OPMxplore

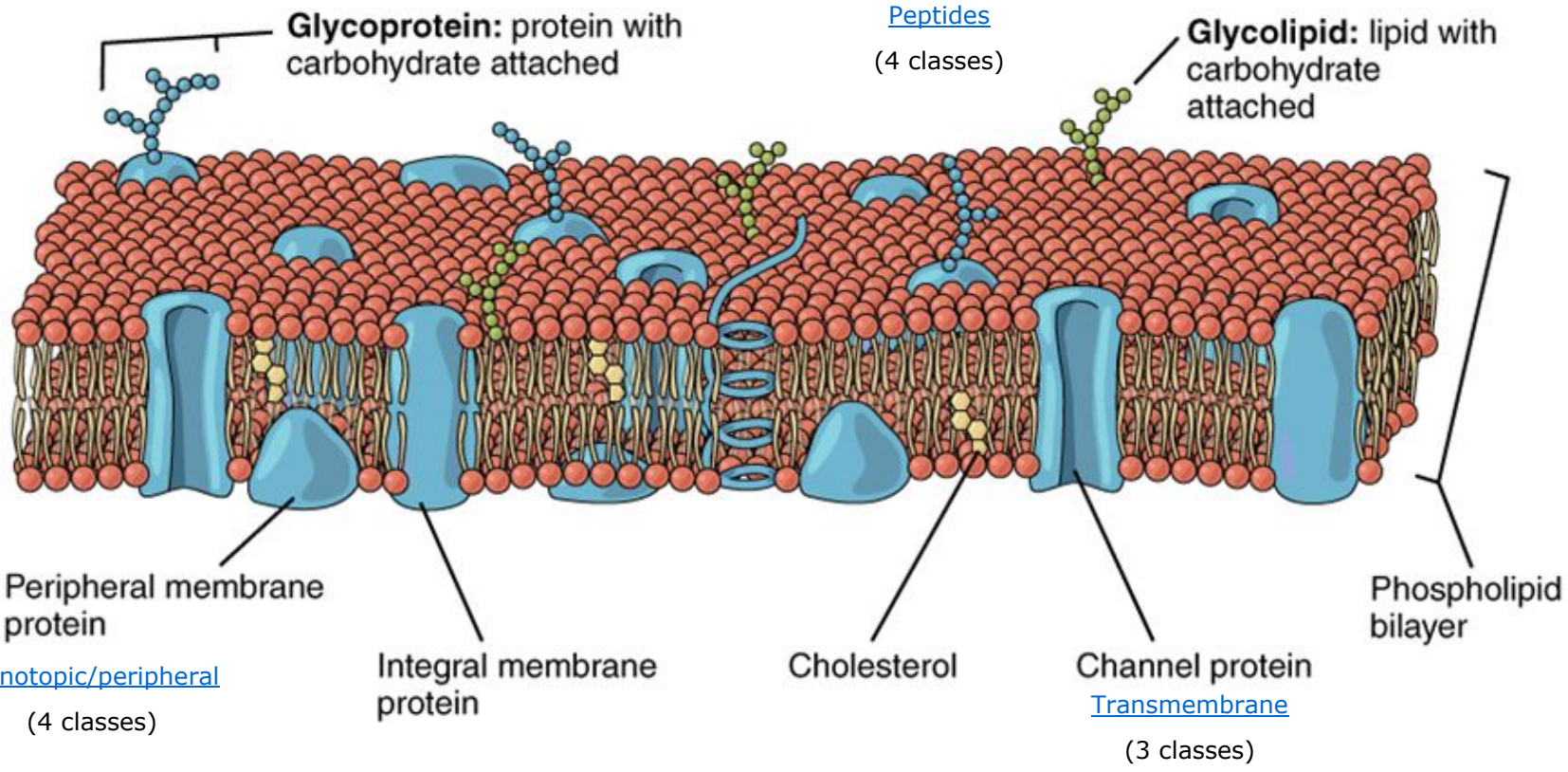
Sinduja, David and Felcy

Agenda

- Background
- Mission statement
- Data
- Users and Use Cases
- Project Structure
- Design
- Demo
- Lessons learned and future work

Background

Membrane Proteins



Protein Data Bank

Orientation of Membrane Proteins (OPM)

opm.phar.umich.edu

orientations of proteins in membranes (OPM) database

UNIVERSITY OF MICHIGAN | COLLEGE OF PHARMACY | LOMIZE GROUP

Search OPM

PDB ID or protein name

HOME | ABOUT OPM | DOWNLOAD OPM FILES | CONTACT US | PPM SERVER | LIPID COMPOSITION ATLAS

Protein Classification

- **Types** (3 types)
- **Classes** (11 classes)
- **Superfamilies** (474 superfamilies)
- **Families** (884 families)
- **Species** (745 species)
- **Localization** (24 types)
- **All proteins in OPM** (3622 proteins)

Orientations of Proteins in Membranes (OPM) database

OPM provides spatial arrangements of membrane proteins with respect to the hydrocarbon core of the lipid bilayer.

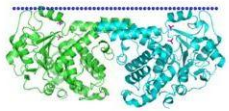
OPM includes all unique experimental structures of transmembrane proteins and some peripheral proteins and membrane-active peptides (Features).

Each protein is positioned in a lipid bilayer of adjustable thickness by minimizing its transfer energy from water to the membrane (Methods).

OPM provides structural classification and sorting according to different criteria (Classification).

Our calculations are in agreement with experimental studies of 24 transmembrane and 39 peripheral peptides and proteins (Comparison with Experiments).

For more information on single-spanning transmembrane proteins please see our Membranome database



4y7i » Myotubularin-related protein 8,

All proteins in OPM (3622 proteins)

Family	Protein Name	PDB ID	Species	Localization	Num. Subunits	Num. TM Sec. Structs.	Hydrophobic Thickness or Depth (Å)	Tilt Angle (°)	$\Delta G_{transfer}$ (kcal/mol)
1.1.01.01.	Archaerhodopsin-2, trimeric	2ei4	*Halobacterium sp.*	Archaebac.	3	21	30.5 ± 1.2	0 ± 1	-143.9
1.1.01.01.	Archaerhodopsin-1	1uaz	*Halobacterium sp.*	Archaebac.	1	7	31.8 ± 1.3	9 ± 2	-65.3
1.1.01.01.	Bacteriorhodopsin	5azd	*Thermus thermophilus*	Bact. Gram-neg inner	4	28	30.0 ± 0.5	3 ± 1	-121.3

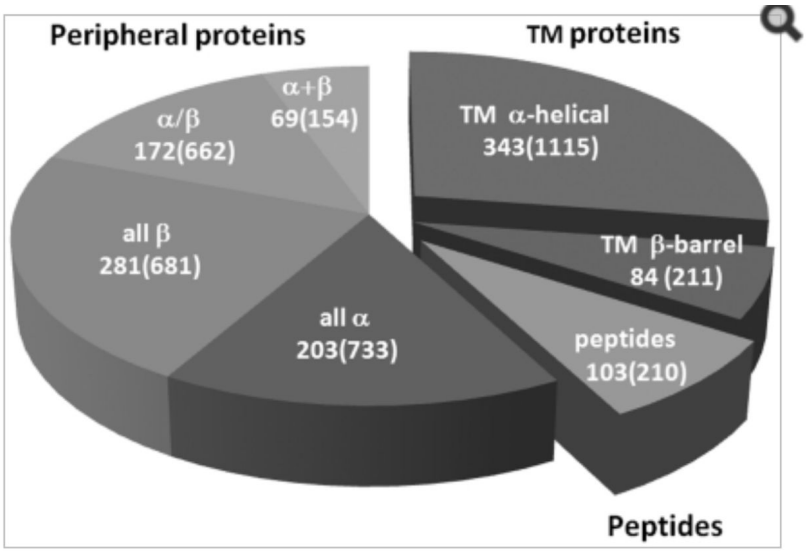
Mission Statement

Develop tools to:

- Query the PDB/OPM database
- Identify relationships between different categories of membrane proteins
- Interactively visualize the results

DATA

1. RCSB-PDB (135,787 experimentally determined protein structures)
2. OPM dataset (3622 membrane proteins)



Current distribution of different OPM entry types (as of 20 July 2011).

User and Use Cases

- Educational setting:
  - Interactive graph to intuitively explore properties of proteins of interest.

- Research setting:
  - Researchers with basic python/jupyter notebook knowledge can ***query*** OPM database for computationally and experimentally derived measures and ***view*** statistical plots.

- User Interface : **Jupyter notebook** and visualizations on **Plotly Dash**.

Design

Components Design

1. Convert database into str searchable format
    a. Python: Sqlite3, pandasql, pypdb
    b. Functions: load_data(), sql_query()
    c. Specs: programmatic access to data and custom query

2. Interactive graphs based on database query
    a. Python: Plotly, seaborn, ipywidgets
    b. Functions: query_hash(), make_plot_*type*()
    c. Specs: guided statistical plots and interactive discovery

Components

1.Searching and displaying desired proteins:

- This will let the user search for desired list of membrane protein(s), usually with a generic name such as "channels" from PDB and match with the list of membrane proteins on OPM. The user will get a data frame with details such as PDBIDs, names of proteins, location of the membrane protein, membrane depth, tilt and transfer energy from water to the membrane.

2.Displaying data visualization:

- This will let the user view the basic statistical features of the desired group of proteins obtained in component 1.

3. Displaying 3D protein Visualization:

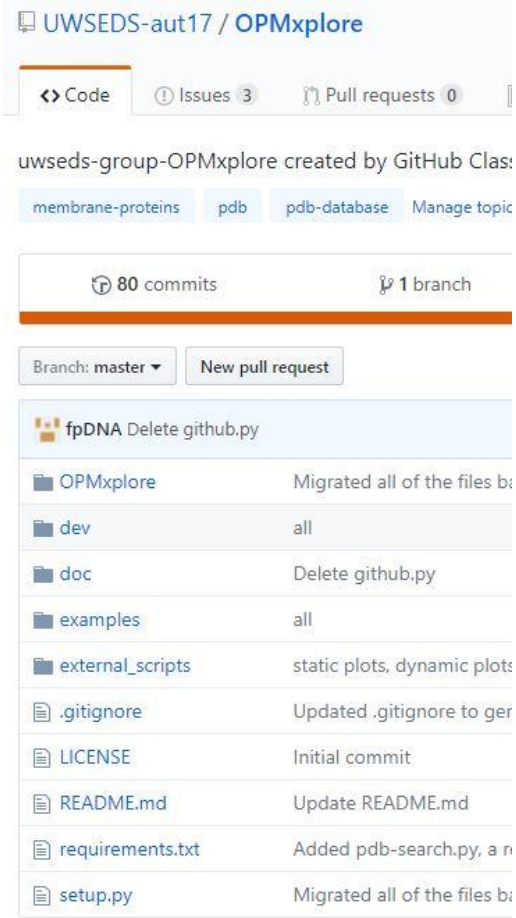
- This will let the user visualize 3D structure of a selected membrane protein from the list obtained in component 1.

4. User Interface:

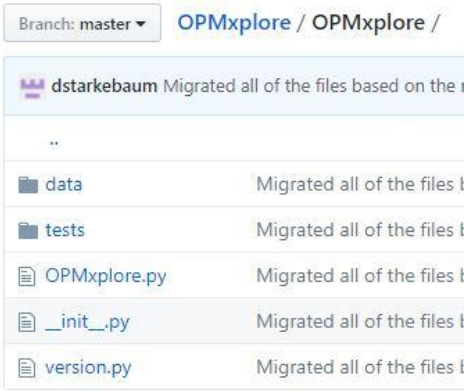
- Building Plotly Dash to allow the user visualize the results from components 2 and 3.

Project Structure

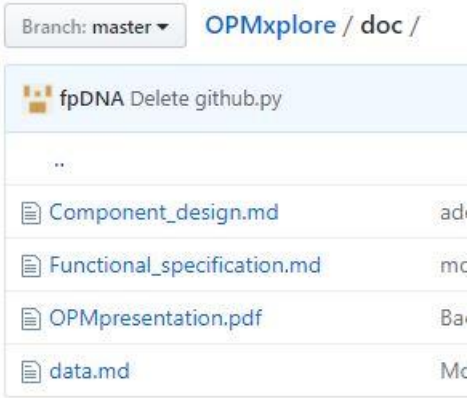
GitHub Repository



Python package folder



Doc folder



Required packages:

Standard tools:
numpy
scipy
matplotlib
pandas
sqlite3

Specialized analysis:
pypdb
biopython
biopandas

Visualization:
nglview
plotly
seaborn
ipywidgets
jupyter-notebook

Demo

Video of the demo: https://youtu.be/8AhEcPVn3ac

Lessons learned and Future Work

Lessons:

- ★ Cyclic Development Process
- ★ Data Munging
- ★ Git Branches

Road Map:

- ★ Stand Alone app for intuitively converging on proteins of interest
- ★ Integrate more computational tools to add to the database
  - ○ Pore profile for channel like proteins

Thanks!