Structural conventions for group I introns

John M.Burke*, Marlene Belfort¹, Thomas R.Cech², R.Wayne Davies³, Rudolf J.Schweyen⁴, David A.Shub⁵, Jack W.Szostak⁶ and Henk F.Tabak⁷

Department of Chemistry, Williams College, Williamstown, MA 01267, ¹Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany, NY 12201, ²Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA

Received August 18, 1987; Accepted August 24, 1987

ABSTRACT

Conventions for nomenclature of structural elements and a standard secondary structure representation for group I introns have been established by workers in the field. These conventions are designed to facilitate effective communication of information concerning the structure and function of these self-splicing introns.

INTRODUCTION

Splicing of group I introns proceeds by a common transesterification mechanism which is initiated by attack of guanosine at the 5' splice site (1). Splicing of group I introns is mediated by the folded structure of the RNA, as evidenced by the characterization of cis-dominant splicing-defective mutations within yeast mitochondrial introns (2,3) and by the self-splicing of several of these introns in vitro (reviewed in reference 1). Group I introns share characteristic structural features (4,5) that direct the splicing reaction. Researchers attending the 1987 Cold Spring Harbor Laboratory Meeting on RNA Processing met to discuss the adoption of conventions concerning nomenclature and secondary structure diagrams for group I introns. Subsequently, we have acted as a committee to refine the recommendations of the larger group, and we now recommend the following conventions.

CONVENTIONS

The group of introns is termed group I. Alternative nomenclatures (i.e. type 1, Class I) are abandoned.

Secondary Structure Nomenclature

Pairing (P) Segments. P1 through P9 (integral values) are the characteristic base-paired segments of group I intron secondary structure, as defined by Waring and Davies (5) and illustrated in Figs. 1 and 2. They are numbered according to their occurrence as one traces the RNA chain from 5' to 3'. P10, a proposed pairing in some introns between 3' exon sequences adjacent to the splice site and the internal guide sequence, must be drawn separately. Note that P2 is the stem closest to the 3' side of P1, and P9 is the stem closest to the 3' side of P7. Some introns do not have P2, in which case their major double-stranded elements would be numbered P1, P3, P4, etc.

Figure 1. Structural diagrams for group I introns. The location of conserved sequence elements P, Q, R and S are indicated. A, group I introns. B, group IA introns. Arrows indicate positions of 5' and 3' splice sites. Sequence elements P, Q, R and S are indicated by dotted lines.

Additional stems which occur between conserved pairing segments *m* and *n* are termed *Pm.1*, *Pm.2* etc., numbered from 5' to 3' (see *P2.1*, *P9.1* and *P9.2* in Fig. 2 for examples). Stems which extend conserved stem *m* are termed *Pma*, *Pmb*, etc. in alphabetical order from 5' to 3' (see stems *P5a*, *P5b*, *P5c*, *P6a*, *P6b*, *P9.1a* in Fig. 2). Stem *Pm* should be split into segments *a* and *b* when each of the following two criteria are met-- bulged bases on both sides of the pairing segment and more than one bulged base on one or both sides. The conserved stem

Figure 2. Secondary structure of intron within *Tetrahymena thermophila* nuclear large rRNA precursor. Exon bases are indicated by lowercase characters, intron sequences by uppercase characters.

which extends P4 is termed P5 rather than P4a for historical reasons (5). Base-paired nucleotides on the 5' side of pairing segment *m* are termed *Pm*[5']; those on the 3' side are *Pm*[3']. For example, in Fig. 2 P4[5'] has the sequence UGCGGG and P4[3'] has the sequence CCACGCA.

Loops (L). Loops are numbered according to the stem which closes them, e.g. L1, L5b, etc. Note that not all pairing segments have associated loops, and that the loop numbering convention does not correspond to previously published schemes.

Joining (J) Segments. Joining segments are those nucleotides which are located

Table I. Recommended and discontinued nomenclature for group I introns.

<u>Discontinued</u>	<u>Recommended</u>
element A	P
element B	Q
element 9L	R
element 2	S
element 9R (or E')	P3[3']
element 9R' (or E)	P3[5']
9R•9R' (or E'•E)	P3
A•B	P4
9L•2	P7

between pairing segments, but are not loops. They are numbered according to the pairing segments they connect, from 5' to 3'. Thus, the segment joining P3 and P4 is J3/4, and that joining P8 and P7 is J8/7. In the *Tetrahymena* rRNA intron, these have the sequences AAU and AUAAGAUA, respectively (Fig. 2).

Nomenclature for Conserved Internal Sequence Elements

The four internal elements which are conserved in nucleotide sequence are called P, Q, R and S, as originally proposed by Waring and Davies (5). Alternative schemes (e.g. A, B, 9L and 2) are discontinued (see Table I). Each of these conserved elements has base-paired and unpaired segments, e.g. element P is composed of part of J3/4, all of P4[5'] and all or part of J4/5. The E'•E pairing (alternatively termed 9R•9R') is not conserved in sequence, and is simply called P3.

Structural Diagrams

Structures are drawn in the general form indicated in Fig. 1A. A specific example of group I (*Tetrahymena* rRNA intron) is provided in Fig. 2. Various group I introns have extra stems to the 3' sides of P2 and P9. In addition, a number have extensions of core stems (e.g. P5, P6 and P8), often containing open reading frames. These are readily accommodated in the diagram without relocating the core elements. Group IA introns constitute a significant subgroup, and contain additional sequences between P7 and P3. They are drawn as indicated in Fig. 1B.

Diagrams are drawn with all stems oriented either vertically or horizontally, for simplicity in preparation. Closely spaced stems may be connected with additional lines (see for example Fig. 2, segments which join P2-P2.1 and P4-P6).

DISCUSSION

The adoption of standard structural representations for group I introns is proposed for the same reasons that the standard cloverleaf diagram and associated conventions for tRNA were adopted some years ago-- clarity and uniformity of presentation. Most aspects of the intron secondary structure model in Fig. 1 are well established on phylogenetic grounds (6) and three pairings (P1, P3 and P7) have been proven on the basis of mutational studies (3, 7-9). The original models (4,5) were proposed on the basis of a few intron sequences from fungal mitochondria and have since been found to fit introns from chloroplasts (10), protozoan nuclei (11), and T4 bacteriophage (12) equally well. As new data on group I intron structure emerge, further refinements of the structural conventions may become necessary. Different research groups may disagree on details of intron structure, particularly outside of the core region. However, just as cloverleaf diagrams of tRNA continue to be useful at a time when detailed crystal structures are available, it is expected that the conventions concerning group I introns will be adequate for the foreseeable future.

ACKNOWLEDGEMENTS

This work was supported by grants to J.M.B. from NSF (DMB-8502691) and NIH (GM36981).

*To whom correspondence and reprint requests should be addressed

³Allelix, Inc., 6850 Goreway Drive, Mississauga, Ontario L4V 1P1, Canada, ⁴Institut für Microbiologie und Genetik, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria, ⁵Department of Biological Sciences, State University of New York, Albany, NY 12222, ⁶Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA, ⁷Section for Molecular Biology, Laboratory of Biochemistry, University of Amsterdam, Kruislaan 318, 1098 SM, Amsterdam, The Netherlands

REFERENCES

1. Cech, T. R. (1987) *Science* **236**, 1532 - 1539.
2. De La Salle, H., Jacq, C. and Slonimski, P. P. (1982) *Cell* **28**, 721-732.
3. Weiss-Brummer, B., Holl, J., Schweyen, R. J., Rodel, G. R., and Kaudewitz, F. (1983) *Cell* **33**, 195-202.
4. Michel, F., Jacquier, A. and Dujon, B. (1982) *Biochimie* **64**, 867 - 861.
5. Davies, R. W., Waring, R. B., Ray, R. A., Brown, T. A. and Scazzocchio, C. (1982) *Nature* **300**, 719 - 724.
6. Waring, R. B. and Davies, R. W. (1984) *Gene* **28**, 277 - 291.
7. Been, M. D. and Cech, T. R. (1986) *Cell* **47**, 207 - 216.
8. Waring, R. B., Towner, P., Minter, S. J. and Davies, R. W. (1986) *Nature* **321**, 133-139.
9. Burke, J. M., Irvine, K. D., Kaneko, K. J., Kerker, B. J., Oettgen, A. B., Tierney, W. M., Williamson, C. L., Zaug, A. J. and Cech, T. R. (1986) *Cell* **45**, 167-176.
10. Rochaix, J. D., Rahire, M. and Michel, F. (1985) *Nucl. Acids Res.* **13**, 975-984.
11. Cech, T. R., Tanner, N. K., Tinoco, I., Jr., Weir, B. R., Zuker, M. and Perlman, P. S. (1983) *Proc. Nat. Acad. Sci. USA* **80**, 3903-3907.
12. Chu, F. K., Maley, G. F., West, D. K., Belfort, M. and Maley, F. (1986) *Cell* **45**, 157-166.