PUCK: a real-time modification of sugar pucker on a PS300

Béatrice Amerein,* Raymond Ripp* and Philippe Dumas

Laboratoire de Cristallographie Biologique, Institut de Biologie Moléculaire et Cellulaire, 15, rue René Descartes, F67084 Strasbourg-Cedex, France

This paper describes a program for a real-time interactive modification of furanose sugar pucker. The resulting change of conformation is visualized instantaneously on the display screen of an Evans and Sutherland PS300, the picture system on which the program was implemented. The program is accessible as a new function, named PUCK, in the molecular modeling system FRODO. This new possibility fills the lack of degrees of freedom for modeling nucleic acids.

Keywords: nucleic acids, sugar pucker, crystallographic refinement, molecular modeling

Received 16 September 1987
Accepted 29 September 1987

Crystallographers who work in the macromolecular field always face the problem of building as precisely as possible the three-dimensional (3D) structure of a molecule from an electron density map. For proteins, torsion angles ϕ and ψ in each residue can reasonably be considered as sufficient degrees of freedom for allowing a correct fit of the backbone to the electron density map. For nucleic acids, the number of variables involved is more important (see Figure 1). Moreover, the furanose sugar conformation plays a key role in determining the whole conformation of the sugar-phosphate chain. Indeed, the pentose ring can undergo conformational changes through a continuum of conformations commonly referred to as the "pseudorotational pathway."[1,2]

Up to now, only two ideal conformations — namely C2′-endo and C3′-endo — were available in the commonly used nucleic acids conformation dictionary of J. Sussman in the modeling graphics software FRODO.[3] However, significant departures from these two conformations have often been described in high-resolution crystal structures, where precise determination of intracyclic torsion angles was possible (see References 4 and 5 for a review). For macromolecular structures, where resolution is generally less extended, the conformation of furanose sugars cannot be determined unambiguously. This is obviously the case for the tRNA structures[6-10] in the 3Å resolution range and, to some extent, for certain deoxyribonucleotide oligomers,[11,12] especially when restricted to a 2.5–3Å resolution range.

In fact, it is possible to draw more information from such structures than the crude separation between C2′- and C3′-endo. Small variations of sugar pucker can affect significantly the orientation of the base relative to the sugar. Experimental observations show that the orientation of a given base can be determined accurately from inspection of a 3Å resolution electron density map (see Figure 2).

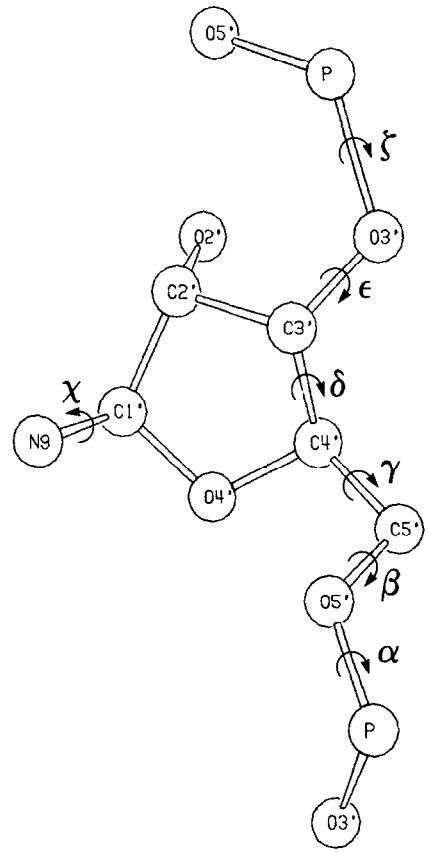

Figure 1. Illustration of the degrees of freedom involved in nucleic acids conformation (IUPAC nomenclature). This figure is an ORTEP drawing.[18]

This paper describes a new option that was implemented in the menu of the PS300 version[13] of FRODO[3] to gain more precision and more versatility in the model-building process. This option, named PUCK, is available in the same way as the other FRODO options, such as FBRT or TOR, and it allows the continuous variation of the pucker of any ribose or deoxyribose ring. It is thus possible to visualize in real time the effect of such a variation upon the attached base as well as upon either the 3′ or the 5′ chain bound to the mobile sugar (the

*Also affiliated with the Département d'Informatique de l'Université Louis Pasteur, 7, rue René Descartes, F67084 Strasbourg-Cedex, France

 0263-7855/87/040184-06 $03.00

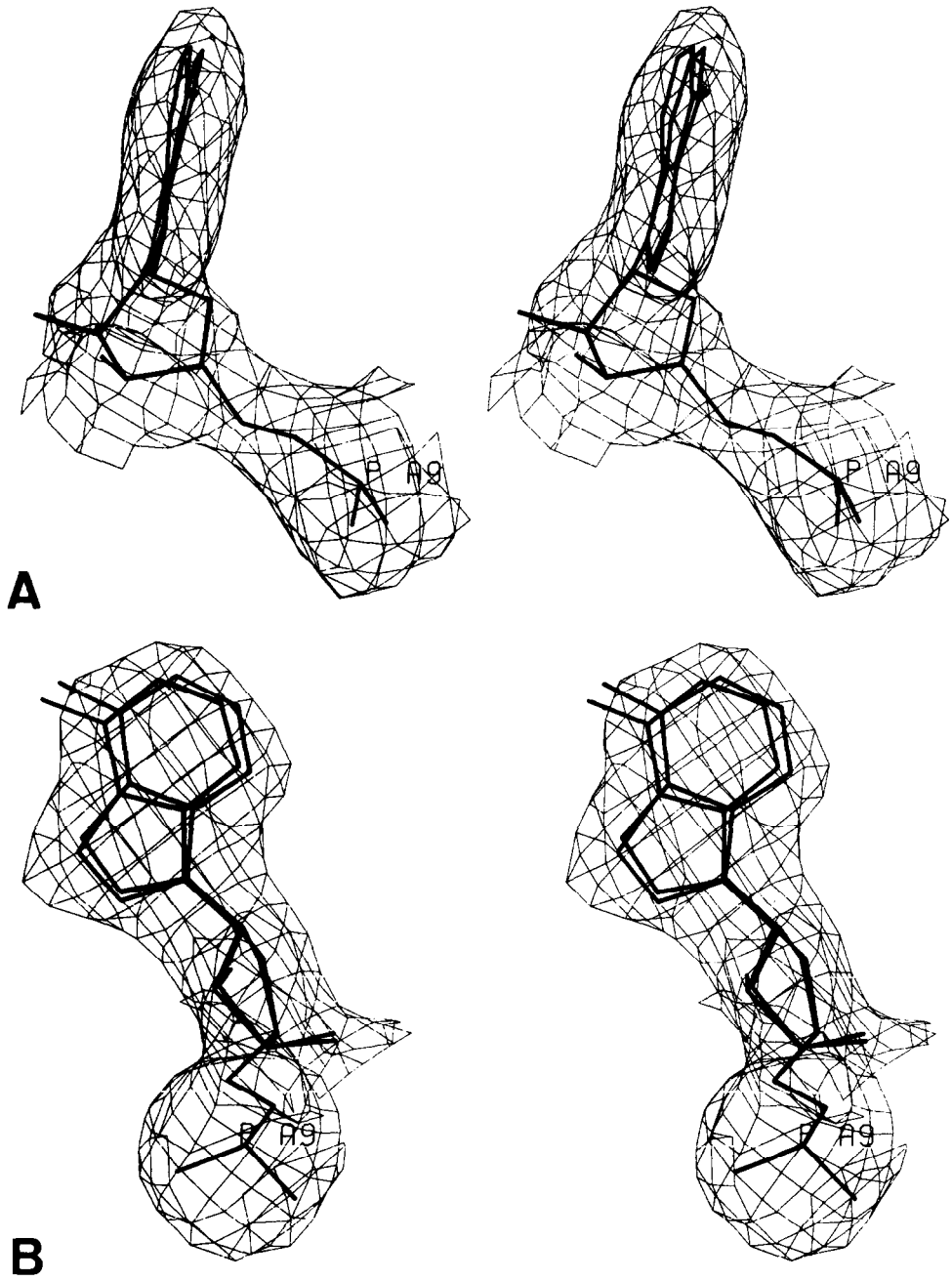

Figure 2. Illustration of accuracy for positioning a base into a 3Å resolution electron density map. This map fragment is the one around residue A9 in the tRNA^{Asp} structure.[9,10] *It corresponds to a well-defined portion of the molecule due to very low thermal agitation at this place. The two views correspond to different orientations of the same part. (A) Effect of a variation of* $\pm 2°$ in the torsion angle χ *defining the orientation of the base. It is clear that larger variation would give a visible lack of goodness of fit. (B) Effect of a variation from 159° to 168° in the phase of pseudorotation for this sugar with C2′-endo conformation. Larger variation would produce a visible lack of fit.*

user has to choose which side will keep fixed). After implementing this new possibility, all possible and reasonable degrees of freedom are available to the model builder in the nucleic acids field. Furthermore, the user has the option to use the refinement program NUCLSQ,[9,10] which makes it possible to restrain sugar conformations close to any desired target pucker.

PRINCIPLE OF THE METHOD

A database of finely sampled sugar conformations is built and stored in the PS300. When the user wants to modify one sugar, he or she selects it by use of the pen and the tablet by hitting onto either O3′ or C5′, which, at the same time, determines respectively the 3′ side or the 5′ side as the moving one (see Figure 3 for details). Then the program analyzes the sugar conformation and determines which is the corresponding conformation in the database. The original sugar structure is then replaced by the one of the database. Afterward, the user can select other conformations "continuously" from the database by rotating a dedicated dial. This way of solving the problem provides a satisfactory feeling of smooth distortion. If the user is satisfied by the new conformation, he can accept it or otherwise reject it (exactly as with the standard options MOVE, FBRT

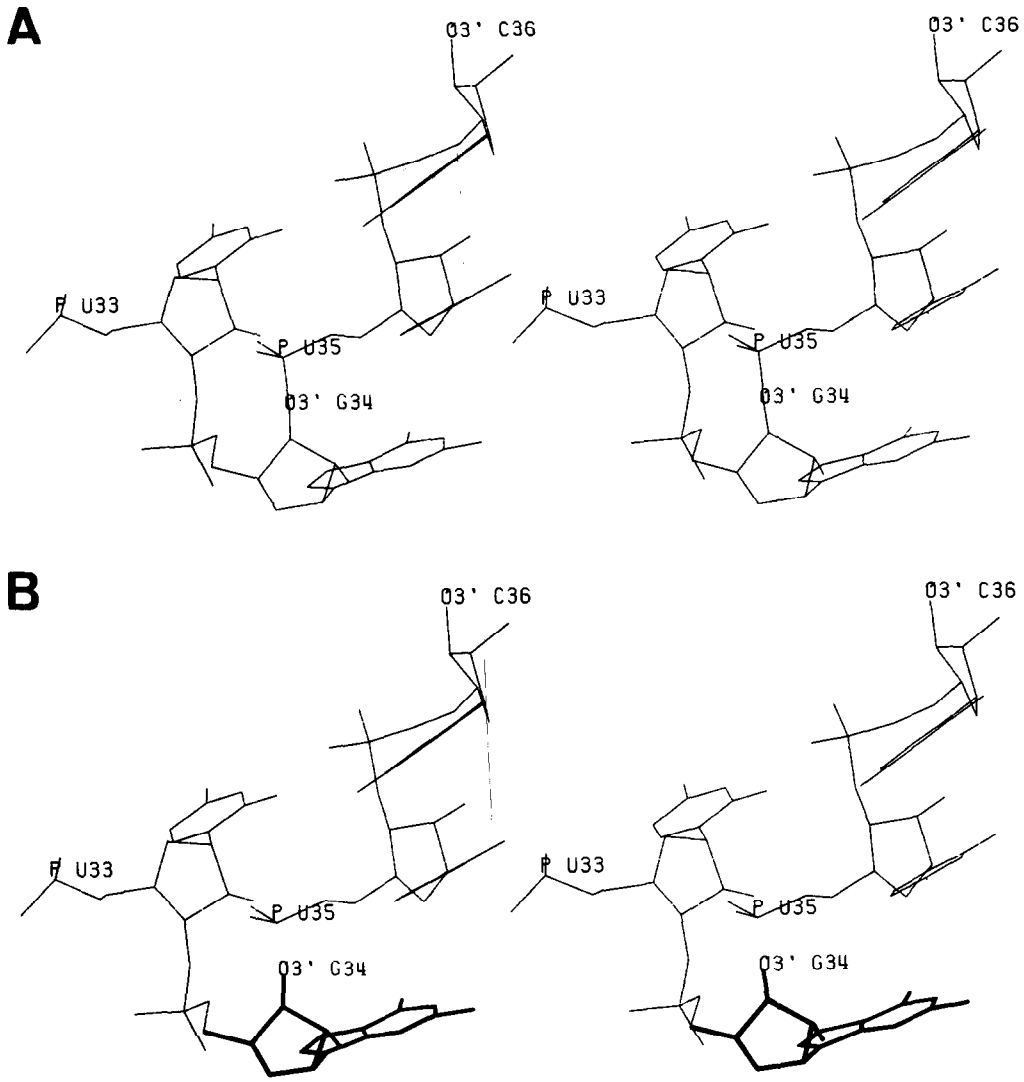

Figure 3. Illustration of the practical way to define both the fixed and the moving parts of the chain during pseudorotation. (a) Residues U33, G34, U35 and C36 of Yeast tRNA^{Asp} before using PUCK. (b) Here the atom O3' of residue G34 has been picked, indicating that G34 will be subject to pseudorotation and that the 3' side of the chain is the moving one during pseudorotation, while the 5' side has to be kept fixed. In the present example, to avoid any movement of the chain, the bond between O3'(G34) and P(U35) was cut previously. The zone that will actually be mobile is therefore reduced to the sugar and to its attached base, as represented in the thick line

or TOR, which are also likely to affect the set of coordinates). The next section describes how to generate the database.

BUILDING THE DATABASE

Only two parameters are necessary to allow a fair description of all possible sugar conformations, namely P and τ_m.[2] P, the phase of pseudorotation, describes the state of pucker in the pseudorotation wheel, and τ_m is an upper limit for the endocyclic torsion angles. The higher τ_m, the less planar the ring. Both P and τ_m can be derived from the intracyclic torsion angles. An exact method for obtaining them, in case of a regular pentagon, was described by Rao *et al.*,[14] as follows:

$$\tau_m^2 = A^2 + B^2 \quad (1)$$

$$tg(P) = B/A \quad (2)$$

$$\text{where } A = 2/5 \sum_{i=1}^{5} \tau_i \cos 2\alpha_i \text{ and } B = 2/5 \sum_{i=1}^{5} \tau_i \sin 2\alpha_i \quad (3)$$

$$\text{with } \alpha_i = 2/5\pi(i - 3) \quad (4)$$

$$\text{and } \tau_i = \tau_m \cos[P + 4/5\pi(i + 2)] \quad (5)$$

are known as τ_i. The five endocyclic torsion angles defined in Figure 4.

Conversely, users should be able to derive the geometry of sugars from any chosen amplitude τ_m and phase of pseudorotation P. A given geometry implies knowledge of the 15 nonindependent values for endocyclic bond angles θ_i, torsion angles τ_i, and bond lengths l_i (see Figure 4). Angles θ_i and bond lengths l_i can be calculated by the following equations, as suggested by Westhof and Sundaralingam[5] and Merritt and Sundaralingam:[15]

$$\theta_i = \theta_{i0} - (\alpha_i + \beta_i[\cos 2P - 2\pi(i - 1/5])\tau_m^2 \quad (6)$$

$$l_i = l_{i0} + b_i \tau_i(P, \tau_m) \quad (7)$$

where θ_{i0}, α_i, β_i, l_{i0} and b_i are empirically determined parameters. The criterion used by these authors for that purpose consists of building a sugar by using all 15

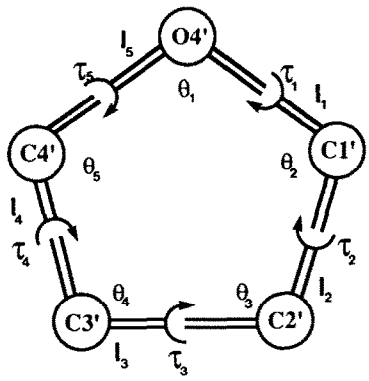

Figure 4. Definition for i = *1,5 of the bond lengths* l_i, *the endocyclic angles* θ_i *and the torsion angles* τ_i, *defining the geometry of the sugar ring*

predicted values, though they are not independent. Indeed, a sugar can be generated by putting one atom (e.g., O4′) at origin, then a bonded atom (e.g., C1′) along the x axis at a predicted distance, then C2′ in the plane xOy at a place respecting predicted bond length and bond angle and, finally, building C3′ and C4′ with respect of the concerned torsion angles, bond angles and bond lengths. In doing so, nine geometrical parameters only, out of 15, have been used. Such a procedure has the disadvantage of giving a specific status to the atom used as the first one and, consequently, to the six remaining values not used for the building. To overcome this difficulty, Merritt and Sundaralingam do not stop the process of building the cycle at atom C4′, but instead continue it and generate all atoms twice. The empirical parameters θ_{i0}, α_i, β_i, l_{i0} and b_i are thus determined to minimize the gaps between homologous atoms. Their values taken from Reference 15 are shown in Table 1. It should be noted that, *a priori*, other empirical sets of equations could have been used for generating the database — for example, the one derived by Pearlman and Kim.[16] Our choice has been driven by the fact that the reciprocal space refinement program NUCLSQ[9] makes use of results exposed in Reference 15, as well as in previous related reports[17] and, subsidarily by the fact that this method takes into account τ_m explicitly.

These empirical equations make it possible to geometrically analyze any sugar structure to obtain P and τ_m, and conversely to build a furanose sugar from these two values, that is to say, to build the necessary database. However, with this data handling, some complications occur. These are examined in the following section.

SELF-CONSISTENCY OF THE METHOD

Let us suppose, after using PUCK, that we accept from the database a conformation for a sugar, this conformation being defined by the doublet P and τ_m. It is conceivable that we could again want to modify its conformation later. The same sugar will thus be picked once more and reanalyzed to obtain the corresponding doublet P' and τ_m'. Clearly, we would like the output values P' and τ_m' to be the same as the input ones P and τ_m. However, this is not exactly the case, since the process of building a sugar and analyzing the resulting conformation is not perfectly cyclic. This fact is easily understandable: It is merely a measure of the inaccuracy of the method. Indeed, empirical parameters have been calculated to minimize the gaps between homologous atoms and not to make them equal to zero. This lack of self-consistency appears not only in the latter problem, but also in the way used for building a sugar. Even with a given set of empirical parameters θ_{i0}, α_i, β_i, l_{i0} and b_i, the exact structure depends on the atom arbitrarily chosen as the first one (the one at origin), as well as upon the sense used to describe the cycle. This problem is easily solved by always building a sugar in the same way. Concerning the lack of self-consistency itself, it is necessary to bring to the output phase an empirical correction. For this purpose, the function $P_i - P_o$, where P_i and P_o are respectively the input and output values of the phase in the building-analyzing process, has been tabulated for P_i varying from 0° to 360° by steps of 3°. A Fourier analysis of this periodical function in P_i allows an analytical correction of P_o to be made (see Figure 5). Since τ_m is practically considered as constant (see below), no correction has been made on this parameter.

Table 1. Values of empirical parameters l_{i0}, b_i, θ_{i0}, α_i and β_i used in equations 6 and 7 (from Reference 15)

i	l_{i0}	b_i	θ_{i0}	α_i	β_i
1	1.409	0.000027	133.14	0.003070	0.000966
2	1.535	0.000500	108.38	0.001446	0.001006
3	1.543	0.000450	105.53	0.002235	0.000868
4	1.543	0.000670	105.86	0.001999	0.000631
5	1.457	0.000470	107.16	0.001453	0.001019

PRACTICAL REALIZATION

Color Plates 1 and 2 illustrate the pseudorotation of

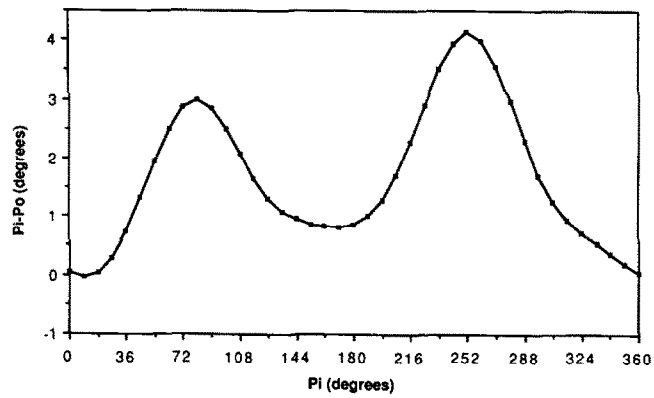

Figure 5. Variation of the empirical correction $P_i - P_o$ *to be made for the sake of self-consistency of the building-analysis process as described in the text. This periodical function in* P_i *has been Fourier-analyzed, which gives:*

$$P_i - P_o = \sum_k [A_k \cos(kP_i) + B_k \sin(kP_i)]$$

with $A_0 = 1.693$, $A_1 = -0.512$, $A_2 = -1.368$, $A_3 = 0.109$, $A_4 = 0.136$, $A_5 = -0.007$, $A_6 = -0.006$, $B_0 = 0.0$, $B_1 = -0.437$, $B_2 = 0.562$, $B_3 = -0.073$, $B_4 = -0.328$, $B_5 = -0.002$ *and* $B_6 = 0.011$. *The sum* $\sum_k (A_k^2 + B_k^2)$ *may be considered as a measurement of the error of the whole sugar constructing method. In this figure, the squares correspond to "experimental" points and the continuous line to Fourier approximation.*

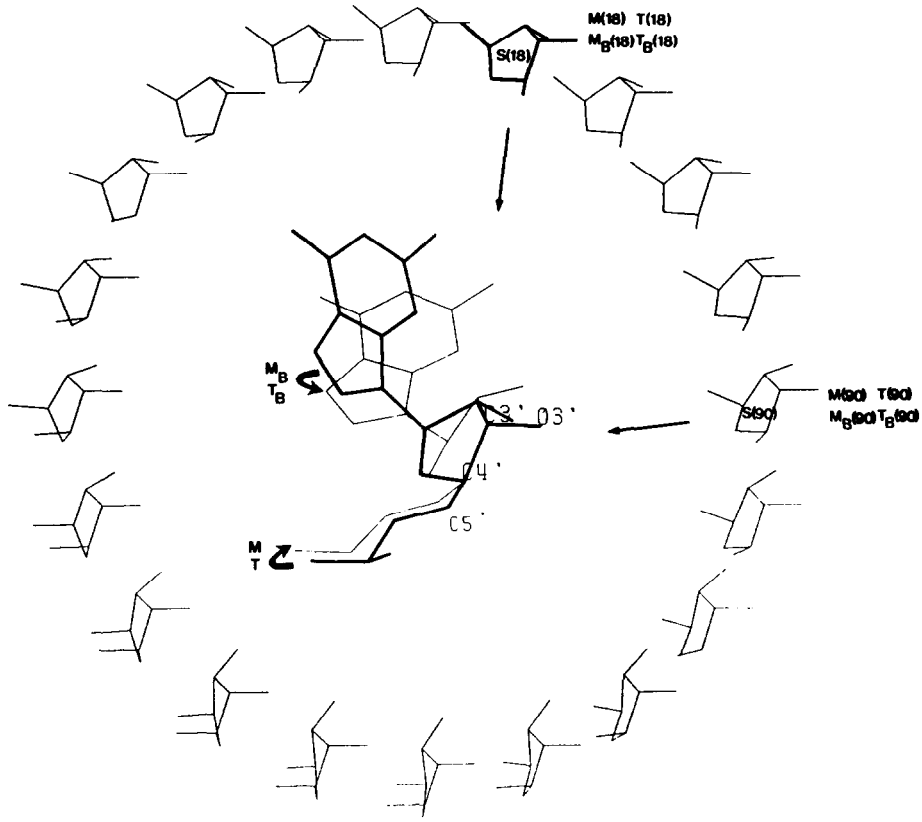

Figure 6. Illustration of the use of the database. Puckering is shown from phase 18° (in thick) to phase 90° (in thin). In this example, the atom C5′ has been picked. This shows selection from the database of the right sugar conformations S(18) and S(90). Matrix $\mathbf{M_B}$ *and vector* $\mathbf{T_B}$ *(respectively, M and T) to apply to the base (respectively, to the moving part of the chain) defining its correct orientation and position during pseudorotation are obtained from the matrices and vectors stored in the database*

as well as the effect obtained when going through the complete pseudorotational pathway.

The aim of PUCK is not only to realize and display the distortion of the sugar during its pseudorotation, but also to show simultaneously the resulting movement of the base and of the moving part of the chain. Each of them is considered as rigidly attached to the glycosidic bond and to either the 3′ or 5′ nucleotide link. Therefore, the database must contain the coordinates of all the different sugar conformations. Also, for each of them, the database must contain a set of matrices of rotation and vectors of translation to define the positions of the base and of the moving part of the chain *versus* an arbitrary fixed part of the sugar structure (see Figure 6). In fact, the whole data set, including rotation matrices, translation vectors and sugar conformations, is calculated *versus* two different sets of axes, each of them corresponding to one of the two possibilities of choosing the mobile side (and consequently the fixed one) of the chain during pseudorotation. More precisely, the set of axes corresponding to the picking of C5′ (O3′) is defined by O3′C3′C4′ (respectively, C5′C4′C3′) defining these three atoms for anchoring one side of the sugar during pseudorotation.

The whole database needs to be calculated once and downloaded to the picture system at the beginning of each work session, along with the PS300 specific graphics code for FRODO extended for PUCK. The addings consist essentially of a PS300 network through which, by rotating the dial dedicated to the phase of pseudorotation, the user generates the index determining which instantaneous sugar conformation should be displayed, as well as which rotations and translations are to be applied simultaneously to the base and to the moving part of the chain.

Practically, the amplitude of deformation τ_m is fixed to 38.7° (see "Restrictions on τ_m") while the phase P can vary. Three possibilities exist:

- Variation of phase from 0° to 180° with an increased precision of 1° around 18° and 162° corresponding, respectively, to C3′-endo and C2′-endo
- Variation of phase from 0° to 357° by 3° steps
- Variation of phase according to any other distribution at will of the user

Determining the matrices and vectors

The following explanation concerns the base, the method used for the moving part of the chain being similar. For each value of the phase P, the matrice $\mathbf{M_B(P)}$ and the vector $\mathbf{T_B(P)}$ contained in the database correspond to a pseudorotation from 0° to P. They are given by superposition of the three atoms O4′C1′X (X being N1 or N9 or C5 according to base type) at phase 0° and of the same atoms at phase P. (This supposes that the considered set of atoms is not subject to distortion during pseudorotation, which can be assumed for the exocyclic bond angle O4′C1′X and the bond length C1′X on the basis of known crystal structures.[4] On the contrary, the length of the vector corresponding to the bond O4′C1′ can change during pseudorotation and, there-

fore, must be normalized before doing the superposition.) Thus, for a pseudorotation from P_1 to P_2, the rotation and translation to be applied to the base are given by $[\mathbf{M_B(P_2)}]\,[\mathbf{M_B(P_1)}]^{-1}$ and $\mathbf{T_B(P_2)} - [\mathbf{M_B(P_2)}]\,[\mathbf{M_B(P_1)}]^{-1}\,\mathbf{T_B(P_1)}$ where $\mathbf{M_B(P_i)}$ and $\mathbf{T_B(P_i)}$ (i = 1 or 2) stand for a matrix of rotation and a vector of translation corresponding to the displacement of the base during pseudorotation from $P = 0°$ to $P = P_i$.

Constructing the sugar conformations

For the different values of phase and knowing τ_m, the parameters τ_i, θ_i and l_i defining the geometry of the sugar ring are calculated. As seen earlier, the exact structure of the sugar depends on the method used to calculate the coordinates from the set of values τ_i, θ_i and l_i. To be consistent with FRODO, we adopted the method used elsewhere in the program to build a sugar (and its ring substituents) from the main conformation dictionary. It consists of putting C5′ at origin, then generating C4′ along the x axis at a predicted distance, then C3′ in the plane xOy respecting the predicted bond length and bond angle, finally building O3′, C2′, O2′, O4′, C1′ and X (X being N9, or N1, or C5) with respect of torsion angles, bond angles and bond lengths. Some of these values are considered as fixed during pseudorotation and, therefore, set to standard values. Others do vary during pseudorotation and are given by τ_i, θ_i and l_i.

Restrictions on τ_m

Though theoretically possible, it is practically difficult to sample the sugar conformations in the two-dimensional (2D) space of the variables P, τ_m. To be efficient the sampling in τ_m should be, for example, from 15° to 45° in 5° steps. The resulting increase in the amount of data would be sevenfold. Presently, 10% of the two-megabyte PS300 memory is needed for holding the database corresponding to a single value in τ_m, and thus 70% of the PS300 memory would be necessary for the 2D database mentioned above. Added to 16% filled by the graphics code of FRODO itself extended for PUCK, only 14% would remain free for holding the user's molecular structure, which does not seem reasonable. For that reason, only a single value of τ_m has been used. It has been taken equal to 38.7°, which is an acceptable mean value for RNA.[16] τ_m variability is very much increased for B-DNA in comparison to RNA, which makes PUCK probably less useful in the former case. However, it is perfectly possible to use any other value for amplitude of pseudorotation.

ACKNOWLEDGEMENTS

We thank Dino Moras for the initiation of this work, as well as for interest and support, and J. C. Spehner for constant interest. We are also indebted to E. Westhof for many discussions about nucleic acid structures. We acknowledge J. W. Pflugrath, M. A. Saper and J. Sacs for making available to us their PS300 version of FRODO.

REFERENCES

1 Kilpatrick, J. E., Pitzer, K. S., and Pitzer, R. *J. Am. Chem. Soc.*, 1947, **64,** 2483–2488
2 Altona, C., and Sundaralingam, M. *J. Am. Chem. Soc.*, 1972, **94,** 8205–8212
3 Jones, T. A. *J. Applied Crystallography*, 1978, **11,** 268–272; Jones, T. A., in *Computational Crystallography*, Sayre, D. (ed), Clarendon Press, Oxford, 1982, 303–317; Jones, T. A. *Methods in Enzymology*, 1985, **115,** 157–171
4 deLeeuw, H. P. M., Haasnoot, C. A. G., and Altona, C. *Isr. J. Chem.*, 1980, **20,** 108–126
5 Westhof, E., and Sundaralingam, M. *J. Am. Chem. Soc.*, 1980, **102,** 1493–1500
6 Sussman, J. L. *et al. J. Mol. Biol.*, 1978, **123,** 607–630
7 Holbrook, S. R. *et al. J. Mol. Biol.*, 1978, **123,** 631–660
8 Hingerty, B., Brown, R. S., and Jack, A. *J. Mol. Biol.*, 1978, **124,** 523–534
9 Westhof, E., Dumas, P., and Moras, D. *J. Mol. Biol.*, 1985, **184,** 119–145
10 Dumas, P., Westhof, E., and Moras, D. (submitted)
11 Fratini, A. V. *et al. J. Biol. Chem.*, 1982, **257,** 14686–14707
12 Brennan, R. G., Westhof, E., and Sundaralingam, M. *J. Biomolecular Structure and Dynamics*, 1986, **3,** 649–665
13 Pflugrath, J. W., Saper, M. A., and Quiocho, F. A. *J. Mol. Graphics*, 1983, **1,** 53–54
14 Rao, S. T., Westhof, E., and Sundaralingam, M. *Acta Crystallogr.*, 1981, **A37,** 421–425
15 Merritt, E. A., and Sundaralingam, M. *J. Biomolecular Structure and Dynamics*, 1985, **3,** 559–578
16 Pearlman, A., and Kim, S. H. *J. Biomolecular Structure and Dynamics*, 1985, **3,** 85–98
17 Westhof, E., and Sundaralingam, M., in *Structure and Dynamics: Nucleic Acids and Proteins*, Clementi, E., and Sarma, R. H. (eds), Adenine Press, New York, 1983
18 Johnson, C. K. ORTEP: a Fortran thermal-ellipsoid plot program for crystal structure illustrations. Report ORNL-3794, 1970, Oak Ridge National Laboratory, TN, USA