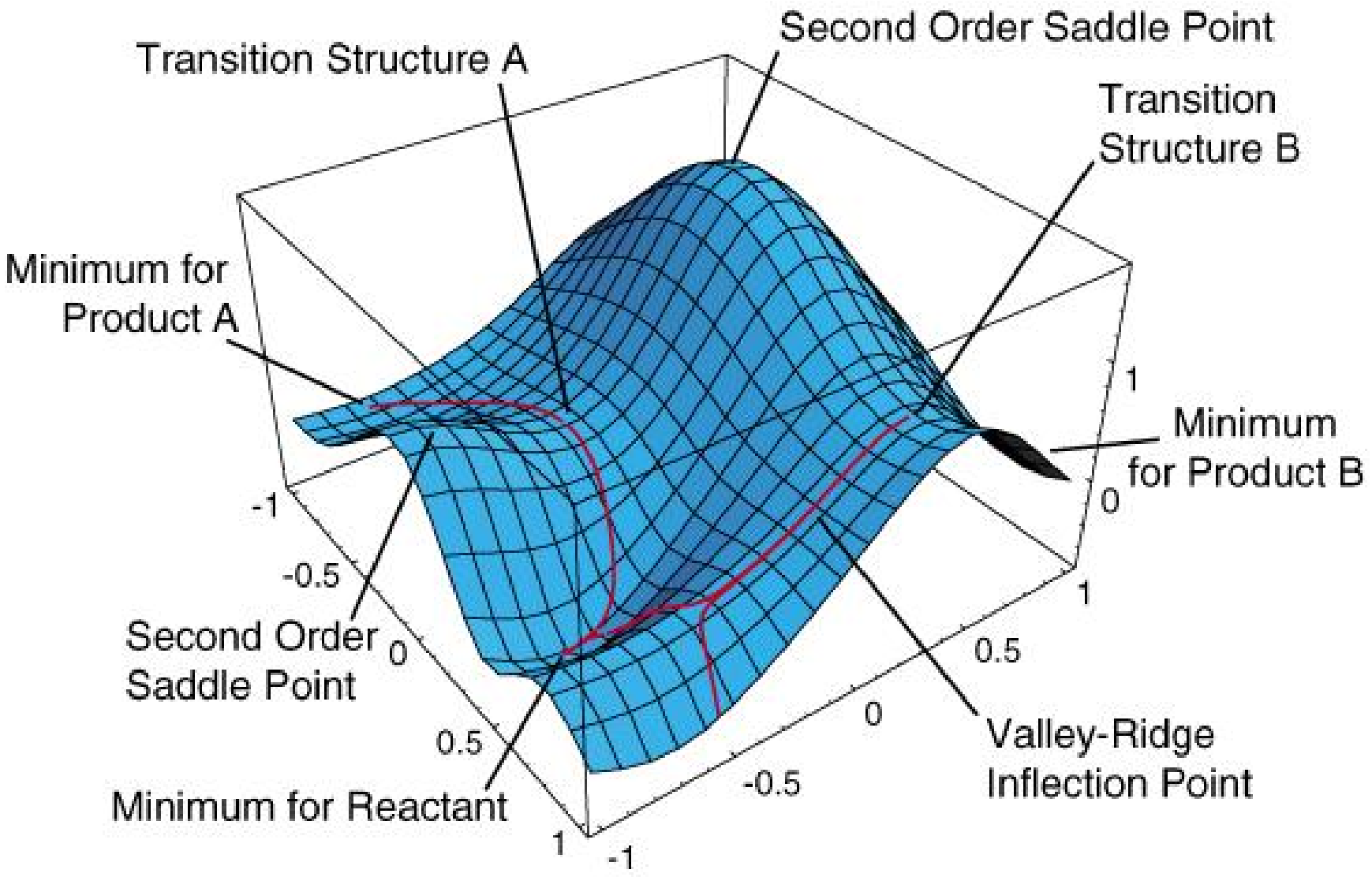
Conformational space and energetics of biomolecules: Physical concepts and performance of DFT-based methods

Alexandre Tkatchenko, Carsten Baldauf, Matti Ropo

Practical Session III / Weekend Project

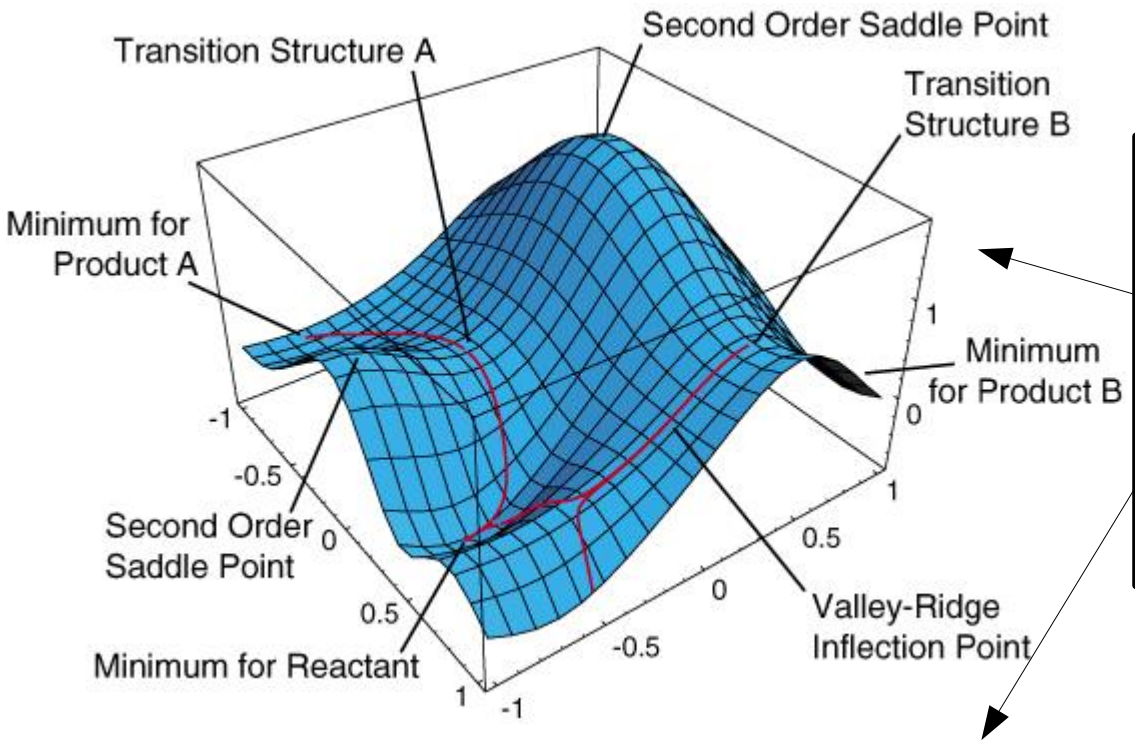
FHI "*DFT and Beyond*" Workshop, Jul. 15, 2011

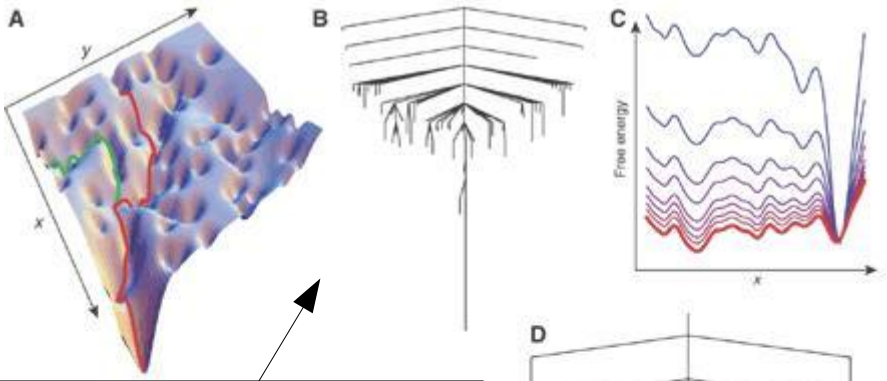
Potential-energy surfaces



Credits: www.chem.wayne.edu

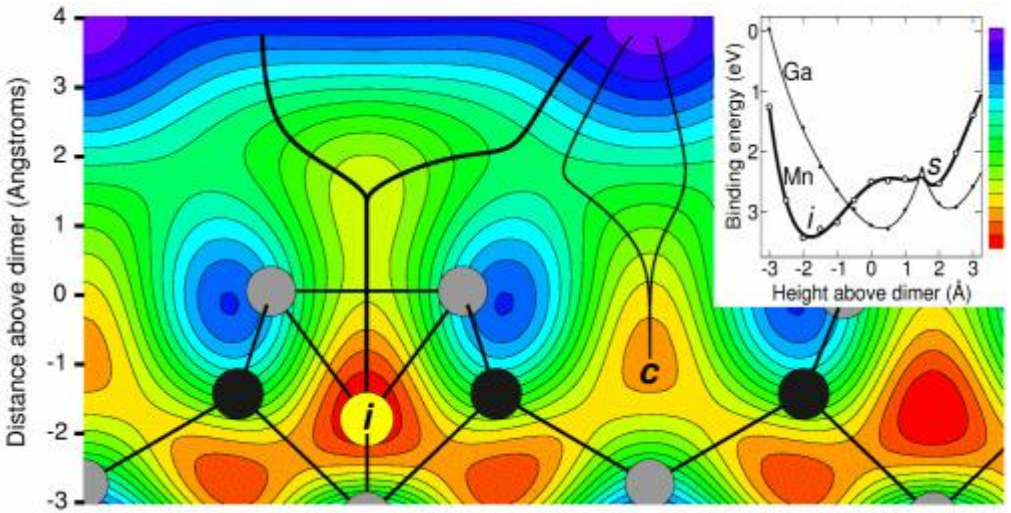
Potential-energy surfaces

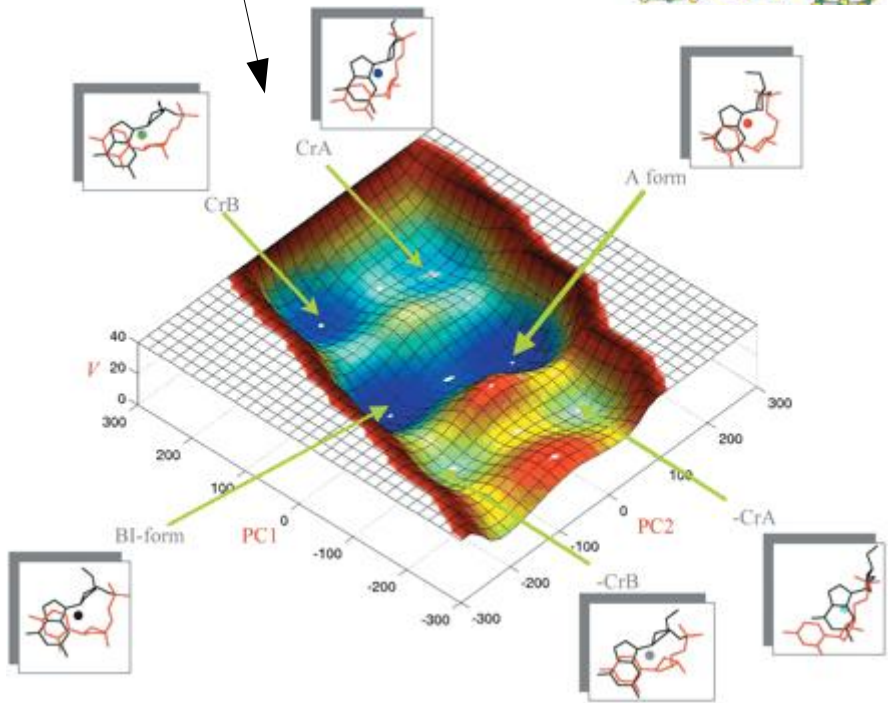




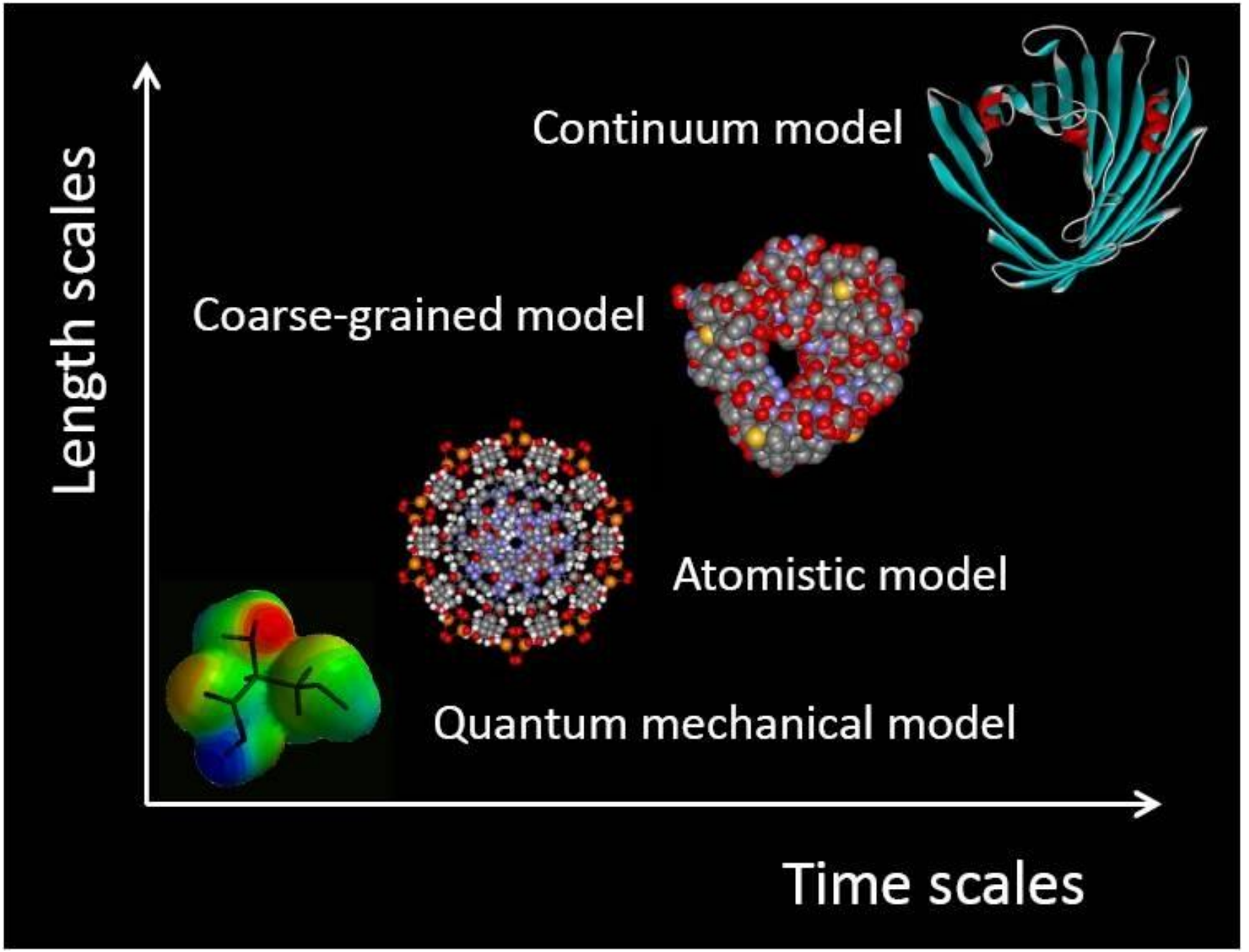
Credits:

- *D. Wales et al.*
- www.chem.wayne.edu
- *S. C. Erwin et al.*, ***PRL.***
- *Elsawi et al.*, ***Nucl. Acids Res.***

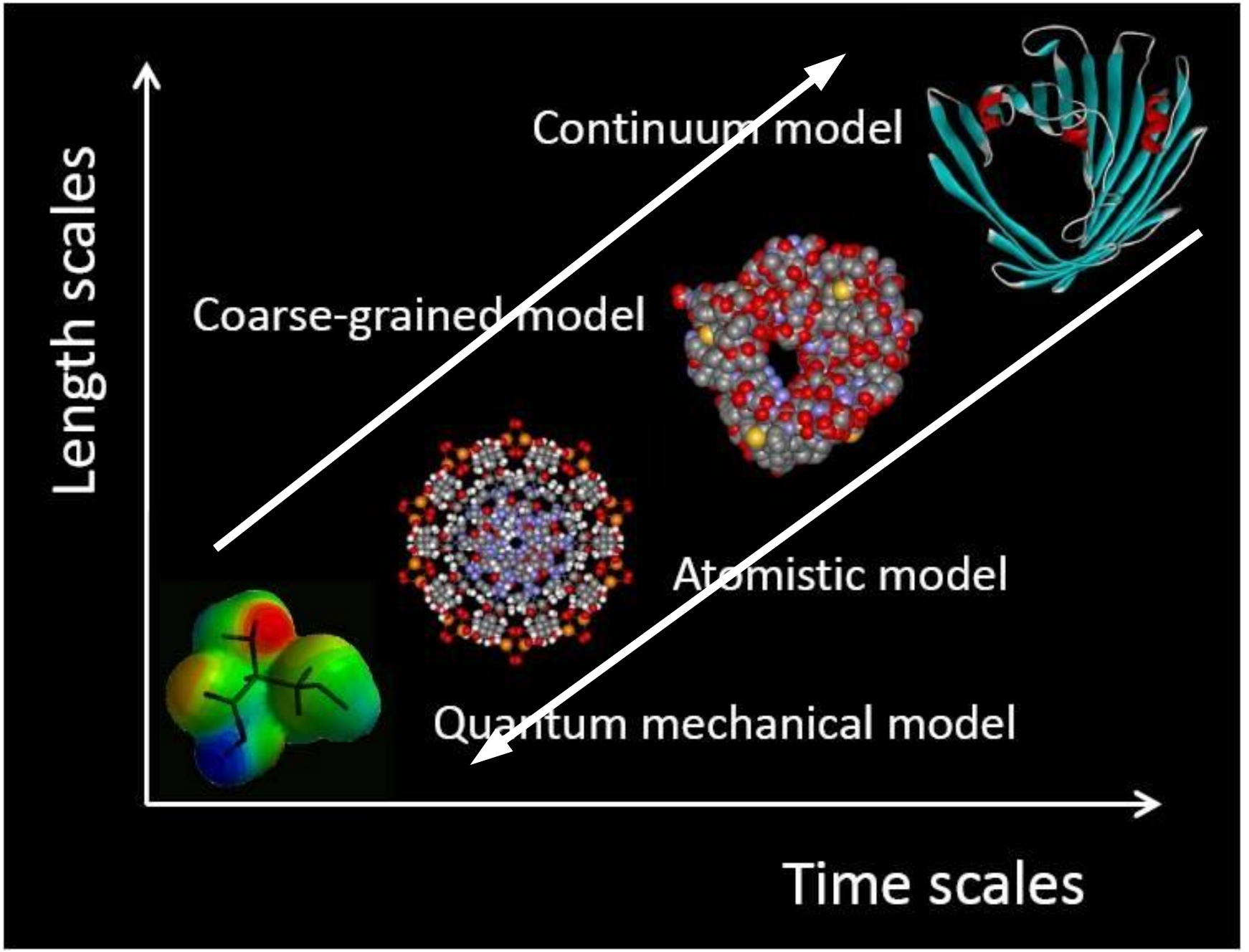




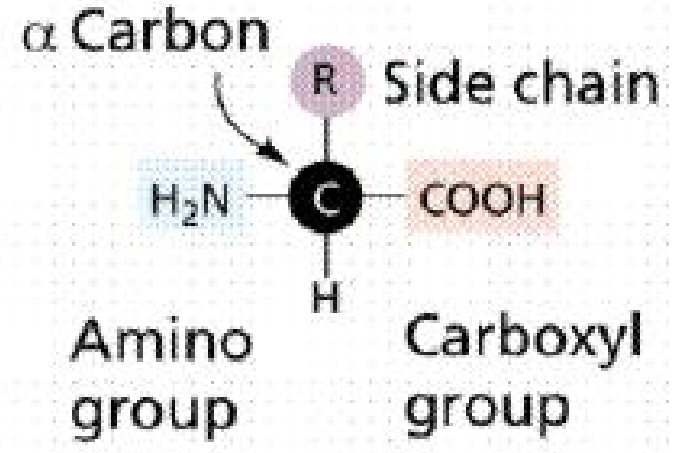
Multiscale modeling

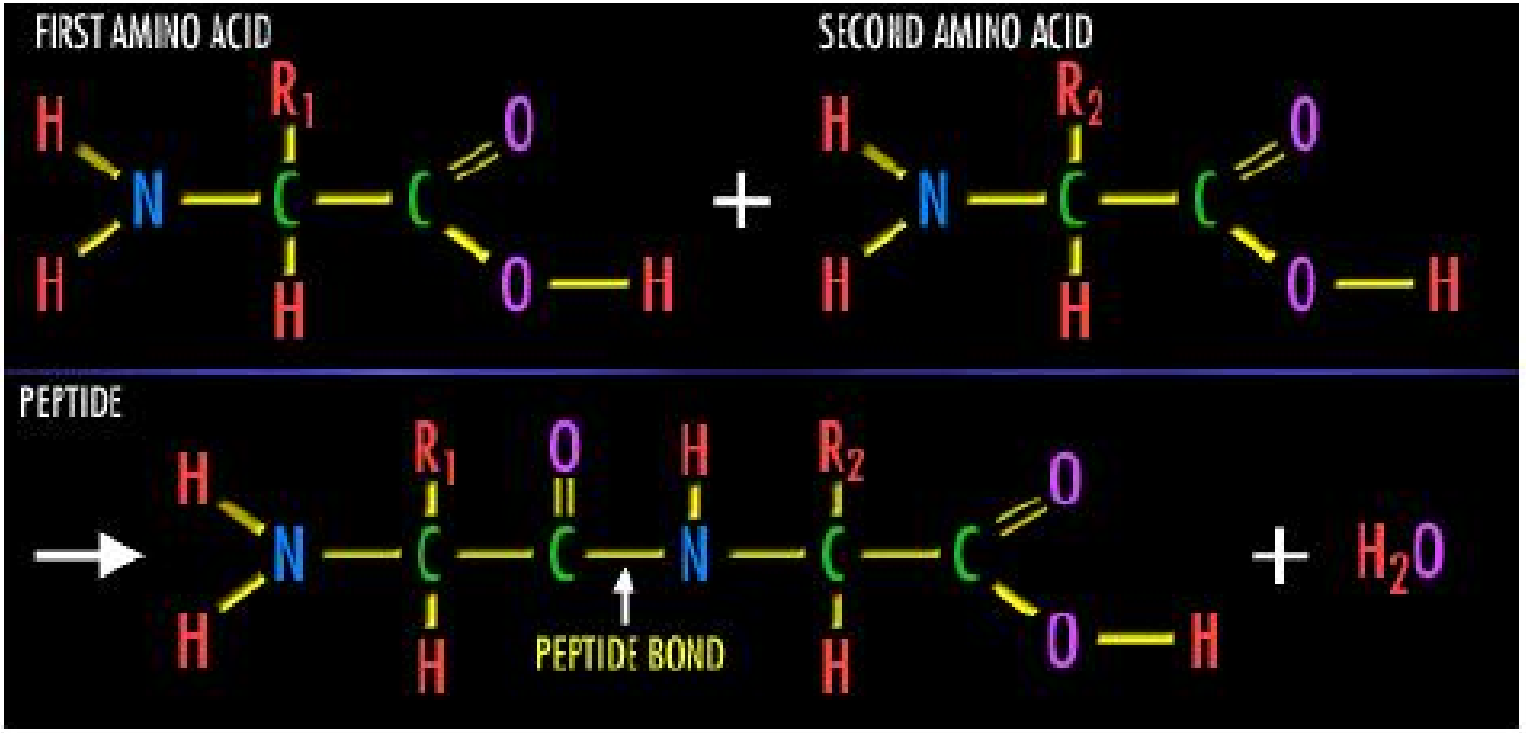


Multiscale modeling

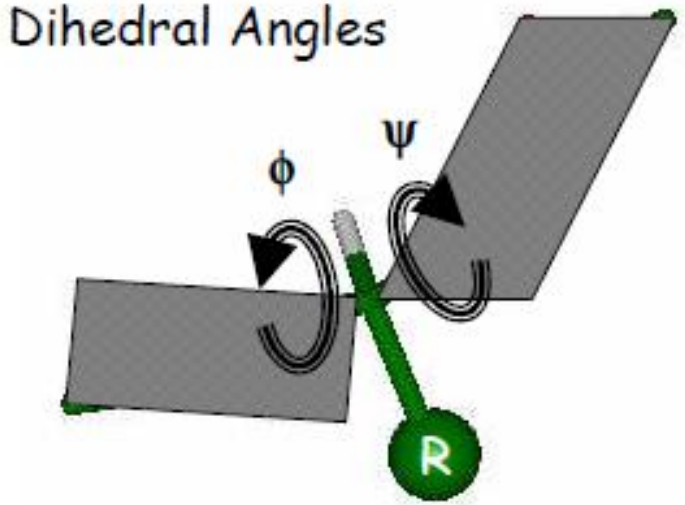


Peptides (and proteins) are made of aminoacids

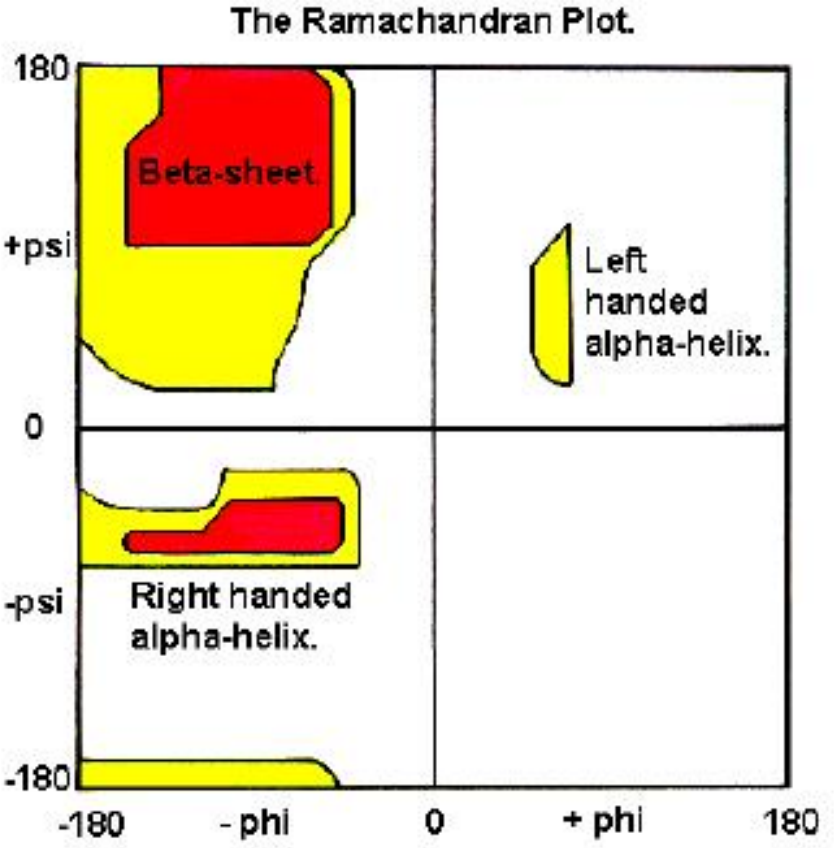


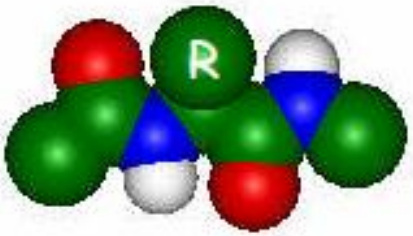


PES for a single aminoacid

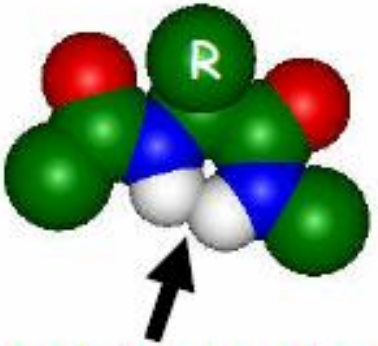


Allowed regions where repulsion among atoms is negligible (theoretical prediction)





Allowed conformation

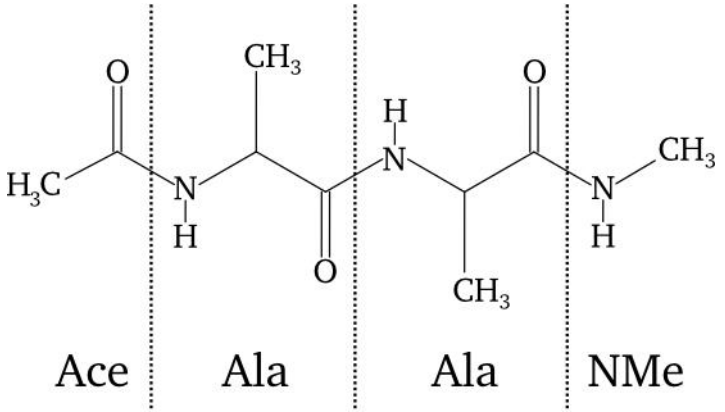


Repulsive overlap

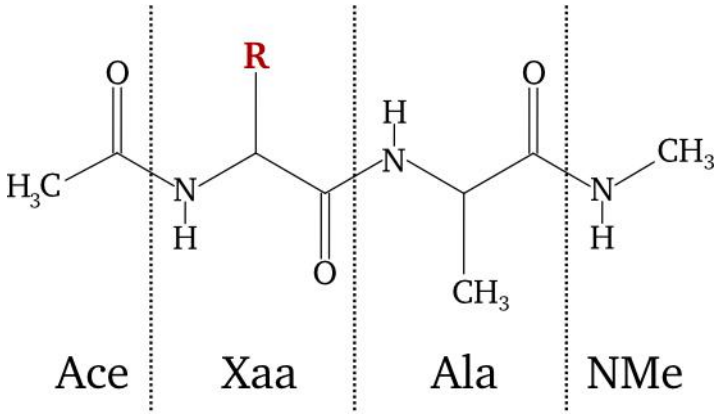
Not allowed conformation

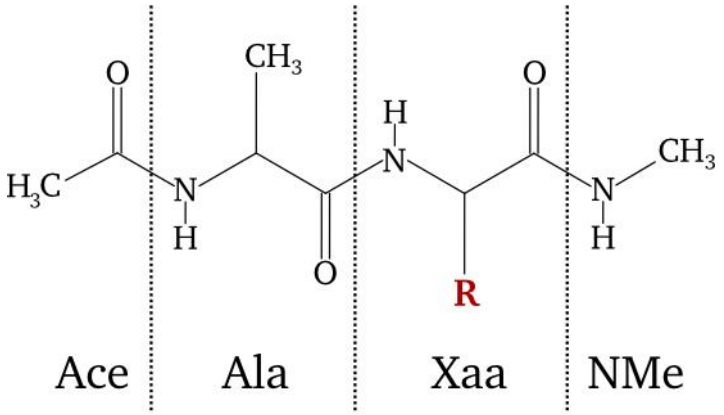
Topic of this tutorial

- Today, we will use a multiscale procedure to analyze the PES of ***Ace-Ala-Ala-NMe***

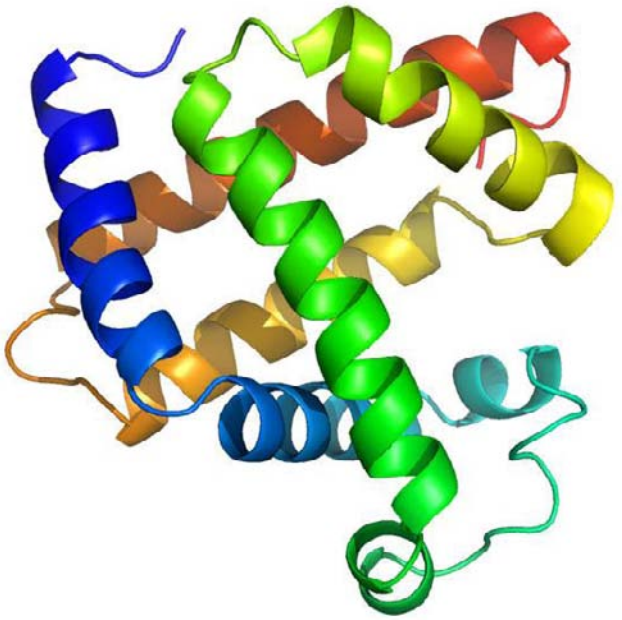


- Over the weekend you will investigate ***Ace-<u>Xaa</u>-Ala-NMe*** *or* ***Ace-Ala-<u>Xaa</u>-NMe***



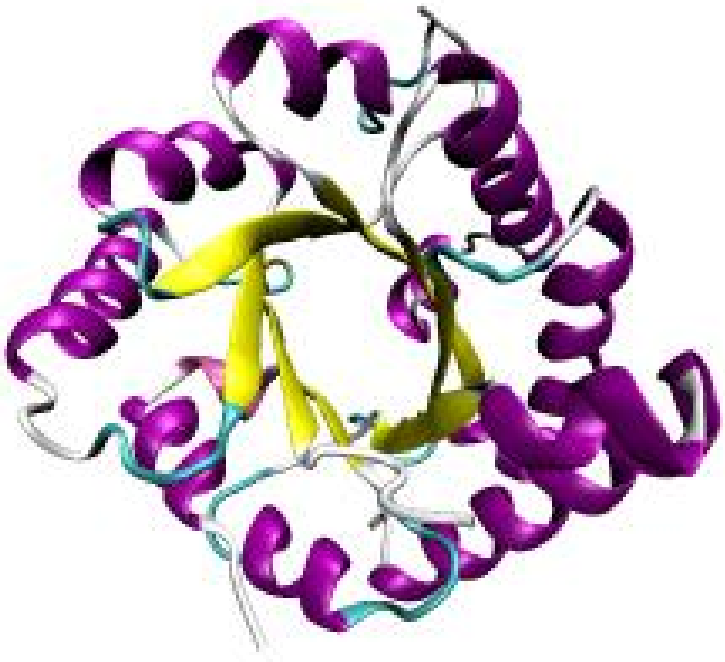


Proteins: The machinery of life



- Proteins account for ~ 16% of human body weight
- Necessary for the growth/repair of muscles, bones, hair, eyes, for creation of antibodies, for metabolism, digestion, *etc*.

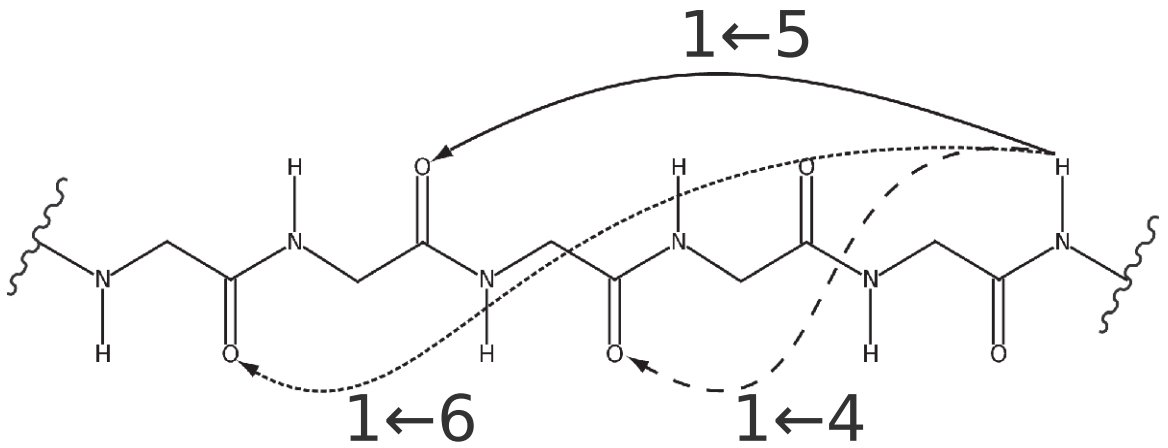
- Proteins are bio-polymers, typically made of 20 natural aminoacids
- Hierarchy of structure:
  - Primary: Aminoacid "code"
  - Secondary: Helices, beta-sheets, loops
  - Tertiary: Functional 3D structure

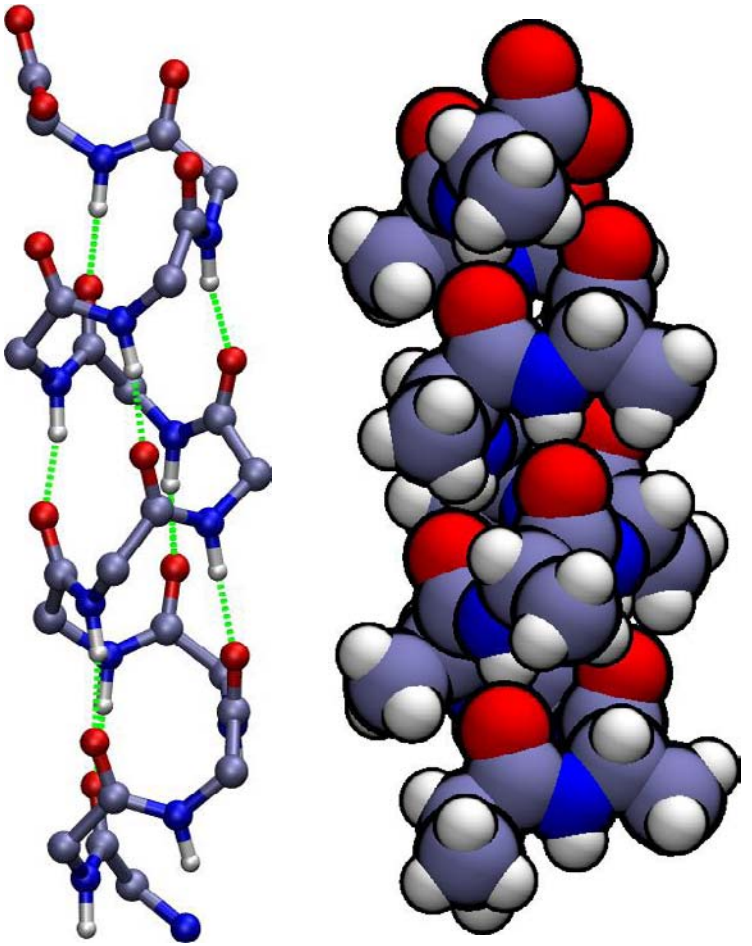


Source: Wikipedia

Secondary structure of proteins: Helices

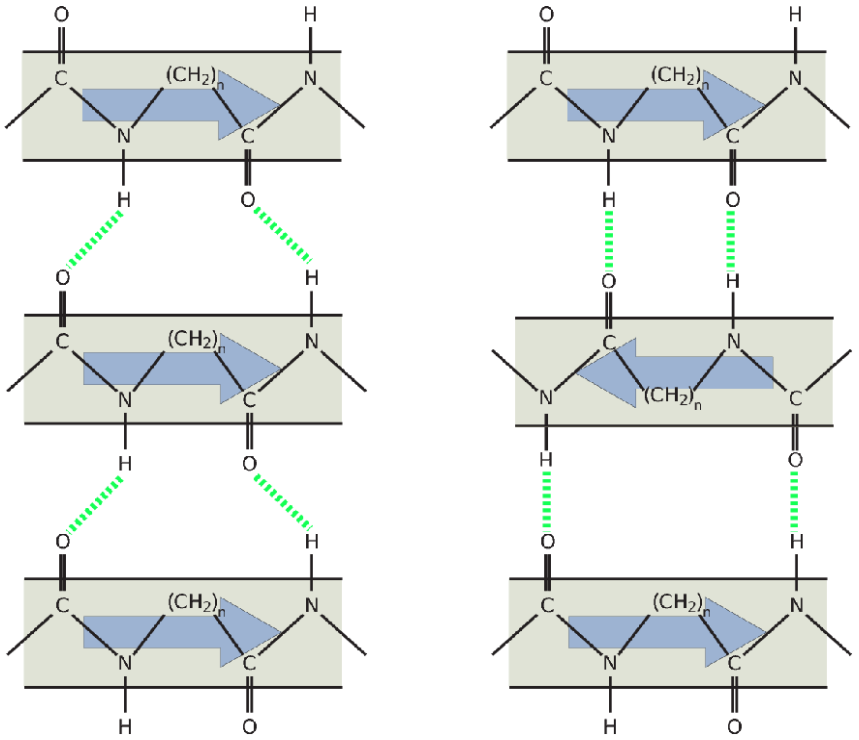
- H-Bond pattern along the helix
- '1-dimensional'
- Periodic repetition of φ/ψ tuples
- 3_{10}-Helix: 1->4
- α-Helix: 1->5
- π-Helix: 1->6

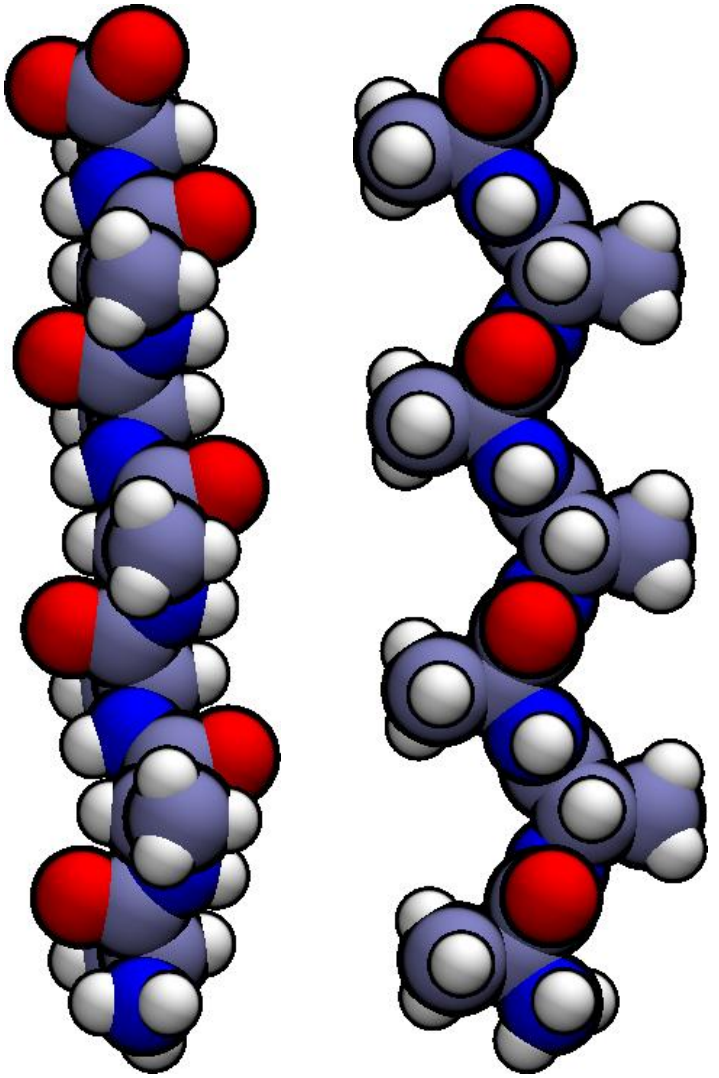




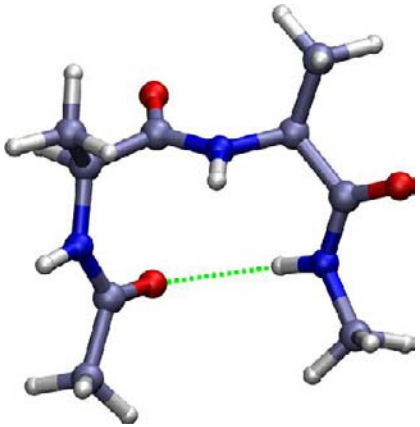
Secondary structure of proteins: Sheets

- Inter-strand H-bonds
- Parallel or anti-parallel
- '1-dimensional'
- Periodic repetition of φ/ψ tuples

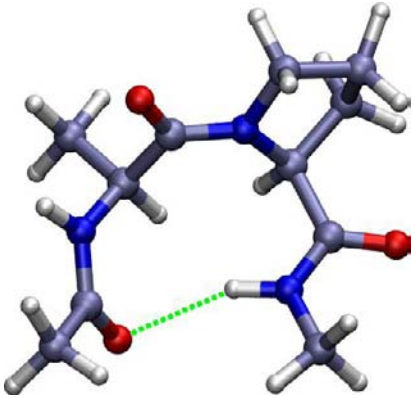




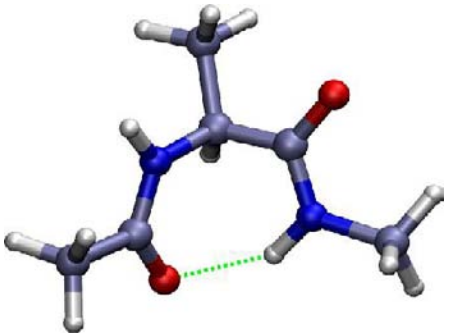
Secondary structure of proteins: Turns and loops



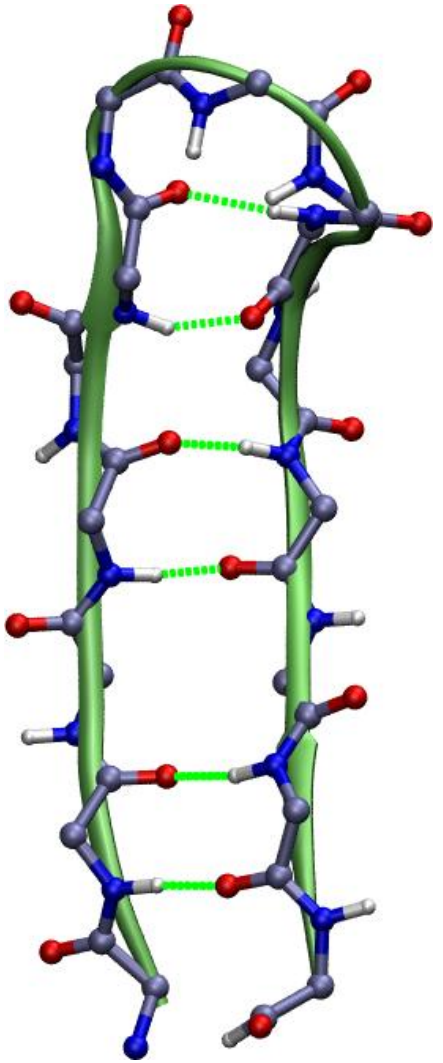
βI-turn:
1←4



βVIa-turn:
1←4



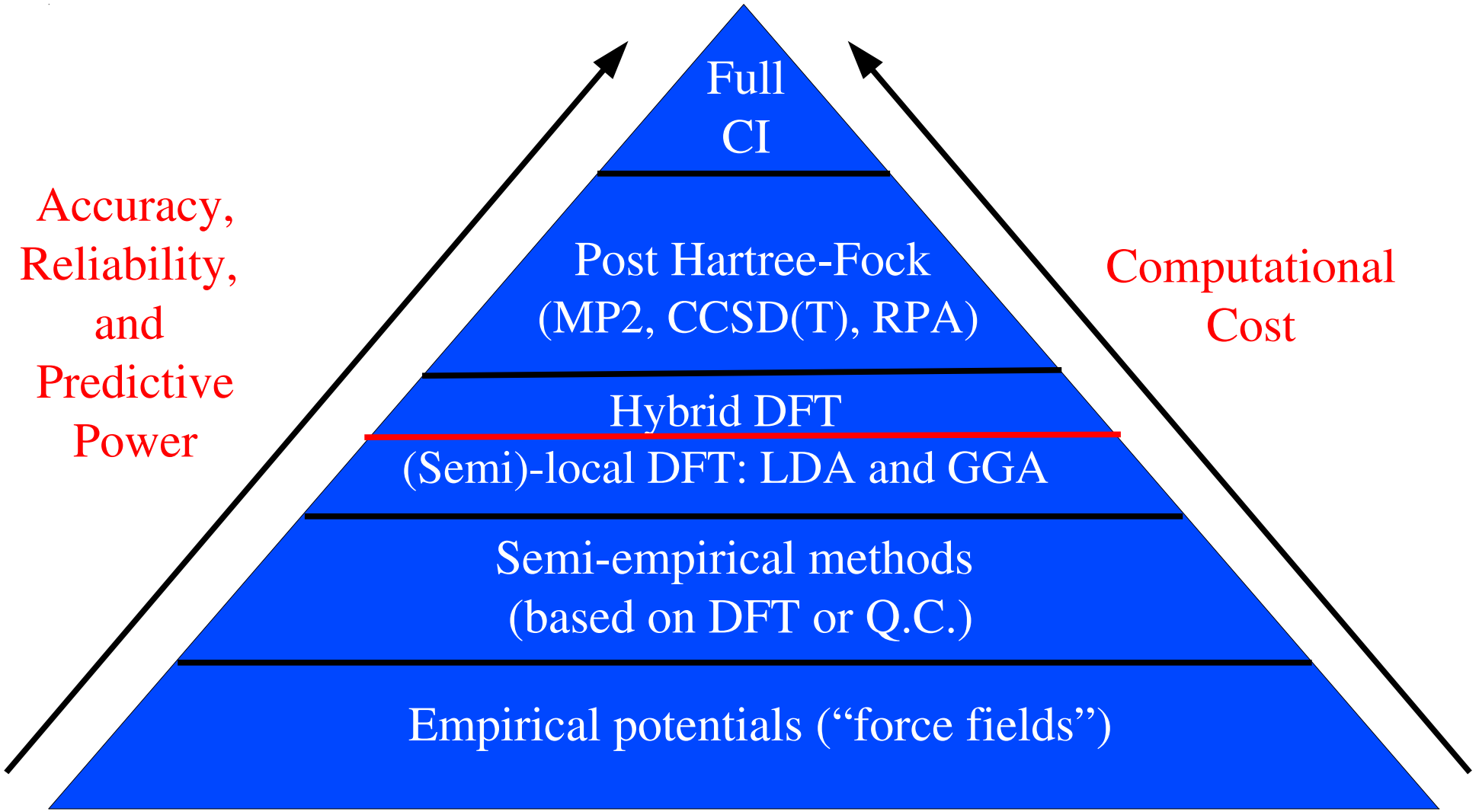
γ-turn:
1←3



- α-, π- and ω-turns
- Reversal of structure direction
- Basis for 'globular' structures

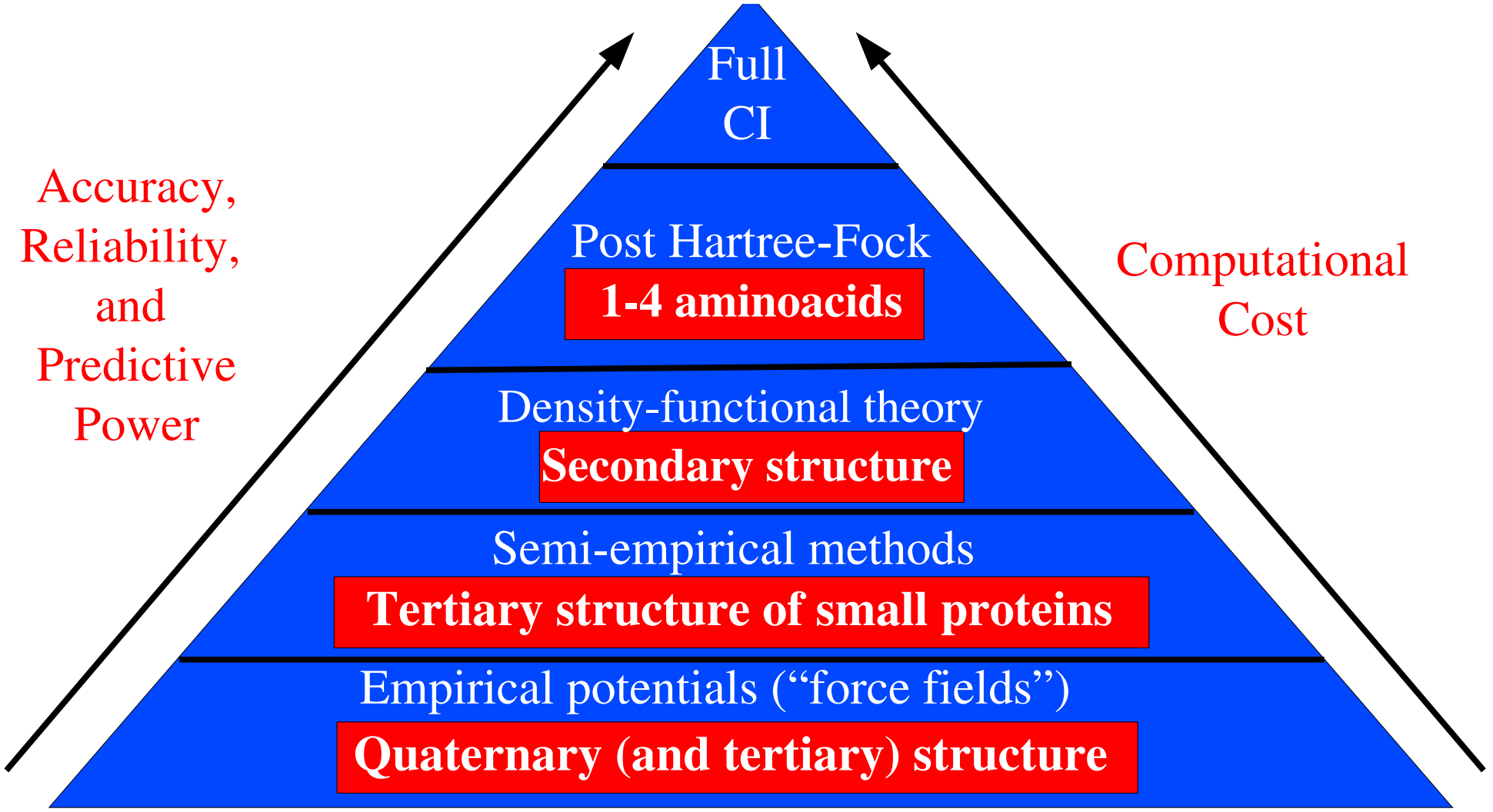
Modeling (predicting) the structure of proteins

$$H\Psi = E\Psi$$



Modeling (predicting) the structure of proteins

$$H\Psi = E\Psi$$

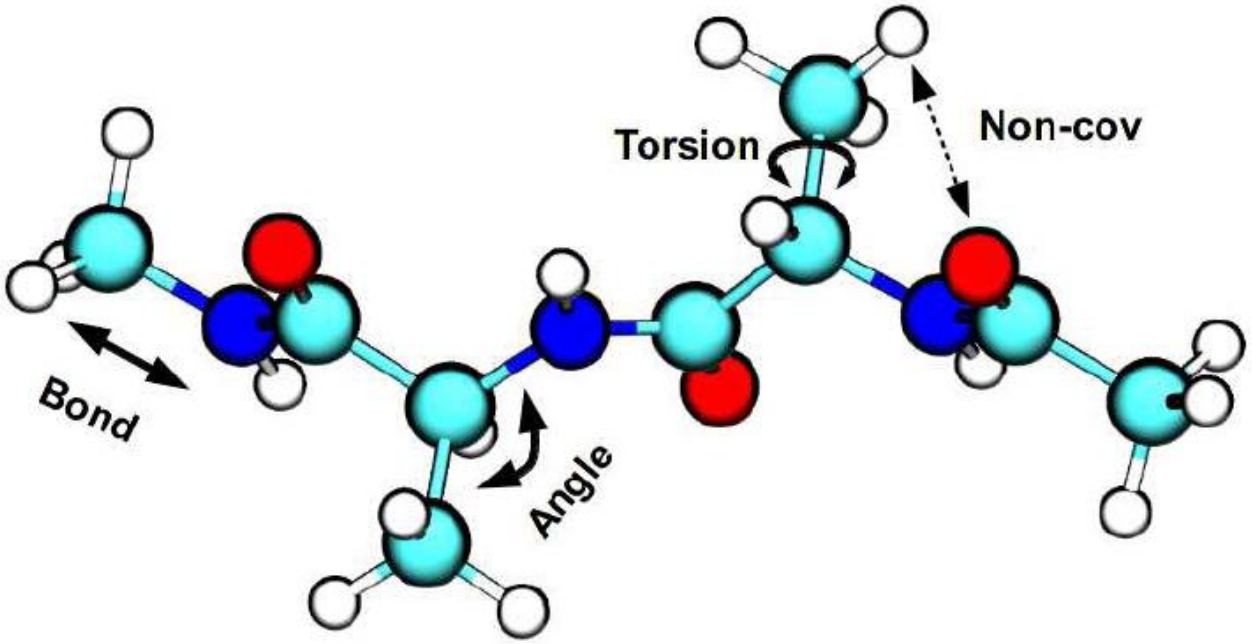


What can we do here?

Two constraints:

Computational resources

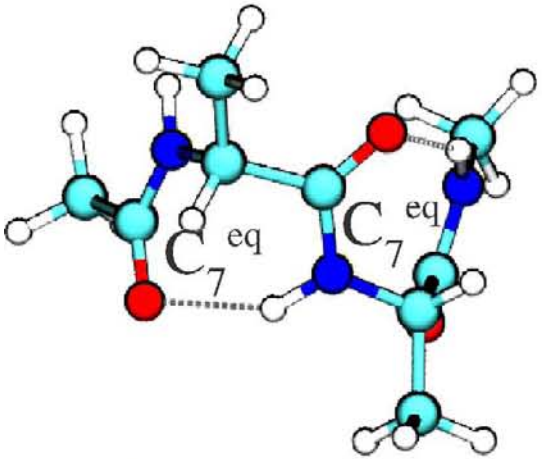
Human time



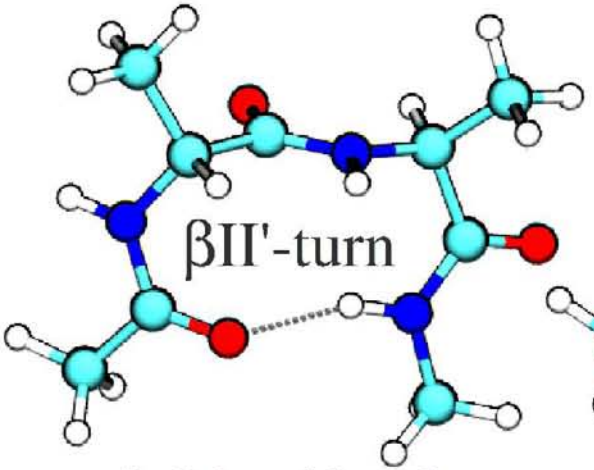
Ace-Ala-Ala-NMe **(Alanine tripeptide)**

Quiz time

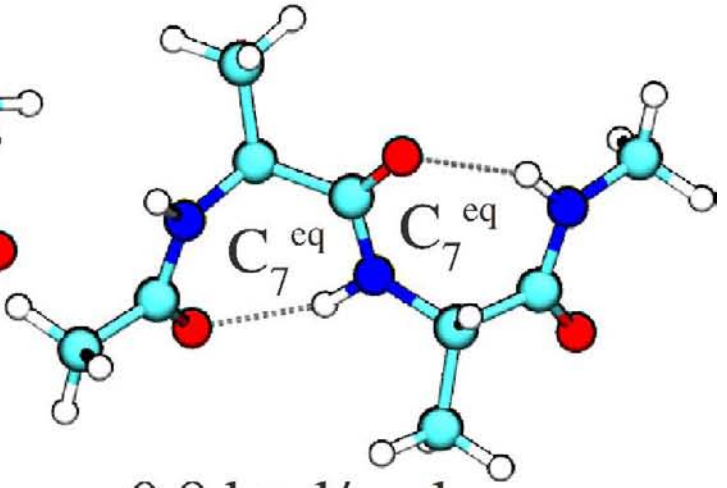
- Q: How many conformations (minima) do you expect for *Ace-Ala-Ala-NMe*?



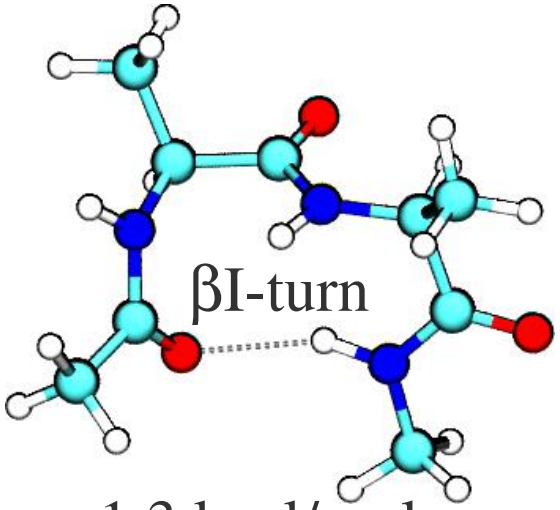
0 kcal/mol



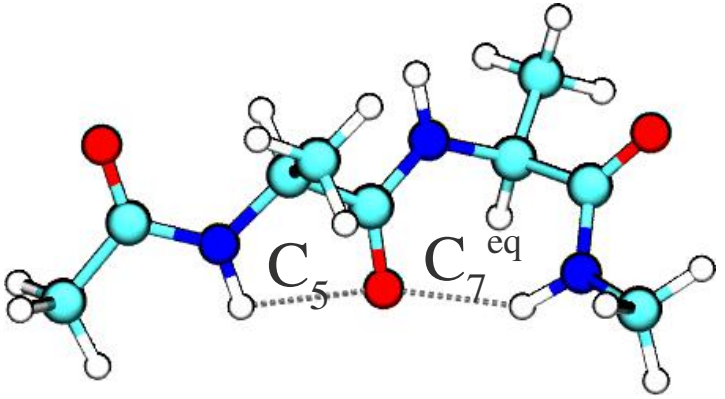
0.5 kcal/mol



0.8 kcal/mol



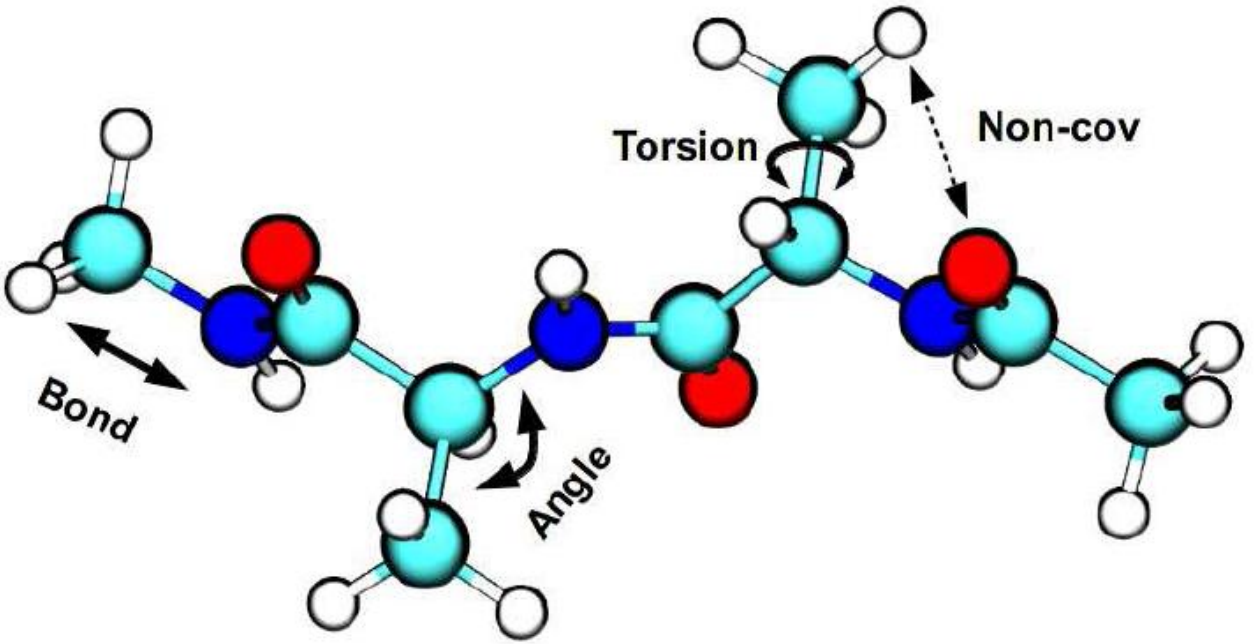
1.3 kcal/mol



1.8 kcal/mol

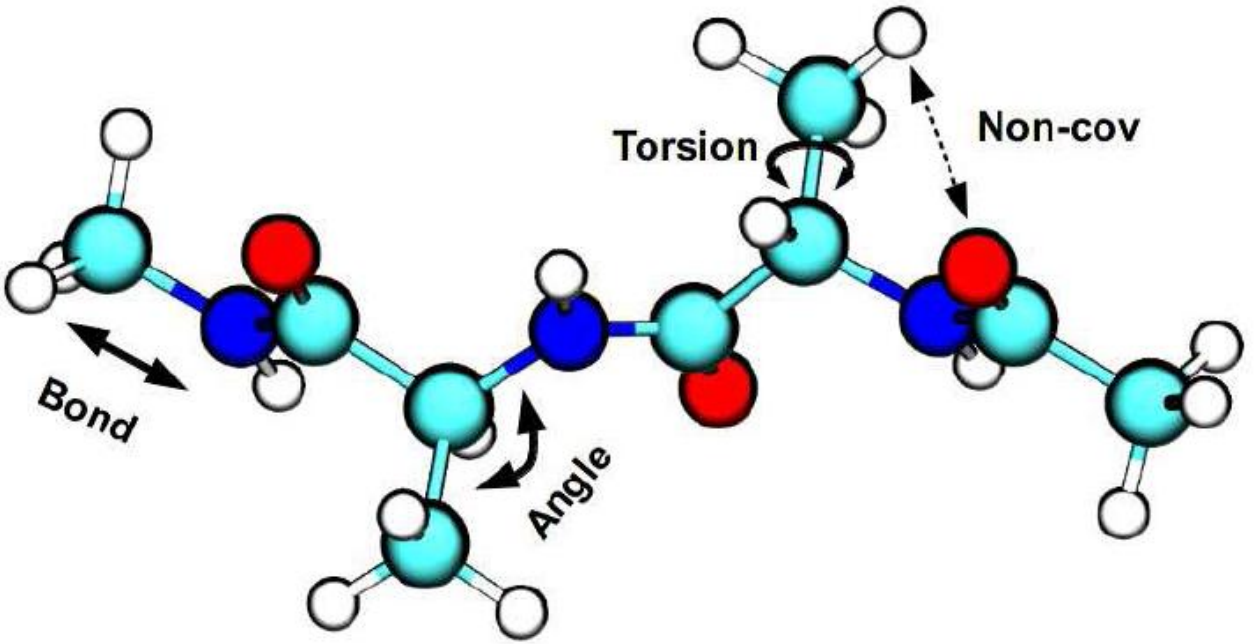
Quiz time

- Q: How many conformations (minima) do you expect for *Ace-Ala-Ala-NMe*? **A: More than 1000.**



Quiz time

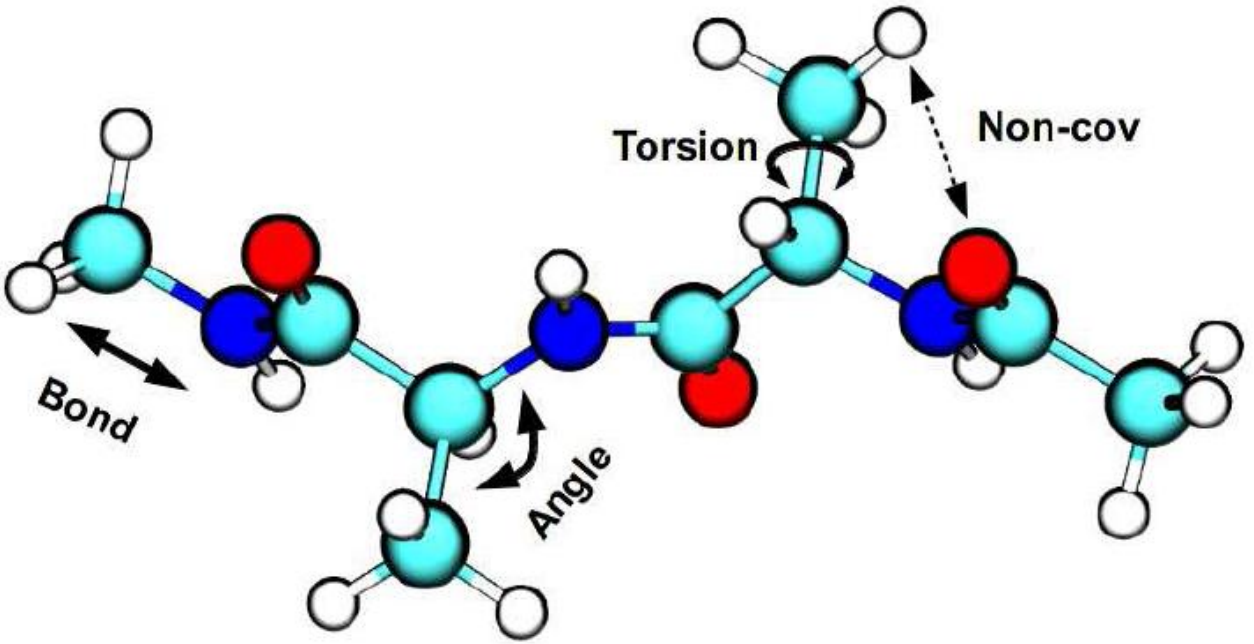
- Q: How many conformations (minima) do you expect for *Ace-Ala-Ala-NMe*? **A: More than 1000.**
- Q: Is it feasible to do full PES scan using DFT?



Quiz time

- Q: How many conformations (minima) do you expect for *Ace-Ala-Ala-NMe*? **A: More than 1000.**
- Q: Is it feasible to do full PES scan using DFT?

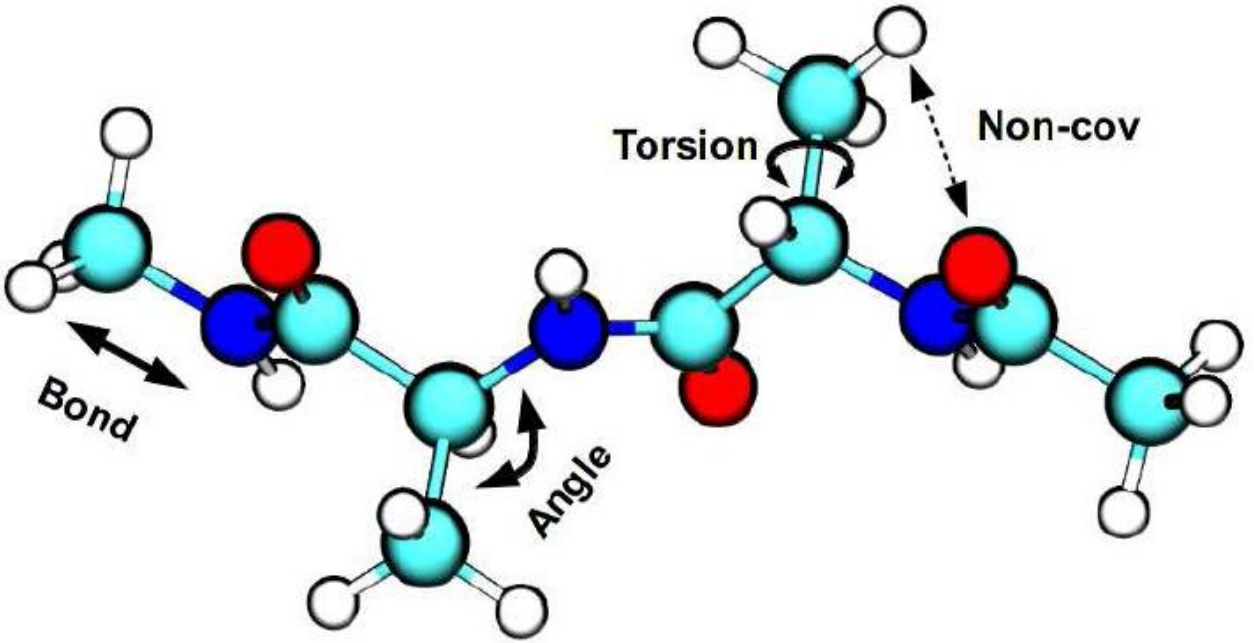
  A: Not during this weekend.



Quiz time

- Q: How many conformations (minima) do you expect for *Ace-Ala-Ala-NMe*? **A: More than 1000.**
- Q: Is it feasible to do full PES scan using DFT?

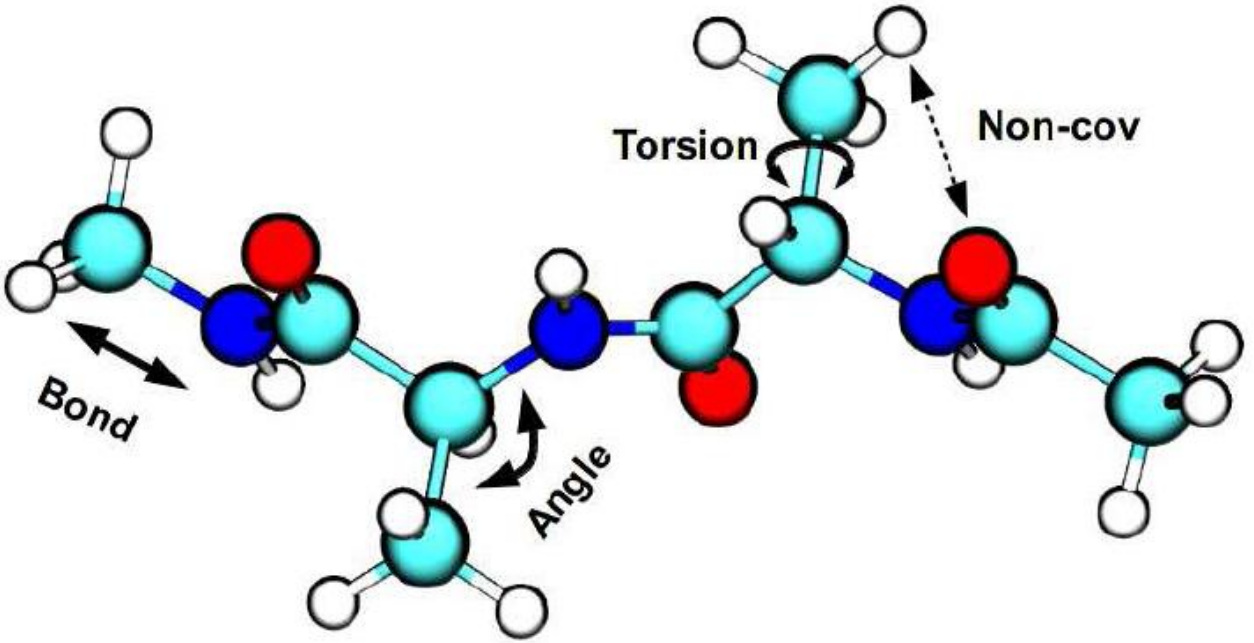
  A: Not during this weekend.
- Q: Do you expect van der Waals interactions to be important for *Ace-Ala-Ala-Nme*?



Quiz time

- Q: How many conformations (minima) do you expect for *Ace-Ala-Ala-NMe*? **A: More than 1000.**
- Q: Is it feasible to do full PES scan using DFT?

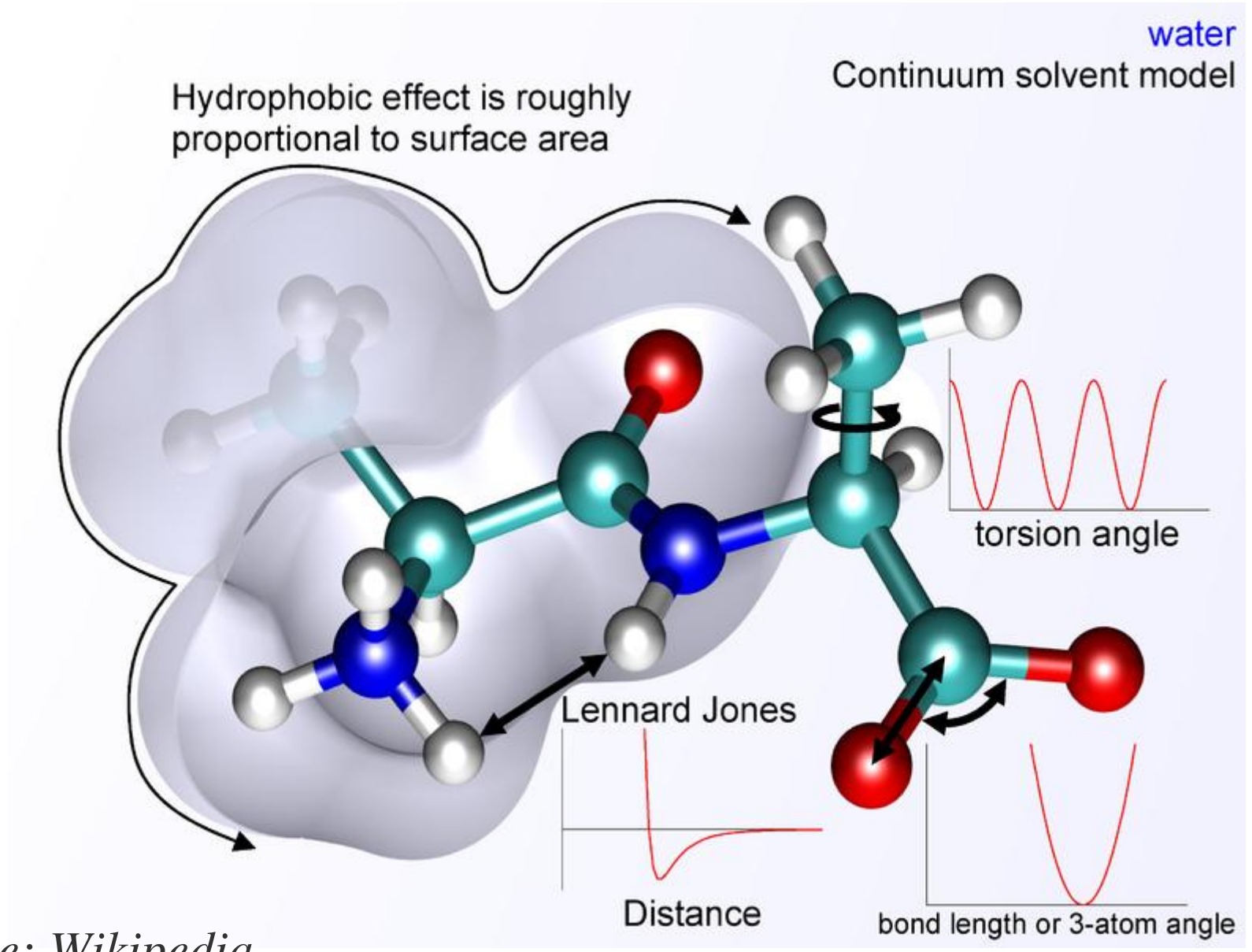
  A: Not during this weekend.
- Q: Do you expect van der Waals interactions to be important for *Ace-Ala-Ala-Nme*? **A: Let's try and see!**



The plot ...

- Generate initial conformations: Scan the PES of *Ace-Ala-Ala-NMe* using classical "force fields"
- Obtain more accurate geometries: Optimize most stable conformations using PBE+vdW
- Reliability and variability of results with approximate DFT: Compare results between different DFT functionals

Classical force fields



Source: Wikipedia

DFT functionals

- PBE
  - Widely used GGA functional
- PBE+vdW
  - PBE + Tkatchenko-Scheffler vdW correction
- PBE0+vdW
  - PBE with 25% HF exchange + vdW correction
- M06-L
  - Highly empirical functional from Truhlar's group, treats middle-range vdW

Questions we (YOU) want to address

- How complex is the PES of tripeptides?
- Do force fields yield reliable starting geometries?
- Are vdW interactions important for the energetics of polypeptides?
- How reliable (and variable) are the conformational energies with different DFT functionals?

Hands on !!!

Conformational space and energetics of (bio)molecules and clusters:

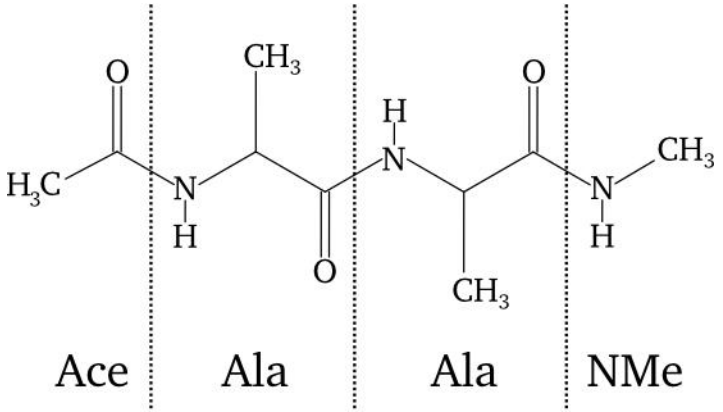
Physical concepts and performance of DFT-based methods

A. Tkatchenko, C. Baldauf, M. Ropo

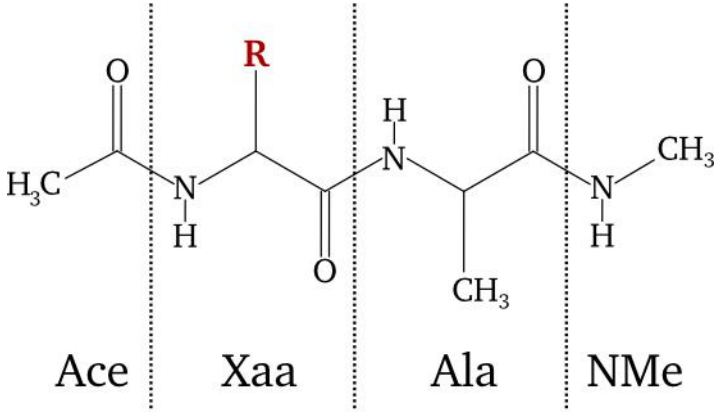
Weekend research project, July 15 to 17, 2011

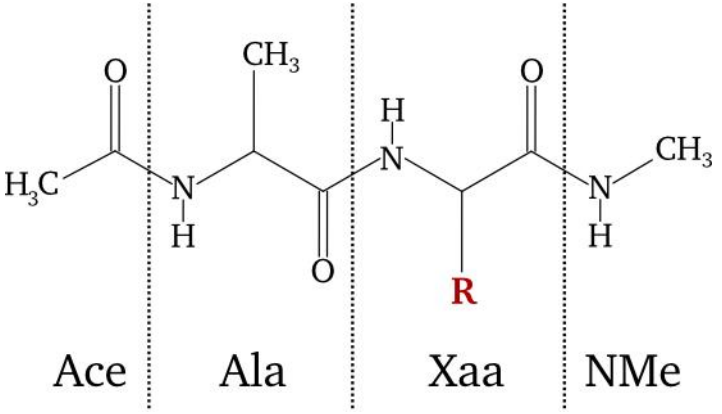
Topic of this tutorial

- We show you how to treat **Ace-Ala-Ala-NMe** today

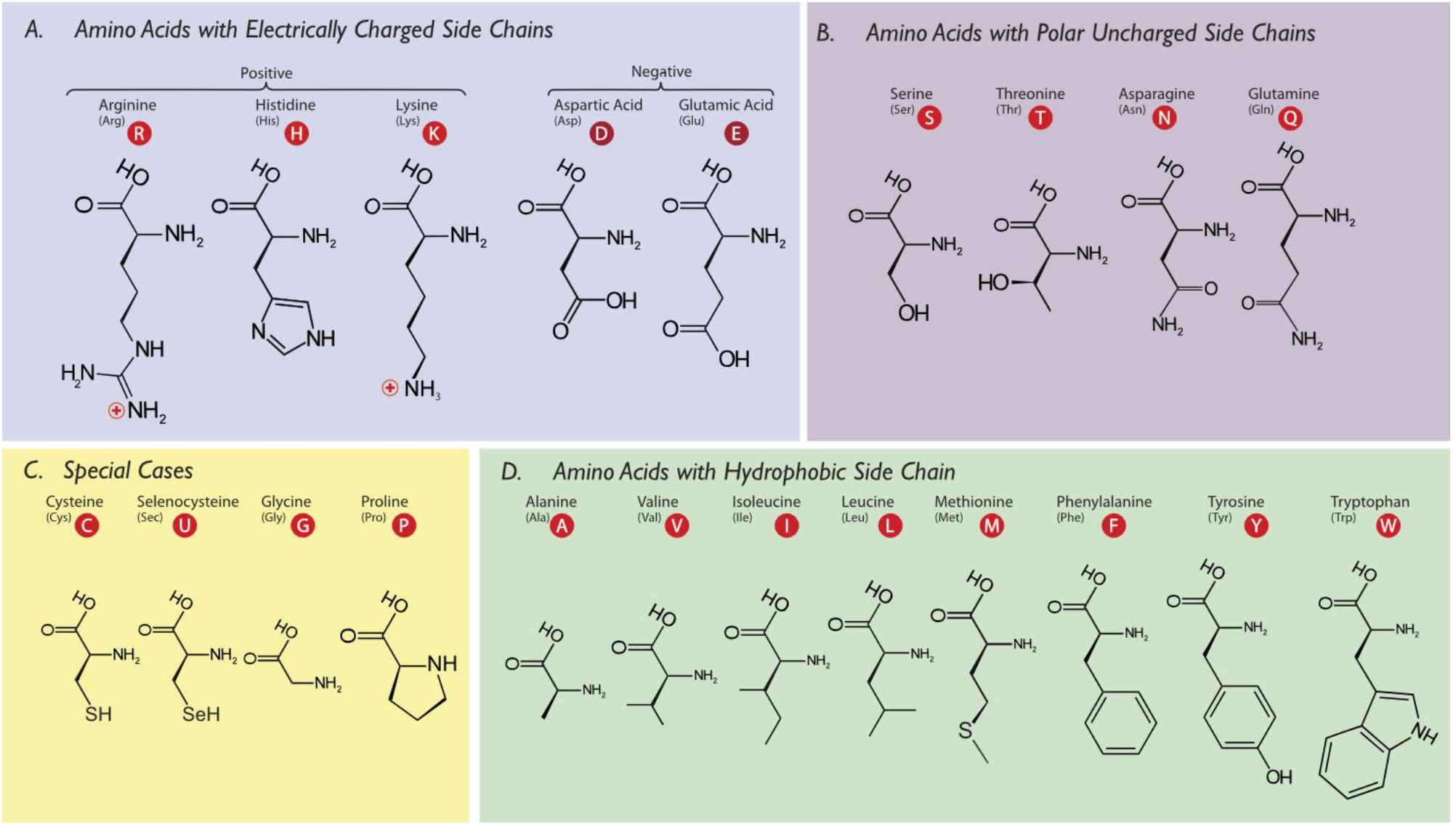


- Over the weekend you investigate **Ace-<u>Xaa</u>-Ala-NMe** or **Ace-Ala-<u>Xaa</u>-NMe**



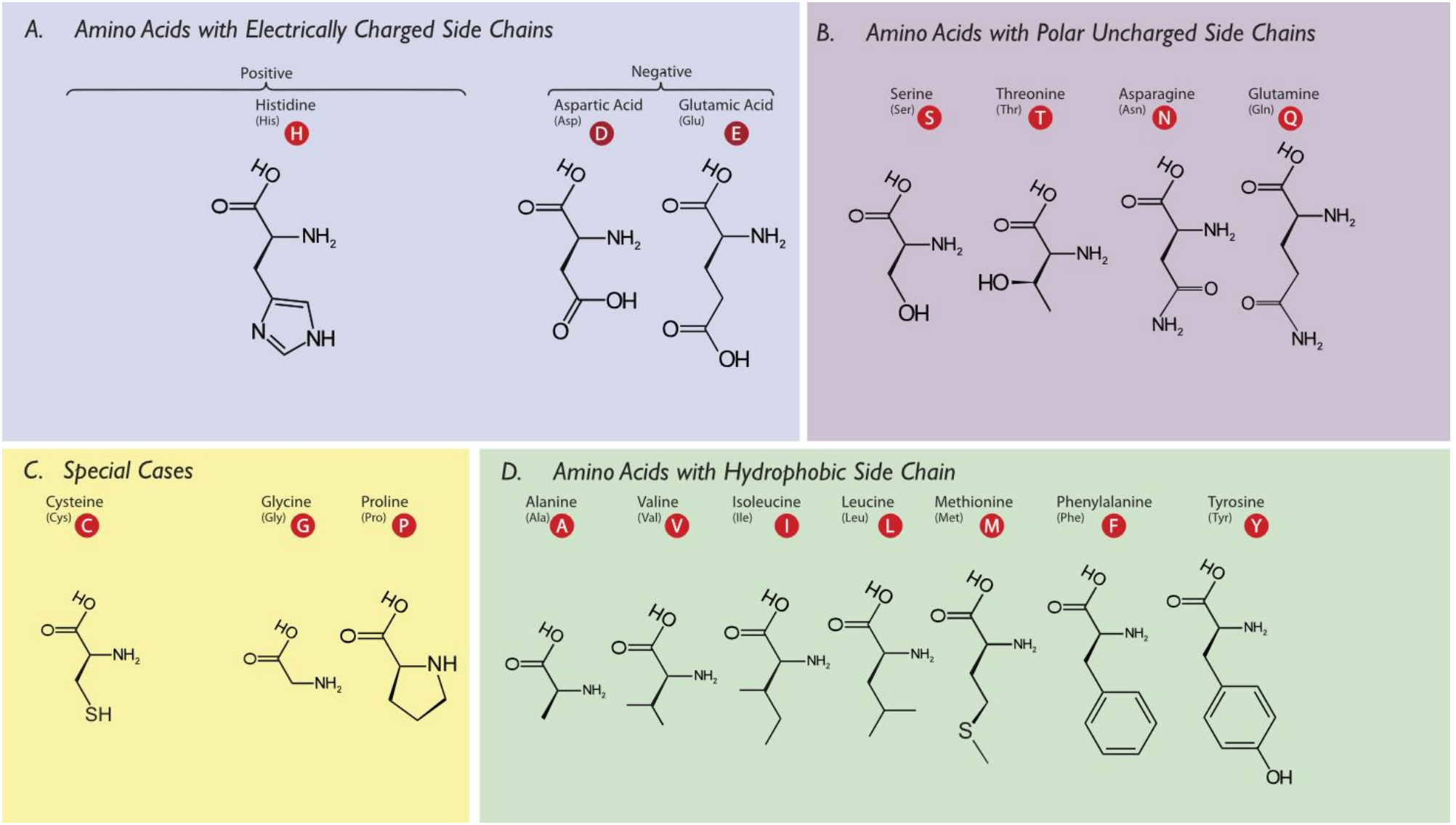


21 proteinogenic amino acids



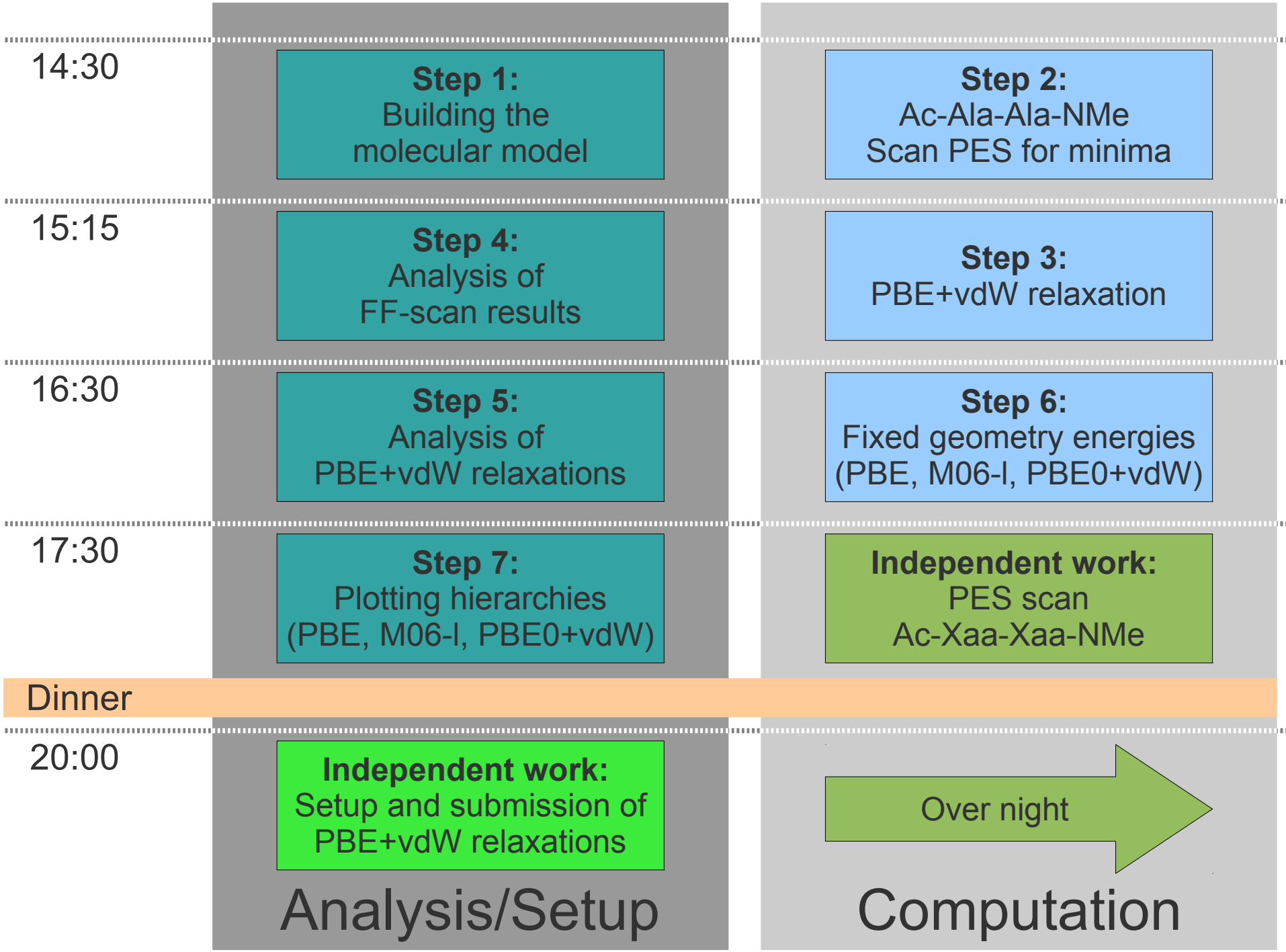
Modified from http://en.wikipedia.org/wiki/Amino_acid, Dan Cojocari

18 proteinogenic amino acids



Modified from http://en.wikipedia.org/wiki/Amino_acid, Dan Cojocari

Timetable: Friday, July 15



Step 0: Setup your environment

```
/pub/TUTORIAL3/instructions
      - The source of the handout (outline-tutorial3.tex)
      - This presentation (Tutorial3_AA-example.odp)

/pub/TUTORIAL3/reference
      - Copy to your home folder
         cp /pub/TUTORIAL3/reference ~/tutorial3
      - Subdirectory 'Params' contains the
         force field parameters
      - Subdirectory 'Scripts' contains
         scripts/script    templates
      - Subdirectories 'Step_*' should be used
         to perform the tutorial

/pub/TUTORIAL3/results
      - pre-computed results

/pub/TUTORIAL3/Weekend_Tasks/workstation{01..49}
      - Files for your weekend research project
```

Step 1: Building the molecular model

- Create model of Ac-Ala-Ala-NMe

```
tinker.protein

AcAANMe                       ; 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]
n               ; do not cyclize the peptide
[hit RETURN]
```

- Visualize with vmd, Jmol, molden
  - `../Scripts/txyz2xyz.sh` to convert to xyz format
  - Follow us on the big screen
  - Measure backbone distances, bond and torsion angles

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

- Copy `AcAANMe.xyz` and `AcAANMe.key` from folder `../Step_1`
- Invoke the `tinker.scan` routine to find minima on the potential energy surface (PES)

```
mpirun -n 4 tinker.scan AcAANMe.xyz 0 10 20 0.0001 0 | \
tee AcAANMe.log
```

0 … automatic selection of torsions
10 … # search directions
20 … energy threshold
0.0001 … energy similarity criterion
0 … start new run

[energies in kcal/mol]

Step 2: Extracting the energies

- Get the energies from `AcAANMe.log` (Tinker uses kcal/mol)

```
grep Map AcAANMe.log | \
awk  '{printf "%03s %6.4f \n", $5,$6}' | \
sort -nk 2 > StructID_Eopls.dat
```

Step 3: Which to minimize with DFT

- Only one relaxation per group, check your workstation number

```
nl StructID_Eopls.dat ; gives a numbered list
```

- Select the conformer with the rank equal to your workstation number
- Create folder `Conf_XX`

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

- The `control.in` file - Calculation setup:

```
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 `control.in` file – The basis:
  - Reduced basis for rapid computation

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

1. Create a directory named `Conf_{01..49}` according to the **number of your workstation**: e.g.: `mkdir Conf_23`
2. Create `geometry.in`:

   Select the **conformation with rank=number of your workstation**

   Convert according file to FHI-aims format:

   ```
   tail -n 32 ../Step_2/AcAANMe.040.xyz | \
   awk '{print "atom",$2,$3,$4,$1}' > \Conf_23/geometry.in
   ```
3. Create/copy `control.in`
4. 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. Proceed to **Step 4** and analyze the output of the force field scan
7. Once the simulation is finished, copy to `/pub/TUTORIAL3/results/Step_3`

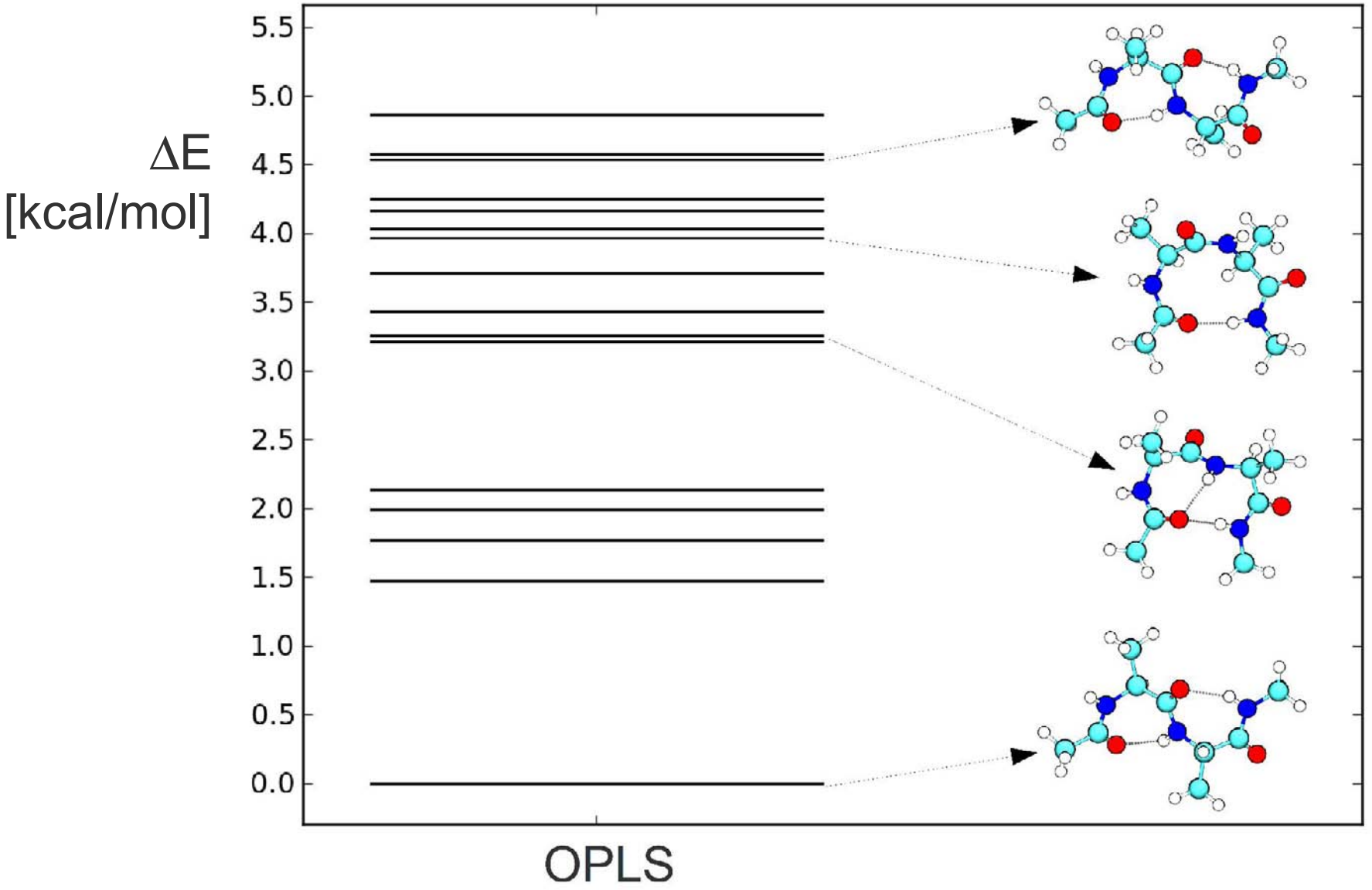
Step 4: Analyzing the force field scan results

- Plot the hierarchy

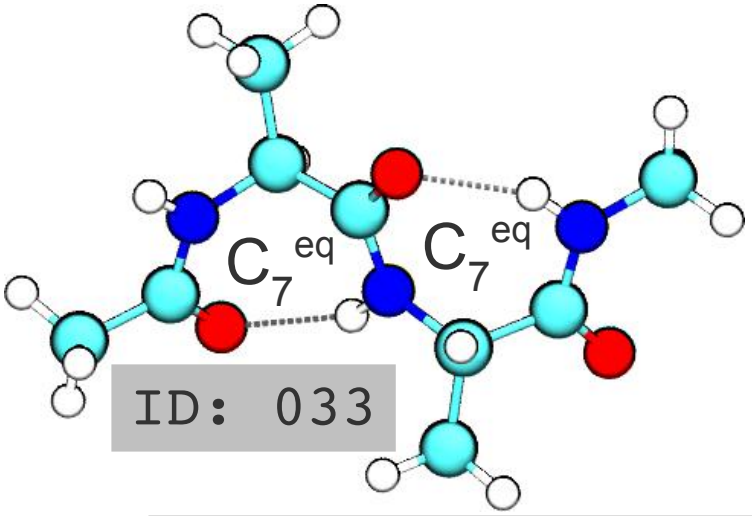
```
python plot_hierarchy.py
```
Remember to modify the script `plot_hierarchy.py`**!**
You find it in `../Scripts`.

- Measure backbone torsion angles
  - `02_get-angles.sh`
- Visualize (e.g. with VMD) the xyz-files in `Step_2`
- Characterize the structures by
  - H bond pattern
  - Backbone torsion angles
  - Compare to the examples in the script

Step 4: OPLS energy hierarchy of Ac-Ala-Ala-NMe

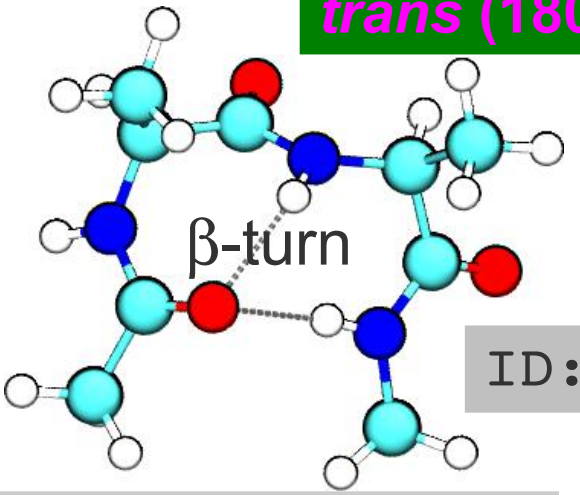


Step 4: Conformers of Ac-Ala-Ala-NMe



ID: 033

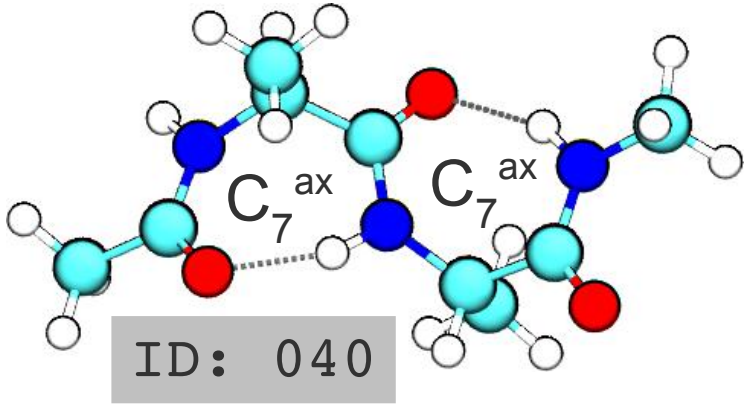
***trans* (180°)**



ID: 039

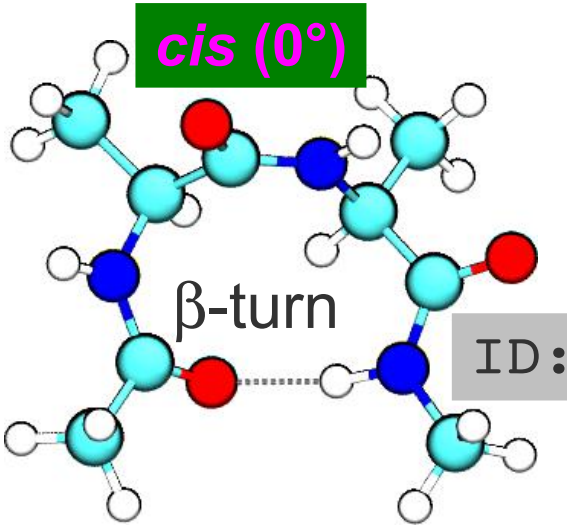
ID	ome1	phi1	psi1	ome2	phi2	psi2	ome3	DE*
033	-180.0	-79.1	67.8	-178.6	-78.4	65.4	-179.3	0.0
039	-177.4	60.8	-73.0	172.4	-95.7	-25.4	-179.9	3.3
021	175.6	-86.6	110.8	-9.5	-119.9	95.0	-178.9	4.0
040	179.9	67.9	-54.5	-179.7	67.0	-53.0	179.8	4.5

* in kcal/mol



ID: 040

***cis* (0°)**



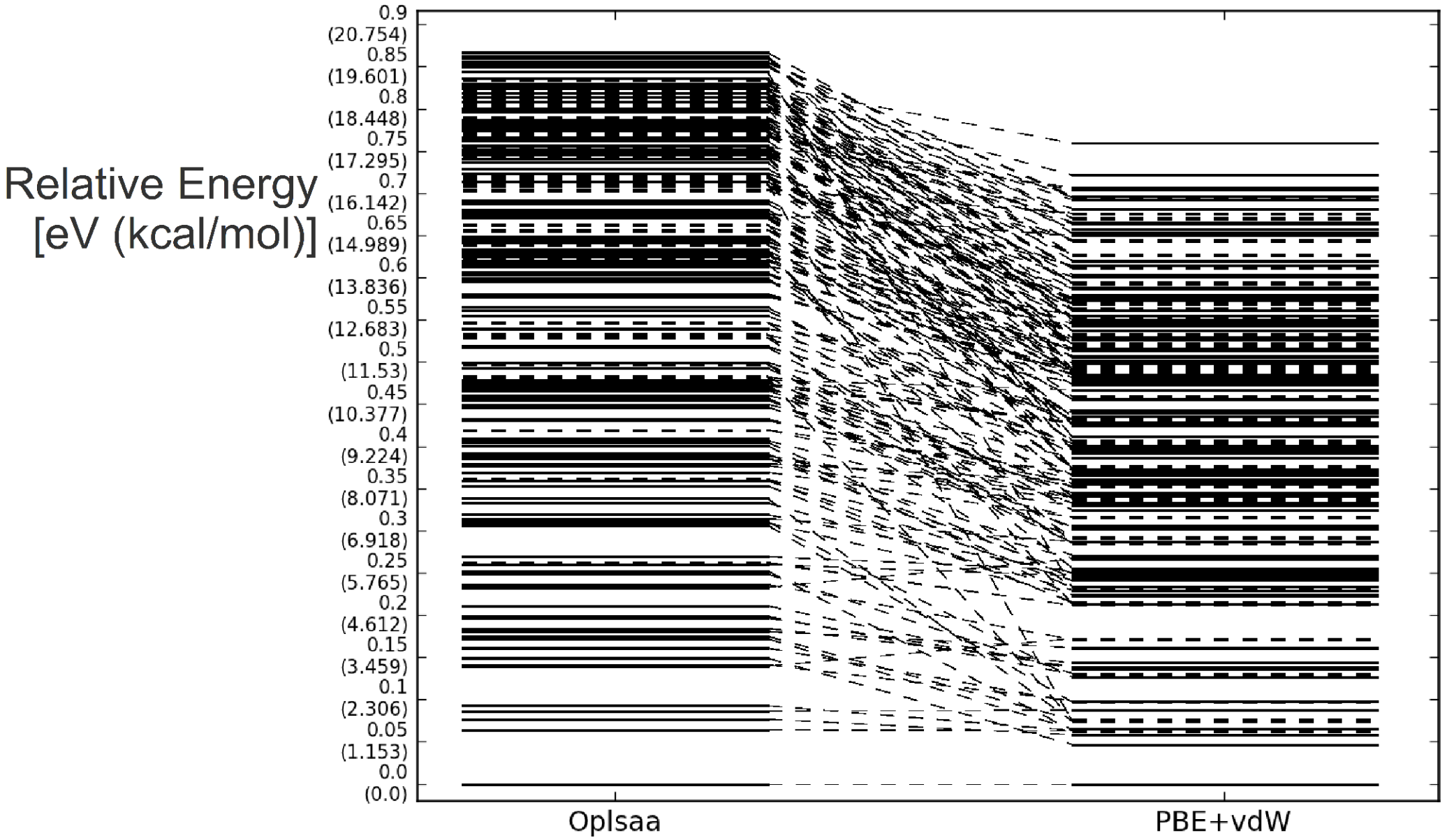
ID: 021

OPLS-AA
Tinker scan results

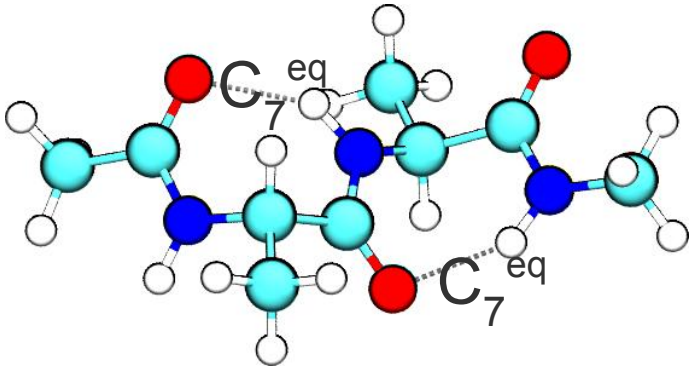
Step 5: Analysing results of the PBE+vdW relaxation

- Get the data from /pub/...
  - Check and use `01_get-results.sh`
- Visualize
- Measure backbone torsion angles
  - Check and use `02_get-angles.sh`
- Characterize and check differences to OPLS results

Step 5: PBE+vdW vs. OPLS

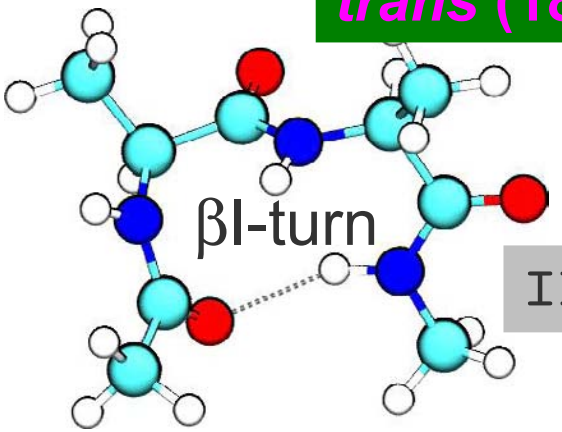


Step 5: Conformers of Ac-Ala-Ala-NMe



ID: 033

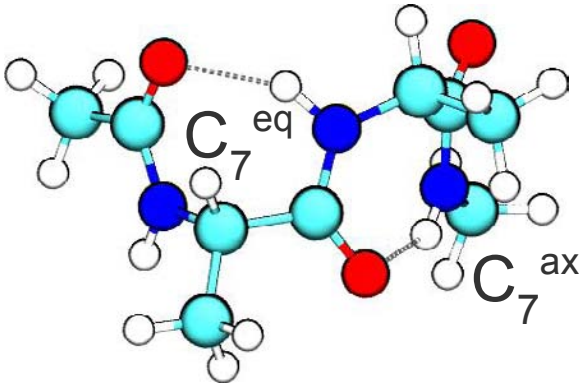
***trans* (180°)**



ID: 007

ID	ome1	phi1	psi1	ome2	phi2	psi2	ome3	DE*
033	-178.5	-82.2	71.4	-171.9	-85.1	71.0	-175.0	0.0
007	-168.6	-75.3	-12.9	172.9	-98.7	8.5	176.7	1.1
034	-179.1	-81.8	72.1	-179.5	68.0	-51.8	-177.3	1.3
005	174.3	-99.9	107.8	-11.9	-114.1	112.8	-176.4	1.7

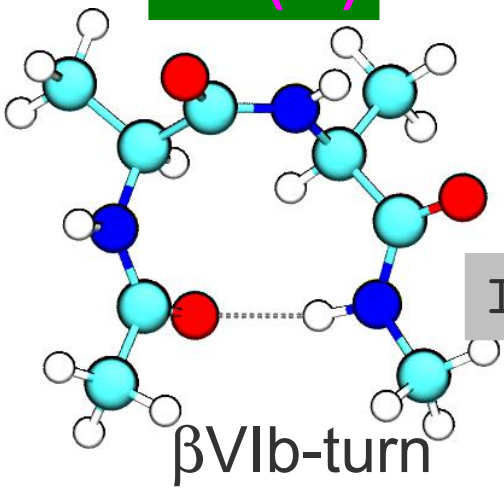
* in kcal/mol



ID: 034

PBE+vdw relaxed
light species defaults

***cis* (0°)**



ID: 005

Step 6: Fixed-geometry energies from different XC-functionals

Modify `control.in`:

- PBE + vdW:

```
xc pbe
vdw_correction_hirshfeld
```

- PBE:

```
xc pbe
```

- PBE0 + vdW:

```
xc pbe0
vdw_correction_hirshfeld
```

- PBE0:

```
xc pbe0
```

- M06-l:

```
xc pbe
total_energy_method m06l
```

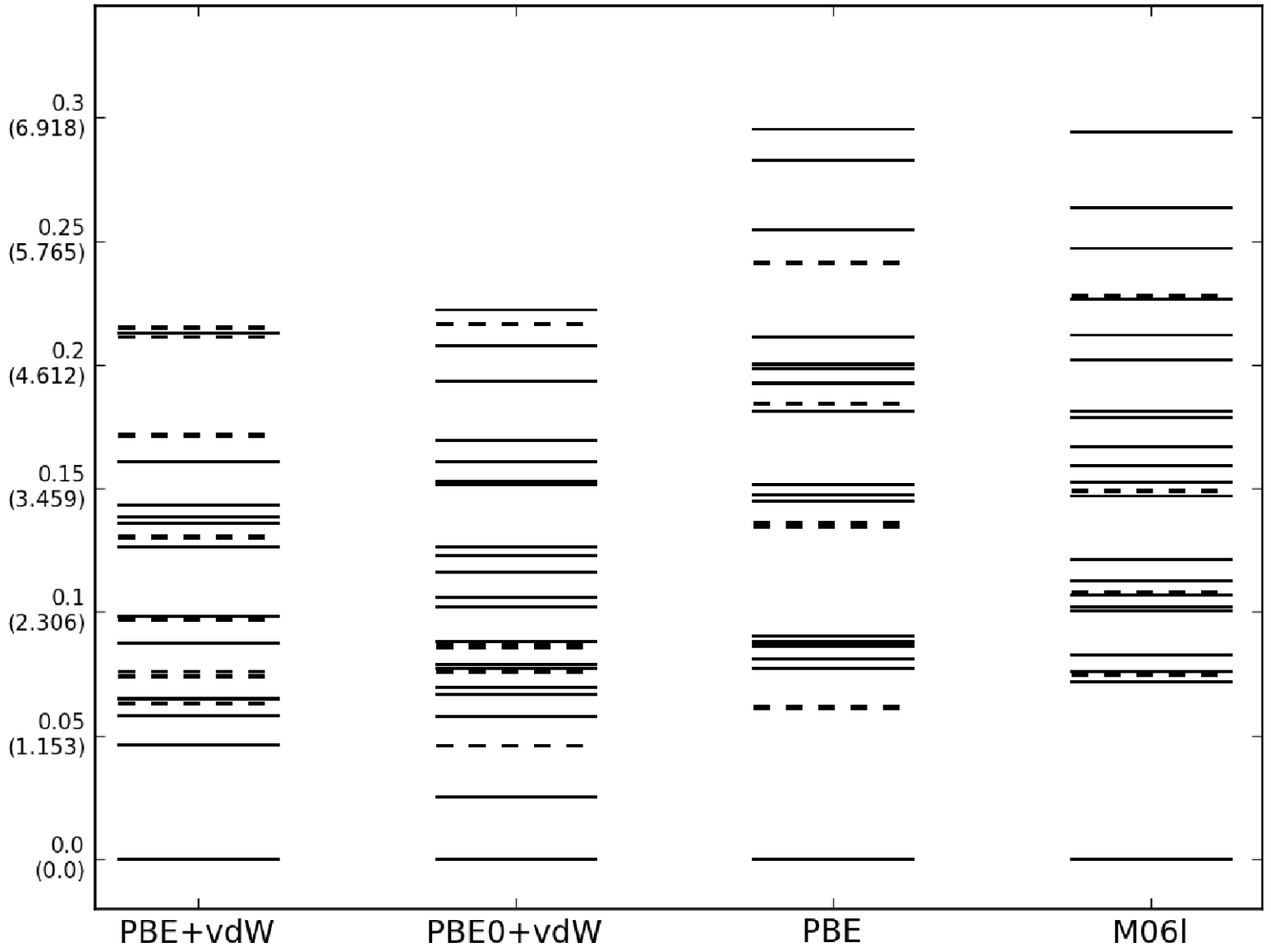
- PBE relaxation*:

```
xc pbe
relax_geometry trm 1.e-2
```

* only for the independent work

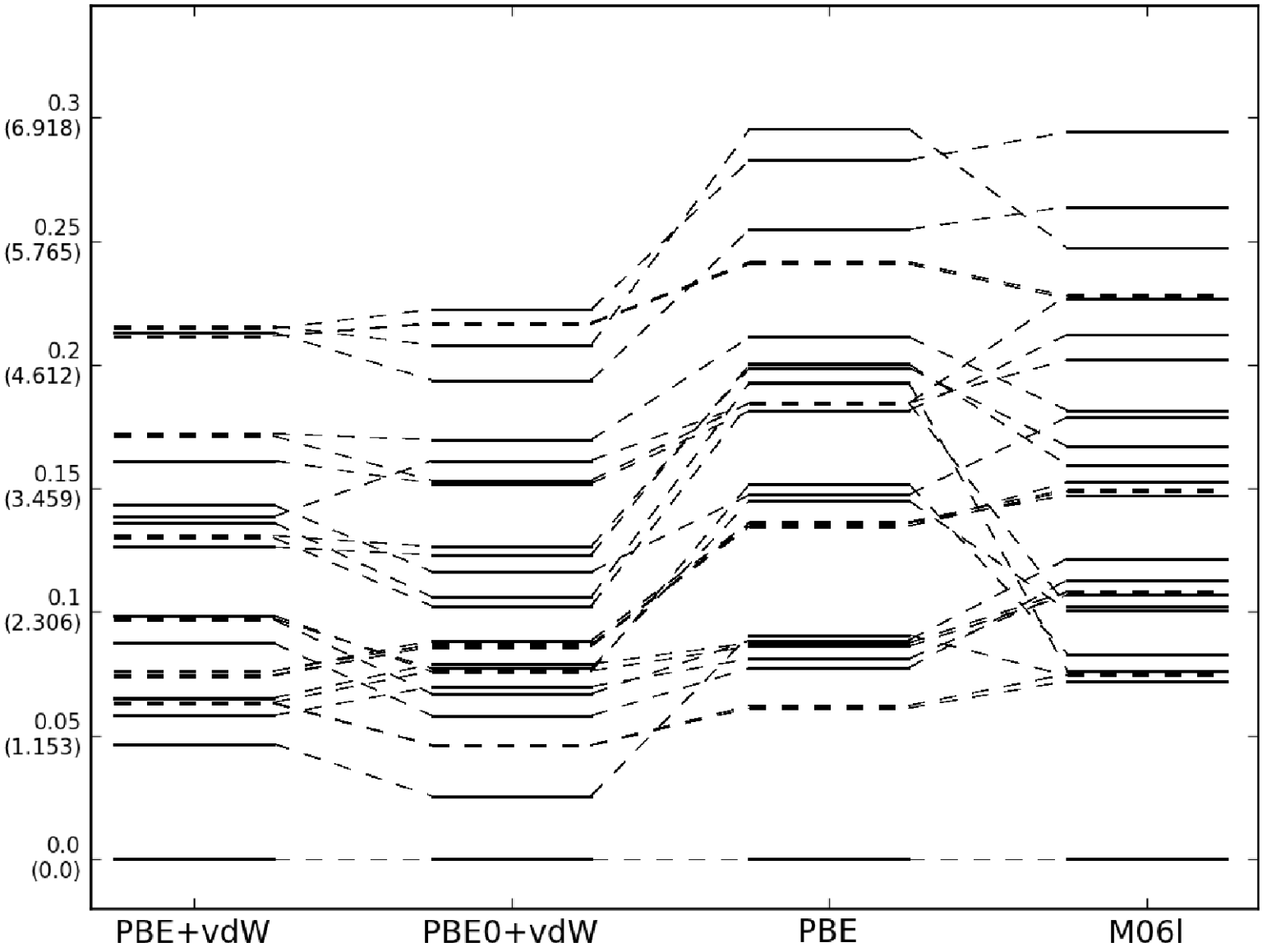
Step 7: Fixed-geometry energies from different XC-functionals

Relative Energy [eV (kcal/mol)]



Step 7: Fixed-geometry energies from different XC-functionals

Relative Energy [eV (kcal/mol)]



Conclusions

- Dimensionality and complicatedness of the PES
- Importance of van der Waals effects for (bio)molecules
- Which energy function to trust?

- How is the relation between secondary structure and amino acid?

Independent work: PES scan

- Your peptide is stored in `[PATH]`
- Scan the potential energy surface (PES)

```
mpirun -n 4 tinker.scan AcXXNMe.xyz 0 10 20 0.0001 0 > \
AcXXNMe.log
```

0 … automatic selection of torsions, 10 … # search directions,
20 … energy threshold, 0.0001 … energy similarity criterion,
0 … start new run

[energies in kcal/mol]

- For each structure to relax: Create a directory and create/copy input files:

```
foreach ENTRY in LIST_OF_STRUCTURES
    do
    [SETUP]
    [SUBMIT]
done
```

Your weekend

IW 1: PES scan Ac-Xaa-Xaa-NMe

IW 2: Setup and submission of PBE+vdW relaxations

Over night

Fri., July 15

IW 3: Analysis of OPLS scan and PBE+vdW relaxations

IW 4: Setup of single-Point calculations (PBE, M06-l, PBE0+vdW)

IW 4*: Setup PBE relaxation without vdW

IW 5: Submission to SGE

14:00 Excursion!!!

Over night

Sat., July 16

IW 6: Analysis of PBE, M06-l, PBE0+vdW single points

IW 6*: Analysis of PBE relaxations without vdW

IW 7: Summarizing your results and filling in the report template

Sun., July 17

Your weekend

IW 3: Analysis of OPLS scan and PBE+vdW relaxations

IW 6: Analysis of PBE, M06-l, PBE0+vdW single points

Please remember, the weekend research project should be fun. You are not obliged to work on it; there will be no negative rating if you decide to do something else.

In case you prepare a **short report** of your work, **send it to us:**
`{tkatchen,baldauf,ropo}@fhi-berlin.mpg.de`

14:00 Excursion!!!

Over night

Over night

Fri., July 15

Sat., July 16

Sun., July 17