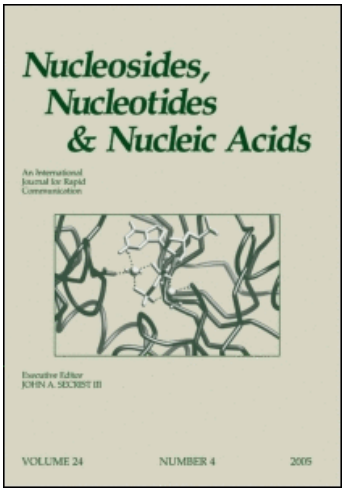

Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

A Procedure for Analysing Nucleic Acid Fragment Geometry and Stacking Interactions

C. C. Wilson[a]
[a] Neutron Science Division, Rutherford Appleton Laboratory, Oxon, UK

To cite this Article Wilson, C. C.(1990) 'A Procedure for Analysing Nucleic Acid Fragment Geometry and Stacking Interactions', Nucleosides, Nucleotides and Nucleic Acids, 9: 2, 163 — 171
To link to this Article: DOI: 10.1080/07328319008045128
URL: http://dx.doi.org/10.1080/07328319008045128

NUCLEOSIDES & NUCLEOTIDES, 9(2), 163-171 (1990)

A PROCEDURE FOR ANALYSING NUCLEIC ACID FRAGMENT GEOMETRY AND STACKING INTERACTIONS

C C Wilson

Neutron Science Division, Rutherford Appleton Laboratory,
Chilton, Didcot, Oxon OX11 0QX, UK.

Abstract

A procedure enabling the rapid location of base-pairs in a nucleic acid fragment and subsequent analysis of base-pair geometry is presented. The programs of Dickerson and Rosenberg for nucleic acid conformation are used in the procedure. In addition to giving information on hydrogen bonding, propeller twisting and other geometric and conformational quantities, an algorithm is presented for quantification of the stacking interactions between planes. The geometrical and stacking calculations are performed in the program PHELIX.

Introduction

There is an enormous amount of data available on the structures of nucleic acid fragments, a considerable proportion of it held in various central databases. In addition, there are many crystal structures of such materials being elucidated on a regular basis. Given this amount of data being generated, it is important that general procedures are available for analysing, comparing and drawing general inferences from the results obtained.

In the field of analysis of the geometry of short chain oligonucleotides, the programs of Dickerson[1] (ROLL, CYLIN, TORAN) and Rosenberg[2] (HELIX) at the University of California at Los Angeles (UCLA) are extensively adopted as standards. However, it is only recently that the geometry of smaller nucleic acid fragment base-pairs has been widely studied[3,4] and an attempt made to standardise such calculations. In addition, there is no simple method available to quantify the degree of base/base stacking in a nucleic acid fragment

structure, or base/drug stacking in an intercalative compound. The definition of a "simple" calculation here is taken as meaning one which involves merely structural geometry calculations, and not of the electronic nature of the materials involved. The procedure to be described performs both geometrical and stacking calculations. The PHELIX program is written in FORTRAN-77 and has been run on the Neutron Division VAX 8650 at RAL.

Calculations of base-pair geometry

The main purpose of the first section of the PHELIX program is to take data from the Cambridge Structural Database, to locate and define the base-pairs, if any, in the structure and to present the data to the Rosenberg/Dickerson UCLA programs in the appropriate format. With coordinate and space group data extracted from the database using CSSR[7] (Figure 1), this task is performed in the following stages:

(i) PHELIX searches for base-pairs within the structure. The criterion used as default at present is the presence of two hydrogen bonds (not involving C-H groups) of less than 3.3Å (donor-acceptor separation), but this criterion can be varied as appropriate. PHELIX works for all crystallographic space groups and provides a quantification of the hydrogen bonding present in the structure (Figure 2);

(ii) The data from the base-pair are manipulated to create a set of coordinates suitable for analysis in HELIX and ROLL. At this stage the appropriate files for input to these latter programs are created;

(iii) the standard helical analysis is performed using HELIX and ROLL, to reveal propeller twist, buckle and C1'-C1' separation. These parameters, defined below, are important in determining base-pair geometry and the possible relations between this and molecular properties in both mono and oligonucleotides.

The geometric parameters determined in this procedure are the following :

Hydrogen bonding parameters - these consist of number of hydrogen bonds, type of base-pairing e.g. Watson-Crick, Hoogsteen etc., hydrogen bond distances and angles. The strength of a hydrogen bond can be roughly approximated in terms of its length, the shorter the donor-acceptor separation, the stronger the bond;

Propeller twist - the angle between the planes of two bases in a base-pair when viewed along the long axis joining them, defined as positive when the rear base must be rotated clockwise with respect to the front base in order to flatten the twist;

Buckle - the dihedral angle between the two base planes along their short axis once propeller twist has been flattened to zero;

C1'-C1' separation - self-explanatory, but this parameter is crucial in the formation of the helical backbone in oligonucleotides.

The procedure as set up is virtually automatic and is rapid (~ a few CPU seconds) for most cases. As an example of what can be achieved using this procedure, over 400 nucleoside structures were analysed for an earlier paper on base-pair geometry[3] in a fairly short period of time. The object of the procedure is to further standardise the classification and quantification of base-pair geometry in known structures. It is desirable that such a procedure be adopted by those studying these structures at present and thus that the relevant data can be presented in a standard and reproducible form.

Stacking interactions

The field of base/base stacking is considerably less well studied than that of in-plane geometry, and the same is true to an even greater extent for base/planar-drug stacking interactions. There have been comprehensive studies of base stacking geometry[8] but these have tended to be rather qualitative. At the other extreme there have been quantum mechanical calculations of these interactions involving complicated empirical or semi-empirical procedures. In order to provide a framework for the simple elucidation of stacking

```
REFERENCE STRUCTURE = 49309   A,B,C =   9.796  18.758   8.829
  ALPHA,BETA,GAMMA =  90.000 100.720  90.000    SPGR =  4 P21
49309 CABSUT01 3',5'-DI-O-ACETYLTHYMIDINE
    BASE 1
  1 N11     0.24680   0.61020   0.41050    2   9  10
  2 C21     0.31820   0.63920   0.30440    1   3   4
  3 O21     0.26990   0.64190   0.16720    2
  4 N31     0.44600   0.66480   0.36860    2   5  47
  5 C41     0.50990   0.66430   0.51870    4   6   7
  6 O41     0.62230   0.69250   0.56210    5
  7 C51     0.43360   0.62790   0.62130    5   8   9
  8 C71     0.49920   0.61890   0.78490    7  48  49  50
  9 C61     0.30670   0.60490   0.56130    1  51   7
    SUGAR 1
 10 C1D1'    0.09670   0.59580   0.36430    1  11  23  52
    BASE 2
 24 N12     0.98380   0.82120   0.33660   25  32  33
 25 C22     0.89970   0.80330   0.44040   24  26  27
 26 O22     0.93150   0.81490   0.57660   25
 27 N32     0.77910   0.77100   0.37460   25  28  65
 28 C42     0.73210   0.75360   0.22060   27  29  30
 29 O42     0.61970   0.72470   0.17870   28
 30 C52     0.82710   0.77310   0.11900   28  31  32
 31 C72     0.78710   0.75340  -0.05170   30  66  67  68
 32 C62     0.94610   0.80430   0.18290   24  69  30
    SUGAR 2
 33 C1D2'    1.12220   0.85050   0.39590   24  34  46  70
    BASE 1
 47 H31     0.47510   0.68700   0.30810    4
 48 H721    0.56820   0.66290   0.82240    8
 49 H731    0.55710   0.56950   0.80050    8
 50 H711    0.42790   0.61780   0.84030    8
 51 H61     0.24700   0.58020   0.63820    9
    BASE 2
 65 H32     0.70860   0.75860   0.42110   27
 66 H712    0.70280   0.78950  -0.09360   31
 67 H722    0.74850   0.69970  -0.07280   31
 68 H732    0.87100   0.76190  -0.11540   31
 69 H62     1.02190   0.82260   0.14090   32
```

FIGURE 1 - PHELIX input file, adapted from CSSR coordinate file, for reference 5.

(a)

PHELIX part 1 - Base-pair geometry

Hydrogen bonding present in the base-pairing scheme in

49309 CABSUT01 3',5'-DI-O-ACETYLTHYMIDINE

Symmetry elements in the form (xm*X+xs, ym*Y+ys, zm*Z+zs)

Base-pair No. 1

Donor	Acceptor	D-A	D-H	H...A	D..H..A	xm	xs	ym	ys	zm	zs
N31	042	2.834	0.772	2.101	158.50	1.	0.00	1.	0.00	1.	0.00
041	N32	2.866	0.897	2.048	151.05	1.	0.00	1.	0.00	1.	0.00

The files for HELIX and ROLL have been prepared.

(b)

PHELIX part 1 - Base-pair geometry

Hydrogen bonding present in the base-pairing scheme in

32030 ARFUAE10 9-ALPHA-D-ARABINOFURANOSYL-ADENINE

Symmetry elements in the form (xm*X+xs, ym*Y+ys, zm*Z+zs)

Base-pair No. 1

Donor	Acceptor	D-A	D-H	H...A	D..H..A	xm	xs	ym	ys	zm	zs
N11	N61	2.858	1.185	1.855	139.00	-1.	-0.5	-1.	0.0	1.	-0.5
N61	N71	3.278	1.072	2.261	157.62	-1.	-0.5	-1.	0.0	1.	-0.5

Base-pair No. 2

Donor	Acceptor	D-A	D-H	H...A	D..H..A	xm	xs	ym	ys	zm	zs
N62	N12	2.937	1.138	1.989	138.19	-1.	-0.5	-1.	-1.0	1.	-0.5
N72	N62	3.012				-1.	-0.5	-1.	-1.0	1.	-0.5

The files for HELIX and ROLL have been prepared.

FIGURE 2 - Part 1 of PHELIX, defining the base-pair geometry in two materials (References (a) 5, (b) 6).

PHELIX part 2 - Plane-plane stacking

(BASE) STACKING analysis on

10289 GUPCYT20 SODIUM GUANYLYL-3',5'-CYTIDINE NONAHYDRATE

OVERLAP (So) parameters > 0.9

Contact No. 1 - Symmetry element represented by
-1.00 1.00 -1.00 0.00 0.00 0.00 1 0 0

Planes BASE 1 and BASE 1 give the overlap parameters
So(i) = 0.94127, So(ii) = 0.93554

OVERLAP (So) parameters > 0.8

Contact No. 2 - Symmetry element represented by
1.00 1.00 1.00 0.00 0.00 0.00 0 0 0

Planes BASE 2 and BASE 1 give the overlap parameters
So(i) = 0.82522, So(ii) = 0.82290

FIGURE 3(a) - Part 2 of PHELIX, evaluating base-base stacking (Reference 11).

interactions, one must devise a simple algorithm based on structural geometry rather than electronic structure. Several simple closely-related algorithms have been devised to calculate the stacking S_o between two planes and they are detailed below.

(i) $$S_o = (3.4/d_{min}) \left[1 - \sum_i (d_i^{ab} - d_{min})/n_i\right] ,$$

where the summation is over the n_i atoms i in the plane containing fewer atoms, d_i^{ab} is the shortest distance between an atom i in plane a and an atom in plane b and d_{min} is the minimum contact of any atom in plane a with one in plane b.

(ii) $$S_o = (3.4/d_{min})^2 \left[1 - \left(\sum_i (d_i^{ab} - d_{min})/n_i\right)^2 \cos\theta_{ab}\right.$$

```
PHELIX part 2 - Plane-plane stacking

(DRUG) STACKING analysis on

 28583 EICGUA  ELLIPTICINE-5-IODOCYTIDYLYL-(3',5')-GUANOSINE ...

OVERLAP (So) parameters > 0.9

Contact No.  1 - Symmetry element represented by
                    1.00  1.00  1.00  0.00  0.00  0.00  0  0  0

                Planes BASE  1 and DRUG  1 give the overlap parameters
                    So(i) = 1.04263, So(ii) = 1.11540

Contact No.  2 - Symmetry element represented by
                    1.00  1.00  1.00  0.00  0.00  0.00  1  0  0

                Planes BASE  1 and DRUG  2 give the overlap parameters
                    So(i) = 1.06966, So(ii) = 1.18720

Contact No.  3 - Symmetry element represented by
                    1.00  1.00  1.00  0.00  0.00  0.00  0  0  0

                Planes BASE  2 and DRUG  1 give the overlap parameters
                    So(i) = 0.98385, So(ii) = 0.98937

Contact No.  4 - Symmetry element represented by
                    1.00  1.00  1.00  0.00  0.00  0.00  0  0  0

                Planes BASE  2 and DRUG  2 give the overlap parameters
                    So(i) = 1.04070, So(ii) = 1.10310

Contact No.  5 - Symmetry element represented by
                    1.00  1.00  1.00  0.00  0.00  0.00  0  0  0

                Planes BASE  3 and DRUG  1 give the overlap parameters
                    So(i) = 0.98094, So(ii) = 0.98782

Contact No.  6 - Symmetry element represented by
                    1.00  1.00  1.00  0.00  0.00  0.00  0  0  0

                Planes BASE  3 and DRUG  2 give the overlap parameters
                    So(i) = 0.94607, So(ii) = 0.90818

Contact No.  7 - Symmetry element represented by
                    1.00  1.00  1.00  0.00  0.00  0.00  0  0  0

                Planes BASE  4 and DRUG  1 give the overlap parameters
                    So(i) = 1.03180, So(ii) = 1.17101

Contact No.  8 - Symmetry element represented by
                    1.00  1.00  1.00  0.00  0.00  0.00  1  0  0

                Planes BASE  4 and DRUG  2 give the overlap parameters
                    So(i) = 1.05012, So(ii) = 1.12169
```

FIGURE 3(b) - Part 2 of PHELIX, evaluating base-drug stacking (Reference 12).

where $\cos\theta_{ab}$ is the angle between the best planes of a and b.

(iii) $S_o = (3.4/d_{min})^2 [1 - (\sum_i (d_i^{ab'} - d_{min})/n_i)^2 \cos\theta_{ab}$

where $d_i^{ab'}$ are the set of distances between atoms in plane a and those in plane b, set up on a one-to-one mapping, which minimise the sum $(d_i^{ab'} - d_{min})$ and hence maximise the stacking parameter.

As can be seen these expressions involve progressively more complex calculations as one moves from (i) to (ii) to (iii). They are all normalised to give a stacking parameter S_o equal to unity for two identical planes (e.g. bases) stacked exactly parallel at 3.4Å separation. There are two practical notes which have emerged from calculations on base/base[9] and base/drug[10] stacking interactions. Firstly, significant stacking is obviously represented by a value of S_o close to 1, but it should be noted that stacking values of less than ~0.75 represent at most weak planar interactions. Secondly, there is very little difference in practice between S_o values calculated using any of the above three expressions. The simplest expressions (i) and (ii) have therefore been used most widely to date.

The results obtained from the PHELIX program using these algorithms to describe the stacking of planar groups (Figure 3 (a) and (b)) represent an initial quantification of this important parameter based purely on structural arguments. There is therefore a very straightforward calculation for evaluating the often imprecise "degree of stacking" and it is hoped that this may be of use in the general assessment of planar group stacking.

Acknowledgements

The author would like to thank R E Dickerson for provision of the UCLA helical analysis programs.

REFERENCES

1. Dickerson R.E. (undated). ROLL, CYLIN, TORAN programs. University of California.

2. Rosenberg J.M. (undated). HELIX program. University of California.
3. Wilson C.C. (1987) **Nucl. Acids Res.**, 15, 8577-8591.
4. Wilson C.C. (1988a) **Nucl. Acids Res.**, 16, 385-393.
5. Wilson C.C., Low J.N., Tollin P. and Wilson H.R. (1984) **Acta Cryst.**, C40, 1712-1715.
6. Cline S.J. and Hodgson D.J. (1980) **Biochim. Biophys. Acta**, 20, 610.
7. Elder M., Hull S.E, Machin P.A. and Mills O.S. (1981) Crystal Structure Search Retrieval. A program for retrieving information from the Cambridge Crystallographic Data Centre databank. SERC Daresbury Laboratory, Warrington, UK.
8. Bugg C.E., Thomas J.M., Sundaralingam M. and Rao S.T. (1971) **Biopolymers** 10, 175-219.
9. Wilson C.C. (1988b) **Nucl. Acids Res.**, 16, 4751-4759.
10. Wilson C.C. (1988c) **Nucl. Acids Res.**, 16, 5229-5240.
11. Rosenberg J.M, Seeman N.C., Day R.O and Rich A. (1976) **J. Mol. Biol.**, 104, 145.
12. Jain S.C., Bhandary K.K. and Sobell H.M. (1979) **J. Mol. Biol.**, 135, 813.

Received January 22, 1989.