



## INTERACTIVE COMPUTER GRAPHICS AND MACROMOLECULAR STRUCTURES

Peter J. Bond  
Department of Biochemical Sciences  
Princeton University  
Princeton, N. J. 08540

In 1964 on Project MAC at MIT, Levinthal and Langridge first applied interactive computer graphics to molecular model building (1, 2). Since then a number of laboratories have developed computer graphics systems to suit their particular needs (3, 4, 5, 6).

I shall begin by describing the system that we have developed in the Princeton University Computer Graphics Laboratory during the past year to investigate macromolecular structures.

The hardware consists of a Digital Equipment Corporation PDP-10 computer with 64K, 36-bit words of 1 microsecond memory, which is interfaced with an Evans and Sutherland Line Drawing System 1. The LDS-1 computer directly accesses the PDP-10 memory. During execution the two central processing units fetch instructions independently. Each computer can alter its own or the other's program and each is accessible by the programmer. Interaction with the display picture is via input devices attached to the LDS-1. These include switches, knobs, buttons, an acoustic tablet, and a joystick. Black and white and color stereoscopic viewing is achieved using a rotating sector device, the so-called "lorgnette". The LDS-1 synchronizes two offset images on the screen with the lorgnette which alternately masks each eye. Real time rotations, translations, and scaling of complex line drawings are performed by a hardware matrix multiplier and lines not in the viewing portion of the display are removed by a hardware clipping divider. The basic facilities that should be provided by a computer graphics system for this kind of work have been listed by Levinthal, *et al.* (3) and Barry, *et al.* (4). These include: (a) the building of molecular models from their component parts in such a way as to preserve their bonded stereochemistry; (b) the definition of a "connectivity-tree" that indicates which atom, or fixed unit of structure, is connected to which. Connectivity in our programs is expressed both explicitly as tables of atom labels and implicitly in ordered lists of coordinates. This should be able to handle branched and ring structures; (c) the translation and rotation of models. Translations are needed to center the image on the screen or to superimpose it on another molecule or an electron density map. Rotation of the whole molecule is required about specified axes, whose origin may be the model's center of gravity or a particular atom in the structure; (d) the simulation of freedom of rotation about single bonds by rotating one part of the model with respect to another, about a specified bond; (e) the economic use of screen size by effective scaling. Also, the scaling from one display to another; (f) the retention of atomic coordinates after performing the required interaction, so that they may be easily output; (g) the separation of the display into two or more fragments, so that each may be fixed or interacted with independently; (h) the indication by dotted and/or colored lines and the listing of pairs of non-bonded atoms whose separation comes within specified ranges that take into account their van der Waals radii; (i) finally, the coupling of the display to routines that calculate functions which depend on the atomic coordinates.

The particular hardware that we access to provide these basic facilities includes the following: (1) sixteen switches with associated lights, each of which controls a "flip-flop" circuit whose "on" or "off" state can be used to signal particular display features or certain algorithms; (2) four knobs which control potentiometers and provide analogue control for such things as rotation rates about axes; (3) a joystick that has two degrees of freedom and supplies two continuous variables for each position. This also has a button "flip-flop" control, and lastly, (4) a three-dimensional acoustic tablet. This takes the form of three mutually perpendicular strip microphones that record the regular sound emitted by a pen with a 3 Kilovolt spark gap. Clocks time the sound and the XYZ distances corresponding to the current pen position are then determined. The addition of the Z dimension has been described by Wipke and Whetstone (7).

Packages of FORTRAN callable display subroutines have been designed for user convenience. They enable programs written in higher level languages to draw displays using only a small number of CALL statements. One of the packages we use was written by W. T. Wipke and T. Dyott and another by A. M. Lesk. If these are used, then approximately 40K of computer memory remains for the other programs. Further storage areas include a 5.2 million word disc, a 9-track tape transport providing for 2.5 million words, and two DEC tape transports, each providing for 0.07 million words.

### Molecular Models in the Study of Macromolecular Structures

Since L. Pauling and R. B. Corey first used space filling molecular models, their use and the use of wire skeletal models have become commonplace in the study of molecular structures. In the determination of protein and polynucleotide structures, crystallographic methods are currently unable to reveal the positions of atoms. One reason for this is that most crystals and polynucleotide fibers, when exposed to X-rays,

diffract only to a resolution of 2 to 3 Å. The inherent disorder in the crystals means that only groups of atoms can be resolved in the maps of electron density that are computed from the spot intensities. A set of stereochemical rules has been established from the study of small related molecules. These small molecules are the amino acid components for protein structures and the bases, nucleosides, and nucleotides for nucleic acid structures. The rules of stereochemistry include the invariance of covalent bond lengths and interbond angles, the planarity of the nucleic acid base atoms, and of the amide group in polypeptide chains and other amino acid side chain groups. Further, the ribose and deoxyribose sugar groups adopt preferred ring pucker and the conformations of dihedral angles about the single bonds in polynucleotide units are observed to be confined to narrow angular ranges. A final geometric constraint is that non-bonded interatomic distances are rarely observed less than the sum of the atomic van der Waals radii.

Mechanical molecular model components are designed to incorporate some or all of these constraints. However, they are rigidly built in and it is not possible to relax the constraints in a reasonable way. The strain in large wire models and the effect of gravity often introduce distortion. The physical support of these mazes of wire rods poses a considerable problem. Richards (8) devised a system of wires held under tension for the attachment of the model. This is currently the most common way of building protein models. Such models usually have been obtained by threading the wire backbone through the three-dimensional electron density maps determined from the X-ray data. A device called a "Richards' Box" (8) is based on a half-silvered mirror which permits sections of electron density maps, plotted on a transparent material, to be seen superimposed on the wire model. A selective system of lights illuminates those parts of the model corresponding to the map sections currently in position. Analysis of helical polynucleotides by the trial-and-error Fourier transform method, involves the building of trial models for which a Fourier transform is calculated (the Fourier transform squared is related to the observed intensity). The construction of regularly repeating wire subunits around a central helix axis suffers from the same imprecision as the protein models. The atomic coordinates measured from wire skeletal models are subject, therefore, to errors resulting from the measuring process and from distortions of the components. The use of computer programs, such as those described by Langridge and MacEwan (2), Diamond (9), and Arnott, *et al.* (10), ensure that those constraints built into the physical model components are imposed on the atomic coordinates.

Clearly, many of the difficulties associated with mechanical models can be overcome if computers with visual display units, which give realistic three-dimensional representations, are used. An illusion of three dimensions can be achieved by continuously rotating or oscillating the image on the screen. We find this adequate for demonstration purposes (such as in the making of movies for teaching purposes), for the development of graphics programs and for the study of the very large structures. However, when actively building models to a precisely constrained three-dimensional geometry, we need to use the lorgnette. Not least of the advantages of such systems is the speed of construction, and that the molecular geometry, once specified, is retained as a set of atomic coordinates stored in the computer.

#### Application of the Princeton System to Macromolecular Models

There are two kinds of interaction that an investigator of macromolecules might require. The first is interaction with molecular and/or crystal structures that have been determined by X-ray diffraction analysis or have been proposed as stereochemically viable models. The second kind of interaction would mimic the action of a manual model-builder.

Implicit in the first kind of interaction is efficient and fast exchange of large data files that represent the results of structural studies. The aim has been to provide programs that give a neophyte maximum interaction with a minimum of concern for the software. First, he is able to select one of the ten protein molecules that we have stored on the disk area. Although time is lost in accessing the disk, it is a necessary sacrifice, since computer core space is limited. Thus, a purely programming problem becomes one of speed as a function of molecule size. A protein with 250 amino acid residues requires about 50 blocks to be read and takes approximately 10 seconds to be displayed. The input/output to the disk is by random access and the data are in binary packed format. Tables of atom labels and connection lists are carried in core and used for drawing and labelling atoms on the display. Real time rotations about and/or translations along the XYZ axes occur with the origin of the coordinate system at the molecule's center of gravity. The molecule can be split into a maximum of nine pre-specified regions. Each region can be manipulated independently for a more detailed analysis, a comparison with another molecular fragment or for a structural modification. Amino acids can be individually labelled and particular residues pointed out using dotted lines which originate from a residue name and number displayed in the corner of the screen. Molecules can be viewed at various levels of complexity; all or part of their polypeptide chain, with or without side chains, or as lines joining the carbon  $\alpha$  atoms of each amino acid residue. A particular atom or residue can be selected and only the neighboring atoms or residues within a specified distance displayed. Questions concerning non-bonded interatomic distances, torsion angle values and strain energies are answered immediately.

Four protein molecules have been chosen to demonstrate the program features just described. We are indebted to Hendrickson and Love (11) for their lamprey hemoglobin coordinates, to Cotton, *et al.* (12) for their staphylococcal nuclease coordinates, to Blow, *et al.* (13) for their chymotrypsin coordinates and to Lipscomb, *et al.* (14) for their carboxypeptidase-A coordinates. Movies have been made of these and other protein molecules for use as teaching aids. Whole molecules, with and without their side chain atoms, are portrayed in three dimensions, some have been broken down to display their more interesting conformational features and then reassembled.

Hemoglobin molecules function by binding molecular oxygen to their heme iron atoms and transporting it to the myoglobin molecules at the muscles, which store it until it is required for metabolic oxidation. A further function is the binding and removal of the carbon dioxide by-product of metabolic oxidation.

Lamprey hemoglobin is a monomeric hemoglobin from sea lamprey. It has been shown by Hendrickson and Love (11) to closely resemble the molecular conformation first found in sperm whale myoglobin (Kendrew, et al., 15) and termed the "myoglobin fold". Figure 1 is a display of the lamprey hemoglobin molecule without its amino acid side chain atoms. The pointer locates the first residue, a proline. The molecule can be seen to possess the eight helical regions (A through H), characteristic of sperm whale myoglobin. Figure 2 displays the sequence of residues leu 118 (located by the pointer) to tyr 145. This includes the GH bend and the H helix, both of which differ from the myoglobin, and the horse hemoglobin chains. A deletion of nine residues has caused the G helix to shorten one turn, the GH corner to assume a novel configuration, and the H helix to be two turns shorter. The carboxyl terminal amino acid, tyrosine (tyr 148) at the end of the H helix, shown in Figure 2 (tyr H22 in myoglobin terminology), is homologous with the penultimate tyrosine in the horse and human hemoglobin chains and with the H22 residue in myoglobin. This tyrosine is expelled from its niche in the FG corner during oxygenation, playing an important role in the proposed mechanism of heme-heme interaction (Perutz, 16).

Staphylococcal nuclease is a globular protein that breaks polynucleotide chains. Cuatrecasas, et al. (17) have shown that esters of thymidine-5'-phosphate or pdTp are hydrolyzed with the cleavage of the 5'-C-O-P bond between the phosphorus and the oxygen. Figure 3 shows the polypeptide backbone atoms of the staphylococcal nuclease molecule. The carbon  $\alpha$  atoms of residues lys 45 and lys 48 have been pointed out and the side chains for residues 45 through 48 have been displayed. Figure 4 shows a close-up view of the loop in the bottom right-hand corner of Figure 3. The dotted lines again locate residues 45 and 48 and the side chains of residues lys 45, his 46, pro 47 (labelled) and lys 48 are seen to extend outward and are relatively exposed to solvent. The main structural features of the molecule are contained in Figures 5a and 5b, where lines have been drawn between the carbon  $\alpha$  atoms of each residue. Figure 5a shows a view of three stretches of helix on the left, the center, and the right of the molecule, a three-tier section of antiparallel pleated sheet forming the upper right side of the molecule, and another section of antiparallel pleated sheet formed by two chains which run vertically up and down the back of the molecule. Figure 5b contains the sequence from lys 49 to the end of the molecule, which is the so-called P3 fragment.

The enzymes of the pancreas,  $\alpha$ -chymotrypsin and carboxypeptidase-A, cleave polypeptide chains. Chymotrypsin cuts in the middle of a polypeptide chain, severing the peptide bond just beyond the carbonyl group of residues with aromatic side chains. Carboxypeptidase-A digests polypeptide chains from the carboxyl terminal end. It removes the last residue most easily when this residue has an aromatic side group. Figure 6a shows the backbone atoms of the chymotrypsin molecule. Residue his 57 has been brightened and can be seen in the top center of the figure. Figure 6b shows the labelled active site residues ser 195, his 57, and asp 102 in the context of a whole molecule. All atoms are labelled for these three residues, while only the carbon  $\alpha$  atoms are displayed for the rest of the molecule. In Figure 6c the three active site residues have been singled out and labelled to demonstrate their relative positions and orientations. Figure 6d illustrates all those residues that have their carbon  $\alpha$  atoms within 10 Å and all other atoms within 7 Å, of the labelled his 57 carbon  $\alpha$  atom. The pointer to the left of his 57 locates the carbon  $\alpha$  of asp 102 and that to the right locates the carbon  $\alpha$  atom of ser 195. Four of the five disulfide bridges of chymotrypsin occur at the same place along the chain as four of the six trypsin bridges. The unique chymotrypsin disulfide bridge, from cys 1 to cys 122, is displayed in Figure 6e. The cystine residues have been labelled and brightened against a background of polypeptide backbone bonds. The suggestion of a twisted sheet in Figure 6a is confirmed when only residues 91-86, 103-108, 55-50, 39-46, and 35-29 are displayed, without their side chain atoms, in Figure 6f. This is to be compared with Figure 7a, where the eight chain twisted  $\beta$  sheet, which forms the structural core of the carboxypeptidase-A molecule, is displayed. The  $\alpha$  to  $\theta$  sheet chains are made up of the sequences 32-37, 46-54, 61-67, 103-111, 190-197, 200-205, 238-243, and 265-271. Figure 7b contains the A to H helices which surround the  $\beta$  sheet. Figure 7c is Figure 7a plus 7b and illustrates the relative positions of the  $\alpha$  helices around the  $\beta$  sheet.

Another class of macromolecule that we are interested in is the nucleic acids. Information a cell uses to build proteins is stored in its nucleus as a coded sequence of purine and pyrimidine bases in a molecule of deoxyribonucleic acid (DNA). Four bases are possible: adenine, thymine (or uracil in ribonucleic acid (RNA)), guanine, and cytosine. Three successive bases along the DNA code for one of the twenty common amino acids. That this information storage device, DNA, is a linear molecule of two intertwined polynucleotide strands was confirmed by the analysis of such X-ray diffraction patterns as shown in Figure 8a. This pattern was obtained by passing X-rays through the highly hydrated crystalline fibers that can be spun from aqueous gels of these polymers. The sample I used here was *Micrococcus lysodeikticus* DNA and it gave the so-called A-DNA conformation. One of the stereochemical subtleties between the A-form and the B-form of DNA is the pucker that the five-membered sugar ring adopts. The orientation of the base to the sugar, defined by the dihedral angle about the glycosidic bond  $\chi$ , is another geometric feature of interest. The so-called anti orientation has  $\chi$  in the angular region  $\sim 90^\circ$  and the so-called syn orientation has  $\chi \sim 270^\circ$ . The more commonly observed anti angle has been postulated to transition to the syn orientation to explain a number of hypothetical molecular mechanisms. Thus, non-standard, "wobble" base pairs have been postulated by Crick to explain the degeneracy of the genetic code. Sobell has proposed an alternative to the Watson-Crick-like wobble base pair of adenine with hypoxanthine. This alternative configuration involves the hydrogen bonding of the imidazole nitrogen of adenine and would require adenine nucleotides on the messenger RNA to be in a syn orientation, if the codon and anticodon phosphate ester chains are to be anti-parallel. Figure 9 shows energy profiles as a function of  $\chi$  for a number of possible sugar puckers. The "pucker" axis in Figure 9 represents a pseudo-rotational coordinate that is related to a particular sugar ring conformation. Each sugar ring pucker can be distinguished as having one or two atoms displaced either on the same or on opposite sides of the plane of the remaining atoms. The computer graphical display of this surface indicates that in this theoretical situation the anti to syn transition could occur if the sugar were first puckered to one of the trough regions.  $\chi$  would then be

free to rotate into the syn orientation where it could be trapped by a further change in the sugar pucker and orientation.

A number of polynucleotide fibers give rise to fiber diagrams which exhibit much continuous diffraction. Figure 8b is one of these, where the few Bragg reflections are confined to the center of the diffractogram. The sample used was that of a triple RNA helical complex formed by two strands of inosinic acid and one strand of adenylic acid. The model refined by Bond and Arnott is shown in Figure 10 and possesses both the adenine-hypoxanthine base pairing schemes discussed above. Figure 10 demonstrates another feature of our system. A model displayed on the screen can be easily manipulated and viewed in any projection. Having arrived at a desirable view, the structure can be permanently recorded using a Calcomp plotter. The well-chosen use of color renders complicated drawings (and displays) less confusing.

Klug, Crick and Wyckoff have shown that streaky patterns like Figure 8b and those that characterize the C-DNA conformation (Marvin, et al., 18) can be explained in terms of screw disorder where molecules pack as interlocking smooth helices. Statistical crystal structures and other kinds of disordering also give rise to streaky patterns. Graphical displays that allow control of the individual helices in their crystallographic unit cells have been tailored to the study of intermolecular interactions that might determine preferred molecular orientations and modes of packing.

Finally, I should mention the model building programs. These accept typed-in sequences of amino acids for whole proteins, or of nucleotides for nucleic acids, like transfer RNA, and display them in some standard conformation. The building blocks are lists of coordinates for the different amino acids, nucleic acid bases and sugar rings or nucleotides and are stored with standard, fixed bond lengths and angles. Each pre-coded unit is joined to the preceding unit by superimposing its first bond onto the last bond drawn. For this connectivity mechanism a "common" bond is defined on entering and leaving each structural unit, such as amino acid residues that are defined beginning and ending with  $NC_{\alpha}$  bonds. In this way bonded stereochemistry is preserved between units.

Dihedral rotations are possible about specified single bonds. These can be selected as required and appear brightened on the screen. The motion of the "bright-bond" pointer through the molecule is stopped and started by a switch. The current position being recorded on the screen has a residue name and index and a value for the conformational angle. Values for angles can be typed in, read as continuously changing dials on the screen or read from lists. This last mode of input can be used to automatically wind a given sequence of amino acids into an  $\alpha$  helix,  $\beta$  sheet configuration, an extended chain, or a particular integral helix. Most planar and puckered ring components are considered as rigid units. Figure 11 shows a sequence of amino acids which have been built in three different conformations. The first section is an extended chain sequence where all the PHI and PSI polypeptide backbone angles have been set to the trans conformation. The bright-bond is in residue leu 3. The middle section has conformational angle values found in  $\beta$ -sheet sequences and the final section takes on  $\alpha$ -helical conformational angles.

The geometric orientation of a helical polynucleotide, with respect to its helix axis, is adjustable in terms of the traditional parameters; tilt, twist, swivel, and distance from the helix axis defined by Langridge, et al. (19). A model builder can specify the kind of helix he wishes to build. Marker positions are located on the screen and if by interaction with the molecular parameters, a model can be built close to the marker position, minimization routines capture the last bond. Figure 12 illustrates different stages in the computer graphical building of the B-DNA helical configuration. The vertical axis is the helix axis, the twist axis is shown passing through a purine base and the dyad or tilt axis completes the orthogonal axial set. In Figure 12a the labelled nucleotide base has been adjusted to the same position and orientation with respect to the helix axis as in the B-DNA model of Arnott, et al. (10). The sugar ring has been fixed with a C2-endo pucker. The unlabelled nucleotide residue exactly mimics the configuration of the first and forms the next repeat unit on a 10-fold integral helix with a pitch of 34 Å. Figure 12b represents the author's attempt at superimposing the O1+C3+ bond onto the marker bond by manually turning knobs which vary the conformational angles; CHI, XI, THET, PSI, and PHI (where CHI is defined by the labelled base atom C4 and atoms N9, C1, C2. XI, THET, PSI, and PHI are the successive dihedral angles defined by the atomic path C3, C4, C5, O4, P, O1+, and C3+, and are about the C4C5, C5O4, O4P, and P O1+ bonds, respectively). A measure of attaining coincidence is reflected in the parameter "H". H is displayed in figures 12a, b, and c, and is related to differences in the coordinates of the O1+ and C3+ atoms and their corresponding marker bond atoms. If a particular trial model can be joined-up in a satisfactory manner, as in Figure 12c, the whole helix is then completed (single, double, triple or quadruple), using the appropriate symmetry operators. Several copies of the resultant helix can be made and used to investigate packing models of the molecules in their unit cell. Finally, the Fourier transform of a given model can be calculated using its cylindrical polar coordinates. The speed of this operation should be emphasized. The time taken to build a polynucleotide model and to plot its Fourier transform on the Calcomp plotter can be as little as two minutes.

The computer graphical aspects of the work presented here were a collaboration of R. Langridge, A. Lesk, S. Stellman, and S. Harbison and the speaker, P. J. Bond.

#### ACKNOWLEDGEMENTS

I would like to acknowledge the technical assistance of Russell Massey and Jane Schroeder. This work was supported by USPHS NIH Grants RR-00578 and GM-16539 and NSF Grant GB-28021.

## References

1. Levinthal, C. 1965. "Computer Construction and Display of Molecular Models". Proceedings of the IBM Scientific Computing Symposium on Computer-Aided Experimentation. 315-321.
2. Langridge, R. and MacEwan, A. W. 1965. "The Refinement of Nucleic Acid Structures". Proceedings of the IBM Scientific Computing Symposium on Computer-Aided Experimentation. 305-311.
3. Levinthal, C., Barry, C. D., Ward, S. A., and Zwick, M. 1968. "Computer Graphics in Macromolecular Chemistry". In Emerging Concepts in Computer Graphics (J. Nievergelt and D. Secret, eds.), W. A. Benjamin, New York, 231.
4. Barry, C. D., Ellis, R. A., Graesser, S. M., and Marshall, G. R. 1969. "Display and Manipulation in Three Dimensions". In Pertinent Concepts in Computer Graphics (M. Faiman and J. Nievergelt, eds.), University of Illinois Press, Urbana, Illinois, 104.
5. Mayer, E. 1970. "Three-Dimensional Graphical Models of Molecules and a Time-Slicing Computer". J. Applied Cryst., 3, 392.
6. Barry, C. D. and North, A. C. T. 1971. "The Use of a Computer-Controlled Display System in the Study of Molecular Conformations". Cold Spring Harbor Symposia on Quantitative Biology, 36, 577.
7. Wipke, W. T. and Whetstone, P. 1971. "Graphic Digitizing in 3-D". SIGGRAPH-ACM, 5, 10.
8. Richards, F. M. 1968. "The Matching of Physical Models to Three-Dimensional Electron-Density Maps: A Simple Optical Device". J. Mol. Biol., 37, 225.
9. Diamond, R. 1966. "A Mathematical Model-Building Procedure for Proteins". Acta Cryst., 21, 253.
10. Arnott, S., Dover, S. D., and Wonacott, A. J. 1966. "Least-Squares Refinement of the Crystal and Molecular Structures of DNA and RNA from X-ray Data and Standard Bond Lengths and Angles". Acta Cryst., B25, 2192.
11. Hendrickson, W. A. and Love, W. E. 1971. "Structure of Lamprey Hemoglobin". Nature New Biology, 232, 197.
12. Cotton, A. F., Bier, C. J., Day, V. W., Hazen, E. E., and Larsen, S. 1971. Cold Spring Harbor Symposia on Quantitative Biology, 36, 243-255.
13. Matthews, B. W., Sigler, P. B., Henderson, R., and Blow, D. M. 1967. Nature, 214, 652.
14. Lipscomb, W. N., Coppola, J. C., Hartsuck, J. A., Ludwig, M. L., Muirhead, H., Searl, J., and Steitz, T. A. 1966. J. Mol. Biol., 19, 423.
15. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. 1958. "A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-ray Analysis". Nature, 181, 662.
16. Perutz, M. F. 1970. "Stereochemistry of Cooperative Effects in Haemoglobin". Nature, 228, 726.
17. Cuatrecasas, P., Wilchek, M., and Anfinsen, C. B. 1969. "The Action of Staphylococcal Nuclease on Synthetic Substrates". Biochemistry, 8, 2277.
18. Marvin, D. A., Spencer, M., Wilkins, M. H. F., and Hamilton, L. D. 1961. J. Mol. Biol., 3, 547.
19. Langridge, R., Marvin, D. A., Seeds, W. E., Wilson, H. R., Hooper, C. W., Wilkins, N. H. F., and Hamilton, L. D. 1960. J. Mol. Biol., 2, 38.

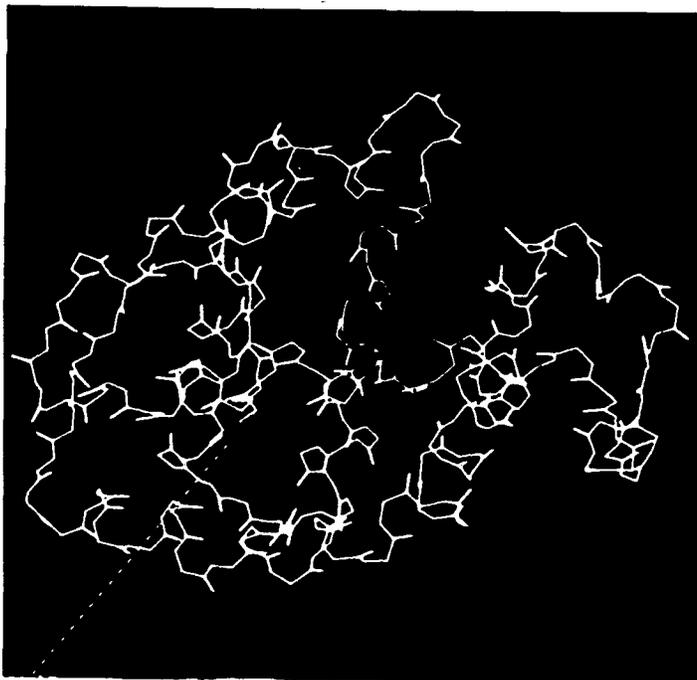


Figure 1

A computer graphical display of the polypeptide backbone of lamprey hemoglobin (11). The pointer locates the carbon  $\alpha$  atom of pro 1. Side chains have not been displayed. The molecule can be seen to possess the eight helical regions characteristic of the "myoglobin fold". The A helix runs horizontally across the bottom of the figure, from left to right, and becomes the vertical B helix. This then loops to the right, through the C and D helices into the E helix, which runs horizontally across the center of the figure, from right to left. The F helix runs across the top left hand corner, to the top center of the molecule. The polypeptide chain drops vertically as the G helix forms the GH loop to the left and then rises vertically as the H helix. At its amino end (located by the pointer), lamprey hemoglobin has a characteristic nonapeptide which runs vertically along the H helix and loops to the left into the A helix.

Figure 2

A close-up view of the GH loop and the H helix, including the amino acid side chains, of the lamprey hemoglobin molecule. Residues from leu 118 (located by the pointer) to tyr 148 have been drawn. The GH corner assumes a novel configuration and the H helix is two turns shorter than in the sperm whale myoglobin and the horse hemoglobin molecules. Tyr 148 at the end of the H helix is homologous with the penultimate tyrosine in horse and human hemoglobins and with tyr H22 in myoglobin. This tyrosine appears to serve a corresponding function in all H chains.

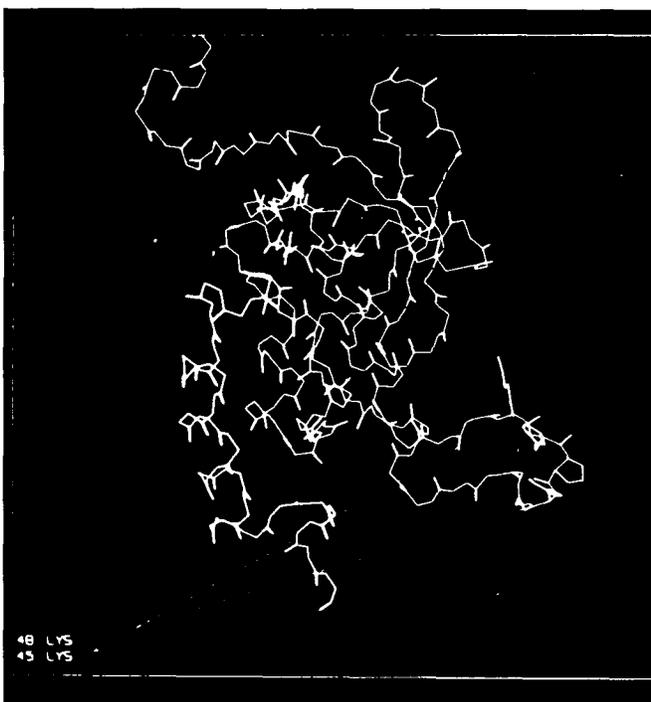
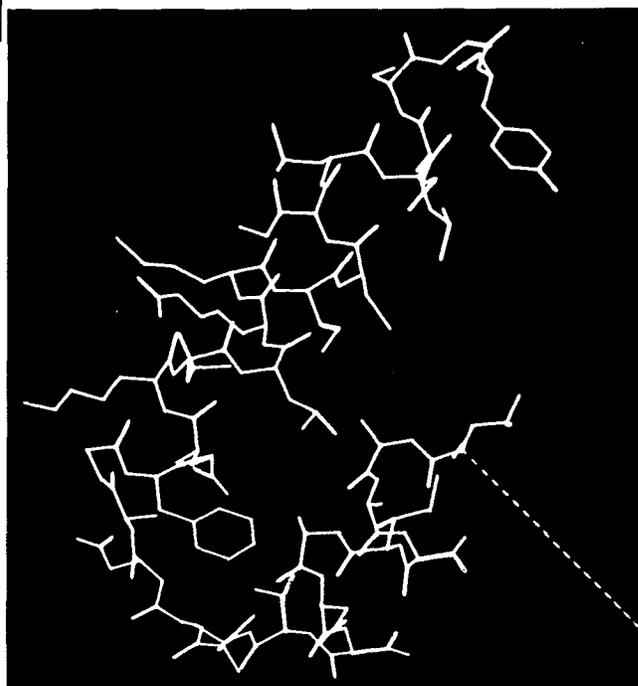


Figure 3

A view of the polypeptide backbone of the staphylococcal nuclease molecule. Residues from lys 45 to lys 48 have been drawn complete with their side chain atoms.

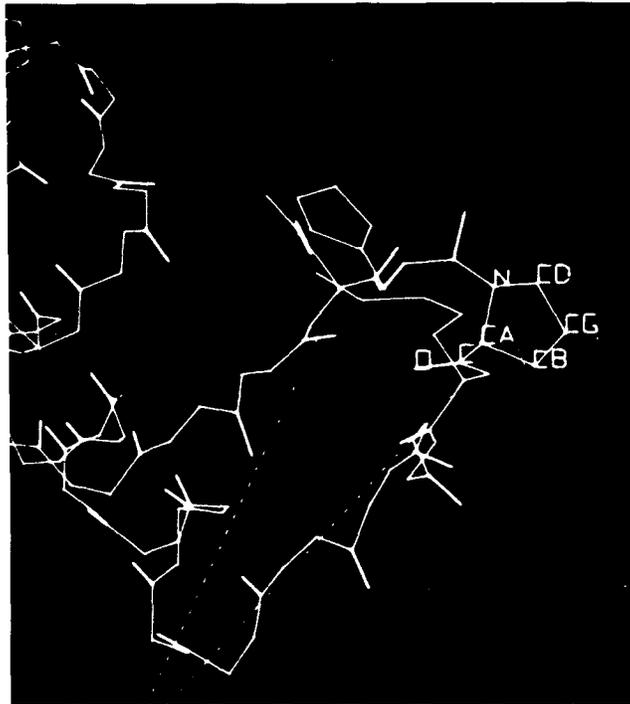
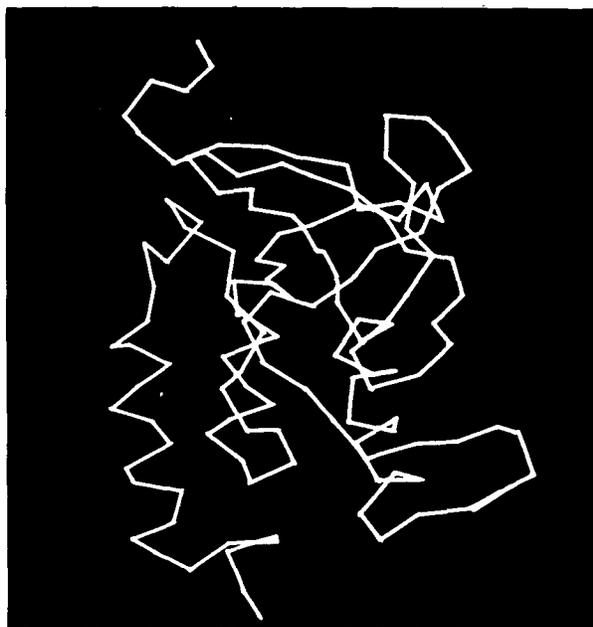
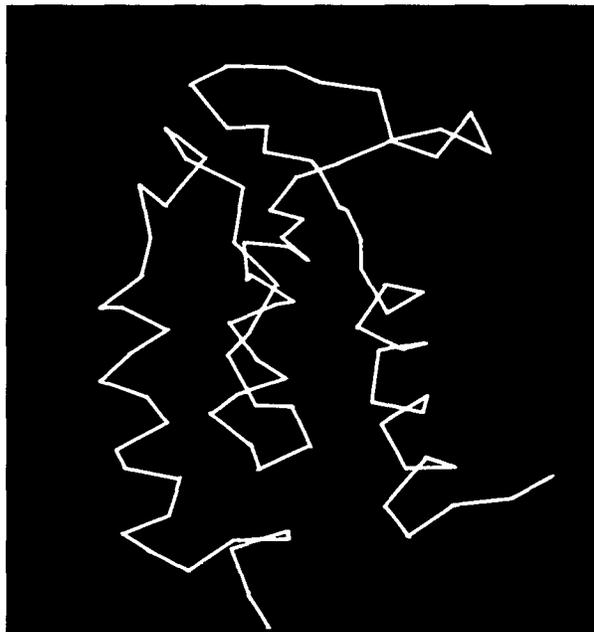


Figure 4

This is a close-up view of the loop in the bottom right hand corner of Figure 3. The pointers locate the carbon  $\alpha$  atoms of residues lys 45 and lys 48. Pro 47 has been labelled. The side chains of the sequence between the pointers extend outward and are relatively exposed to solvent.



(a)



(b)

Figure 5

a) A display of the carbon  $\alpha$  backbone of the staphylococcal nuclease molecule. This view shows the three short stretches of helix at the left, center, and right of the molecule, a three-tier section of anti-parallel pleated sheet forming the upper right side of the molecule and another section of anti-parallel pleated sheet formed by the two parts of the chain going vertically up and down the back of the molecule.

b) This figure shows the carbon  $\alpha$  backbone from lys 49 to the end of the molecule; the so-called P3 fragment.

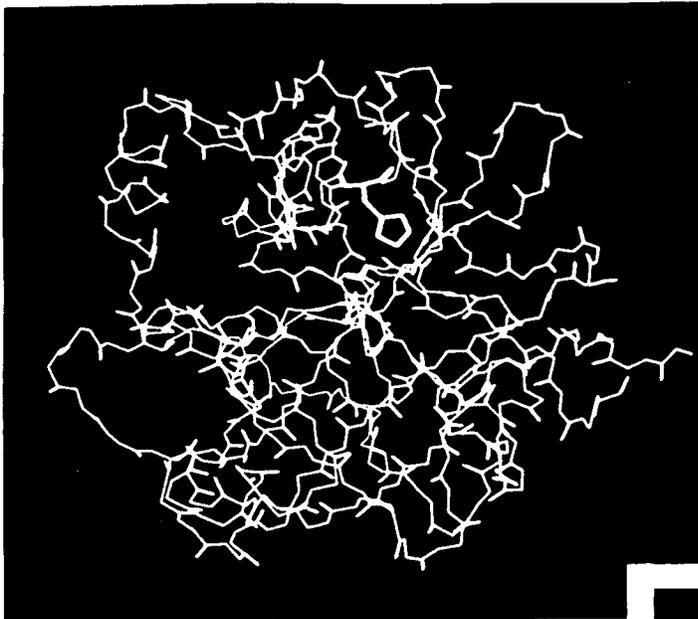


Figure 6(a)

The polypeptide backbone of  $\alpha$  chymotrypsin. The side chain of his 57 has been added and the whole residue brightened to emphasize its position in the molecule. Note the suggestion of a twisted  $\beta$  sheet in the center (this has been singled out and displayed in Figure 6(f)).

Figure 6(b)

The carbon  $\alpha$  backbone of chymotrypsin. The catalytically important residues his 57, asp 102 and ser 195 have been displayed and labelled.

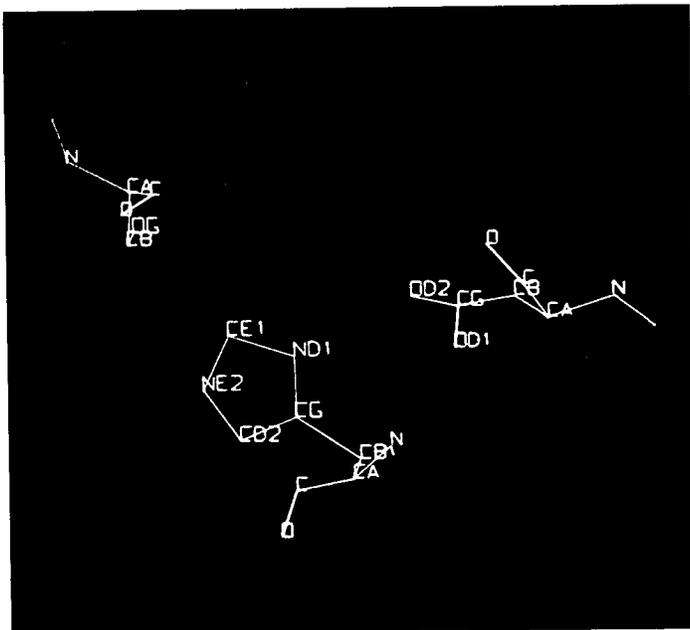
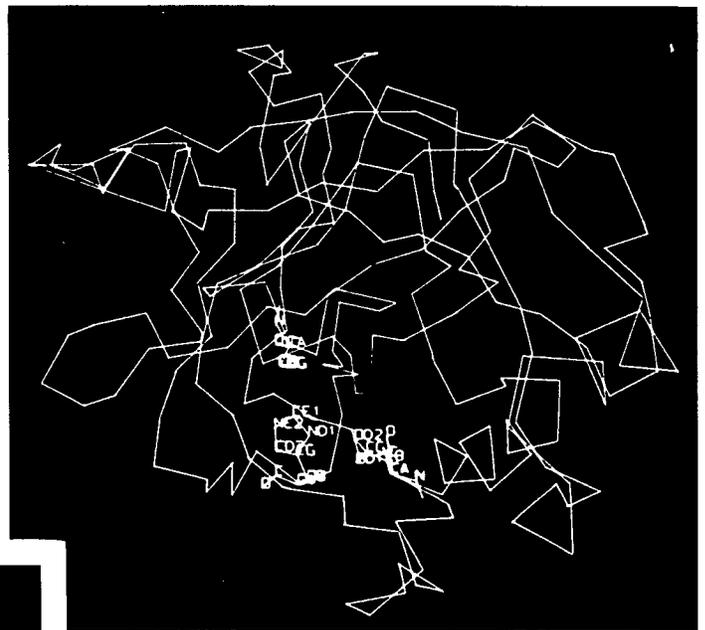


Figure 6(c)

The relative positions of the three residues involved in the proposed cleavage mechanism of chymotrypsin. The mechanism involves ser 195 (top left) attacking a bond by donating a proton. Interaction with his 57 (center) and with the buried asp 102 (right center) gives ser 195 a negative charge making it a strong nucleophile.

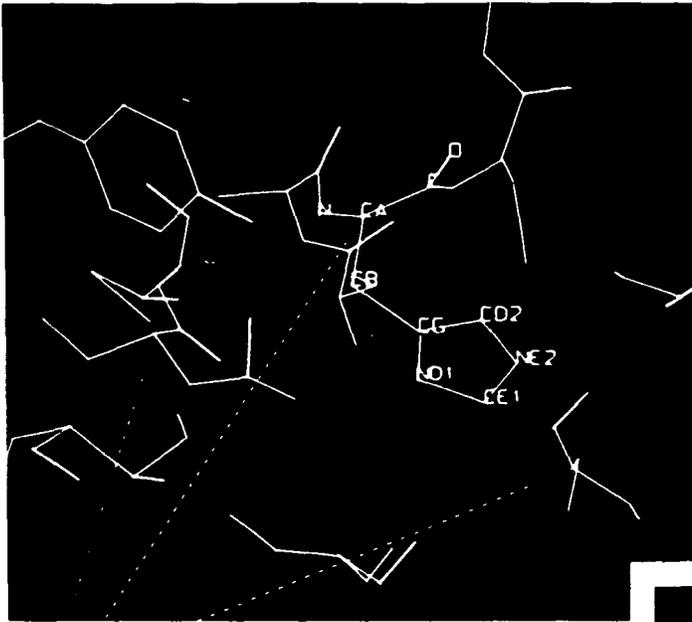


Figure 6(d)

A display of residues that have their carbon  $\alpha$  atoms within 10 Å, and of all other atoms within 7 Å, of the carbon  $\alpha$  of the his 57 residue (labelled and located by the central pointer). The pointer to the left locates asp 102 and that to the right, ser 195.

Figure 6(e)

The disulfide bridge from cys 1 to cys 122. These two residues have been brightened and labelled against a background of polypeptide backbone bonds.

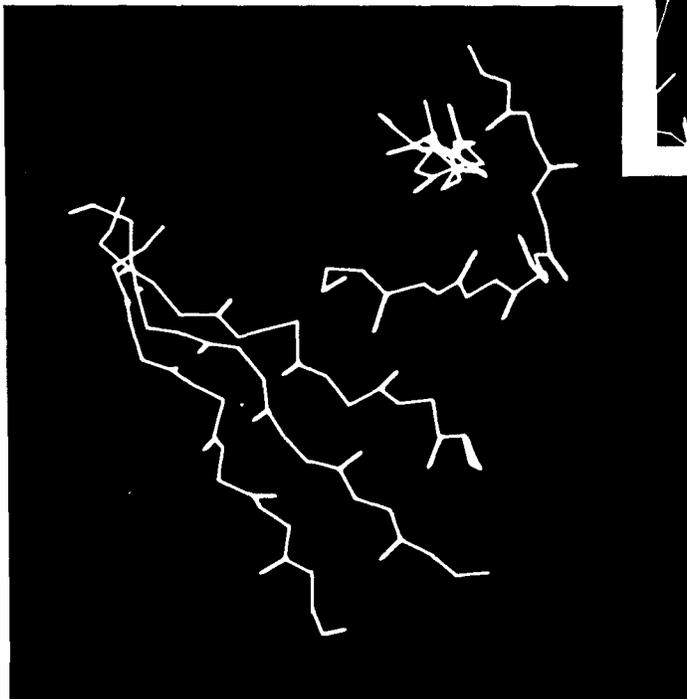
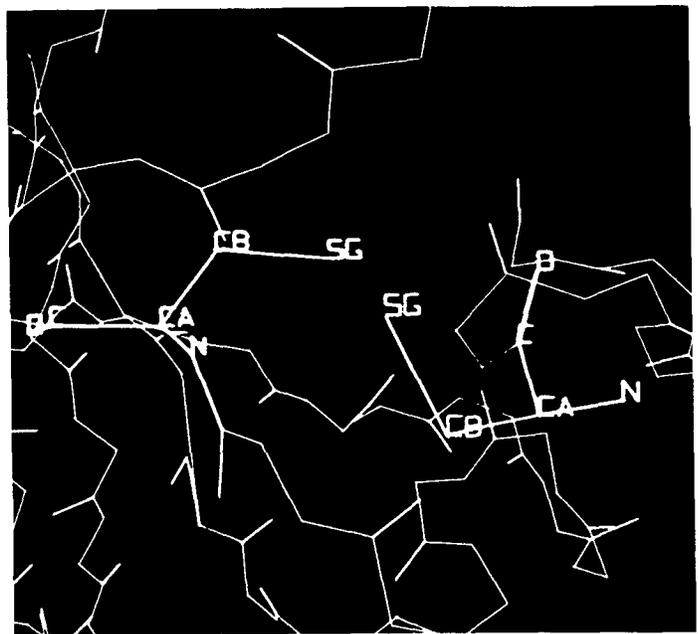


Figure 6(f)

The polypeptide backbone strands from residues 91-86, 103-108, 55-50, 39-46 and 35-29 form a twisted sheet.

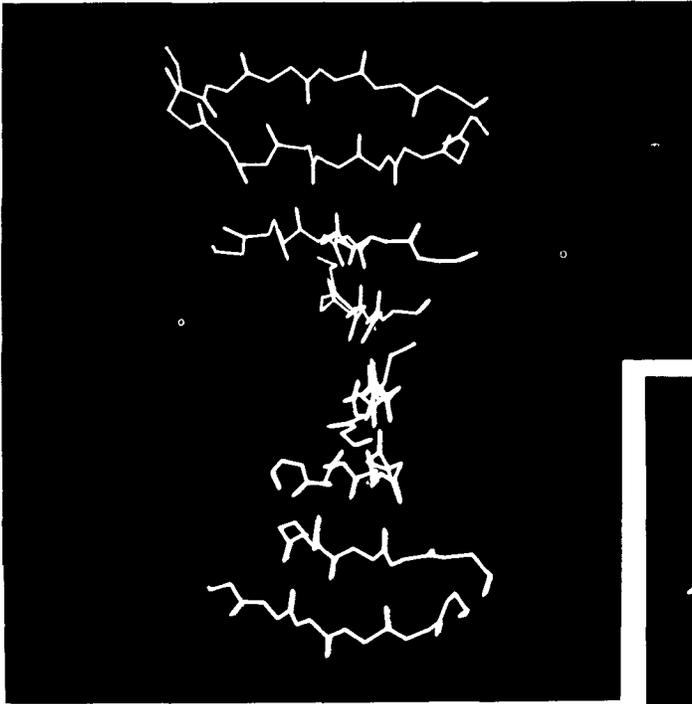


Figure 7(a)

The eight chain twisted  $\beta$  sheet that forms the core of carboxypeptidase A. The  $\alpha$  to  $\theta$  strands are from residues 32-37, 46-54, 61-67, 103-111, 190-197, 200-205, 238-243, 265-271. No side chains have been displayed.

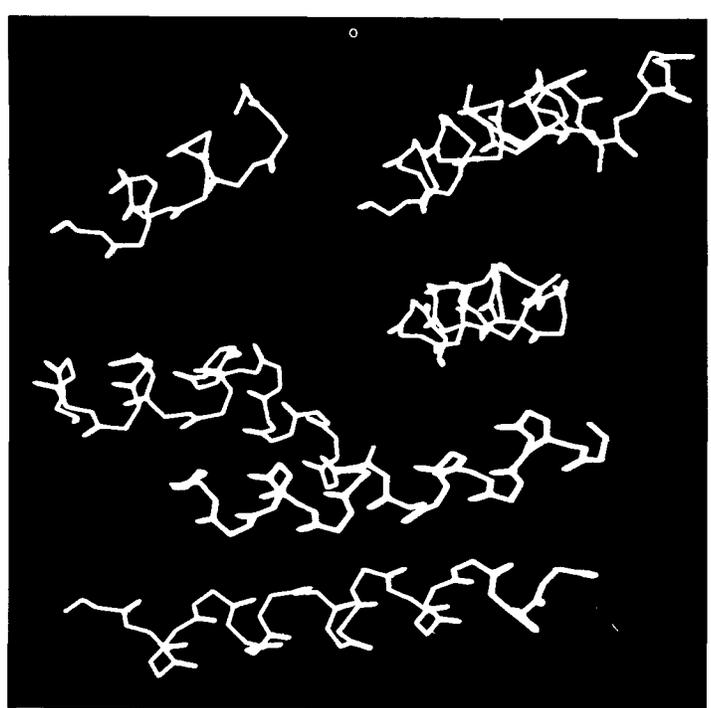


Figure 7(b)

The eight  $\alpha$  helices that cluster around the  $\beta$  sheet of figure 7(a). The A to H helices are formed by residues 14-29, 72-88, 94-103, 115-122, 174-184, 215-233, 254-262 and 288-305.

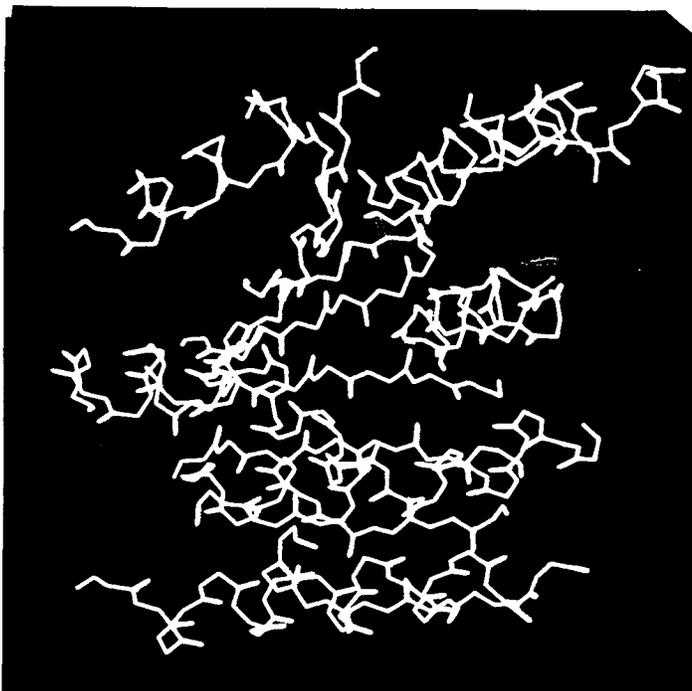


Figure 7(c)

The twisted  $\beta$  sheet packed with  $\alpha$  helices.

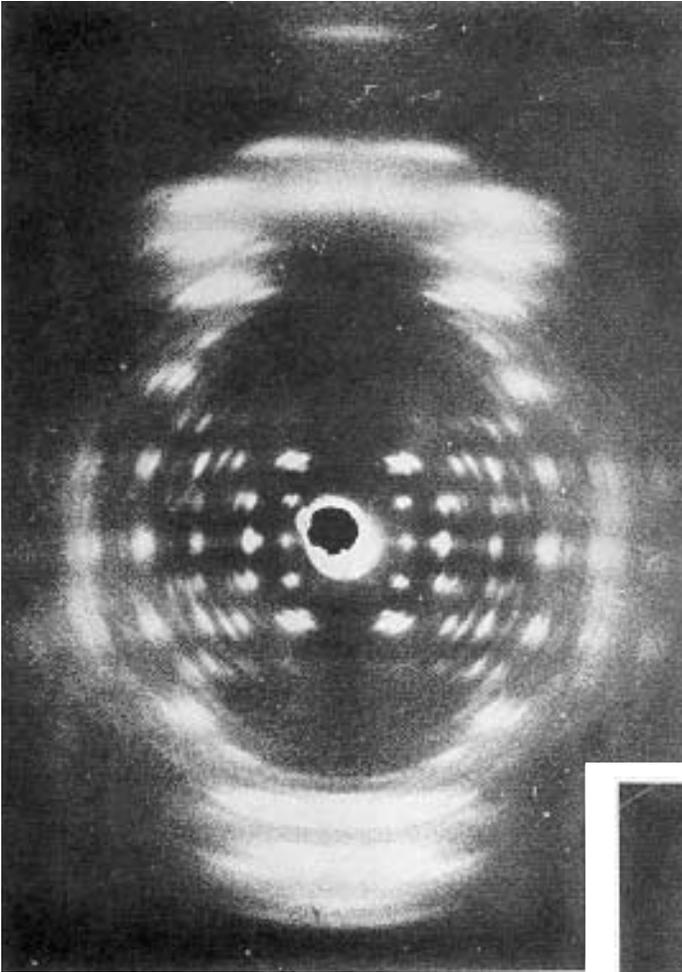


Figure 8(a)

X-ray fiber diagram obtained from a sample of Micrococcus lysodeikticus DNA. The fiber was tilted from the perpendicular to the X-ray beam, to record the meridional reflection on the eleventh layer line. The pattern was obtained at 75% relative humidity and is characteristic of the A-DNA conformation.

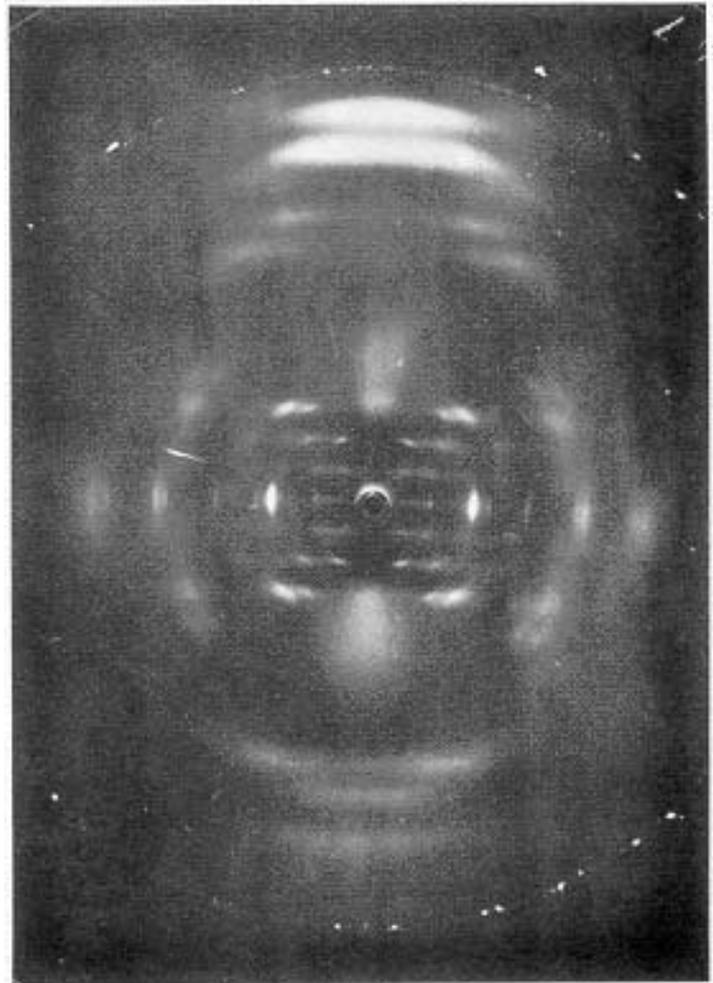


Figure 8(b)

X-ray fiber diagram obtained from a sample of triple RNA helical complex formed by two strands of inosinic acid with one of adenylic acid. The fiber was tilted to record the meridional reflection on the twelfth layer line. The pattern was obtained at 75% humidity.

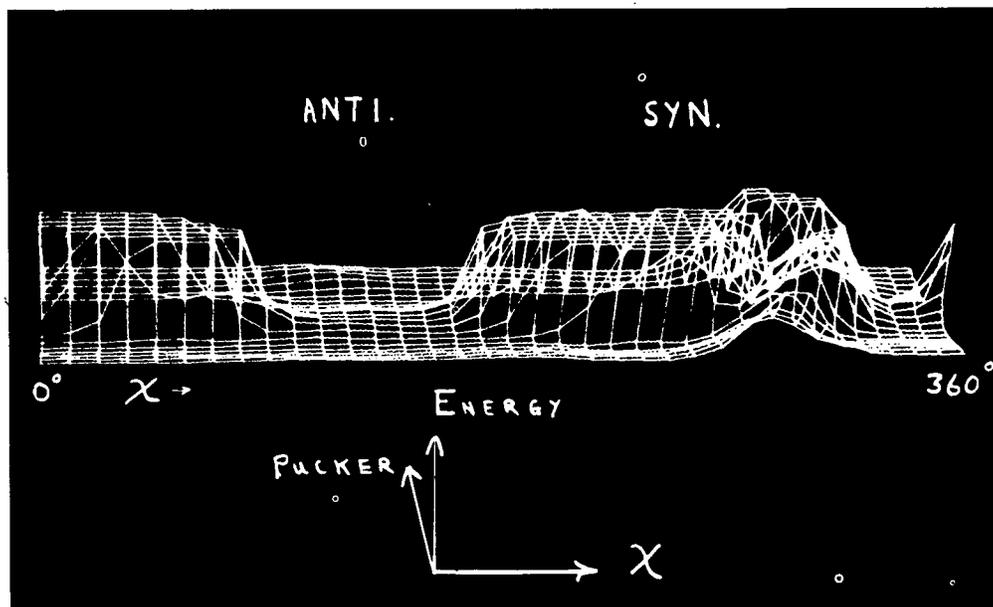


Figure 9

A computer graphical energy surface. The energy has been calculated for an adenine ribonucleoside using a function which accounts for non-bonded interactions, electrostatic interactions and a strain term.

$$E = \sum_{i<j} [-ar_{ij}^{-6} + br_{ij}^{-12}] + \sum_{i<j} \frac{c q_i q_j}{r} + \sum_{k=1}^5 V_{k, \text{strain}}$$

(where summations for  $i<j$  are for non-bonded atoms.  $r_{ij}$  is the distance between atom  $i$  and  $j$  and  $a$ ,  $b$ , and  $c$  are constants.  $q_i$  is the electrostatic charge on atom  $i$  and  $V_{k, \text{strain}}$  is the Hook's law potential for strain about the pentose atoms). Energy has been plotted for  $\chi$  varying from  $0^\circ$  to  $360^\circ$  and for twenty different puckers of the ribose sugar. The "pucker" axis is a pseudo-rotational coordinate related to a particular pucker. The anti and syn regions correspond to  $\chi$  in the angular ranged  $\sim 90^\circ$  and  $\sim 270^\circ$ , respectively.

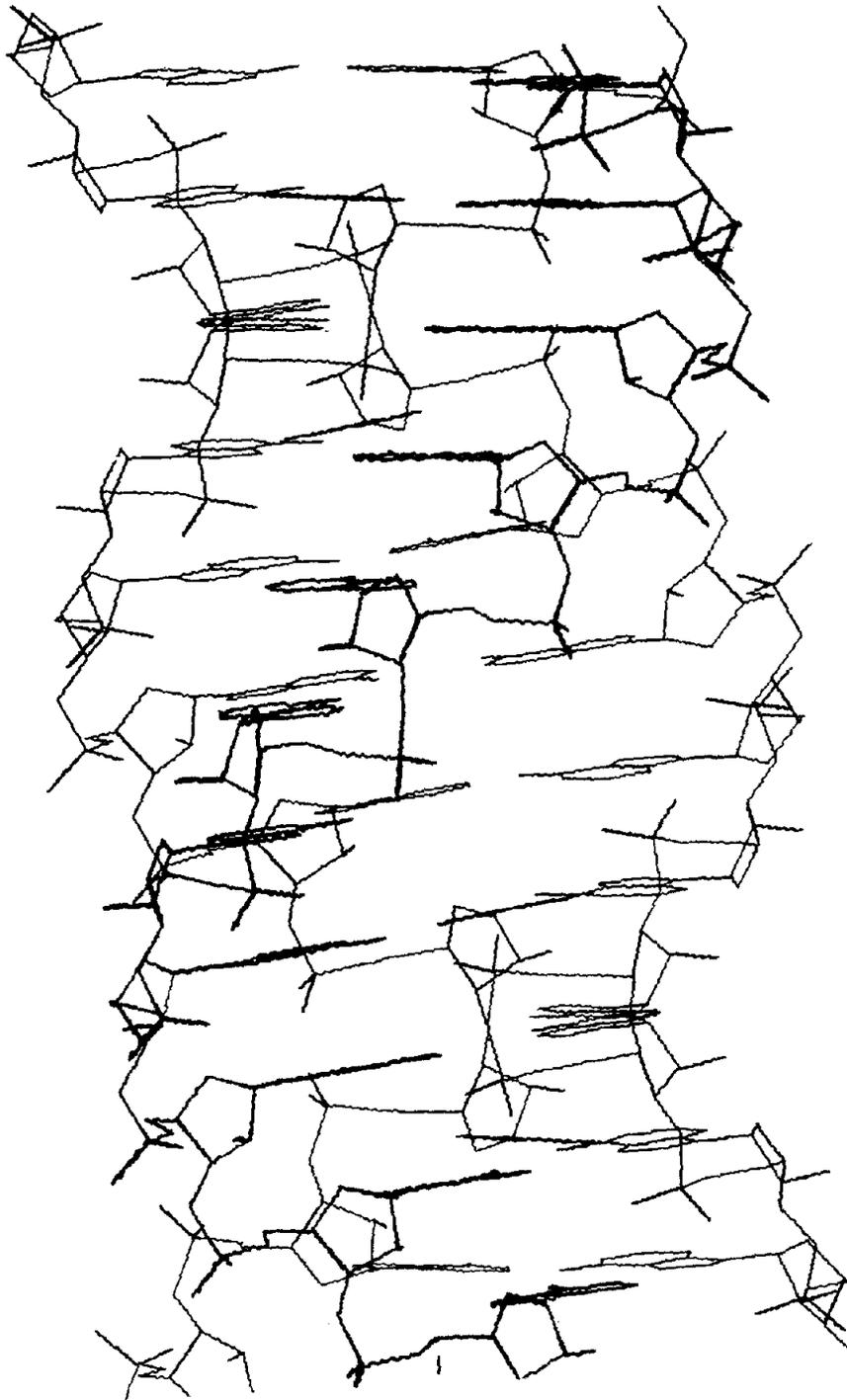


Figure 10  
The refined triple helical model for poly riboadenylic acid with two strands of poly riboinosinic acid. The model is from the work of S. Arnott and P. J. Bond.

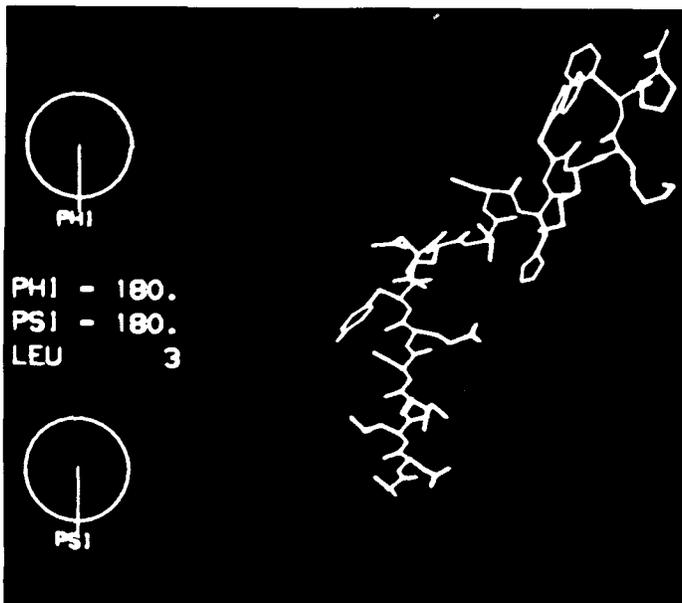
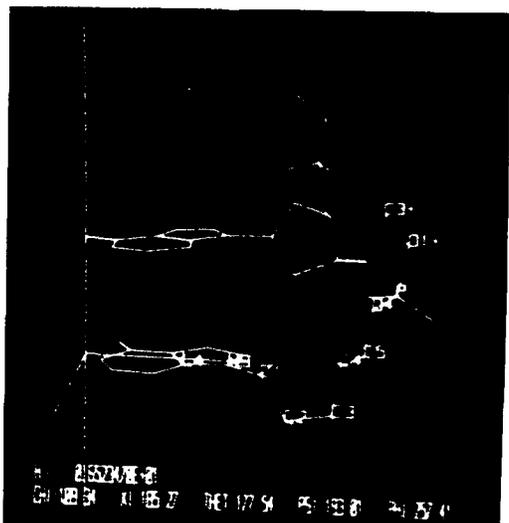


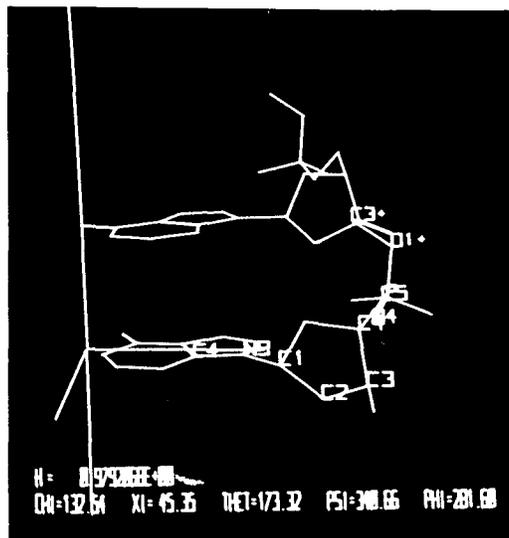
Figure 11

A computer graphical display generated by the protein model building program. A sequence of amino acids has been connected together with standard, fixed bond lengths and angles. The vertical section has the polypeptide backbone angles PHI and PSI in the trans conformation. The "bright-bond" marker locates residue leu 3, in this sequence, and reports numerically and graphically the values of its dihedral angles. The middle sequence has conformational angle values found in  $\beta$ -sheet configurations, and the sequence in the top right hand corner has been given  $\alpha$ -helical conformational angles.

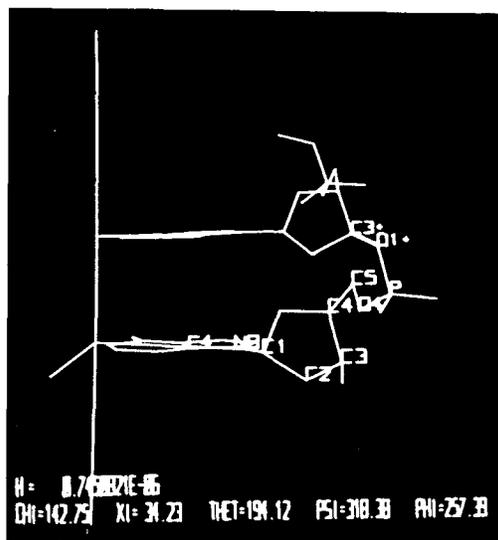


Figures 12 Stages in the computer graphical building of the B-DNA helix.

12(a) The vertical axis is the helix axis. The position and orientation of the nucleotide units with respect to the helix axis have been adjusted to those of the B-DNA model of Arnott, *et al.* (10).



12(b) By turning knobs the investigator can adjust the dihedral angles CHI, XI, THET, PSI and PHI until the O1+C3+ bond is close to its marker position. These adjustments may be used to bias the model in any way; such as to ensure that conformational angles are in preferred "standard" ranges.



12(c) The O1+C3+ bond has been captured, by a minimization routine to its marker bond. The model is a 10-fold helix with a pitch of 34 Å and conformational angles very similar to those of the published model (10).