THE UNIVERSITY of EDINBURGH

How is python used in biomolecular sciences?

L. Hedges

Antonia Mey

antonia.mey@ed.ac.uk

@ppxasjsm

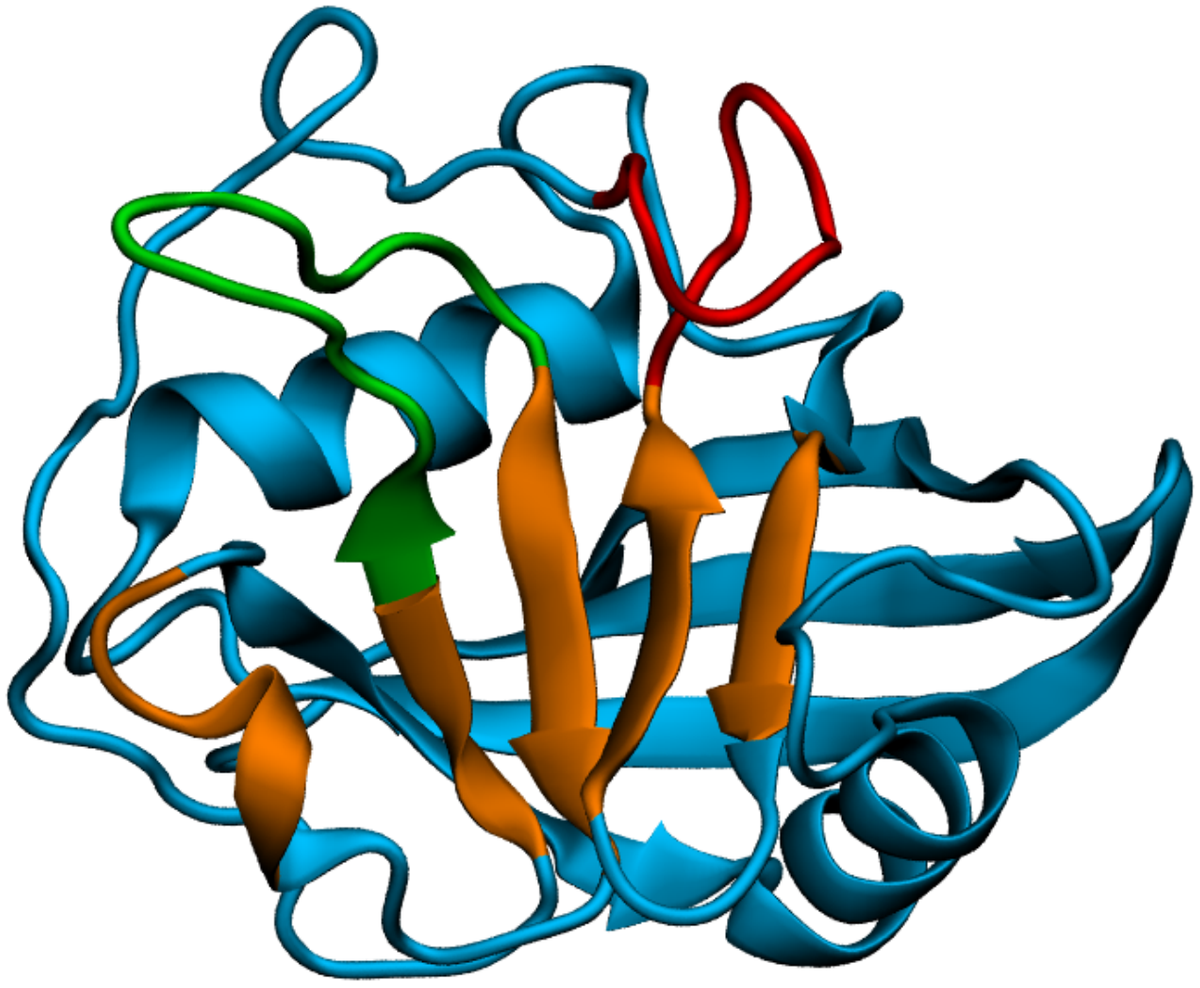
C. Woods

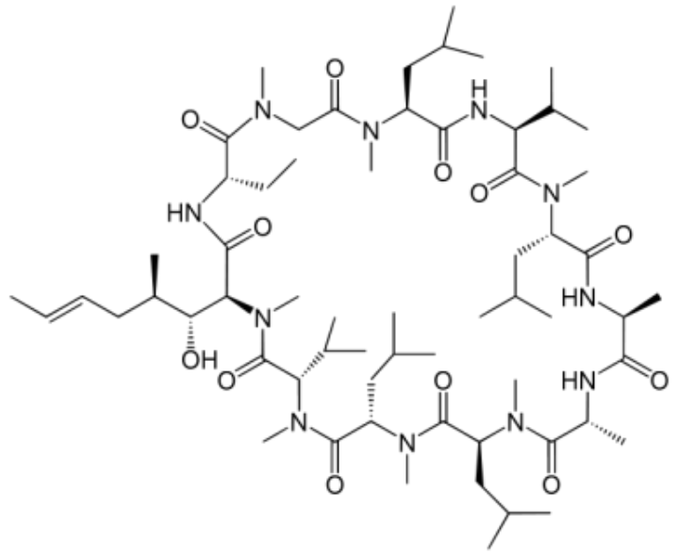
Europython 2018 — Edinburgh

25/07/2018

What is computational biosomething?

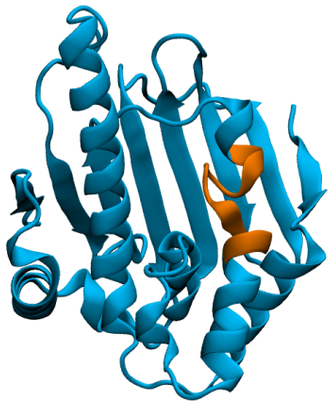
42 x10^{-10} m



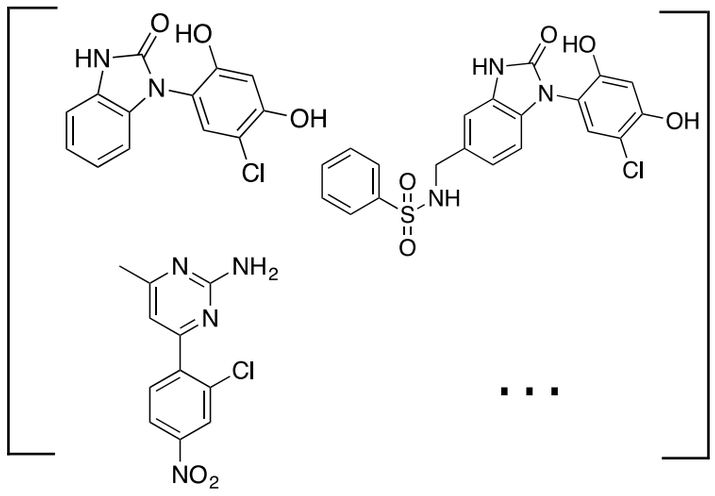


X-ray structure of a protein:
set of x,y,z spatial coordinates of atoms

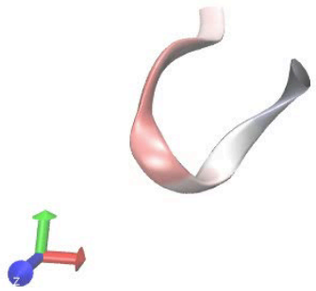
Typical questions tackled with MD



How well do small molecules inhibit a protein's functionality?



many potential compounds

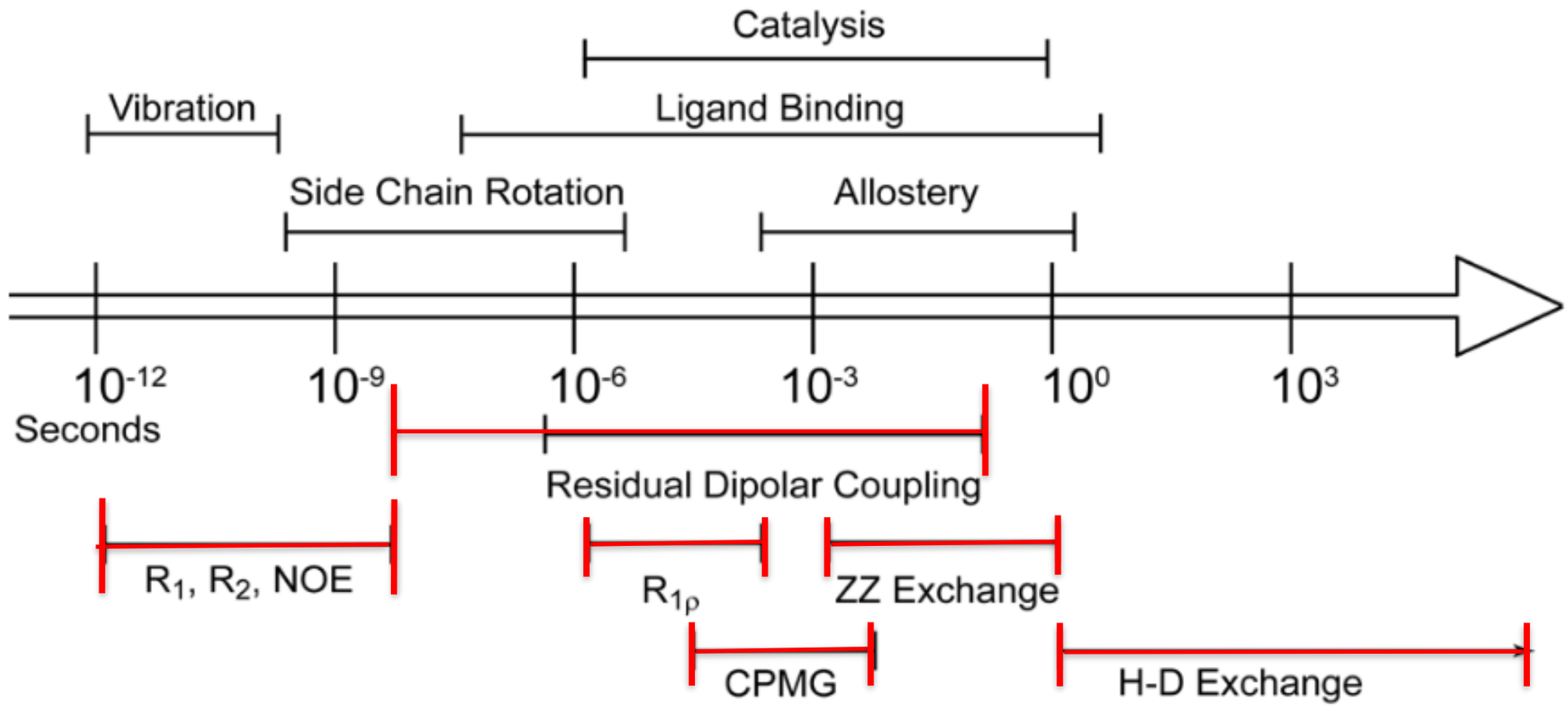


How do proteins fold?

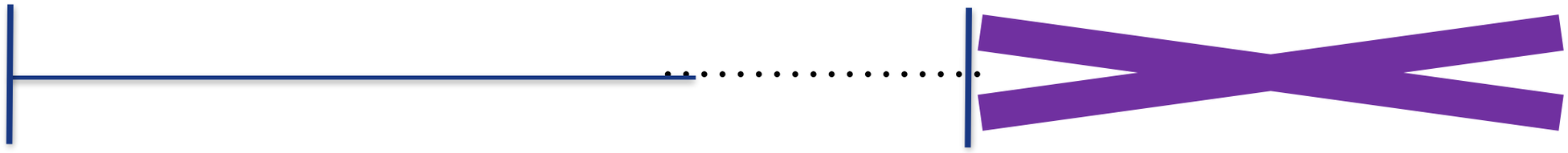
How do proteins interact with each other?

Timescales in biology

Timescales of processes

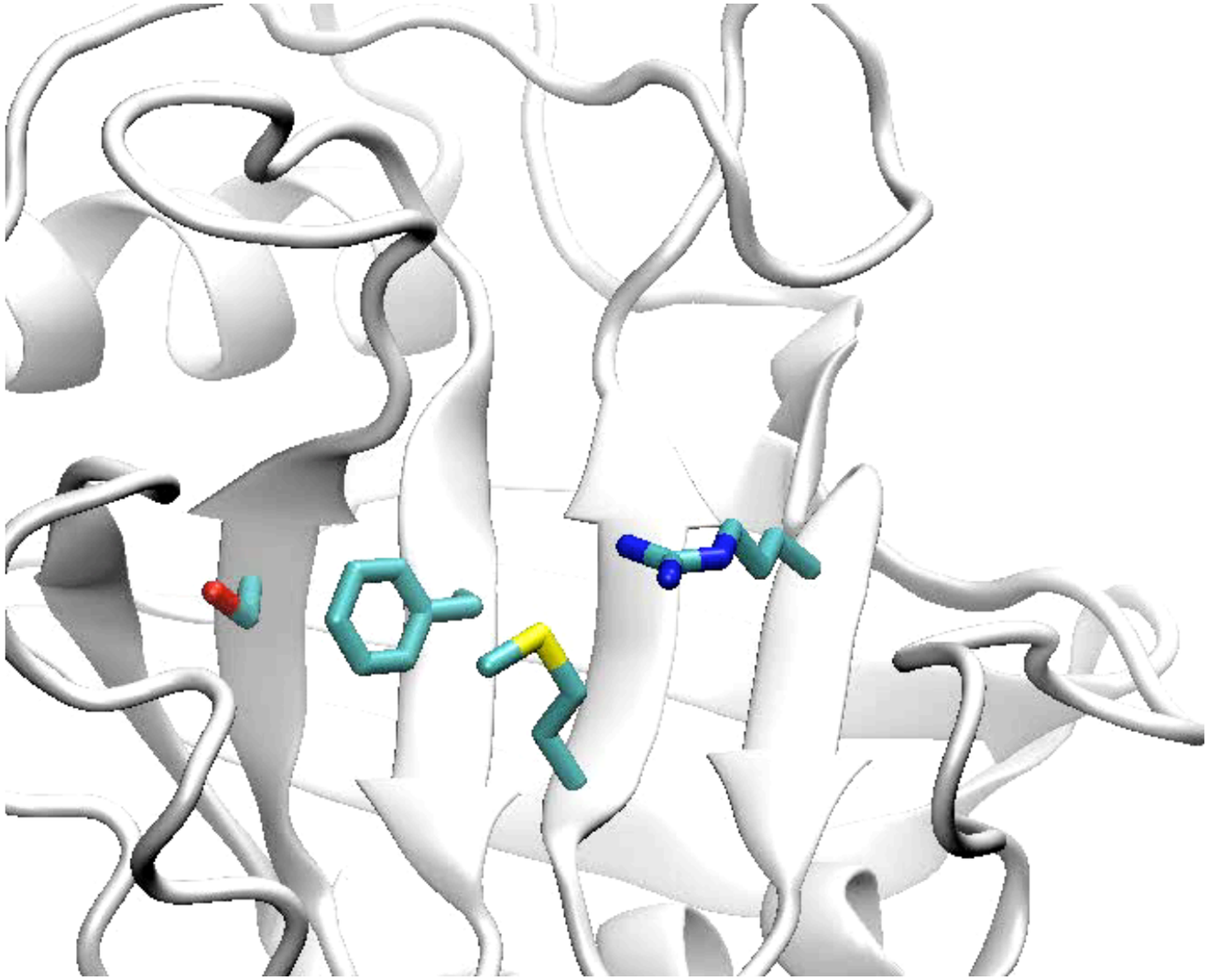


Molecular Dynamics

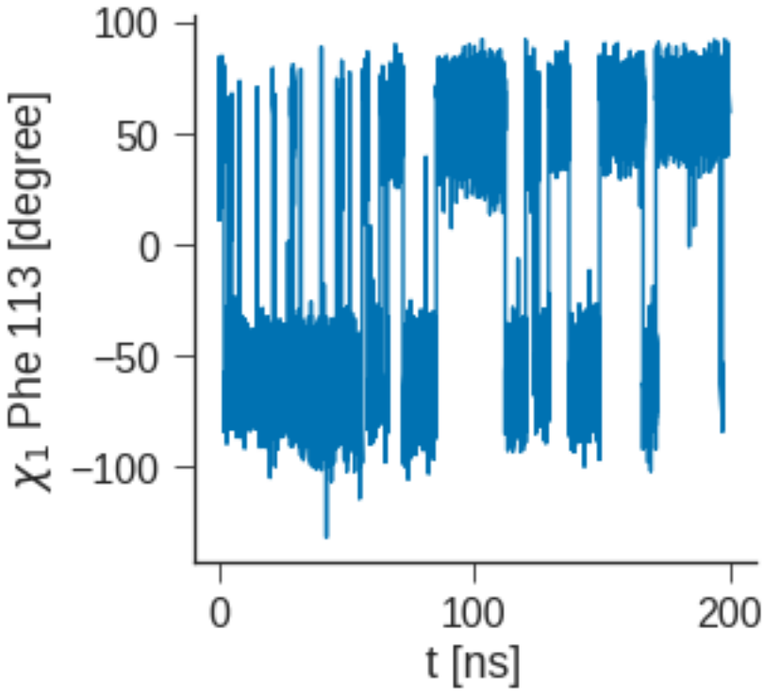


Other experimental techniques: e.g. cryo EM

Generate timeseries to compare to experiments



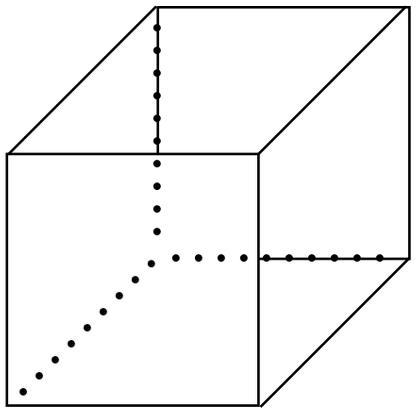
200 ns of protein dynamics



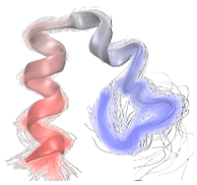
time trace of an angle

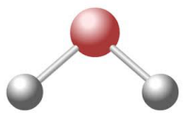
What is MD?

box



+





+

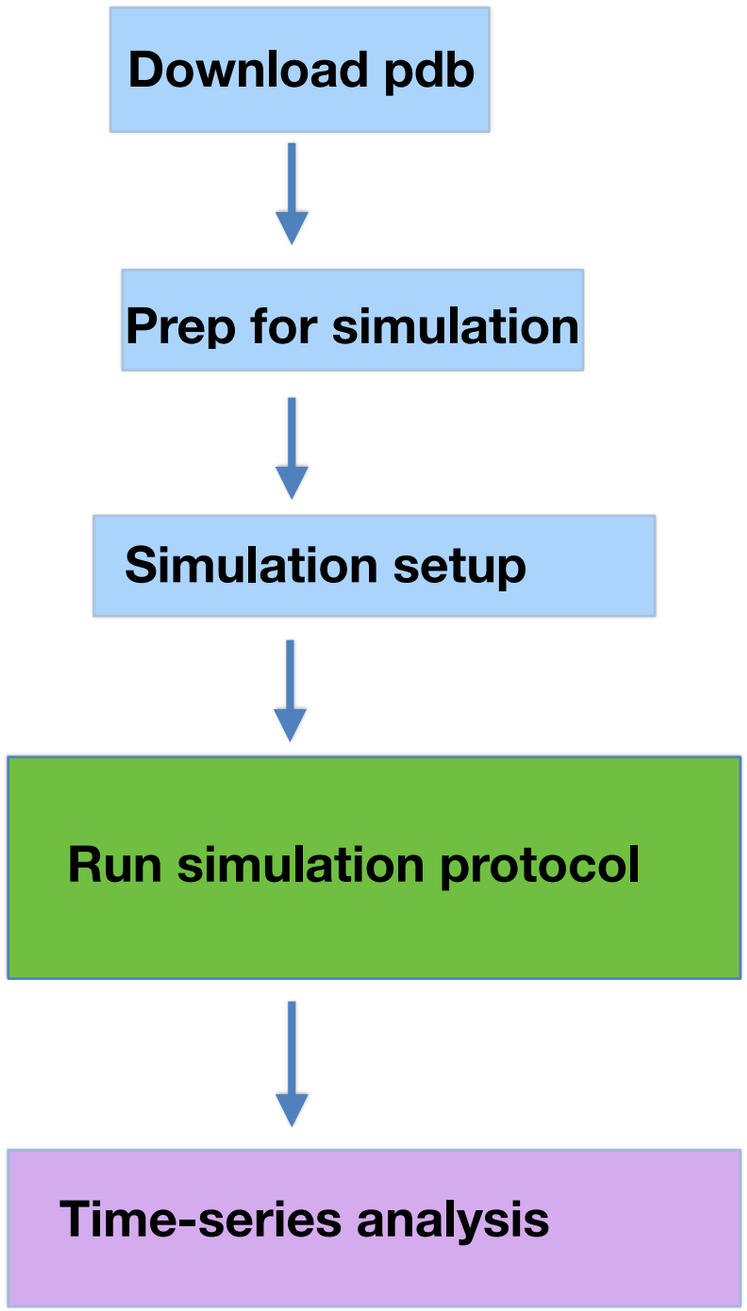
$$
\begin{aligned}
U &= \sum_{\text{bonds}} k_r (r - r_0)^2 \\
&+ \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \\
&+ \sum_{\text{dihedrals}} k_\phi \left[1 + \cos(n\phi + \phi_0)\right] \\
&+ \sum_{\text{atom } i} \sum_{j \neq i} 4\epsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right] \\
&+ \sum_{i} \sum_{j \neq i} \frac{q_i q_j}{\epsilon_0 r_{i,j}}
\end{aligned}
$$

+ integrator

$\longrightarrow \langle A \rangle_{\text{ensemble}} = \langle A \rangle_{\text{time}}$

Physics engine usually written in C++, to integrate Newtons equation of motion using leap frog type algorithms.

Typical MD workflow



pdb api

pdb2gmx

Cresset suite

Schrödinger suite

OpenMM tools

Amber tools

RDkit

[…]

Commercial

GUI, TCL, bash, python, Perl … get creative

OpenMM

Gromacs

NAMD

DLPoly

Amber

SOMD

Charmm

[…]

Python API

mostly C++
commandline

PyEMMA

MDAnalysis

md-analysis-tools

MDTraj

MDAnalysis

MMTools

[…]

Python

TCL, bash, python, Perl … get creative

Typical file formats

coordinates	.pdb	.crd	.gro	.xyz	.mol2
trajectories	.dcd	.xtc	.trr		
forcefields	.psf	.parm7	.itp		

And of course all MD programs can read all these file formats? — No

Scenario: Simulate .gro file with Amber

I have a coordinate .gro (Gromacs) file I want to simulate with Amber, how do I do this?

visualise coordinates

convert to pdb format

use a setup tool

run simulation

Scenario: Simulate .gro file with Amber

I have a coordinate .gro (Gromacs) file I want to simulate with Amber, how do I do this?

visualise coordinates

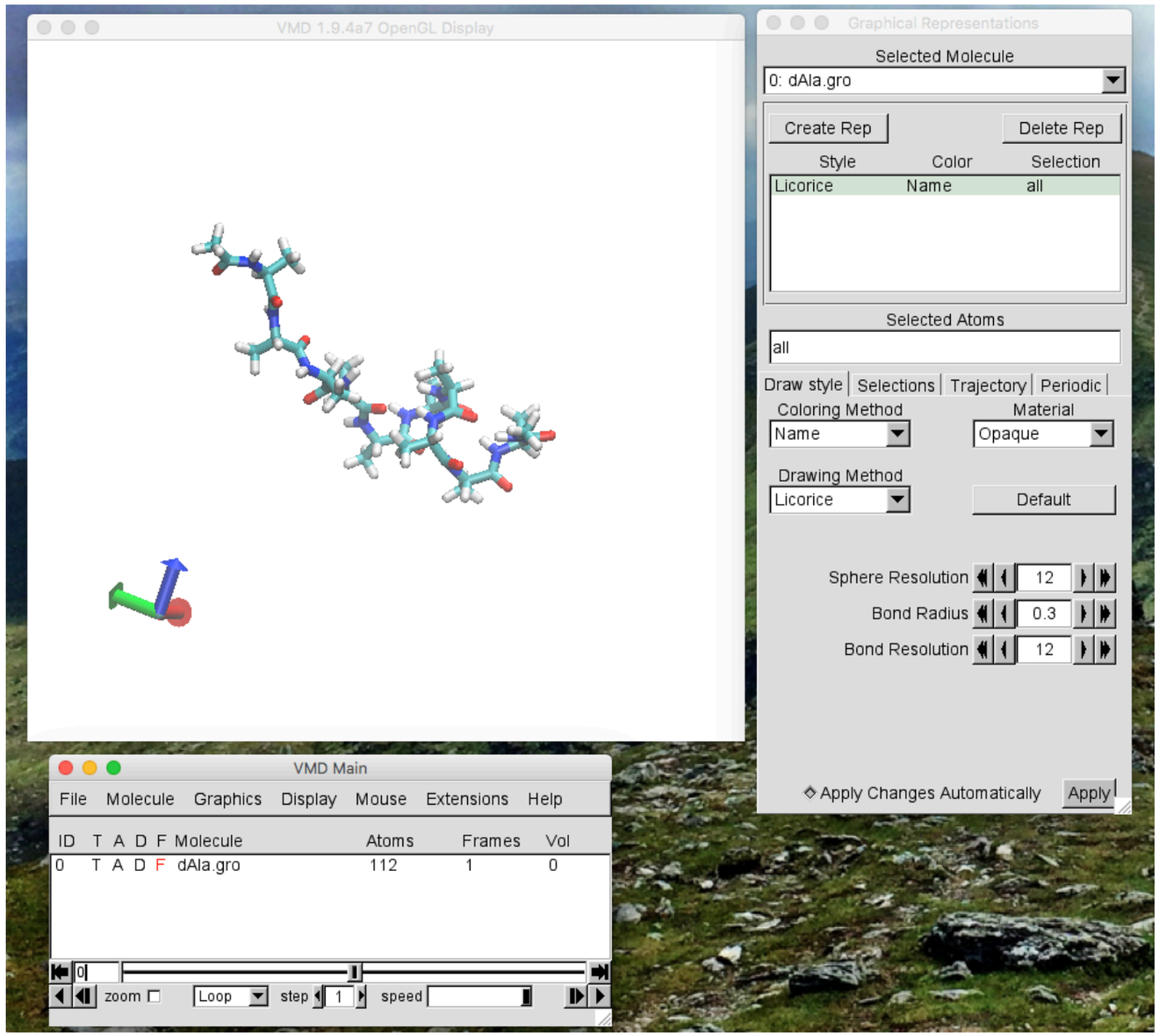
convert to pdb format

use a setup tool

run simulation

or, use one of the many other tools…

```
ppxasjsm::azuma { ~/Documents/}-> vmd dAla.gro
```



Scenario: Simulate .gro file with Amber

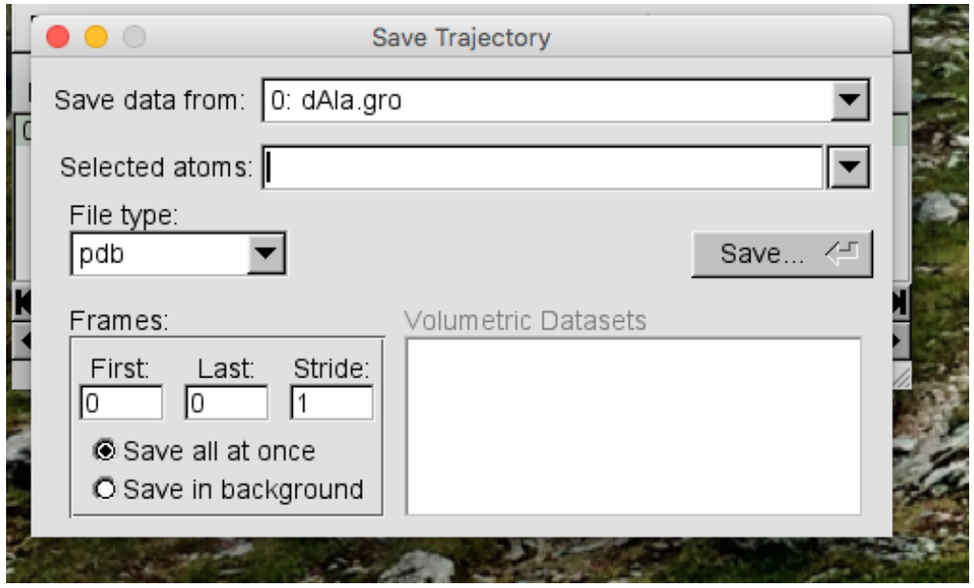
I have a coordinate .gro (Gromacs) file I want to simulate with Amber, how do I do this?

visualise coordinates

convert to pdb format

use a setup tool

run simulation



or

```
In [7]: import MDAnalysis as md

In [8]: u = md.Universe('dAla.gro')
        coors = u.select_atoms('all')
        coors.write('md_analysis.pdb')
```

or ....

Scenario: Simulate .gro file with Amber

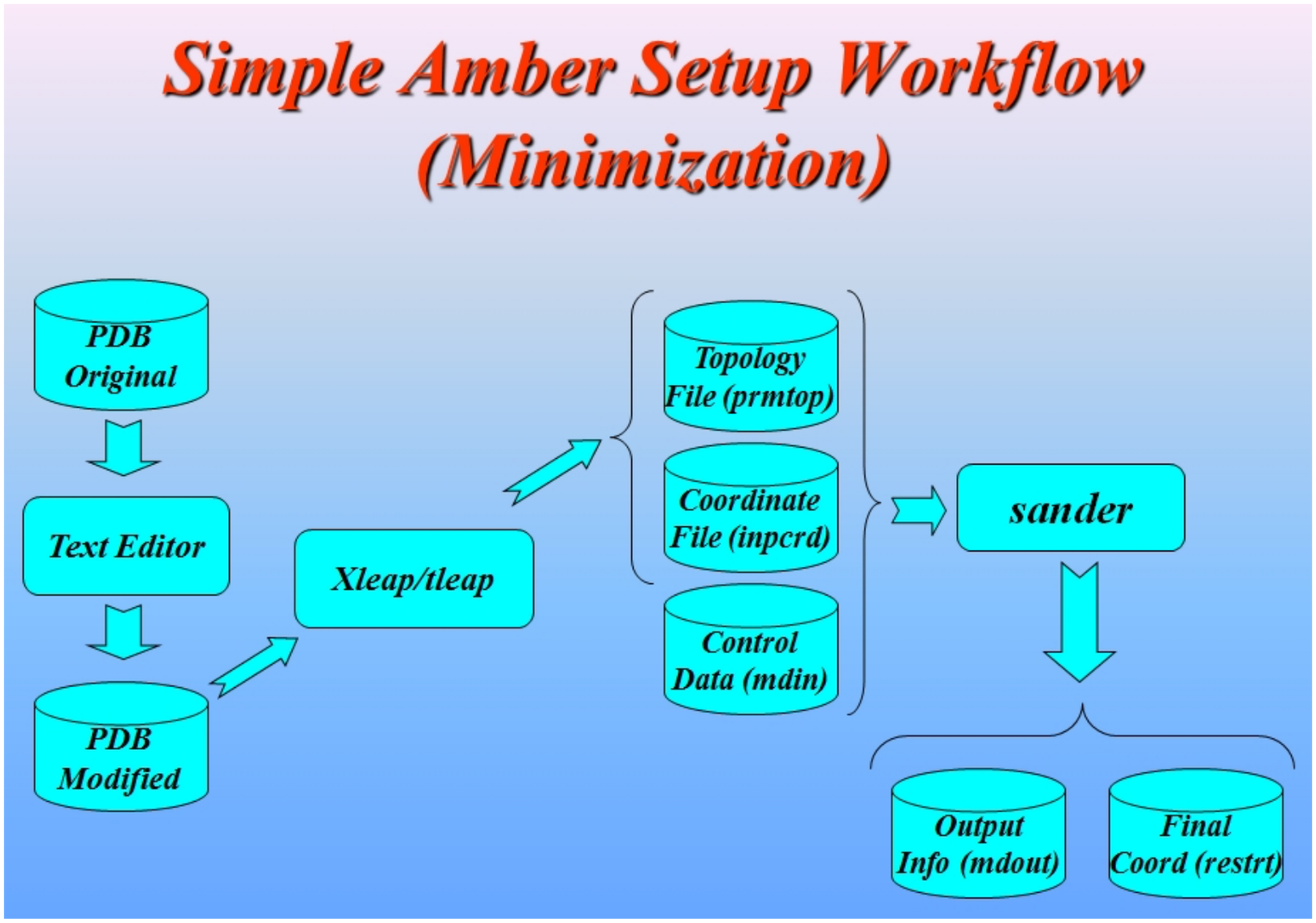
I have a coordinate .gro (Gromacs) file I want to simulate with Amber, how do I do this?

visualise coordinates

convert to pdb format

use a setup tool

run simulation



taken from AMBER tutorial

pretty much all the arrows are bash scripts with input files

Scenario: Simulate .gro file with Amber

I have a coordinate .gro (Gromacs) file I want to simulate with Amber, how do I do this?

visualise coordinates

convert to pdb format

use a setup tool

run simulation

```
ppxasjsm::azuma { ~/Documents/}-> FESetup setup.in

[globals]
forcefield = amber, ff99SBildn, tip3p

[protein]
basedir = protein
file.name = protein.pdb
molecules = ala

box.type = rectangular
box.length = 10.0
align_axes = True
neutralize = yes

min.nsteps = 1000
min.restr_force = 10.0
min.restraint = notsolvent

md.heat.nsteps = 1000
md.heat.restr_force = 10.0
md.heat.restraint = notsolvent

md.constT.nsteps = 1000
md.constT.restr_force = 10.0
md.constT.restraint = notsolvent

md.press.T = 298.0
md.press.nsteps = 50000
md.press.p = 1.0
```

Scenario: Simulate .gro file with Amber

I have a coordinate .gro (Gromacs) file I want to simulate with Amber, how do I do this?

visualis

convert

use a se

run simu

General | System | Integrator | Simulation

Input coordinates	input.pdb
Input topology	input.prmtop
Forcefield	AMBER99sb-ildn
Water Model	TIP3P
Platform	CUDA
Precision	mixed
Device index	
OpenCL platform indx	

```
##############################################################################
#
# this script was generated by openmm-builder. to customize it further,
# you can save the file to disk and edit it with your favorite editor.
##############################################################################
#

from __future__ import print_function
from simtk.openmm import app
import simtk.openmm as mm
from simtk import unit
from sys import stdout

pdb = app.PDBFile('input.pdb')
forcefield = app.ForceField('amber99sbildn.xml', 'tip3p.xml')

system = forcefield.createSystem(pdb.topology, nonbondedMethod=app.PME,
    nonbondedCutoff=1.0*unit.nanometers, constraints=app.HBonds,
rigidWater=True,
    ewaldErrorTolerance=0.0005)
integrator = mm.LangevinIntegrator(300*unit.kelvin, 1.0/unit.picoseconds,
    2.0*unit.femtoseconds)
integrator.setConstraintTolerance(0.00001)
```

Scenario: Simulate .gro file with Amber

I have a coordinate .gro (Gromacs) file I want to simulate with Amber, how do I do this?

visualise coordinates

convert to pdb format

use a setup tool

run simulation

And another command line tool…

```
pmemd -O -i ala.in -p ala.parm7
-c ala.rst7 -o myout.mdout -x
myout.nc -e myout.mdene -r
myout.rst7
```

Zoo of applications leads to hacky workflows

- Most tools have **organically grown** from academic software with poor software practices
- In order for tools to work with each other in complicated workflows a lot of **hacky bash scripting** is used by academic users
  - Users need to be **experts in many different software** with command line interfaces to interlink them
  - Many tools can do similar things and there may not be **obvious solutions for one problem** (google trap: try suggestions until one works)

-> Loss of focus on the science

The same example as before BioSimSpace

In [ ]:
```
import BioSimSpace as BSS
```

Reading in the gro file as before

In [ ]:
```
system = BSS.IO.readMolecules('input/dAla.gro')
```

Simply view the molecule in the notebook rather than using external tool

In [ ]:
```
BSS.viewMolecules('input/dAla.gro')
```

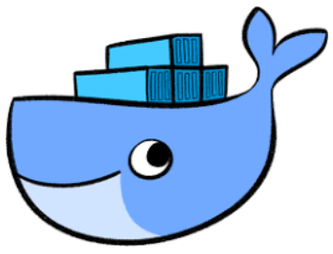
Selecting the first and only molecule to parametrise it and then solvating it

In [ ]:
```
molecule = system.getMolecules()[0]
```

In [ ]:
```
molecule = BSS.Parameters.parameterise(molecule, 'ff99SB').getMolecule()
```

In [ ]:
```
system = BSS.Solvent.solvate('tip3p', molecule=molecule,
                             box=3*[3*BSS.Units.Length.nanometer],ion_conc=0)
```

Cloud server demo — using BioSimSpace



```
docker run chryswoods/biosimspace
```

http://130.61.69.221/hub/tmplogin

http://130.61.69.221/

ORACLE CLOUD

Azure

BioSimSpace phase I

Download pdb

↓

Prep for simulation

↓

Simulation setup

↓

Run simulation protocol

↓

Time-series analysis

some manual prep

Commercially available software

BioSimSpace

L. Hedges

C. Woods

API overview

Core: Sire — Molecular library in C++

Gateway

Trajectory

IO

Process

MD

Protocol

What is BioSimSpace — summary

BioSimSpace

System Setup

Protein
Protein + ligand
ligand in water
ligand in organic solvent
…

Software

pdb2gmx
tleap
FESetup
…

Interoperable tool for biomolecular simulations

Trajectory generation

MD
MC
Enhanced sampling
…

Software

Amber
Gromacs
OpenMM
HTMD
Plumed
…

Simulation Analysis

MSM
Perturbation
Map
Reweighting
…

Software

pyemma
gmx wham
pymbar
…

https://github.com/michellab/BioSimSpace

Conclusions

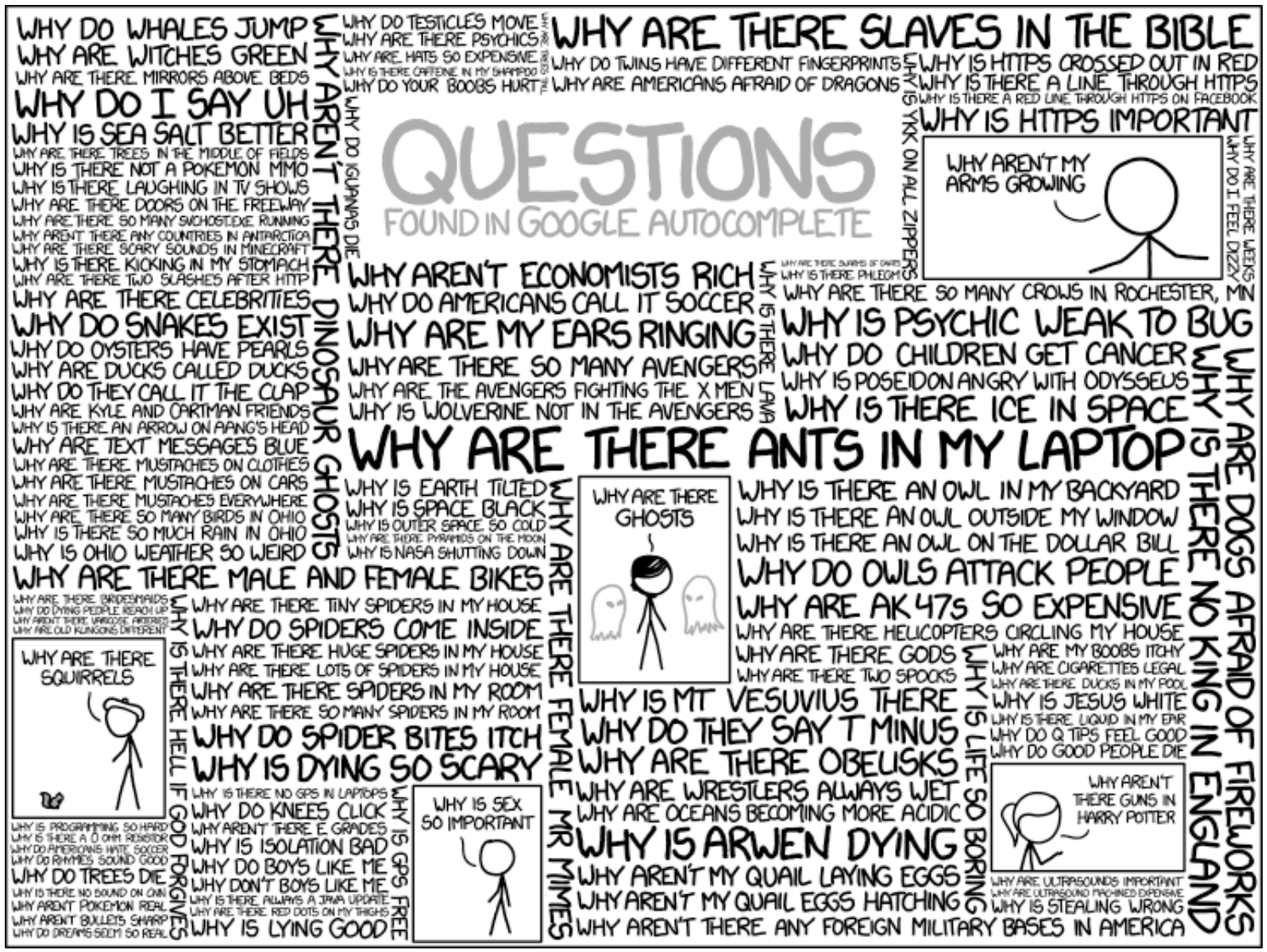
BioSimSpace

- Python API to write workflow components for Biomolecular simulations
- Allows to focus on science and not software: No need to become an expert at different MD packages, setup or analysis tools
- Ease of use in the cloud, with scalable computing resources and future academic fool proof pricing model
- Planned support for Knime and CWL workflow managers

~~**BioSimSpace**~~

- A workflow engine
- A top down approach by trying to reinvent the wheel again
- A ‘finished’ piece of software: It is very much in an alpha development stage with a large list of features and capabilities to be added in the future

Questions


QUESTIONS
FOUND IN GOOGLE AUTOCOMPLETE
WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY ARE THERE MALE AND FEMALE BIKES
WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARICOSE AMBERS
WHY ARE OLD KLINGONS DIFFERENT
WHY ARE THERE SQUIRRELS
WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON OVNI
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL
WHY DO TESTICLES MOVE
WHY ARE THERE PSYCHICS
WHY ARE HATS SO EXPENSIVE
WHY IS THERE CAFFEINE IN MY SHAMPOO
WHY DO YOUR BOOBS HURT
WHY ARE THERE SLAVES IN THE BIBLE
WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT
WHY AREN'T MY ARMS GROWING
WHY IS YKK ON ALL ZIPPERS
WHY ARE THERE SWARMS OF DAVID
WHY IS THERE PHLEGM
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS
WHY IS PSYCHIC WEAK TO BUG
WHY DO CHILDREN GET CANCER
WHY IS POSEIDON ANGRY WITH ODYSSEUS
WHY IS THERE ICE IN SPACE
WHY ARE THERE ANTS IN MY LAPTOP
WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN
WHY ARE THERE GHOSTS
WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS
WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA
WHY ARE MY BOOBS ITCHY
WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE
WHY AREN'T THERE GUNS IN HARRY POTTER
WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG
WHY IS SEX SO IMPORTANT
WHY IS LYING GOOD
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY
WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY ARE THERE SPIDERS IN MY HOUSE
WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO GUINEA PIGS DIE
WHY AREN'T THERE DINOSAUR GHOSTS
WHY ARE THERE FEMALE MR MIMES
WHY IS THERE LAVA
WHY IS GPS FREE
WHY IS LIFE SO BORING
WHY ARE THERE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND
WHY DO I FEEL DIZZY
WHY ARE THERE WEEKS