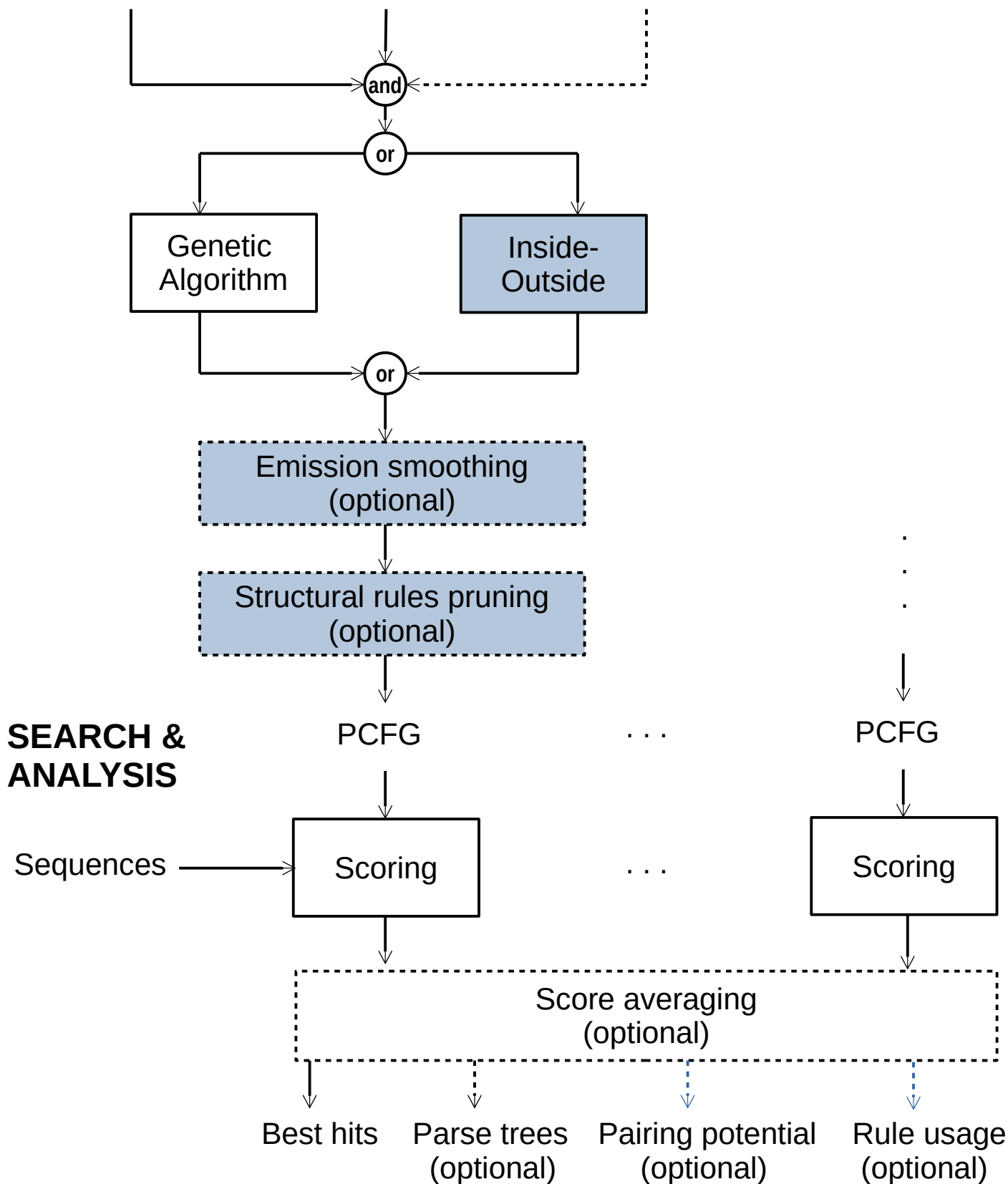


TRAINING

Covering CFG

Sequences

Contact lists (optional)



Supplementary Figure 1: The outline of the processing pipeline in the probabilistic context-free grammar-based framework for protein sequence analysis. Optional features, inputs and outputs are marked with dotted lines. Newly introduced features and outputs are marked with grey blue.