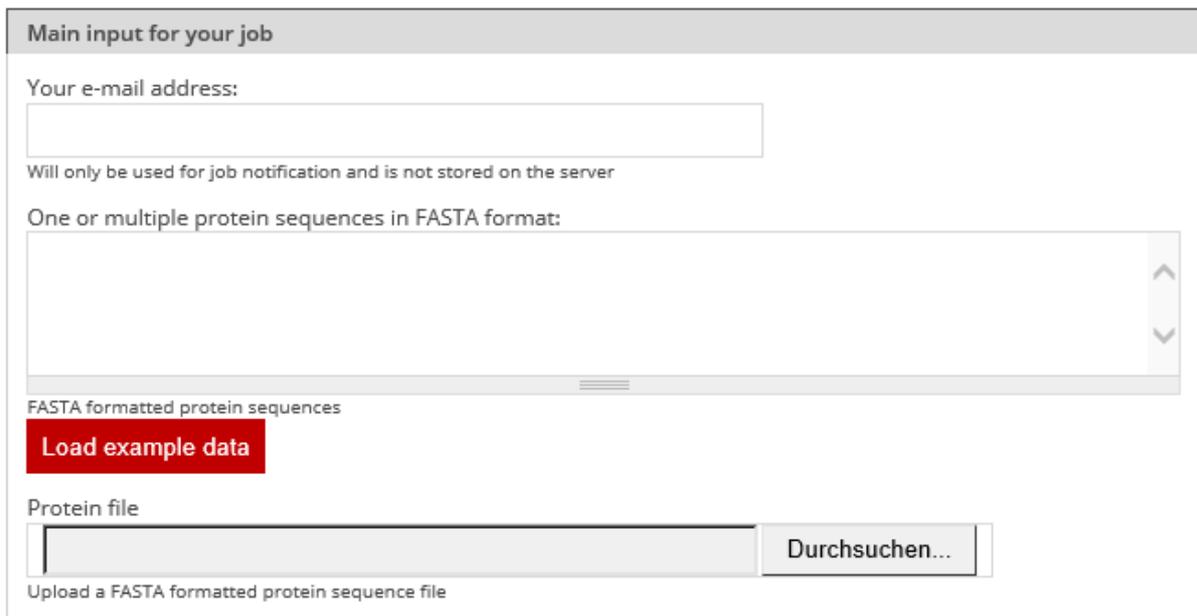# User guide – Job submission

The first box of the submission form contains the main input information for the calculations. With EffectiveDB it is possible to have a single multi-FASTA protein file as one and only input data to run all provided predicting methods.

Figure 1 The option of stating an e-mail address is useful, especially when enabling the time-consuming "genome-mode". The major box in the center should contain the data that is going to be analyzed. It can be either a single protein, a group of proteins and/or proteins from a whole genome in multi-FASTA protein format.

Within the workflow of EffectiveDB, there are two major parts (protein mode / genome mode), which can be enabled either separately or combined. It depends on the number of submitted proteins, which of the two modes is appropriate.

If you submit a single protein or a small group of proteins, the protein mode should be used. EffectiveDB is then able to predict whether the uploaded protein(s) are putatively secreted (EffectiveT3, T4SEpre, EffectiveCCBD, EffectiveELD). Additionally, if enabled by the user, Predotar, gives information about the target location in the specific host-cell. Each of the prediction methods can be manually en/disabled, according to the interests of the user.
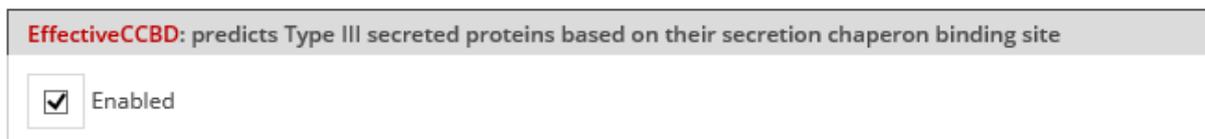
Figure 2 Enabling/Disabling of a specific function, shown at the example of EffectiveCCBD

The genome mode is disabled by default, due to its long calculation time. If it is enabled, the user must upload the (whole) genomic information of an organism.

Figure 3 The genome mode must be manually enabled. It searches for type III, IV and VI bacterial secretion systems within the uploaded genome. Additionally the user can check the input data for genomic completeness, done by CheckM. When enabling the genome mode, automatically the Z-score of EffectiveELD will be refined.

Although it is more time consuming than secretion prediction, also predictions of bacterial secretion systems (EffectiveS346) can be obtained. The genome mode has to be enabled for that purpose. The genome mode assumes that the submitted proteins represent an entire genome, which is as complete as possible. This can optionally be checked by CheckM. Although the user has the opportunity to change the minimal threshold of genomic completeness, our recommendation is 85%.

As an additional feature to reduce the calculation time substantially, there is the option to upload an already prepared genotype-file containing the COG/NOG-profile of the interested organism (arrow #1 in Fig. 3). It is crucial to upload it in the proper format (example file can be found by following the link; arrow #2 in Fig. 3)

- Only COG/NOG from eggNOG 4.0 that are present in the interested genome are valid input data
- All COG/NOG must be on root level. It is not possible to upload bactNOG for example.
- Only one single COG/NOG accession per line is allowed.

There can be an optional second column in the genotype-file that includes the protein name. In this case, the possible number of secretion systems can be estimated from the proteins associated to the most relevant COGs/NOGs.

If a valid COG/NOG profile is uploaded, the time consuming COGnitor run is omitted automatically. Thus, no additional disabling is necessary.