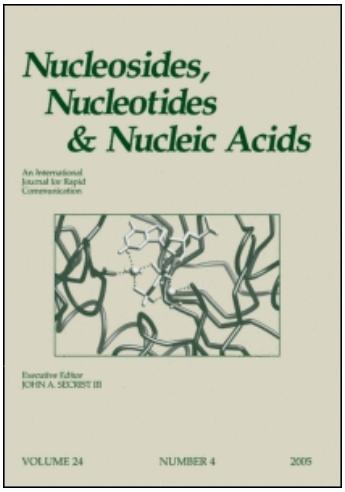

Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

X-Ray Studies of DNA Aiming at Elucidating the Sequence-Structure Code

Udo Heinemann[a]
[a] Claudia Alings and Hanspeter Lauble Abteilung Saenger, Institut für Kristallographie, Freie Universiät Berlin, Berlin 33, FRG

To cite this Article Heinemann, Udo(1990) 'X-Ray Studies of DNA Aiming at Elucidating the Sequence-Structure Code', Nucleosides, Nucleotides and Nucleic Acids, 9: 3, 349 — 354
To link to this Article: DOI: 10.1080/07328319008045146
URL: http://dx.doi.org/10.1080/07328319008045146

NUCLEOSIDES & NUCLEOTIDES, 9(3), 349-354 (1990)

X-RAY STUDIES OF DNA AIMING AT ELUCIDATING THE SEQUENCE-STRUCTURE CODE

Udo Heinemann*, Claudia Alings and Hanspeter Lauble
Abteilung Saenger, Institut für Kristallographie, Freie Universität
Berlin, Takustraße 6, D-1000 Berlin 33, FRG.

Abstract. A survey of DNA fragments crystallized as A- or B-form double helices reveals that, while a simple code relating dinucleotide code words to DNA structure does not exist, certain predictions concerning helicoidal parameters can be made based on sequence information.

Although crystallographic studies of synthetic DNA fragments have been carried out over the entire last decade, the field is still in its beginning stages. This becomes especially evident when one compares the several hundred known crystal structures of proteins with the small number of high-resolution X-ray studies of oligonucleotides: there are about 10 reports each of structure determinations of DNA fragments in the right-handed A- and B-forms. In addition to unmodified, purely Watson-Crick base paired DNA, a fairly large number of DNA molecules with mismatched or unpaired bases, chemical modifications and bound drug or protein molecules have been investigated. Here, we are not concerned with these perturbed DNA helices. Instead, we would like to ask whether the limited data base on "normal" DNA helix structure allows predictions relating base sequence to conformational features of the DNA to be made.

Given the rather limited success usually met with predicting protein conformation, and considering the much smaller amount of structure information available for DNA helices, it might be deemed doubtful whether meaningful rules relating sequence to conformation could possibly be derived. On the other hand, with nucleic acids the task of defining a sequence-structure code is easier since all structures have to fall within the framework of the known global DNA helix structures (A, B, Z) and there are only 4 structural units, the bases, as opposed to the 20 amino acids constituting proteins.

An evaluation of sequence-determined DNA structure must focus on the geometry of base pair stacking in the double helix. Disregarding distorsions of individual base pairs, the problem is reduced to describing the transformation in space required to superimpose a base pair onto its next neighbour. According to a generally accepted nomenclature[1] the three rotations involved are named *twist* (about an axis perpendicular to the base pairs, the local helix axis), *roll* (about the common long axis of the base pair step) and *tilt* (about the common short axis, the pseudo-dyad axis). The corresponding translations along these axes are *rise*, *slide* and *shift*. Here we shall restrict ourselves to analysing *twist*, *roll* and *slide*, those parameters that vary most in DNA helices.

The crystallographic data base at our disposal consists of A- and B-DNA double helices investigated in our laboratory[2-4] and elsewhere for which structural parameters have been reported[5]. The size of this data base determines the nature of code words to be used in developing a sequence-structure code. The simplest code possible would just consist of purine/pyrimidine (Y/R) steps. Predictions of DNA conformation based on that type of code and a considerably smaller data base[6] have failed to stand the test of recently determined DNA structures. Thus, a more complicated code based on base pair identity and/or longer code words must be employed.

To determine the length of code words that can be applied to test the data base we have to consider the number of individual double stranded dimers, trimers and longer oligomers that can be built up with the structural units present. For odd-numbered oligomers the number of possible sequences is $n_c=N^L/2$ where L is the length of the fragment and N is the number of building blocks (4 considering bases A,C,G,T or 2 considering Y,R). For even-numbered oligomers we find $n_c=(N^L+N^{L/2})/2$. Our data base contains 45 base pair dimers (dinucleotide steps) in A-DNA and 57 in B-DNA. Since there are $n_c=10$ theoretically possible base pair dimers, on average each is observed about 5 times. We would obtain ratios of 36/32 and 50/32 between observed and possible trimers. Therefore, trimer or longer code words are useless for an analysis of the presently available data base, as they do not occur with sufficient frequency to yield statistically meaningful results.

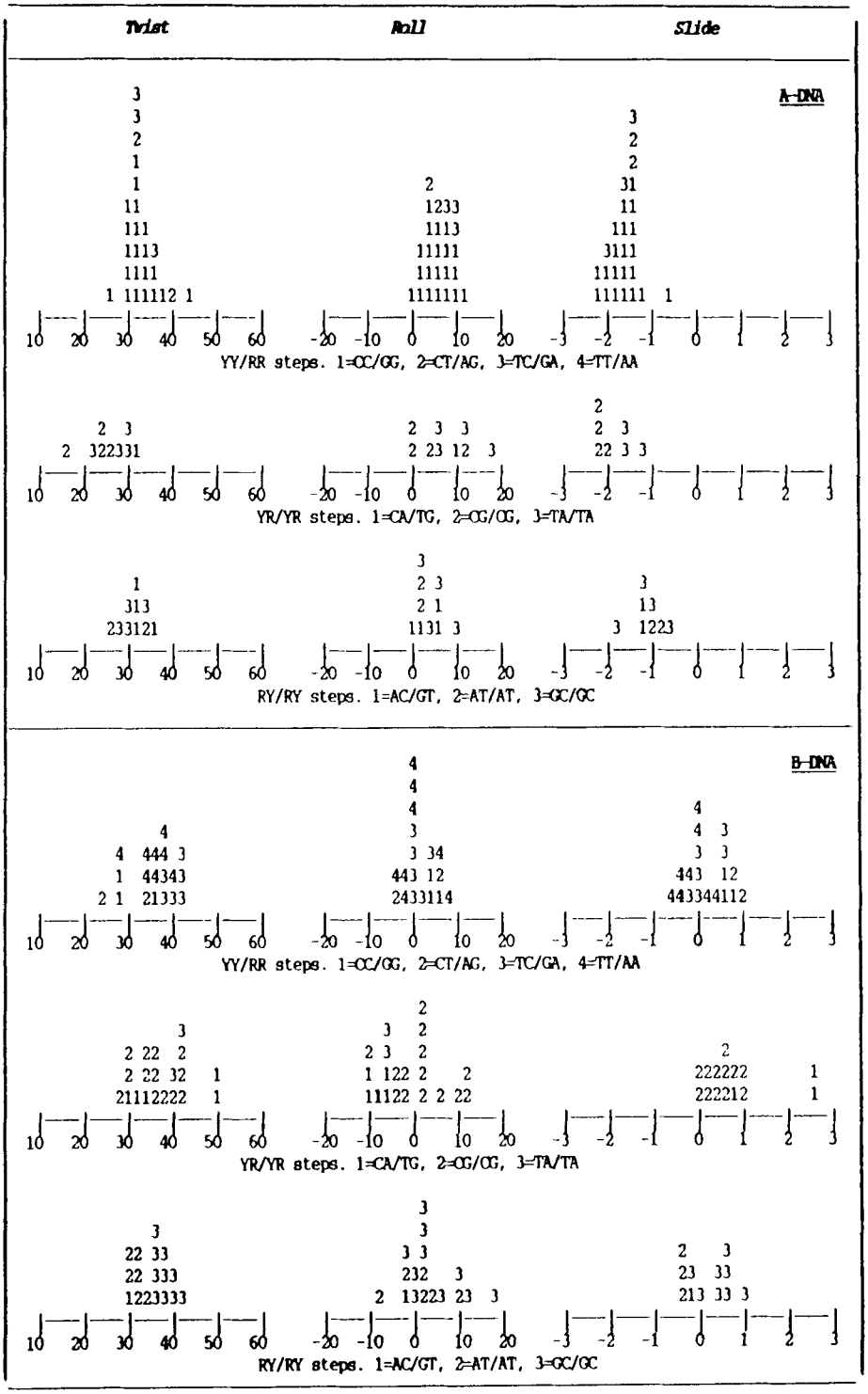

FIG. 1. Histograms of the helicoidal parameters *twist*, *roll* and *slide* in crystalline DNA fragments. *Twist* and *roll* are given in (°), *slide* in (Å).

FIG. 1 gives histograms of *twist, roll* and *slide* values observed in A- and B-DNA fragments analysed by X-ray crystallography. For clarity, these histograms separate YY/RR, YR/YR and RY/RY base pair steps. Some trends pertaining to both A- and B-DNA helices become immediately apparent. *(a)* TT/AA steps are avoided and CC/GG steps are over-represented in A-DNA helices while the reverse holds true for B-DNA. This observation is in agreement with the known preferences of TT/AA and CC/GG to direct DNA double helices into the B- and A-forms, respectively, under suitable conditions[7]. *(b)* For all three parameters, the distributions of adopted values tend to be wider in B-DNA than in A-DNA. Apparently, the A-form helix is locally more rigid than the B-helix. *(c)* YY/RR steps tend to show a narrow "normal" distribution around their standard values (*twist* = 33°, *roll* = 5°, *slide* = -1.5 A in A-DNA and *twist* = 36°, *roll* = 0°, *slide* = 0 A in B-DNA). *(d)* For all three parameters, YR/YR steps tend to adopt a wide range of values and occasionally show extreme behaviour. Examples for this are the CG/CG step in A-DNA and the CA/TG step in B-DNA (see below). Sequence-directed local structural perturbations in DNA helices may thus be expected to occur at YR/YR steps and not at YY/RR steps. *(e)* The ranges of *twist* and *roll* for A- and B-DNA overlap significantly, although the distributions are not centered around the same mean value. In contrast, *slide* shows a bimodal change with little overlap between A- and B-DNA. Hence, *slide* may be a useful parameter to discriminate between the two right-handed helix forms along with the lateral displacement of base pairs from the helix axis and sugar pucker[8]. *(f)* Most importantly, there is no simple sequence-structure code based on dinucleotide code words. In both A- and B-DNA each of the 10 possible base pair stacks adopts a range of conformations, here illustrated by the parameters *twist, roll* and *slide,* which may be, but not always is wide. However, certain trends are discernible.

First, we shall look at A-DNA. *(a)* CG/CG differs from other base pair stacks by displaying unusually small *twist, roll* and *slide* which is accompanied by an extended sugar-phosphate backbone at guanosines and pronounced interstrand purine-purine stacking. This has been noted before[9] and has been the basis for a new hypothesis concerning the B-Z transition of the DNA polymer poly d(GC)[10]. Recently, the influence of crystal packing on the geometry of CG/CG has been demonstrated[11],

suggesting that environmental factors may play an important role in determining the conformation of this dinucleotide step and others. *(b)* CG/CG may not be as unique as it seems at first sight, since YR/YR steps have a tendency to show reduced *twist* in general. *(c)* RY/RY steps, on the other hand, tend to display small values of *slide*.

In B-DNA, *(a)* YR/YR steps are clustered around larger than average values of *slide*. *(b)* Especially, some (but not all) CA/TG steps have quite unusually large values of *twist* and *slide*. This stacking geometry effectively abolishes base-base overlap on both strands of the double helix and leads to a shielding of carbonyl functions by the purine and pyrimidine six-membered rings[4]. *(c)* All CA/TG steps tend to have a more negative *roll* than other YR/YR. *(d)* Lastly, among RY/RY steps GC/GC tends to show the largest values of *twist*, *roll* and *slide*.

It is concluded that a simple code relating dinucleotide sequence with DNA helical structure does not exist. However, the presently available data base already allows some useful generalizations based on conformational preferences of dinucleotide steps. Often, simple geometrical arguments may be proposed to explain the observed structural features of base pair stacks.

ACKNOWLEDGEMENT

We would like to express our gratitude to Ronald Frank and Helmut Blöcker (GBF, Gesellschaft für Biotechnologische Forschung, Braunschweig) and to Sabine Schulze and Rolf Bald (Institut für Biochemie, Freie Universität Berlin) for DNA synthesis and to Wolfram Saenger for reading the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 9/B7.

REFERENCES

1. Dickerson, R.E. (1989) J. Biomol. Struct. Dyn. **6**, 627-634
2. Heinemann, U., Lauble, H., Frank, R. and Blöcker, H. (1987) Nucleic Acids Res. **15**, 9531-9550
3. Lauble, H., Frank, R., Blöcker, H. and Heinemann, U. (1988) Nucleic Acids Res. **16**, 7799-7816
4. Heinemann, U. and Alings, C. (1989) J. Mol. Biol., in press
5. Kennard, O. and Hunter, W.N. (1989) in "Landolt-Börnstein", New Series, VII/1a (Saenger, W., ed.) Springer, Berlin, Heidelberg, New York, pp. 255-360
6. Calladine, C.R. (1982) J. Mol. Biol. **161**, 343-352
7. Peticolas, W.L., Wang, Y. and Thomas, G.A. (1988) Proc. Natl. Acad. Sci. USA **85**, 2579-2583

8. Heinemann, U., Alings, C. and Lauble, H. (1989) in "Proceedings of the Sixth Conversation in Biomolecular Stereodynamics" (Sarma, R.H., ed.) Adenine Press, Schenectady, in press
9. Heinemann, U., Lauble, H., Frank, R. and Blöcker, H. (1988) Nucleosides and Nucleotides **7**, 699-702
10. Saenger, W. and Heinemann, U., submitted
11. Shakked, Z., Guerstein-Guzikevich, G., Frolow, F. and Rabinovich, D., submitted