

# A nomenclature of junctions and branchpoints in nucleic acids

David M. J. Lilley, Robert M. Clegg<sup>1</sup>, Stephan Diekmann<sup>2</sup>, Nadrian C. Seeman<sup>3</sup>, Eberhard von Kitzing<sup>4</sup> and Paul J. Hagerman<sup>5</sup>

CRC Nucleic Acid Structure Research Group, Department of Biochemistry, The University, Dundee DD1 4HN, UK,

<sup>1</sup>Abteilung Molekulare Biologie, Max-Planck-Institut für Biophysikalische Chemie, Am Fassberg 11, D-37077

Göttingen, Germany<sup>2</sup>Institut für Molekulare Biotechnologie e.V, Beutenbergstraße 11, D-07745 Jena, Germany,

<sup>3</sup>Department of Chemistry, New York University, 100 Washington Square East, New York, NY 10003, USA,

<sup>4</sup>Abteilung Zellphysiologie, Max-Planck-Institut für Medizinische Forschung, D-69028 Heidelberg, Germany and

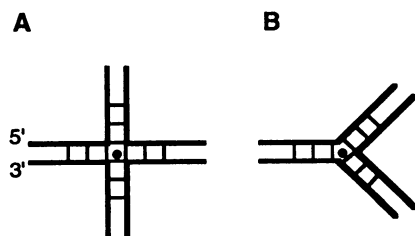
<sup>5</sup>Department of Biochemistry, Biophysics and Genetics, University of Colorado Health Sciences Center, Denver, CO 80262, USA

Branchpoints are a common feature of folded nucleic acids. They may be created from double stranded DNA during processes generating sequence rearrangements such as homologous or site-specific recombination, or by the secondary and tertiary folding of single-stranded nucleic acids. The latter is especially important in RNA, and the branched structure of such molecules may be central to their function. The structures of branchpoints in DNA molecules are known to be recognised by a number of enzymes that may bind to and manipulate DNA junctions. As the complexity of the junctions studied in laboratories has increased, a need has arisen to find a nomenclature that will allow the unambiguous description of any given branchpoint.

We therefore propose a method for describing the connectivity of any branchpoint or junction in the secondary structure of a nucleic acid. Such a branchpoint forms the point of connection between a number of different helical segments, with or without the inclusion of formally single stranded regions.

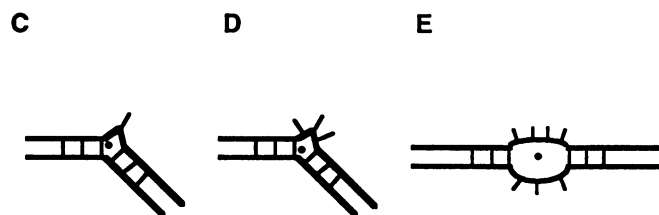
At a chosen place in the molecule, write a dot • about which we will describe the connectivity. Then pass 360° about this dot, through either backbone or basepair (suggested direction 5'→3' through the backbone, i.e. clockwise as drawn below), Write H for each helix and S for a single stranded section traversed. Basepairs are viewed from their minor groove side. Choose a startpoint to maximise the initial number of helices. Single-stranded sections may have a subscript indicating the number of unpaired bases.

## Fully basepaired junctions



According to the scheme, the simple four-way junction (A) becomes HHHH, which could be shortened to 4H, while the three-way junction (B) becomes HHH or 3H. An additional possibility might be to indicate the length of the helix as a subscript, e.g. a three-way junction comprising three arms each of 10 bp might be described by 3H<sub>10</sub>. However, this remains optional because the length of the helix may be ambiguous (these helices can have their own features such as bulges and bubbles), or even unknown.

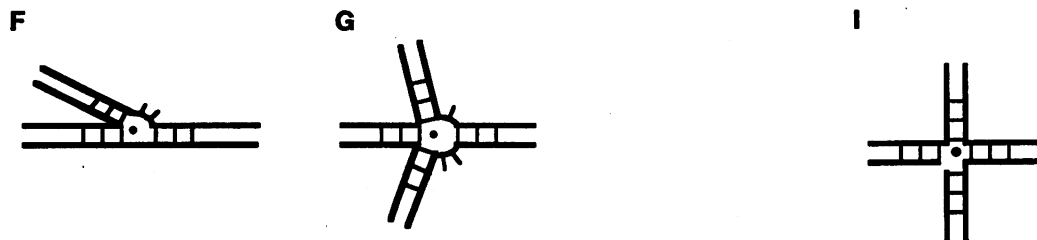
## Duplexes with single-stranded sections



Here we have two molecules with base bulges (C and D), and a duplex interrupted by a single stranded bubble, or internal loop (E). According to the rules, the bulged duplexes becomes 2HS<sub>1</sub> (molecule C) and 2HS<sub>3</sub> (molecule D), while the bubble becomes HS<sub>3</sub>HS<sub>4</sub>. An extra rule is introduced where there are multiple single-stranded regions—the start point is chosen so that the shorter single-stranded region is written first. A single-base mismatch could be described by the formal description HS<sub>1</sub>HS<sub>1</sub>.

## More complex junctions containing single-stranded sections

These will be described using an addition of these rules. They will be written so as to place the maximum number of helical sections first. Taking the examples of three- and four-way junctions with one or more single-stranded sections:



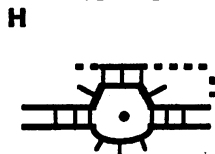
According to the rules, junction (F) becomes  $3HS_2$ , while junction (G) is described by  $3HS_1HS_2$

### Other helical junctions

In principle, this method could be extended to include other types of helical junction. For example, the interface between a section of right-handed B-DNA and one of left-handed Z-DNA might be described by  $H_BH_Z$ . H would be retained as a conventionally basepaired duplex of unspecified character.

### Tertiary interactions

The single-stranded sections (loops or bulges) might make long-range interactions, such as the formation of pseudoknots. The resulting double-stranded sections will be indicated either by D or by  $D_n$ , where  $n$  is the number of base pairs (if unambiguous). Thus a typical pseudoknot



would be described as  $HS_1D_3S_1HS_3$ .

### Breaks in continuity

In some cases the path around the point may be broken at some point, by a nick in the backbone. This could be indicated by a slash /, e.g. the junction

becomes  $4H/$ .

### Branch migration

In some structures, notably the four-way junction, some sequences can undergo branch migration. It is hard to allow explicitly for this, although unless arm lengths are specified this need not alter the formal description of the junction. But it should be borne in mind that in some circumstances the formal description might apply only to one isomer of a given structure.

### NOTE

This is a document of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB), whose members are A. Cornish-Bowden (chairman), A. Bairoch, Helen Berman, C. R. Cantor, C. Liébecq, G. P. Moss, K. F. Tipton and P. Venetianer. NC-IUBMB thanks a panel whose members were David M. J. Lilley (convenor), Robert M. Clegg, Stephan Diekmann, Paul J. Hagermann, Nadrian C. Seeman and Eberhard von Kitzing, for drafting these recommendations, as well as former members of NC-IUBMB and present and former members of the IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN), namely, J. F. G. Vliegthart (past-chairman), A. J. Barrett, M. A. Chester, H. B. F. Dixon, Maria Kaplan, P. Karlson, K. L. Loening, J. C. Rigg, W. Saenger and N. Sharon, for consultation. Comments may be sent to any member of NC-IUBMB, or to the convenor of the panel, David M. J. Lilley.

This document has previously been published in the *European Journal of Biochemistry*.