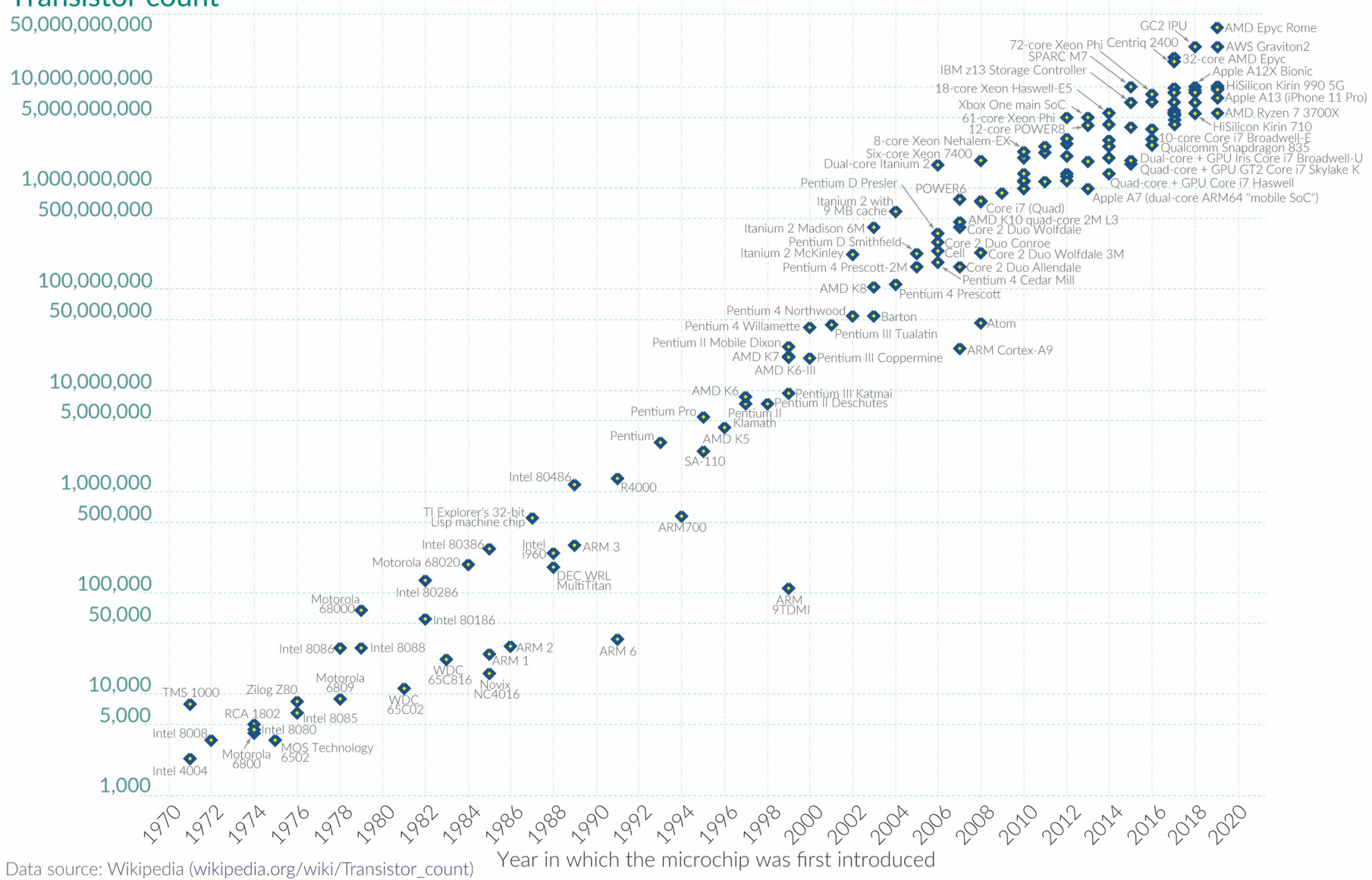
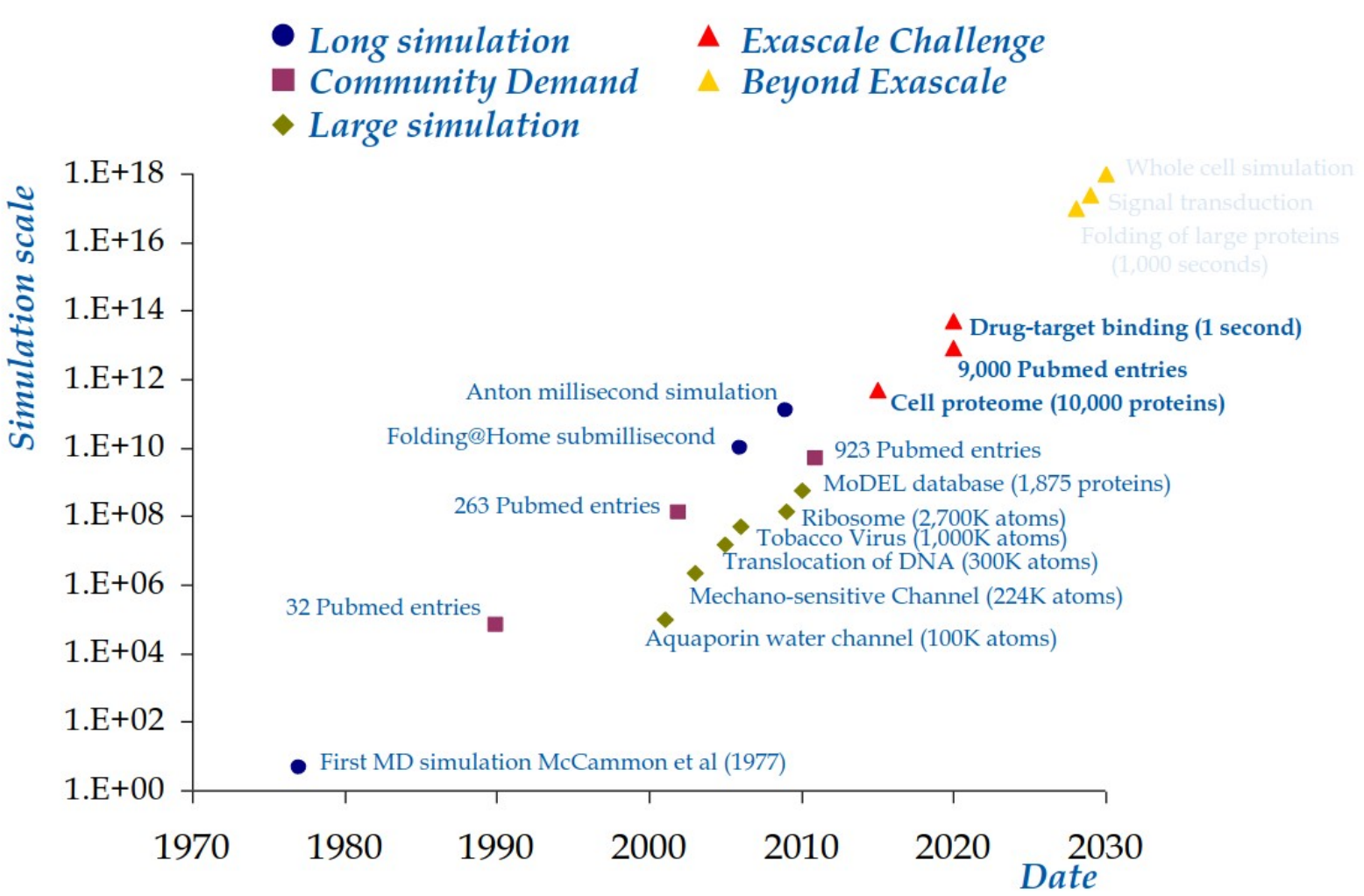
Molecular Dynamics Data Managment

Transistor count



Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)



Simulation scale for long and large simulations is calculated multiplying simulation length (nanoseconds) with size (number of atoms). For Community Demand we calculate simulation scale multiplying the number of entries un Pubmed related to protein MD by the standards of that year (nanoseconds x number of atoms).

Scientific community demands

The ABCs of molecular dynamics simulations on B-DNA, circa 2012

David L Beveridge[1,*], **Thomas E Cheatham III**[2], and **Mihaly Mezei**[3]

[1]Department of Chemistry and Molecular Biophysics Program, Wesleyan University Middletown, CT 06459, USA

[2]Department of Medicinal Chemistry, University of Utah, Salt Lake City, UT 84112, USA

[3]Department of Structural and Chemical Biology, Mount Sinai School of Medicine, New York, NY 10029, USA

MoDEL (Molecular Dynamics Extended Library): A Database of Atomistic Molecular Dynamics Trajectories

Tim Meyer,[1,2,5] Marco D'Abramo,[1,5] Adam Hospital,[1,3,5] Manuel Rueda,[1] Carles Ferrer-Costa,[1] Alberto Pérez,[1,2] Oliver Carrillo,[1] Jordi Camps,[1,2,3] Carles Fenollosa,[1,3] Dmitry Repchevsky,[1,2,3] Josep Lluis Gelpí,[1,2,3,4] and Modesto Orozco[1,2,3,4,*]

[1]Joint IRB-BSC Computational Biology Programme, Institute of Research in Biomedicine, Parc Científic de Barcelona, Baldiri Reixac 10, Barcelona 08028, Spain

[2]Barcelona Supercomputing Center, Jordi Girona 31, Edifici Torre Girona. Barcelona 08034, Spain

[3]National Institute of Bioinformatics, Parc Científic de Barcelona, Baldiri Reixac 10, Barcelona 08028, Spain

[4]Departament de Bioquímica i Biología Molecular, Facultat de Biología, Avgda Diagonal 645, Barcelona 08028, Spain

[5]These authors contributed equally to this work

*Correspondence: modesto@mmb.pcb.ub.es

DOI 10.1016/j.str.2010.07.013

Scientific community demands

BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data

Adam Hospital[1,2], Pau Andrio[2,3], Cesare Cugnasco[3,4], Laia Codo[2,3], Yolanda Becerra[3,4], Pablo D. Dans[1,2], Federica Battistini[1,2], Jordi Torres[3,4], Ramón Goñi[2,3], Modesto Orozco[1,2,3,5,*] and Josep Ll. Gelpí[2,3,5,*]

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac 10-12, 08028 Barcelona, Spain, [2]Joint BSC-IRB Research Program in Computational Biology, Baldiri Reixac 10-12, 08028 Barcelona, Spain, [3]Barcelona Supercomputing Center, Jordi Girona 29, 08034 Barcelona, Spain, [4]Dept. Computer Architecture, Technical University of Catalonia (UPC-BarcelonaTech), 08034 Barcelona, Spain and [5]Department of Biochemistry and Molecular Biology, University of Barcelona, 08028 Barcelona, Spain

Received August 27, 2015; Revised October 30, 2015; Accepted November 02, 2015

GPCRMD

How Do Molecular Dynamics Data Complement Static Structural Data of GPCRs

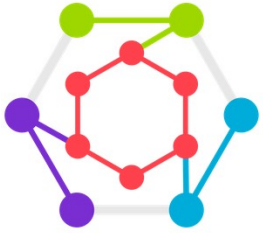
Mariona Torrens-Fontanals [1], Tomasz Maciej Stepniewski [1,2,3], David Aranda-García [1], Adrián Morales-Pastor [1], Brian Medel-Lacruz [1] and Jana Selent [1,*]

1 Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM) —Department of Experimental and Health Sciences, Pompeu Fabra University (UPF), 08003 Barcelona, Spain; mariona.torrens@upf.edu (M.T.-F.); tm.stepniewski@gmail.com (T.M.S.); darandagar@gmail.com (D.A.-G.); drnmoralespastor@gmail.com (A.M.-P.); brianmedelmo@gmail.com (B.M.-L.)
2 InterAx Biotech AG, PARK innovAARE, 5234 Villigen, Switzerland
3 Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, 02-093 Warsaw, Poland
* Correspondence: jana.selent@upf.edu

Received: 24 June 2020; Accepted: 15 August 2020; Published: 18 August 2020

check for updates

Scientific community demands



OPTIMADE
Open Databases Integration for Materials Design

Article | Open Access | Published: 12 August 2021

OPTIMADE, an API for exchanging materials data

Casper W. Andersen, Rickard Armiento, Evgeny Blokhin, Gareth J. Conduit, Shyam Dwaraknath, Matthew L. Evans, Ádám Fekete, Abhijith Gopakumar, Saulius Gražulis, Andrius Merkys, Fawzi Mohamed, Corey Oses, Giovanni Pizzi, Gian-Marco Rignanese ✉, Markus Scheidgen, Leopold Talirz, Cormac Toher, Donald Winston, Rossella Aversa, Kamal Choudhary, Pauline Colinet, Stefano Curtarolo, Davide Di Stefano, Claudia Draxl, … Xiaoyu Yang + Show authors

Molecular Dynamics Data Bank. The European Repository for Biosimulation Data

What is the MD data?

Topology

- Atom names
- Atom elements
- Atom charges
- Atom bonds (sometimes)
- Residue names
- Residue numeration
- Chain names (sometimes)
- Other constants

Trajectory

- Coordinates
- Other non-constants

Simulation parameters

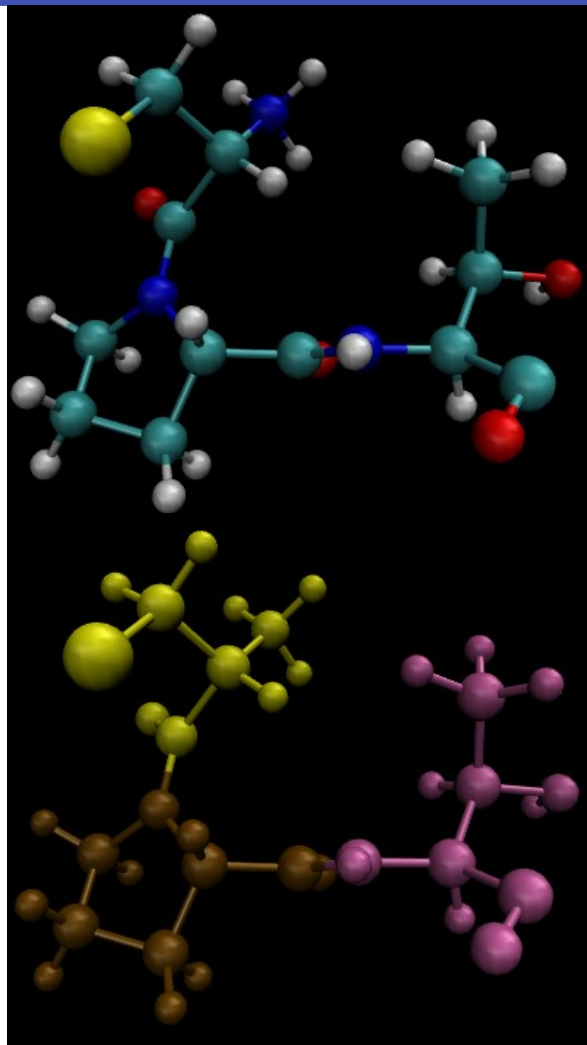
- Force field(s)
- Temperature
- Pressure
- Other constants

Topology

```
ATOM      1  N    CYS A  32
ATOM      2  HT1  CYS A  32
ATOM      3  HT2  CYS A  32
ATOM      4  HT3  CYS A  32
ATOM      5  CA   CYS A  32
ATOM      6  HA   CYS A  32
ATOM      7  CB   CYS A  32
ATOM      8  HB1  CYS A  32
ATOM      9  HB2  CYS A  32
ATOM     10  SG   CYS A  32
ATOM     11  C    CYS A  32
ATOM     12  O    CYS A  32
ATOM     13  N    PRO A  33
ATOM     14  CD   PRO A  33
ATOM     15  HD1  PRO A  33
ATOM     16  HD2  PRO A  33
ATOM     17  CA   PRO A  33
ATOM     18  HA   PRO A  33
ATOM     19  CB   PRO A  33
ATOM     20  HB1  PRO A  33
ATOM     21  HB2  PRO A  33
ATOM     22  CG   PRO A  33
ATOM     23  HG1  PRO A  33
ATOM     24  HG2  PRO A  33
ATOM     25  C    PRO A  33
ATOM     26  O    PRO A  33
ATOM     27  N    THR A  34
ATOM     28  HN   THR A  34
ATOM     29  CA   THR A  34
ATOM     30  HA   THR A  34
ATOM     31  CB   THR A  34
ATOM     32  HB   THR A  34
ATOM     33  OG1  THR A  34
ATOM     34  HG1  THR A  34
ATOM     35  CG2  THR A  34
ATOM     36 HG21  THR A  34
ATOM     37 HG22  THR A  34
ATOM     38 HG23  THR A  34
ATOM     39  C    THR A  34
ATOM     40  O    THR A  34
```

Each line defines an atom

Residues are defined by several lines (atoms)



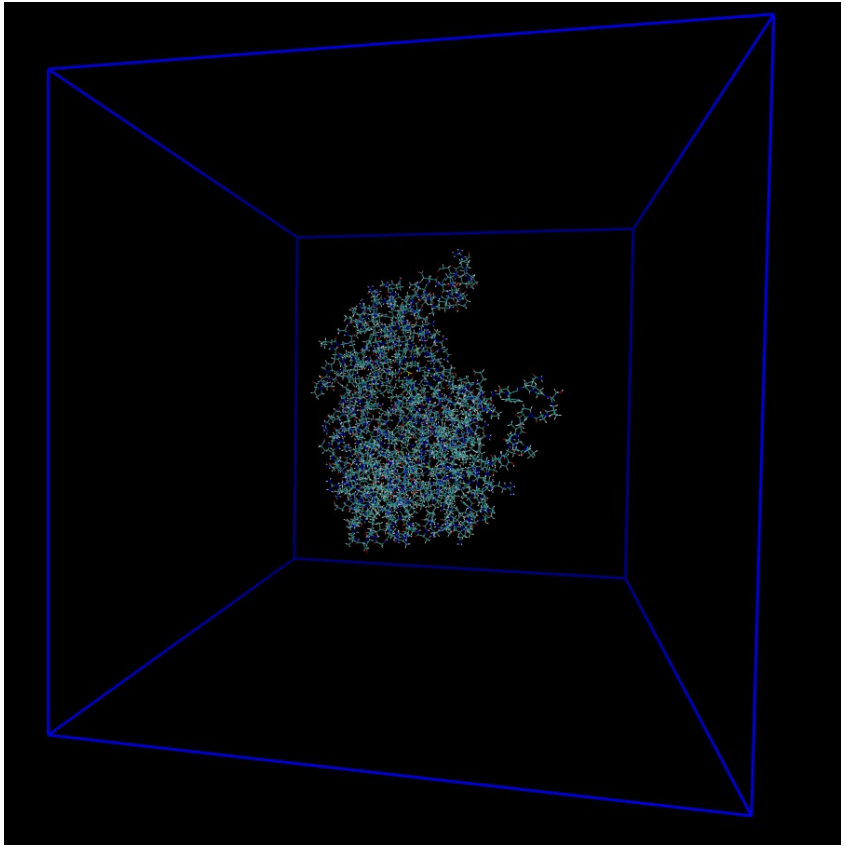
Trajectory

```
Cpptraj Generated trajectory
 101.310 111.290  27.372 101.200 110.640  28.070 100.320 109.630  27.580 100.806
 109.104  26.756  99.418 110.092  27.180  99.890 108.570  28.610  98.992 108.090
  28.220 100.880 107.560  28.780 101.100 107.330  30.160 100.422 106.562  30.530
 102.520 106.920  30.350 103.640 107.690  30.250 103.610 108.750  30.050 104.760
 107.040  30.390 104.360 105.710  30.550 105.120 104.500  30.680 106.340 104.340
  30.660 104.330 103.370  30.830 104.800 102.478  30.890 102.970 103.400  30.840
 102.350 102.260  30.990 102.868 101.404  31.130 101.344 102.280  31.080 102.230
 104.490  30.690 102.980 105.630  30.550  99.560 109.140  30.000  99.510 110.228
  29.990 100.760 108.660  30.810 101.570 109.380  30.690 100.538 108.520  31.868
  98.300 108.620  30.430  97.720 108.880  31.900  96.250 108.880  31.800  98.410
 110.050  32.490  98.210 107.590  32.680  97.680 106.300  32.400  97.940 106.010
  31.382  96.594 106.330  32.490  98.210 105.240  33.370  97.648 104.316  33.230
  99.600 105.000  33.150 100.270 105.110  34.400 100.220 104.164  34.938 101.690
 105.490  34.210 102.210 106.750  34.080 101.604 107.644  34.040 103.520 106.780
  34.010 103.880 105.440  34.080 105.170 104.810  34.050 106.270 105.370  33.980
 105.120 103.430  34.090 105.990 102.920  34.050 103.960 102.730  34.170 104.062
 101.430  34.140 104.962 100.974  34.100 103.204 100.896  34.180 102.740 103.270
  34.250 102.760 104.630  34.190  98.060 105.690  34.840  97.326 106.486  34.960
  99.460 106.160  35.170  99.610 107.168  34.780  99.660 106.120  36.240  97.820
 104.600  35.700  96.440 104.430  36.450  95.400 104.130  35.460  96.270 105.570
  37.380  96.810 103.100  37.260  96.960 101.880  36.540  97.200 102.110  35.502
  96.008 101.350  36.550  98.070 100.960  37.050  97.980 100.018  36.508  99.370
 101.480  36.810 100.130 101.430  38.010 100.680 100.490  38.070 101.080 102.570
  38.010 100.840 103.920  38.080  99.846 104.318  38.230 101.900 104.680  37.930
 102.930 103.750  37.750 104.340 103.930  37.510 104.970 104.980  37.380 105.030
 102.730  37.410 106.032 102.760  37.280 104.440 101.510  37.480 105.210 100.470
  37.360 106.208 100.560  37.230 104.788  99.558  37.460 103.140 101.300  37.670
 102.440 102.460  37.810  97.940 100.670  38.540  96.974 101.022  38.902  99.100
 101.460  39.140  98.792 102.484  39.348  99.482 100.988  40.044  98.010  99.280
  38.760  97.570  98.620  40.150  96.780  97.430  39.780  96.990  99.670  41.000
  98.990  98.190  40.720  99.820  97.290  40.000  99.820  97.550  38.942  99.432
  96.276  40.100 101.260  97.340  40.480 101.830  96.600  39.920 101.840  98.620
  40.240 102.590  99.000  41.380 103.574  98.532  41.350 102.750 100.480  41.410
```

Each 3 coordinates define the position of 1 atom (x, y, z)

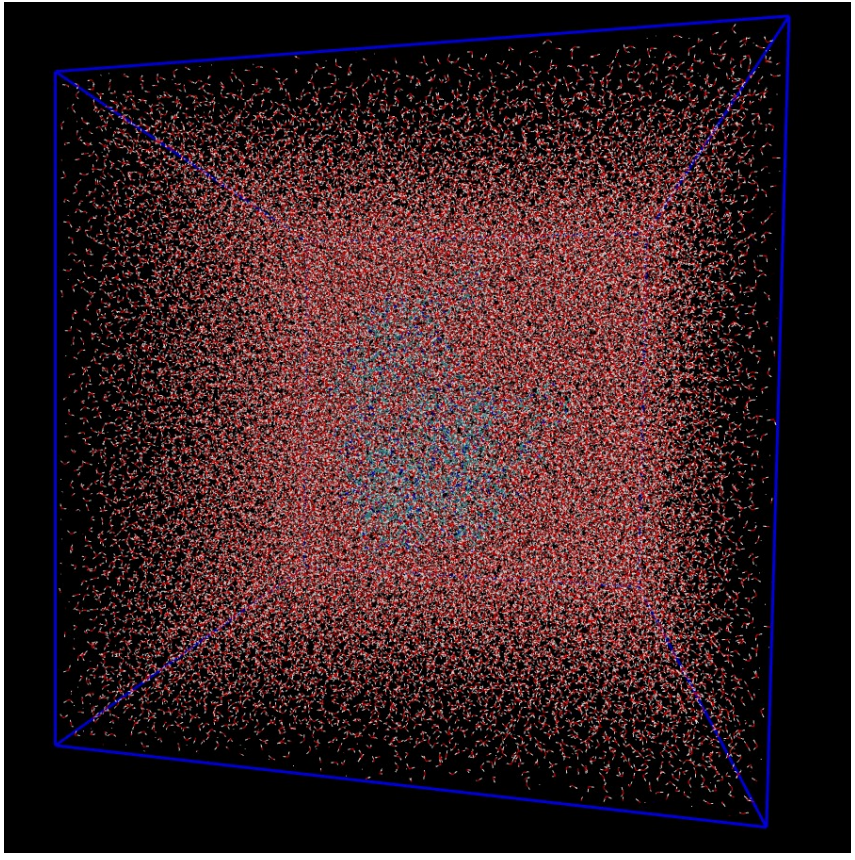
Trajectory

Protein: 7217 atoms



Add water
and ions

System: 187575 atoms



Trajectory

187575 atoms x 3 coordinates per atom = **562725 coordinates**

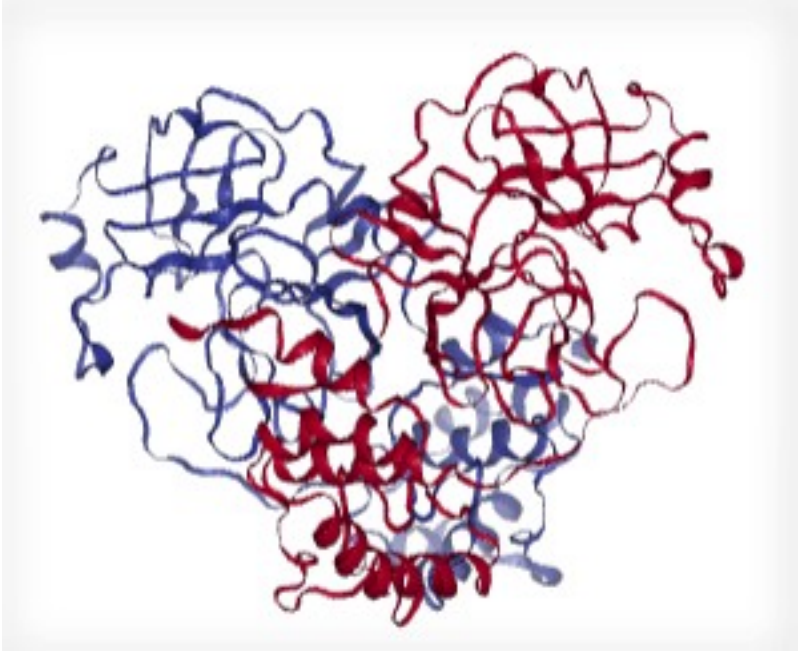
1000 ns of total simulation / 0.1 ns per frame = **10000 frames**

562725 coordinates per frame * 10000 frames =
5627250000 coordinates

5627250000 coordinates * 4 bytes per coordinate =
22509000000 bytes = 22.509 Gb

Topology + Trajectory

If we have the topology and the trajectory we can combine both to display the simulation results



Simulation parameters

```
; RUN CONTROL PARAMETERS
integrator               = md
; Start time and timestep in ps
tinit                    = 0.0
dt                       = 0.002
nsteps                   = 50000000
; For exact run continuation or redoing part of a run
init-step                = 0
; Part index is updated automatically on checkpointing
simulation-part          = 1
; Multiple time-stepping
mts                      = no
; mode for center of mass motion removal
comm-mode                = Linear
; number of steps for center of mass motion removal
nstcomm                  = 100
; group(s) for center of mass motion removal
comm-grps                =

; LANGEVIN DYNAMICS OPTIONS
; Friction coefficient (amu/ps) and random seed
bd-fric                  = 0
ld-seed                  = -1

; ENERGY MINIMIZATION OPTIONS
; Force tolerance and initial step-size
emtol                    = 10
emstep                   = 0.01
; Max number of iterations in relax-shells
niter                    = 20
```

```
; TEST PARTICLE INSERTION OPTIONS
rtpi                     = 0.05

; OUTPUT CONTROL OPTIONS
; Output frequency for coords (x), velocities (v) and forces (f)
nstxout                  = 500000
nstvout                  = 500000
nstfout                  = 0
; Output frequency for energies to log file and energy file
nstlog                   = 5000
nstcalcenergy            = 100
nstenergy                = 5000
; Output frequency and precision for .xtc file
nstxout-compressed       = 5000
compressed-x-precision   = 1000
; This selects the subset of atoms for the compressed
; trajectory file. You can select multiple groups. By
; default, all atoms will be written.
compressed-x-grps        =
; Selection of energy groups
energygrps               = Protein Lipid SOL_ION

; NEIGHBORSEARCHING PARAMETERS
; cut-off scheme (Verlet: particle based cut-offs)
cutoff-scheme            = Verlet
; nblist update frequency
nstlist                  = 10
; Periodic boundary conditions: xyz, no, xy
pbc                      = xyz
periodic-molecules       = no
```

Additional data

- Burocratic metadata
- Protein sequence references
- Non-classical MD annotations
- A lot of analyses

Data life cycle

Data life cycle

- Download raw data and have a look

Data life cycle

Data life cycle

- Download raw data and have a look
- Merge files and parse/convert formats

Data life cycle

Amber-MD/**cpptraj**

Biomolecular simulation trajectory/data analysis.

ProDy

Protein Dynamics & Sequence Analysis

MD ANALYSIS

MDTRAJ

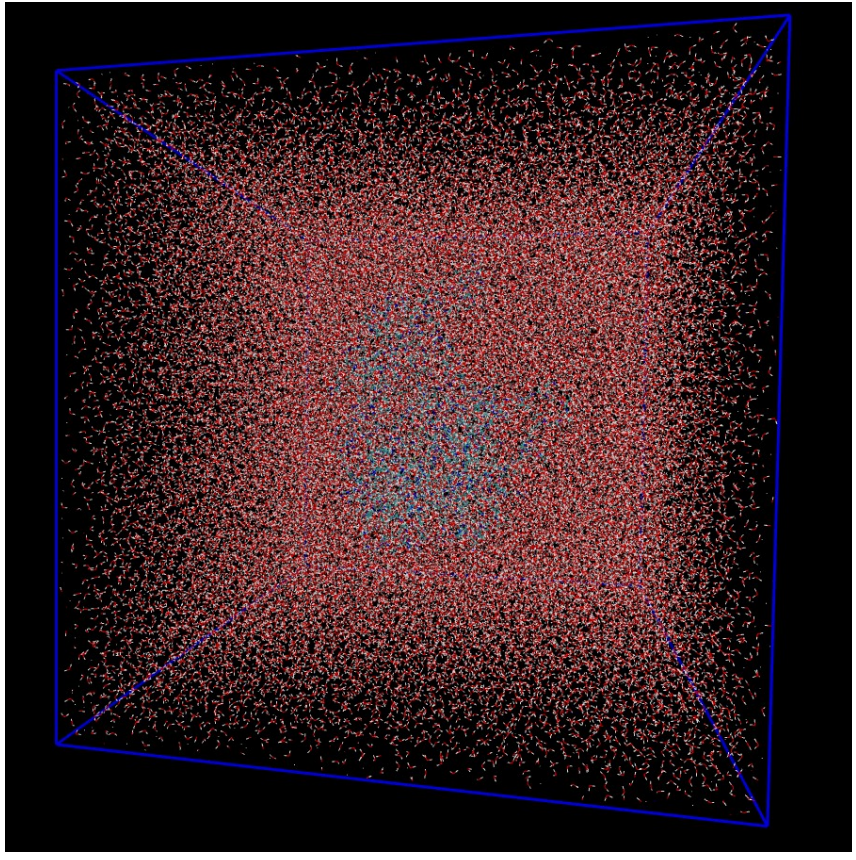
VMD

Data life cycle

- Download raw data and have a look
- Merge files and parse/convert formats
- Filter target atoms

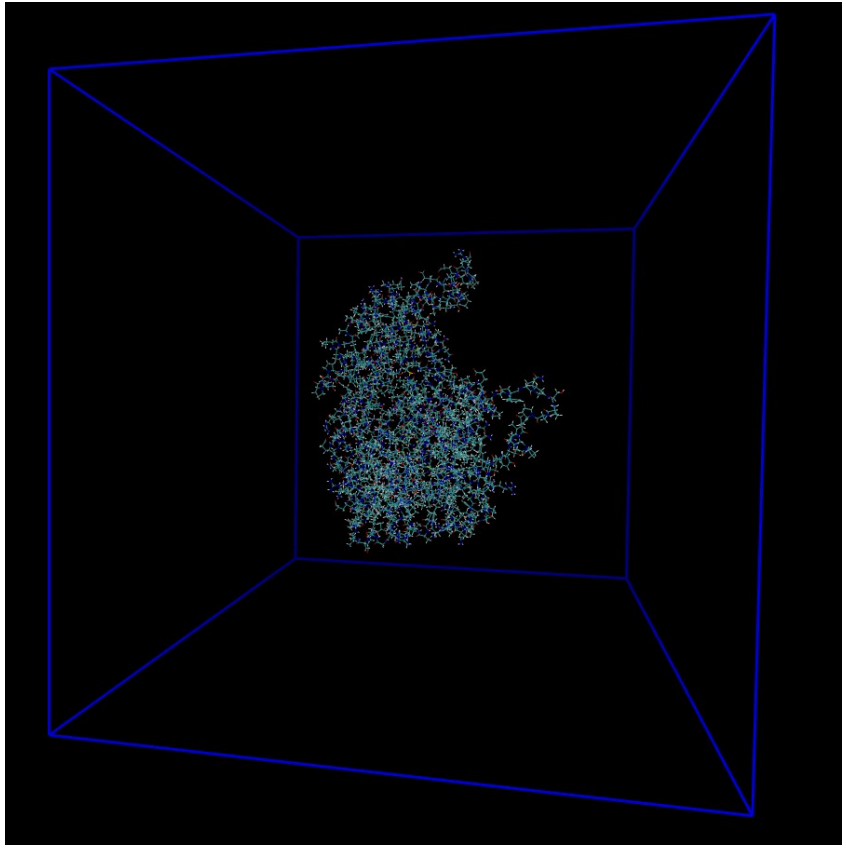
Data life cycle

System: 187575 atoms



Remove water and ions

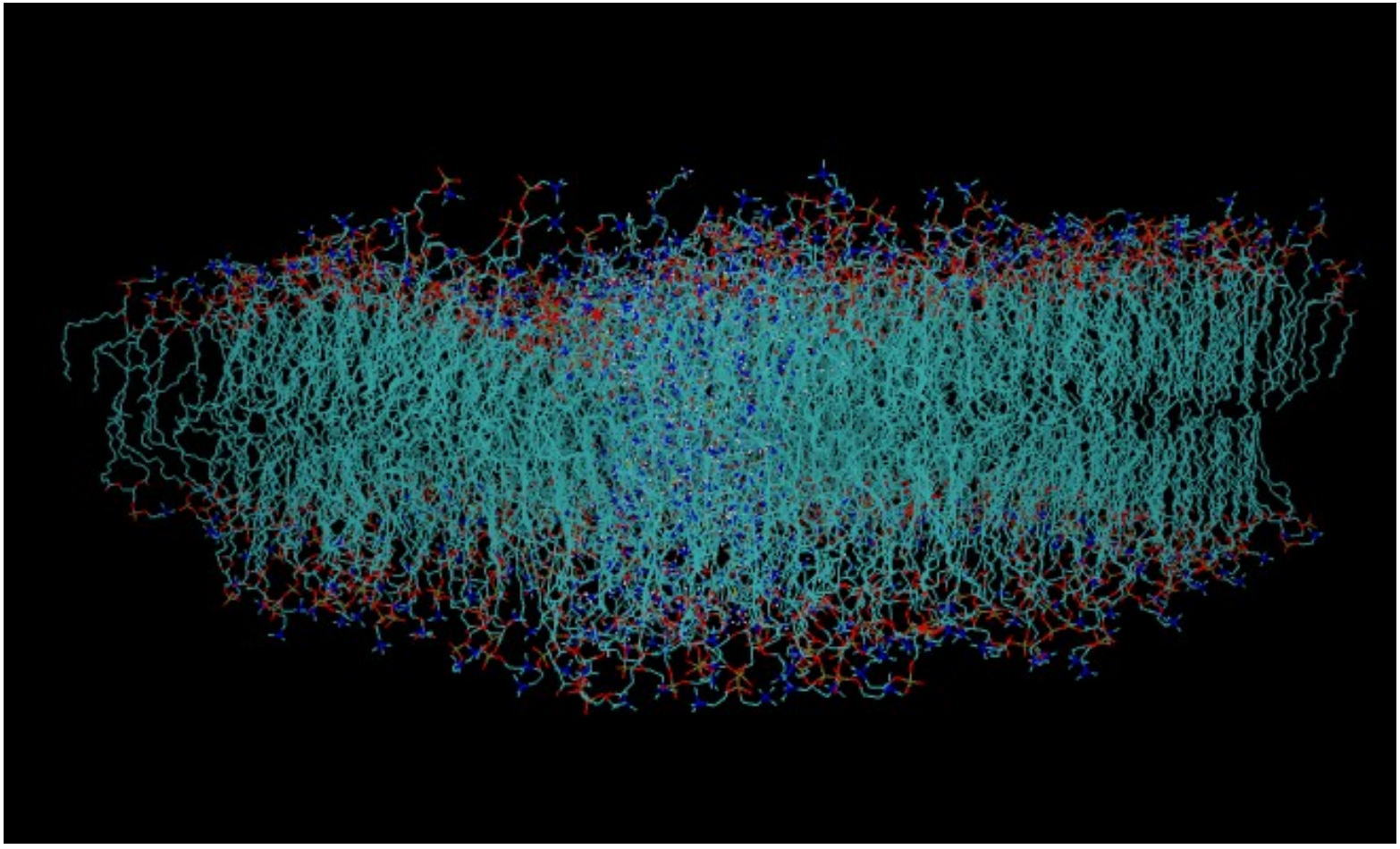
Protein: 7217 atoms



Data life cycle

- Download raw data and have a look
- Merge files and parse/convert formats
- Filter target atoms
- Imaging and fitting

Data life cycle



Data life cycle

- Download raw data and have a look
- Merge files and parse/convert formats
- Filter target atoms
- Imaging and fitting
- Check and fix/standarize topology and trajectory

Data life cycle

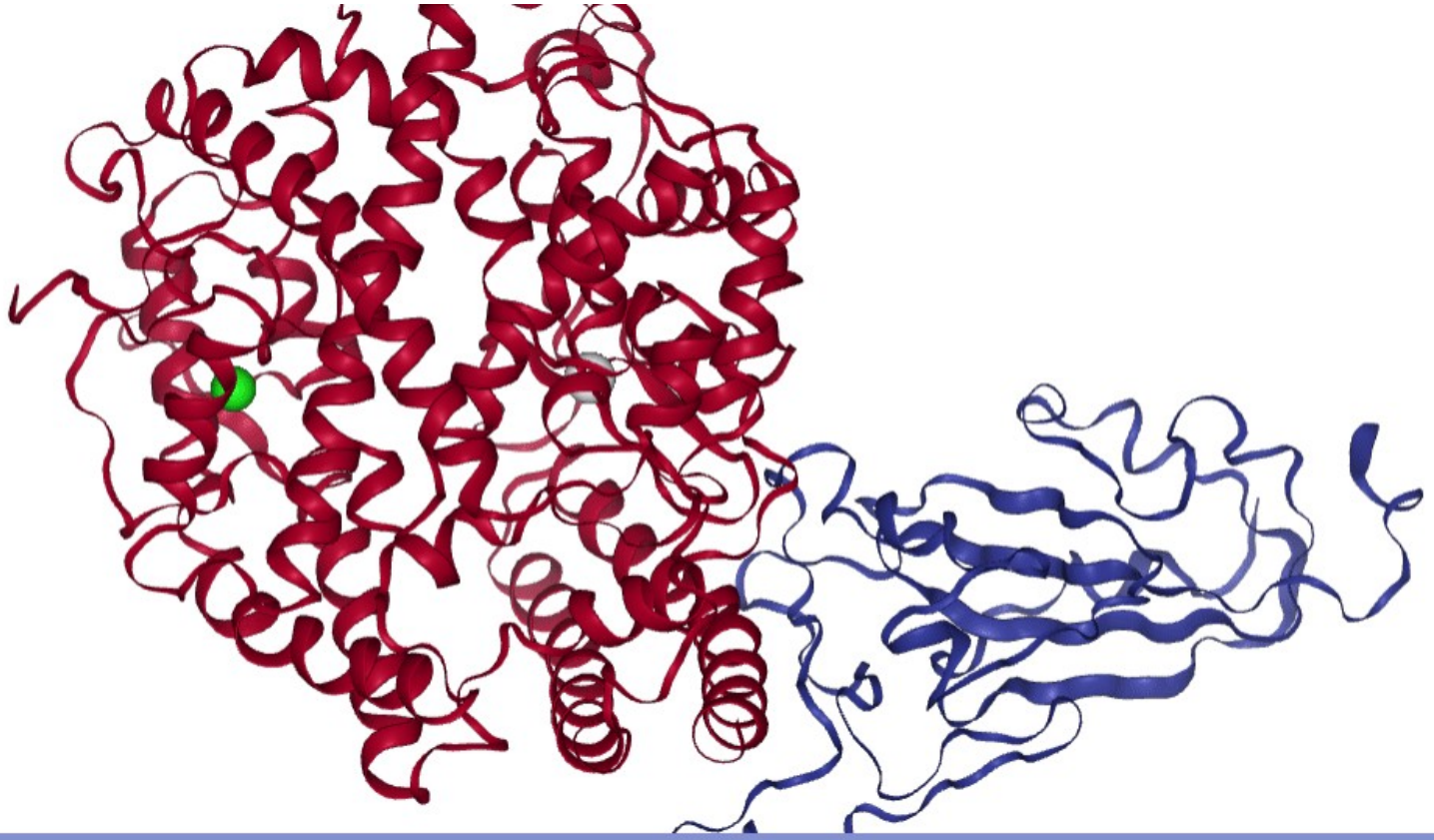
- Download raw data and have a look
- Merge files and parse/convert formats
- Filter target atoms
- Imaging and fitting
- Check and fix/standarize topology and trajectory
- Analyze data

Data life cycle

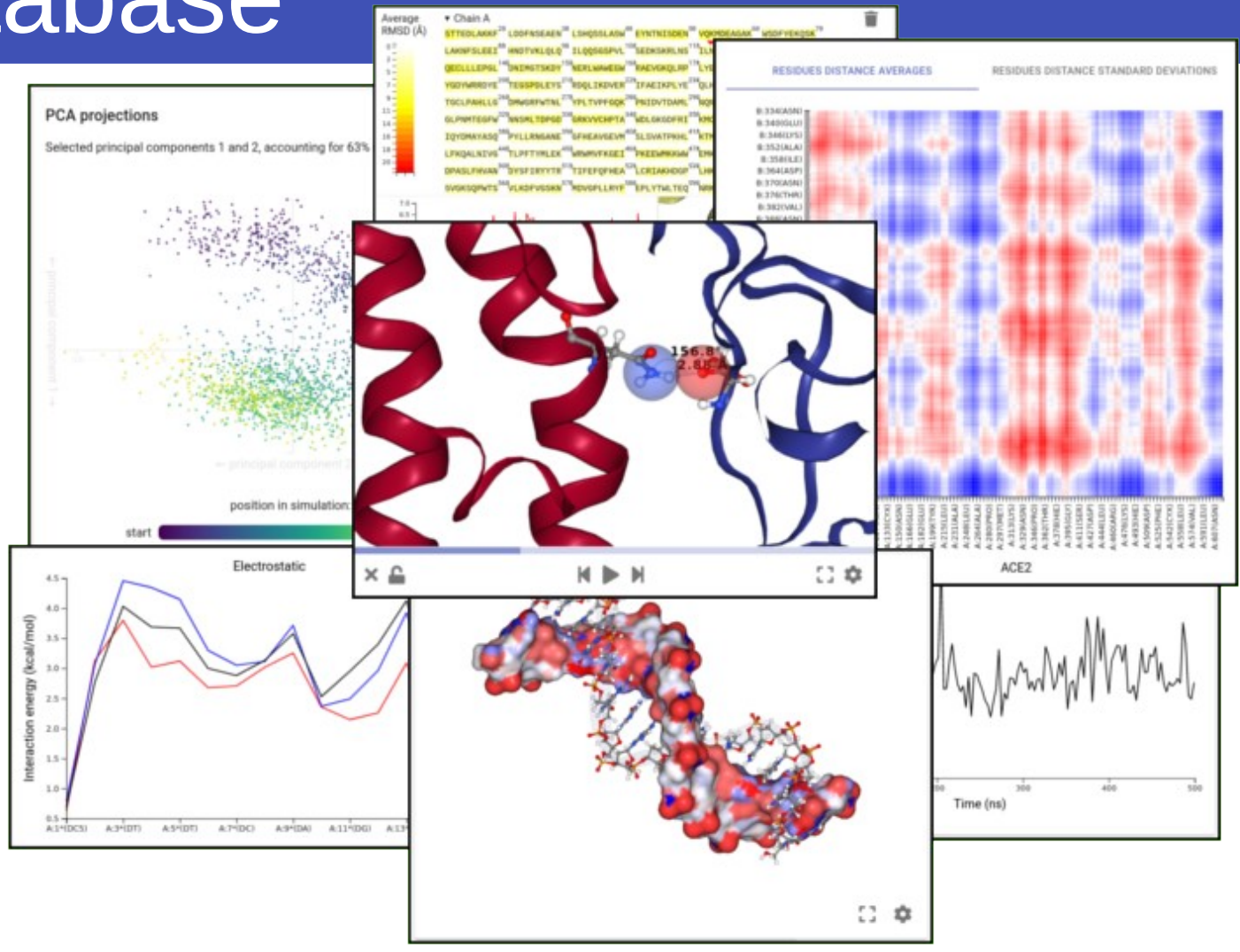
- Download raw data and have a look
- Merge files and parse/convert formats
- Filter target atoms
- Imaging and fitting
- Check and fix/standarize topology and trajectory
- Analyze data
- Publish curated data and analyses

The MD database

The MD database



The MD database



The MD database

Hands on: Using the API