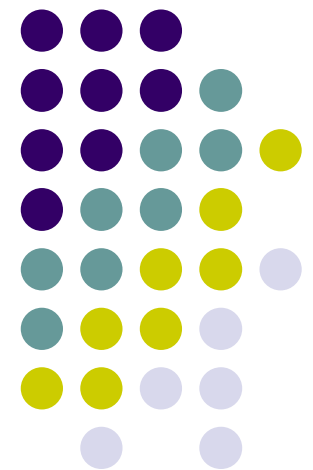


Project Descriptions

BINF5240
Lecture 18



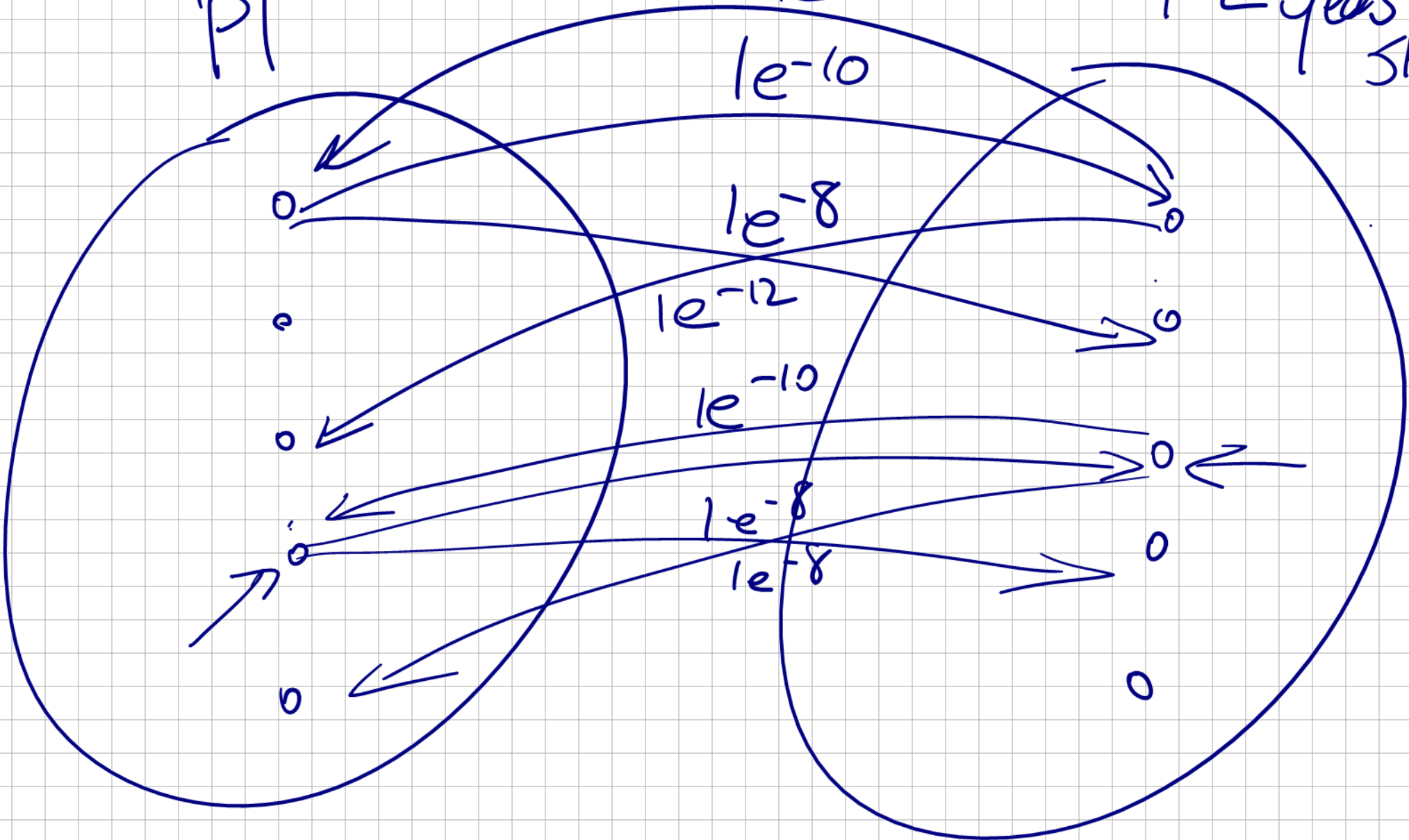
Class Project: Blast Database



1. Write a program that computes all pairwise blast alignments for two species' proteomes and stores the alignments in a relational database.
2. Write a program that retrieves the blast alignment for two proteins (specified by their accessions) from the relational database.
3. Write a program that *all* finds pairs of orthologous proteins that are mutually best hits in the species' proteomes.

P1 dros 8K

P2 yeast 5K



$1e^{-10}$

$1e^{-10}$

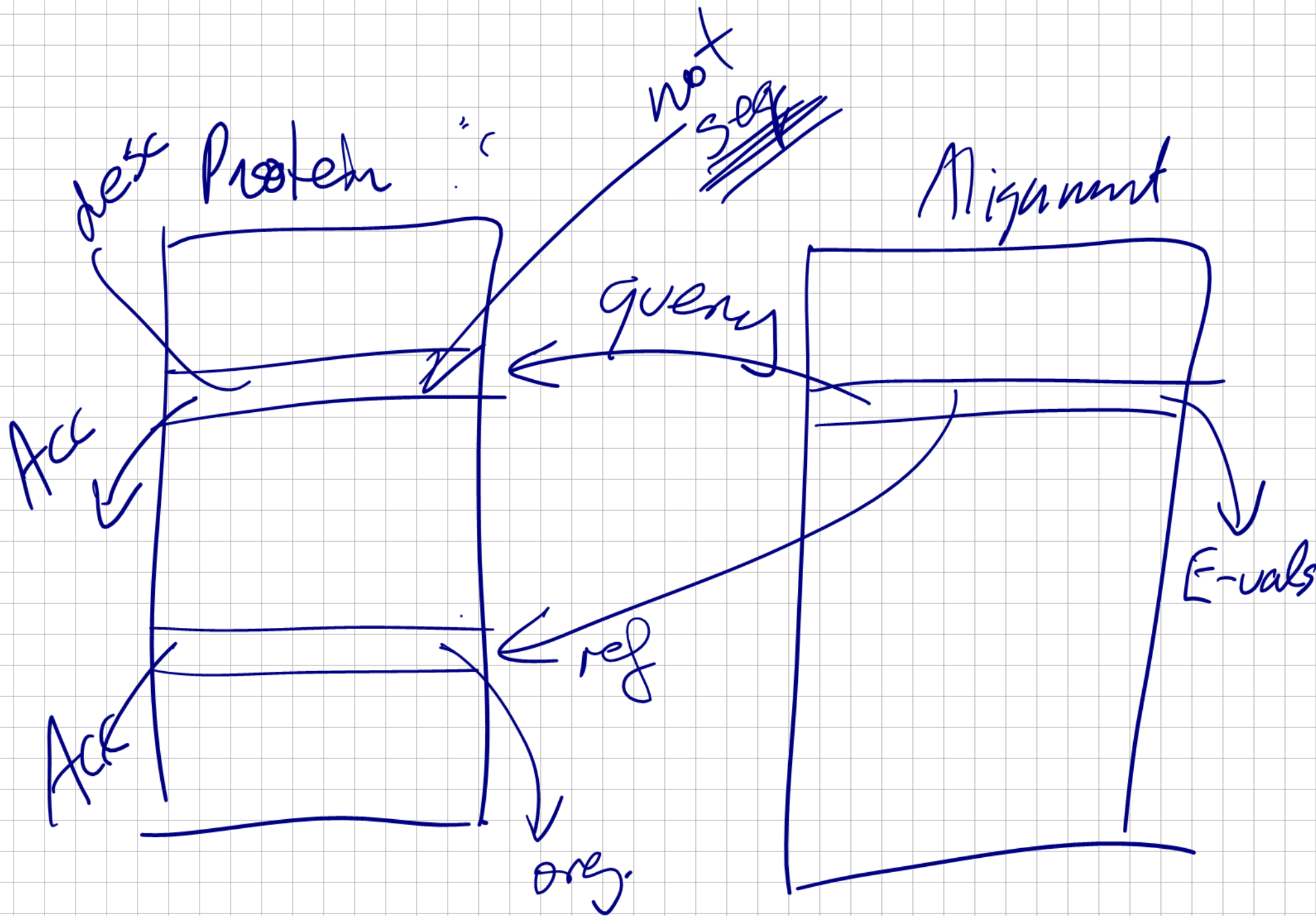
$1e^{-8}$

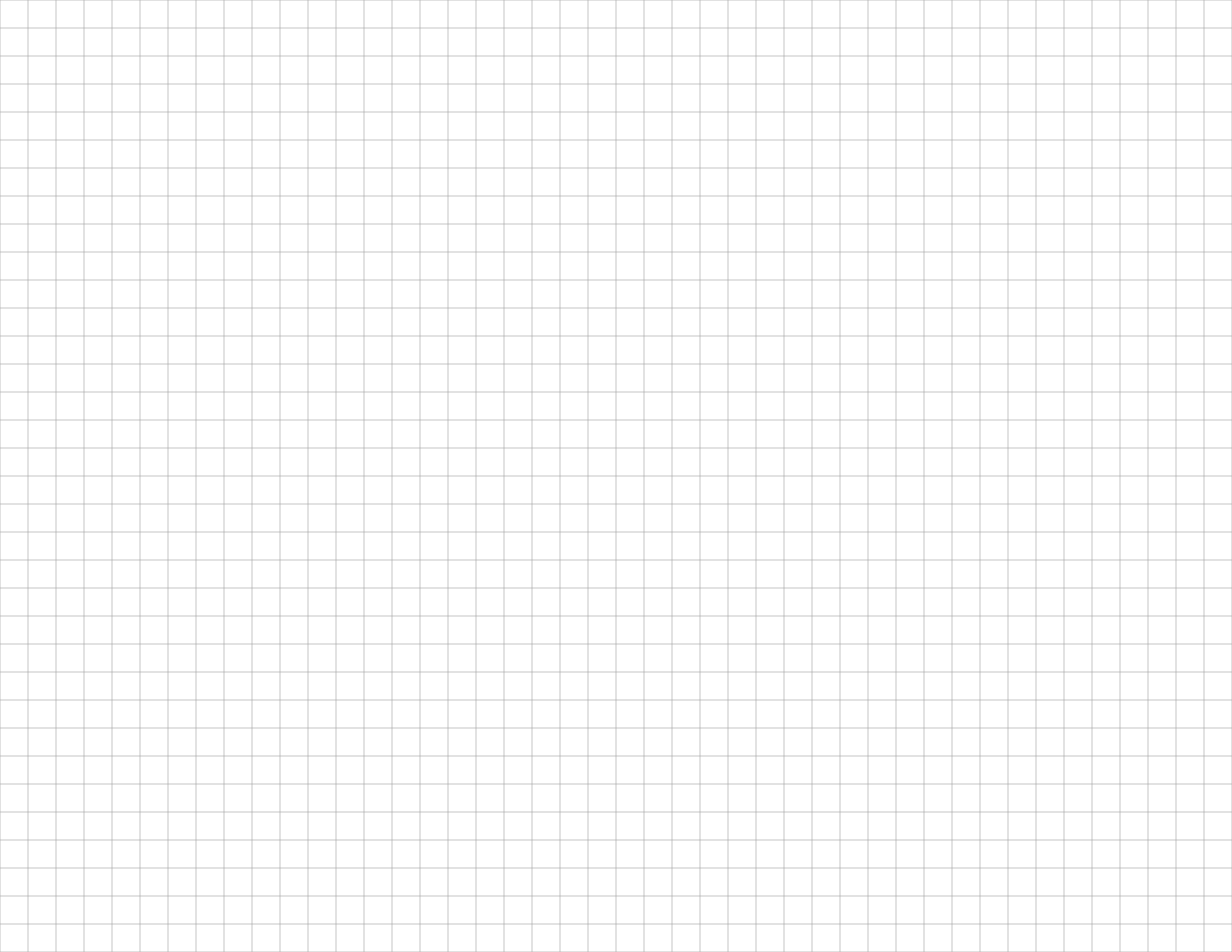
$1e^{-12}$

$1e^{-10}$

$1e^{-8}$

$1e^{-8}$



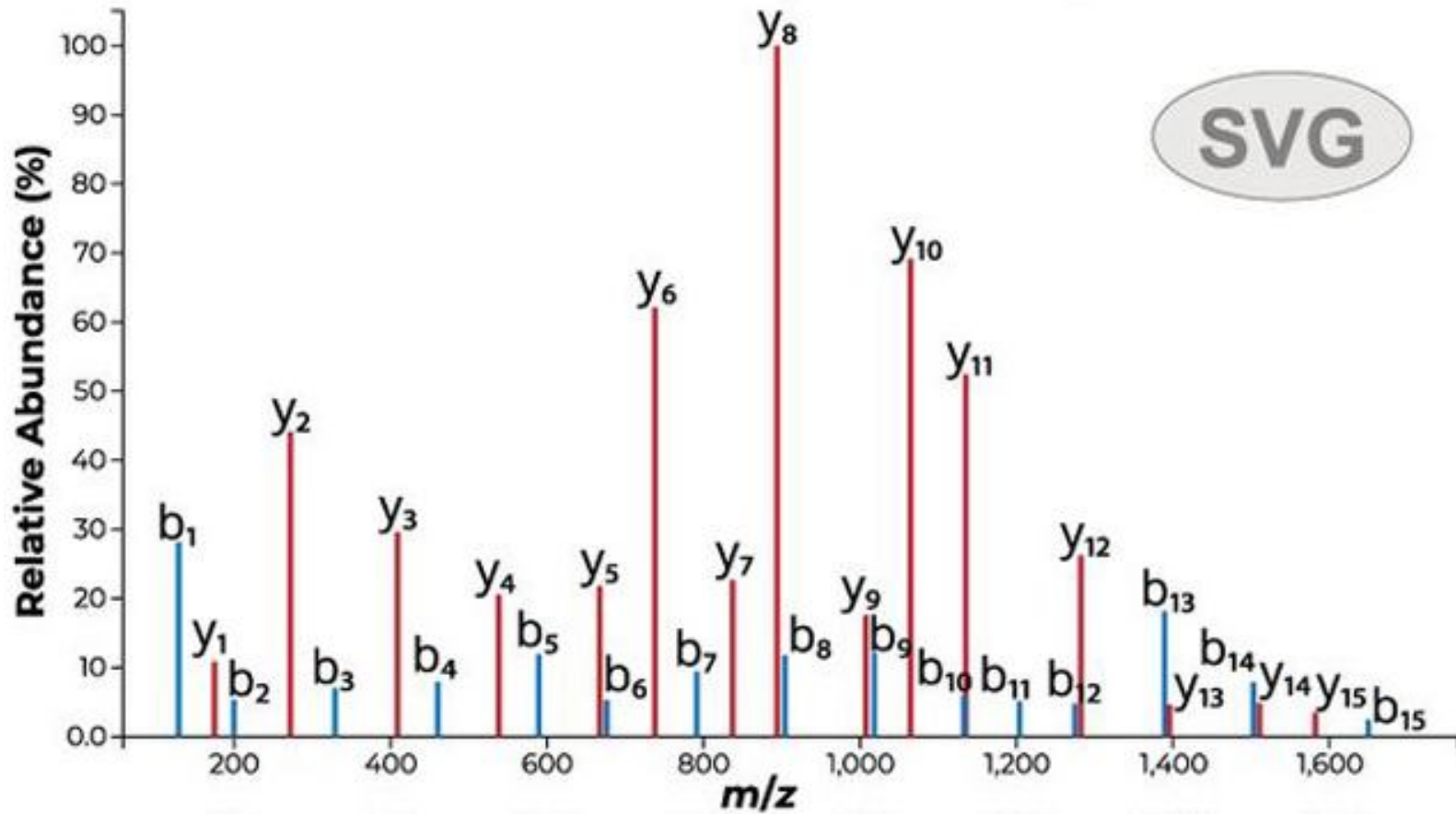


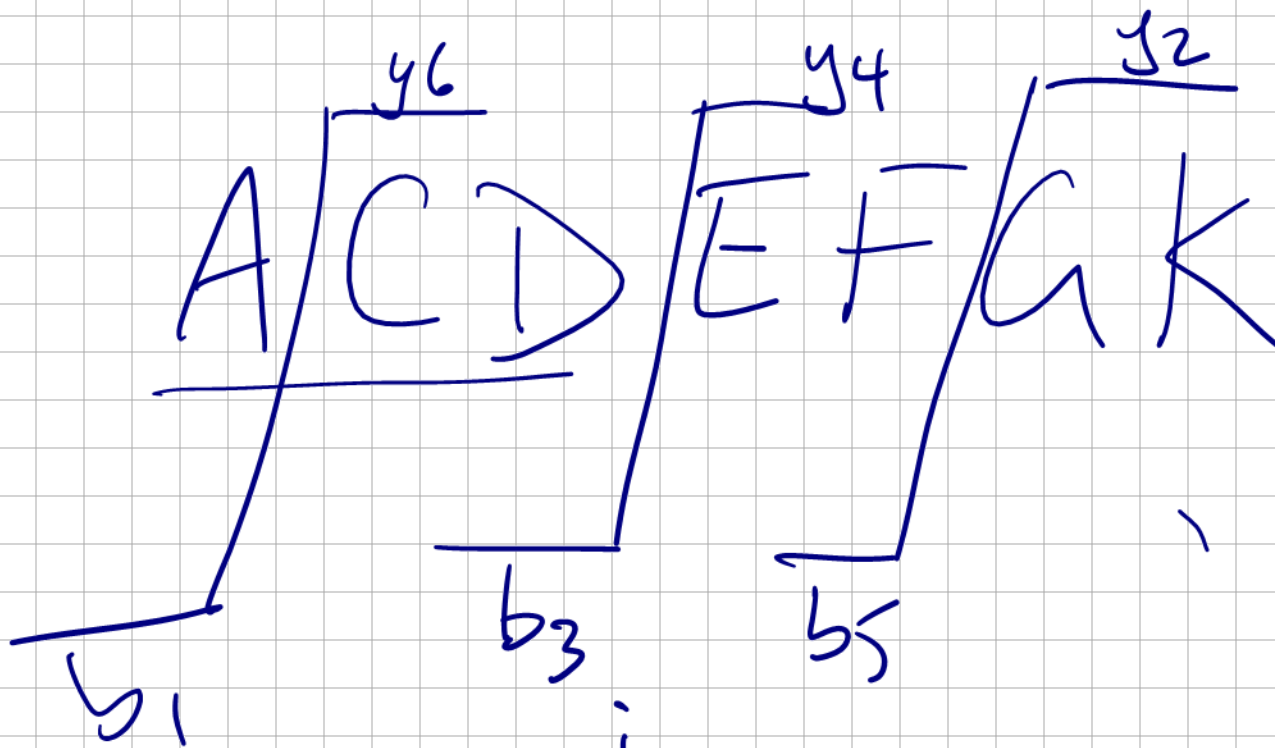
Class Project: MS/MS Viewer



- Write a program to display peptide fragmentation spectra from an mzXML file.
 - The program will take an mzXML file, a scan number, and a peptide sequence as input.
 - The peptide's b-ion and y-ion m/z values should be computed, and peaks matching these m/z values annotated with appropriate labels.
 - The output figure/plot should aid the user in determining whether or not the peptide is a good match to the spectrum.

Example of annotated spectrum

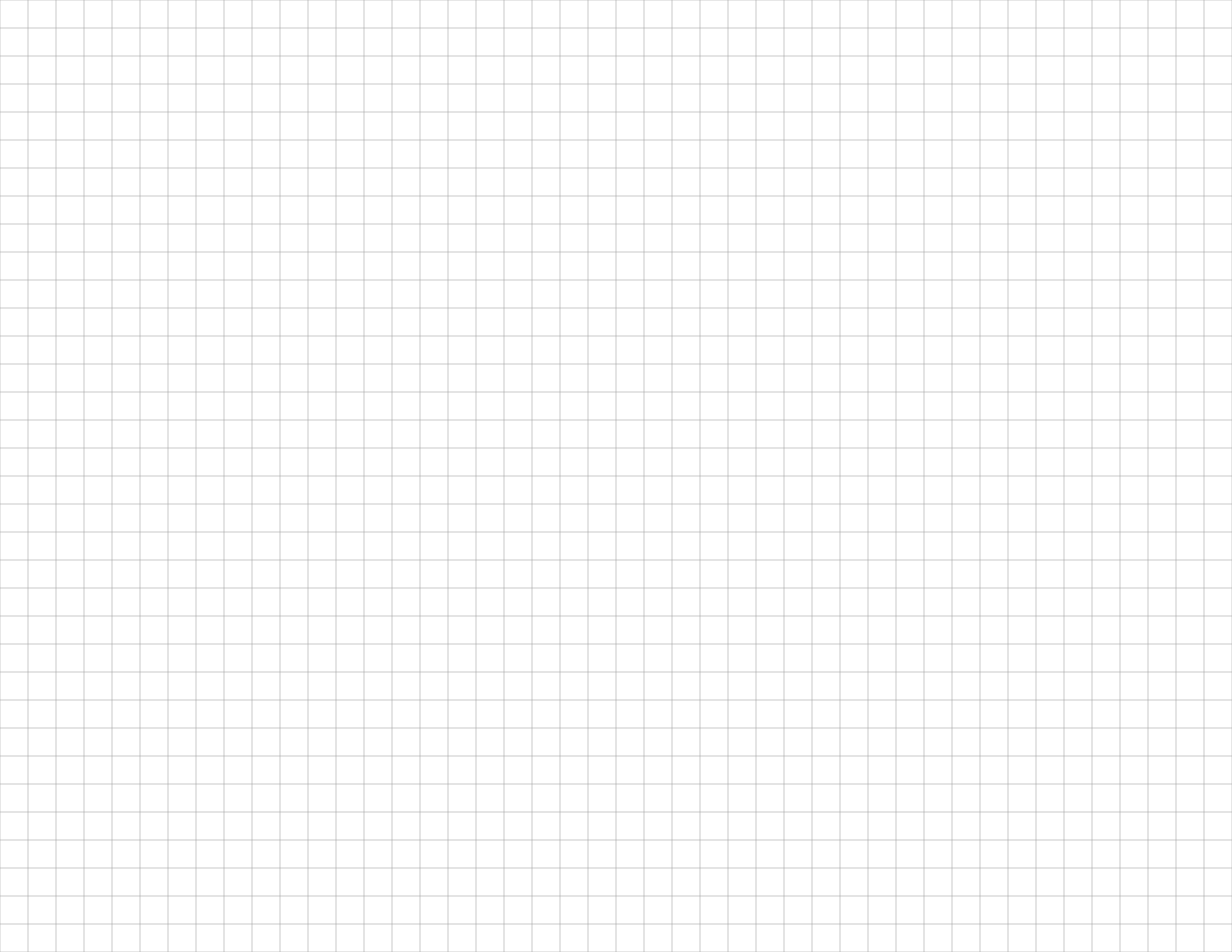




$$mw_G = 57$$

$$b_i = \sum_{j=1}^i mw_{AA_j} + 1$$

$$y_i = \sum_{j=1}^i mw_{AA_{(n+1-j)}} + 19$$







Class Project: Protein Digest

- Write a simple web-server application using TurboGears to carry out an *in silico* enzymatic digest of a user-provided protein sequence.
 - Users should be able to specify min and max length, min and max molecular weight, # of missed cleavages, and specific enzyme.
 - Output should be a table of peptides, with their length, molecular weight, # of missed cleavages, and amino-acids to left and right of each peptide in the protein sequence.

at least 5!

Example: PeptideMass



PeptideMass

PeptideMass [\[references\]](#) cleaves a protein sequence from the UniProt Knowledgebase (Swiss-Prot and TrEMBL) or a user-entered protein sequence with a chosen enzyme, and computes the masses of the generated peptides. The tool also returns theoretical isoelectric point and mass values for the protein of interest. If desired, PeptideMass can return the mass of peptides known to carry post-translational modifications, and can highlight peptides whose masses may be affected by database conflicts, polymorphisms or splice variants.

[Instructions](#) are available.

Enter a UniProtKB protein identifier, ID (e.g. ALBU_HUMAN), or accession number, AC (e.g. P04406), or an amino acid sequence (e.g. 'SELVEGVIV'; you may specify post-translational modifications, but [PLEASE read this document first!](#)):

the fields. the cleavage of the protein.

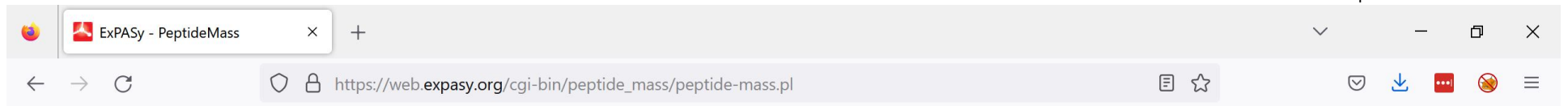
The peptide masses are

with cysteines treated with:

with carboxymide adducts

https://web.expasy.org/peptide_mass/

Example: PeptideMass



1149.5759	25-34	1	DAHKSEVAHR
698.3580	29-34	0	SEVAHR
973.5214	29-36	1	SEVAHRFK
1226.6051	35-44	1	FKDLGEENFK
951.4418	37-44	0	DLGEENFK
3365.6874	37-65	1	DLGEENFKALVLIAFAQYLQ QCPFEDHVK
2433.2635	45-65	0	ALVLIAFAQYLQCPFEDHV K
3563.8606	45-75	1	ALVLIAFAQYLQCPFEDHV KLVNEVTEFAK
1149.6150	66-75	0	LVNEVTEFAK
2515.1326	66-88	1	LVNEVTEFAKTCVADESAEN CDK
1384.5355	76-88	0	TCVADESAENCDK
2383.0540	76-97	1	TCVADESAENCDKSLHTLFG DK
1017.5363	89-97	0	SLHTLFGDK
1875.0156	89-105	1	SLHTLFGDKLCTVATLR
876.4971	98-105	0	LCTVATLR
2177.9698	98-117	1	LCTVATLRETYGEMADCCAK
1320.4905	106-117	0	ETYGEMADCCAK
1959.7881	106-122	1	ETYGEMADCCAKQEPER
658.3155	118-122	0	QEPER
1657.7751	118-130	1	QEPERNECFLQHK
1018.4775	123-130	0	NECFLQHK
1939.9079	123-138	1	NECFLQHKDDNPNLPR
940.4483	131-138	0	DDNPNLPR
2514.0700	131-138	1	DDNPNLPRVDEEYVAVQTA EYVDEEYVAVQTA

A ~~K~~ A C R E F K A D E F R P A D E

$$MW = \sum_{i=1}^n MW_{A_i} + 19$$

2 m.c.

