

23.3. NUCLEIC ACIDS

TAAAACGTTTTAAAACCCCC is much less curved because the roll kink at CG is counterbalanced by roll kinks in the opposite direction at the two flanking TA steps. In both cases, A-tracts are straight and completely unbent. (Note that both roll kinks can involve compression of the major groove, as expected, because the kink sites are a half turn of helix apart.)

This similarity of behaviour of DNA in crystals and in protein–DNA complexes should come as no surprise, since the local molecular environments – close intermolecular contacts, partial dehydration, low water activity, low local dielectric constant, high ionic strength, presence of divalent cations – are similar in these two cases and quite different from that of free DNA in dilute aqueous solution. Far from being unwanted ‘crystal deformations’, the local changes in structure resulting from intermolecular contacts in DNA crystals provide positive information about sequence-dependent deformability that is relevant to the protein recognition process. With regard specifically to A-tract behaviour, Occam’s Razor would argue in favour of model (3) above for the behaviour of A-tracts in solution. The situation in dilute aqueous solution becomes of secondary importance if what is wanted is an understanding of A-tract B-DNA behaviour in protein–DNA complexes. Here, the answer is unambiguous: A-tracts in their biological setting are inherently rigid structural elements, chosen by natural selection when bending should be avoided.

23.3.5. Summary

Three families of nucleic acid double helix have been found – A, B and Z – with widely different structures and usages. The A and B

helices are right-handed and have no limitations on base sequence. Z is left-handed and effectively limited to alternating purines and pyrimidines, with G and C overwhelmingly favoured. B is the biologically significant helix for DNA and is used in genetic coding. A is the helix of preference for RNA because it can accommodate the C2'-OH group of ribose, which produces steric clash in the B helix. The Z helix has, as yet, no well established biological function. A left-handed DNA configuration can be induced in longer DNA segments by negative supercoiling in solution, but it is not clear that this left-handed configuration is identical to the Z-DNA seen in short crystalline oligomers, because of the reversed orientation of backbone strands in Z-DNA.

B-DNA is an inherently malleable or deformable duplex. Its sugar ring conformations are much more variable than those of A-DNA. The base sequence of B-DNA is expressed directly *via* hydrogen bonds between bases of a pair, and indirectly *via* hydrogen-bond donors and acceptors along the floor of the major and minor groove. Sequence is also expressed as a *differential deformability* of different regions of the duplex. The two most obvious parameters affected by base sequence are minor groove width and helix bendability. Certain sequences of B-DNA are not statically bent, but are more bendable under stress than are other sequences. Bending occurs *via* roll, usually in the direction that compresses the broad major groove. Pyrimidine-purine or Y-R steps are most conducive to roll bending, and purine-purine steps are least bendable, particularly A-tracts of four or more AT base pairs without the weak T-A step. Natural selection has engineered Y-R steps into a DNA sequence where a sharp roll bend is wanted, and short A-tracts into a sequence where bending is not desired.

Appendix 23.3.1. X-ray analyses of A, B and Z helices

Table A23.3.1.1. X-ray analyses of A helices, DNA and RNA

This table and the two that follow are intended as a historical background and a focus on the geometry of the intact double helix. References are current as of late 1997; sequences marked ‘to be published’ in 1997 that still are unpublished two years later have been deleted. Also omitted are sequences with fewer than four base pairs in the asymmetric unit, complexes with intercalating drugs, helices with bulges or looped-out bases, unusual structures such as quadruplexes, hammerhead ribozymes and tRNA. For information on these and for more recent results, consult the Nucleic Acid Database (NDB) at <http://ndbserver.rutgers.edu/>. An NDB number in parentheses indicates that the authors have never made coordinates available to the public. These structures are of little scientific value, but have been included for historical reasons.

Notes: Overhanging, unpaired bases are double underlined. Single underlining calls attention to mismatched bases or other interesting or relevant sequence aspects. Z = number of asymmetric units per cell. Ubp = number of base pairs per asymmetric unit. NDB No. = Nucleic Acid Database serial number. Abbreviations: 2am = 2-amino; 5br = 5-bromo; 6ame = 6'- α -methyl; 4mo = 4-methoxy; 5me = 5-methyl; 6aOH = 6'- α -hydroxyl; 6mo = 6-methoxy; 8oxo = 8-oxo; 6et = 6-ethyl; ara = arabinosyl; ps = phosphorothioate; (P) = leading phosphate; A, T, G, C = DNA; a, u, g, c = RNA; Py = pyrrole; Im = imidazole.

(a) Dodecamers

Sequence	Space group	Z	Ubp	Date, institution	NDB No.	Reference
CCCCGCGGGGG	$P3_221$	6	12	1991, Barcelona	ADL025	(A38)
CCGTACGTACGG	$P6_122$	12	6	1992, Ohio State	ADL045	(A41)
GCGTACGTACGC	$P6_122$	12	6	1992, Ohio State	ADL046	(A39)

(b) Decamers

Sequence	Space group	Z	Ubp	Date, institution	NDB No.	Reference
GCGGGCCCGC	$P6_122$	12	5	1993, Ohio State	ADJ051	(A46)
GCACGCGTGC	$P6_122$	12	5	1996, Ohio State	ADJ075	(A60)
ACCGGCCGGT	$P6_122$	12	5	1989, MIT	ADJ022	(A26)
ACCGGCCGGT	$P6_122$	12	5	1995, MIT	ADJ065	(A55)
ACCGGCCGGT	$P6_122$	12	5	1995, MIT	ADJ066	(A55)
CCCGGCCGGG	$P2_12_12_1$	4	10	1993, Ohio State	ADJ049	(A47)
CCIGGCC ^{5me} CGG	$P2_12_12_1$	4	10	1995, Ohio State	ADJB61	(A58)