ABC transporter structural data

PDB querying

MHryc

PDB API

problem

I couldn't find a structure search attribute that would allow for querying by domains or transporter classification and `struct_keywords.pdbx_keywords` are not standardized.

abc_query.json

```
"parameters": {
    "attribute": "struct_keywords.text",
    "operator": "contains_phrase",
    "value": "ABC transporter"
  }
"request_options": {
  "results_content_type": [
    "experimental"
  ],

"return_type": "entry"
```

885 results

TCDb

ABC classification

Classifies ABCs into ABC1, ABC2 and ABC3 Superfamilies, contains structural data but doesn't allow to easily parse and download larger quantities.

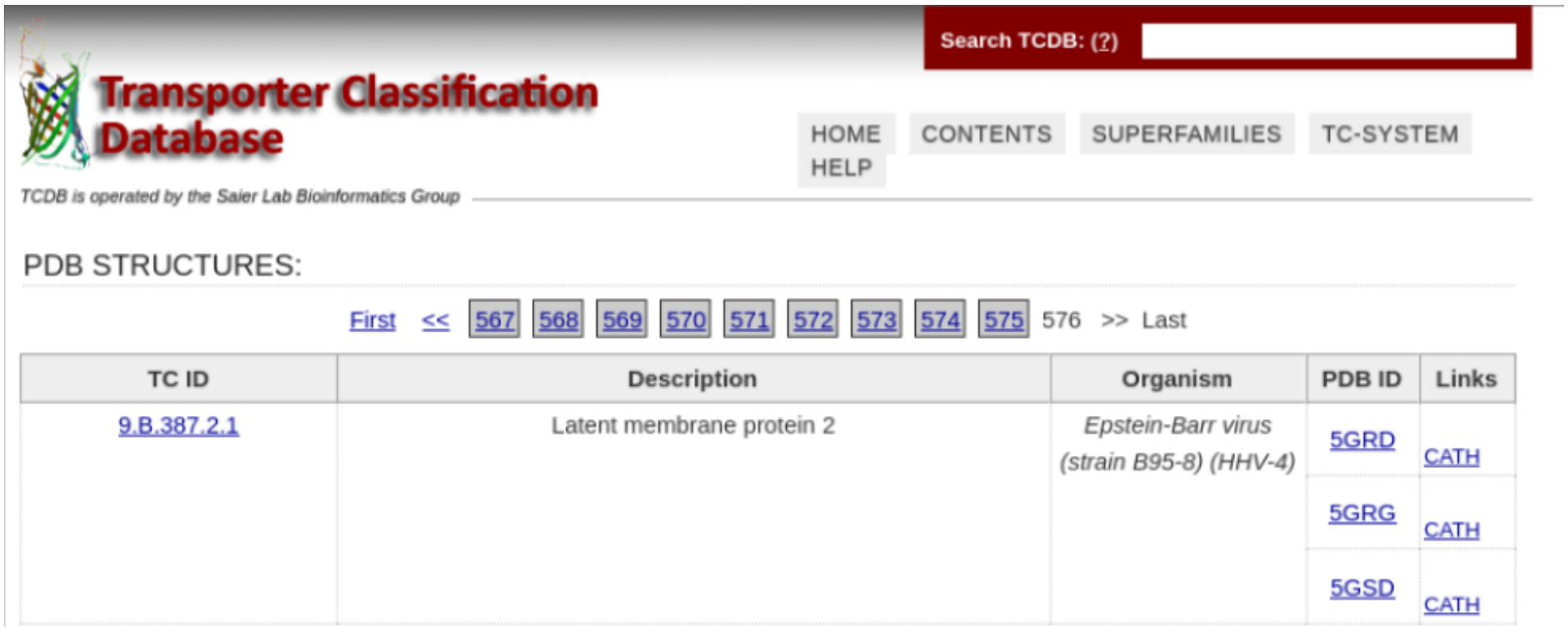


Figure 1: TCDb structural data

Foldseek & Foldmason

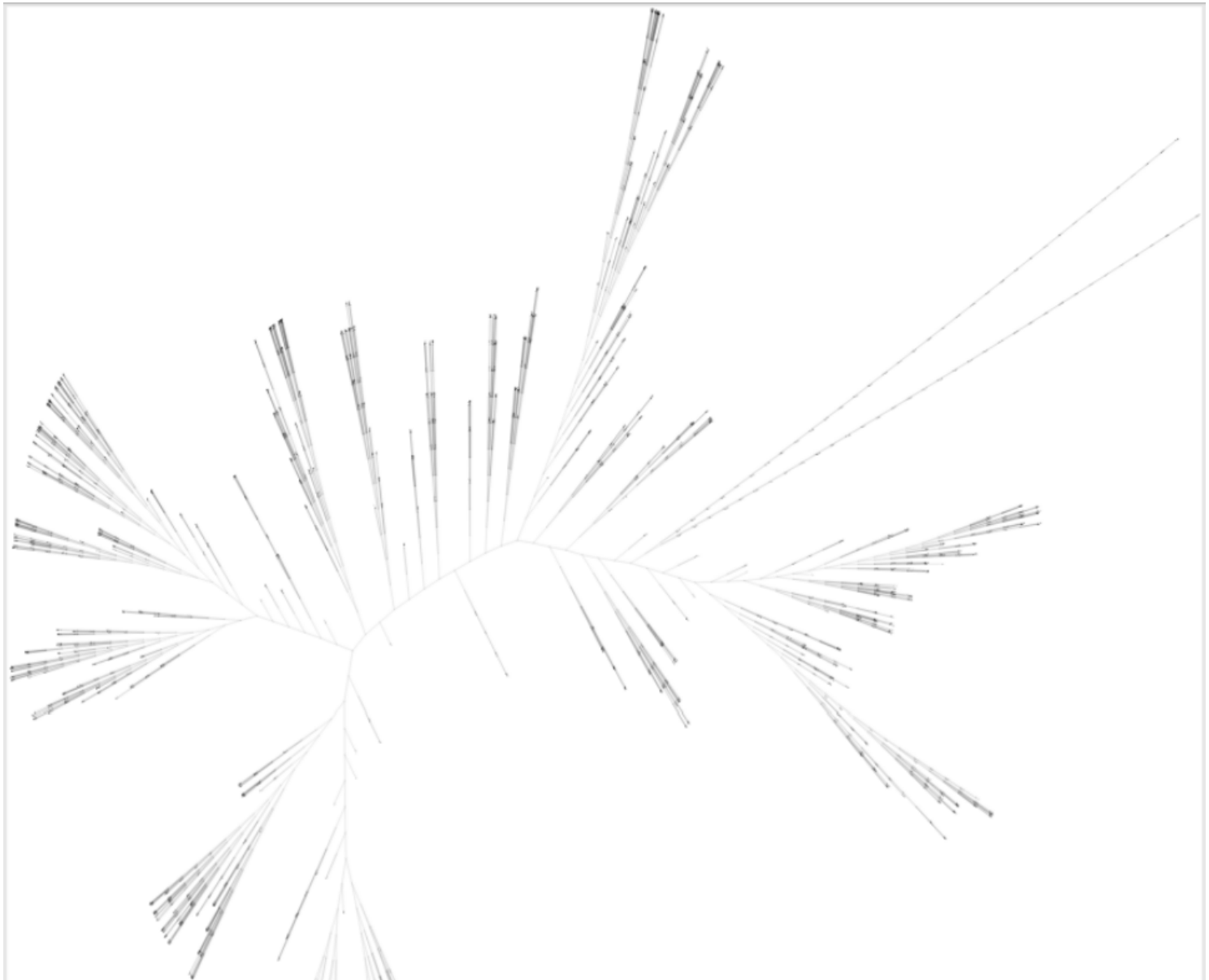
Foldseek

```
# generated by PyMOL 3.1.0
#
data_obj01
_entry.id obj01
#
_cell.entry_id obj01
_cell.length_a 73.260
_cell.length_b 95.860
_cell.length_c 109.990
_cell.angle_alpha 87.13
_cell.angle_beta  82.43
```

Using Foldseek Search Server on whole structures seems to yield too specific results but also some *noise*, eg. human NBD1 of CFTR

Clustering with Foldmason

Run with 558 structures.



ABC type representants

Type	representant
Type I	2R6G
Type II	4FI3
Type III	4HUQ
Type IV	5TV4
Type V	6AN7
Type VI	5X5Y
Type VII	5LJ7

[https://doi.org/10.1146/annurev-biochem-011520-105201]

TMD extraction

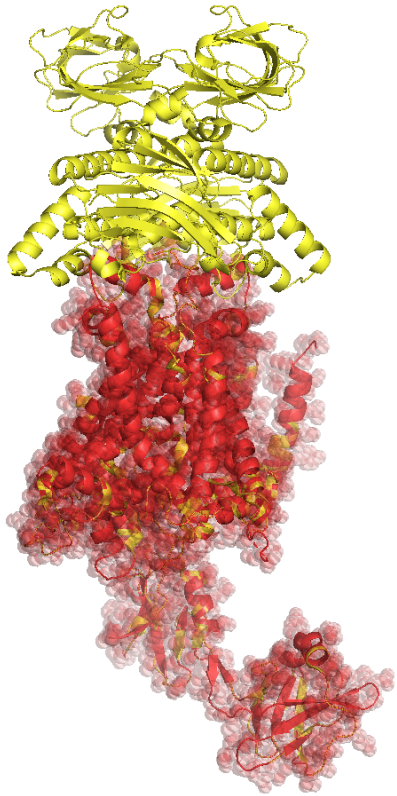


Figure 3: Type I, TMD in red

Searching based on extracted TMDs

Yields ABCs, but also ferroporins, CLC transporters, ATPases, GLUTs etc. because they share structural similarity, eg.

CFTR is a Cl– channel that bears structural and sequence homology to ABC transporters
[https://doi.org/10.1016/j.bbamem.2010.02.022]

UniProt

query

```
"abc transporter" AND (structure_3d:true)
```

845 records, 347 reviewed (SwissProt) with 1832 PDB accesion numbers after deduplication.

statistics

top 5 species (record count)

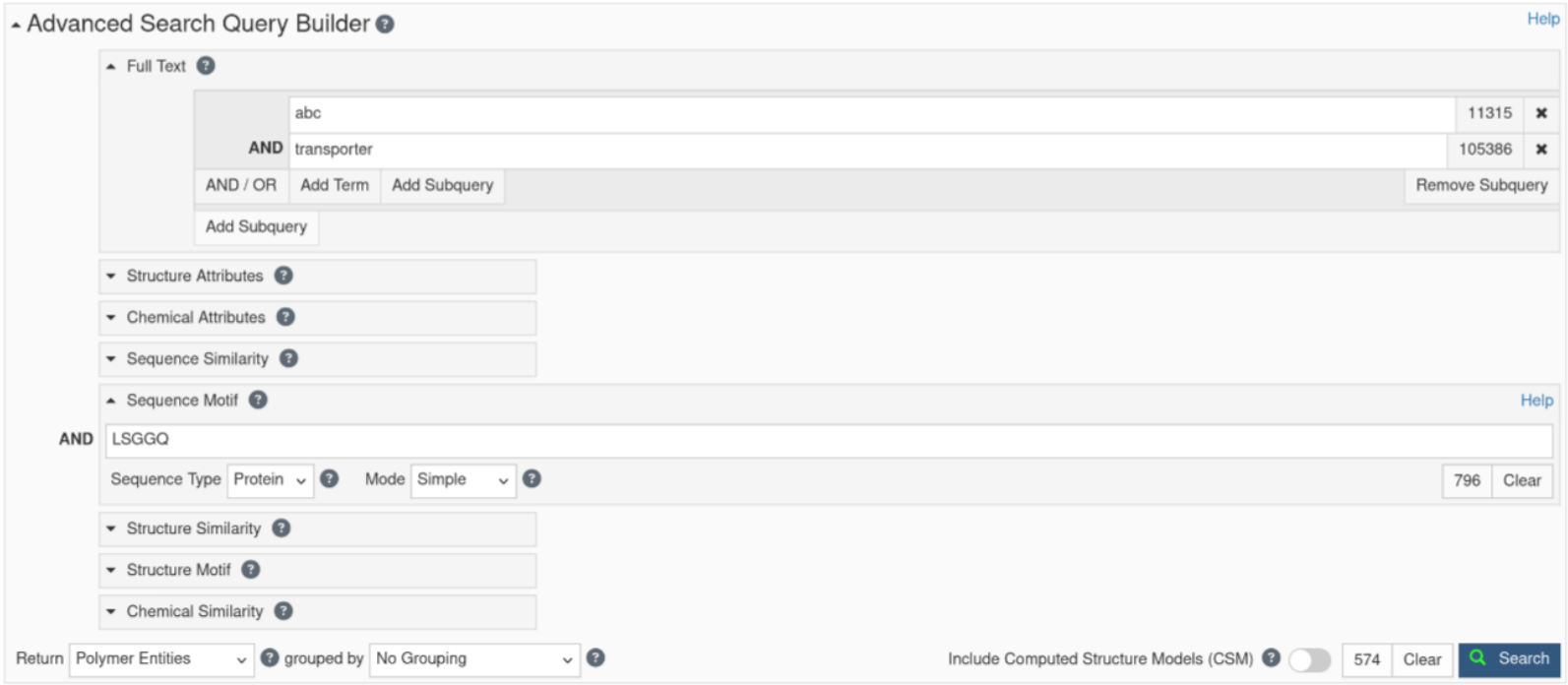
1. 86 Escherichia coli
2. 35 Homo sapiens
3. 26 Bacillus subtilis
4. 21 Mycobacterium tuberculosis
5. 16 Saccharomyces cerevisiae

top 5 species (*not exact* PDB ID count)

1. 957 Escherichia coli
2. 305 Homo sapiens
3. 80 Bacillus subtilis
4. 77 Saccharomyces cerevisiae
5. 74 Salmonella typhimurium

Advanced PDB search

query by NBD signature motif (LSGGQ)



Structure counts

574 polymer entities, but only 203 representatives if grouped by 100% sequence identity.

Final remarks

todo

easiets to hardest (?)

1. create a complete list of family or exact transporter names (eg. from HGNC, or TCDb)
2. find overlap between obtained by different methods
3. group structures by presence of ATP/ADP or other molecules
4. remove incomplete structures (subunits, “associating proteins”, ribosome bound subunits)
5. group structures by ABC type
6. determine transporter state (when not stated by authors)