

MSc in **Bioinformatics**

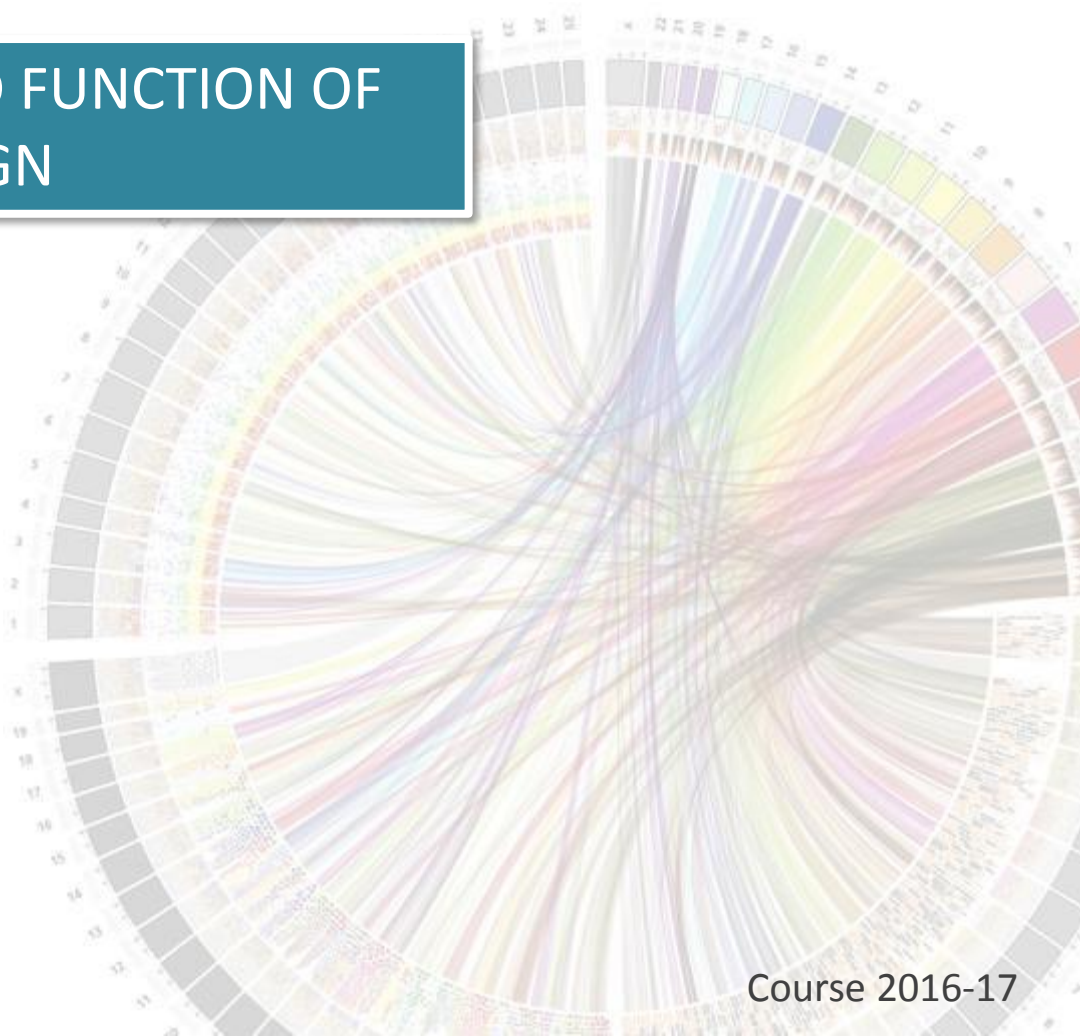
Module 4: STRUCTURE AND FUNCTION OF PROTEINS AND DRUG DESIGN

Scientific software
development with
modern Python

Jaime Rodríguez-Guerra

InSilichem

Department of Chemistry (UAB)





Why Python for science

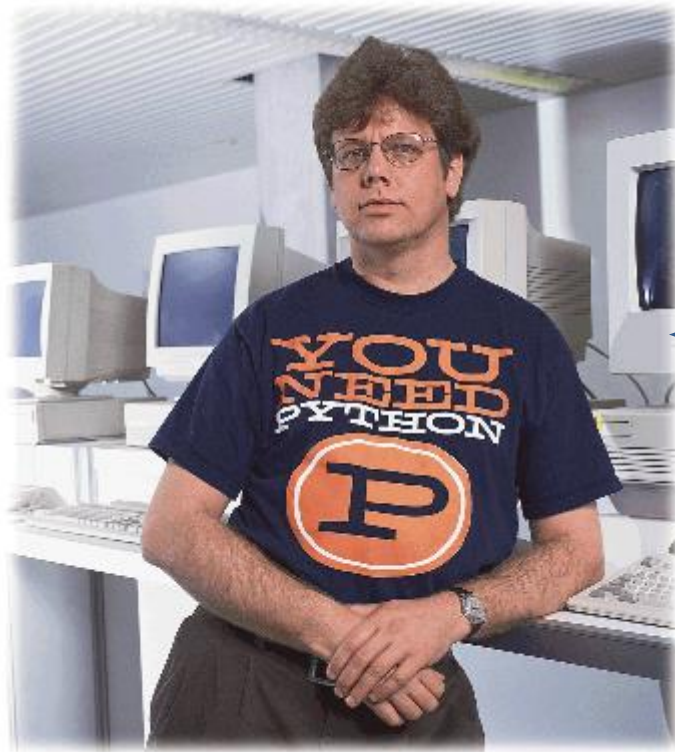
“Code is read much more often than it is written”

Guido van Rossum, aka BDFL

- Readability is key to understandability, reproducibility, extensibility... SCIENCE!
- Lots of free, high quality packages (“batteries incl.”)
- The perfect glue language



Why Python for science



<http://legacy.python.org/~guido/pics.html>

“Code is read much more often than it is written”



Guido van Rossum, aka BDFL

- Readability is key to understandability, reproducibility, extensibility... SCIENCE!
- Lots of free, high quality packages (“batteries incl.”)
- The perfect glue language



Why not

- Python is slow!
 - Accelerate critical parts with Numba, Cython or a C extension
- Python 2 vs Python 3
 - Use Python 3 and stop complaining (or `conda`!)
- Meaningful indentation is silly
 - You'll take that back on Day 2
- I don't like it!
 - I don't take emotionally-driven arguments seriously



Did I mention readability?

```
animals = ['dog', 'cat', 'lion', 'bird']  
colors = ['yellow', 'red', 'black', 'blue']
```

```
for ($i=0; $i<len(animals); $i++) {  
    animal = $animals[$i];  
    color = $colors[$i];  
    print "$color $animal";  
}
```

```
for animal, color in zip(animals, colors):  
    print color, animal
```



Batteries included

- All kinds of modules ship together with Python
- If not, the Python ecosystem is awesome!
- For Science, there's nothing like the SciPy stack

Numpy	Base N Dimensional Array Package
Scipy Library	Fundamental Library For Scientific Computing
Matplotlib	Comprehensive 2d Plotting
Ipython (Jupyter)	Enhanced Interactive Console
Sympy	Symbolic Mathematics
Pandas	Data Structures & Analysis



My favourite: Jupyter Notebooks!

- A whole new paradigm to software development
- Combines the flexibility of a dynamic interpreter with the depth & focus of a text editor...
- ... and adds all kinds of awesomeness!
- You can even send the resulting notebook file to your colleagues with the results included:
REPRODUCIBILITY!



Really! Pure awesomeness!

jupyter spectrogram (autosaved)



File Edit View Insert Cell Kernel Help

Python 3



Simple spectral analysis

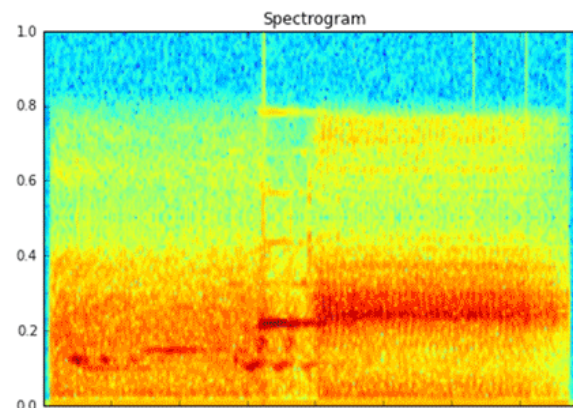
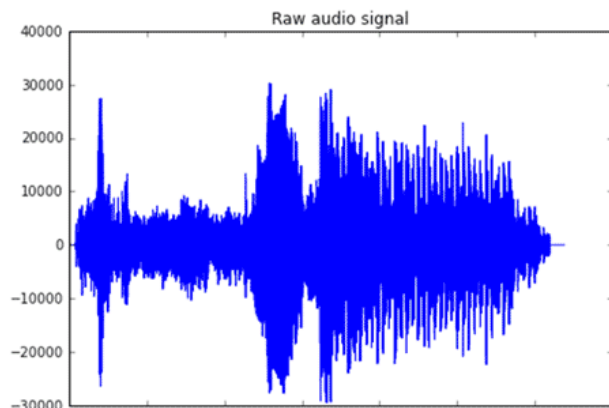
An illustration of the [Discrete Fourier Transform](#)

$$X_k = \sum_{n=0}^{N-1} x_n \exp\left(\frac{-2\pi i}{N} kn\right) \quad k = 0, \dots, N-1$$

```
In [2]: from scipy.io import wavfile
rate, x = wavfile.read('test_mono.wav')
```

And we can easily view it's spectral structure using matplotlib's builtin specgram routine:

```
In [5]: fig, (ax1, ax2) = plt.subplots(1,2,figsize(16,5))
ax1.plot(x); ax1.set_title('Raw audio signal')
ax2.specgram(x); ax2.set_title('Spectrogram');
```





Install all this stuff!

- These packages can take long to compile from source... Like hours!
- Thankfully, we have the Anaconda Project!
- A distribution of Python with all the scientific packages you will ever need AND a package manager that downloads compiled builds for your system.
- It even provides environments!



Environments




- Isolated compilation of packages.
- Create an environment for each project, with versions written down! (Reproducibility!)
 - Also prevents incompatibilities between versions.
 - You can even have Python 2 and 3 in the same system! Just create different envs!
- “Why would I need that?”
 - If you’re doing serious dev, trust me, you will.



Ready?

- Go to <https://conda.io/miniconda.html>

Miniconda

	 Windows	 Mac OS X	 Linux
Python 2.7	64-bit (exe installer) 32-bit (exe installer)	64-bit (bash installer)	64-bit (bash installer) 32-bit (bash installer)
Python 3.5	64-bit (exe installer) 32-bit (exe installer)	64-bit (bash installer)	64-bit (bash installer) 32-bit (bash installer)

Fork me on GitHub