# Nomenclature for T-cell receptor (TCR) gene segments of the immune system*

WHO-IUIS Nomenclature Sub-Committee on TCR Designation[1]

*The recommended procedures and criteria for T-cell receptor (TCR) designations are described. The officially adopted designations are for the TCR A, B, D and G loci and for V, D, J and C segments.*

## Principles of TCR nomenclature

The following principles and criteria should be observed.

(1) The intention is to develop a simple set of rules that can be applied to all TCR V, D, J and C loci and which are consistent with current trends in gene mapping nomenclature.

(2) The nomenclature should be applicable to all species and to systems characterized at any level from the first sequence to full mapping.

(3) To satisfy (2), the official name should not contain detailed information about gene order, pseudogenes, cDNA or genomic sequence, and the productive or non-productive nature of polymorphisms. Such additional information should be in parentheses after the official name.

(4) Gene segments should be named only when a full cDNA or genomic sequence is available. Naming is at the level of gene segments but data from the cDNA sequence may be used since for most genes only the cDNA sequence is available. A name for the rearranged gene product should be assembled from the loci names (see page 114, TCRD).

(5) The locus names (A, B, G, D) and genetic elements (V, D, J, C) are self-explanatory. S refers to gene segments and is used to enumerate and distinguish subfamily members. The TCR A and D V-segments present a problem since the same V region can be used for A or D TCR chains. In the present proposal one set of names is given for all A and D V-genes; the term A or D or both can be used in the name depending on whether the context is for TCR alpha or delta chains. Thus one might use TCRAV1S1 or TCRDV1S1 or TCRADV1S1 for the same V-gene in different contexts. This may seem unorthodox but once the A/D rule is known there is no problem since there will never be two A/D V-genes with the same V-S- name. The numbering of the few V-genes used *mainly or exclusively* in delta TCRs begin with 101, i.e., V101S1, V102S1, etc., allowing identification of unique delta families.

(6) An asterisk (*) will separate alleles from loci, consistent with gene mapping rules.

## Nomenclature of human TCR gene segments

The officially adopted designations for TCRA, TCRB, TCRG, TCRD (no hyphen separates TCR from A, B, G or D) loci are described below.

(1) *TCRA*

| | |
|---|---|
| TCRAV1S1<br>TCRAV1S2<br>TCRAV1S3 | Distinct loci but these members are of the same family, where S refers to family member. See note 1 (Annex). |
| TCRAV2S1 | Second family. |
| TCRADV1S1 | When the V region can be used by alpha or delta, A, D or AD can be used. See principle 5 (above). |
| TCRAV1S1*1<br>TCRAV1S1*2 | Alleles at the same locus. See notes 2 to 8 (Annex). |

* This article was drafted by a group of experts working under the auspices of the International Union of Immunological Societies (IUIS) and has been approved by the Nomenclature Committee of IUIS. Requests for reprints and all correspondence should be addressed to the Chairman of the IUIS Nomenclature Committee, Professor Michel Kazatchkine, Unite d'Immunologie, Hôpital Broussais, 96 rue Didot, 75014 Paris, France. A French translation of this Terminology Note will appear in a later issue of the *Bulletin*.

[1] Members of the Nomenclature Sub-Committee: A.F. Williams (United Kingdom) (*Chairman*), J.L. Strominger (USA) (*Co-Chairman*), and J. Bell (United Kingdom) (*Co-Chairman*), T.W. Mak (Canada), J. Kappler (USA), P. Marrack (USA), B. Arden (Germany), M.P. Lefranc (France), L. Hood (USA), S. Toneqawa (USA), and M. Davis (USA). Standing Committee on TCR Designation: T.W. Mak (*Chairman*), Ontario Cancer Institute, Princess Margaret Hospital, Toronto, Ontario, Canada, and B. Arden, Paul Ehrlich Institute, Langen, Germany.

Reprint No. **5362**

| | |
|---|---|
| TCRAJ1S1<br>S2 | To designate J region families and their members. |
| TCRAC1 | |
| *1<br>*2 | To designate AC locus. For AC alleles if found. |

Further description, when available, could be provided in parentheses. See note 5 (Annex).

(2) *TCRB*

| | |
|---|---|
| TCRBV1S1<br>BV1S2<br>BV2S1 | Exactly as for TCRAV. |
| *1<br>*2<br>*3 | If alleles are found. |
| TCRBD1<br>TCRBD2 | For the 2 D loci. |
| *1<br>*2 | If alleles are found. |
| TCRBJ1S1<br>TCRBJ1S2<br>TCRBJ1S3<br>TCRBJ2S1<br>TCRBJ2S2<br>TCRBJ2S3 | For the J region loci. See note 9 (Annex). |
| *1<br>*2 | If alleles are found. |
| TCRBC1<br>TCRBC2 | For the two BC loci |

(3) *TCRG*

| | |
|---|---|
| TCRGV1S1<br>S2 | For the V family with multiple segments. |
| TCRGV2S1 | For all of the other segments including pseudogenes. See note 4 (Annex). |
| *1<br>*2 | To be added if alleles are found. |
| TCRGJ1S1<br>TCRGJ1S2<br>TCRGJ1S3<br>TCRGJ2S1 | |
| *1<br>*2 | To be added if alleles are found. |
| TCRGC1 | |
| GC2*1<br>*2<br>*3 | For alleles of the 2nd GC segment. |

(4) *TCRD*

| | |
|---|---|
| TCRDV101S1 | As for the AV segments or with the 101 etc. names. See principle 5 (above). |
| TCRADV101S1 | Where the sequence used in both alpha and delta AD can be used. |
| TCRDV101S2 | |
| *1<br>*2 | To be added if alleles are found. |
| TCRDD1<br>TCRDD2<br>TCRDD3 | For the D region genomic segments. |
| *1<br>*2 | See note 10 (Annex). |
| TCRDJ1S1<br>TCRDJ1S2<br>TCRDJ2S1 | For the 3 J segments. |
| *1<br>*2 | For alleles as before. |
| TCRDC1 | For the only DC segment. |

### *Name for a rearranged gene product*

A complete name for a rearranged TCRA gene (for example) could be:

TCRAV1S1J1S2C1.

Various shorthand versions could be used within a paper after the first complete naming:

V1S1J1S2C1
or V1S1J1S2
or VA1S1J1S2
or VA1S1
or VA1.1
or VB2S2J1S2 (for a TCRB gene)
or VB2S2
or VB2.2

i.e., in a paper this last abbreviation would describe TCRBV2S2J1S2C1 (thereafter called VB2S2 or VB2.2).

# Annex

## *Notes*

(1) Criteria for distribution to families have been given detailed consideration and will be included in a forthcoming publication.

(2) Alleles should only be named if it is certain they are true alleles. If it is possible that they are pseudoalleles (product of a distinct locus), then they should be initially named as a family member but product of a distinct locus, e.g., AV1S1, AV1S2, AV1S3, etc. When proven to be truly allelic, the designation can be changed, e.g., AV1S3 could be changed to AV1S1*2 if it were allelic to AV1S1*1. Alleles are defined at the nucleic acid level.

(3) Care must be taken to *avoid* using N addition nucleotides to designate a V, D or J segment as an allele.

(4) If two identical sequences exist at different loci, these could bear the same name but be distinguished by a or b after the name, i.e. TCRAV1S1*1a or TCRAV1S1*1b.

(5) At the end of the official nomenclature a parenthesis after the official name would contain information about:

| | |
|---|---|
| (a) gene order | (23) |
| (b) pseudogenes | (P) |
| (c) orphon genes | (O) |
| (d) nonproductive substitutions | (N) |
| (e) tentative designation | (T) |

(5a) The gene order would require complete genomic mapping of a complex and would also address duplicate genes.

(5b) Pseudogenes may be shown as a P after the official name. P would only be used when the sequence indicating a functional TCR could not be formed using this segment (e.g., frame shift from a single base deletion, stop codon, loss of an essential amino acid, etc.); i.e., TCRAV1S3*1P.

(5c) For example, a processed pseudogene. Similarly, additional information about loci mapping outside the complex could be denoted by the letter O (orphon).

(5d) An allelic nucleotide substitution that does not result in an amino acid change.

(5e) When an incomplete sequence is obtained and clearly represents a new gene segment, it can be designated as tentative (T). This designation can be removed when the sequence is completed.
For example:

| | | |
|---|---|---|
| | TCRAV1S1*1(43, P) | Representing in order the locus number from the C region, in this case a pseudogene. |
| or | TCRAV1S1*2(43, N) | In this case a nonproductive allelic variant. |
| or | TCRAV1S7(O) | An orphon gene. |

(6) Where the allelic polymorphism is the deletion of a gene, this could be designated by V1S1*O.

(7) In order to be assigned a name, the sequence would have to be referenced (published or in press). Once the nomenclature is introduced and accepted, then all new sequences should be submitted to the Standing Committee by authors and/or editors for naming prior to publication (as is now done in the HLA field).

(8) It would be difficult to insist that a cell line be deposited and available for verification (as is done for HLA alleles) because many TCR sequences are derived from factor-dependent lines (some easily lost) by PCR (polymerase chain reaction), etc. There does not seem to be an easy way to deal with this problem.

(9) The J region loci are classed according to genomic localization, not sequence similarity. This is the one exception with the rest of the nomenclature.

(10) Where multiple D segments are used in the transcript, this can be expressed as follows:

TCRDV1S2D2D3J2C1.