Interactive design and manipulation of macromolecular architecture utilizing nucleic acid junctions

Nadrian C Seeman

Department of Biological Sciences, State University of New York at Albany, Albany, NY 12222, USA

A program is introduced for the design of macromolecular structures using branched nucleic acid junctions as building blocks. This can result in unexpected features in the designed molecules, several examples of which are illustrated. The program allows interactive construction of models constrained by double-helical connecting rods between vertices. The logical structure of the design process is based on the single helical segment, rather than on the junction, as the fundamental unit. The file structure permits storage of both molecules and instruction protocols.

Keywords: macromolecular architecture, nucleic acid junctions, sticky-ended ligation

received 25 March 1985, accepted 4 April 1985

Biotechnologists have exploited the high specificity of sticky-ended nucleic acid ligation for over a decade[1]; linear unbranched constructions (particularly plasmid circles) have been used in, for instance, molecular cloning and gene expression. Branched or junctional nucleic acids, 'junctions', appear to be attractive for use in design of 3D macromolecular structures. Nucleic acid junctions can be constructed from oligonucleotides[2,3] and are described as 'immobile' when all opportunities for branch point migration have been eliminated. An algorithm can be used which forms, from starting molecules, stable complexes with selected sequences that have minimal sequence symmetry[2,3], and have alternative pairing arrangements thermodynamically considered by the algorithm[4,5].

An example of an immobile nucleic acid junction is shown in Figure 1. The concept of nucleic acid junctions has already received attention[7,8], and junction complexes have been characterized by physical and electrophoretic techniques[9,10]. These studies have yielded structural and physicochemical information about nucleic acid junctions and have shown them to be analogues of important intermediates in the process of recombination[11].

In principle, nucleic acid junctions may be ligated in a similar manner and with the same ease as the long, linear duplexes utilized by biotechnologists[2,3]. Nucleic

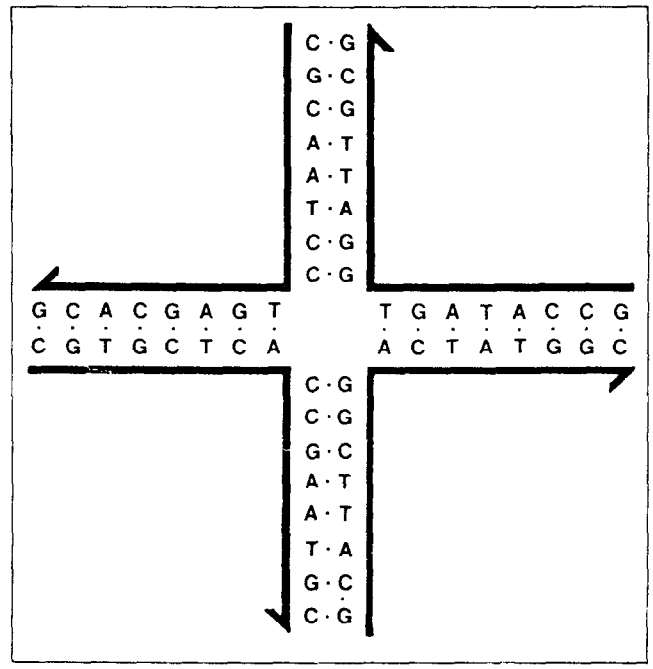

Figure 1. Immobile nucleic acid junction composed of four Hexadecanucleotides. This sequence has been designed by using sequence symmetry minimization rules, supplemented by equilibrium distribution optimization[2–5]. Note the lack of two-fold symmetry around the centre, so that branch point migration[6] is not possible. This sequence also contains no repeating GpG sequence longer than two, in order to minimize G–G non-Watson–Crick base pairing

acid junctions may be thought of as vertices that may be joined together in molecular architecture to construct *N*-connected networks[2,3,12]. Specificity of products is assured by the specificity associated with annealing sticky ends that allows control of the ligation process. The number of arms emanating from an immobile nucleic acid junction can range from three to eight[2,3], the appropriate values deciding that of *N*, the number of connections. Because junctions are composed of oligonucleotides, the interjunctional spacings are significantly below the persistence length of DNA. This suggests that, to a first approximation, they may be

 0263-7855/85/020034-06 $03.00

treated as mechanically stiff macromolecular-scale Dreiding[13] or framework[14] molecular models. Ligation experiments involving junctions support the notion that they are indeed stiff[15].

The most important way in which these 'valence clusters' of nucleic acid junctions differ from valence clusters of atoms is the correlation between intervertex distance and torsion angle. The 'bonds' which result from ligating nucleic acid junctions are double helices, and the relative torsional orientation of two ligated clusters is a function of the number of residues which separate them[12]. It has been reasonably simple to design structures expected to be planar [15]:this has been readily accomplished by separating junctions by integral or half-integral multiples of the nucleic acid periodicity. However, the generation of more complex, 3D structures, including periodic networks[2,3,12,16], is not amenable to such a simple rule. One must be able to estimate the effects of altering the lengths of the double helices, the junction 'valence angles', and the changes in twist which result from traversing a junction by different pathways. For this purpose, the author has written a program in FORTRAN which allows the user to perform interactive molecular model building with nucleic acid junction as components.

PROGRAM DESIGN

The basic steps in designing macromolecular architecture[12] are outlined in Figure 2. The first step is the selection of the structure to be built, the second deals with the adaptation of geometry to the chosen design, and the third is the selection of the sequences which will maximize the probability of achieving the desired architecture. The selection of sequences has been treated elsewhere[2–5], as have the geometric details of the middle step[12]. This program aids the designer in both the first and second steps.

It is possible to design the 3-connected quadrilateral shown in Figure 2 in a number of different ways: the case shown results in five strands after ligation, but a 4-stranded complex may also be devised[12]. Thus, while nucleic acid junctions can be readily treated as valence clusters[12], this is not particularly useful in a program for designing their concatenates. Here, the fundamental unit is the single helical strand, which the program allows to be extended, paired with, or bent.

It should be noted that little is known about details of geometry and flexibility of the junction at present. Thus it should not be treated as though its angles were fixed, although this may be an effective approach in the future, in the light of crystallographic determinations of the structure of junctions. Similarly, it is not yet worthwhile to assign strain estimates to deformations involving junction geometry. Studies currently in progress may allow us to do this in the near future; and transform the purely geometrical approach taken here into a physical system based on favourable structures and the potential surfaces which maintain them.

The system used in the program treats molecules differently to an ordinary molecular model building proram and works at a very low 'working resolution'. Its subjects are such large molecules that the normal requirement in programs for display of molecular structures — that every atom or perhaps most atoms in a virtual bond formulation should be displayed — is not

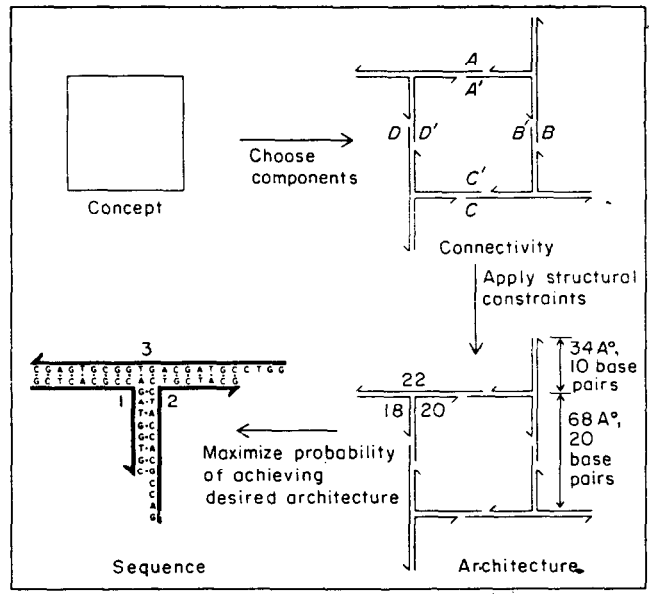

Figure 2. Basic architecture program for macromolecular design. This illustration shows the basic steps involved in design of macromolecular structures. The example is taken from a closed square which has been experimentally constructed[15]. The first step asks the question: what should be made? Having decided on a square, the second step is the choice of components, exclusively 3-arm junctions in this case. One must also choose the sticky-ends, and where they are going to go. In the case shown, all the sticky ends are unique. However, it is possible to do the reaction more cheaply, if one can tolerate a mixture of products, by letting A = B = C = D, and A′ = B′ = C′ = D′. The fundamental topology (eg four final strands, or five final strands, as shown) is also selected at this point. These decisions yield the second stage, connectivity. In the third step, the structural constraints are applied to the connectivity, to give the third stage, architecture. This results in the edges being 20 nucleotide pairs long, the exocyclic arms ten base pairs long, and the sticky-ends four base pairs long. Thus, asssuming a ten-fold helix repeat, the side of the square will be about 68 Å, with 34 Å exocyclic double helices attached for stability. The lengths of the strands which correspond to this architecture are 18, 20 and 22 nucleotides long, as indicated. Finally, one must choose sequences which will maximize the probability of achieving this architecture, through sequence symmetry minimization supplemented by thermodynamic calculations[2–5]. The sequence shown is for the case A = B = C = D[15]

needed, and only one or two atoms per residue are more than sufficient to give the user adequate information to make choices and understand trends. More input information causes confusion rather than insight. On the other hand, the extension of old strands and the generation of new strands requires knowledge of the helical axis of every segment, and the orientation of every residue. Thus, a consistent coordinate system is maintained on each residue.

The phosphorus atom is treated as the only atom whose coordinates are retained for each nucleotide, and is used as the origin of an orthogonal coordinate system, the basis vectors of which point: towards the helical axis, parallel to the helical axis and tangentially to a circle intersecting the helical axis in the plane of the

phosphorus atom. The lengths of these three vectors are, respectively, the helical radius of the phosphorus atom, the helical repeat distance and again, the helical radius of the phosphorus atom. These convenient lengths give the programmer and user the position of the helical axis and the height within the helix of the next residue. These four sets of coordinates constitute a virtual *atomic cluster* which is useful for both internal and user oriented applications.

PROGRAM FEATURES

The program was written for the Sperry 1100/82 mainframe computer, with Tektronix software, but is designed to be adaptable to any other mainframe computer. Available to the user is a menu of 55 commands through which four different types of function may be implemented:

- molecular construction and deformation
- geometrical measurements
- file handling
- display functions.

These are described briefly below.

Molecular construction and deformation

A new helical segment may be created arbitrarily at either end of an existing helical segment, or opposite an existing segment. One group of segments may be fused to another group of segments, preserving intergroup geometry, by utilizing the Nyburg[17] algorithm. For example, one junction can be fused to another, by superimposing any two atomic clusters which they contain. The program allows clusters to be overlapped at their ends or in the middle of strands, thereby allowing different separations to be tested with the same building blocks. The lengths of segments can be altered, and segments excised. Groups of segments can be translated at will, and helices rotated about their axes, with other segments following the rotation. A physical segment may correspond to several logical segments which have been joined. Altering the orientation of any logical segment is possible from either the 3′ or 5′ end, thus created the bends necessary for junction formation. The parameters of individual phosphorus atoms (and their associated vectors) are conveniently changed. Helical parameters are also available for modification, so that the user may select from B-DNA, A-DNA, Z-DNA, RNA-11 or other helices.

Geometrical measurements

Examples of quantitative data associated with the molecule, created by the program, include the coordinates of any phosphorus atom or atom cluster, distances between atoms, bond angles and torsion angles. These descriptive parameters may be defined at the level of individual atoms or helix segments. A particularly useful measurement is the helical relationship between the atomic clusters representing any two residues[18]. This tells the user how much the structure has to be deformed in order to achieve ring closure or spatial periodicity[11].

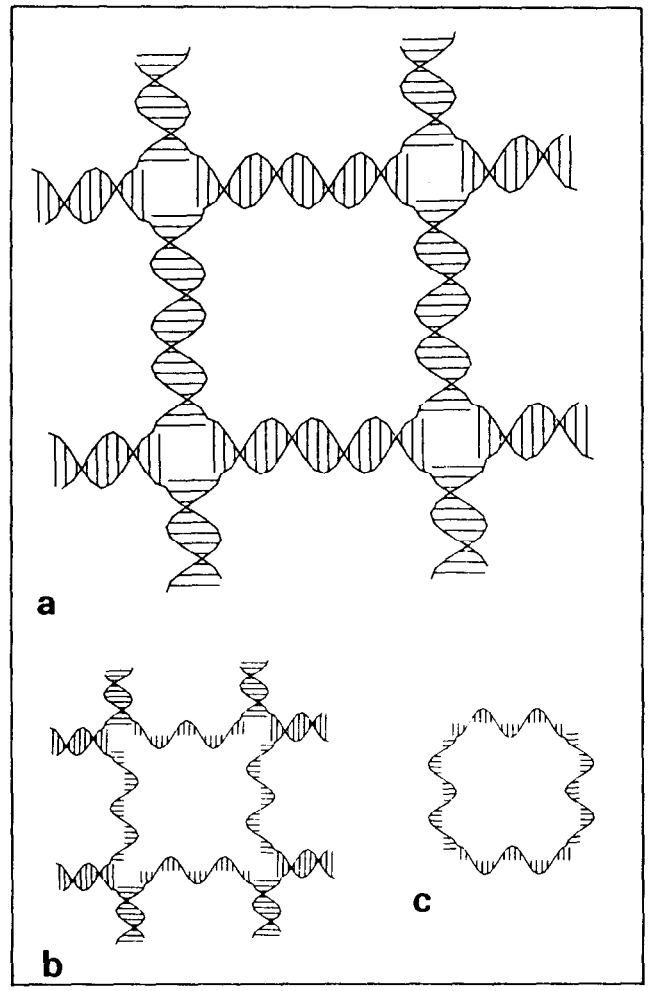

Figure 3. Square composed of four planar 4-arm junctions with an integral edge length. The junctions used in this example are assumed to be planar, and the interjunction separation is 20 base pairs, exactly two turns of 10-fold DNA. (a) The complete square. (b) The linear strands which contribute to this structure. (c) The closed circular component of this structure. Whatever the 'valence angles' at the junction corners, the overall double helical structure is expected to be essentially flat

File Handling

The program contains two sets of files, the molecule file and the macrofile. In the former, the user can keep labelled molecular constructions and avoid *de novo* rebuilding of them when they are needed; thus, a given fragment may be retrieved and fused with a growing structure. The program also contains a macro capability: instructions may be stored in the macrofile and reused when desired. This allows the user immediate reconstruction of large concatemers, and permits the use of different molecules in the same scenario and saves on storage space. Both sets of files are in card-image format, so that modifications can be made and errors can be removed using the system text editor.

Display functions

The program has been written for a 2D display system. Rotations are controlled by incorporated Software

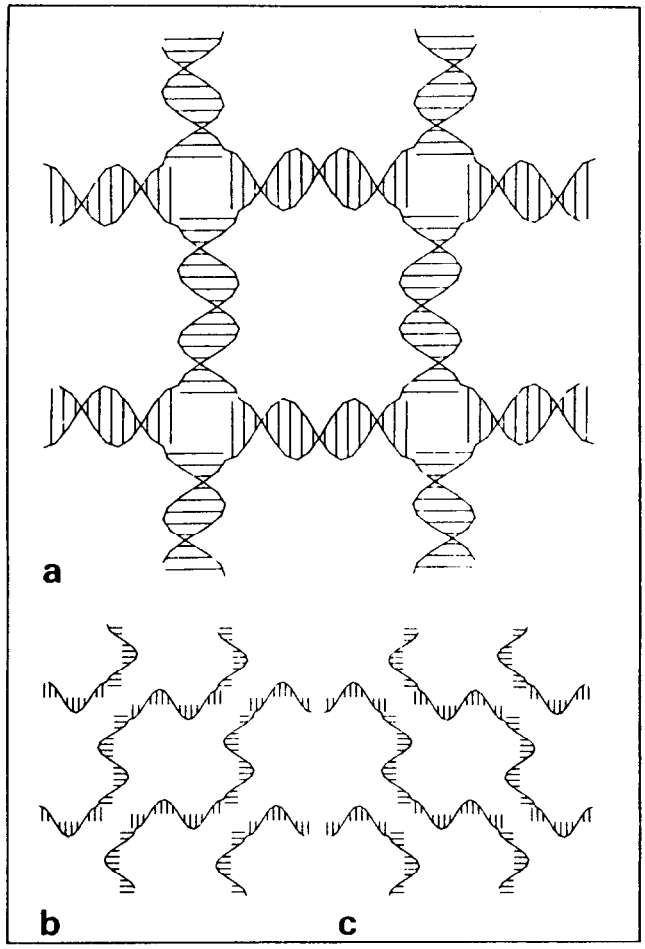

Figure 4. Square composed of four planar 4-arm junctions with a half-integral edge length. This square is identical to that in Figure 3, except that the edges are a turn and a half of DNA. (a) The complete square. (b) One half of the linear strands which form the square. (c) The other half of the linear strands which are components of the square. There are no closed circles in this structure. Rather, the two sets of strands form an interweaving meshwork

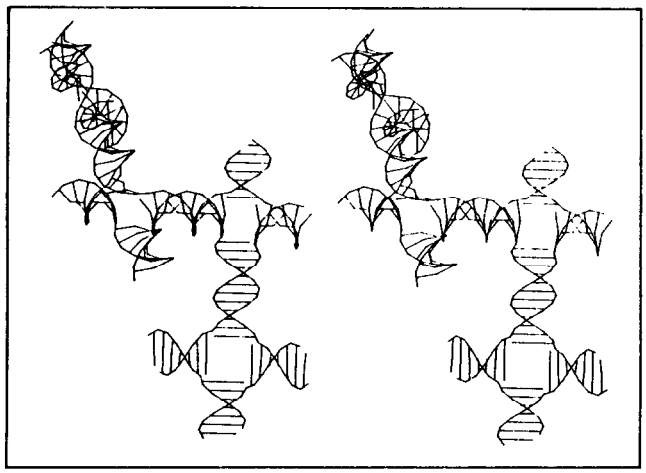

Figure 5. Structure formed from four planar junctions separated by base pairs. These planar junctions have the same structure, but shorter arms, than those in Figures 3 and 4. The molecule is drawn in stereoscopic projection to highlight the three-dimensionality of the structure

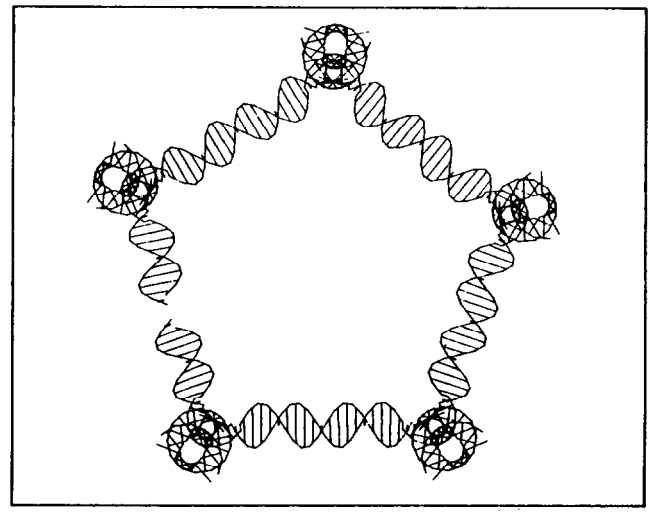

Figure 6. Pentagonal figure formed when five tetrahedral junctions are linked 1–3. The separation of the junctions is 20 base pairs, exactly two turns. The lack of closure is the result of the difference between a tetrahedral angle and a pentagonal angle. The two arms of each junction which do not participate in forming the pentagon are seen obliquely at each vertex

which also instructs for other display functions. Stereoscopic viewing is available, and segments may be masked (not drawn) by choice. Normally, only the phosphorus atoms are connected, but it is possible to display a line representing the bases (see Figures 3–9). The coordinate system centred on each phosphorus atom may also be displayed optionally. Atoms may be individually highlighted. It is also possible to look down the axis of the helix through which any two phosphorus atom clusters are related[18], thereby obtaining accurate symmetrical views, which are otherwise difficult to achieve.

APPLICATIONS

Design of macromolecular architecture nucleic acid junctions is still in its earliest stages. Some of the basic concepts are unintuitive; a few will be illustrated here, with a pictorial clarification provided by the program. Figure 3 shows a 'square' constructed from four planar 4-arm junctions. Figure 3(a) shows the intact square, while Figures 3(b) and 3(c) show how it can be 'decomposed' into linear strands and a cyclical strand respectively. The separation of junctions is equal to 20 base pairs in this example, which uses a classical ten-fold double helix. In Figure 4(a), a square is constructed from the same four junctions, but separated by 15 base pairs. The decomposition of this system into its constituent strands, all of which are linear, is shown in Figures 4(b) and 4(c). This difference between the squares in Figures 3 and 4 is important, for resistance to exonuclease provides an experimental key to the formation of junction figures[15] — only the square shown in Figure 3 would have such resistance. A choice of a non half-integral junction-to-junction separation results in a 3D structure, as shown stereoscopically in Figure 5, where the separation is eleven base pairs.

Several investigators have found that 4-arm junctions without 4-fold symmetry exhibit important asymmetric

properties[19,20]. This point may be illustrated dramatically, by ligating tetrahedrally shaped junctions together. These junctions have been formed from the component planar junctions of Figures 3 and 4 by bending opposite arms symmetrically out of the plane; no attempt has been made to alter the relative twists of individual arms. When they are linked 1–3, ie through opposite arms, the twist alteration is very small. The nearly perfect closure of a pentagon (to make a macromolecular scale version of cyclopentane) is shown in Figure 6. When they are linked 1–2, i.e. through adjacent arms, the result is the helical structure shown in Figures 7(a) and 7(b). The differences are due to the difference in the intrinsic azimuthal twist associated with traversing the tetrahedral junction by qualitatively different routes.

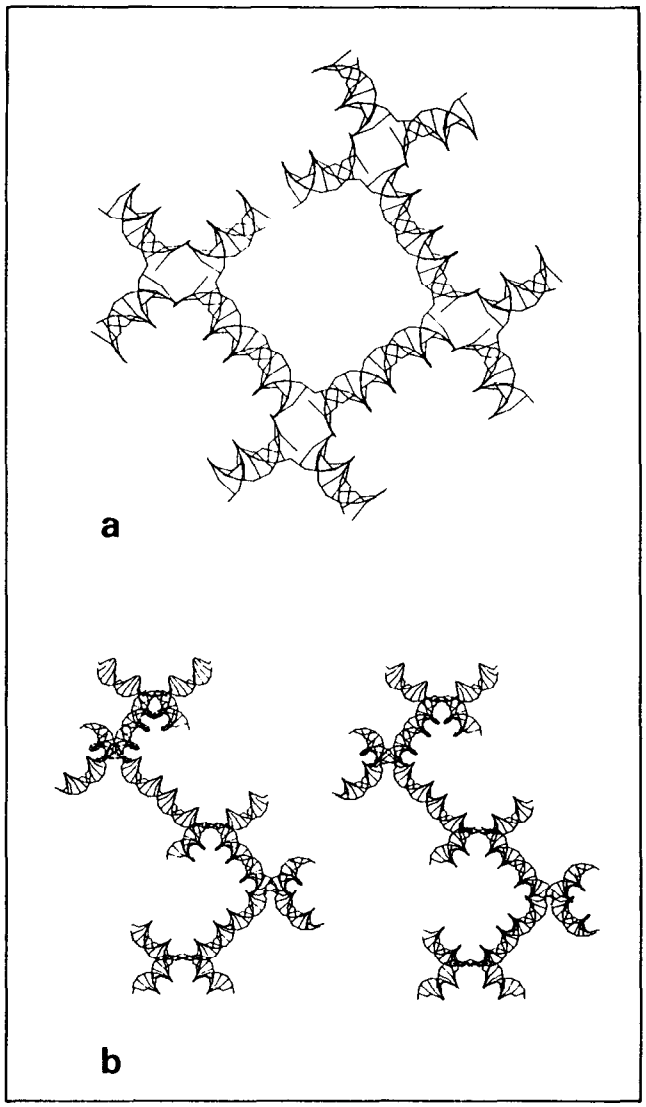

Figure 7. Helical structure formed when tetrahedral junctions are joined 1–2 (a) The view down the axis of the helix . One of the five junctions has been removed for clarity. The helical relationship is about 84° per junction. The torsion angle formed by three arms is about 56°. (b) A stereoscopic projection perpendicular to the axes of the helix. Comparison with Figure 5 demonstrates the great difference between the two kinds of repeated linkage. The views in both Figure 5 and Figure 6(a) were obtained with the same command, requesting a view down the axis of the helix that relates the first residue of two different junctions

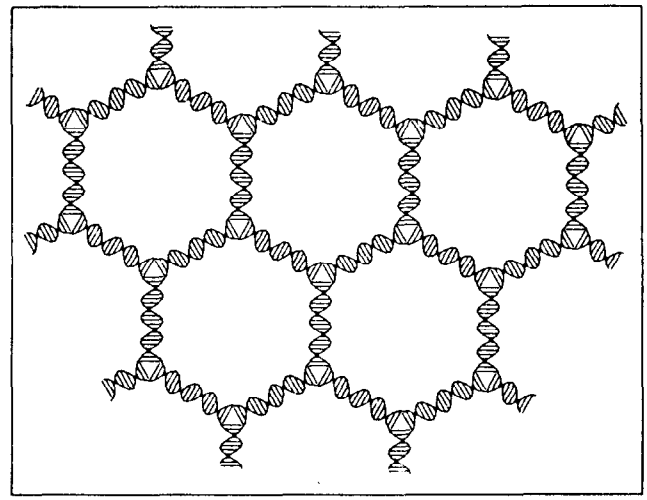

Figure 8. Planar repeating lattice composed of 3-arm junctions. The separation of the junctions is exactly two turns. While this lattice is easy to construct by graphics, preliminary experimental evidence suggests that the trigonal cluster is probably not the correct structure for the 3-arm junction[15]. Thus, one must beware of the pitfalls which can occur when geometry is the only criterion

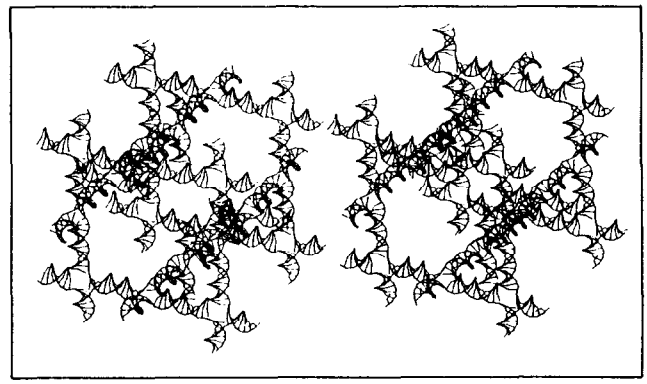

Figure 9. Portion of a repeating 3D, 4-connected network. The helix used for this particular construction is 10.4-fold helical. Junctions are separated by 13 base pairs, resulting in an effective 90° torsion angle per edge. This structure is not the complete repeat unit, but it illustrates the basic features of such a network composed of junctions. The 3D character of the construction is highlighted by the stereoscopic projection

DISCUSSION

Described here is an interactive graphics system which facilitates the design of macromolecular architecture involving nucleic acid junctions. The program utilizes only geometrical considerations, because of the paucity of physical data relating to the structures of nucleic acid junctions. Nevertheless, geometry provides many important features involved in the design of nucleic-acid-based macromolecular architecture. One of the most important features of the nucleic acid junction as a building block is the enormous specificity of interac-

tion, intrinsic in technology of sticky-ended ligation. Thus, specific cross-links, rather than random polymerization can be engineered into the final product. The use of sticky-ended ligation may also be able to enforce a desired geometry on certain junctions[12].

It is to be hoped that the program, and its more physically based successors, will eventually result in the successful design of highly specific 3D closed figures and *N*-connected networks based on nucleic acids. An example of a 2D lattice is shown in Figure 8. A portion of a 3D, 4-connected network is illustrated stereoscopically in Figure 9. The right angles were achieved by use of DNA with 10.4 residues per turn[21], and separations of 13 residues per edge. Such lattices will be useful tools for preparing diffracting material in molecular biology[2,3], and may find applications in areas of biotechnology as well.

ACKNOWLEDGEMENTS

This work has been supported by Grants GM-29554 and ES-00117 from the National Institutes of Health. The author would like to thank the computer centre of the State University of New York at Albany for the use of its Sperry 1100/82 computer. The author is the recipient of a Research Career Development Award from the National Institute of Health USPHS.

REFERENCES

1 **Cohen, S N, Chang, A C Y, Boyer H W and Helling, R B** *Proc. Natl. Acad. Sci. USA* Vol 70 (1973) p 3240
2 **Seeman, N C, in Sarma, R H (ed)** *Biomolecular Stereodynamics* Adenine Press, New York Vol 1 (1981) p 269
3 **Seeman, N C,** *J. Theor. Biol.* Vol 99 (1982) p 237
4 **Seeman, N C and Kallenbach, N R in Pullman B and Jortner J (eds)** *Nucleic Acids: The Vectors* D Reidel, Dordrecht (1983) p 183
5 **Seeman, N C and Kallenbach, N R** *Biophys. J.* Vol 44 (1983) p 201
6 **Thompson, B J, Camien, M N and Warner, R C** *Proc. Natl. Acad. Sci. USA* Vol 73 (1976) p 2299
7 **Kallenbach, N R, Ma, R-I and Seeman, N C** *Nature* Vol 305 (1983) p 829
8 **Kallenbach, N R, Ma, R-I, Wand, A J, Veeneman, G H and van Boom, J H** *J. Biomol. Str. Dyn.* Vol 1 (1983) p 159
9 **Seeman, N C, Maestre, M F, Ma, R-I, and Kallenbach, N R** in **Rein R (ed)** *The Molecular Basis of Cancer* Alan Liss, New York (1985) p 99
10 **Wemmer, D E, Wand, A J, Seeman, N C and Kallenbach, N R** (1985) submitted for publication
11 **Dressler, D and Potter, H** *Ann. Rev. Biochem.* Vol 51 (1982) p 727
12 **Seeman, N C** *J. Biomol. Struct. Dyn.* Vol 3 (1985) in press
13 **Dreiding, A S** *Helv. Chim. Acta* Vol 42 (1959) p 1339
14 **Brumlik, G C, Barrett, E J and Baumgarten, R L** *J. Chem. Ed.* Vol 41 (1964) p 221
15 **Seeman, N C, Ma, R-I and Kallenbach, N R** (1985) submitted for publication
16 **Wells, A F** *Three Dimensional Networks and Polyhedra* John Wiley and Sons, New York (1977)
17 **Nyburg, S C** *Acta Crystallogr., Sect. B* Vol 30 (1974) p 251. The routines used in the program are adapted from a program written by S C Nyburg
18 **Rosenberg, J M, Seeman, N C, Day, R O and Rich, A** Vol 69 (1976) p 979. The routines used in the program are adapted from a program written by J M Rosenberg
19 **Wilson, J H** *Proc. Natl. Acad. Sci. USA* Vol 76 (1979) p 3641
20 **Seeman, N C and Robinson, B H in Sarma R H (ed)** *Biomolecular Stereodynamics* Adenine Press, New York, Vol 1 (1981) p 279
21 **Wang, J C** *Proc. Natl. Acad. Sci. USA* Vol 76 (1979) p 200