

**Proteins (including monoclonal antibodies):
Mandatory information for INN selection and publication**

ANNEX TO INN APPLICATION FORM

Geneva, May 2020

Programme on International Nonproprietary Names (INN)

***INN Programme and Classification of Medical Products Unit
Health Products Policy and Standards Department (HPS)
Access to Medicines and Health Products Division (MHP)
World Health Organization, Geneva***

© World Health Organization 2020

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by the World Health Organization to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall the World Health Organization be liable for damages arising from its use.

Mandatory information for INN selection and publication for all proteins (including monoclonal antibodies):

Annex to INN application form

(Please note that incomplete requests will not be considered)

For all proteins (including monoclonal antibodies):

- Complete mature amino acid sequence **in a format that can be copied for analysis** (Word or in the text of an e-mail), using the single-letter code for each amino acid with spaces between groups of ten characters, five groups per line and with a number indicating the position of the last amino acid at the end of each line.
- Complete precursor nucleotide sequence with spaces between codons and translation (including the stop codon) and with numbers per line, and **in a format that can be copied for analysis** (Word or in the text of an e-mail).
- If applicable, state and explain the purpose of having amino acid differences with the native sequence (for a monoclonal antibody: constant region amino acid changes by comparison with the closest genomic C gene and allele) (e.g. mutations introduced to alter receptor binding or change the isoelectric point, to prevent C1q binding, enhance FcRn binding, etc.).
- Positions of all disulfide bridges (specify if they are determined or predicted).
- Post-translational modifications (specify if they are determined or predicted).
- Expression system (the cell type, specific strain and the clone name used for the expression).
- If available, the three-dimensional structure in Protein Data Bank format or the Protein Data Bank accession code.
- If **glycosylated**, the glycosylation profile (the types of sugar, the location of glycosylation site(s), etc.); specify if they are determined or predicted; if the cell line in which the protein/peptide is produced is engineered, detailed information if the glycosylation pattern is affected.
- If **conjugated**, the mean numbers of molecules of the conjugated part, and if known, positions where the conjugate is attached.

For monoclonal antibodies:

- IG class and subclass and light chain type.
- Antibody format (e.g. complete antibody, Fab, scFv, etc.) For non-standard formats, including bispecifics, be as specific as possible (e.g. scFv fusion to the beginning of the VH domain, IgG format with two different light chains and two different heavy chains with electrostatic mutations...).
- Source of the original antibody (or antibodies) that provided the binding affinity [e.g., hybridoma (including species origin such as mouse, rat, etc.), EBV immortalization of human B-cells, transgenic mice with human genes, artificial human phage display library, naive human phage display library, immunized human phage display library], be as specific as possible.
- Subsequent engineering of V domains. e.g. none, humanization by CDR grafting (specify the CDR definition that was used), humanization by resurfacing framework mutations, etc. This must be provided for each V domain, if different.
- A graphic representation/drawing of the arrangement of the domains or linkage of the domains.
- CDR-IMGT (sequence and residue range).
- CDR-Kabat (sequence and residue range).
- The closest genomic germline V, J and C genes and allele using IMGT germline names:
 - For the V-domains, if the domains are nominally human (e.g. produced from human antibodies, EBV immortalization of human B-cells, human phage display libraries, transgenic mice with human V-domain genes, or similar), the closest human gene/allele should be given;
 - If the V-domains have been humanized by CDR-grafting onto a human framework, the closest human gene/allele to the parent human framework should be given;
 - Otherwise the closest germline (human or other species) should be given.
- Name/structure of the antigen against which the monoclonal antibody is directed and the official gene name that encodes the target. Where an antibody binds across two or more components of a hetero-multimer, gene names for all relevant components should be provided.
- Laboratory code name(s) and/or code name used in publications and clinical trials.
- If the terminal lysine is absent in the heavy chain amino acid sequence, a statement of the fabricant confirming that indeed there is no lysine codon in the nucleotide sequence (if not the lysine should be added in the amino acid sequence mentioning the posttranslational modification clipping).

For pegylated proteins:

- The details of pegylation: the **end group** and the **polymer chain** with the average number of repeat units (to 2 significant figures).
- The details of the **linker** (not the reagent used): where the linker is attached to the active moiety, and, ideally, if multiple sites are involved, in what proportion they are modified.

Please be aware that sequence information will be published either electronically (Mednet) or in both print and electronic format, depending on the size of the structure.

Examples can be found in published INN lists:

<http://www.who.int/medicines/publications/druginformation/innlists/en/>