The FASTA format

The FASTA format is a general format for storing sequences in text files. It is described in the text book on page 23 and 25. It consists of one single initial line starting with a greater-than-symbol (>) and usually followed by sequence identifiers and a short description. The following lines contain the actual amino acid or nucleotide sequence using a single letter per residue. The sequence may be broken onto several lines. There are usually not more than 70 residues per line.

Several sequences may be stored in the same file, one sequence after the other, but each sequence must of course begin with the description line starting with the greater-than-symbol (>).

An example:

```
>gi|15832355|ref|NP_311128.1| alkylated DNA repair protein [Escherichia coli O157:H7 str. Sakai]
MLDLFADAEPWQEPLAAGAVILRRFAFNAAEQLIRDINDVASQSPFRQMVTPGGYTMSVAMTNCGHLGWT
THRQGYLYSPIDPQTNKPWPAMPQSFHNLCQRAATAAGYPDFQPDACLINRYAPGAKLSLHQDKDEPDLR
APIVSVSLGLPAIFQFGGLKRNDPLKRLLLEHGDVVVWGGESRLFYHGIQPLKAGFHPLTTDCRYNLTFR
QAGKKE
>gi|87298840|ref|NP_006011.2| alkylated DNA repair protein alkB homolog [Homo sapiens]
MGKMAAAVGSVATLATEPGEDAFRKLFRFYRQSRPGTADLEGVIDFSAAHAARGKGPGAQKVIKSQLNVS
SVSEQNAYRAGLQPVSKWQAYGLKGYPGFIFIPNPFLPGYQWHWVKQCLKLYSQKPNVCNLDKHMSKEET
QDLWEQSKEFLRYKEATKRRPRSLLEKLRWVTVGYHYNWDSKKYSADHYTPFPSDLGFLSEQVAAACGFE
DFRAEAGILNYYRLDSTLGIHVDRSELDHSKPLLSFSFGQSAIFLLGGLQRDEAPTAMFMHSGDIMIMSG
FSRLLNHAVPRVLPNPEGEGLPHCLEAPLPAVLPRDSMVEPCSMEDWQVCASYLKTARVNMTVRQVLATD
QNFPLEPIEDEKRDISTEGFCHLDDQNSEVKRARINPDS
```