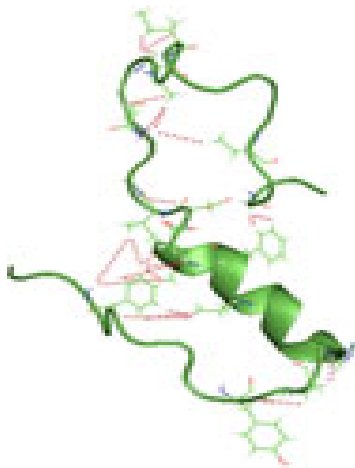# Tutorial: Protein analysis using the GlyGen

Nuerye Ainiwan

Professor Raja Mazumder

September 10, 2019

The George Washington University

Introduction

The purpose of protein analysis is to understand the data types and information about specific protein-encoded genes or to map to typical protein sequences that represent the gene and related information.

Protein analysis using the GlyGen

i) What is GlyGen?

According to the GlyGen (2019), "GlyGen is an international project funded by The National Institutes of Health Computational and Informatics Resources for Glycoscience. GlyGen is a data integration and dissemination project for carbohydrate and glycoconjugate related data."

ii) GlyGen portal Homepage https://glygen.org. Its navigation bar contains

EXPLORE: drop-down menus include Glycan, Protein, Glycoprotein searches by choosing simple or advanced search.

QUICK SEARCH AND TRY ME: can answer user-specific questions. For example, what are the functions of protein X?

DATA: GlyGen data can be accessed from the Data page.

HELP: drop-down menus include Resources, Feedback, Contact us, About.

### iii) Using GlyGen to do Protein analysis

My biological question: What chemicals are used to detect amyloid-beta in Alzheimer's disease?

In order to find the answer, I will use the following steps.

step-by-step instructions

1.Getting the targeted protein

Go to **Homepage https://glygen.org -> click on** EXPLORE -> select Protein from drop-down menus, or you can directly click on Protein Explore on the interface. (Figure 1)
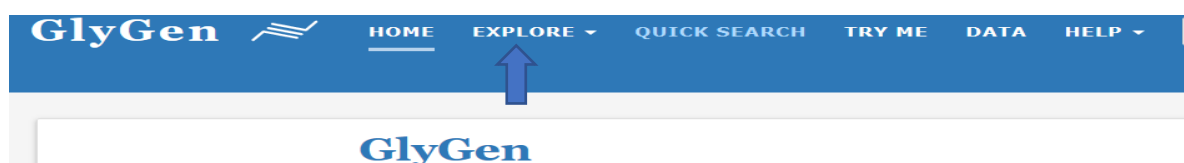
(Figure 1)

You'll be allowed to go to the Protein Search interfaces. It includes Simple Search, Advanced Search, and Tutorial. (Figure 2)

(Figure 2)

->select protein from the drop-down category and type in (**amyloid-beta**) and hit Search. ->This will allow you to see the Summary of your Protein Search. (Figure 3)
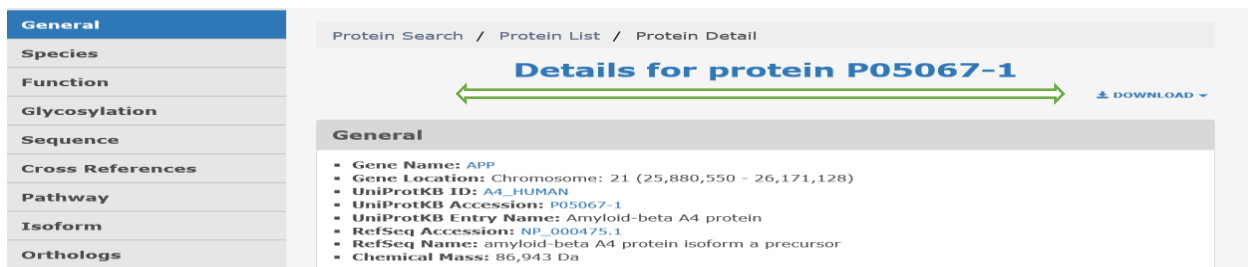

(Figure 3)

2. Getting the targeted protein details information

Blow the "Summary of your Protein Search," there are Protein List. You can find the protein you want and click the UniProtKB Accession number. (Figure 4) (For example, I find the organism is Homo sapiens, and UniProtKB Name is Amyloid-beta A4 protein, and UniProtKB Accession Number is P05067-1.)->click UniProtKB Accession number P05067-1. You will get the Details for protein P05067-1. (Figure 5)

| UniProtKB Accession	Gene Name	UniProtKB Name	Chemical Mass (Da)	Organism	RefSeq Name	RefSeq Accession
P12023-1	App	Amyloid-beta A4 protein	86722	Mus musculus	amyloid-beta A4 protein isoform 1 precursor	NP_001185752.1
P05067-1	APP	Amyloid-beta A4 protein	86943	Homo sapiens	amyloid-beta A4 protein isoform a precursor	NP_000475.1

(Figure 4)

(Figure 5)

3. Getting the chemicals information which is used to detect amyloid-beta in Alzheimer's disease.

Once you click UniProtKB Accession P05067-1->you will be navigated to uniprot.org website->https://www.uniprot.org/uniprot/P05067-1/-> click Entry menus. Under an Entry menu there is a list of sub-menus(which can be check or uncheck) ->click Pathology & Biotech. (Figure 6)
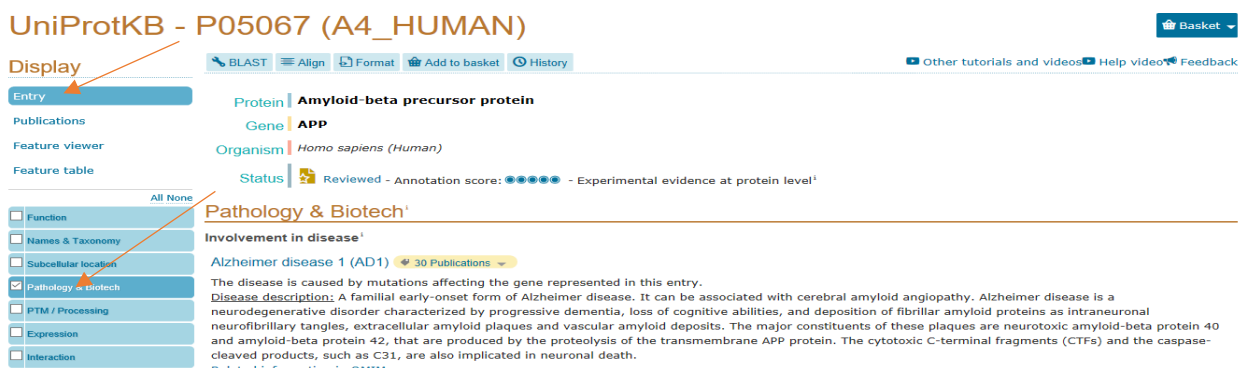
(Figure 6)

(Figure 7)

You can get access to the Chemistry databases, which include ChEMBL and Drugbank. Under the Drugbank menu, there is a list of Chemistries that can be used to detect amyloid-beta in Alzheimer's disease. (Figure 7) (For example, I click one of the chemistries and find detail information about Flutemetamol (18F) and references.) (Figure 8)

(Figure 8)

Critique the resource

Problems found during the search steps:

1. There is not enough information on a particular topic.

In order to use the disease name to find the corresponding protein that caused the disease, I tried to use the search category as disease and search term as Alzheimer's disease. However, there was only four protein, and I cannot find the gene name APP protein I need in the protein list.

(Figure 8)

2. The existing content was created without an organized and documented content strategy. For example, in order to find protein (amyloid-beta) information, I tried to use the search category as protein and search term (amyloid-beta), but 91 proteins were found. In the absence of a further screening process, this makes it difficult for the protein seeker.

How I will improve it:

1. To make it easier to find and improve the efficiency of finding, I may add filtered programs. For example, the protein can be filtered by disease, organisms, gene ontology.

2. When I tried to use the search category as disease and search term as Alzheimer's disease, I did not find the gene name APP protein I need in the protein list. So I may use a more comprehensive data integration platform to handle a variety of different types of data.

Reference

Alzheimer's Association Reports. (2016). 2016 Alzheimer' Disease facts and figures. Alzheimer's & Dementia: *The Journal of the Alzheimer's Association, 12*(4), 459–509. Retrieved from https://dot.org/10.1016/j.jalz.2016.03.001

Johnson, K., Minoshima, S., Bohnen, N., Donohoe, K., Foster, N., Herscovitch P.,… Thies, W. (2013). Appropriate use criteria for amyloid PET: A report of the amyloid imaging task force, the society of nuclear medicine and molecular imaging, and the Alzheimer's association. *The Journal of Nuclear Medicine, 54*, 476–490.

Rabinovici, G., Constantine, G., Charles, A., Kiran, C., Ilana, G., Lucy, H.,…Maria, C. (2019). Association of amyloid positron emission tomography with subsequent change in clinical management among medicare beneficiaries with mild cognitive impairment or dementia. *JAMA, 321*, 1286-1294. https://doi:10.1001/jama.2019.2000