

## GenomeQuest Quick Start Guide

### Small Peptide Searches

#### How to enter a search into GenomeQuest

1. Select **IP Search** as search type from the main GenomeQuest page.
2. Select the query sequence type (either nucleotide or protein) sequence in the **Query** box by clicking on the appropriate tab and paste the desired sequence in the box. Remember that it must be in FASTA format, and all query sequences must be identified in the header line. It's best to not have spaces in the header.

**Example:**

```
>US20100223293-16257
CTGCAGGCGAACTGTTCTGTTCCGGTAGGGAATTCACCGCAGGCGAGAACGCGGCCGGCT
GCAAAGCTTGCGGAGTACGGAGTAAAACTATTTTTTGAGGGCTGCCGCCTTTGGACAAAT
CCAGTAAACCCACCGAGCTTCGGAAATGTGGGACTGAGAAGGGACGGCGATCCCAGATCA
CACAGAGGACAGGGGAAAACGAAGCCACCGAGCCCCACACGTCGCCATCCATCGCCGTA
ATCGATCACCGCCGTCTCCTCCCCACACACCCACCGGAACCGTCGTCCTGACCTCTGCCA
GCGATAAGCAAATCCCCTCCCCACTTTATCGTCCACAAAGCCTTCTTCCCGCCCTCCCGAATC
GCTCCCTCTGTCCCTCCGCTCCAGCCGCCGCCGCTCCGCCCTCCGAATCCCAATGG
CGTCCGCGGCCGCTCCTCCAACCTCCCTCTCCCTCGCGGCCCGCGCGGCCACCAGGGCAGC
CGCCGCCGCCGCCGCCCGC
```

3. For all search types other than Patentability, select the option of **patents databases only** under the **Type of search**. For patentability searches please select **Patents and public reference databases**.
4. Provide a suitable and easily recognizable name under **Result name**, such that one can easily identify the desired search results.

The screenshot below illustrates how the window for entry of nucleotide query sequences will look. Note that the **NUCLEOTIDE SEQUENCES** tab is selected for this example of a nucleotide query sequence.

The screenshot shows the 'IP Search' interface. At the top, there is a 'Query' section with two tabs: 'Nucleotide Sequences' (which is highlighted with a red box) and 'Protein Sequences'. Below the tabs is a text input area with the prompt 'Paste or Choose your query.' and a sub-prompt 'Paste your nucleotide sequences to begin.' Below the input area, there are search options: 'Type of Search' with radio buttons for 'Patents Databases Only' (selected) and 'Patents and Public Reference Databases'. There is also a 'Result Name' field containing 'IP 2012-10-19 08:00:51' and a checkbox for 'Send E-mail on completion'. At the bottom, there is a checkbox for 'Compare to both nucleotide and protein databases'.


5. Optionally, you may select the box **“compare to both nucleotide and protein databases”** for protein or nucleotide query sequences. Some examples when one might NOT to compare against both nucleotide and proteins are if:
  - a. nucleotide sequence is not expressed: e.g. promoter, terminator, intron, or other non-expressed sequence;
  - b. nucleotide sequence is genomic, or a whole genome (very unlikely to have such a query sequence, but it can happen);
  - c. sequences are very short, such as primers, probes;
  - d. fusion proteins or other synthetic proteins not from nature;
  - e. vectors, plasmids, vector containing a gene (although you may want to search the gene against protein as well separately)'
  - f. any other query sequence where it doesn't make scientific sense to search it against both.

6. Select the appropriate algorithm under the **“search strategy”** option:

Query	Algorithm
CDR or other short peptide	GenePast
Sequence longer than 25-50 residues	Blast or GenePast
Sequence with variable positions	MOTIF

**For all algorithms, you have the choice to limit subject length.**

## GenePast options (best for CDRs, other short peptides or nucleic acid sequences):

**Search Strategy** 

**GenePAST**   Blast   Fragment Search   Motif

Percent identity

Over the length of

**Additional Strategy Parameters**

Limit subject length from  to  residues

Keep a maximum of  results (per query)

---

**Percent identity:** Minimum % identity in order to keep a result. If you are looking for an almost exact match, you can set this very high (e.g. 95% or even higher).

If you are looking for sequences where you have a reasonable idea of what length to expect, then set the “limit subject length” to reflect this. For example, if you are searching a group of CDR sequences and you want to find out if they are in a given antibody molecule, setting the subject length range between x (whatever the length of the shortest CDR is) and perhaps 900 (or whatever makes sense to you, most likely much shorter) will help focus your resultset, even before you begin analysis.

Consider these pre-filters.

---

**General help information:** <https://docs.genomequestlive.com/>

**For further assistance:** [support@gqlifesciences.com](mailto:support@gqlifesciences.com)