APPLICATION OF THREE-DIMENSIONAL INTERACTIVE GRAPHICS IN X-RAY CRYSTALLOGRAPHIC ANALYSIS*†

STEVEN D. STELLMAN

American Health Foundation, 1370 Avenue of the Americas, New York, N.Y. 10019 U.S.A.

Abstract—A program called XpY was written for the PDP-10/LDS-1 at the Princeton University Computer Graphics Laboratory, for generating and displaying models of dinucleoside phosphates. The molecule GpC, a member of this class and a fragment of the nucleic acid RNA, was subjected to X-ray diffraction analysis.

The paper describes the importance of model building in X-ray analysis, and shows step by step how XpY was used to deduce the atomic coordinates of GpC from the experimental data. The program documentation is also included as an Appendix.

A subjective critique of graphics is made in the Conclusions section.

1. INTRODUCTION

Chemists have traditionally been counted among the most enthusiastic and successful developers of new technological methods, and have been quick to spot and exploit those features of new technology which might be beneficial to their research. Therefore, it is somewhat disappointing that in the application of computer graphics to chemistry and chemical physics, the promises and expectations have for a long time been considerably in excess of concrete accomplishments. One of the main problems along the way is probably the fact that, while for a chemist, the structure of a molecule is a simple concept, the expression of structure in terms suitable for efficient computer manipulation depends greatly on the nature of the specific problem to be solved. Since there is no such thing as a 'standard' computer representation of chemical structure, every laboratory has had to work out its own coordinate generation and manipulation algorithms, resulting in a great duplication of effort and nearly complete incompatibility of programs and techniques. In addition, the scope of usage of graphics in chemistry has been mostly 'look and see,' consisting of simple displays of molecules whose structures and properties had been previously elucidated by non-graphical methods. Until quite recently there has been no significant integration of graphics into the sophisticated array of other instrumental methods for molecular analysis.

Fortunately, this picture is rapidly changing. Computer graphics is just now experiencing a welcome advancement at the fertile hands of chemists. Katz and Levinthal have recently published a review of graphics work done at Columbia[1], in which they deal in detail with their molecular structure algorithms. They cite as two applications the use in 'picking peaks' (separating signal from noise) in crystallographic electron density maps, applied to UpA—a molecule similar to our GpC; and also the use of graphics in biology for serial section anatomy.

Anatomical sectioning is also being developed by Sheryl Glasser at Princeton and at the University of California at San Diego. At Princeton we have also examined the analogy between serial sections and the

so-called 'Richards Box' for matching crystallographic peaks with molecular models. A. Lesk has recently reported on pattern recognition in crystallographic analysis[2], and has written graphics programs for visual matching of calculated and observed Paterson (autocorrelation) functions.

Work is being done at Cornell to compare homologous or alternative sub-structures in similar proteins[3]. Along these lines, in our own lab, S. Harbison has written CAAPS (Computer Aided Analysis of Protein Structure), for displaying and manipulating structures of known proteins.

Professor W. T. Wipke and coworkers at Princeton have developed the system SECS (Simulation and Evaluation of Chemical Synthesis) for discovering synthetic routes to desired organic compounds. Prof. Wipke's group is also responsible for developing the GIGL language used in the XpY program.

My purpose in this paper is to show how our group, the Computer Graphics Laboratory at Princeton University, has developed interactive three-dimensional interactive graphics as a unique and useful research tool in X-ray crystallographic analysis. In the example I have chosen, the structure of the RNA fragment called GpC, the graphics played an essential role in determining from the experimental data the precise atomic locations and molecular conformations.

2. CHOICE OF THE MOLECULE GpC

The molecule guanylyl-3',5'-cytidine monophosphate (GpC) was selected for study for a variety of reasons, not the least of which was simply its availability in pure crystalline form suitable for X-ray diffraction analysis. (Molecules like this are generally extremely difficult to obtain and crystallize.) Of greater interest is its importance as a biological molecule. GpC is a fragment of the nucleic acid RNA, the polymer which transmits genetic information from DNA of the cell's nucleus to the cytoplasm, where that information is decoded and translated into the synthesis of the proteins and other molecules necessary for the metabolic functioning of the cell[4].

Several varieties of RNA are known, and each plays some vital part in this information-transmission process. The mechanism of this process is a target of intense study.

It is well known that the *shape* or *conformation* of each form of RNA is a crucial factor in its correct biological functioning. Hence, X-ray analysis is a natural choice for studying these molecules, because it is the physical technique best capable of revealing the relative positions of atoms and thus the conformation.

Why not, then, perform X-ray analysis directly on RNA rather than its sub-units? In fact, this is often done with a great measure of success[5], but because of the tremendous size of these polymers and the random repetition of their sub-units, single crystals cannot be obtained. This procedure inherently limits the resolution of coordinates to about 3Å at best, which is not sufficient to distinguish individual atoms. Crystallography of low molecular weight compounds, by contrast, can routinely locate atoms within 1Å resolution or less, and can even account for anisotropic thermal atomic vibrations.

Finally, the choice of GpC as a representative of its class was important because knowledge of its conformation would be helpful in constructing a model of a smaller type of RNA for which single crystal data *is* available, namely, transfer RNA or tRNA. Model-building will be an essential preliminary task in the ultimate solution of the tRNA's atomic structure, for reasons now to be explored.

3. IMPORTANCE OF MODEL-BUILDING

To obtain the atomic structure of any molecule, it should in principle be possible to simply subject the crystal to an X-ray beam and measure the intensities and phases of the diffracted rays. Since the atoms (or rather their electrons) are responsible for the scattering, a three-dimensional Fourier inversion of the diffraction pattern must reveal the atomic coordinates:

$$\rho(x, y, z) = \sum_{h,k,l} |F(h,k,l)| \exp[i\alpha(h,k,l)] \exp[2\pi i(hx + ky + lz)]$$

where $\rho(x, y, z)$ is the electron density at a point in space (x, y, z), h, k and l are integral Miller indices for planes passing through the crystal (h is the reciprocal of the x intercept of a given plane; k and l are the y and z intercepts), $|F(h, k, l)|$ is the square root of the intensity of the diffracted beam in the direction h, k, l, and $\alpha(h, k, l)$ is the phase of that beam.

Computer analysis of diffraction data involves much more than simply evaluating the above formula, however, for the following reason. Any electromagnetic radiation (such as an X-ray) can be represented mathematically by two components, the amplitude and phase, given in the equation by $|F|$ and α, respectively. Unfortunately, the phase is not measurable with present diffraction equipment, so that the experimenter is left with only half the necessary data, viz. the amplitudes. The problem of restoring the phases is one of the most significant problems of molecular biology. For both fiber and crystal work, the most reliable method is to construct a trial model of the molecule under study and calculate the amplitudes and phases from the molecular model. By use of iterative least-squares procedures, the correct phases can be converged upon, and the true structure determined[5].

Use of least-squares methods, as in other applications of numerical analysis, depends critically on the closeness of the trial structure to the actual solution. An average error of about 0·5 Å ($=5 \times 10^{-9}$ cm) in every atomic coordinate is often sufficient to prevent convergence of the trial structure to the correct molecular structure of a molecule.

Thus, it is quite important, in the case of complex molecules, to be able to predict *ab initio* a nearly perfect trial structure even before analysis of data is begun. In the past this was generally done using stiff wire models, plus the knowledge of likely interatomic bond distance and angles acquired from solution of many molecules.

However, this procedure is not altogether satisfactory for molecules which have a low resistance to deformation and which could conceivably exist in a wide variety of conformations. Wire models, which are excellent when it comes to holding rigid bond lengths and angles, are notoriously poor for modeling flexible torsion angles between successions of bonds. Besides giving inaccurate dihedral angles, the metal variety have to be untightened and retightened with Allen wrenches for every slight adjustment, and the plastic-tube-and-pin models begin to slip after very little manipulation. Furthermore, for chemical structures containing many bonds and occupying great volumes of space, such as nucleic acids and proteins, external supporting devices are required which nearly always render the interior of the molecule inaccessible both to adjustment and view. Finally, under these circumstances, the all-important coordinates are quite difficult to measure accurately enough to serve as input for the least-squares procedure. Enter graphics.

4. DESCRIPTION OF THE COMPUTER

The Princeton University Computer Graphics Laboratory is equipped with a Digital Equipment Corporation PDP-10 computer with 64K 36-bit words of memory plus a 5·2 million word disc, coupled to an Evans and Sutherland Line Drawing System 1, a second computer specialized for the generation of interactive displays.

Data for display by the LDS-1 pass through a hardware matrix multiplier and clipping divider, to a line generator producing visible output on a CRT whose viewing area measures 10×10 in. The matrix multiplier produces coordinates for a projected view of a three-dimensional structure. The clipping divider eliminates those portions of lines lying outside a window defined by the programmer.

The LDS-1 can display several thousand lines, flicker-free, with real-time manipulative control.

The design of the LDS-1 emphasizes three-dimensional displays. Its special hardware features include: (1) internal registers and instructions that operate directly with three-dimensional coordinates; (2) hardware matrix multiplication; (3) depth queueing, or diminishing of intensity as the simulated distance of the image from the observer increases; (4) automatic perspective; (5) the capacity to generate stereoscopic displays, by synchronizing two offset images with a rotating wheel that alternately masks each eye. (Multicolored displays are achieved by a similar technique.)

Interaction between programmer and display is posible by means of input devices attached to the LDS-1. These consist of a bank of switches, a bank of knobs, a 'joystick' supplying simultaneously two continuous variables, and a three-dimensional acoustic tablet, which senses the x, y, and z coordinates of a sparking penpoint. A bank of lights coupled with the switches is available for output. The sixteen switches can each be set to one of three positions: permanent off, permanent on, and spring-loaded on. A button atop the joystick can serve as a seventeenth switch.

All display programs must: (1) generate a data base of coordinates in the form of packed 18 bit fixed point integers; (2) set registers in the LDS-1 governing the appearance of the display, such as the portion of the screen on which the display appears (registers in the matrix multiplier control translation, rotation, and scaling); (3) execute instructions to generate an image on the screen; and (4) read the interactive input devices for possible modification of the display. For example, a displacement of the joystick might be converted into a displacement of part of the display. Any desired assignment of the parameters supplied by the interactive input devices to kinematic or other modification of the display may be programmed in a suitable language. Several such high level languages have been tested. One has been described by its author, Arthur Lesk[6].

The program was written entirely in FORTRAN. The graphics manipulation package was GIGL v. 2·0, a FORTRAN-like processer wirtten by T. Dyott and W. T. Wipke for this machine. Our experience has shown this to be a superb tool for scientific applications on this machine. A major advantage is that, as the language is entirely compatible with FORTRAN, graphic statements may be freely mixed with non-graphical ones for maximum convenience.

5. XpY: THE MODEL-BUILDING PROGRAM

The philosophy of the program to be described is simple: to give the graphics user the capability, from the console, to construct models of the desired molecule consistent with his pre-conveived assumptions about its structure. Essential to success is the ability to immediately test a given model against the X-ray data to see whether it is correct. (There is a built-in irony in crystallography that, while there is no trivial way to deduce a structure directly from the data, any guess at a structure can be trivially and immediately tested to see whether it is indeed correct).

The action of the program will now be described. A concise instruction manual is included as an Appendix. At the computer's request, the user types two choices from the set A, G, C, or U, corresponding to the desired chemical components desired in the dinucleoside phosphate. Sixteen different molecules can be constructed this way, selected from Adenine, Guanine, Cytosine, or Uracil residues, and otherwise having analogous chemical structure to two consecutive bases of RNA. The selected molecule is generated and displayed on the screen. The user then has the option to change values of parameters, such as internal torsional angles and sugar puckering,

either by rotation of a knob or by typing in a specific value on the teletype. The displayed molecule instantly adopts the selected conformation, and the values of the dihedral angles are simultaneously displayed as text at the bottom of the screen. By pressing some of the seventeen switches, the user can at any time request the computer to perform certain tasks, such as typing the current coordinates, or drawing the molecule on the Calcomp plotter for hard copy. The image on the screen may be manipulated so as to show any desired view or perspective.

6. SPECIAL FEATURES FOR GpC

The program functions described above are applicable to any of the sixteen possible XpY molecules. While GpC is the one of particular interest here, the capability for the fifteen others was included because their coordinates were simple to generate (chemically analagous to GpC) and were needed for other biophysical studies made in this laboratory[7].

For GpC, there are certain extra options which relate directly to the existing X-ray data for this molecule, and which were incorporated specifically for use with this data. These options have mostly to do with the overall orientation of the rigid GpC molecule, rather than with the flexing of internal bonds. The X-ray data gave at the outset the physical dimensions of the crystalline unit cell, which is the smallest geometrically repeating unit, and independent density measurements told us the number of GpC molecules (four) to be expected within this volume. These four molecules, furthermore, had to obey certain symmetry relationships among themselves, with the proviso that they also not be intertangled. Thus, it was as important to manipulate the overall mode of molecular packing for GpC as to deform the molecule about its flexible internal bonds.

7. SOLUTION OF THE CRYSTAL STRUCTURE OF GpC

The chemical primary structure of the molecule is shown in Fig. 1, and definitions of the dihedral angles are in Table 1. The solution took place in two stages:

1. The conformation of a single GpC molecule was determined *in vacuo*, that is, without regard for solvent or neighboring molecules, by numerically minimizing the intramolecular potential energy[8].

2. The minimum energy structure was packed into the experimental unit cell via graphics and numerical
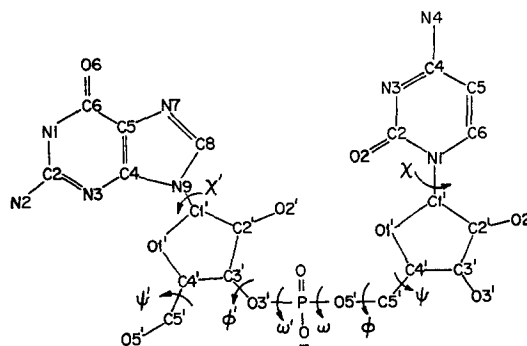

Fig. 1. Chemical structure of GpC.

Table 1. Definition of dihedral angles for GpC

Angle	Bonds
χ'	O1'–C1'–N9–C8
ψ'	C3'–C4'–C5'–O5'
ϕ'	P–O3'–C3'–C4'
ω'	O5'–P–O3'–C3'
ω	C5'–O5'–P–O3'
ϕ	C4'–C5'–O5'–P
ψ	C3'–C4'–C5'–O5'
χ	C6–N1–C1'–O1'

All angles A–B–C–D are measured clockwise from A to D when viewed along B–C. A eclipsing D is 0°.

methods, subject to additional experimental constraints. Each packing mode so found was visually examined and those free of intermolecular entanglements were compared against the X-ray data by Fourier inversion. That packing mode yielding the smallest discrepancy index (a number computed somewhat like a least-squares residual) was deemed correct and further refined by non-graphical methods[9].

An explanation of the rationale for systematically choosing various packing schemes for study has been presented elsewhere, as have details of the *in vacuo* calculation[8,9]. As these details are largely in the chemical realm, they are of minimal interest here and are omitted.

A selection of the various packing schemes found and their discrepancy indexes relative to the experimental data are given in Table 2; the schemes are pictured in Fig. 2. In particular, the one labeled A had the lowest R-factor, 39·0 per cent, and was tentatively thought correct. A simple independent test was available (SWITCH 4) by graphically superimposing the 4,0,2 plane on the displayed model, and observing that a great many atoms lay in this plane. Since the X-ray data showed a very strong reflection in that direction, implying that much scattering material lies in this plane, this was an important confirmation of the trial structure.

Finally, the coordinates of all the atoms were transferred to tape and taken to a much faster computer (360/91) for subsequent least-squares refinement. This work is now complete and is described elsewhere[10].

8. HOW THE PROGRAM WORKS

After initializing a number of variables, setting up the display logic in terms of items, instances, and pictures (required by GIGL), calculating initial coordinates of the
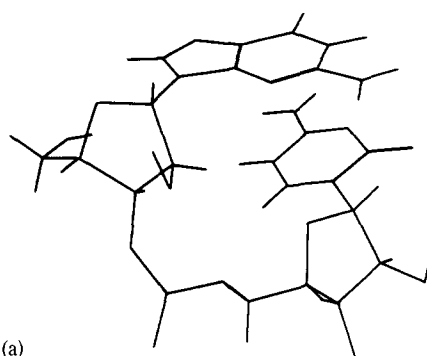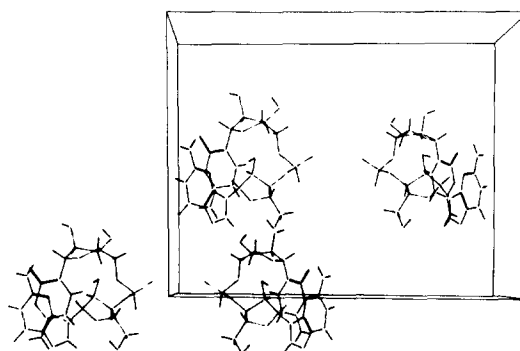

(a)
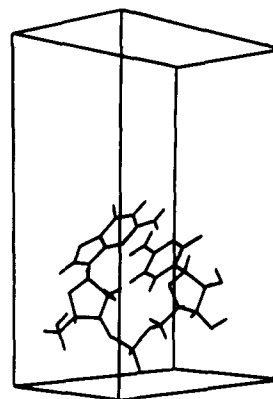
(b)

(c)

Fig. 2.

Table 2. Three packing schemes found for the molecule GpC and their conformations

Form	Fig. No.	Dihedral angles, degrees								Eulerian angles		Energy, kcal/mole	Discrepancy index, R
		χ'	ψ'	ϕ'	ω'	ω	ϕ	ψ	χ	Θ	Ξ		
A	2d	−1	62	200	296	271	194	57	21	223	280	−38·0	0·393
B	2f	−1	62	203	294	277	192	52	21	8	100	−38·0	0·590
C	2g	269	61	155	114	119	69	25	97	163	246	−15·7	0·627

(d)

XSI 259 00   PSI 54 00   PHI 187 00   OME 280 00
OME 259 00   PHI 207 00   PSI 62 00   XSI -2 00
C3 ENDO 108   ORIENT 221 0 780 0
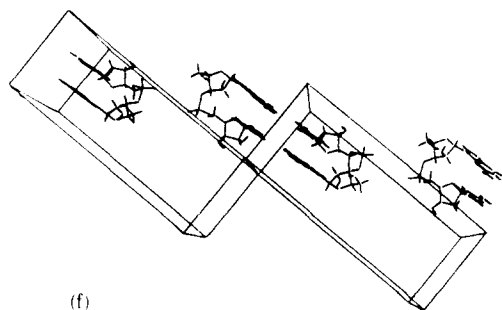

(g)

XSI 337 00   PSI 25 00   PHI 69 00   OME 119 00
OME 114 00   PHI 155 00   PSI 61 00   XSI 269 00
C3 ENDO 108   ORIENT 163 0 246 0

Fig. 2. (a) Lowest minimum energy conformation calculated for
an isolated GpC molecule. sugar pucker C3'-endo. E =
–39·0 kcal/mole. (b) Lowest energy packing scheme for GpC,
sugar pucker C3'-endo. denoted Form A. This trial structure was
successfully refined against our X-ray data. E = –38·0 kcal/mole.
The figure shows molecule 1 (see the table accompanying Switch
6) in the unit cell. (c) Same as (b). but showing entire contents of
unit cell. (d) Two contiguous unit cells viewed along crystallog-
raphic b axis. For clarity, only two of the four molecules in the
unit cell are shown. (e) Same view as (d). with all four molecules in
the unit cell shown. (f) Form B packing scheme for the lowest
energy single molecule conformation. C3'-endo. E =
–38·0 kcal/mole. This trial structure could not be refined. Same
view as (d). (g) Possible packing scheme C. sugar pucker
       C3'-endo. E = 15·7 kcal/mole. Same view as (d).

(e)

XSI 259 00   PSI 54 00   PHI 187 00   OME 280 00
OME 290 00   PHI 207 00   PSI 62 00   XSI -2 00
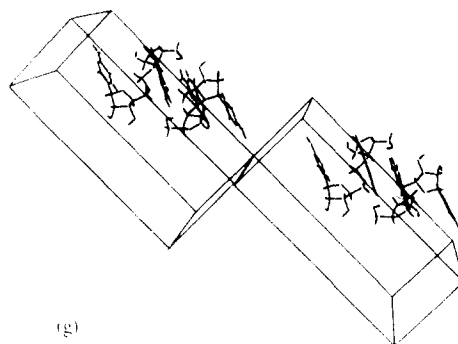C3 ENDO 108   ORIENT 221 0 280 0

molecule, and drawing it on the screen. the program then
enters the following loop:

```
1000 CONTINUE
     CALL UPKEEP
     IF (SWITCH(1)).....
     IF (SWITCH(2)).....
                .
                .
                .
     GO TO 1000
     END
```

The subroutine UPKEEP reads all the input devices for
the graphics (17 switches, X–Y joystick, 4 potentiometric
knobs, 3-D acoustical tablet) and stores their values. By
convention. if switches 1 and 11 are on simultaneously,
the overall image is controlled by the following table:

Device	Action
Knob 1	X-rotation
Knob 2	Y-rotation
Knob 3	Z-rotation
Knob 4	Z-translation (zoom)
Joystick	X–Y translation

(f)

XSI 260 00   PSI 55 00   PHI 191 00   OME 276 00
OME 293 00   PHI 203 00   PSI 62 00   XSI -1 00
C3 ENDO 108   ORIENT 7 0   180 0

Besides this overall control, and independently, the
switch readings are interpreted as OFF or ON (i.e. they
have values of .T. or .F. in the logical IF statements), and

program execution is controlled accordingly. For example, the interpretation of

IF (SWITCH(5)) CALL ADJCEL

is that is SW 5 is turned on, the program displays one of the 26 unit cells contiguous with the basic one always shown.

As long as no deformation of the molecule is requested, the coordinates of the original are kept as the data base and are never changed. Effective rotation, translation, and clipping are performed on this original by LDS-1 hardware.

Whenever deformation is requested, the PDP-10 must re-compute an entire set of coordinates from scratch, scale them, issue new drawing instructions, and pass them on to the graphics. This turns out to be very fast and scarcely interferes with the interactiveness at all.

A. Initialization of coordinates

The fastest way of generating a molecule is by a linked-atom matrix method[11], provided that the molecule can topologically be represented as a polymer with not too long side chains. Preliminary examination showed that this was possible for all the dinucleoside phosphates. Figure 3 shows the structure of GpC when represented this way.

zero, and are later varied through knobs or teletype interaction.

B. Display structure

The GIGL language allows creation of items, instances, and pictures within a logical hierarchy. The structure of the XpY program is shown in Fig. 4.

The meaning of the picture is that INSA is an instance of the item ITXT. ITXT is the textual display of names and values of angles, and INSA is an instance which maps

Fig. 4. Display structure of the program XpY.

that text onto the bottom of the screen. The meanings and uses of the other items are given in the next section. The
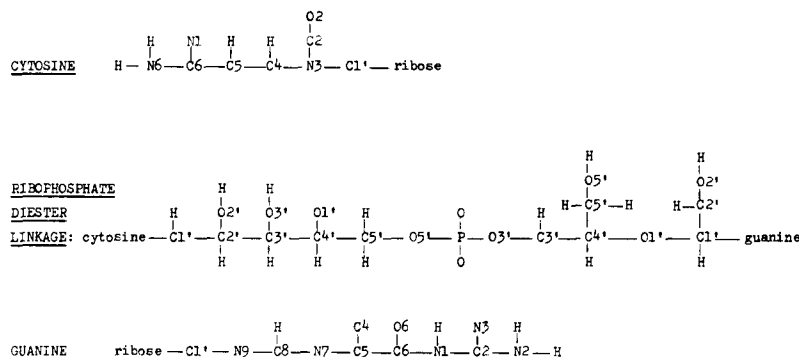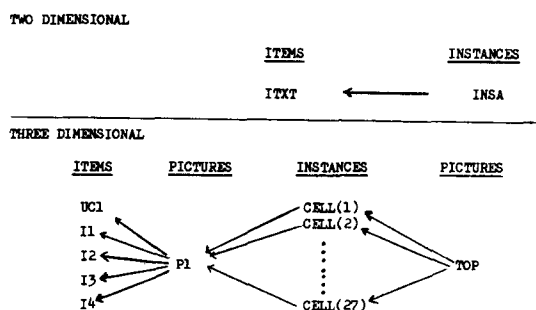

Fig. 3. Linear polymeric structure of GpC molecule.

The algorithm given by Scott and Scheraga requires three types of parameters: bond lengths, bond angles, and dihedral angles[12]. A linked-atom algorithm is simply a way of using these parameters to compute Cartesian coordinates.

Values of bond lengths and angles for the base ring structures and ribophosphate fragments were taken from Arnott[5], and held constant. Dihedral angles in the planar bases are either 0° or 180°, and were determined by inspection. Dihedral angles in the five-membered sugar rings are variable and depend on the state of sugar pucker. We represent the sugar pucker (i.e. the out-of-plane deformation of the five-member furanose ring) by T. Sato's pseudo-rotation coordinate system, so that the coordinates of all five atoms can be generated from a single parameter, Φ[13].

Eight parameters now remain unspecified, viz. the dihedral angles described in Fig. 1. These are initialized to

ultimate display instruction, which is executed once in the program, is

CALL DISPLAY ('INSA', 'TOP')

where the first argument is the highest order two-dimensional construct and the second is the highest in three dimensions. When deformations or other changes are desired, the display structure originally created is not disturbed, but the fresh drawing instructions are generated in dummy items and at the last minute are substituted into their appropriate logical places in 'INSA' or 'TOP'.

C. How the options work, switch by switch

1. IF (.NOT. SWITCH(1)) CALL PLOTD('TOP')

PLOTD is a subroutine which digitizes the screen image currently represented in the picture 'TOP', and draws it on-line on the CALCOMP plotter.

2. IF (SWITCH(2)) CALL BLANQ

.

.

```
SUBROUTINE BLANQ
CALL BLANK ('I2')
CALL BLANK ('I3')
CALL BLANK ('I4')
RETURN
END
```

The items I2, I3, and I4 contain symmetry-generated copies of the original molecule, I1. (See Switch 6) Setting Switch 2 on blanks out these images, leaving only the original. Extra code not shown unblanks them when the switch is turned off. They are initially invisible.

3. IF (SWITCH(3)) CALL QUERY

The computer retrieves the coordinates of three atoms entered by the user and calculates distances and an angle by conventional geometry.

4. IF (SWITCH(4)) CALL GETPUK

.

.

```
SUBROUTINE GETPUK
COMMON PHIPUK
CALL UPKEEP................... reads the knobs
PHIPUK = DEG(DKNOB(3))  takes the reading of
                        knob 3 and converts
                        it to an angle in
                        degrees
CALL DINUK..................... reconstructs a new
                        molecule with sugar
                        puckering corres-
                        ponding to the
                        pseudo-rotation
                        value given as
                        PHIPUK
RETURN
```

5. IF (SWITCH(5)) CALL ADJCEL

During initialization, twenty-seven instances of the unit cell and its contents were made, each differing from the others by a translation along each of the three crystallographic axes. The original image is 'CELL(14)', and the remainder are its 26 contiguous neighbors. The subroutine ADJCEL simply blanks all of these instances except for the one requested.

6. IF (SWITCH(6)) CALL QUIV

The subroutine QUIV calculates the coordinates of molecules 2, 3, and 4 in the unit cell, based on molecule 1 and the following symmetry scheme.*

Molecule	Fractional coordinates		
1	x	y	z
2	$-x$	y	$-z$
3	$\frac{1}{2}+x$	$\frac{1}{2}+y$	z
4	$\frac{1}{2}-x$	$\frac{1}{2}+y$	$-z$

*This corresponds to the space group C2. For further details concerning molecular symmetry, see [9].

†It can be seen that, as with least-squares residuals, the smaller R is, the more nearly correct the trial structure is.

Fractional coordinates are a way of representing the symmetry positions of molecules relative to the unit cell. They are more convenient than Cartesian coordinates, because the unit cell is not always coincident with an orthogonal coordinate system. In the case of GpC, the unit cell is monoclinic, meaning that its x, y, and z lengths are all different, and its x and z axes are inclined at an angle other than 90° to each other.

Once the Cartesian coordinates of molecules 2, 3, and 4 are computed, drawing instructions for molecule 2 are placed in item 'I2', 3 in 'I3', and 4 in 'I4'.

7. IF (SWITCH(7)) CALL ORIENT

The subroutine ORIENT reads two Eulerian angles from the teletype and uses them as global rotation angles for the molecule about x and z axes, relative to the fixed unit cell. A third Eulerian angle is unnecessary, because the X-ray data tells us the bases should be parallel to the y axis, thus removing one degree of freedom; in practice, the Cytosine base is generated parallel to the y axis.

8. IF (SWITCH(8)) CALL GETPM

Subroutine GETPM reads the knobs and translates each value to an angle in degrees. If switch 9 is simultaneously off, the order of knobs and angles is knob $1 = \chi$; knob $2 = \psi$; knob $3 = \phi$; knob $4 = \omega$. If switch 9 is on, the order is knob $1 = \omega$; knob $2 = \phi'$; knob $3 = \psi'$; knob $4 = \chi'$.

The subroutine then calls DINUK, which reconstructs the entire molecule based on these new dihedral angles, and redraws the molecule.

10. IF (SWITCH(10)) CALL PLANE

The computer requests three integers (Miller indices h, k, l) and draws the corresponding plane over the unit cell. It simply floats each integer, takes its reciprocal, and uses that number as the fractional coordinate in a drawing instruction.

12. IF (SWITCH(12)) CALL TYVAL

The computer requests an integer, n, and an angle, sets the nth dihedral angle to that value, leaving the rest alone, and reconstructs the molecule by calling DINUK.

14. IF (SWITCH(14)) CALL GETORT

This operates like switch 7, except that input is from the knobs instead of the teletype.

17. IF (SWITCH(17)) CALL CALCSF

The subroutine CALCSF is a modification of an existing program for computing structure factors from coordinates. The output lists the calculated and observed structure factors side by side for the strongest reflections, and computes the R-factor or discrepancy index:

$$R = \Sigma|(|F_o| - |F_c|)|/\Sigma|F_o|$$

where F_o is the observed structure factor and F_c is its calculated value.‡

9. CONCLUSIONS

Since the purpose of this meeting is to record advances in computer graphics, it is appropriate for me to outline what I think are our genuine accomplishments, and at the

same time draw attention to pitfalls and shortcomings which I have met.

The most important accomplishment I can name has been the desctruction of the prejudice held by many chemists against graphics as a research tool. Under the principle 'nothing succeeds like success,' I can safely predict a mushrooming interest in graphics in chemistry, encourage equally by diminishing costs of suitable hardware and by increasing support from traditional sources.

On a more technical level, it should be pointed out that the graphics did not automate the crystallographic process the way, say, an improved Fourier routine would have, but rather added in significant measure to the judgment capabilities of the chemists themselves. In other words, the graphics did not help to 'crank out' the answer, but rather extended the capability and creativity of the users so that we could make better judgments about how to apply our chemical intuition.

To go back a bit, even before the graphics was brought in, the problem was assessed chemically, and certain constraints were deduced from the data. These were then reduced to mathematical form and programmed into the graphics. For example, a technique called the Patterson method[2] located the position of the single phosphorus atom of GpC. One function of the subroutine ORIENT is therefore to translate the GpC molecule as a rigid body to place the phosphorus atom at that location. Another chemical consequence was the likelihood that the G and C bases of different molecules would be mutually oriented in one particular way (Watson-Crack base pairing)[4]. This was not programmed, but served as a visual check on trial solutions suggested by graphical manipulation.

I am not the least embarrassed to say that many promising avenues were not explored simply because we came across the correct answer so soon after beginning, and therefore were not motivated to go further. For example, we were prepared to write a lengthy subroutine to compute intermolecular energies, if visual assessment of entanglements was not good enough. To our relief (and delight) the visual quality was so good we scrapped the idea.

Of the graphics method itself, we learned that interaction is desirable and possible for quite complex molecules. The algorithms chosen for coordinate generation and manipulation proved to be efficient, and the execution of drawing instructions was fast. The only step which broke the interactivity was the Fourier transform; I do not know how to speed that up. I hasten to add that, because of the desirability of comparing calculated and observed structure factors side by side, this step gives the illusion of being I/O bound, so the time drawback is not serious.

An obvious criticism is that we have to recompute and redraw the entire molecule every time one of its dihedral angles is changed. Why not just rotate about that bond in hardware? The answer is that, while we technically could, there seems to be no easy way on the LDS-1 to recover the transformed coordinates needed for evaluation of R and for a permanent record.

The use of switch 9 to indicate by being on or off which four of the eight dihedral angles to attach to the knobs was unsatisfactory. It proved harder than we first thought to make the knobs remember where they were when switch 9 was reset. However, we like the idea of using one to multiply the number of others. The stereo and color options, though available, did not seem to enhance the image and were not often used.

In conclusion, the graphics does an outstanding job for structural problems in which simple geometric constraints introduce great difficulties, especially when analytical solutions become hopelessly unworkable. As I showed in a recent work on graphics and helical polymers[14], it provides necessary mathematical backup for situations where the user needs to rely most on intuition.

REFERENCES

1. L. Katz and C. Levinthal, Interactive computer graphics and representation of complex biological structures, *Ann. Rev. Biophys. Bioeng.* **1**, 465 (1972).
2. A. Lesk, Pictorial pattern recognition and the phase problem of X-ray crystallography, *Commun. ACM* **15**, 3 (1972).
3. P. K. Warme, R. W. Tuttle and H. A. Scheraga, Comparison of three-dimensional structures of macromolecules, *Comput. Prog. Biomed.* **2**, 248 (1972).
4. J. D. Watson, *Molecular Biology of the Gene.* Benjamin, New York, 2nd Ed (1970).
5. S. Arnott, J. Wonacott and S. D. Dover, 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 (1969).
6. A. Lesk, Generation of interactive displays from FORTRAN using the PDP-10/LDS-1 computer graphics system, *Software: Practice Experience* **2**, 259 (1972).
7. S. B. Broyde, S. D. Stellman, B. Hingerty and R. Langridge, Conformational stability in dinucleoside phosphate crystals, *Biopolymers* **13**, 1243 (1974).
8. S. D. Stellman, B. Hingerty, S. B. Broyde, E. Subramanian, T. Sato and R. Langridge, Conformation of guanosine-3',5'-cytidine monophosphate (GpC), *Macromolecules* **6**, 652 (1973).
9. S. D. Stellman, B. Hingerty, S. B. Broyde, E. Subramanian, T. Sato and R. Langridge, Structure of guanosine-3',5'-cytidine monophosphate. I. semi-empirical potential energy calculations and model building, *Biopolymers* **12**, 2731 (1973).
10. B. Hingerty, E. Subramanian, S. D. Stellman, S. B. Broyde, T. Sato and R. Langridge, Structure of guanosine-3',5'-cytidine monophosphate. II. structure of the calcium derivative in space group P2₁, *Biopolymers* (1974, in press).
11. S. D. Stellman, Computer simulation of polymer conformation and conformational thermodynamics, Ph.D. Thesis, New York University (1971).
12. R. A. Scott and H. A. Scheraga, Conformational analysis of macromolecules. II. The rotational isomeric states of the normal hydrocarbons, *J. chem. Phys.* **44**, 3054 (1966).
13. T. Sato, Private communication.
14. S. D. Stellman, Computer graphics in the solution of the Chain Deformation Problem, *Macromolecules* **7**, 296 (1974).

Documentation for XPY

Introduction

I. Program XPY is a generalized program for building and displaying models of dinucleoside 3′-5′ phosphates, where X and Y can be any of the bases A, G, C, or U.

The user can simultaneously and independently adjust the values of eight commonly referred-to-backbone dihedral angles, as well as the puckering of the sugar. The conformational energy of the molecule can be computed and typed out.

Special features for G$_p$C:

The unit cell dimensions, which are known from crystallographic analysis, can be used to generate equivalent molecules at symmetrically located positions with the cell. The molecules can then be oriented within the cell to maximize observed hydrogen bonding or minimize packing conflicts. The structure factors can be calculated and compared with the X-ray data.

Method

The parameters for the four bases, A, C, G, and U, are computed at the beginning of the program. The molecule is constructed with a Scott-Scheraga successive matrix transform scheme, working from the 5′ end to the 3′ end: base-sugar-phosphate-sugar-base.

New values of the dihedral angles are constantly input to provide fresh conformational coordinates, as recalculated under the matrix scheme.

For G$_p$C, the three equivalent molecules in the unit cell are calculated under the symmetry of the C2 space group.

To execute

1. Log in on the PDP-10.
2. RUN DSK XPY[30, 14].
3. Computer will type:

DINUCLEOSIDE PHOSPHATE XPY
X:

User must then type in desired 3′-linked base. E.g.

X: G (User typed in the latter G)

Computer types

Y:

User must type in desired 5′-linked base, such as C:

Y: C

(All user responses, of course, are followed by a carriage return)

AT START OF PROGRAM, SWITCHES 1 and 11 MUST BE TURNED ON.

4. *Interaction*

Image of the molecule now appears on the screen. (If G$_p$C, the unit cell will be outlined around the molecule).
The following switches are now interactive.

SWITCH 1: Turn off briefly to plot the image on the screen onto the Calcomp plotter.
SWITCH 2: Turn on to blank out the three equivalent molecules—which are generated by SWITCH 6.
SWITCH 3: Turn on briefly for QUERY.
The computer types "Q:" The user responds with a string of the form XIIXIIXII, where X = either M or S (for main or side chain), and II = a two-digit integer index number within the desired array (main or side).

A list of index numbers for the G$_p$C molecule is available from S. Stellman.
Computer then types out the bond distances between the first and second, and second and third, atoms, along with the included bond angle.
SWITCH 4: When turned on, knob 3 can change the sugar puckering, Φ, in increments of 9° from 0° to 360°. Each 9° increment corresponds to a different mode of puckering, whose name is displayed on the screen.

Φ	Puckering	Φ	Puckering
342°–9°	C1′–exo	198°–225°	O1′–exo
18°–45°	O1–endo	234°–261°	C4′–endo
54°–81°	C4′–exo	270°–297°	C3′–exo
90°–117°	C3′–endo	306°–333°	C2′–endo
126°–153°	C2′–exo	342°–9°	C1′–exo
162°–189°	C1′–endo		

SWITCH 5: Turn on briefly. Computer types out
CELL NO:
User then types in an integer (12) from 01 to 27, corresponding to one of 27 unit cells continuous to the displayed (which is itself No. 14).
SWITCH 6: Turn on briefly. Computes and displays three molecules in the unit cell equivalent to the first by C2 symmetry. This may be left on in conjunction with Switches 7 and 14.
SWITCH 7: Turn on briefly. Computer types out

ORIENT:

User types in two angles, Θ and Ξ (2F) format, as decimal numbers between 0 and 360 degrees, to orient molecule 1 within the unit cell. Orientation is constrained so that phosphorous atom lies at (3, 0, −1), and plane of cytosine base is parallel to the y-axis.
SWITCH 8: Leave on as desired. Attaches dihedral angles to knobs, and displays their values as text at the bottom of the screen. The user can thus deform the molecule as if it were a wire model. This option works in conjunction with SWITCH 9 as follows.

SWITCH 9 ON		SWITCH 9 OFF
Knob 1	ω'	χ
Knob 2	ϕ'	ψ
Knob 3	ψ'	ϕ
Knob 4	χ'	ω

SWITCH 10: Turn on briefly. Computer types:

H, K, L(3I):

Type in three integers separated by blanks or commas. The crystallographic plane corresponding to H, K, and L supplied will appear. To turn it off, hit switch 10, follow immediately with CR.
SWITCH 12: Turn on briefly. Computer types "TY:" User types an integer, followed by a space, followed by an angle (F10.0) from 0 to 360 degrees. One of the dihedral angles is then set to value just typed in as follows:

INTEGER	ANGLE	INTEGER	ANGLE
1	χ'	5	ω
2	ψ'	6	ϕ
3	ϕ'	7	ψ
4	ω'	8	χ

SWITCH 13: ON for stereo.
SWITCH 14: ON to attach orientation angles for whole

molecule (as with SWITCH 7) to knobs 1 and 2, and to display their values on the screen.

    Knob 1 Θ
    Knob 2 Ξ

SWITCH 17: Depress joystick button briefly. Computer will type the calculated and observed structure factors for the current model, followed by R (3Å data only).