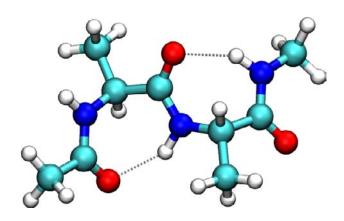
Hands-On Tutorial on Ab Initio Molecular Simulations

Berlin, July 12 - 21, 2011



Tutorial III: Weekend Research Project
Conformational space and
energetics of biomolecules:
Physical concepts and performance of
DFT-based methods

Today we investigate:

Over the weekend you will investigate:

Xaa stands for one of these:

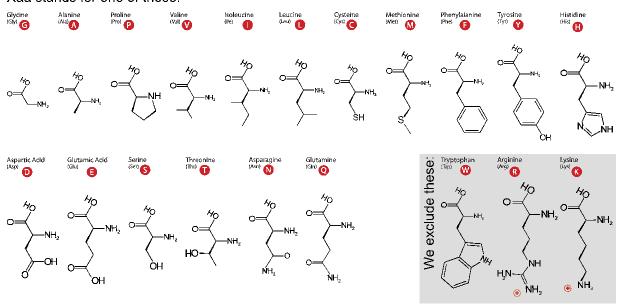


Figure 1: Chemical formula of the Ace-Ala-NMe peptide to be investigated during the tutorial today; peptides of the type Ace-Xaa-Ala-NMe and Ace-Ala-Xaa-NMe will be subject of the independent research projects for the weekend. The 20 proteinogenic amino acids are given with full name, three-letter-code (in parentheses), single-letter-code (red circles), and chemical formula. Trp, Arg, and Lys are excluded from the tutorial for technical reasons.

1 Introduction

1.1 Aim of this tutorial

The weekend research project aims at a detailed analysis of the potential-energy surface (PES) of model biomolecules, specifically tripeptides with the formula Ace-Ala-Xaa-NMe and Ace-Xaa-Ala-NMe with Xaa being one of the twenty canonical amino acids (Figure 1. The goal is understanding the complexity of the PES of even such rather "small" systems, and analyzing the description of the PES with different DFT-based methods. We start with a scan of the PES using classical molecular mechanics. Afterwards, selected conformations will be post-relaxed using van der Waals energy corrected density-functional theory (PBE+vdW). Finally, we will investigate a set of low-energy conformations using a variety of density functionals (PBE, PBE+vdW, M06-L, and PBE0+vdW). This approach, systematically performed for the 17 canonical amino acids regarded here, will shed light on how a particular amino acid within a sequence triggers the formation of a specific structure type.

1.2 Theoretical background

1.2.1 Amino acids, peptides, and their structures

The 20 proteinogenic amino acids¹ in Figure 1 share the common feature of a carboxy and an amino function. Peptides are oligomers of amino acids linked together by peptide (a.k.a. amide) bonds that are formed by a reaction between carboxy and amino groups. The sequence of these amino acids (i.e. ADPA (single letters), Ala-Asp-Pro-Ala (three-letter-code)) is referred to as the primary structure. The overall conformation of an amino acid or a peptide can be described by backbone torsion angles. The backbone of a peptide consists of the atoms forming the peptide bonds and the atoms linking them (C^{α} , C, N, cf. Figure 3a). An illustrative example for the peptide torsion angles ϕ and ψ is given in Figure 2. For each amino acid within a sequence, three backbone torsion angles ϕ , ψ , and ω (cf. Figure 3a) are measured along specific backbone atoms:

```
\omega: C^{\alpha} and C of residue (i-1) and N and C^{\alpha} of residue (i).

\phi: C of residue (i-1) and N, C^{\alpha}, and C of residue (i).

\psi: N, C^{\alpha}, and C of residue (i) and N of residue (i+1).
```

In case of repeating triplets of backbone torsion angles for the amino acids of a (sub)sequence, this part is called periodic, else aperiodic In the introductory talk you were already introduced to the general concept of peptide/protein structure formation. Here we will focus on the structures that you can actually expect to find for the model systems of this weekend research project. The secondary structure elements under investigation can be classified by the pattern of H bonds. An H bond consists of an H donor (e.g. the backbone NH group) and an acceptor (e.g. the backbone C=O group). A naming like this ' $i \leftarrow i$ ' or ' $1 \leftarrow 1$ ' describes H bonds between the same amino acid within a chain. ' $i \leftarrow i + 2$ ' or ' $1 \leftarrow 3$ ' (like in C_7^{ax} and C_7^{eq} , cf. Table 1 and Figure 3) stands for the interaction of the NH group of a third residue with a CO of a first residue, the H bond is formed backward along the sequence. A description by backbone torsion angles is helpful to distinguish between structures with H bond patterns that span more residues like ' $i \leftarrow i + 3$ ' or ' $1 \leftarrow 4$ ', namely the examples for β turns and the 3_{10} helix shown in Figure 3c. The nomenclature C_7^{ax} and C_7^{eq} refers to the formation of a 7-membered pseudocycle closed by an H bond, the side chain of the amino acid residue can be oriented parallel to the axis ('ax') or parallel to the plane ('eq', like equatorial). Idealized values for the backbone torsion angles of prominent secondary structures are given in Table 1.

In the first part of the tutorial we introduce you to the case of the alanine tripeptide Ace-Ala-Ala-NMe. It is called tripeptide because of the three peptide bonds between the two Ala residues and the Acetyl(Ace)-and Aminomethyl(NMe)-cappings. The Ala building block features a methyl 'side chain' that does not act as H bond donor or acceptor. In case of other building blocks, more complicated H bonding patterns can arise due to side chain backbone interactions as you will experience during the weekend research project (cf. Figure 1).

List of references:

DOI: 10.1002/jcc.21211 "Extensive Conformational Searches of 13 Representative Dipeptides and an Efficient Method for Dipeptide Structure Determinations Based on Amino Acid Conformers" by Yu Wenbo et al.

DOI: 10.1142/S0219633609005118 "Conformational preferences of N-acetyl-glycine-glycine-N'-methylamide: a theoretical study" by Ho-Jin Lee et al.

DOI: 10.1002/(SICI)1096-987X(199708)18:11<1415::AID-JCC8>3.0.CO;2-O "Structural and energetic relations between β turns" by Möhle et al.

DOI: 10.1002/pro.5560031206 "A revised set of potentials for β -turn formation in proteins" by Hutchinson et al.

¹http://en.wikipedia.org/wiki/Amino_acid

Table 1: Idealized backbone torsion angles of the different secondary structure motifs possible for peptides Ace-Ala-Xaa-NMe and Ace-Xaa-Ala-NMe. The structures marked with an asterisk are drawn in Figure 3. *Please note*: The values are idealized and collected from empirical data; you will find only approximate agreement with the results you produce here.

3.5.4.6	1	1		1	1
Motif	ϕ_{i+1}	ψ_{i+1}	ω	ϕ_{i+2}	ψ_{i+2}
$1 \leftarrow 1 \text{ H bond}$					
C_5^*	-160	-160			
$1 \leftarrow 3 \text{ H bond}$					
C_7^{ax*}	60	-80	180		
$C_7^{eq}*$	-80	80	180		
$1 \leftarrow 4 \text{ H bond}$					
periodic					
3_{10} -helix*	-50	-25	180	-50	-25
$1 \leftarrow 4 \text{ H bond}$					
aperiodic					
βI	-60	-30	180	-90	0
β I'	60	30	180	90	0
βII	-60	120	180	80	0
β II'*	60	-120	180	-80	0
βIII	-60	-30	180	-60	-30
β III'	60	30	180	60	30
$\beta ext{VIII}$	-60	-30	180	-120	120
$\beta VIa1*$	-60	120	0	-90	0
$\beta VIa2$	-120	120	0	-60	0
β VIb	-120	120	0	-60	150

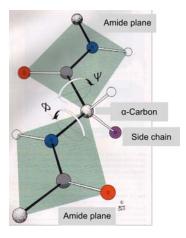


Figure 2: Illustration of peptide backbone torsion angles by Irving Geiss, taken from the textbook Voet and Voet: "Biochemie" (1994).



c) Example structures to be expected for Ac-Ala-Ala-NMe

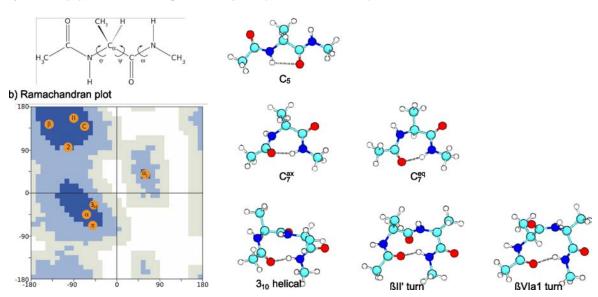


Figure 3: The schematic formula (a) illustrates the backbone torsion angles ϕ , ψ , and ω . Plotting the occurrence of ϕ , ψ combinations in proteins leads to a Ramachandran type plot that shows preferred and forbidden regions (b). Examples of structures (c) to be expected for the Ace-Ala-Ala-NMe model under investigation.

1.2.2 Potential-energy surface scan using a force field

The scan of the PES using DFT is prohibitively expensive, even for small peptides (on the available hardware). Thus, here we determine initial minimum geometries with a systematic conformational search algorithm using the scan program from the Tinker molecular mechanics package. Molecular mechanics force fields are based on classical interatomic potentials, without explicit treatment of electrons. The energy model is divided into two parts: bonded and non-bonded interactions. The first correspond to bond stretching (2-atoms), angular bending (3-atoms), and torsional (4-atoms) degrees of freedom. The non-bonded interactions correspond to electrostatic (Coulomb) and van der Waals attraction and repulsion (see figure 4) and the energy of the system with N atoms, E(N) is given by:

$$E(N) = E_{bonded} + E_{non-bonded} \tag{1}$$

$$E_{bonded} = E_{bonds} + E_{angles} + E_{torsions} \tag{2}$$

$$E_{non-bonded} = E_{electostatics} + E_{vdW} \tag{3}$$

In this exercise the so-called OPLS-AA force field is employed. The OPLS-AA force field is optimized for peptides and proteins solvated in explicit water molecules. The actual method to generate the initial geometries is called basin hopping search. In this method potential energy surface is scanned by generating trial structures around local minima and relaxing obtained structure to obtain possible new local minima (see figure 5). The procedure is then repeated for each newly found local minimum. Particularly in the Tinker package this is done by generating new trial structures from the normal modes of the molecules and relaxing the obtained trial structures using the specified force field.

This procedure of a force field-based conformational search is used solely as a pre-sampling. In order to obtain a more accurate energy hierarchy subsequent DFT calculations are performed.

1.2.3 Density-functional theory

To test the variability of the results on the DFT functional, we will use various flavors of DFT:

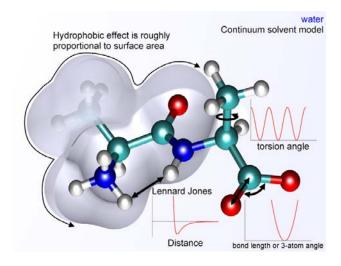


Figure 4: Schematic figure of the individual contributions to the molecular mechanics force field energy model (reproduced from Wikipedia).

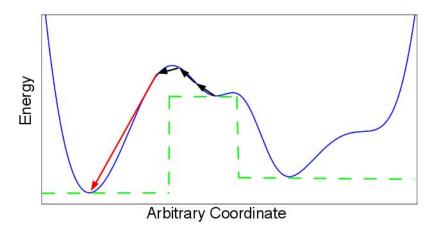


Figure 5: Schematic picture of the basing hopping algorithm in tinker scan. Blue line is potential energy surface, black arrows are trial steps taken, red arrow is the minimization of energy and dashed green line is obtained potential energy surface.

- **PBE** Widely used generalized gradient appproximated (GGA) functional by Perdew, Burke, and Ernzerhof.
- PBE+vdW PBE functional including the Tkatchenko-Scheffler van der Waals (vdW) correction.
- **M06-L** Highly empirical meta generalized gradient approximated functional from the Truhlar's group. Includes an improved description of middle-range vdW interactions compared to GGAs.
- **PBE0** Hybrid functional with PBE correlation and 25% Hartree-Fock (HF) exchange, the HF exchange partially removes the electron self-interaction present in PBE.
- **PBE0+vdW** Hybrid functional with PBE correlation and 25% Hartree-Fock (HF) exchange including the Tkatchenko-Scheffler vdW correction. The HF exchange partially removes the electron self-interaction present in PBE.

2 Practical work

In the first part of the tutorial we will guide you through the analysis of the Ace-Ala-NMe tripeptide. This includes the PES scan using the Tinker molecular mechanics package for initial geometries, the relaxation of obtained geometries with FHI-aims using the PBE+vdW method, characterizing obtained structures, and comparison of conformational energy hierarchies between different DFT functionals.

In the second part of the tutorial, the weekend research project, you will investigate a similar system on your own. Please find at the very end of this script a schedule for today and a suggested schedule for the weekend.

Step 0: Prerequisites

The folder /pub/TUTORIAL3/ contains:

reference – Copy this folder to your home

cp /pub/TUTORIAL3/reference/ ~/tutorial3

[Params] – contains the force field parameters and the modified species defaults for the relaxations.

[Scripts] – contains a selection of scripts (bash, python, awk) that can be used during the tutorial. More details can be found in the Appendix.

[Step_1..7] – inside these directories, the steps of the tutorial have to be performed. Each of them contains a file ReadMe.txt with information and commands and some of them contain bash scripts that automatize tasks.

instructions - contains the source of this document (outline-tutorial3.tex) and the presentation
 (Tutorial3_AA-example.odp)

results – contains pre-computed results

Weekend_Tasks/workstation {01..49} - files needed for the weekend project are stored here.

Step 1: Building the tripeptide model

Working directory: Step_1

The tinker.protein program requires information about the force field parameters and on the treatment of chiral centers. This information is given in the file:

AcAANMe.key – it contains:

parameters \sim /hands-on-2011-tutorials/tutorial3/reference/Params/oplsaa.prm enforce-chirality

Tinker can automatically generate three-dimensional models of peptides from a given sequence. Call the program tinker.protein and give the options as listed below to generate the initial model of the *Ace-Ala-NMe* peptide.

Create the model by running tinker.protein

Tinker requires input of several parameters:

Acaanme; Gives the basic filename

Alanine tripeptide; title of the molecule ACE; three letter code of the first residue ALA; three letter code of the second residue

ALA; ... third ...

NME; termination

[hit RETURN]; indicates the end of the sequence

n; do **not** cyclize the peptide

As a result you obtain the file Acaanme.xyz with cartesian coordinates in Tinker-format. In order to make it readable with a molecule viewer, convert it to a standard xyz format with a script from the Scripts-Folder:

../Scripts/txyz2xyz.sh AcAANMe.xyz

as a result you will get a file Acaanme.xyz.xyz, which you can visualize with your favorite viewer (available are: jmol, molden, vmd). On the big screen we will use vmd.² Follow us on the screen, on how to measure simple geometric properties.

Load molecule – VMD Main window | File | New Molecule ...

Change representation – VMD Main window | Graphics | Representations ...

Measure a distance - VMD Graphics window | Hit '2' on the keyboard | click two atoms ...

Measure a bond angle - VMD Graphics window | Hit '3' on the keyboard | click three atoms ...

Measure a torsion angle - VMD Graphics window | Hit '4' on the keyboard | click four atoms ...

Revert to rotation mode – VMD Graphics window | Hit 'r' on the keyboard

Delete/Manage labels - VMD Main window | Graphics | Labels ...

Step 2: Scan the PES of Ac-Ala-Ala-NMe for minima

Working directory: Step_2

First, copy the files Acaanme.key and Acaanme.xyz from directory Step_1. Next step is to scan potential energy surface using tinker.scan program with following command:

Scan the PES mpirun -n 4 tinker.scan AcAANMe.xyz 0 10 20 0.0001 0 | tee AcAANMe.log
The Tinker tinker.scan routine performs a conformational search on the molecule given by (AcAANMe.xyz).
The second argument points to automatic selection (0) of the torsion angles to vary during the search. The third argument specifies the number of search directions to follow from every newly found minimum (here 10). During the search, only conformations within a certain energy window

will be accepted (the fourth argument, here: 20 kcal/mol). The decision if a structure has already been found during the scan is based on a simple criterion of energy similarity (fifth argument, here:

²www.ks.uiuc.edu/Research/vmd/

 $^{^3}$ http://www.ks.uiuc.edu/Training/Tutorials/vmd/tutorial-html/index.html

0.0001). Sixth argument states that we do not continue a previously stopped search. Make sure to save the output ('| tee Acaanme.log') as it contains the conformational energies of the minima found.

Extract the energies — extract the energy (in kcal/mol) from the output with grep, only keep columns 5 (Structure ID) and 6 (OPLS energy) with awk, and sort according to the second column (energy) with sort. Save in the file StructID_Eopls.dat: grep 'Map' AcAANMe.log | \
awk '{printf "%03s %6.4f \n", \$5,\$6}' | \
sort -n -k 2 > StructID_Eopls.dat
or invoke the script '01_get-energies.sh'.

Convert the energies – to convert from kcal/mol to eV:

```
awk '{print $1,($2*0.04336)}' StructID_Eopls.dat > StructID_Eopls.eV.dat
```

Convert the coordinate files - by using the script '02_convert2xyz.sh'.

Step 3: DFT relaxation of force field structures with PBE+vdW

Working directory: Step_3

In this step, selected low-energy force field geometries will be relaxed with FHI-aims. The usage of FHI-aims and the syntax of the control.in and geometry.in input files were already discussed during the workshop. The options in control.in:

```
xc pbe
vdw_correction_hirshfeld
sc_accuracy_rho 1E-4
sc_accuracy_eev 1E-3
sc_accuracy_etot 1E-6
sc_iter_limit 100
empty_states 3
relax_geometry trm 1.e-2
```

The computing times prohibit the relaxation of more than one or two OPLS-conformers with PBE+vdW during this afternoon session. Check the input file control.in.PBEvdW_relax, we work with reduced species defaults due to the time restrictions.

The task will be distributed, please proceed as follows:

- 1. Create a directory named Conf_[01..49] according to the number of your workstation: e.g.: mkdir Conf_23
- 2. Check the file StructID_Eopls.dat, where the conformations are sorted according to their energy. Select the conformation (three-digit number, e.g.: rank 23 is Structure number 040) with the rank (in rising energy) equal to your workstation number. Convert the according coordinate file to an FHI-aims geometry.in file in Conf_[Number of your workstation]:

```
tail -n 32 ../Step_2/AcAANMe.040.xyz | \
awk '{print "atom", $2,$3,$4,$1}' > Conf_23/geometry.in
```

- Create a control.in in the directory:
 cp control.in.PBEvdW_relax Conf_23/control.in
- 4. Enter the directory (cd Conf_23) and start the FHI-aims run:

 mpirun -np 4 aims.hands-on-2011.scalapack.mpi.x >& run.PBEvdW_relax &
- 5. You can monitor the convergence of the simulation: get_relaxation_info.pl run.PBEvdW_relax

- 6. During the waiting time you can proceed to Step_4 and analyze the output of the force field conformational scan.
- 7. Copy the final directory (once the simulation is finished) to /pub/TUTORIAL3/results/Step_3

Step 4: Analyzing the force field scan results

Working directory: Step_4

Plot energy hierarchy

First copy the energy file from Step 2 (StructID_Eopls.eV.dat) and the plotting script (plot_hierarchy.py) here. Edit plot_hierarchy.py (description within the file) and run with: python plot_hierarchy.py

Visualize low energy conformers

The files in folder Step_2 have already been converted to standard xyz format. Visualize the most stable, the least stable, and some intermediate conformations. What are the interactions (hydrogen bonds, van der Waals) that stabilize Ace-Ala-NMe conformations? Check the structure types from Figure 3 and try to assign some of them to the structures you just visualize. Tips on the usage of VMD were given under Step 1.

Measure backbone torsion angles

A further way to characterize short peptides is based on their backbone geometries. The awk-script xyz-geometry.awk is employed by 02_get-angles.sh to measure torsion angles.

Compare the obtained backbone torsion angles with the idealized values given in Table 1.

Step 5: Analysing results of the PBE+vdW relaxation

Working directory: Step_5

Results directory: /pub/TUTORIAL3/results/Step_3

Get the data

All of you should now have stored the results in /pub/TUTORIAL3/results/Step_3 Take a look at the script 01_get-results.sh and run it to get coordinate files and an energy list.

Plot energy hierarchy

The script above gives you a energy list (OPLSrank_Epbevdw.dat), use plot_hierarchy.py from the Scripts directory to plot the hierarchy. Edit plot_hierarchy.py (description within the file) and run with:

python plot_hierarchy.py [Are you using eV or kcal/mol?]

In order to compare the hierarchies from OPLS and PBE+vdW, you can use plot_hierarchy.py with two input files. Create a appropriate file with the command:

head -n 50 ../Step_2/StructID_Eopls.eV.dat | nl | awk '{printf "%02s %6.4f \n", \$1,\$3}' > OPLSrank_Eopls.dat

Now edit the Python script accordingly and run it.

Visualize low energy conformers

Visualize the conformations (stored in xyz format in this directory). What are the interactions (hydrogen bonds, van der Waals) that stabilize Ace-Ala-Ala-NMe conformations? Check the structure types from Figure 3 and try to assign some of them to the structures you just visualize.

Measure backbone torsion angles

Again, the awk-script xyz-geometry.awk is employed by 02_get-angles.sh to measure torsion angles. How do the results differ from the ones obtained with the force field? Do you find patterns that resemble the idealized angles in Table 1?

Step 6: Fixed-geometry calculations for with different exchange-correlation functionals

Working directory: Step_4

We have used species defaults with reduced accuracy for the relaxations. Now, perform energy evaluations on these fixed geometries with standard 'light species defaults'. Use the following functionals:

```
PBE+vdW:
    xc pbe
    vdw_correction_hirshfeld

PBE:
    xc pbe

PBEO:
    xc pbe0

PBEO+vdW:
    xc pbe0
    vdw_correction_hirshfeld

M06-1:
    xc pbe
    total_energy_method m06l
```

Check the template control.in files for correct keywords and species defaults. The scripts O1_Create_Input.sh and O2_Submit_to_SGE.sh can be used to generate input files and to submit them to the Sun Grid Engine (SGE, now also Oracle Grid Engine)⁴ running on your workstations. SGE is scheduling and managing jobs. Please check Sub_Temp for an example SGE script. Some useful commands:

```
qsub Task - Submits a task to the SGE
```

qstat - Shows the status, r indicates a running job, qw a job waiting in the queue

qdel Job-ID – deletes a job from the queue

Alternatively to using SGE, you can run the jobs 'by hand' in the console or write a short script that takes care of it.

⁴http://en.wikipedia.org/wiki/Oracle_Grid_Engine

Step 7: Plotting and comparing the conformational energy hierarchies of different functionals

Last step in this task is to analyse the results from the different methods and answer the following questions:

- Does the conformational hierarchy agree between PBE, PBE+vdW, M06-L, and PBE0+vdW?
- Could you explain (or at least speculate on) the origin of deviations between different methods?
- What is the difference in computational cost between different methods? [Are you using eV or kcal/mol?]

Finally: Preparation for the independent research work

In the weekend project you will apply the steps 2 to 7 to a different model peptide according to Figure 1. You will find 'your peptide' in the one of the folders:

/pub/TUTORIAL3/Weekend_Tasks/workstation{01..49}.

Please select according to the number of the workstation you are working at. We already prepared the coordinates for you (step 1). In order to manage finishing the small project within this weekend, best run step 2 now and submit the PBE+vdW relaxations (20 to 50 most stable OPLS conformers) after dinner, to let them run over night.

Please check the Scripts folder for tools for the automated generation of FHI-aims input and submission to SGE. The scripts and tools provided work with the example shown on Friday, for your research project they have to be changed.

Please report your data:

- Fill in your results (PBE+vdW backbone torsion angles, energies, your classification) into your copy of the template OpenOffice spreadsheet:

 /pub/TUTORIAL3/Weekend_Tasks/tutorial3_report_template.ods
- Save the energy hierarchy data files and png output.
- Prepare figures in jpeg or png format of the 5 lowest-energy structures at the PBE+vdW level.
- \bullet You might want to prepare a short report on your findings and interpretations and send it via e-mail to to us 5 .
- Copying all data and analysis obtained back to the folder where you got the input file from: /pub/TUTORIAL3/Weekend_Tasks/workstation{01..49}.

 $^{^{5}}$ {tkatchen,baldauf,ropo}@fhi-berlin.mpg.de

3 Appendix: Tools for work

plot_hierarchy.py Creates energy hierarchy plots for or more energy hierarchy. This plotting tools made using python using pyplot package and to change names of the input files one has to modify the inputfiles list in the programs with text editor.

xyz-geometry.awk Computes distance, angle, or dihedral angle from xyz coordinates. Input are the line numbers from an xyz file. 2, 3, or 4 input values result in the calculation of distance (2), and angle (3), or a dihedral angle (4). Usage: xyz-geometry.awk -v c=a,b[,c[,d]] xyz-file

txyz2xyz.sh Converts the Tinker xyz format to standard xyz format.

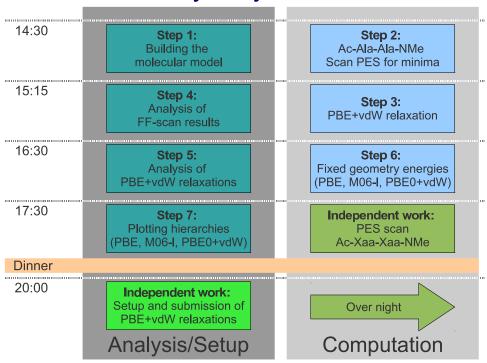
xyz2aims.sh Converts standard xyz format to the FHI-aims geometry.in format.

get_final_geometry.sh Extracts the final cartesian coordinates from FHI-aims output and gives standard xyz format.

submit_template.sh Template script for generation and submission of FHI-aims runs.

basic bash example 'for [item] in [list] do' Useful template to automatize task.

Timetable: Friday, July 15



Your weekend

