



Introduction

The Human Proteome Organization (HUPO) Proteomics Standards Initiative (PSI) has just finalized the PSI Extended FASTA Format (PEFF). It is based on the widely-used FASTA format for encoding the protein sequence databases used by most tandem mass spectrometry (MS/MS) search engines, but adds a uniform mechanism for encoding substantially more metadata about the sequence collection as a whole as well as individual entries, including support for encoding known sequence variants, PTMs, and proteoforms. PEFF is defined by a full specification document, controlled vocabulary terms, a set of example files, software libraries, and a file validator [1].

Benefits of using PEFF

- Consistent interpretation of sequence annotations by all supporting tools.
- Enables ability to search for known amino acid substitutions.
- Enables ability to search for certain PTMs at known positions without needing to search for all via search parameters, increasing sensitivity, search speed and reducing false positives. The current neXtProt PEFF database includes a wide set of modifications that are typically not searched for.
- Enables display in search results of the known and considered variants/PTMs, and the known and detected variants/PTMs.
- It can be used to fully describe individual proteoforms.
- Includes metadata about databases in the file.

Next Steps

Encourage more widespread implementations of PEFF in sequence database producers and consumers.

More information

Web site:

psidev.info/peff

Preprint:



Specification

The PEFF format defines a file header section and individual sequence entries section. The file header includes information such as the database name, source, version, etc. PEFF sequence entries are similar to standard FASTA sequence entries but with extended description lines.

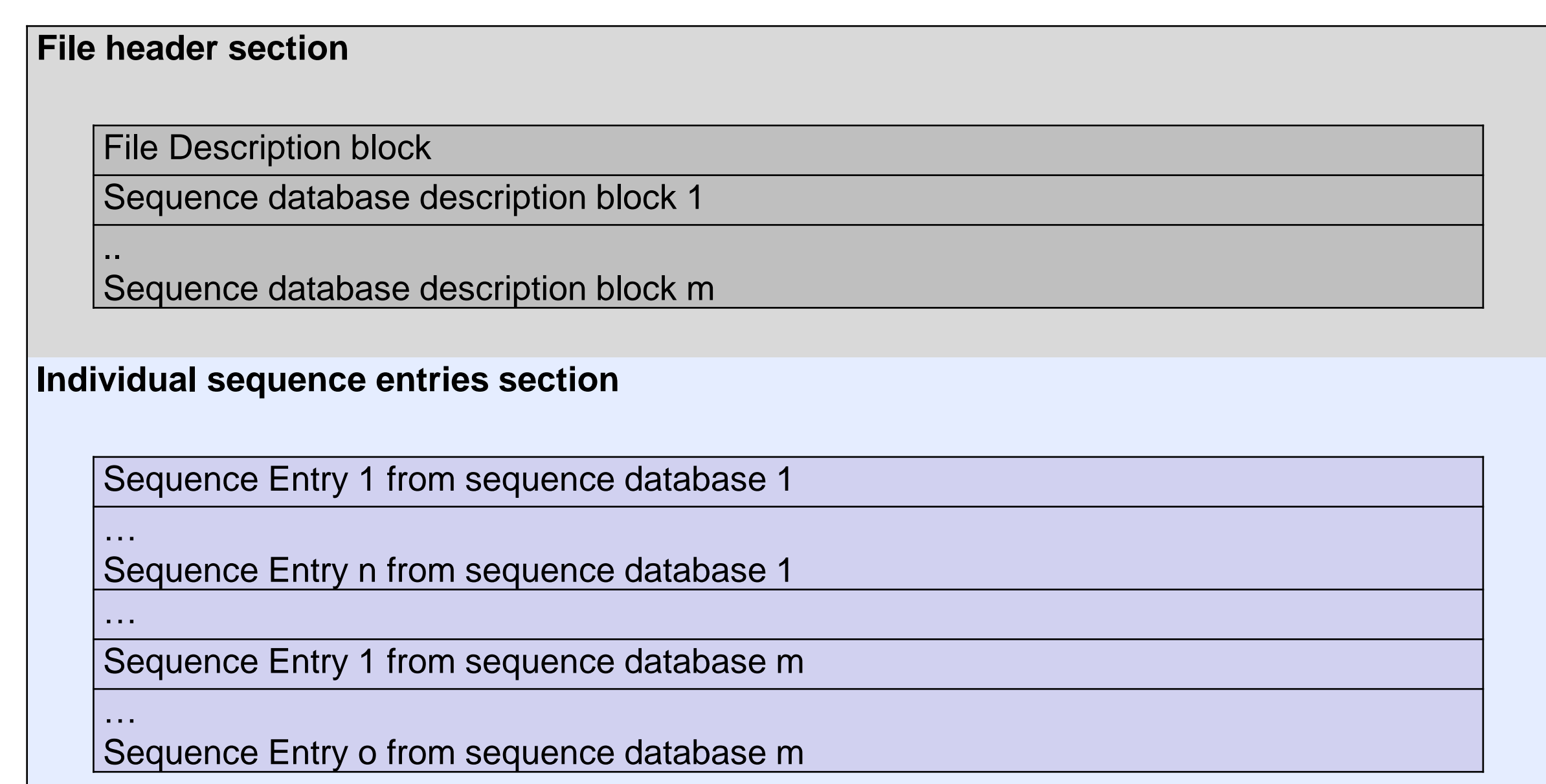


Figure 2. Graphical representation of the PEFF file structure

The **file description block** declares the format version and optionally general comments

As a PEFF file may include multiple sequence databases, metadata about each of them are described in the corresponding **sequence database description blocks**. These include DB name, version, source, sequence type, among others, all encoded using the following format:

key=value

Where the *key* elements are controlled vocabulary (CV) terms and the *value* term contains one or more items whose format is defined for each key in the CV.

```
# DbName=neXtProt-extract
# Prefix=nxp
# DbDescription=extract of neXtProt with manual modifications
# GeneralComment= A selection of protein entries
# Decoy=false
# DbVersion=2019-01-11
# DbSource=http://www.nextprot.org
# NumberOfEntries=62
# SequenceType=AA
# //
```

Figure 2. Example of a sequence database description block. This block can be repeated if the file contains multiple databases (such as forward and decoy)

In the **individual sequence entries section**, each of the sequence entries have a description line and a sequence line.

The PEFF description lines are in the following format

>Prefix:DbUniqueId \key=value \key=value ...

Where the *key* elements are controlled vocabulary (CV) terms and the *value* term contains one or more items whose format is defined for each key in the CV.

In addition to the mandatory unique entry identifier, information such as protein name, sequence length, sequence variants, PTMs, sequence maturation or post-translational processing and taxonomy can be described.

```
>nxp:NX_Q13217-1 \DbUniqueId=NX_Q13217-1 \PName=DnaJ homolog subfamily C member 3 isoform Iso 1 \GName=DNAJC3
\NcbiTaxId=9606 \TaxName=Homo sapiens \Length=504 \SV=156 \EV=471 \PE=1 \ModResPsi=(274|MOD:00046|O-phospho-L-
serine) (290|MOD:00064|N6-acetyl-L-lysine) \ModRes=(248|Disulfide) (258|Disulfide) (313|Disulfide)
(329|Disulfide) \VariantSimple=(4|S) (34|S) (49|S) (50|S) (51|S) (52|S) (53|S) (54|S) (55|S) (56|S) (57|S) (58|S)
(59|S) (60|S) (61|S) (62|S) (63|S) (64|S) (65|S) (66|S) (67|S) (68|S) (69|S) (70|S) (71|S) (72|S) (73|S) (74|S) (75|S)
(76|S) (77|S) (78|S) (79|S) (80|S) (81|S) (82|S) (83|S) (84|S) (85|S) (86|S) (87|S) (88|S) (89|S) (90|S) (91|S) (92|S)
(93|S) (94|S) (95|S) (96|S) (97|S) (98|S) (99|S) (100|S) (101|S) (102|S) (103|S) (104|S) (105|S) (106|S) (107|S) (108|S)
(109|S) (110|S) (111|S) (112|S) (113|S) (114|S) (115|S) (116|S) (117|S) (118|S) (119|S) (120|S) (121|S) (122|S) (123|S)
(124|S) (125|S) (126|S) (127|S) (128|S) (129|S) (130|S) (131|S) (132|S) (133|S) (134|S) (135|S) (136|S) (137|S) (138|S)
(139|S) (140|S) (141|S) (142|S) (143|S) (144|S) (145|S) (146|S) (147|S) (148|S) (149|S) (150|S) (151|S) (152|S) (153|S)
(154|S) (155|S) (156|S) (157|S) (158|S) (159|S) (160|S) (161|S) (162|S) (163|S) (164|S) (165|S) (166|S) (167|S) (168|S)
(169|S) (170|S) (171|S) (172|S) (173|S) (174|S) (175|S) (176|S) (177|S) (178|S) (179|S) (180|S) (181|S) (182|S) (183|S)
(184|S) (185|S) (186|S) (187|S) (188|S) (189|S) (190|S) (191|S) (192|S) (193|S) (194|S) (195|S) (196|S) (197|S) (198|S)
(199|S) (200|S) (201|S) (202|S) (203|S) (204|S) (205|S) (206|S) (207|S) (208|S) (209|S) (210|S) (211|S) (212|S) (213|S)
(214|S) (215|S) (216|S) (217|S) (218|S) (219|S) (220|S) (221|S) (222|S) (223|S) (224|S) (225|S) (226|S) (227|S) (228|S)
(229|S) (230|S) (231|S) (232|S) (233|S) (234|S) (235|S) (236|S) (237|S) (238|S) (239|S) (240|S) (241|S) (242|S) (243|S)
(244|S) (245|S) (246|S) (247|S) (248|S) (249|S) (250|S) (251|S) (252|S) (253|S) (254|S) (255|S) (256|S) (257|S) (258|S)
(259|S) (260|S) (261|S) (262|S) (263|S) (264|S) (265|S) (266|S) (267|S) (268|S) (269|S) (270|S) (271|S) (272|S) (273|S)
(274|S) (275|S) (276|S) (277|S) (278|S) (279|S) (280|S) (281|S) (282|S) (283|S) (284|S) (285|S) (286|S) (287|S) (288|S)
(289|S) (290|S) (291|S) (292|S) (293|S) (294|S) (295|S) (296|S) (297|S) (298|S) (299|S) (300|S) (301|S) (302|S) (303|S)
(304|S) (305|S) (306|S) (307|S) (308|S) (309|S) (310|S) (311|S) (312|S) (313|S) (314|S) (315|S) (316|S) (317|S) (318|S)
(319|S) (320|S) (321|S) (322|S) (323|S) (324|S) (325|S) (326|S) (327|S) (328|S) (329|S) (330|S) (331|S) (332|S) (333|S)
(334|S) (335|S) (336|S) (337|S) (338|S) (339|S) (340|S) (341|S) (342|S) (343|S) (344|S) (345|S) (346|S) (347|S) (348|S)
(349|S) (350|S) (351|S) (352|S) (353|S) (354|S) (355|S) (356|S) (357|S) (358|S) (359|S) (360|S) (361|S) (362|S) (363|S)
(364|S) (365|S) (366|S) (367|S) (368|S) (369|S) (370|S) (371|S) (372|S) (373|S) (374|S) (375|S) (376|S) (377|S) (378|S)
(379|S) (380|S) (381|S) (382|S) (383|S) (384|S) (385|S) (386|S) (387|S) (388|S) (389|S) (390|S) (391|S) (392|S) (393|S)
(394|S) (395|S) (396|S) (397|S) (398|S) (399|S) (400|S) (401|S) (402|S) (403|S) (404|S) (405|S) (406|S) (407|S) (408|S)
(409|S) (410|S) (411|S) (412|S) (413|S) (414|S) (415|S) (416|S) (417|S) (418|S) (419|S) (420|S) (421|S) (422|S) (423|S)
(424|S) (425|S) (426|S) (427|S) (428|S) (429|S) (430|S) (431|S) (432|S) (433|S) (434|S) (435|S) (436|S) (437|S) (438|S)
(439|S) (440|S) (441|S) (442|S) (443|S) (444|S) (445|S) (446|S) (447|S) (448|S) (449|S) (450|S) (451|S) (452|S) (453|S)
(454|S) (455|S) (456|S) (457|S) (458|S) (459|S) (460|S) (461|S) (462|S) (463|S) (464|S) (465|S) (466|S) (467|S) (468|S)
(469|S) (470|S) (471|S) (472|S) (473|S) (474|S) (475|S) (476|S) (477|S) (478|S) (479|S) (480|S) (481|S) (482|S) (483|S)
(484|S) (485|S) (486|S) (487|S) (488|S) (489|S) (490|S) (491|S) (492|S) (493|S) (494|S) (495|S) (496|S) (497|S) (498|S)
(499|S) (500|S) (501|S) (502|S) (503|S) (504|S) (505|S) (506|S) (507|S) (508|S) (509|S) (510|S) (511|S) (512|S) (513|S)
(514|S) (515|S) (516|S) (517|S) (518|S) (519|S) (520|S) (521|S) (522|S) (523|S) (524|S) (525|S) (526|S) (527|S) (528|S)
(529|S) (530|S) (531|S) (532|S) (533|S) (534|S) (535|S) (536|S) (537|S) (538|S) (539|S) (540|S) (541|S) (542|S) (543|S)
(544|S) (545|S) (546|S) (547|S) (548|S) (549|S) (550|S) (551|S) (552|S) (553|S) (554|S) (555|S) (556|S) (557|S) (558|S)
(559|S) (560|S) (561|S) (562|S) (563|S) (564|S) (565|S) (566|S) (567|S) (568|S) (569|S) (570|S) (571|S) (572|S) (573|S)
(574|S) (575|S) (576|S) (577|S) (578|S) (579|S) (580|S) (581|S) (582|S) (583|S) (584|S) (585|S) (586|S) (587|S) (588|S)
(589|S) (590|S) (591|S) (592|S) (593|S) (594|S) (595|S) (596|S) (597|S) (598|S) (599|S) (600|S) (601|S) (602|S) (603|S)
(604|S) (605|S) (606|S) (607|S) (608|S) (609|S) (610|S) (611|S) (612|S) (613|S) (614|S) (615|S) (616|S) (617|S) (618|S)
(619|S) (620|S) (621|S) (622|S) (623|S) (624|S) (625|S) (626|S) (627|S) (628|S) (629|S) (630|S) (631|S) (632|S) (633|S)
(634|S) (635|S) (636|S) (637|S) (638|S) (639|S) (640|S) (641|S) (642|S) (643|S) (644|S) (645|S) (646|S) (647|S) (648|S)
(649|S) (650|S) (651|S) (652|S) (653|S) (654|S) (655|S) (656|S) (657|S) (658|S) (659|S) (660|S) (661|S) (662|S) (663|S)
(664|S) (665|S) (666|S) (667|S) (668|S) (669|S) (670|S) (671|S) (672|S) (673|S) (674|S) (675|S) (676|S) (677|S) (678|S)
(679|S) (680|S) (681|S) (682|S) (683|S) (684|S) (685|S) (686|S) (687|S) (688|S) (689|S) (690|S) (691|S) (692|S) (693|S)
(694|S) (695|S) (696|S) (697|S) (698|S) (699|S) (700|S) (701|S) (702|S) (703|S) (704|S) (705|S) (706|S) (707|S) (708|S)
(709|S) (710|S) (711|S) (712|S) (713|S) (714|S) (715|S) (716|S) (717|S) (718|S) (719|S) (720|S) (721|S) (722|S) (723|S)
(724|S) (725|S) (726|S) (727|S) (728|S) (729|S) (730|S) (731|S) (732|S) (733|S) (734|S) (735|S) (736|S) (737|S) (738|S)
(739|S) (740|S) (741|S) (742|S) (743|S) (744|S) (745|S) (746|S) (747|S) (748|S) (749|S) (750|S) (751|S) (752|S) (753|S)
(754|S) (755|S) (756|S) (757|S) (758|S) (759|S) (760|S) (761|S) (762|S) (763|S) (764|S) (765|S) (766|S) (767|S) (768|S)
(769|S) (770|S) (771|S) (772|S) (773|S) (774|S) (775|S) (776|S) (777|S) (778|S) (779|S) (780|S) (781|S) (782|S) (783|S)
(784|S) (785|S) (786|S) (787|S) (788|S) (789|S) (790|S) (791|S) (792|S) (793|S) (794|S) (795|S) (796|S) (797|S) (798|S)
(799|S) (800|S) (801|S) (802|S) (803|S) (804|S) (805|S) (806|S) (807|S) (808|S) (809|S) (810|S) (811|S) (812|S) (813|S)
(814|S) (815|S) (816|S) (817|S) (818|S) (819|S) (820|S) (821|S) (822|S) (823|S) (824|S) (825|S) (826|S) (827|S) (828|S)
(829|S) (830|S) (831|S) (832|S) (833|S) (834|S) (835|S) (836|S) (837|S) (838|S) (839|S) (840|S) (841|S) (842|S) (843|S)
(844|S) (845|S) (846|S) (847|S) (848|S) (849|S) (850|S) (851|S) (852|S) (853|S) (854|S) (855|S) (856|S) (857|S) (858|S)
(859|S) (860|S) (861|S) (862|S) (863|S) (864|S) (865|S) (866|S) (867|S) (868|S) (869|S) (870|S) (871|S) (872|S) (873|S)
(874|S) (875|S) (876|S) (877|S) (878|S) (879|S) (880|S) (881|S) (882|S) (883|S) (884|S) (885|S) (886|S) (887|S) (888|S)
(889|S) (890|S) (891|S) (892|S) (893|S) (894|S) (895|S) (896|S) (897|S) (898|S) (899|S) (900|S) (901|S) (902|S) (903|S)
(904|S) (905|S) (906|S) (907|S) (908|S) (909|S) (910|S) (911|S) (912|S) (913|S) (914|S) (915|S) (916|S) (917|S) (918|S)
(919|S) (920|S) (921|S) (922|S) (923|S) (924|S) (925|S) (926|S) (927|S) (928|S) (929|S) (930|S) (931|S) (932|S) (933|S)
(934|S) (935|S) (936|S) (937|S) (938|S) (939|S) (940|S) (941|S) (942|S) (943|S) (944|S) (945|S) (946|S) (947|S) (948|S)
(949|S) (950|S) (951|S) (952|S) (953|S) (954|S) (955|S) (956|S) (957|S) (958|S) (959|S) (960|S) (961|S) (962|S) (963|S)
(964|S) (965|S) (966|S) (967|S) (968|S) (969|S) (970|S) (971|S) (972|S) (973|S) (974|S) (975|S) (976|S) (977|S) (978|S)
(979|S) (980|S) (981|S) (982|S) (983|S) (984|S) (985|S) (986|S) (987|S) (988|S) (989|S) (990|S) (991|S) (992|S) (993|S)
(994|S) (995|S) (996|S) (997|S) (998|S) (999|S) (1000|S) (1001|S) (1002|S) (1003|S) (1004|S) (1005|S) (1006|S) (1007|S)
(1008|S) (1009|S) (1010|S) (1011|S) (1012|S) (1013|S) (1014|S) (1015|S) (1016|S) (1017|S) (1018|S) (1019|S) (1020|S)
(1021|S) (1022|S) (1023|S) (1024|S) (1025|S) (1026|S) (1027|S) (1028|S) (1029|S) (1030|S) (1031|S) (1032|S) (1033|S)
(1034|S) (1035|S) (1036|S) (1037|S) (1038|S) (1039|S) (1040|S) (1041|S) (1042|S) (1043|S) (1044|S) (1045|S) (1046|S)
(1047|S) (1048|S) (1049|S) (1050|S) (1051|S) (1052|S) (1053|S) (1054|S) (1055|S) (1056|S) (1057|S) (1058|S) (1059|S)
(1060|S) (1061|S) (1062|S) (1063|S) (1064|S) (1065|S) (1066|S) (1067|S) (1068|S) (1069|S) (1070|S) (1071|S) (1072|S)
(1073|S) (1074|S) (1075|S) (1076|S) (1077|S) (1078|S) (1079|S) (1080|S) (1081|S) (1082|S) (1083|S) (1084|S) (1085|S)
(1086|S) (1087|S) (1088|S) (1089|S) (1090|S) (1091|S) (1092|S) (1093|S) (1094|S) (1095|S) (1096|S) (1097|S) (1098|S)
(1099|S) (1100|S) (1101|S) (1102|S) (1103|S) (1104|S) (1105|S) (1106|S) (1107|S) (1108|S) (1109|S) (1110|S) (1111|S)
(1112|S) (1113|S) (1114|S) (1115|S) (1116|S) (1117|S) (1118|S) (1119|S) (1120|S) (1121|S) (1122|S) (1123|S) (1124|S)
(1125|S) (1126|S) (1127|S) (1128|S) (1129|S) (1130|S) (1131|S) (1132|S) (1133|S) (1134|S) (1135|S) (1136|S) (1137|S)
(1138|S) (1139|S) (1140|S) (1141|S) (1142|S) (1143|S) (1144|S) (1145|S) (1146|S) (1147|S) (1148|S) (1149|S) (1150|S)
(1151|S) (1152|S) (1153|S) (1154|S) (1155|S) (1156|S) (1157|S) (1158|S) (1159|S) (1160|S) (1161|S) (1162|S) (1163|S)
(1164|S) (1165|S) (1166|S) (1167|S) (1168|S) (1169|S) (1170|S) (1171|S) (1172|S) (1173|S) (1174|S) (1175|S) (1176|S)
(1177|S) (1178|S) (1179|S) (1180|S) (1181|S) (1182|S) (1183|S) (1184|S) (1185|S) (1186|S) (1187|S) (1188|S) (1189|S)
(1190|S) (1191|S) (1192|S) (1193|S) (1194|S) (1195|S) (1196|S) (1197|S) (1198|S) (1199|S) (1200|S) (1201|S) (1202|S)
(1203|S) (1204|S) (1205|S) (1206|S) (1207|S) (1208|S) (1209|S) (1210|S) (1211|S) (1212|S) (1213|S) (1214|S) (1215|S)
(1216|S) (1217|S) (1218|S) (1219|S) (1220|S) (1221|S) (1222|S) (1223|S) (1224|S) (1225|S) (1226|S) (1227|S) (1228|S)
(1229|S) (1230|S) (1231|S) (1232|S) (1233|S) (1234|S) (1235|S) (1236|S) (1237|S) (1238|S) (1239|S) (1240|S) (1241|S)
(1242|S) (1243|S) (1244|S) (1245|S) (1246|S) (1247|S) (1248|S) (1249|S) (1250|S) (1251|S) (1252|S) (1253|S) (1254|S)
(1255|S) (1256|S) (1257|S) (1258|S) (1259|S) (1260|S) (1261|S) (1262|S) (1263|S) (1264|S) (1265|S) (1266|S) (1267|S)
(1268|S) (1269|S) (1270|S) (1271|S) (1272|S) (1273|S) (1274|S) (1275|S) (1276|S) (1277|S) (1278|S) (1279|S) (1280|S)
(1281|S) (1282|S) (1283|S) (1284|S) (1285|S) (1286|S) (1287|S) (1288|S) (1289|S) (1290|S) (1291|S) (1292|S) (1293|S)
(1294|S) (1295|S) (1296|S) (1297|S) (1298|S) (1299|S) (1300|S) (1301|S) (1302|S) (1303|S) (1304|S) (1305|S) (1306|S)
(1307|S) (1308|S) (1309|S) (1310|S) (1311|S) (1312|S) (1313|S) (1314|S) (1315|S) (1316|S) (1317|S) (1318|S) (1319|S)
(1320|S) (1321|S) (1322|S) (1323|S) (1324|S) (1325|S) (1326|S) (1327|S) (1328|S) (1329|S) (1330|S) (1331|S) (1332|S)
(1333|S) (1334|S) (1335|S) (1336|S) (1337|S) (1338|S) (1339|S) (1340|S) (1341|S) (1342|S) (1343|S) (1344|S) (1345|S)
(1346|S) (1347|S) (1348|S) (1349|S) (1350|S) (1351|S) (1352|S) (1353|S) (1354|S) (1355|S) (1356|S) (1357|S) (1358|S)
(1359|S) (1360|S) (1361|S) (1362|S) (1363|S) (1364|S) (1365|S) (136
```