Princeton GROMACS Users meeting

Why would you choose GROMACS over other MD engines? (and some GROMACS tips)

Hint: It is about water.

Gül Zerze

GMX is not the only fast & free MD engine

— But it's been advertised for "biomolecular" simulations at atomistic resolution.
— GMX is particularly good at atomistic simulations, ranging from 100 atoms to 10,000,000 atoms, for two main reasons (subjective):

- GMX processes human readable, easily modifiable (not hard-coded), and COMPACT descriptors of necessarily complicated interactions between atoms. That is called a force field file.

- A lot of biomolecular simulations are "large solute in solvent" kind. Solvent is typically water (or some aqueous solution). Water is particularly fast in GMX. On top of that, GMX provides PBC treatment for the most compact possible space-filling boxes, like truncated octahedron or rhombic dodecahedron, which are particularly useful for large solute in solvent simulations. Don't be afraid of using them. Post-processing of those boxes are out of ordinary but definitely not rocket science. And the best part is, GMX has many analysis tools already, you don't have to post-process anything yourself.

GROMACS: Great Red Ostrich Makes All Chemists Sane in water

GMX is not the only fast & free MD engine

— But it's been advertised for "biomolecular" simulations at atomistic resolution.
— GMX is particularly good at atomistic simulations, ranging from 100 atoms to 10,000,000 atoms, for two main reasons (subjective):

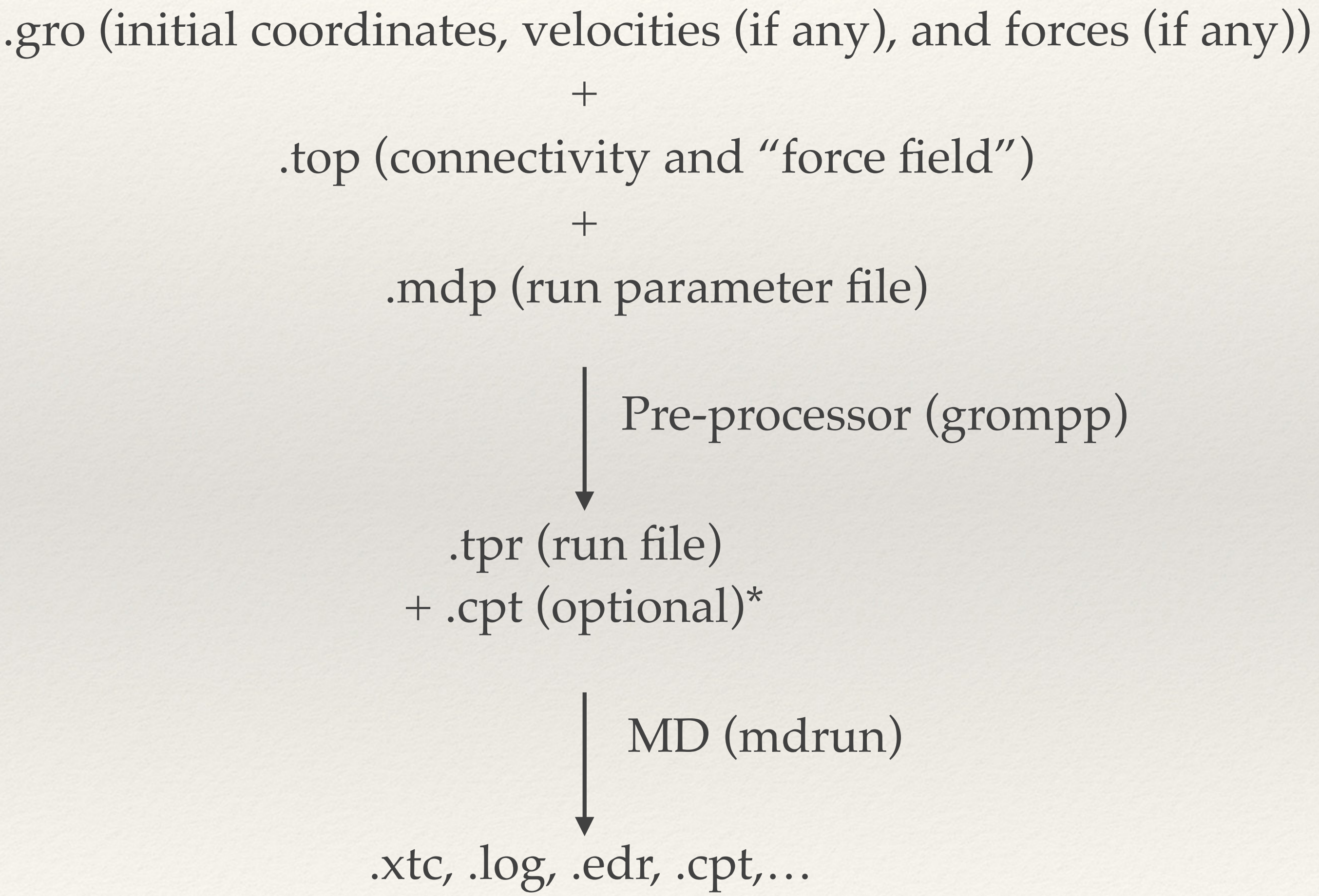
- GMX processes human readable, easily modifiable (not hard-coded), and COMPACT descriptors of necessarily complicated interactions between atoms. That is called a force field file.

- A lot of biomolecular simulations are "large solute in solvent" kind. Solvent is typically water (or some aqueous solution). Water is particularly fast in GMX. On top of that, GMX provides PBC treatment for the most compact possible space-filling boxes, like truncated octahedron or rhombic dodecahedron, which are particularly useful for large solute in solvent simulations. Don't be afraid of using them. Post-processing of those boxes are out of ordinary but definitely not rocket science. And the best part is, GMX has many analysis tools already, you don't have to post-process anything yourself.

GROMACS: ~~Great Red Ostrich Makes All Chemists Sane in water~~
GROMACS: GROningen MAchine for Chemical Simulations

Some file formats

GMX processes human-readable files but reads and writes compressed binaries.



.gro (initial coordinates, velocities (if any), and forces (if any))

+

.top (connectivity and “force field”)

+

.mdp (run parameter file)

Pre-processor (grompp)

.tpr (run file)
+ .cpt (optional)*

MD (mdrun)

.xtc, .log, .edr, .cpt,…

*.cpt is a full precision checkpoint file. Please remember to back up .cpt files before resubmitting jobs. grompp cannot reproduce cpt file if it is broken for whatever reason.

What if you don't have "GMX-friendly" files?

pdb2gmx! (for .gro and .top)

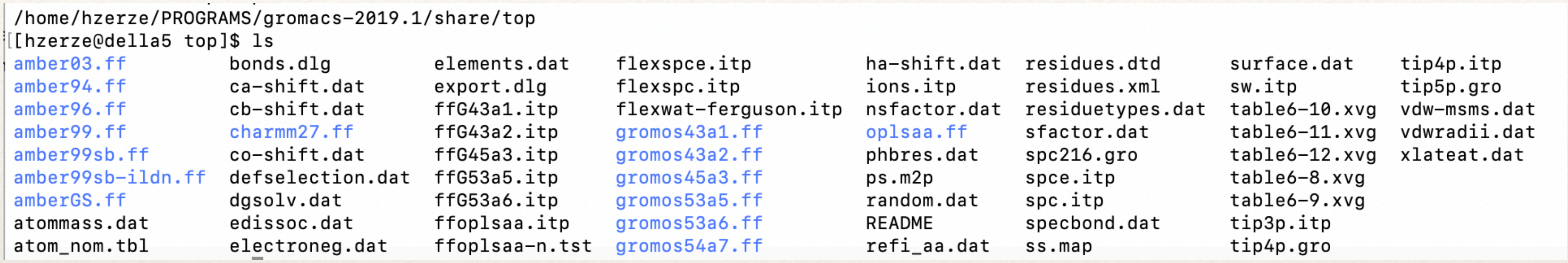
It actually reads any type of coordinate file and produce a .gro file and .top file as long as the "residues" are defined in the "force field".

Run parameter file (.mdp) is a custom file. One can specify all parameters are the run parameters of interest. Beware of defaults. It's a long list of parameters. If you'll use the defaults, make sure that you like them.

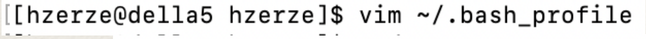
(Further reading on .mdp options: Chapter 7 of reference manual)

What is a force field file?

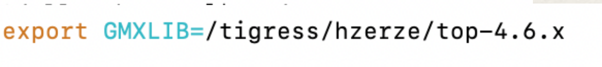
Force fields and environment variables



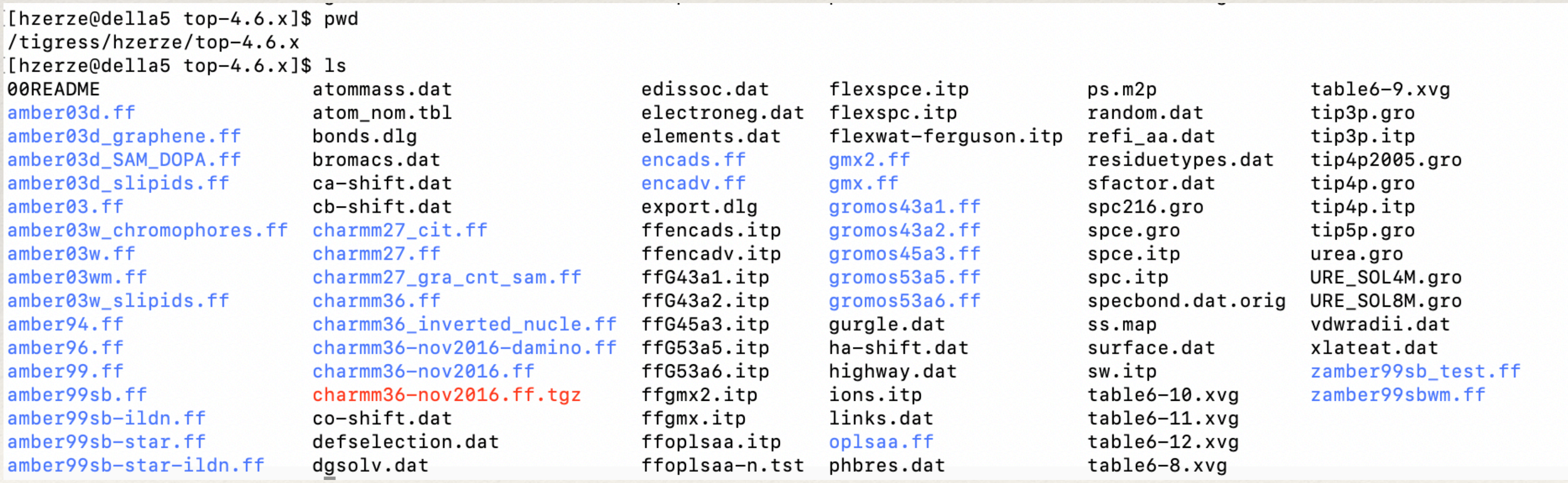
```
/home/hzerze/PROGRAMS/gromacs-2019.1/share/top
[[hzerze@della5 top]$ ls
amber03.ff          bonds.dlg         elements.dat     flexspce.itp           ha-shift.dat  residues.dtd         surface.dat       tip4p.itp
amber94.ff          ca-shift.dat      export.dlg       flexspc.itp            ions.itp      residues.xml         sw.itp            tip5p.gro
amber96.ff          cb-shift.dat      ffG43a1.itp      flexwat-ferguson.itp   nsfactor.dat  residuetypes.dat     table6-10.xvg     vdw-msms.dat
amber99.ff          charmm27.ff       ffG43a2.itp      gromos43a1.ff          oplsaa.ff     sfactor.dat          table6-11.xvg     vdwradii.dat
amber99sb.ff        co-shift.dat      ffG45a3.itp      gromos43a2.ff          phbres.dat    spc216.gro           table6-12.xvg     xlateat.dat
amber99sb-ildn.ff   defselection.dat  ffG53a5.itp      gromos45a3.ff          ps.m2p        spce.itp             table6-8.xvg
amberGS.ff          dgsolv.dat        ffG53a6.itp      gromos53a5.ff          random.dat    spc.itp              table6-9.xvg
atommass.dat        edissoc.dat       ffoplsaa.itp     gromos53a6.ff          README        specbond.dat         tip3p.itp
atom_nom.tbl        electroneg.dat    ffoplsaa-n.tst   gromos54a7.ff          refi_aa.dat   ss.map               tip4p.gro
```



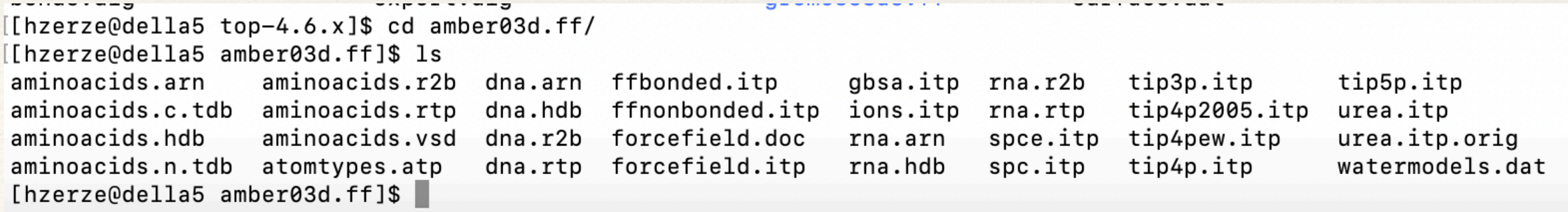
```
[[hzerze@della5 hzerze]$ vim ~/.bash_profile
```



```
export GMXLIB=/tigress/hzerze/top-4.6.x
```

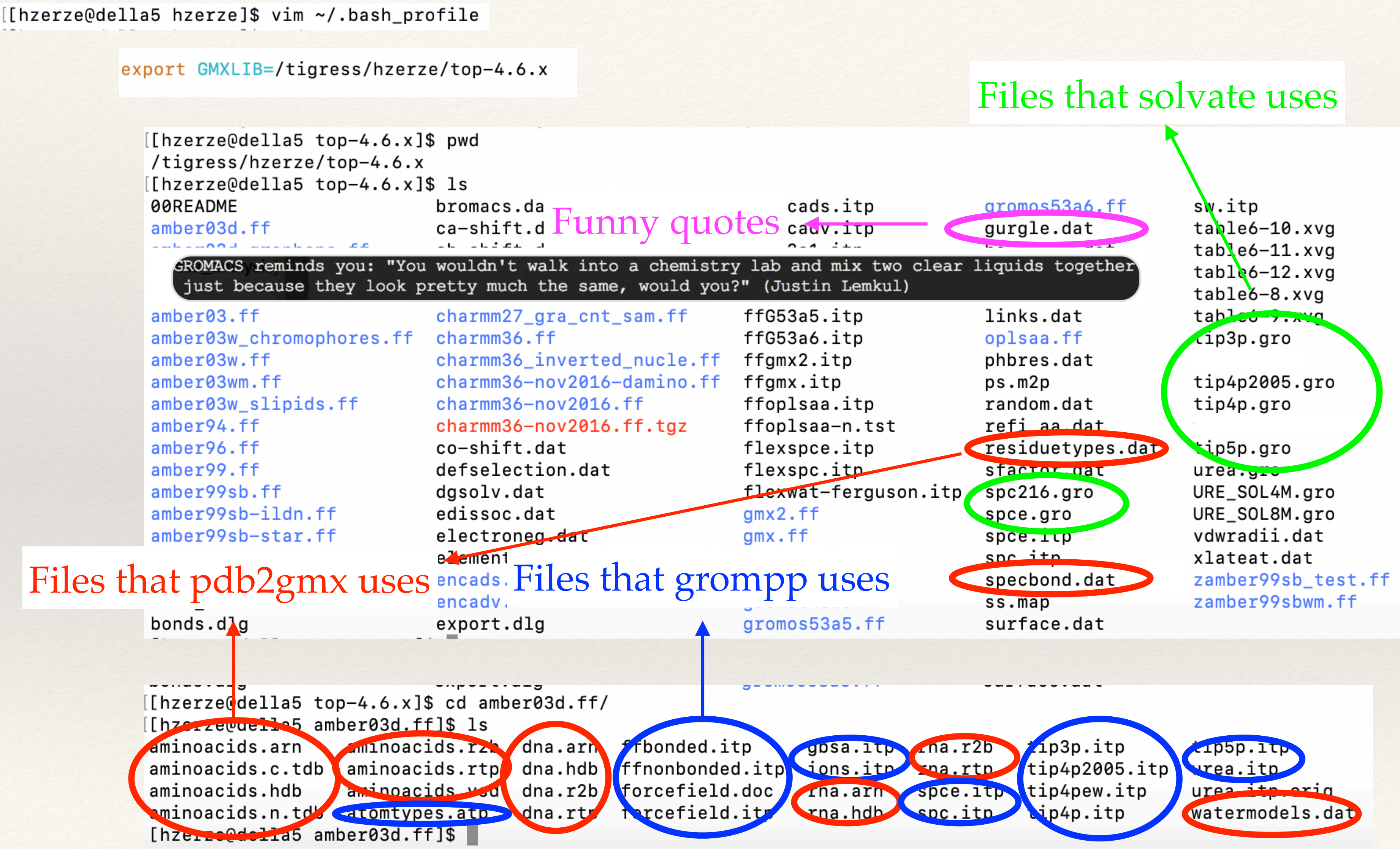


```
[hzerze@della5 top-4.6.x]$ pwd
/tigress/hzerze/top-4.6.x
[hzerze@della5 top-4.6.x]$ ls
00README                    atommass.dat                    edissoc.dat      flexspce.itp          ps.m2p               table6-9.xvg
amber03d.ff                 atom_nom.tbl                    electroneg.dat   flexspc.itp           random.dat           tip3p.gro
amber03d_graphene.ff        bonds.dlg                       elements.dat     flexwat-ferguson.itp  refi_aa.dat          tip3p.itp
amber03d_SAM_DOPA.ff        bromacs.dat                     encads.ff        gmx2.ff               residuetypes.dat     tip4p2005.gro
amber03d_slipids.ff         ca-shift.dat                    encadv.ff        gmx.ff                sfactor.dat          tip4p.gro
amber03.ff                  cb-shift.dat                    export.dlg       gromos43a1.ff         spc216.gro           tip4p.itp
amber03w_chromophores.ff    charmm27_cit.ff                 ffencads.itp     gromos43a2.ff         spce.gro             tip5p.gro
amber03w.ff                 charmm27.ff                     ffencadv.itp     gromos45a3.ff         spce.itp             urea.gro
amber03wm.ff                charmm27_gra_cnt_sam.ff         ffG43a1.itp      gromos53a5.ff         spc.itp              URE_SOL4M.gro
amber03w_slipids.ff         charmm36.ff                     ffG43a2.itp      gromos53a6.ff         specbond.dat.orig    URE_SOL8M.gro
amber94.ff                  charmm36_inverted_nucle.ff      ffG45a3.itp      gurgle.dat            ss.map               vdwradii.dat
amber96.ff                  charmm36-nov2016-damino.ff      ffG53a5.itp      ha-shift.dat          surface.dat          xlateat.dat
amber99.ff                  charmm36-nov2016.ff             ffG53a6.itp      highway.dat           sw.itp               zamber99sb_test.ff
amber99sb.ff                charmm36-nov2016.ff.tgz         ffgmx2.itp       ions.itp              table6-10.xvg        zamber99sbwm.ff
amber99sb-ildn.ff           co-shift.dat                    ffgmx.itp        links.dat             table6-11.xvg
amber99sb-star.ff           defselection.dat                ffoplsaa.itp     oplsaa.ff             table6-12.xvg
amber99sb-star-ildn.ff      dgsolv.dat                      ffoplsaa-n.tst   phbres.dat            table6-8.xvg
```



```
[[hzerze@della5 top-4.6.x]$ cd amber03d.ff/
[[hzerze@della5 amber03d.ff]$ ls
aminoacids.arn     aminoacids.r2b  dna.arn  ffbonded.itp     gbsa.itp  rna.r2b  tip3p.itp      tip5p.itp
aminoacids.c.tdb   aminoacids.rtp  dna.hdb  ffnonbonded.itp  ions.itp  rna.rtp  tip4p2005.itp  urea.itp
aminoacids.hdb     aminoacids.vsd  dna.r2b  forcefield.doc   rna.arn   spce.itp tip4pew.itp    urea.itp.orig
aminoacids.n.tdb   atomtypes.atp   dna.rtp  forcefield.itp   rna.hdb   spc.itp  tip4p.itp      watermodels.dat
[hzerze@della5 amber03d.ff]$
```

Force fields and environment variables



```
[hzerze@della5 hzerze]$ vim ~/.bash_profile
```

```
export GMXLIB=/tigress/hzerze/top-4.6.x
```

Files that solvate uses

```
[hzerze@della5 top-4.6.x]$ pwd
/tigress/hzerze/top-4.6.x
[hzerze@della5 top-4.6.x]$ ls
00README                     bromacs.da                   cads.itp                gromos53a6.ff   sw.itp
amber03d.ff                  ca-shift.d                   cadv.itp                gurgle.dat      table6-10.xvg
                                                                                                  table6-11.xvg
GROMACS reminds you: "You wouldn't walk into a chemistry lab and mix two clear liquids together   table6-12.xvg
 just because they look pretty much the same, would you?" (Justin Lemkul)                         table6-8.xvg
amber03.ff                   charmm27_gra_cnt_sam.ff      ffG53a5.itp             links.dat       table6-9.xvg
amber03w_chromophores.ff     charmm36.ff                  ffG53a6.itp             oplsaa.ff       tip3p.gro
amber03w.ff                  charmm36_inverted_nucle.ff   ffgmx2.itp              phbres.dat
amber03wm.ff                 charmm36-nov2016-damino.ff   ffgmx.itp               ps.m2p          tip4p2005.gro
amber03w_slipids.ff          charmm36-nov2016.ff          ffoplsaa.itp            random.dat      tip4p.gro
amber94.ff                   charmm36-nov2016.ff.tgz      ffoplsaa-n.tst          refi_aa.dat
amber96.ff                   co-shift.dat                 flexspce.itp            residuetypes.dat tip5p.gro
amber99.ff                   defselection.dat             flexspc.itp             sfactor.dat     urea.gro
amber99sb.ff                 dgsolv.dat                   flexwat-ferguson.itp    spc216.gro      URE_SOL4M.gro
amber99sb-ildn.ff            edissoc.dat                  gmx2.ff                 spce.gro        URE_SOL8M.gro
amber99sb-star.ff            electroneg.dat               gmx.ff                  spce.itp        vdwradii.dat
                             element                                              spc.itp         xlateat.dat
                             encads.                                              specbond.dat    zamber99sb_test.ff
                             encadv.                                              ss.map          zamber99sbwm.ff
bonds.dlg                    export.dlg                   gromos53a5.ff           surface.dat
```

Funny quotes

Files that pdb2gmx uses

Files that grompp uses

```
[hzerze@della5 top-4.6.x]$ cd amber03d.ff/
[hzerze@della5 amber03d.ff]$ ls
aminoacids.arn     aminoacids.r2b   dna.arn  ffbonded.itp     gbsa.itp  rna.r2b   tip3p.itp      tip5p.itp
aminoacids.c.tdb   aminoacids.rtp   dna.hdb  ffnonbonded.itp  ions.itp  rna.rtp   tip4p2005.itp  urea.itp
aminoacids.hdb     aminoacids.vsd   dna.r2b  forcefield.doc   rna.arn   spce.itp  tip4pew.itp    urea.itp.orig
aminoacids.n.tdb   atomtypes.atp    dna.rtp  forcefield.itp   rna.hdb   spc.itp   tip4p.itp      watermodels.dat
[hzerze@della5 amber03d.ff]$
```

(Further reading on interaction parameters and force fields: Chapter 4&5 of reference manual)

GPU acceleration in GMX

It supports only Verlet neighbor searching scheme. And this was a fundamental change as Verlet neighbor searching is about twice slower than optimized group neighbor searching when running solely on CPUs.

Things one can’t do with GPUs in GMX:
Essentially anything that cannot use Verlet neighbor searching scheme. What are they:
— “Frozen” groups
— Any energy exclusion requiring simulation (for whatever reason)
— Twin-range cutoff requiring force fields
— Tabulated potentials

GPU acceleration in GMX

It supports only Verlet neighbor searching scheme. And this was a fundamental change as Verlet neighbor searching is about twice slower than optimized group neighbor searching when running solely on CPUs.

Things one can't do with GPUs in GMX:
Essentially anything that cannot use Verlet neighbor searching scheme. What are they:
— "Frozen" groups
— Any energy exclusion requiring simulation (for whatever reason)
— Twin-range cutoff requiring force fields
— Tabulated potentials

Things that will (unnecessarily) slow GMX down:

— Double-precision (about twice slower)
— Updating neighbor list too frequently (especially for group neighbor searching)
— Using Verlet neighbor searching (instead of group) when running GMX solely on CPUs (about twice slower)

Post-processing GMX trajectories

trjconv
There is (almost) always a trjconv solution into the particular xtc post-processing problem that you might have.

GMX is not the only MD engine outputting trajectories in the form xtc files. Also importantly, it is not the only software that can read xtc files. However, GMX is for sure the fastest to process xtc. GMX has a ton of analysis tools (further read: Chapter 8 of reference manual) but has no visualization interface.

The way to read xtc files with other softwares for analysis:
Python (MDAnalysis or MDTraj)
C++ (XTCReader)
… probably more that I don't know …

For visualization:
Open the .gro file in VMD, Chimera, PyMol… and load .xtc file on top of the .gro file.

“This is not the end,
This is not even the beginning of the end,
This is, perhaps, the end of the beginning.”
— Reference manual: http://manual.gromacs.org/2016/manual-2016.pdf

GROMACS

Fast, Free, Flexible (and kind of funny)

— Tutorials:
http://www.mdtutorials.com/gmx/
http://www.gromacs.org/Documentation/Tutorials

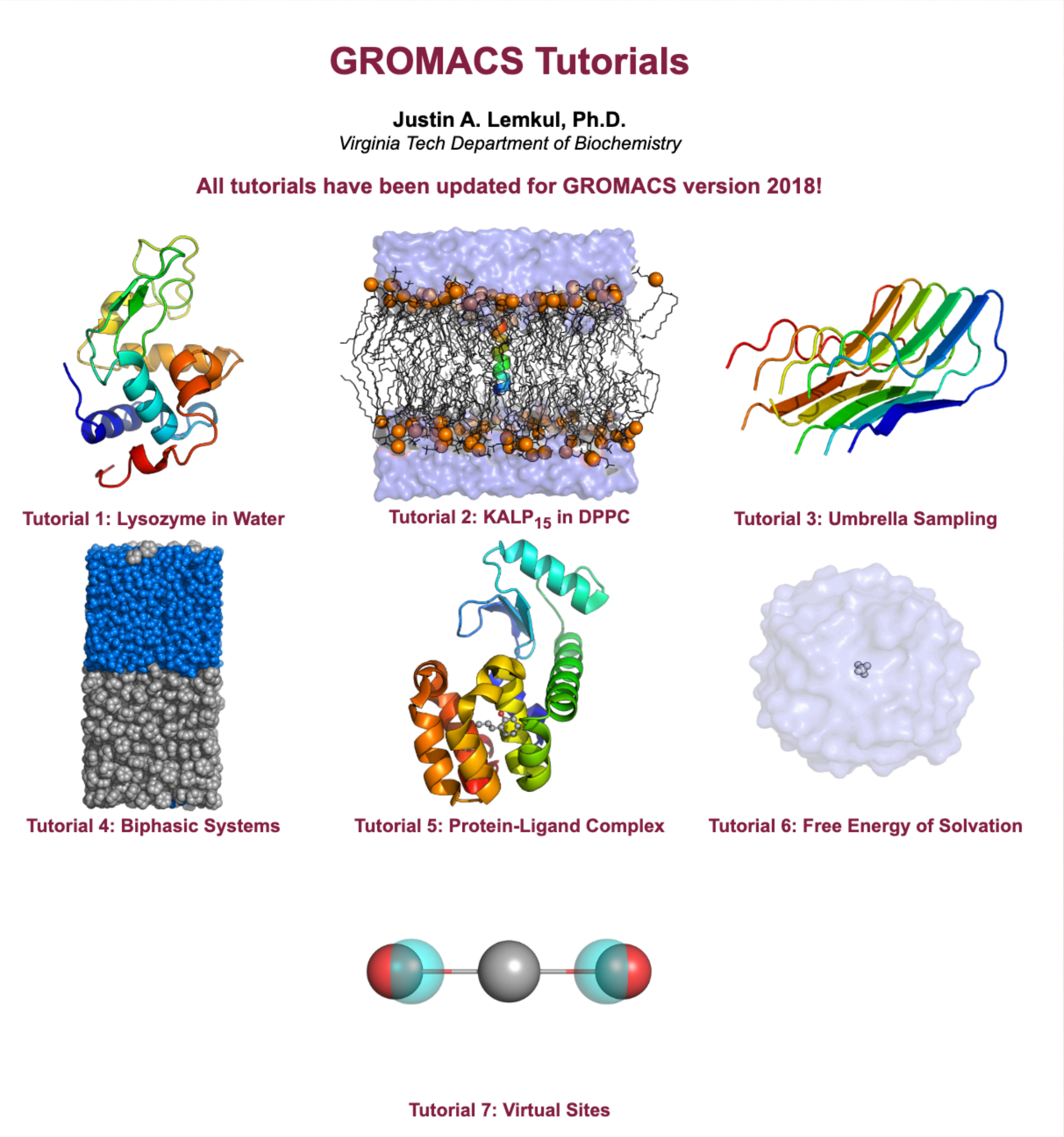
Or simply google “Justin Lemkul GROMACS tutorials”

— Ask questions:
http://www.gromacs.org/Support/Mailing_Lists

Google your problem before asking! Chances are ~95% it’s been already answered before.

— Report bugs:
https://redmine.gromacs.org/

— Feel free to reach out to me:
hzerze@princeton.edu



GROMACS Tutorials

Justin A. Lemkul, Ph.D.
Virginia Tech Department of Biochemistry

All tutorials have been updated for GROMACS version 2018!

Tutorial 1: Lysozyme in Water

Tutorial 2: $KALP_{15}$ in DPPC

Tutorial 3: Umbrella Sampling

Tutorial 4: Biphasic Systems

Tutorial 5: Protein-Ligand Complex

Tutorial 6: Free Energy of Solvation

Tutorial 7: Virtual Sites