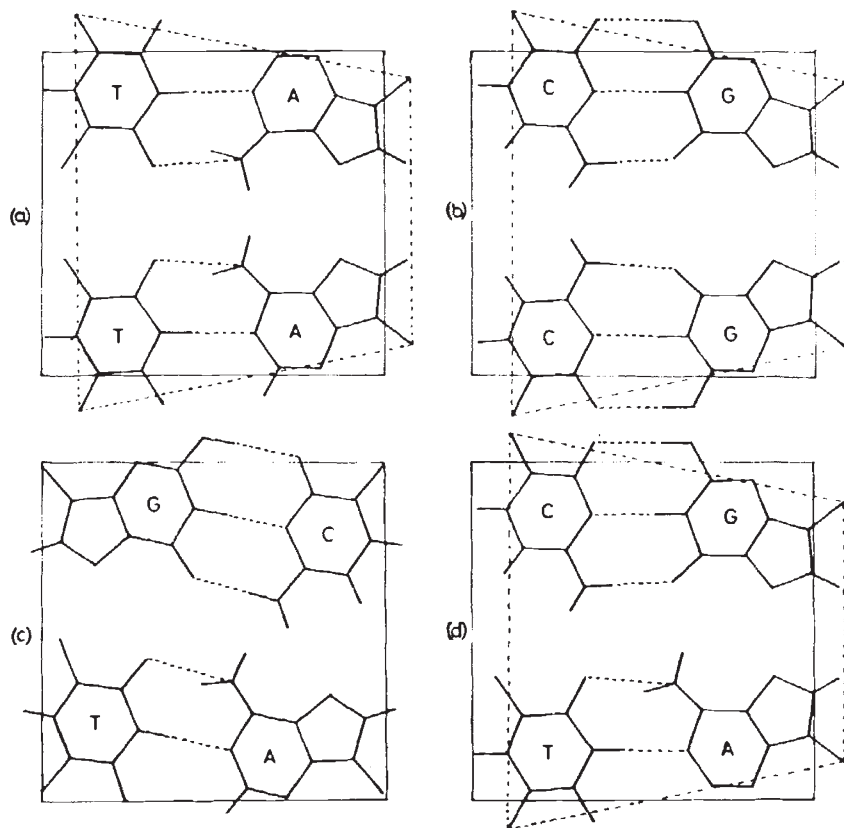review of such work edited by Sundaralingam and Rao, 1975).

FIG. 3.—This figure shows how attempts to pair unlike base pairs, or to pair like base pairs
    other than about a dyad axis results in irregular arrangements or in the impossibility of
    forming hydrogen bonds.

The specificity of the model lies in the pairing of identical base pairs,
about dyad axes, and on the fact that this pairing results in the glycosyl
bonds lying close to the corners of squares of the same size allowing regular
helical structures to be formed. It is of interest to compare this basis for
specificity with that of the double helix itself which depends on pairing
between bases and on the fact that this results in an equal separation of
glycosyl bonds.

Specificity of the four-strand structure also lies in the fact that in it the
sugar phosphate chains can be in good van der Waals contact along their
entire length. This contact would contribute to stability but it also, though,
suggests that for formation and stability of the structure that ionisation of the
phosphate groups would have to be suppressed in some way. There is,
therefore, the possibility of the structure being stabilised by the binding

of other substances such as ions or basic protein. This is an interesting point in the models' favour as it would be desirable for a structure with the role which is proposed for it to depend for stability on at least one factor which was significantly different from those required for the stability of the double helix itself.

(i) *Symmetry of the model*

The model is interesting from the point of view of its symmetry and possibilities for symmetry given certain base sequences. It has, or can have, the highest possible symmetry of any nucleic acid model so far described.

The model can always have 222 symmetry relating the sugar phosphate chains. In other words it can have three mutually perpendicular two-fold axes of rotation, or dyad axes. If, however, the structure involves palindromic (reverse tandem) base sequences then the whole structure can have this 222 symmetry.

Any given symmetry for a nucleic acid structure suggests the possibility of the binding of a protein with the same or related lower symmetry. One can envisage the structure being clasped by an enzyme with 222 symmetry. Similarly one can envisage one class of enzyme sensitive to the sugar phosphate symmetry only, another sensitive to base sequence and yet another sensitive to a palindromic base sequence. Sobell (1973) in an interesting paper discusses symmetry in protein nucleic acid interactions and its genetic implications.

The deep grooves of the constituent double strands are used in formation of the four-strand structure, therefore only the " shallow " or " narrow " grooves of the structure expose the base sequence to the outside of the molecule. The structure has two such identical grooves, one provided by each of the constituent double helices. The four-strand model therefore suggests that any specific association with protein would involve these " shallow " grooves.

(ii) *Possible modifications to the model*

The structure has been described as a helix with the parameters of a Watson-Crick double helix. What has been suggested is a structure of a general type. I believe, however, that it is valuable to think of nucleic acid structures as, at least under certain conditions, being anything but rigid and indeed as widely deformable (McGavin, 1973). One only has to look at some of the many published electron micrographs to see that this is true.

There are several ways in which the four-strand structure could be modified, for example by using slightly different base pairing schemes, by using U rather than T and ribose rather than deoxyribose and by using different numbers of repeating units per turn of the helix.

Space filling models of a left-handed version of the specifically paired four-strand structure can be built. Such models include of course left-handed double helices with Watson-Crick base pairs. Such double-helical models have been built before but the weight of opinion is against them and in favour of the right-handed structure because of the impossibility of building structures with the parameters of A form DNA in a left-handed form (see for example Fuller *et al.*, 1965) and as it is known that transformations between the A and B forms can occur. Single crystal analysis of a dinucleoside shows right-handed structure. (Rosenberg *et al.*, 1973).

From a limited amount of model building of this kind it seems that neither the right nor the left-handed " form " of B DNA can be unwound to any extent without first being extended. It seems, therefore, that a process of transformation from a right to a left-handed helix would be rather complicated.

3. APPLICATION OF THE MODEL TO GENETIC RECOMBINATION

(i) *Exact matching and recognition of double strands, initiation of recombination*

It should be easy to see how the structure just described might explain the very close matching of nucleic acid molecules which occurs in recombination and which is known to be exact right down to the level of single base pairs. The structure similarly might explain, at least part of, the initial recognition process.

Initial recognition might be analogous to the process where complementary single strands find and pair with each other, although there is a question over the flexibility of double strands, but as mentioned above some of the many published electron micrographs show that considerable flexibility is possible.

What the structure can apparently do best is explain the precision of matching of double strands and hence the precision of the breakage and reunion process. A four-strand region, or a length over which such a region was transmitted, might be the region of effective pairing discussed by Pritchard (1955). Such a region need not be particularly long.

A simple picture of breakage and reunion, resulting in cross-over, can be introduced at this point. In this, a four-strand structure simply shears at one level and then rejoins after the two parts are rotated with respect to each other by 180 degrees. If the rotation is by 0 or by 360 degrees then there is no cross-over. There would, in the simplest model of this kind, be equal chances of there being, or not being, crossing over.

None of this is inconsistent with the possibility of specific base sequences being involved in recombination. Such sequences might be recognised by the parts of a multimeric protein which could then associate with the resulting formation of a length of four-strand pairing (fig. 6). It is not intended to suggest here that this is a complete description of initiation. More than one factor seems likely to be involved in the initial process of recognition. In addition to mechanisms such as the one discussed here, it seems worth considering the kind of mechanism suggested by Mosevitsky (1973), who suggests that varying sequences may result in small conformational changes and hence in unique structural profiles.

(ii) *The possibility of including mismatches in the structure*

As shown above, the non-specific pairing of double strands is unlikely. The inclusion in the model of separate mismatches of base pairs is, however, a different matter. Details of the consequences of four-strand involvement in recombination would depend on the extent to which this was possible. I do not intend to discuss the several possibilities which there are. Mismatches, though, might occasionally be included in a length of four-strand structure, but they also might limit the length of a four-strand region or

limit the length of a pair of double-strand molecules along which a length of four-strand complex was transmitted.

(iii) *The possibility of the transmission of a four-strand region along a pair of double-strand molecules*

As just mentioned, a length of four-strand structure might be transmitted along a pair of double-stranded molecules. Such transmission would be rather like the transmission of a cross-strand exchange by rotary diffusion (Meselson, 1972). Detailed study of such a process would be difficult but it seems that, given a moderate degree of flexibility in double-stranded molecules, it would be possible. Such a process would be hindered or stopped by mismatches between base pairs.

(iv) *The problem of plectonemic coiling*

While plectonemic coiling, in relation to separation of strands and in relation to structure formation, is a problem, I do not think that it is funda-mentally different to the problem met with in the double helix itself and it can be ignored, surmounted or avoided in similar ways. One can, for example, postulate " nicks ", left-handed structure involvement or the involvement of a specific protein.
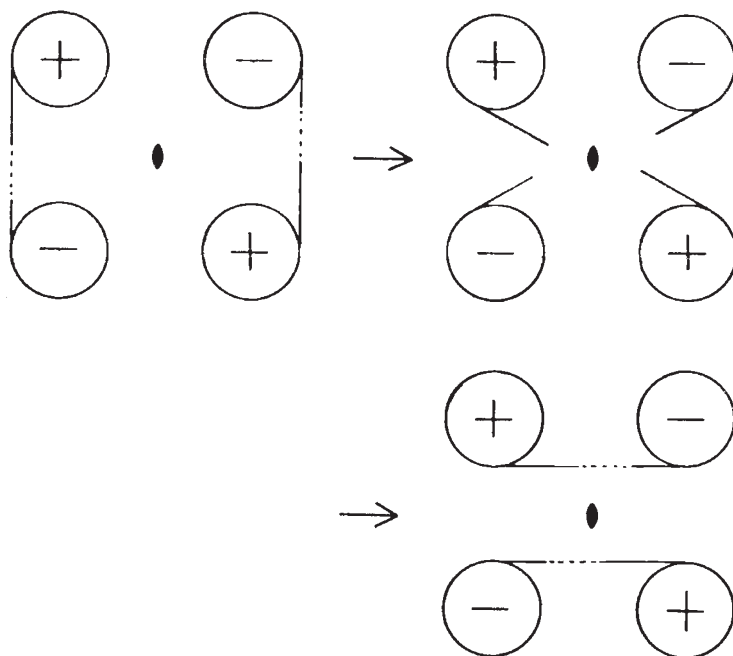

FIG. 4.—Diagram of " exchange of base pairing " at one level in a four-strand structure. A pair of linked circles represent a Watson Crick base pair. The signs indicate the polarity of the strands. (*a*) Shows two base pairs arranged as in the four-strand structure with " deep " grooves together; (*b*) shows a stage in a process of exchange of base pairing where the " Watson-Crick " bonding is broken; (*c*) formation of the alternative Watson-Crick bonding, note, however, that there is no pairing between the two base pairs as the " shallow grooves " are now together.

(v) *Interference, marker effects, polarisation*

A complete model for recombination must explain several phenomena. Some of these are Interference, Marker Effects, and Polarisation of Effects. While these are not considered in detail here, an explanation of some of them might be sought in the extent of formation or the extent of transmission of lengths of four-strand structure.   Such an approach to the explanation of interference seems similar to that of Pritchard (1955) except for the mode of association.
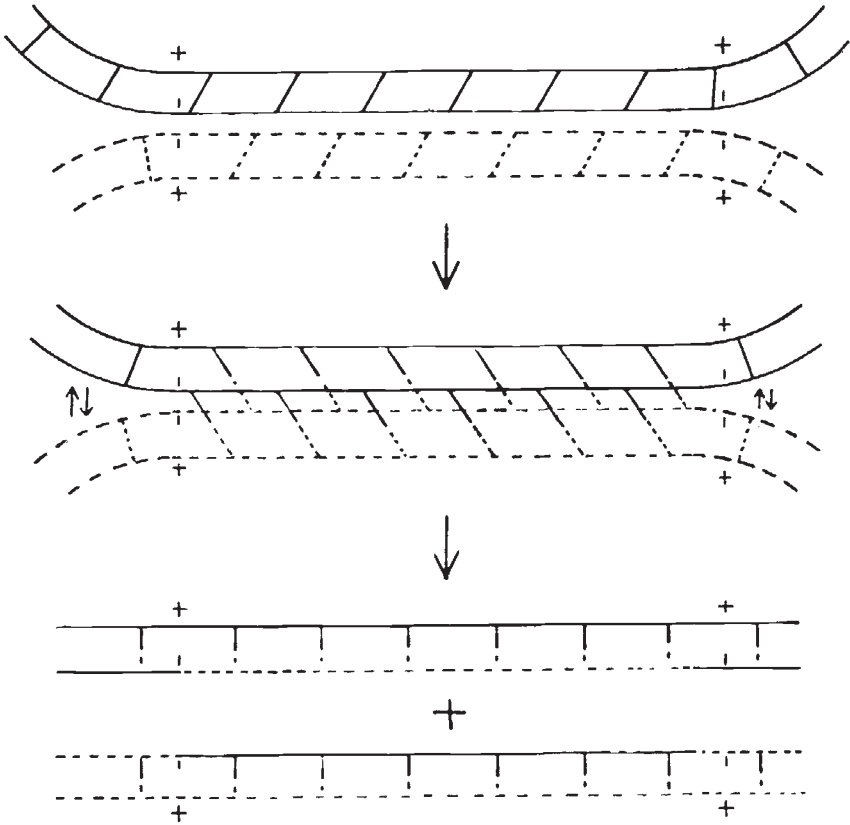

Fig. 5.—This figure shows the process of exchange of base pairing viewed from the side; (*a*) is a pair of double-stranded molecules involving a region of four-strand structure, (*b*) shows the same pair following " exchange of base pairing ", (*c*) shows the separated " hybrid " molecules.   No crossover is illustrated; had breakage and reunion involved the other pair of strands at one end of the four-strand region then cross-over would have occurred.

Some of the effects observed in work on recombination could therefore depend for explanation on the extent of formation or transmission of a region of four-strand pairing.   Explanation of several other effects such as gene conversion and post-meiotic segregation might follow if an " exchange of base pairing " is assumed, resulting in hydrid DNA formation.

(vi) *Hybrid DNA formation and the possibility of the exchange of base pairing*

The formation of a region of four-strand structure might be followed by an exchange of base pairing resulting in several different effects. Such an exchange is shown in figs. 4 and 5. Fig. 4 shows the exchange diagrammatically at one level in a four-strand region. Fig. 5 shows a region of pairing and exchange, looking at right angles to the length of a pair of double-strand molecules. Note that fig. 4 shows how the process could promote the separation of the newly formed hybrid double strands. Fig. 6 illustrates the possible involvement of an enzyme in the process.
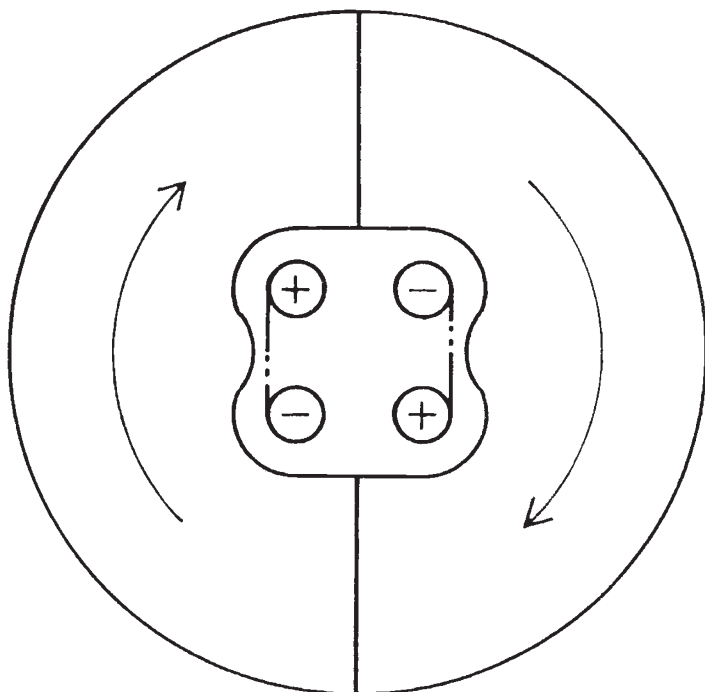

Fig. 6.—This shows how a four-strand region might be clasped by a symmetrical enzyme molecule. The arrows are included to show the symmetry of the enzyme when viewed along one of the dyad axes.

Such a process is closely analogous to splitting down into single strands followed by the mutual recognition of complementary single strands which is a feature of several of the published models. Recognition, however, is at the earlier stage of pairing of the homologous double strands, no separation of single strands ever takes place and the structure remains symmetrical throughout the formation of two heteroduplexes. Model building shows that there are no obvious stereochemical problems associated with a process of this kind.

Such an exchange of pairing would, of course, have to be followed by the breakage and reunion of strands to allow the separation of the two double-stranded molecules as indicated in fig. 5. Depending on which pairs of single strands are broken and rejoined there either would or would not be crossing over.

4. DETECTION OF THE STRUCTURE

If four-stranded structures of the kind discussed here do exist then how could they be detected? Several experimental approaches might point towards the presence of four-strand structures. These would depend on the dyad axis, on the presence of four strands and perhaps particularly on the mass per unit length being twice that of a normal double helix. For example, an oriented sample of a four-strand structure would give a distinctive X-ray diffraction pattern (McGavin, 1971; Arnott, Chandrasekaran and Selsing, 1975) and even a sample which was not oriented might give distinctive continuous scatter. Density and sedimentation properties of four-stranded material would be distinctive. It would be possible to obtain material in which 0, 1, 2, 3 or 4 strands were labelled. Certain optical properties such as hypochromicity and optical rotation would probably be distinctive. Guschlbauer (1975), for example, mentioned that he believed poly G to be four-stranded on the basis of circular dichroism. It seems also worth considering techniques involving the kinetics of breaking of strands (Cavalieri and Rosenberg, 1963). I am not convinced that Cavalieri and Rosenberg's results were ever satisfactorily explained other than by four-strand structures.

While specifically paired four-strand structures could, therefore, probably be detected there are a number of reasons for believing that they might be missed. These are, the structure is very similar to the double helix in many ways, being simply a specific association of two such helices, many of its properties would therefore be close to those of the double helix itself. The structure, for example, has a diameter, perhaps at first sight rather surprisingly, not much greater than that of the double helix itself. Although the mass per unit length of the structure is twice that of the double helix this puts it rather far outside the normal range of observation and if structures with this property were noticed then unless one was looking for four-strand structures they might be rejected as something totally different. In addition, as suggested already, the structure would depend for stability on the presence of specific or non-specific basic protein, small molecules or ions, in other words on conditions widely different from those which are sufficient for the stability of the double helix itself.

5. CONCLUSION

In conclusion, how seriously should we take models of this kind?

Consider it first as a molecular model. There is perhaps an unfortunate tendency to give all molecular models the same weight. Compact regular models, however, as this one is, fall into a rather special category. Several compact regular molecular models have subsequently been shown to occur naturally. The best example of this kind is perhaps the alpha helix of proteins, which was suggested first on the basis of model building by Pauling, Corey and Branson (1951). There are also in the literature a number of other polypeptide helices which appear to be stereochemically satisfactory but which have not yet been shown to occur. The four-strand model seems at least to come into this last category, and its specificity makes it of rather more interest.

From the point of view of genetic recombination the model is at least as detailed as most others and is the only one I know of which suggests in detail how specific pairing of intact homologous double strands might occur.

6. References

ARNOTT, S., CHANDRASEKARAN, R., AND SELSING, E. 1975. The variety of polynucleotide helices In Sundaralingam and Rao, S. T. *Structure and Conformation of Nucleic Acids and Protein Nucleic Acid Interactions.* University Park Press, Baltimore, U.S.A.

CAVALIERI, L. F., AND ROSENBERG, B. H. 1963. *Progr. Nucleic Acid Res., 2,* 1.

FINCHAM, J. R. S., AND DAY, P. R. 1971. *Fungal Genetics,* 3rd Edition. Blackwell, Oxford.

FULLER, W., WILKINS, M. H. F., WILSON, H. R., AND HAMILTON, L. D. 1965. The molecular configuration of DNA IV; X-ray diffraction study of the A form. *J. Mol. Biol., 12,* 60-80.

GIERER, A. 1966. Model for DNA and protein interaction and the function of the operator. *Nature, 212,* 1480-1481.

GUSCHLBAUER, W. 1975. In Sundaralingham and Rao, p. 594.

HOLLIDAY, R. 1964. A mechanism for gene conversion in fungi. *Genet. Res., 5,* 282-304.

KUBITSCHEK, H. E., AND HENDERSON, T. R. 1966. DNA replication. *Proc. Nat. Acad. Sci., Wash., 55,* 512-519.

KUSHEV, V. V. 1975. *Mechanisms in Genetics Recombination.* Plenum.

LÖWDIN, P. 1964. In *Electronic Aspects of Biochemistry,* ed. B. Pullman, p. 167. Academic Press, New York.

MCGAVIN, S. 1971a. Models of specifically paired (homologous) nucleic acid structures. *J. Mol. Biol., 55,* 293-298.

MCGAVIN, S. 1971b. A four strand nucleic acid model with a specific pairing of like Watson-Crick double helices and its properties. *Proc. First European Biophysics Congress,* pp. 259-262.

MCGAVIN, S. 1973. An attitude to nucleic acid models. *Bull. Math. Biol., 35,* 407-409.

MCGAVIN, S. 1977. In preparation.

MESELSON, M. 1972. Formation of hybrid DNA by rotary diffusion during genetic recombination. *J. Mol. Biol., 71,* 795-798.

MOSEVITSKY, M. I. 1973. Conformational diversity of the double stranded DNA as a basis for the homologous recognition and synapsis. *Studia Biophysica, 40,* 93-98.

O'BRIEN, E. J. 1967. Crystal structures of two complexes containing guanine and cytosine derivatives. *Acta Cryst., 23,* 92-106.

PAULING, L., COREY, R. B., AND BRANSON, H. R. 1951. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Nat. Acad. Sci., Wash., 37,* 295-211.

PRITCHARD, R. H. 1955. The linear arrangement of a series of alleles of *Aspergillus nidulans. Heredity, 9,* 343-371.

SIGAL, N., AND ALBERTS, B. 1972. Genetic recombination: the nature of a crossed strand exchange between two homologous DNA molecules. *J. Mol. Biol., 71,* 789-793.

SIMUNDZA, A., SAKORE, T. D., AND SOBELL, H. M. 1970. Base pairing configuration between purines and pyrimidines in the solid state. *J. Mol. Biol., 48,* 263-278.

SOBELL, H. M. 1972. Molecular mechanism for genetic recombination. *Proc. Nat. Acad. Sci., U.S.A., 69,* 2483.

SOBELL, H. M. 1973. Symmetry in protein—nucleic acid interaction and its genetic implications. *Advan. Genet., 17,* 411-490.

SUNDARALINGAM, M., AND RAO, S. T. 1975. Editors. *Structure and Conformation of Nucleic Acids and Protein Nucleic Acid Interactions.* University Park Press, Baltimore, U.S.A.

WATSON, J. D., AND CRICK, F. H. C. 1953. Molecular structure of nucleic acids: a structure of deoxyribose nucleic acid. *Nature, 171,* 737-738.

WHITEHOUSE, H. L. K., AND HASTINGS, P. J. The analysis of genetic recombination on the polaron hybrid DNA model. *Genet. Res., 6,* 27-92.