

File contents

The files contain the following columns:

ID and structure column names:

- Hit molecule name
- Hit SMILES

Annotation column names

- protein_class_level_3_pref_name
- protein_class_level_4_pref_name
- protein_class_level_5_pref_name
- protein_class_level_6_pref_name

Property column names:

- mol_mass
- logp
- psa
- h_bond_donors
- h_bond_acceptors

Interpretation of the annotations, suggested file usage

The file contains duplicate compounds, as one compound can have more than one annotation.

The *annotation columns* mark a path from one level below the root to the leaf node of the ProteinClassification tree. The class_level of some leaf nodes is <6, so some of these fields might be empty. Below we included the whole ProteinClassification tree in the following format: **chemble_db_id name**. The uppermost class ("Kinase") has a class_level=2. The class_level grows by 1 with every indentation, up to level 6. Depending on your project, you can filter the file by filtering the appropriate *annotation column* by the ProteinClassification **name** of interest. Not all ProteinClassification names present in the tree have compounds for them in the data file.

Example 1: Get all compounds that are expected to act on "AGC protein kinase CRIK subfamily". From the tree file, based on the indentation you can see that the class_level of this receptor group is 6, so to get all compounds, you can filter the csv for rows where

protein_class_level_6_pref_name = "AGC protein kinase CRIK subfamily"

Example 2: Get all compounds that are expected to act on "Protein kinase regulatory subunit". Based on the indentation, the class_level of this receptor group is 3, so you can filter the csv for rows where

protein_class_level_3_pref_name = "Protein kinase regulatory subunit"

ProteinClassification tree

6 Kinase

- 1100 Protein Kinase
 - 1289 AGC protein kinase group
 - 203 AGC protein kinase AKT family
 - 1668 AGC protein kinase DMPK family
 - 314 AGC protein kinase CRIK subfamily
 - 301 AGC protein kinase GEK subfamily
 - 315 AGC protein kinase ROCK subfamily
 - 1618 AGC protein kinase GRK family
 - 302 AGC protein kinase BARK subfamily
 - 283 AGC protein kinase GRK subfamily
 - 1682 AGC protein kinase MAST family
 - 204 AGC protein kinase NDR family
 - 232 AGC protein kinase PDK1 subfamily
 - 205 AGC protein kinase PKA family
 - 206 AGC protein kinase PKC family
 - 317 AGC protein kinase PKC alpha subfamily
 - 318 AGC protein kinase PKC delta subfamily
 - 286 AGC protein kinase PKC eta subfamily
 - 303 AGC protein kinase PKC iota subfamily
 - 304 AGC protein kinase PKC theta
 - 490 AGC protein kinase PKC unclassified
 - 207 AGC protein kinase PKG family
 - 208 AGC protein kinase PKN family
 - 1648 AGC protein kinase RSK family
 - 287 AGC protein kinase MSK subfamily
 - 289 AGC protein kinase RSK subfamily
 - 382 AGC protein kinase RSKp70
 - 288 AGC protein kinase p70 subfamily
 - 209 AGC protein kinase SGK family
 - 233 AGC protein kinase YANK family
 - 1452 Atypical protein kinase group
 - 1664 Atypical protein kinase ABC1 family
 - 471 Atypical protein kinase ABC1-A subfamily
 - 291 Atypical protein kinase BCR family
 - 119 Atypical protein kinase PDHK subfamily
 - 305 Atypical protein kinase PIKK family
 - 450 Atypical protein kinase FRAP subfamily
 - 1647 Atypical protein kinase RIO family
 - 437 Atypical protein kinase RIO1 subfamily
 - 438 Atypical protein kinase RIO2 subfamily
 - 439 Atypical protein kinase RIO3 subfamily
 - 320 Atypical protein kinase alpha family
 - 1339 CAMK protein kinase group
 - 1702 CAMK protein kinase CAMK1 family

- 409 CAMK protein kinase AMPK subfamily
 - 410 CAMK protein kinase BRSK subfamily
 - 411 CAMK protein kinase CHK1 subfamily
 - 383 CAMK protein kinase LKB subfamily
 - 412 CAMK protein kinase MARK subfamily
 - 413 CAMK protein kinase MELK subfamily
 - 414 CAMK protein kinase NIM1 subfamily
 - 415 CAMK protein kinase NuaK subfamily
 - 416 CAMK protein kinase PASK subfamily
 - 384 CAMK protein kinase QIK subfamily
- 269 CAMK protein kinase CAMK2 family
- 245 CAMK protein kinase DAPK family
- 1710 CAMK protein kinase DCAMK1 family
- 1713 CAMK protein kinase MAPKAPK family
 - 472 CAMK protein kinase MAPKAPK subfamily
 - 428 CAMK protein kinase MNK subfamily
- 246 CAMK protein kinase MLCK family
- 234 CAMK protein kinase PHk family
- 235 CAMK protein kinase PIM family
- 236 CAMK protein kinase PKD family
- 270 CAMK protein kinase RAD53 family
- 1690 CAMK protein kinase RSKb family
 - 385 CAMK protein kinase MSKb subfamily
 - 386 CAMK protein kinase RSKb subfamily
- 247 CAMK protein kinase TSSK family
- 1715 CAMK protein kinase unique family
- 1301 CK1 protein kinase group
 - 210 CK1 protein kinase CK1 family
 - 321 CK1 protein kinase CK1-a
 - 322 CK1 protein kinase CK1-g
 - 946 CK1 protein kinase VRK family
- 148 CMGC protein kinase group
 - 237 CMGC protein kinase CDK family
 - 323 CMGC protein kinase CDC2 subfamily
 - 324 CMGC protein kinase CDK2
 - 325 CMGC protein kinase CDK5 subfamily
 - 326 CMGC protein kinase CDK7 subfamily
 - 327 CMGC protein kinase CDK8 subfamily
 - 328 CMGC protein kinase CDK9 subfamily
 - 429 CMGC protein kinase PCTAIRE
 - 430 CMGC protein kinase PFTAIRE
 - 431 CMGC protein kinase PITSLRE subfamily
 - 388 CMGC protein kinase TAIRE subfamily
 - 497 CMGC protein kinase unclassified CDK subfamily
- 248 CMGC protein kinase CDKL family
- 211 CMGC protein kinase CK subfamily
- 297 CMGC protein kinase CLK family

- 1669 CMGC protein kinase DYRK family
 - 465 CMGC protein kinase Dyrk1 subfamily
 - 466 CMGC protein kinase Dyrk2 subfamily
 - 390 CMGC protein kinase HIPK subfamily
- 239 CMGC protein kinase GSK family
- 249 CMGC protein kinase MAPK family
 - 329 CMGC protein kinase ERK subfamily
 - 391 CMGC protein kinase ERK1
 - 392 CMGC protein kinase ERK3
 - 393 CMGC protein kinase ERK5
 - 330 CMGC protein kinase JNK subfamily
 - 331 CMGC protein kinase nmo subfamily
 - 332 CMGC protein kinase p38 subfamily
- 1640 CMGC protein kinase RCK family
 - 306 CMGC protein kinase MAK
- 250 CMGC protein kinase SRPK family
- 158 Other protein kinase group
 - 251 Other protein kinase AUR family
 - 292 Other protein kinase Bud32 family
 - 1701 Other protein kinase CAMKK family
 - 432 Other protein kinase Meta subfamily
 - 271 Other protein kinase CDC7 family
 - 252 Other protein kinase CK2 family
 - 293 Other protein kinase Haspin family
 - 253 Other protein kinase IKK family
 - 254 Other protein kinase MOS family
 - 255 Other protein kinase NAK family
 - 256 Other protein kinase NEK family
 - 395 Other protein kinase Nek1
 - 417 Other protein kinase Nek11
 - 272 Other protein kinase NKF1 family
 - 1686 Other protein kinase NKF4 family
 - 1629 Other protein kinase PEK family
 - 396 Other protein kinase GCN2 subfamily
 - 334 Other protein kinase HRI
 - 335 Other protein kinase PEK subfamily
 - 336 Other protein kinase PKR
 - 257 Other protein kinase PLK family
 - 397 Other protein kinase PLK2
 - 1657 Other protein kinase TLK family
 - 273 Other protein kinase TOPK family
 - 308 Other protein kinase TTK family
 - 258 Other protein kinase ULK family
 - 259 Other protein kinase WEE family
 - 398 Other protein kinase WEE1
 - 260 Other protein kinase Wnk family
 - 440 Other protein kinase unique family

- 1331 STE protein kinase group
 - 261 STE protein kinase STE11 family
 - 340 STE protein kinase ASK
 - 419 STE protein kinase MEKK2
 - 262 STE protein kinase STE20 family
 - 400 STE protein kinase FRAY subfamily
 - 341 STE protein kinase KHS subfamily
 - 342 STE protein kinase MSN subfamily
 - 343 STE protein kinase MST subfamily
 - 420 STE protein kinase NinaC subfamily
 - 401 STE protein kinase PAKA subfamily
 - 402 STE protein kinase PAKB subfamily
 - 344 STE protein kinase SLK subfamily
 - 345 STE protein kinase TAO subfamily
 - 346 STE protein kinase YSK subfamily
 - 298 STE protein kinase STE7 family
 - 1714 STE protein kinase unique family
- 1285 TK protein kinase group
 - 130 Tyrosine protein kinase Abl family
 - 131 Tyrosine protein kinase Ack family
 - 132 Tyrosine protein kinase Alk family
 - 133 Tyrosine protein kinase Axl family
 - 134 Tyrosine protein kinase Csk family
 - 1614 Tyrosine protein kinase DDR family
 - 149 Tyrosine protein kinase EGFR family
 - 135 Tyrosine protein kinase Eph family
 - 150 Tyrosine protein kinase FGFR family
 - 136 Tyrosine protein kinase Fak family
 - 137 Tyrosine protein kinase Fer family
 - 151 Tyrosine protein kinase InsR family
 - 152 Tyrosine protein kinase JakA family
 - 153 Tyrosine protein kinase JakB family
 - 139 Tyrosine protein kinase Met family
 - 1684 Tyrosine protein kinase Musk family
 - 159 Tyrosine protein kinase PDGFR family
 - 140 Tyrosine protein kinase Ret family
 - 1649 Tyrosine protein kinase Sev family
 - 141 Tyrosine protein kinase Src family
 - 242 Tyrosine protein kinase SrcA
 - 219 Tyrosine protein kinase Srm
 - 142 Tyrosine protein kinase Syk family
 - 143 Tyrosine protein kinase Tec family
 - 144 Tyrosine protein kinase Tie family
 - 146 Tyrosine protein kinase Trk family
 - 160 Tyrosine protein kinase VEGFR family
- 1334 TKL protein kinase group
 - 1674 TKL protein kinase IRAK family

- | — 1678 TKL protein kinase LISK family
 - | — 348 TKL protein kinase LIMK subfamily
 - | — 441 TKL protein kinase TESK subfamily
- | — 240 TKL protein kinase LRRK family
- | — 1623 TKL protein kinase MLK family
 - | — 349 TKL protein kinase HH498 subfamily
 - | — 294 TKL protein kinase ILK subfamily
 - | — 295 TKL protein kinase LZK subfamily
 - | — 421 TKL protein kinase MLK subfamily
 - | — 307 TKL protein kinase TAK1 subfamily
- | — 1638 TKL protein kinase RAF family
- | — 241 TKL protein kinase RIPK family
- | — 1693 TKL protein kinase STKR family
 - | — 404 TKL protein kinase STKR Type 1 subfamily
 - | — 405 TKL protein kinase STKR Type 2 subfamily
 - | — 403 TKL protein kinase STKR1
- | — 129 Protein kinase regulatory subunit