FASTA Format

National Cancer Institute

Source

National Cancer Institute. FASTA Format. NCI Thesaurus. Code C47845.

A sequence in FASTA format consists of a single-line description, followed by lines of sequence data. The first character of the description line is a greater-than (">") symbol in the first column. Sequences are represented in the standard IUB/IUPAC single letter amino acid and nucleic acid codes, with a single hyphen or dash being used to represent a gap of indeterminate length; in amino acid sequences asterix ("*") can represent a translation stop.