

Supplementary Material

Table 1: The number entries extracted from several sources. Repeated data record with different accession number in same source was removed

Source name	Number of entries
UniProtKB/Swiss-Prot	186
GenBank	110
Protein Data Bank (PDB)	102
Protein Information Resource (PIR)	4
Protein Resource Format (PRF)	3