

psnGPCRdb

Downloadable Data Guide

Table of Contents

Introduction.....	1
Data from PSN analysis on a single receptor structure.....	2
Data Files.....	2
info.csv.....	2
links.csv.....	3
hubs.csv.....	3
comm.txt.....	3
Ligand Shell files.....	4
Path distribution files.....	4
Ligand Shell files.....	4
metapath.csv.....	4
Plots.....	6
Path distribution plots.....	6
corrpairs_dist.png.....	6
Ligand Shell files.....	6
mpaths2d.png.....	6
3D Script Files.....	7
links.pml, links.vmd, hlinks.pml, hlinks.vmd.....	7
hubs.pml, hubs.vmd.....	7
comms.pml, comms.vmd.....	7
metapath.pml, metapath.vmd.....	7
Additional Files.....	8
Data from GPCR PSN difference.....	9
Data Files.....	9
info.csv.....	9
links_tab.csv.....	9
hubs_tab.csv.....	10
mpdiff_info.csv.....	10
mpdiff_tab.csv.....	11
Plots.....	12
Network Difference plots.....	12
3D Script Files.....	14
Additional Files.....	15
Data of Consensus networks.....	16
How to Cite.....	17
Contacting Us.....	18
Copyright and Disclaimer Notices.....	19

Introduction

This document describes the format and meaning of all files that can be downloaded from psnGPCRdb database.

All data files contained in the downloadable zip file, are named after the network they refer to (e.g. 1F88_hubs.pml). For the sake of clarity and readability in the following pages all data files will be listed without the network name prefix (i.e. paths_summary.csv instead of 1F88_summary.csv).

Almost all data files present in the zip archive are in csv format, that can be opened ed edited using common spreadsheet software (e.g. LibreOffice Calc, MS Excel and Google Sheets). Other data files are simple text file that can be opened with any text editor (e.g. Notepad, TextEdit, etc). Among output files, there are also several plots in png format that can be visualized with any image viewers and several script files that can be used to produce 3D representation of your network. The latter are present for both, [pyMol](#) and [VMD](#) molecular visualization software. Please, refer to their respective manuals for more information about how to use the scripts.

Data from PSN analysis on a single receptor structure

This section provides a detailed description of the data files available for PSN analysis carried out on a single GPCR structure and present in the downloadable compressed zip archive.

Data Files

The following sections describe the content of all data files. As stated above, almost all are csv files, that can be opened with a spreadsheet software, while those with the txt extension are simple text files.

info.csv

This csv file contains a table with some information about the network and is identical to that called “Net Summary” present in the “Summary Tables” section. In more detail, this file reports the following information:

Property	Description
NetName	Name of the network.
PDBFile	Coordinates file in PDB format.
NetFile	Network file that can be used with psnntools or WebPSN webserver.
MinFreq	Always 100, irrelevant for the user.
MinCons	Always 0, irrelevant for the user.
MergeClust	Cluster merging criterion, always “imin”.
Gly	Whether the network includes Glycine or not, always “yes”.
Imin	Network Imin.
LNodes	Number of nodes with at least one link.
Links	Number of links.
Hubs	Number of hubs (i.e. nodes with at least 4 links).
HLinks	Number of links mediated by at least one hub.
Communities	Number of communities.
CommNodes	Number of nodes involved in a community.
CommLinks	Number of links involved in a community.

links.csv

This csv file contains a table that lists all the links present in the network and the columns arrangement is almost identical to that present in the “Links” table of the “Viewer & Tables” section of the result page. The columns in this table have the following meanings:

- **N**: progressive link id, irrelevant for the user.
- **Node1, Node2**: the two interacting nodes that form the link.
- **OutValue1, OutValue2**: these columns report the conservation grades of the two nodes involved in a link.
- **Freq**: always 0 or 100, irrelevant for the user.
- **Force**: the interaction strength between the two nodes.
- **IsNode1Hub?, IsNode2Hub?**: "Yes" if the corresponding node is a hub (i.e. the node has more than 3 links), otherwise "No".
- **Clust**: the cluster of nodes the link belong to, see the online documentation page for more information about this topic.
- **Comm**: the id of the community the link belong to, otherwise 0.

hubs.csv

This csv file contains the table that lists all the hubs present in the network and the columns arrangement is almost identical to that present in the “Hubs” table of the “Viewer & Tables” section of the result page. The columns in this table have the following meanings:

- **N**: progressive link id, irrelevant for the user.
- **Hub**: the hub being considered.
- **OutValue**: the conservation grade of each hub.
- **Freq**: always 0 or 100, irrelevant for the user.
- **Force**: the average interaction strength of the links of the corresponding hub.
- **Clust**: the cluster of nodes the hub belong to, see the online documentation for more information about this topic.
- **Comm**: the id of the community the hub belong to, otherwise 0.

comm.txt

This plain text file contains information about the communities present in the network. For each community, the number of links, nodes, hubs and links mediated by at least one hub are reported. Additionally, the ratio between the number of links and nodes (L/N Ratio), the ratio between hubs and nodes (H/N Ration), hub mediated links (HL/N Ratio) and hub mediated links and all links (HL/L Ratio) are listed. Finally, for each community, a table with all community links is also present.

Ligand Shell files

For receptors bound to one or more ligands, a data file called `ligshell.csv` is also present in the downloadable zip file. These csv files report the links between each ligand and the surrounding nodes and are organized as the csv file with all the links of the network. A 2D representation of these links are also present as a png image file (`ligshell.png`).

Path distribution files

These are a set of 5 csv files all organized in the same manner and that store the same type of information and for this reason are all detailed in this section.

These tables report the distribution of some descriptors in the global pool of paths and are used to generate the corresponding plots visible on the `psnGPCRdb` database in the “Plots” section of the network page and also present in the downloadable zip archive as png files.

All tables share the same structure composed by 3 columns: the first, specific for each file, reports an increasing value of the specific descriptor, while the remaining two columns report the number and the percentage of the global/filtered pool of paths with the corresponding descriptor value. The paths distribution files are:

- **paths_length_dist.csv:** the distribution of the number of nodes in each path of the global/filtered path pool.
- **paths_force_dist.csv:** the distribution of the average interaction strength of the links in each path of the global/filtered path pool.
- **paths_corr_dist.csv:** the distribution of the average correlation between the each node and the first and last nodes in each path of the global/filtered path pool.
- **paths_corrfract_dist.csv:** the distribution of the percentage of nodes with a correlation with the first and/or the last node in each path present in the global/filtered path pool.
- **paths_hubs_dist.csv:** the distribution of the percentage of hubs in each path present in the global/filtered path pool.

Ligand Shell files

This csv file contains the % of node pairs, in the 2nd column, as a function of their correlation, in the 1st column. The corresponding image file is called `corrpairs_dist.png`.

metapath.csv

This csv file contains a detailed report of the links and nodes that contribute to the global/filtered metapath. This table is almost identical to that called “Global MetaPath” in the “Viewer & Tables” section of the network page. The columns in this table have the following meanings:

- **N:** progressive link id, irrelevant for the user.
- **Node1, Node2:** the two interacting nodes that form the metapath link.
- **OutValue1, OutValue2:** these columns report the conservation grades of the two nodes involved in a link.

- **LinkFreq:** always 0 or 100, irrelevant for the user.
- **LinkForce:** the interaction strength between the two nodes.
- **LinkRec:** the relative recurrence of the link in the filtered pool of shortest paths.
- **Node1Rec, Node2Rec:** the average recurrence of the links of each nodes.
- **IsNode1Hub, IsNode2Hub:** "Yes" if the corresponding node has more than 3 links, otherwise "No".

Plots

This section briefly describes the plot files present in the downloadable zip archive. As for other output files these plots are available for the global paths pool and, if a path filter was applied, for the resulting filtered path pool.

Path distribution plots

These plots are produced using the data present in the path distribution data files detailed in [Path distribution files](#) section. For more details about these plots refer to the description given about the corresponding data files.

corrpairs_dist.png








This plot reports the % of node pairs as a function of their correlation. The corresponding data file is called corrpairs_dist.csv.

Ligand Shell files

This is a 2D representation of the links between each ligand and the surrounding nodes.

mpaths2d.png

This is a 2D representation of the global metapath also available on the psnGPCRdb database. Regular nodes are represented as circles while hubs are drawn as pentagons. Links are represented as lines connecting two nodes. Each node is colored according to its conservation grade calculated using the ConSurf Server, while the color of each link is the average of the conservation grades of its nodes. These colors are the same used in the corresponding 3D representations present on the psnGPCRdb database and in the 3D script files. The following table reports the colors used for each conservation grades:

ConSurf Grade	n/a	1	2	3	4	5	6	7	8	9
Color										

3D Script Files

This section lists the 3D script files present in the downloadable zip archive. These scripts map the network on the receptor structure and are equivalent to those present on the psnGPCRdb database in the “Viewer & Tables” section of the network page.

In these representations each node is represented as a sphere centered on the C α carbon atom of standard amino acids and on the atom nearest to the geometric center for all other molecules present in the structure. There are scripts for both PyMol and VMD molecular visualization software. The two versions of the same script are equivalent and use the same colors and styles.

links.pml, links.vmd, hlinks.pml, hlinks.vmd










These scripts reproduce all the links present in the receptor network and their nodes. A stripped down version of these files with only those links mediated by at least one hub is also present (e.g. hlinks.pml and hlinks.vmd). As for metapaths, nodes are colored according to their conservation grades and links according to the average conservation grade of their nodes. The color scheme is the same used for the [2D metapaths](#) representation.

hubs.pml, hubs.vmd

These scripts show the hubs of the network colored according to their conservation grades.

comms.pml, comms.vmd

These scripts display communities of nodes identified in the calculated network. Each community is represented with a unique color and nodes and links belonging to the same community share the same colors. The first nine most populous communities are colored as follow:

Community	1 st	2 nd	3 rd	4 th	5 th	6 th	7 th	8 th	9 th
Color									

metapath.pml, metapath.vmd

These scripts represent methpath calculated on the global path pool. Nodes and links are colored using the same coloring scheme detailed for the [2D metapaths](#) representation representation.

Additional Files

Finally, all downloadable zip archives contain a series of additional files.

In particular, the receptor coordinates file in **PDB** format, which is needed to correctly use PyMol and VMD script files. A binary network file with extension **.psn** is also present. This file can be used to further analyze the corresponding network with our [psntools](#) program and/or [WebPSN](#) server.

A series labels files, with extension **.lab**, which are used when calculating the difference between two networks or a consensus among a pool of networks. In these circumstances it is of fundamental importance to unambiguously identify structurally equivalent residues/ligands among processed receptors. A unique identifier, called label, is then associated to these equivalent residues so that their interactions can be compared among the analyzed networks. The labels are generated from information published by [GPCRdb](#).

Finally, the conservation grades of each amino acid is stored in a text file with **.cns** extension. The conservation are calculated using [ConSurf](#) webserver

For more detail about labels files and conservation grades, refer to the corresponding [documentation pages](#).

Data from GPCR PSN difference

This section provides a detailed description of the data files of network difference divided by file type. Most of these files share some similarities with those presented in the previous sections.

Data Files

The following sections describe the content of all data files. As stated above almost all are csv files, which can be opened with a spreadsheet software, while those with the txt extension are simple text files.

info.csv

This csv file contains a three columns table almost identical to that present in the “Summary Tables” section of the result page. The first column lists a some network properties and the following two columns the corresponding values in the two analyzed networks. After the header line there are 18 rows with the following meaning:

- **Freq:** always 100, irrelevant for the user.
- **Imin:** a network specific value that indicates the lowest interaction strength needed to link two nodes. Please refer to the theory section of the online documentation for more details about this value.
- **LNodes:** the number nodes with at least one link.
- **Links:** the number of links.
- **Hubs:** the number of hubs.
- **HLinks:** the number of links mediated by hubs.
- **SpecLinks, SpecLinks%:** the number of links, and the corresponding percentage, present only in one of the two analyzed networks.
- **SharedLinks, SharedLinks%:** the number of links, and the corresponding percentage, shared by both networks.
- **SpecNodes, SpecNodes%:** the number of nodes with at least one link, and the corresponding percentage, present only in one of the two analyzed networks.
- **SharedNodes, SharedNodes%:** the number of nodes with at least one link, and the corresponding percentage, shared by both networks.
- **SpecHubs, SpecHubs%:** the number of hubs, and the corresponding percentage, present only in one of the two analyzed networks.
- **SharedHubs, SharedHubs%:** the number of hubs, and the corresponding percentage, shared by both networks.

links_tab.csv

This csv table is arranged similarly to that present in the “Links” table of the “Viewer & Tables”

section of the result page. The columns in this table have the following meanings:

- **N**: progressive link id, irrelevant for the user.
- **Node1, Node2**: the two interacting nodes.
- **Owner**: the name of the network the link belong to or "Shared" if the corresponding link is present in both networks.
- **Freq1, Freq2**: always 0 or 100, irrelevant for the user.
- **Force1, Force2**: the interaction strength between the two nodes or 0 if the link is not present.
- **OutVal1, OutVal2**: these columns report the conservation grades of the two nodes involved in a link.
- **IsNode1HubInNet1, IsNode1HubInNet2**: "Yes" if the corresponding node has more than 3 links in the first network, otherwise "No".
- **IsNode2HubInNet1, IsNode2HubInNet2**: "Yes" if the corresponding node has more than 3 links in the second network, otherwise "No".

hubs_tab.csv

This csv table reproduces that present in the "Hubs" table of the "Viewer & Tables" section of the result page. The columns in this table have the following meanings:

- **N**: progressive link id, irrelevant for the user.
- **Hub**: the hub being considered.
- **Owner**: the name of the network the hub belong to or "Shared" if is present in both networks.
- **Degree1, Degree2**: the number of links the considered hub has in the first and second network.
- **Freq1, Freq2**: always 0 or 100, irrelevant for the user.
- **Force1, Force2**: the average interaction strength of the links of considered hub in the two networks.

mpdiff_info.csv

This csv table contains a comparison summary of the metapaths calculated on the two structure networks. The first column lists a series of network properties and the following two columns the corresponding values in the two analyzed networks. After the header line there are 14 rows with the following meaning:

- **MinFreq**: always 0, irrelevant for the user.
- **MinCorr**: always 0.7. This is the minimum ENM correlation that at least one node in each path must have with the first or last node in the path. This value is fixed and is set so high to ensure a solid correlation of the atomic fluctuation among the nodes in the paths.
- **MinRec**: always 10. This is the minimum relative recurrence in the path pool needed by a link to be represented in a metapath.
- **TotPaths**: the total number of global/filtered shortest paths.

- **MPLinks**: the number of links in the global/filtered metapaths.
- **MPNodes**: the number of nodes in the global/filtered metapaths.
- ***SpecLinks, SpecLinks%**: the number of metapath links, and the corresponding percentage, present only in one of the two analyzed networks.
- ***SharedLinks, SharedLinks%**: the number of metapath links, and the corresponding percentage, shared by both networks.
- ***SpecNodes, SpecNodes%**: the number of metapath nodes, and the corresponding percentage, present only in one of the two analyzed networks.
- ***SharedNodes, SharedNodes%**: the number of metapath nodes, and the corresponding percentage, shared by both networks.


mpdiff_tab.csv

This table is almost identical to those called “Global MetaPath” in the “Viewer & Tables” section of the result page. The columns in this table have the following meanings:

- **N**: progressive metapath link id, irrelevant for the user.
- **Node1, Node2**: the two interacting nodes that form the metapath link.
- **Owner**: the name of the network the link belong to or "Shared" if the corresponding link is present in both metapaths.
- **Rec1, Rec2**: the link recurrence in both metapaths.
- **Freq1, Freq2**: always 0 or 100, irrelevant for the user.
- **Force1, Force2**: the interaction strength between the two nodes or 0 if the link is not present.
- **OutVal1, OutVal2**: these columns report the conservation grades of the two nodes involved in a link.
- **IsNode1HubInNet1, IsNode1HubInNet2**: "Yes" if the corresponding node has more than 3 links in the first network, otherwise "No".
- **IsNode2HubInNet1, IsNode2HubInNet2**: "Yes" if the corresponding node has more than 3 links in the second network, otherwise "No".

Plots

This section briefly describes the plots available for a networks difference analysis. The colors used in all plots are the same of those used on the online database and have the following meaning:

Present in	1 st network	both networks	2 st network
Color			

Network Difference plots

These plots graphically describe the difference between the two analyzed networks.

- **avg_corrfract_hist.png**: The average % of correlated residues if the shortest communication paths in the two networks
- **avg_corr_hist.png**: The average correlation of the nodes in the shortest communication paths in the two networks
- **avg_hubfract_hist.png**: The average % of hub residues if the shortest communication paths in the two networks
- **avg_len_hist.png**: The average length of the shortest communication paths in the two networks
- **corr_dist_pcn.png**: The % of shortest communication paths as a function of the correlations of their nodes
- **corr_dist.png**: The number of shortest communication paths as a function of the correlations of their nodes
- **corrfract_dist_pcn.png**: The % of shortest communication paths as a function of the fraction of correlated nodes
- **corrfract_dist.png**: The number of shortest communication paths as a function of the fraction of correlated nodes
- **corrpairs_dist.png**: The % of nodes as a function of their correlation
- **force_dist_pcn.png**: The % of shortest communication paths as a function of the average interaction strength of their links
- **force_dist.png**: The number of shortest communication paths as a function of the average interaction strength of their links
- **hlinks_hist.png**: The number of links mediated by at least one hub in each network
- **hubs_dist_pcn.png**: The % of shortest communication paths as a function of the % of their hubs
- **hubs_dist.png**: The number of shortest communication paths as a function of the % of their hubs
- **hubs_hist.png**: The number of hubs in each network
- **len_dist_pcn.png**: The % of shortest communication paths as a function of their length

- **len_dist.png**: The number of shortest communication paths as a function of their length
- **ligshell.png**: A 2D representation of the difference in the interconnectivity of each ligand in both network
- **links_hist.png**: The number of links mediated by at least one hub
- **lnodes_hist.png**: The number of nodes with at least one link in each network
- **spec_comm_hubs.png**: The number of specific and shared hubs in the two networks
- **spec_comm_links.png**: The number of specific and shared links in the two networks
- **spec_comm_lnodes.png**: The number of specific and shared nodes with at least one link in the two networks
- **totpathsdiff_hist.png**: The number of paths in both networks
- **mpdiff2d.png**: A 2D representation of difference in the two global metapath

3D Script Files

This section briefly lists the 3D script files present among the network difference data files. These scripts compare links, hubs and global metapaths: linksdiff, hubsdiff, mpdiff, respectively. There are two versions for each 3D script, for both PyMol and VMD molecular visualization software and all use the same colors based on conservation grades detailed in previous sections.

Additional Files

As for single receptor networks, a series of pdb, psn, lab and cns files are present in the downloadable zip files. See the corresponding [section](#) for more detail.

Data of Consensus networks

All data files, plots and 3D scripts of consensus networks shares the same names, format and organization of those produced in a single network analysis. Please refer to those sections for more details.

How to Cite

Thank you for using PSNTools, we really appreciate it.

Please, remember to cite at least one of the following papers in all published works which utilize this software:

NEW_PAPER_CITATION_HERE

Angelo Fellingine, Michele Seeber and Francesca Fanelli
webPSN v2.0: a webserver to infer fingerprints of structural communication in biomacromolecules
Nucleic Acids Res, Web Server Issue, 19 May 2020
<https://doi.org/10.1093/nar/gkaa397>

Contacting Us

If you have any questions or if you encounter any problems with this software please do not hesitate to contact us at the following email addresses:

@ [Prof. Francesca Fanelli](#)

@ [Angelo Felling](#)

Copyright and Disclaimer Notices

PSNTools is copyright 2017-2020 the University of Modena and Reggio Emilia (Italy).

PSNTools is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 3 of the License, or (at your option) any later version.

PSNTools is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with Wordom. If not, see <http://www.gnu.org/licenses>