A computer graphics study of multistranded DNA models

Stewart McGavin

Chemistry Department, University of Dundee, Dundee, Scotland

H.E. Bosshard, C. Carlson and J. Postma

EMBL, Heidelberg, FRG

A range of related nucleic acid models has been built using the computer graphics system at EMBL (Heidelberg). The relationship between these models is considered in terms of a simple "core" structure. One of these models is a four-strand structure in which two duplexes of the Watson-Crick kind are specifically related by a twofold rotation axis and which has already been discussed in some detail.

The models fall into two principal classes. In one, each sugar phosphate chain is attached to every layer of bases in the "core" structure; in the other, each chain is attached to alternate layers of bases.

Several of the models discussed have specificity both of complementarity and of identity.

Keywords: DNA models, multistranded DNA models

INTRODUCTION

We have been interested for some time in a range of DNA models that are closely related and that involve specificity of interaction between components. Most of these models have not so far been considered quantitatively. In this paper several such models built in detail by computer modeling are described and discussed. The paper places emphasis on the use of a "core" structure of base "tetrads" in building and relating these models (McGavin[1,6]).

Some of these models are of biological interest (McGavin[1–6,8]). One such model, shown in Color Plate 2a, which has been discussed already in some detail, is a four-strand structure in which Watson-Crick duplexes are paired specifically about a dyad axis coincident with a common long molecular axis and with major grooves in continuous and specific contact. This has proved to be of considerable interest, perhaps particularly in relation to models for genetic recombination (see, for example, McGavin,[4,8] Nash *et al.*,[9] Kikuchi and Nash,[10] Nash and Pollock,[11] West *et al.*[12,13]).

It has also been suggested that the basic four-strand structure (Color Plate 2a) might be formed from the arms of cruciform structures (McGavin[4,19]). For further discussion of cruciform models, see, for example, Gierer[14] and several papers by Lilley (e.g., 15).

This paper is related to an earlier paper published in this journal on models for recombination junctions at the molecular level (McGavin[16]). The four-strand model—which, in a sense, is the basis of this paper—is also the basis of these models of junctions. This paper, however, covers quite different ground. Here, we discuss a range of regular helical models and describe work where we have taken advantage of the possibilities of computer model building and graphics.

(a) to improve and discuss in more detail the basic four-strand model and variants of this
(b) to build in quantitative detail several models that were built earlier only as space-filling models or that were merely outlined or surmised in terms of the core structure (McGavin[6]).
(c) to discuss some models that are completely new.

We believe that DNA and nucleic acids in general are of such interest that any regular, stereochemically reasonable and compact models that can be constructed are worth studying, especially where these involve specificity of interactions between strands. The models discussed here involve complementary specificity of the normal Watson-Crick kind. In addition to this, some of them involve what we have called a specificity of identity. An example of this is the basic four-strand model itself (Color Plate 2a), where two identical duplexes are paired together about a dyad axis.

APPARATUS AND METHODS

Relationship between models

The basis of the range of models discussed in this paper can be thought of as the core of the specific four-strand

Address reprint requests to Dr. McGavin at the Chemistry Department, University of Dundee, Dundee, Scotland, UK
Received 6 June 1988; accepted 21 June 1989

model mentioned above. We have indeed come to think of the use of this core as a model building "game." The use of this core was first discussed in a study restricted to work with space-filling models (McGavin[6]).

The "core" is a stack of base "tetrads" in which the glycosyl bonds are distributed approximately at the corners of squares. Each of these "tetrads," or layers in the core, is formed from two identical Watson-Crick base pairs related to each other by a twofold axis. These pairings have been discussed fairly extensively.[3,25,26] In pairing of GC with GC, each pyrimidine is hydrogen bonded through N4 to either O6 or N7 of the purine of the dyad related base pair, whereas in the pairing of AT with AT O4 of each pyrimidine is hydrogen bonded to either N6 or N7 of a purine of the symmetrically related base pair.

The "core" is an impressively solid structure that appears in some configurations to be almost cylindrical (see Color Plate 1 and Reference 6). A range of models can be built by rearranging the essentially square layers of bases relative to each other by rotation about the long molecular or stack axis. Following such rearrangements, the sugar phosphate chains are reattached.

Two principal classes of models can be built in this way. In the first, the connections of sugar phosphate chains are to successive layers as in the original four-strand model. In the second class, connections are made to alternate layers, and the resulting models therefore involve extended chains and a systematic mutual intercalation of layers of bases.

All the immediately resulting models, since they use the same base tetrads, have both complementary specificity of the Watson-Crick kind and specificity of identity. The models can though be broken down into parts in different ways, and there is some interest in considering the parts themselves. If it is accepted that a model is stereochemically reasonable, then any of its parts will also be stereochemically reasonable, even though they may be unlikely, at least in isolation, from an energetic point of view.

Model building

In this work we frequently had to make major alterations to models. Where this resulted in unsatisfactory distances or angles between residues (the break between residues is at O3...P), these were brought back as close as possible to the standard values by changing torsion angles "manually." Least-squares regularization was then carried out by use of the stereochemically restrained refinement program NUCLSQ, to correct distances and angle distances deviating from ideal values by more than two standard deviations (one standard deviation is .02 Å for bond lengths and .03 Å for angle distances).

By "manually" we mean use of a facility "TOR" within FRODO. This makes no reference to torsional barriers. The main point of its use here was to bring O3 and P atoms of successive residues sufficiently close together to allow use of NUCLSQ. Tables 1 and 2 are provided to allow assessment of torsion angles resulting from the use of NUCLSQ.

It was usually found that the regularization process did not destroy the helical regularity except for small changes at the ends of chains and that, where required, a satisfactory regular helical structure could be constructed from the central one of the five layers that we normally worked with.

In this work there are no electron density maps or parameters from fiber diffraction patterns to fit that would focus attention on limited regions of conformational space. Several fairly rigid simplifications were therefore introduced and maintained in this model building. In spite of these rigidities, the models described come close to being stereochemically acceptable. Relaxation of these rigidities might allow improvement of models but would introduce complications that would at present be too time-consuming.

Simplifications

The main simplifications are as follows:

(a) C2′ endo sugars were used normally. While C2′ endo is a commonly used pucker for B-form DNA models, it seems likely that in other structures, such as some of the models described here, other sugar puckers would occur.

Clearly a great deal more work could be done to investigate the effect of different puckers and of combinations of puckers, whereas in Color Plate 3 nonequivalent residues are involved.

(b) For simplicity and also because there was no obvious reason for tilting, planar and coplanar bases with their planes normal to the helix axis were normally used.

It is probable that some contacts could be improved by tilting, and it should be noted that tilting is a feature of most nucleic acid models. Tilt of bases in duplex structures may, however, be related to the optimizing of base pair stacking, and exactly similar considerations do not apply where interduplex interactions also occur.

(c) A rectangular, essentially square distribution of glycosyl bonds is assumed (some discussion of models involving relatively sheared base pairs is however given below).

(d) Most work was done with GC tetrads in which GC base pairs are related by a twofold axis coincident with the long molecular axis. The assumption is made that such tetrads might be replaceable by tetrads involving CG, AT or TA base pairs (see McGavin[3]).

Graphics system

All the models described in this paper were built using the computer graphics facility at EMBL (Heidelberg). The initial coordinates for the basic four-strand model, however, were obtained by use of wire models (see papers cited above).

The Graphics System at EMBL, when most of this work was done, was based on an Evans and Sutherland Multi Picture System (MPS) hosted by a Digital Equipment Corporation VAX-11/785. Most of the work described involved use of an Evans and Sutherland CSM calligraphic color scope. Interaction was via a digitizing tablet and a normal computer terminal.

Full tachystoscopic stereo is available at EMBL. In this, rapidly alternating left and right eye views are observed by means of a slotted rotating cylinder. Each eye is provided with the appropriate view of the model, and the three-dimensional (3D) structure is easily perceived. Use of stereo proved to be of considerable value in interactive modeling.

Table 1. Torsion angles for some of the models discussed with sugar phosphate chains attached to successive layers

Model		α	β	∂	δ	ϵ	ξ	χ
32 degrees	C	−60	−159	37	146	168	−98	−112
	G	−59	−158	35	145	168	−96	−111
Sheared	C	−65	−165	45	144	174	−105	−112
	G	−48	−164	27	145	172	−100	−110
Alternating angle	C	−50	−177	36	145	168	−99	−111
	G	−45	−156	25	144	−179	−110	−110
Alternating syn and anti Antiparallel	C3	−56	−157	35	144	26	123	−113
	G4	−37	−158	−121	145	164	−83	+65
	G8	−51	−156	28	144	26	122	−112
	C9	−36	−157	−121	145	163	−79	+64
Alternating syn and anti Parallel	C3	−31	−149	6	145	163	−84	+71
	G4	−47	−152	22	143	170	−89	−108
	G8	−47	−152	22	144	171	−91	−108
	C9	−31	−149	6	145	164	−85	+71

Note: Alpha, beta, gamma, delta, epsilon and zeta are O3′ − P − O5′ − C5′, P − O5′ − C5′ − C4′, O5′ − C5′ − C4′ − C3′, C5′ − C4′ − C3′ − O3′, C4′ − C3′ − O3′ − P and C3′ − O3′ − P − O5′ respectively. chi is O4′ − C1′ − N9 − C4 for purines and O4′ − C1′ − N1 − C2 for pyrimidines

Table 2. Torsion angles for the series of extended models with differing rotation angles

	α	β	∂	δ	ϵ	ξ	χ
16	149	152	−158	147	−175	−102	−100
12	132	153	−137	147	−180	−109	−92
8	119	156	−120	146	179	−117	−84
4	106	160	−109	146	−180	−124	−77
0	100	167	−104	146	−177	−131	−71
−4	100	166	−107	146	−177	−132	−66
−8	108	163	−117	146	−178	−131	−62
−12	121	159	−132	146	180	−129	−58

Programs

Programs that were readily available at EMBL were used even though they may not always have been the best conceivable for the task at hand. They were:

(1) FRODO (originated by T.A. Jones; see, for example, Jones[17]). The version in use at EMBL is an extensive redevelopment of the program.

(2) NUCLIN and NUCLSQ. These are programs for the stereochemically restrained least-squares refinement of nucleic acid structures. They were modified by E. Westhof (Strasbourg) from G. Quigley's REFIN (MIT) and W. Hendrickson's PROLSQ, which are for the refinement of protein structures. NUCLIN sets up the stereochemical restraints for the refinement of nucleic acids by the Hendrickson-Konnert method. NUCLSQ is the refinement program (for discussion of stereochemically restrained refinement, see, for example, References 36, 37, and 38).

(3) Other utility programs developed at EMBL and elsewhere.

Photography

The diagrams reproduced were photographed from the screen of the MPS color scope. They were taken using a Rolleiflex SLX camera with Kodak Ektachrome Professional Film (daylight).

The MPS lets users choose from a wide range of color hues and saturations. Blue and orange were chosen to contrast the duplexes. O4′ atoms were colored the conventional red for oxygen because they acted as arrows giving an immediate indication of the polarity of chains. C1′ atoms were colored green because they were often used as guides in modeling.

Reproduction of models

The models discussed here were developed sporadically over a period of years and in different ways. We have coordinates for all of them, and torsion angles for some of them are given in Tables 1 and 2. Coordinates for a version of the four-strand model were published some time ago.[3]

The distribution of residues in all the models, however, could be generated from a single Watson-Crick base paired pair of residues by translations and other symmetry operations. Connections between residues could be made as described here or in other ways.

RESULTS: MODELS OF THE FIRST CLASS

The basic four-strand model

Color Plates 2a–2d show the basic four-strand model and subdivisions of it. The essentials of this four-strand model have already been discussed, and therefore discussion of it here is limited to recent attempts to improve it.

Some time was spent in trying to find an optimum rotation angle between successive layers of bases. The best structures found have a rotation angle between successive layers of between 31 and 32 degrees. Note that an angle in the 31- to 32-degree range was found by Arnott *et al.*,[21] in four-strand structures of the synthetic polynucleotides poly G and poly I. These polymer structures, however, have parallel chains.

In Color Plate 2a with a 32-degree rotation angle, a few contacts involving C2′ remain short, in particular C2′ to C6 is 2.89 Å (a single torsion contact associated with the angle chi) and C2′ to P is 3.15 Å (a multiple torsion contact between successive residues). See an earlier discussion of this (McGavin[16]).

Structures with relatively sheared base pairs

Retaining the dyad axis relating base pairs, models were built in which the angle of shear between base pairs in a tetrad was systematically changed. This was done by shearing the base pairs of an unwound helix and then reconstructing a helix of the desired rotation angle. Base pairs were moved equally and oppositely with respect to the helix axis.

Most of this work was done with GC paired to GC. A .6 Å total shear such that C1′ of C (the pyrimidine) is at the vertex of an acute C1′ C1′ C1′ angle (formed by C1′ atoms of purine, pyrimidine and purine, respectively) allowed improvement in the H bond vector angles relating the base pairs. The resulting structure involved a base pairing that was close to that actually observed in a crystal structure by O'Brien[22] (see also McGavin[3]). A difference, though, is that in that structure there was some tilt of base pairs relative to each other. This model also showed an improvement in torsion contacts; thus the C2′ to C6 contacts are 2.9 Å, and there are no short multiple torsion contacts.

This sheared structure tended to move on regularization to a smaller radius; thus the interduplex C1′ to C1′ distance was reduced to about 10.5 Å. This would not be satisfactory for AT to AT pairing.

Pairing between AT and AT can also be optimized by shearing of one base pair with respect to the other. Here a concern is the accommodation of the methyl group of thymine. Again, satisfactory structures result in which the bases are translated (sheared) with respect to each other by .6 Å. As just implied, a rather larger radius is required than with GC to GC pairing. An inter duplex C1′ to C1′ distance of 10.71 Å gave a methyl C to N7 distance of 2.95 Å.

Alternating structures (alternation of C1′ C1′ C1′ angles)

Models were built in which dyad-related sheared GC base pairs were alternated along the length of the four-strand structure with dyad-related sheared CG base pairs. These models were constructed by flipping over alternate layers of the model discussed in the preceding section (alternate tetrads of the sheared structure described in the above section were rotated by 180 degrees about axes perpendicular to the C1′ C1′ vectors of the base pairs and to the long helix axis of the structure).

In these models G and C alternate along each strand, and the C1′ C1′ C1′ angles are alternately obtuse and acute along each strand.

Models with alternating syn and anti base orientations

Models were built in which bases have alternating syn and anti orientations along each chain by rotating alternate tetrads of the core of the basic four-strand model by 90 degrees about the long axis of the structure. (Note that the rotation of all base tetrads of the core by 90 degrees, that is, rotation of the entire core by 90 degrees, results in models in which all base orientations become syn. For a discussion of such models, see papers by Hopkins, Ref 20.)

A striking feature of models of this kind (see Color Plates 3a and 3b) is the alternation of Watson-Crick base pairing from one strand to two other strands.

Color Plate 3a retains the polarity of strands of the parent model, whereas 3b has all strands parallel.

The transformation from 3a to 3b was achieved by rotation of one sugar phosphate chain of each duplex by 180 degrees about axes perpendicular to the length of the molecule. This changes the polarity of these chains and also changes base orientations from syn to anti and vice versa.

In 3b all four strands are equivalent. The structure has, like the Z DNA structure (Wang *et al.*[23]) and like 3a, an alternation of syn and anti orientations of bases along each strand. A significant difference to 3a however is that whereas 3b has two syn and two anti orientations at each level, in 3a all base orientations at any one level are the same.

As in building 3a and 3b, GC base pairs were used throughout, and the models have an alternation of G and C along each chain. Because of the similarity of base tetrads and the flexibility of sugar phosphate chains, it is likely that any sequence of bases could be accommodated in 3a and 3b.

As 3a stands at present, the angle at 03′ is strained (the C3′ to P angle distance is 2.75 Å, compared with the "ideal" value of 2.65 Å).

In 3b the torsion angle gamma for the C (ie the syn) residues is too small and there are short contacts associated with this.

In this work we have not tried to relate models directly to Z DNA structure. There might, therefore, be some interest in extending work to left-handed screws and in changing syn orientations to anti and vice versa.

Torsion angles for some of the models in class 1 are given in Table 1.

RESULTS: MODELS OF THE SECOND CLASS

Extended models and their flexibility

Color Plates 4a to 4h show a series of results obtained with an extended four-strand model in which the spacing between layers is 6.80 Å. The starting structure for the series was the model with zero degrees rotation between successive layers (4e).

A four-strand model with 32 degrees rotation between successive layers was unwound (stacked), and the separation between base tetrads was then increased from 3.40 to 6.80 Å. Connections between residues were then made as well as possible by the manipulation of torsion angles. Other members of the series were made by changing the rotation angle between layers in steps of four degrees. No further direct manipulation of torsion angles was found necessary, but each member of the series was put through least-squares regularization.

For this work auxiliary programs were used that extract the central layer of the five used and then produce a helix of a specified rotation about, and translation (in this case, 6.8 Å) along the helix axis. NUCLSQ does not readily move planes and did not significantly change the axial translation of the repeated units (note, for example, that the eight-strand model that involves intercalation was constructed from the 12-degree member of this series).

The intention of this work was to confirm the unlikelihood of any great hindrance to the rotation of such an extended structure between limits that we believe would be about +16 and −12 degrees. Bearing in mind that we are dealing with essentially separated and extended sugar phosphate chains, this indication is given by the torsion angles shown in Table 2. In addition, a rough assessment of the quality of these models is given by the number of distances (bond length and angle distances) that were output by the program as deviating from "ideality" by more than two standard deviations. The +12, +8, +4, 0, and −4 degree models showed no distances in this category, while the +16, −8 and −12 degree models showed indications of more strain.

It is unlikely that such extended structures, in the absence of flat spacers (that is, of intercalation of some kind), would exist as such, although it is interesting to speculate on conditions under which they could so exist. These models were built to demonstrate the properties that extended models such as those discussed below might have. They are also of interest in consideration of "neighbor exclusion" in the intercalation of small molecules by DNA (see below).

One of the best models in the series is the 12-degree one. This was used as the basis of the models discussed in the next section.

An eight-stranded model

An eight-stranded model was built by translating one of two coincident 12-degree models (4b) by 3.4 Å along the common helix axis and rotating this by 45 degrees about the same axis relative to the other.

In the resulting eight-strand model (5a), the base tetrads of one four-strand component are systematically intercalated by those of the other similar component.

With 45 degrees rotation between the two components, the sugar phosphate chains are all clear of each other, and the two components (each as 4b) are free to rotate relative to each other from this 45-degree position within limits set by the formation of contacts between the sugar phosphate chains. Were such eight-strand structures to occur, one would expect such contacts to be optimized.

While the eight-strand model can be twisted between limits of the kind just described with no change of the helical parameters of the four-strand mutually intercalated components, the whole structure could also change in rotation angle shown in the above section for one of the four-strand components.

DISCUSSION: MODELS OF THE FIRST CLASS

The basic four-strand model

Possible biological roles for the basic four-strand model (2a) have already been discussed fairly extensively (see, for example, References 16 and 19 and other citations given in these papers; Hopkins[20] also provides a useful recent review of possible roles of multistranded models). Such models have already been useful in the formulation of theories, particularly on parts of genetic recombination and in suggesting experiments.

Alternating structures

The two different kinds of alternating structures that we discuss are both a consequence of four-strand structure.

In the first kind, where C1′ C1′ C1′ angles are alternately acute and obtuse, these angles become defined only on the addition of a third or of a fourth strand to a duplex. In the second, where there is an alternation of syn and anti bases along each strand, the alternation of pairing between two strands that is a feature of these models again requires four strands.

3a and 3b are particularly interesting. 3b is, among the models which we have built, the one that comes the closest to the four-strand structure proposed by Arnott *et al.*[21] for certain homopolynucleotides in that it has parallel strands. Note also in this connection recent work by Sen and Gilbert.[32]

A feature of 3b is that the strands are equivalent in the sense that they cannot be distinguished unless base sequence is taken into account.

Although the following suggestion would be inconsistent with the exact symmetry of the core structure, 3b gives rise to the idea of a loose four-strand structure in which diagonally opposed strands are of different sequence. In this each strand would be paired specifically by Watson-Crick pairing to alternate residues of its two nearest neighbors but would not be related directly in sequence to the opposed strand.

Subdivisions of the basic four-strand model

We have become increasingly interested recently in the duplex structure of Color Plate 2c and the three-strand structure of Color Plate 2d.

Recently we suggested the use of 2c as an intermediate allowing the formation of four-strand structure by Watson-Crick pairing. This idea was developed (McGavin[19]) to produce a scheme for recombination consistent with the topological implications of recent results of Griffith and Nash[31] obtained with the lambda bacteriophage system.

Similarly, a three-strand structure (2d) might be formed by Watson-Crick pairing of a single strand to 2c. This seems of interest in relation to models for recombination that involves the formation of three-stranded structure (see for example West *et al.*[12] and Radding *et al.*[18]) and also in relation to models for transcription.

Watson-Crick pairing is clearly a more strongly specific interaction than the additional interaction that we have used to form base tetrads from Watson-Crick base pairs (see the discussion in McGavin[3]). *While it is perhaps difficult to envisage the direct specific formation of 2c from separate single strands, it is much less difficult to envisage its specific formation from a Watson-Crick duplex in which complementary strands are already exactly in register.*

Although only one H bond or possibly one bifurcated H bond is involved in 2c, there is base stacking, there can be a good contact between the sugar phosphate chains and the structure might be stabilized by ions or by basic protein or other materials.

2c can be optimized as a duplex in its own right rather than as a component of the four-strand model. Models in fact were built earlier by one of us in which the base pairs of this duplex were placed approximately at the radius of those of B form DNA (McGavin, unpublished).

It is of some interest to compare 2c with the model that Hopkins[20] describes as "configuration 2" and that involves Watson-Crick base pairing but has the same strand polarities as 2c.

DISCUSSION: MODELS OF THE SECOND CLASS

Tape-like flexibility

The tape-like flexibility (shown in 4a to 4h) that the extended models have would be useful in the processing of information. It would allow essentially side-by-side specific pairing of both complementary and identical structures. Note that a model with 0 degrees rotation between layers can be built with little difficulty.

Neighbor exclusion

The flexibility of the models shown in Color Plate 4 leads to a possible explanation of neighbor exclusion in the intercalation of duplex DNA by small planar molecules (see, for example, Lerman,[27] Crothers,[28] Sobell *et al.*[29] and Neidle[30] for discussion of this). Neighbor exclusion is sometimes explained by the idea that certain features of a neighbor-excluded structure are positively favorable (e.g., an alternation of sugar puckers). The work described above, however, indicates that there is enough room for the intercalation of small molecules between every pair of base pairs along a duplex. We suggest that the flexibility (lack of rigidity) of such structures would make the retention of small molecules in every slot unlikely in a linear molecule, although this would be stereochemically possible.

Eight-stranded model

While the eight stranded model (5a) has both a specificity of complementarity and of identity, there is no obvious specificity between the mutually intercalated components of the kind shown in Color Plate 4b. These could be identical, but the structure at the same time provides a means of carrying two completely different sequences together intimately without any covalent connection between them. The same indeed is true of the four-strand components 5b and 5c themselves.

Note that this eight-strand model, like 2a, has twice the mass per unit length of a normal Watson-Crick duplex and that the four-strand components (5b and 5c) have the normal duplex mass per unit length.

Parts of the eight-strand model

This model can be broken down into parts in several ways. These parts are not discussed in detail here. 5b and 5c, however, show two possible dyad-related components. These component structures are of the kind first mentioned by McGavin, Wilson and Barr.[24]

Consideration of the surface of such models led one of us (McGavin[1]) earlier to consider the pairing of base pairs about dyad axes to form regular helical structures as in 2a. The base pairings themselves had in fact been considered earlier by Kubitschek and Henderson[25] and by Löwdin.[26] This diverted interest at that time to the use of such pairing in the construction of the basic four-strand model (2a).

The pairing within the eight-strand models of 5b and 5c involves this same dyad relationship of base pairs. It also involves an interaction of a meshing kind.

As mentioned above, several other structures can be produced by breaking down the eight-strand model in different ways. Thus structures with very highly specific surfaces can be formed by splitting 5a along the Watson-Crick bonding system, as in the formation of 2c from 2a.

Other structures result if the base tetrads within 5a are rotated relative to each other, as in the formation of 3a and 3b above, to form syn as well as anti base orientations. Clearly a number of structures, some of considerable complexity but which nevertheless remain regular and compact, can be formed by manipulating the core within the eight-stranded model and by subsequent subdivision.

Energy minimization

This paper was written to describe a computer modeling study on a range of related DNA models using the program FRODO and along with this, for refinement (regularization) of the models, the programs NUCLIN and NUCLSQ.

Interest, however, has been expressed in the possibility of extending this work to energy minimization and molecular dynamics. We therefore add this short section on preliminary work using the molecular simulation package GROMOS (see References 33 and 34). The effect of the

energy minimization is to release possible local strain by the conjugate gradient method.

In most of the work described above, models with five layers of bases were used. The results of regularization were then examined to see if at least the central three layers had remained essentially equivalent, that is, consistent with an extended helical structure. A continuous regular helical structure could then be constructed from the central layer if this was required. The energy minimization work, so far, is limited to fragments involving five layers.

Energy minimization results

The program GROMOS outputs energy figures (kJ mol^{-1}) for the starting structure and for the structures that result from increasing numbers of cycles of the energy minimization process. The figures of most interest are those for total energy, energy associated with bond angles, electrostatic energy, Lennard Jones (van der Waals) energy and energy associated with dihedral (torsion) angles.

Of particular interest are figures for the initial structure, that is, for the structure that had already been subjected to stereochemically restrained least-squares refinement (regularization) using NUCLIN and NUCLSQ and also for the structure following that number of GROMOS cycles required to produce actual energy minimization. In all cases we were able to achieve such minimization, with some, usually small, reduction in total energy.

No material changes in the models were suggested by this preliminary work. In view though of the fact that the models are closely related and that comparisons between models and with the well-established Watson-Crick duplex could be productive, we believe that further detailed energy minimization work would be worthwhile.

The basic four-strand model and its parts (2a, 2b and 2c)

The total energy of the four-strand model itself is slightly lower than twice that of either of its duplex components taken separately (that is, than two components of the Watson-Crick kind or two components of the "alternative" kind (2c)). The four-strand structure involves more interactions of the van der Waals kind, and also of the electrostatic kind, than the sum of two components of either kind taken separately.

The energies of the two kinds of duplex component (2b and 2c) and of a B form Watson-Crick duplex that was included for comparison are similar to each other. It should, though, be remembered that here we are studying fragments and that GROMOS does not identify hydrogen bonds as such and deals with all electrostatic interactions in the same way.

The series of extended duplexes

This series was built (see above) to confirm the idea based on simple (noncomputer) model building that models of the second class that involve extended chains and systematic mutual intercalation of duplex components would be flexible in the sense that the rotation angle between successive layers of the structure could be varied easily and continuously within limits. The fact that the energy values obtained for all the structures in the series are similar bears out this argument.

The limits, however, are not shown clearly by these preliminary energy minimization figures because of accommodation of strain at the ends of the fragments. The earlier study based on the use of NUCLIN and NUCLSQ involved examination of the output given by the program for the central layer of bases where that was essentially consistent with an extended helix.

A molecular dynamics simulation (MD) (for a review, see Reference 35) on the zero structure of this series resulted in a collapsed molecular conformation within a few picoseconds, which is what would be expected from the open structure of the initial model.

ACKNOWLEDGEMENTS

S. McG. wishes to thank the staff at EMBL (Heidelberg) for help and hospitality during several visits. Particular thanks are due to D. Banner and D. A. Marvin. Thanks are also due to E. Westhof (Strasbourg).

S. McG. also wishes to thank the European Molecular Biology Organization for the award of a short-term fellowship which supported part of this work, and also SERC for a travel grant in support of the more recent work described here.

REFERENCES

1 McGavin, S. *J. Mol. Biol.* 1971, **55**, 293–298
2 McGavin, S. *Proc. First European Biophysics Congress* 1971, 259–262
3 McGavin, S. *J. Theor. Biol.* 1979, **77**, 83–99.
4 McGavin, S. *Heredity* 1977, **39**, 15–25
5 McGavin, S. *Nature* 1973, **242**, 330
6 McGavin, S. *J. Theor. Biol.* 1980, **85**, 665–672
7 McGavin, S. *J. Theor. Biol.* 1981, **91**, 33–40
8 McGavin, S. *J. Theor. Biol.* 1984, **107**, 37–56
9 Nash, H.A., Mizuuchi, K., Enquist, L.W. and Weisberg, R.A. *Cold Spring Harbor Symp.*, 1980, **45**, 417–428.
10 Kikuchi, Y. and Nash, H.A. 1980, *Proc. Natn. Acad. Sci. U.S.A.* 1979, **76**, 3760–3764
11 Nash, H.A. and Pollock, T.J. *J. Mol. Biol.* 1983, **170**, 19–83
12 West, S.C., Cassuto, E. and Howard-Flanders, P. *Proc. Natn. Acad. Sci. U.S.A.* 1981, **78**, 6149–6153
13 West, S.C., Cassuto, E. and Howard-Flanders, P. *Nature* 1981, **290**, 29–33
14 Gierer, A. *Nature* 1966, **212**, 1480–1481
15 Lilley, D.M.J. *Biochem. Soc. Trans.* 1984, **12**, 127–140
16 McGavin, S. *J. Molecular Graphics* 1985, **3**, 93–100
17 Jones, T.A. *J. Appl. Cryst.* 1978, **11**, 268–272
18 Das Gupta, C., Shibata, T., Cunningham, R.P. and Radding, C.M. *Cell* 1980, **22**, 437–446
19 McGavin, S. *J. Theor. Biol.* 1989, **136**, 135–150
20 Hopkins, R.C. *Comments on Molecular and Cellular Biophysics* 1984, **2A**, 153–178
21 Arnott, S., Chandrasekaran, R. and Martilla, C.M. *Biochem. J.* 1974, **141**, 537–543

22 O'Brien, E.J. *Acta. Cryst.* 1967, **23**, 92–106
23 Wang, A.H.J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. and Rich, A. *Nature* 1979, **282**, 680–686
24 McGavin, S., Wilson, H.R. and Barr, G.C. *J. Mol. Biol.* 1966, **22**, 187–191
25 Kubitschek, H.E. and Henderson, T.R. *Proc. Natn. Acad. Sci. U.S.A.* 1966, **55**, 512–519
26 Löwdin, P. In *Electronic Aspects of Biochemistry* (B. Pullman, ed.). New York, Academic Press, 1964, 167
27 Lerman, L.S. *J. Mol. Biol.* 1961, **3**, 18–30
28 Crothers, D.M. *J. Mol. Biol.* 1968, **55**, 293–298
29 Sobell, H.M., Tsai, C.C., Gilbert, S.G., Jain, S.G., and Sakore, T.D. *Proc. Natn. Acad. Sci. U.S.A.* 1976, **73**, 3068–3072
30 Neidle, S. *Progress in Medicinal Chemistry* 1979, **16**, 151–218
31 Griffith, J.D. and Nash, H.A. *Proc. Natl. Acad. Sci. U.S.A.* 1985, **82**, 3124–3128
32 Sen, D. and Gilbert, W. *Nature* 1988, **334**, 364–366
33 Groningen Molecular Simulation Package. BIOMOS. Lab. of Phys. Chem. Univ. of Groningen, the Netherlands
34 van Gunsteren, W.F. *Protein Engineering* 1988, **2**, 5–13
35 McCammon, J.A. and Harvey, S.C. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, 1987
36 Hendrickson, W.A. and Konnert, J.H. In *Biomolecular Structure, Function, Conformation and Evolution* (R. Srinivasan, ed.). Vol. 1, p43–50, Pergamon Press, Oxford, 1980
37 Konnert, J.H. and Hendrickson, W.A. *Acta Cryst.* 1980, **A36**, 344–349
38 Westhof, E., Dumas, P. and Moras, D. *J. Mol. Biol.* 1985, **184**, 119–145