File contents

The files contain the following columns:

ID and structure column names:

- Molecule name
- SMILES

Annotation column names

- protein_class_level_2_pref_name
- protein_class_level_3_pref_name
- protein_class_level_4_pref_name
- protein_class_level_5_pref_name

Property column names:

- logp
- psa
- refractivity
- rotatable_bonds
- h_bond_donors
- h_bond_acceptors
- aromatic_rings
- aliphatic_rings
- acidic_group_count
- basic_group_count
- non_cyclic_amide_count
- fsp3
- mol_mass
- halogen_atom_count
- nso_atom_count
- heteroatom_ratio

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 <5, 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 (“Membrane receptor”) has a class_level=1. The class_level grows by 1 with every indentation, up to level 5. 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 “Hydroxycarboxylic acid receptor”. From the tree file, based on the indentation you can see that the class_level of this receptor group is 5, so to get all compounds, you can filter the csv for rows where

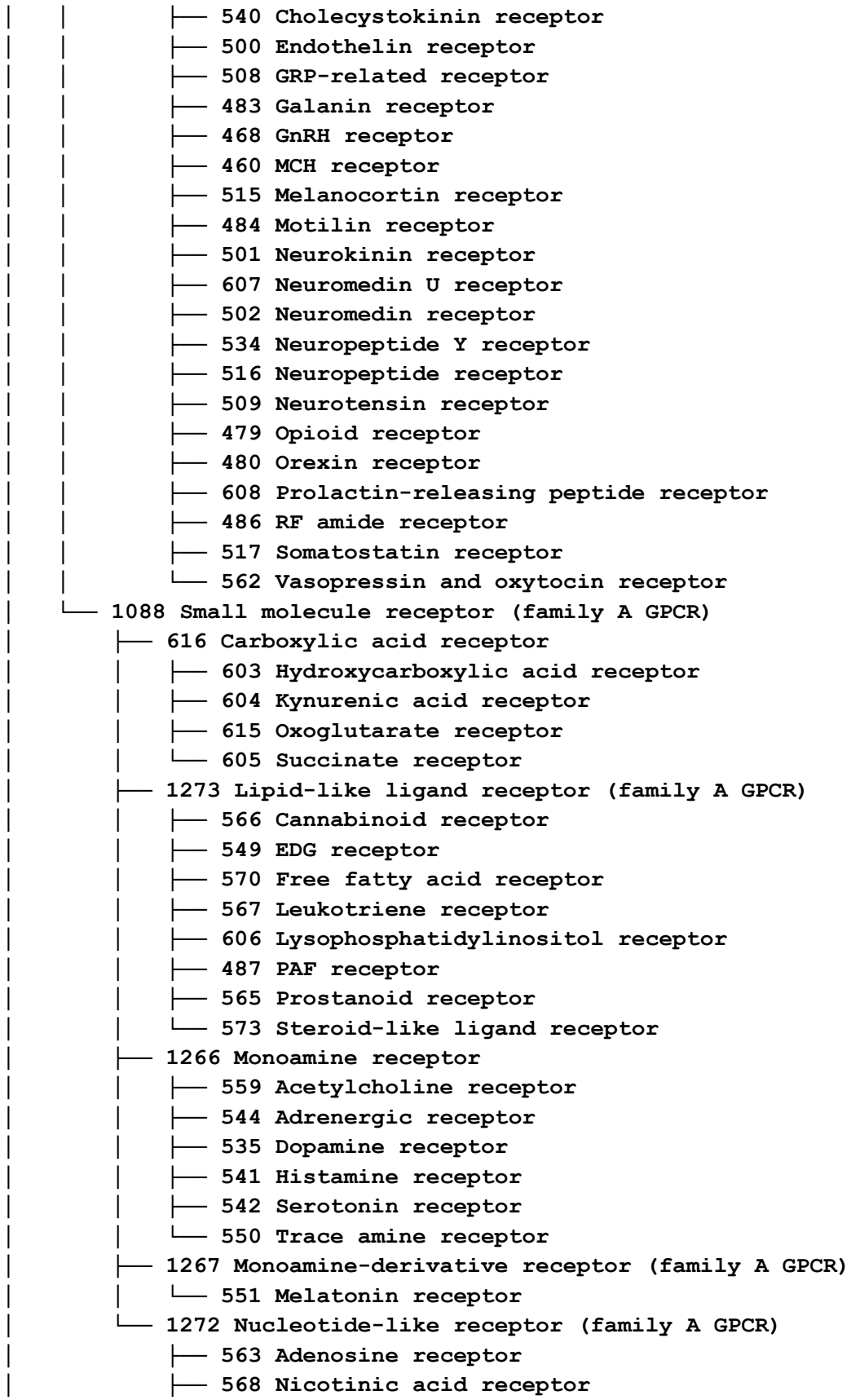
protein_class_level_5_pref_name = “Hydroxycarboxylic acid receptor”

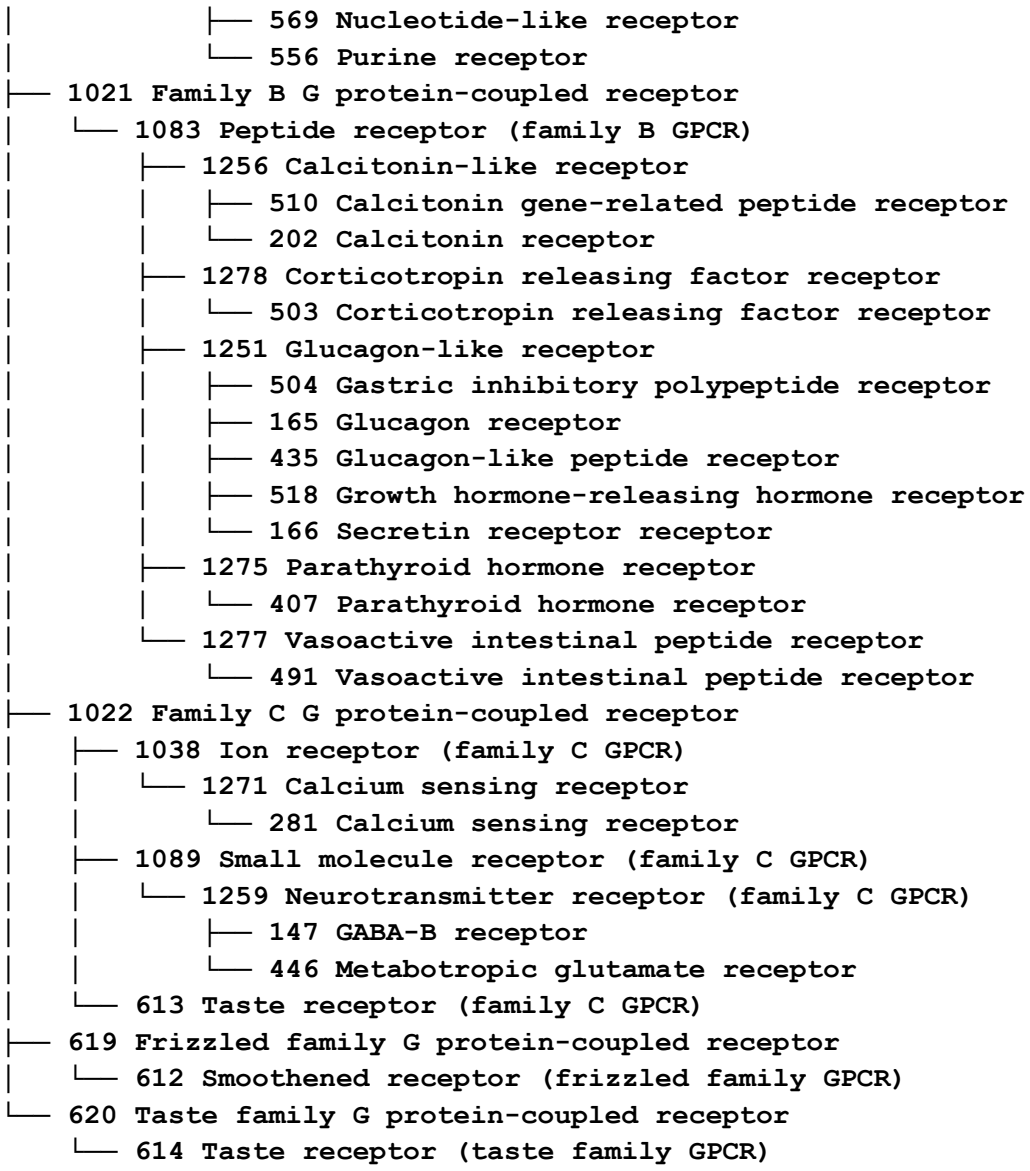
Example 2: Get all compounds that are expected to act on “Small molecule receptor (family A GPCR)”. Based on the indentation, the class_level of this receptor group is 3, so you can filter the csv for rows where

protein_class_level_3_pref_name = “Small molecule receptor (family A GPCR)”

ProteinClassification tree

```
11 Membrane receptor
├── 1020 Family A G protein-coupled receptor
│   ├── 621 Opsin
│   │   └── 611 Rhodopsin
│   ├── 1082 Peptide receptor (family A GPCR)
│   │   ├── 1269 Anaphylatoxin receptor family
│   │   │   └── 528 Anaphylatoxin receptor
│   │   ├── 1265 Chemokine receptor
│   │   │   ├── 547 CC chemokine receptor
│   │   │   ├── 558 CX3C chemokine receptor
│   │   │   ├── 554 CXC chemokine receptor
│   │   │   └── 548 XC chemokine receptor
│   │   ├── 1270 Chemokine receptor-like
│   │   │   └── 555 Chemerin receptor
│   │   ├── 1268 Glycohormone receptor
│   │   │   └── 561 Glycohormone receptor
│   │   ├── 506 N-formyl methionyl peptide receptor
│   │   │   └── 574 N-formyl methionyl peptide receptor
│   │   ├── 617 Peptide growth factor receptor (family A GPCR)
│   │   │   └── 609 Prokineticin receptor
│   │   ├── 1274 Protease-activated receptor
│   │   │   └── 571 Protease-activated receptor
│   │   ├── 618 Relaxin-like peptide receptor (family A GPCR)
│   │   │   └── 610 Relaxin receptor
│   │   └── 1253 Short peptide receptor (family A GPCR)
│   │       ├── 533 Adrenomedullin receptor
│   │       ├── 507 Angiotensin receptor
│   │       ├── 499 Bradykinin receptor
```


540 Cholecystokinin receptor
500 Endothelin receptor
508 GRP-related receptor
483 Galanin receptor
468 GnRH receptor
460 MCH receptor
515 Melanocortin receptor
484 Motilin receptor
501 Neurokinin receptor
607 Neuromedin U receptor
502 Neuromedin receptor
534 Neuropeptide Y receptor
516 Neuropeptide receptor
509 Neurotensin receptor
479 Opioid receptor
480 Orexin receptor
608 Prolactin-releasing peptide receptor
486 RF amide receptor
517 Somatostatin receptor
562 Vasopressin and oxytocin receptor
1088 Small molecule receptor (family A GPCR)
616 Carboxylic acid receptor
603 Hydroxycarboxylic acid receptor
604 Kynurenic acid receptor
615 Oxoglutarate receptor
605 Succinate receptor
1273 Lipid-like ligand receptor (family A GPCR)
566 Cannabinoid receptor
549 EDG receptor
570 Free fatty acid receptor
567 Leukotriene receptor
606 Lysophosphatidylinositol receptor
487 PAF receptor
565 Prostanoid receptor
573 Steroid-like ligand receptor
1266 Monoamine receptor
559 Acetylcholine receptor
544 Adrenergic receptor
535 Dopamine receptor
541 Histamine receptor
542 Serotonin receptor
550 Trace amine receptor
1267 Monoamine-derivative receptor (family A GPCR)
551 Melatonin receptor
1272 Nucleotide-like receptor (family A GPCR)
563 Adenosine receptor
568 Nicotinic acid receptor


569 Nucleotide-like receptor
556 Purine receptor
1021 Family B G protein-coupled receptor
1083 Peptide receptor (family B GPCR)
1256 Calcitonin-like receptor
510 Calcitonin gene-related peptide receptor
202 Calcitonin receptor
1278 Corticotropin releasing factor receptor
503 Corticotropin releasing factor receptor
1251 Glucagon-like receptor
504 Gastric inhibitory polypeptide receptor
165 Glucagon receptor
435 Glucagon-like peptide receptor
518 Growth hormone-releasing hormone receptor
166 Secretin receptor receptor
1275 Parathyroid hormone receptor
407 Parathyroid hormone receptor
1277 Vasoactive intestinal peptide receptor
491 Vasoactive intestinal peptide receptor
1022 Family C G protein-coupled receptor
1038 Ion receptor (family C GPCR)
1271 Calcium sensing receptor
281 Calcium sensing receptor
1089 Small molecule receptor (family C GPCR)
1259 Neurotransmitter receptor (family C GPCR)
147 GABA-B receptor
446 Metabotropic glutamate receptor
613 Taste receptor (family C GPCR)
619 Frizzled family G protein-coupled receptor
612 Smoothened receptor (frizzled family GPCR)
620 Taste family G protein-coupled receptor
614 Taste receptor (taste family GPCR)