1. Swiss-Prot (all entries on Oct 25 2010)

2. Non-bacterial entries removed

3. Old entries removed (integration date < 2005)

4. Entries annotated with uninformative words removed

5. Entries with unchanged TrEMBL annotation removed

6. Similarly annotated entries removed

7. Entries similar in sequence removed (ID > 90%, COV > 90%)

8. Test proteins compared against UniProtKB with BLAST

9. Upon characterization created BLAST neighbours removed

Annotation of remaining BLAST neighbours restored using UniSave