

Programming with Matlab

Day 4: Reading and Saving Data

Save and load Data: Matlab's format

- A session can be saved or loaded, eg. all variables that are stored in cache.

>> save('mySave.mat')  Saves all variables to file mySave.mat

>> save('mySave.mat', 'var1', 'var2', 'perrito')  Saves variables var1, var2 and perrito to file mySave.mat

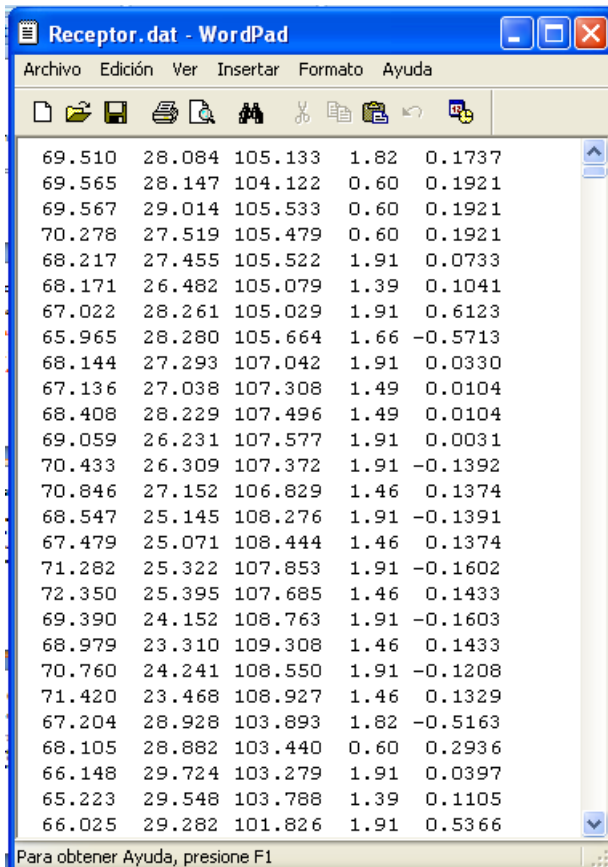
>> load('mySave.mat')  Loads variables stored in file mySave.mat

Data stored in text files

- **Advantage:** Highly compatible

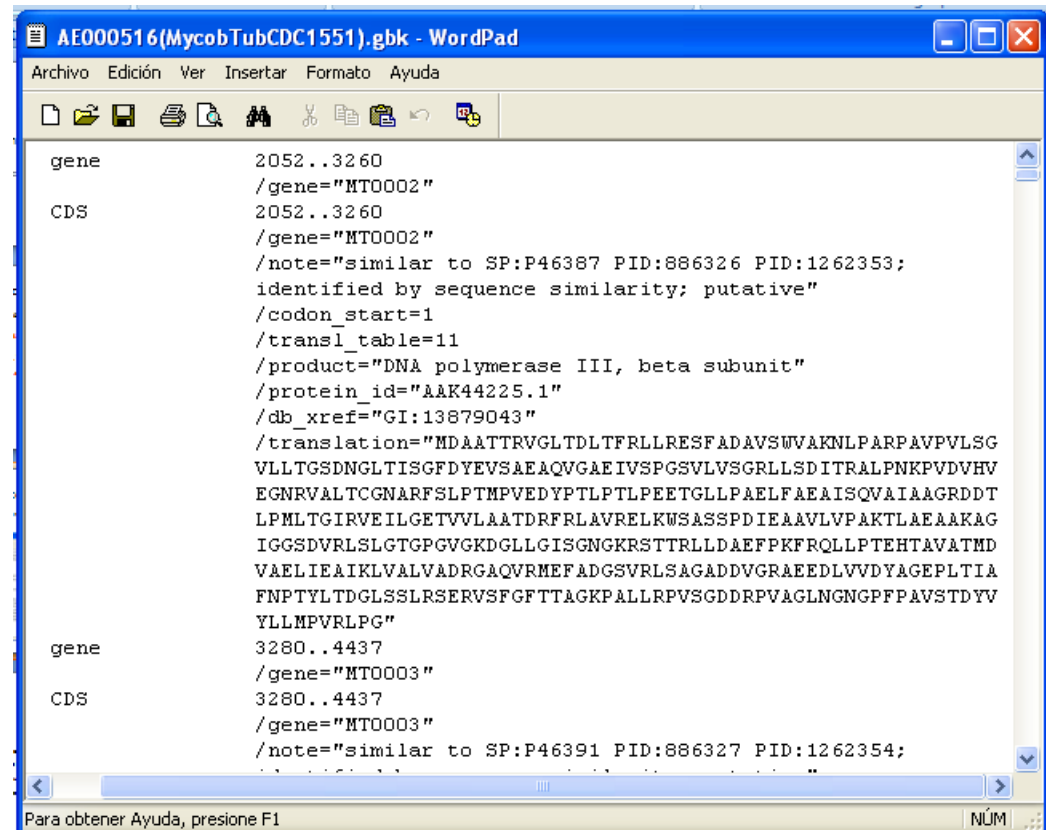
Main types

Matrix-formatted



69.510	28.084	105.133	1.82	0.1737
69.565	28.147	104.122	0.60	0.1921
69.567	29.014	105.533	0.60	0.1921
70.278	27.519	105.479	0.60	0.1921
68.217	27.455	105.522	1.91	0.0733
68.171	26.482	105.079	1.39	0.1041
67.022	28.261	105.029	1.91	0.6123
65.965	28.280	105.664	1.66	-0.5713
68.144	27.293	107.042	1.91	0.0330
67.136	27.038	107.308	1.49	0.0104
68.408	28.229	107.496	1.49	0.0104
69.059	26.231	107.577	1.91	0.0031
70.433	26.309	107.372	1.91	-0.1392
70.846	27.152	106.829	1.46	0.1374
68.547	25.145	108.276	1.91	-0.1391
67.479	25.071	108.444	1.46	0.1374
71.282	25.322	107.853	1.91	-0.1602
72.350	25.395	107.685	1.46	0.1433
69.390	24.152	108.763	1.91	-0.1603
68.979	23.310	109.308	1.46	0.1433
70.760	24.241	108.550	1.91	-0.1208
71.420	23.468	108.927	1.46	0.1329
67.204	28.928	103.893	1.82	-0.5163
68.105	28.882	103.440	0.60	0.2936
66.148	29.724	103.279	1.91	0.0397
65.223	29.548	103.788	1.39	0.1105
66.025	29.282	101.826	1.91	0.5366

Other formats



```
gene                2052..3260
                    /gene="MT0002"
CDS                 2052..3260
                    /gene="MT0002"
                    /note="similar to SP:P46387 PID:886326 PID:1262353;
                    identified by sequence similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="DNA polymerase III, beta subunit"
                    /protein_id="AAK44225.1"
                    /db_xref="GI:13879043"
                    /translation="MDAATTRVGLTDLTFRLLRESFADAVSWVAKNLPARPVAVPVL
                    VLLTGSNDGLTISGFDYEVSAEAQVGAIEIVSPGSVVLVSGRLLSDITRALPNKPV
                    DHHV
                    EGNRVALTTCGNARFSLPTMPVEDYPTLPTLPEETGLLPAELFAEAI
                    SQVAIAAGRDDT
                    LPMLTGIRVEILGETVVLAATDRFRLAVRELKWSASSPDEAAVLVPAKTL
                    AEAAKAG
                    IGGDVRLSLGTGPGVGKDGLLGISGNGKRSTTRLLDAEFPKFRQLLPT
                    EHTAVATMD
                    VAELIEAIKLVADVDRGAQVRMEFADGSVRLSAGADDVGRAEEDLVVDY
                    AGEPLTIA
                    FNPTYLTDGLSSLRSERVSFGFTTAGKPALLRPVSGDDRPVAGLNGNGP
                    PFAVSTDYV
                    YLLMPVRLPG"
gene                3280..4437
                    /gene="MT0003"
CDS                 3280..4437
                    /gene="MT0003"
                    /note="similar to SP:P46391 PID:886327 PID:1262354;
                    identified by sequence similarity; putative"
                    /codon_start=1
                    /transl_table=11
                    /product="DNA polymerase III, beta subunit"
                    /protein_id="AAK44226.1"
                    /db_xref="GI:13879043"
                    /translation="MDAATTRVGLTDLTFRLLRESFADAVSWVAKNLP
                    ARPVAVPVL
                    SVLLTGSNDGLTISGFDYEVSAEAQVGAIEIVSPGSVVLVSGRLLSDIT
                    RALPNKPV
                    DHHV
                    EGNRVALTTCGNARFSLPTMPVEDYPTLPTLPEETGLLPAELFAEAI
                    SQVAIAAG
                    RDDT
                    LPMLTGIRVEILGETVVLAATDRFRLAVRELKWSASSPDEAAVLVPAKTL
                    AEAAKAG
                    IGGDVRLSLGTGPGVGKDGLLGISGNGKRSTTRLLDAEFPKFRQLLPT
                    EHTAVATMD
                    VAELIEAIKLVADVDRGAQVRMEFADGSVRLSAGADDVGRAEEDLVVDY
                    AGEPLTIA
                    FNPTYLTDGLSSLRSERVSFGFTTAGKPALLRPVSGDDRPVAGLNGNGP
                    PFAVSTDYV
                    YLLMPVRLPG"
```

Reading matrix-formatted data

- Look for the file in the “current directory” menu, right click, “import data”
- In case the file is column-formatted, in order to integrate it in a script,

```
>> Matrix = importdata( 'filename' )
```

Reading other files

- **Step 1:** Open file so that Matlab can access it.

```
>> id = fopen ('filename'); → Opens the file in read mode.
```



File identifier. It will represent the file from now on.



Name of the file, including extension.

- **Step 2:** Read the file. One useful instruction (see also fread):

```
>> cadena = fgetl(id); → Reads one line of the file, and stores it in variable cadena. Each time we execute fgetl, we get the following line.
```

- **Step 3:** Close the file

```
>> fclose(id);
```

Store data in text files

- **Step 1:** Open file so that Matlab can access it.

```
>> id = fopen ('filename','w');
```

Opens the file in write mode. If the file does not exist, creates the file.

Write mode.

File identifier. It will represent the file from now on.

Name of the file, including extension.

- **Step 2:** Write on the file (see also fwrite):

```
>> cadena='first';
```

```
>> respuesta=7.82;
```

```
>> fprintf(id,'The answer to the %s problem is %g.\n',cadena,respuesta)
```

String to be written

\n = new line

Variables whose values will be inserted.

%s = Insert text

%g = Insert number (see the help for sprintf for other number formats)

File identifier. If none is specified, it will be printed on the screen.

- **Step 3:** Close the file

```
>> fclose(id);
```

Exercise

1) Implement the following functions:

- Max and Min functions, which find the maximum and minimum of a given vector.
- A Function that computes the coulombic potential between two atoms.

$$\text{Atomo 1: } p_1 = (x_1, y_1, z_1), Q_1$$

$$\text{Atomo 2: } p_2 = (x_2, y_2, z_2), Q_2$$

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2};$$

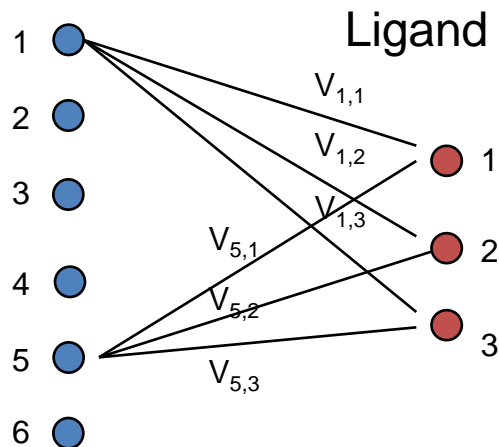
$$V_{1,2} = \frac{Q_1 Q_2}{d};$$

Exercise

2) Write a script that:

- Calculates the dimensions of a box which contains the proteing (use the max and min functions!)
- Read the ligand file (ligand_01.dat)
- Computes the coulombic interaction between the whole protein and the whole ligand. (hint: 2 vector “all against all”).

Protein



$$TotalCoul = V_{1,1} + V_{1,2} + V_{1,3} + \dots + V_{6,1} + V_{6,2} + V_{6,3};$$

Exercise

3) Write a script that:

- Computes the coulombic interaction between each ligand (ligand_01.dat, ligand_02.dat, ...) and the protein.
- Finds the ligand with the lowest coulombic interaction energy.
- Plot the energies.

