Open PHACTS

Deliverable 1.5.1

Transporter Taxonomy established

Prepared by UNIVIE, EBI
Approved by UNIVIE, EBI, GSK, CD, Pfizer

August 2014
Version 1.0

Project title: An open, integrated and sustainable chemistry, biology and pharmacology
        knowledge resource for drug discovery
Instrument: IMI JU
Contract no: 115191

Start date: 01 March 2011
Duration: 60 months

Nature of the Deliverable	
Report	
Prototype	x
Other	
Dissemination level	
Public dissemination level	x
For internal use only	

Definitions

Partners of the Open PHACTS Consortium are referred to herein according to the following codes:

1 - GSK – GlaxoSmithKline – **Coordinator**
2 - UNIVIE – Universität Wien – **Managing Entity of IMI JU funding**
3 - DTU – Technical University of Denmark
4 - UHAM – University of Hamburg, Center for Bioinformatics
5 - BIT – BioSolveIT GmbH
6 - PSMAR – Consorci Mar Parc de Salut de Barcelona
601 - FIMIM – Fundacio Institut Mar d'Investigacions Mediques
602 - UPF – Universitat Pompeu Fabra
7 - LUMC – Leiden University Medical Centre
8 - RSC – Royal Society of Chemistry
801 - RSCWW – RSC World Wide Ltd
9 - VUA – Stichting VU-VUMC
10 - CNIO – Centro Nacional de Investigaciones Oncológicas
11 - UNIMAN – University of Manchester
12 - UM – Universiteit Maastricht
13 - ACK – ACKnowledge
14 - USC – Universidade de Santiago de Compostela
15 - UBO – Rheinische Friedrich-Wilhelms-Universität Bonn
16 - AZ – AstraZeneca AB
17 - Pfizer – Pfizer Limited
18 - Esteve – Laboratorios del Dr. Esteve, S.A.
19 - Novartis – Novartis Pharma AG
20 - ME – Merck
21 - HLU – H. Lundbeck A/S
22 - Lilly – Eli Lilly and Company Limited
23 - NBIC – Stichting Netherlands Bioinformatics Centre
24 - SIB – Swiss Institute of Bioinformatics
25 - CD – ConnectedDiscovery
26 - EMBL-EBI – European Molecular Biology Laboratory
27 - Janssen – Janssen Pharmaceutica NV
28 - OGL – OpenLink Group Ltd
29 - OPF – The Open PHACTS Foundation
30 - ALM – Laboratorios Almirall S.A.
31 - SciBite – SciBite Limited

1   Introduction

The first public release of the Open PHACTS Discovery Platform allowed browsing the Enzyme hierarchy to retrieve proteins from a given enzyme family as well as the associated pharmacology. However, the Enzyme hierarchy does not cover all available proteins/targets, and therefore our plan was to include additional target class taxonomies. This deliverable concentrates on the implementation of hierarchies for transporters.

In a first step, available transporter classification schemes (including mainly IUPHAR/BPS, TCDB, GO and ChEMBL) were investigated and compared to their usefulness to be included into Open PHACTS. The results of this study were published recently [1]. The ChEMBL target classification and GO are available from the Open PHACTS Discovery Platform starting with Version 2.0. For transporters, we were interested in expanding the hierarchies to information available from either IUPHAR/BPS or TCDB. To ensure sustainability the decision was made to improve the ChEMBL classification to include this information, rather than to create a taxonomy available in Open PHACTS only.

2   Choice and design of the classification

To allow the comparison of an IUPHAR/BPS or TCDB based classification scheme, a list of 1444 membrane transport proteins (human transport proteins plus all transport proteins available in ChEMBL) was compiled. We predicted the classification of these proteins which were currently unclassified in IUPHAR/BPS (500) or TCDB (604). Most of the lower levels in TCDB did not have a name, which is however a prerequisite for browsing. We created names according to the proteins contained in these groups, which might create difficulties if proteins from different organisms are to be included in the future.

Finally, IUPHAR/BPS was used as basis for the Ion channels, including only some subclasses of TCDB. For the transporters a combination of IUPHAR/BPS and TCDB was used, following TCDB for the first and second level, and afterwards using an IUPHAR/BPS based classification (including the concept of solute carriers SLC). This introduced some contradictions, which were accepted as the SLC classification is well known, thus increasing the usability. In addition, a level of auxiliary transport proteins was introduced according to TCDB class 8.A. Figure 1 shows the new classification in ChEMBL compared to the IUPHAR/BPS and TCDB classifications. Levels marked in blue originate from IUPHAR/BPS, levels in red are from TCDB.
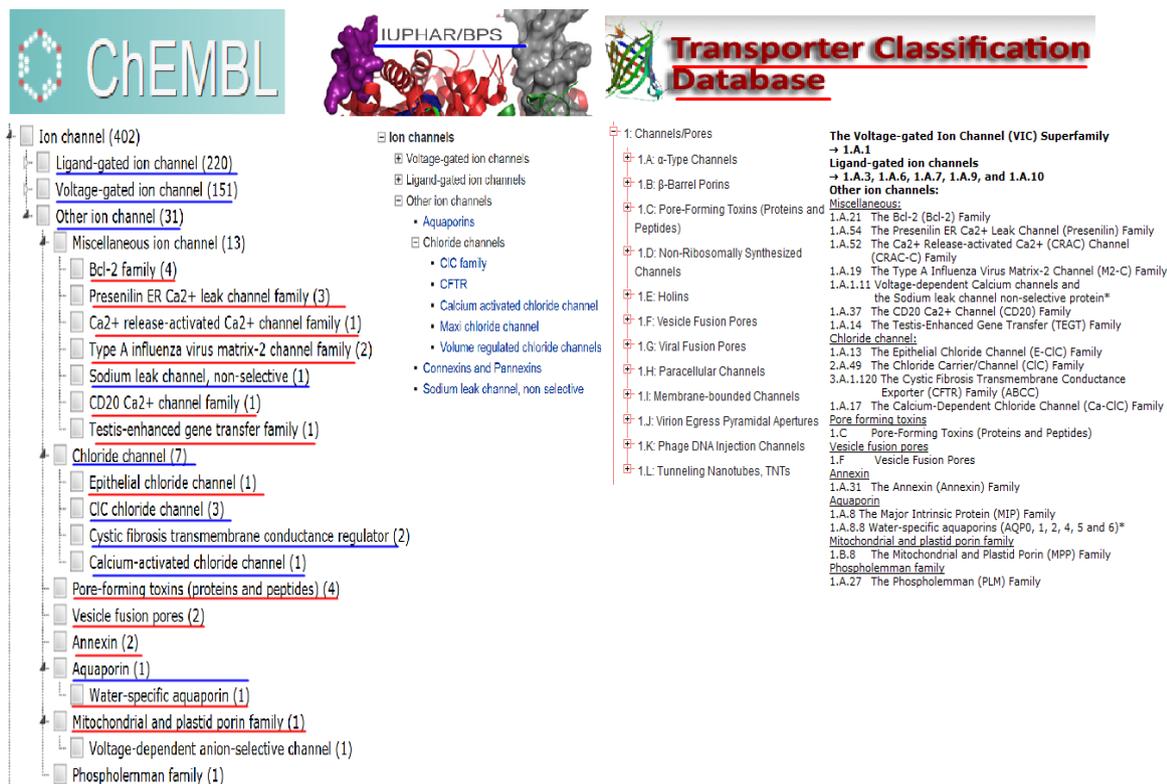

Figure 1a - Ion channel classification of ChEMBL following mainly an IUPHAR/BPS-like classification scheme.

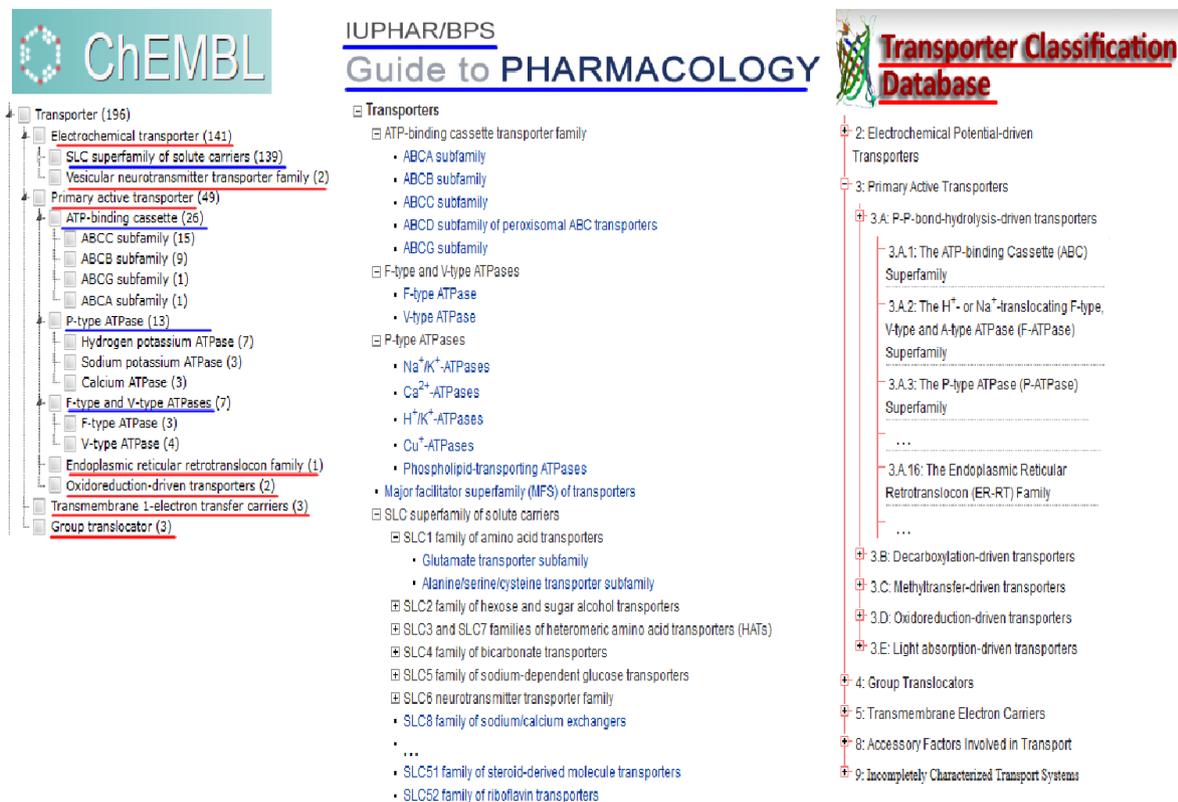Figure 1b - Transporter classification of ChEMBL, a combination of an IUPHAR/BPS-like and a TCDB-like classification scheme.

3 Implementation in ChEMBL

Following the construction of the final transporter classification scheme, the ChEMBL database schema was modified to allow storage of the new protein families and better integration with ConceptWiki. Existing protein classes for transporters and ion channels within ChEMBL were downgraded and replaced with the new classification, which includes 198 protein classes in total. Existing ChEMBL targets were re-classified according to the new classification scheme.

The new transporter classification scheme is now available within the ChEMBL database and RDF data set, can be browsed via the interface at https://www.ebi.ac.uk/chembl/target/browser. The classification is also supplied as an OBO format ontology and has been integrated into ConceptWiki, to allow searching within the Open PHACTS system.

The new classification scheme provides multiple advantages over the original classification provided by ChEMBL and used previously within Open PHACTS. Firstly, the classification is more comprehensive, and has allowed classification of existing proteins for which there was no appropriate protein class defined previously. Secondly, the classification is more easily maintained, being based on a regularly updated, community trusted, resource (Guide to Pharmacology). Thirdly, newly identified family members can more easily be classified according to sequence similarity and ion/substrate specificity rather than requiring detailed knowledge of their transport mechanism (i.e. antiporter/symporter/uniporter), as was the case previously. Finally, the new classification is intended to be more intuitive to users and has eliminated a number of redundant nodes that were present in the old scheme. The advantages of this work are supported by the fact that a total of 610 targets and 638 individual proteins are now classified as ion channels, transporters, or auxiliary transport proteins in ChEMBL (version 19), compared with only 438 targets and 470 proteins prior to implementation of the new scheme (version 17).

4 References

[1] Michael Viereck, Anna Gaulton, Daniela Digles, Gerhard F. Ecker, Transporter taxonomy – a comparison of different transport protein classification schemes, Drug Discovery Today: Technologies, Volume 12, June 2014, Pages e37-e46