

Figure S1: **The Biozon hierarchical document and relation classification model.** Datasets are mapped to nodes and edges whose class is one of those displayed in either hierarchy. When incorporating new datasets, the hierarchy may be extended by adding the appropriate subclass if a given data object or relation is currently not represented in Biozon.

Mapping ESTs to proteins with Biozon

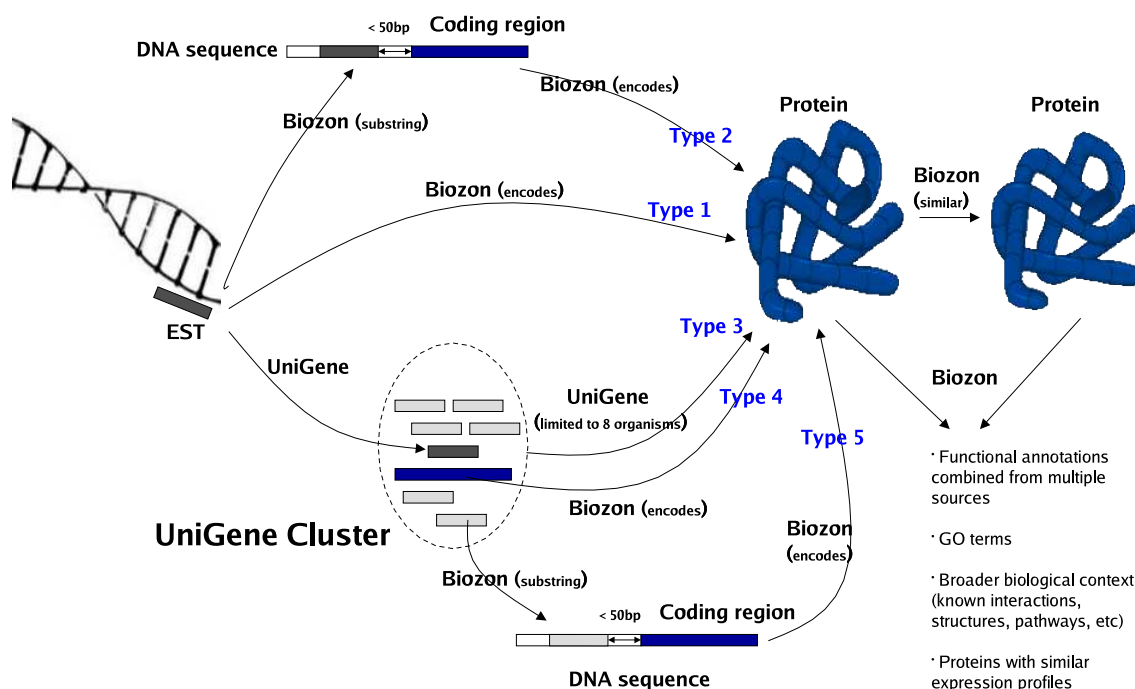


Figure S2: **The Biozon's EST2Prot system.** An EST is mapped to a protein using one of five possible paths in the Biozon data graph. We say that EST s is mapped to protein P if (i) s encodes P (ii) s is a substring of DNA s' near an encoding region of s' which encodes for P , (iii) s is in a Unigene cluster to which NCBI assigns P , (iv) s is in a Unigene cluster containing s' and s' encodes P , (v) s is a member of a UniGene cluster containing s' and s' is a substring of s'' near an encoding region of s'' which encodes for P . To enrich the set of functional descriptors associated with each EST we also utilize similarity relations between proteins.