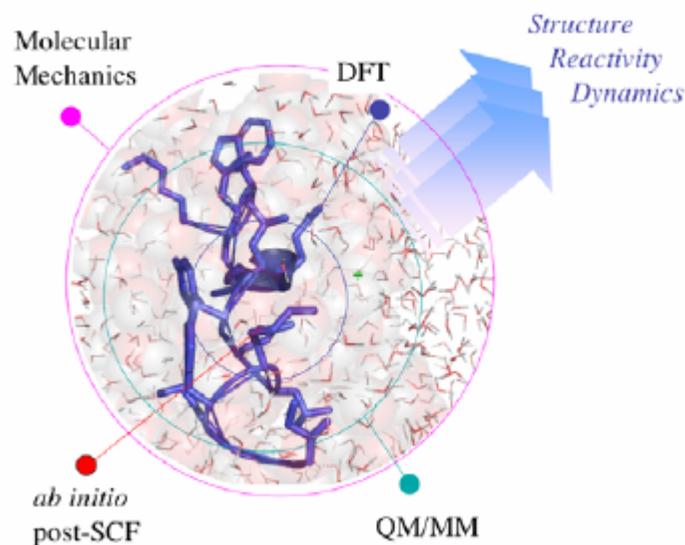


Societa' Chimica Italiana – Gruppo Interdivisionale di Chimica Computazionale

## Scuola di Chimica Computazionale

Introduzione, per Esercizi, all'Uso del Calcolatore  
in Chimica Organica e Biologica



**Modellistica Molecolare:  
un'introduzione alla meccanica  
molecolare e all'analisi  
conformazionale**

*Laura Belvisi – Universita' di Milano  
Dip. di Chimica Organica e Industriale*



**Siena, Aula Informatica del  
Complesso Didattico di San Miniato  
25-29 Settembre 2006**

Some of the material presented in this lecture was rielaborated from

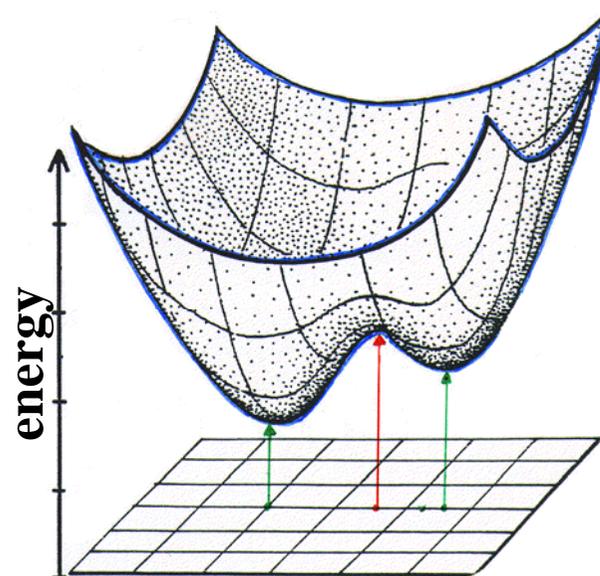
[http://bioinfo.tau.ac.il/~hanochs/cerius2/2004-2005/cerius2\\_winter\\_2004.html](http://bioinfo.tau.ac.il/~hanochs/cerius2/2004-2005/cerius2_winter_2004.html)

a course on Molecular Modeling by Hanoch Senderowitz at Tel Aviv University

# Sampling the PES

- **Energy minimization**
  - Single minimum

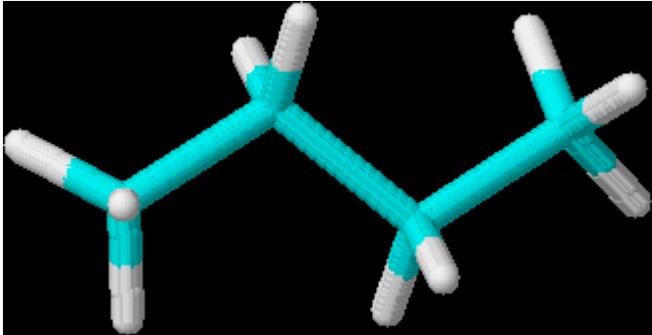
- **Conformational search**
  - Multiple minima



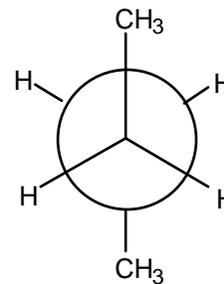
# Conformational Analysis

## ■ Conformers

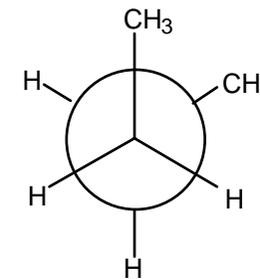
- Structures differing only by rotation around one or more bonds.
- Constitute stationary points on the PES.



**anti butane**



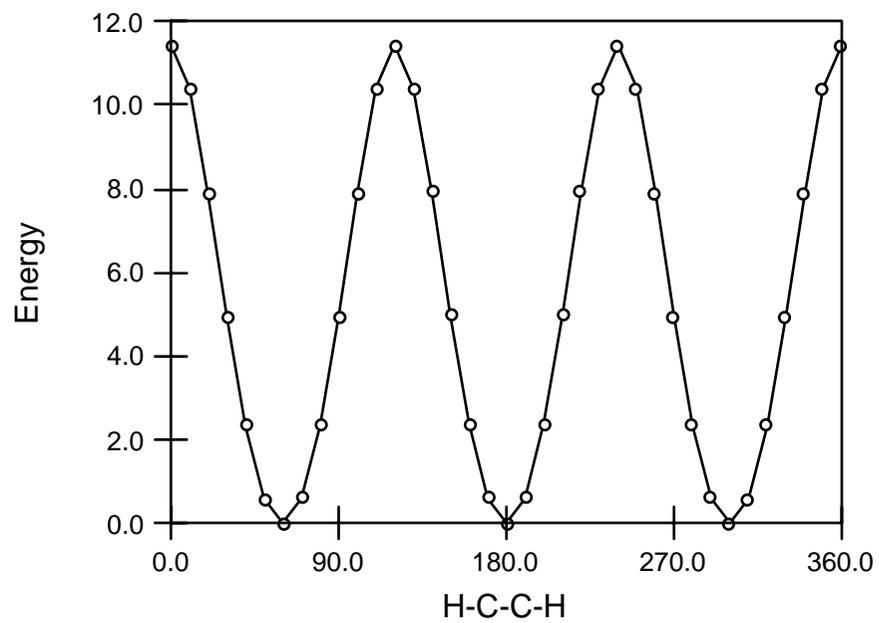
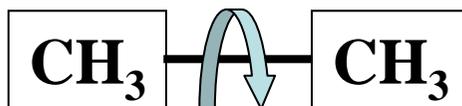
**gauche butane**



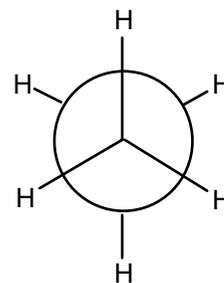
## ■ Conformation

Any point on the PES.

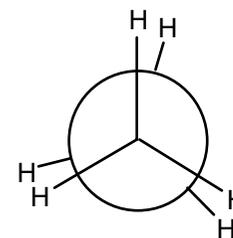
# Ethane



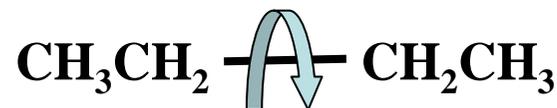
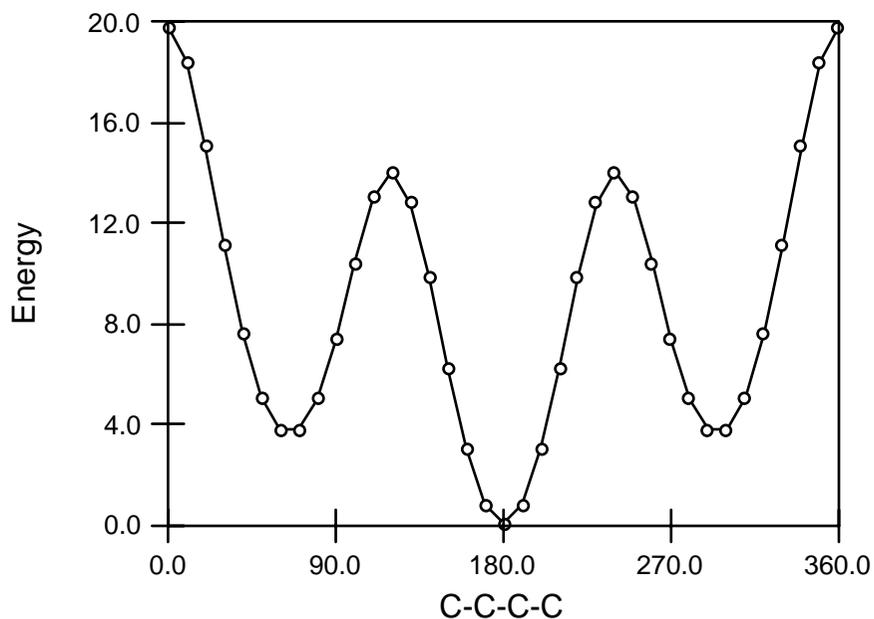
**staggered**  
**H-C-C-H = 180**



**eclipsed**  
**H-C-C-H = 0**

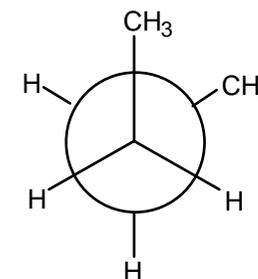
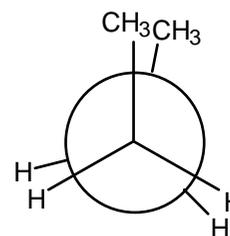


# Butane



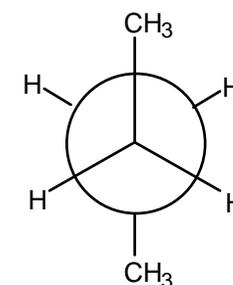
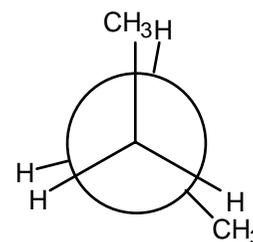
**eclipsed 1**  
C-C-C-C = 0

**gauche+**  
C-C-C-C = +60



**eclipsed 2**  
C-C-C-C = +120

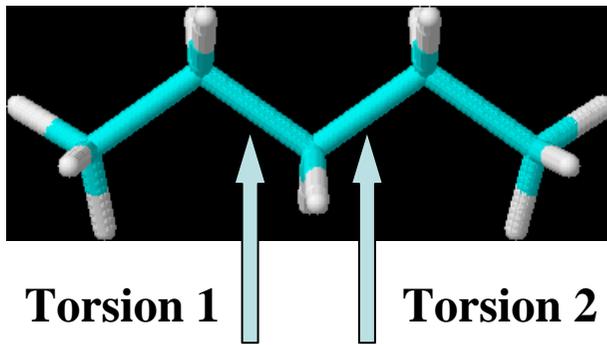
**anti**  
C-C-C-C = 180



Name	Value of C-C-C-C	Type
eclipsed 1	0	Highest TS
gauche <sup>+</sup>	60	Minimum
eclipsed 2	120	TS
anti	180	Global minimum
eclipsed 3	-120	TS
gauche <sup>-</sup>	-60	Minimum

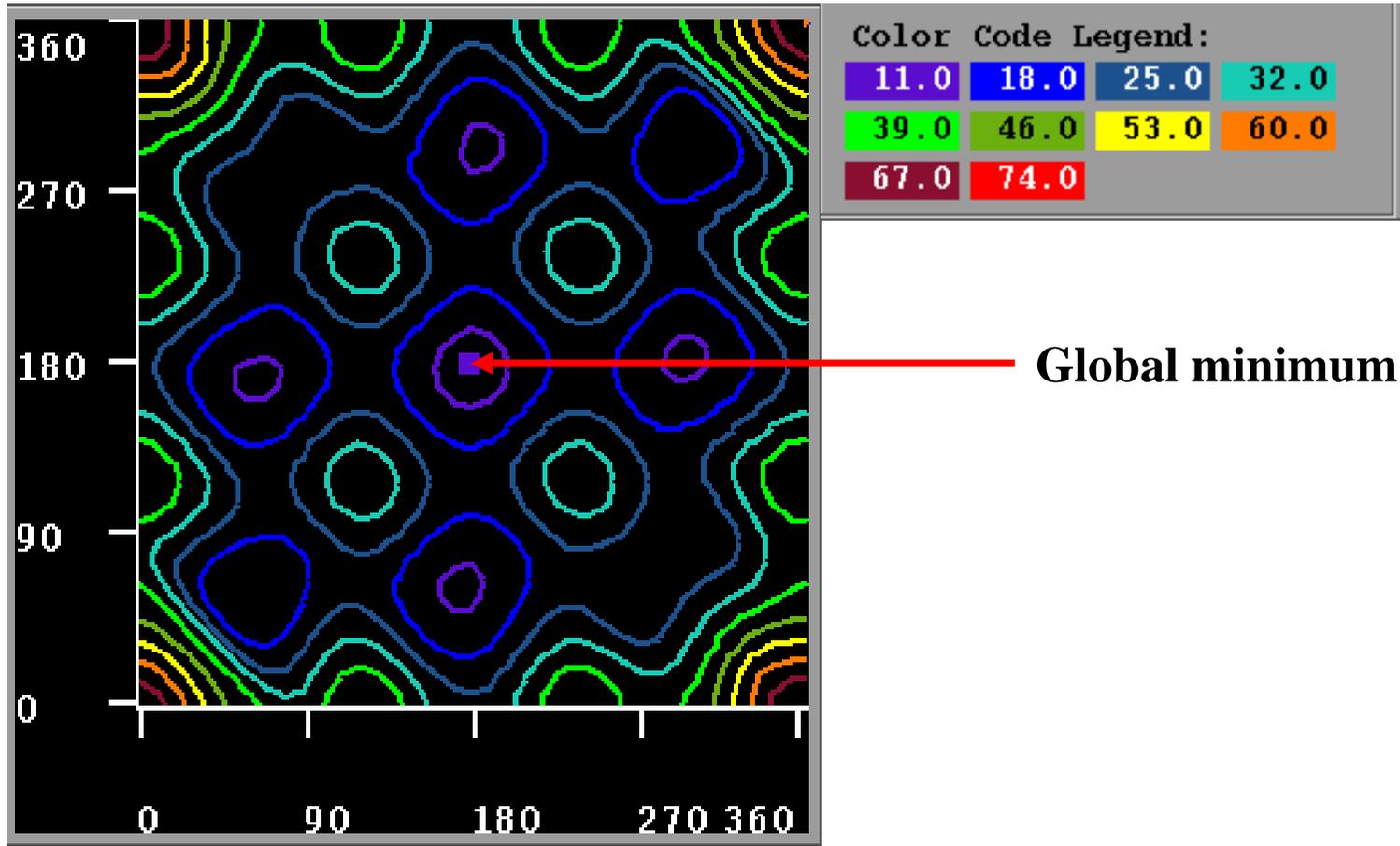
# Pentane

- For a molecular system with  $N$  rotatable bonds, there are  $3^N$  potential minima.



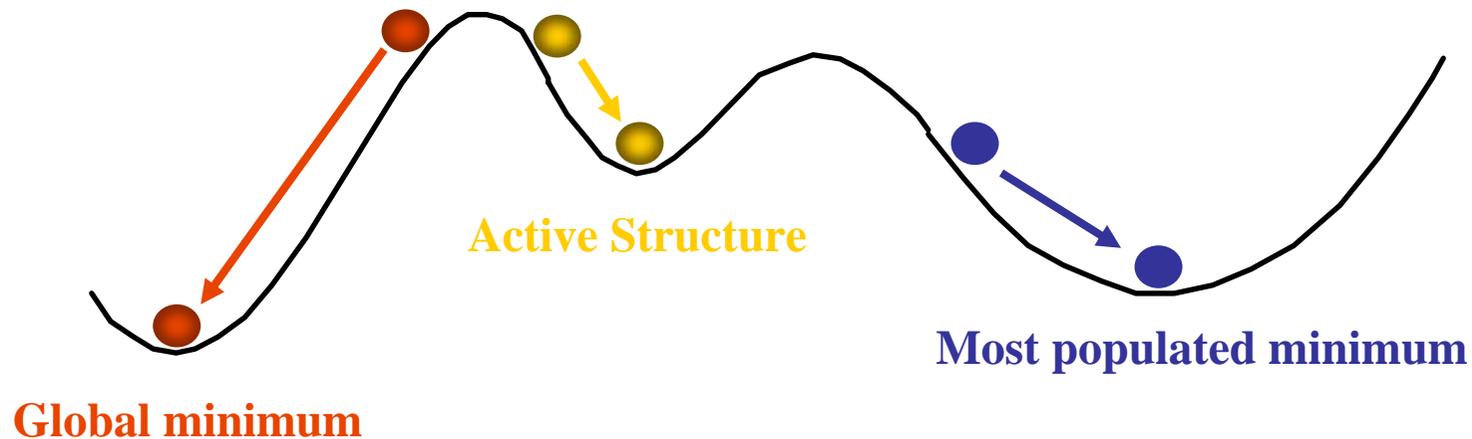
Name	Torsion1	Torsion2	Type
aa	180	180	Global minimum
ag <sup>+</sup>	180	60	minimum
ag <sup>-</sup>	180	-60	minimum
g <sup>+</sup> a	60	180	minimum
g <sup>-</sup> a	-60	180	minimum
g <sup>+</sup> g <sup>+</sup>	60	60	minimum
g <sup>-</sup> g <sup>-</sup>	-60	-60	minimum
g <sup>+</sup> g <sup>-</sup>	60	-60	minimum
g <sup>-</sup> g <sup>+</sup>	-60	60	minimum

# Pentane: Contour Map



# Population of Minima

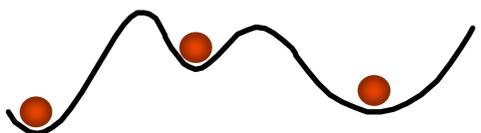
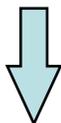
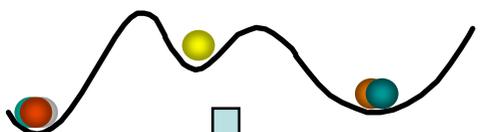
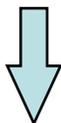
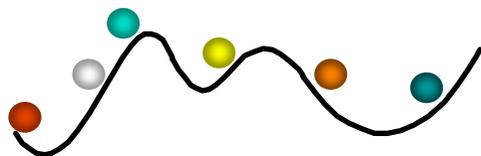
- No minimization method can guarantee the location of the global energy minimum.
- But, is this what we really want?



# Boltzmann Averaging and Conformational Search

- Boltzmann averaging is exact
- Approximations come about from the set of conformers (or conformations),  $i$ , which is considered.
- If we choose to go with conformational searching, we should at least look for a set of energetically accessible minima.

# Conformational Search Outline



- Randomly or systematically generated starting geometries

- Energy minimization

- Duplicates elimination

- Representative structures for each potential minimum

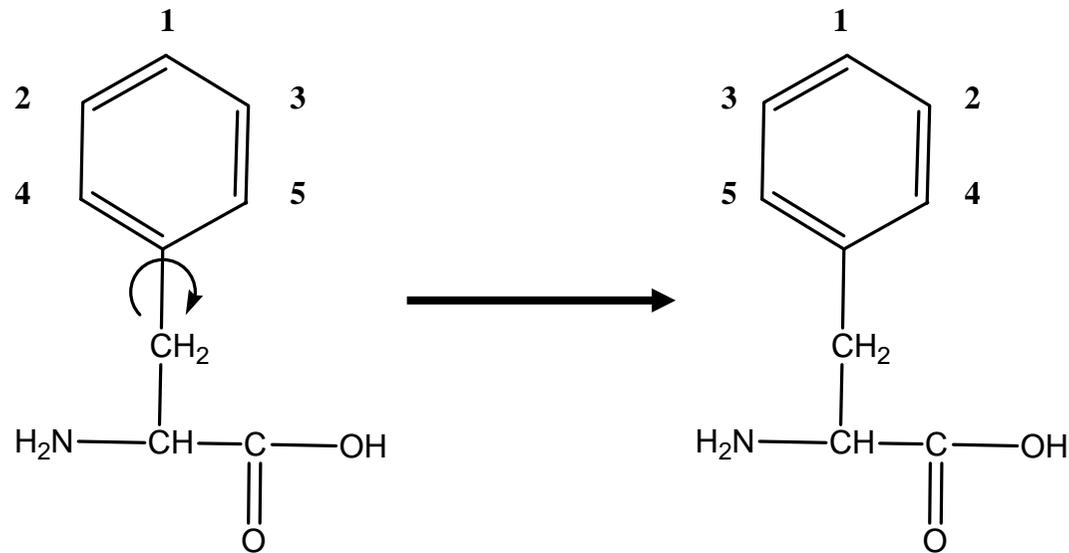
⇕ Gives coverage of potential surface.

⇓ Combinatorial growth.

⇓ Resulting ensemble reflects enthalpy only.

# Elimination of Duplicates

- Usually based on RMS superposition.
- Symmetry considerations should be taken into account.



# Systematic Methods

- **Systematically vary each of the (N) rotatable bonds in the molecule.**

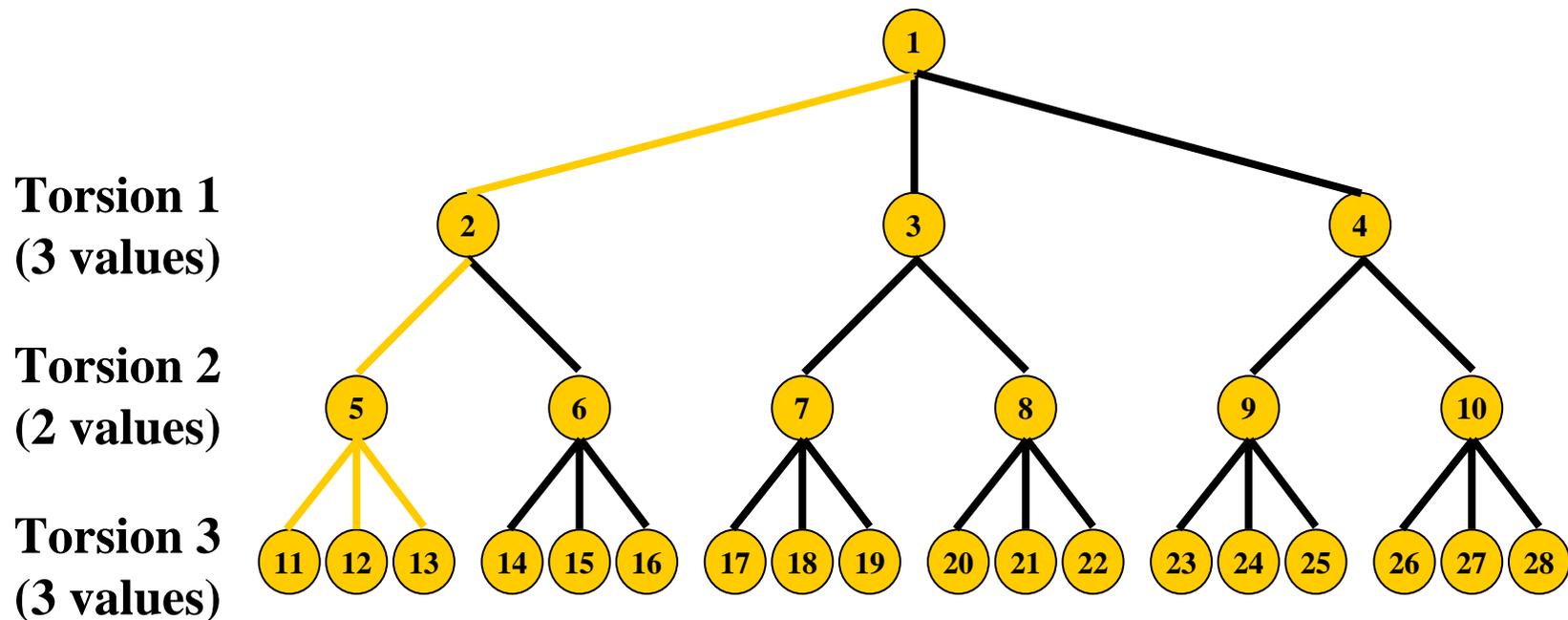
- **Number of conformations =  $\prod_{i=1}^N \frac{360}{\theta_i}$**

- **Rings are transformed into pseudo-acyclic structures.**
  - **Ring closure constraints needed to be checked.**

- **Limitations**

- **Number of processed structure rapidly increases with N.**
  - **Problematic structures may be removed prior to minimization.**
- **Poor coverage of conformational space at the beginning of the run.**

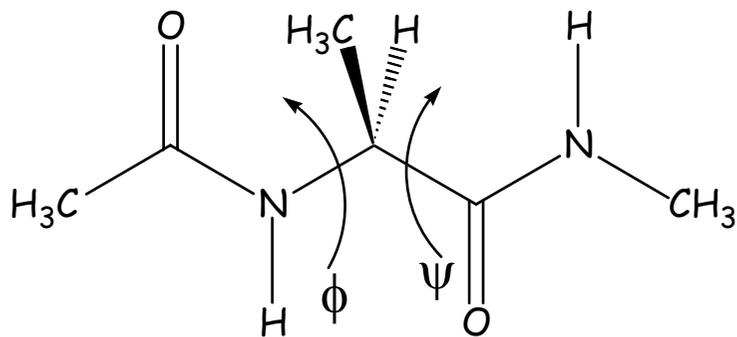
# Removing Problematic Structures



- Tree search proceeds as 1,2,5,11; 1,2,5,12; 1,2,5,13.
- If *e.g.*, the first value of torsion 1 combined with the second value of torsion 2 give rise to a problem (*e.g.*, high steric energy), all conformations containing this combination can be immediately rejected (14, 15, 16).

# Systematic Methods: Example

Ac-Ala-NHMe

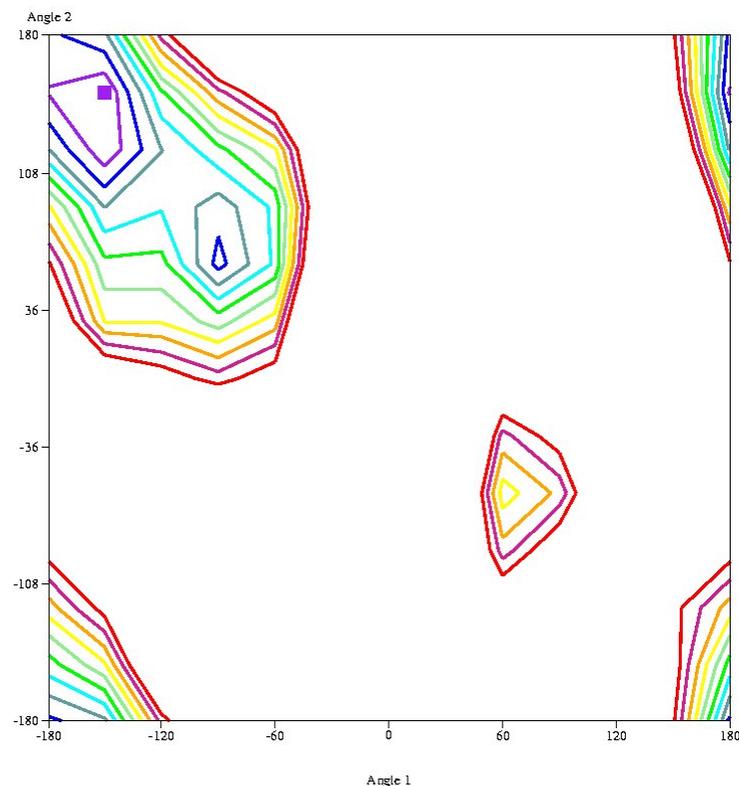


## Ramachandran plot

The grid is generated by systematic variation of the  $\phi$  (angle 1),  $\psi$  (angle 2) torsional angles, followed by partially constrained minimization of the molecule.

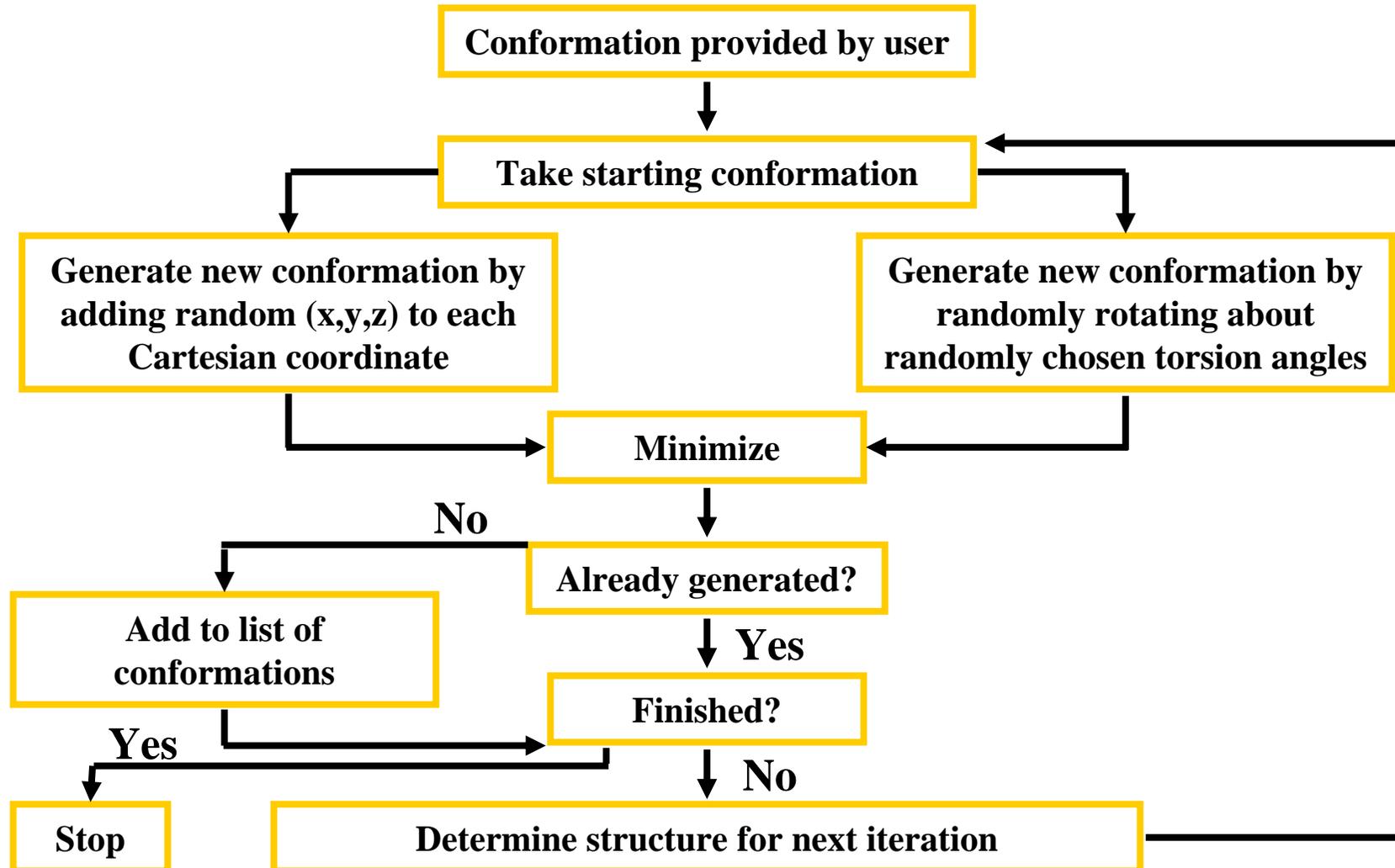
- $\phi, \psi$   $-160^\circ, 140^\circ$  ( $E = 0$  KJ/mol)
- $\phi, \psi$   $-80^\circ, 70^\circ$  ( $E = 3$  KJ/mol)
- $\phi, \psi$   $70^\circ, -60^\circ$  ( $E = 12$  KJ/mol)

$\phi, \psi$  conformational map of N-acetyl-alanine-N-methylamide



AMBER\*, MacroModel V9.1

# Random Search (e.g., MCMM)



# Random Search: Points for consideration

- **Choosing structure for next iteration**
  - Last structure
  - Random structure
  - Least used structure
  - Lowest energy structure
- **Termination criteria**
  - Pre-defined number of steps
  - Convergence (no new structures generated)
- **Potential limitations**
  - Poor convergence at the end of the run.

# Random Search: Points for consideration

Random search methods are often called “Monte Carlo search”, because the random conformational change is driven by randomly generated numbers.

This is not to be confused with Monte Carlo Metropolis simulations of chemical systems.

In a Monte Carlo (MC) conformational search the random move is followed by an energy minimization (EM) step and not by the application of Metropolis algorithm.

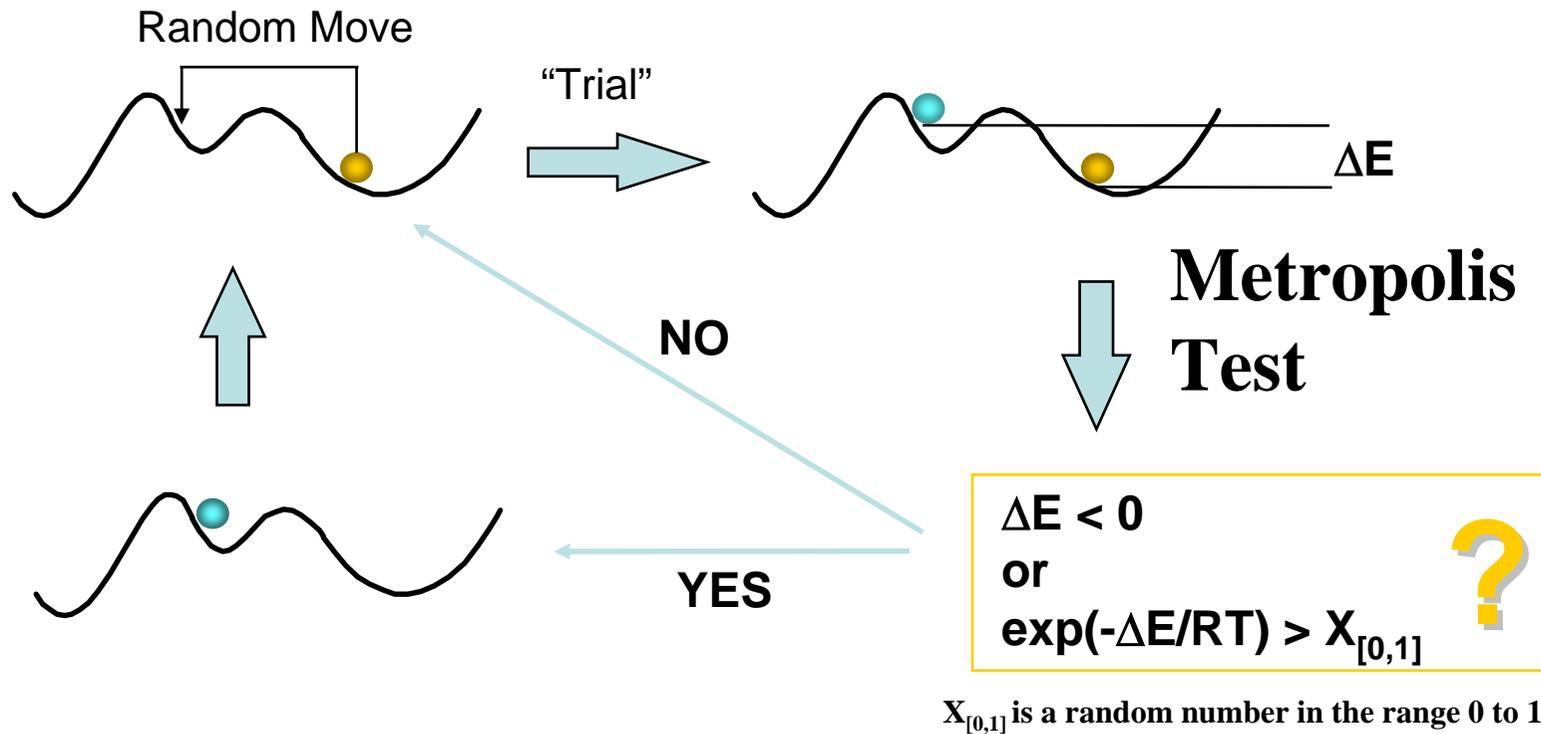
So the output of a MC/EM conformational search is not a thermodynamic ensemble, but a collection of minimum energy conformations with their associated relative energy.

# Additional Methods

- **Distance geometry**
  - **derive molecular conformations from experimental data**
- **Metropolis Monte Carlo**      **explore the entire conformational space of molecules.**
- **Molecular dynamics simulations**

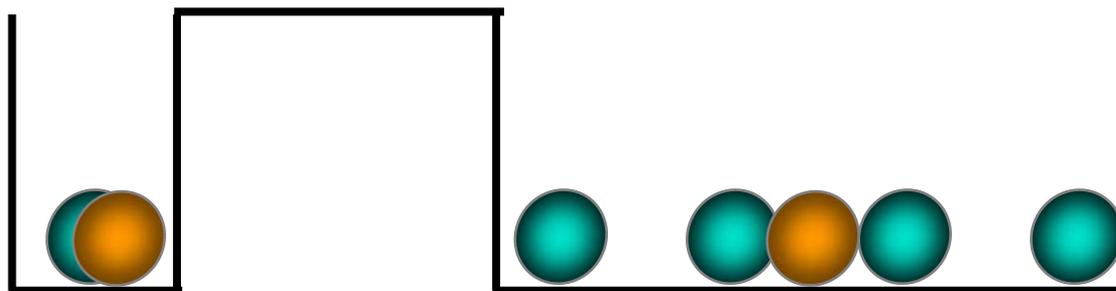
# Metropolis Monte Carlo

- A method for sampling the PES.
- Allows for uphill moves.



# Deriving Molecular Properties from the PES

- Any molecular property is calculated by averaging the properties of all points (conformations) on the PES. The average weights are given by the relative free energies of the conformations.



- **Energy minimization and conformational search**

- The two wells have equal enthalpies and thus contribute equally to each molecular property

- **Molecular simulations (MD, MC, MD/MC)**

- The two wells have different entropies and consequently free energies. The wider well contributes more to molecular properties

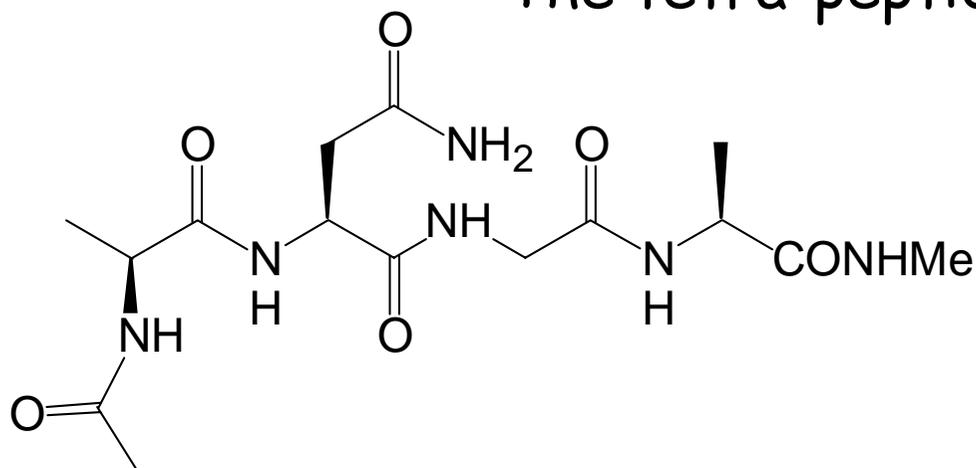
# Moving on (Sampling) the PES

- **Each point on the PES represents a molecular conformation characterized by its structure and energy.**
- **By sampling the PES we can derive molecular properties.**
- **Sampling energy minima only (energy minimization) will lead to molecular properties reflecting the enthalpy only.**
- **Sampling the entire PES (molecular simulations) will lead to molecular properties reflecting the free energy.**
- **In both cases, molecular properties will be derived from the PES.**
- **In order for our results to be meaningful, our sampling of the PES must be converged.**

# Our chemical problem

To obtain a 3D model of a medium size, flexible molecule

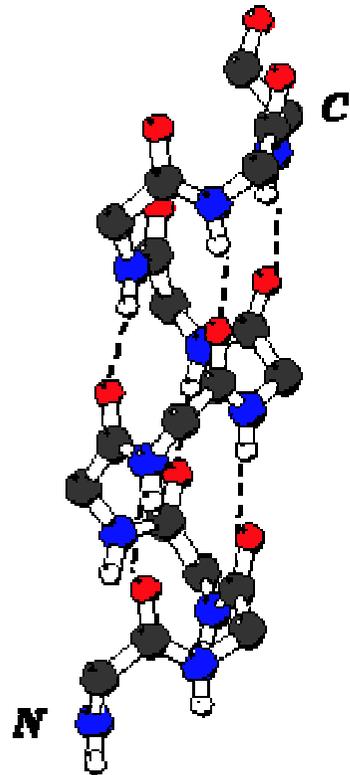
The tetra-peptide *Ala-Asn-Gly-Ala* (ANGA)



Characterize the conformational space of a model linear tetrapeptide by means of molecular mechanics methods and conformational search techniques

# Secondary structure elements of peptides and proteins

↪ local inter-residue interactions mediated by hydrogen bonds

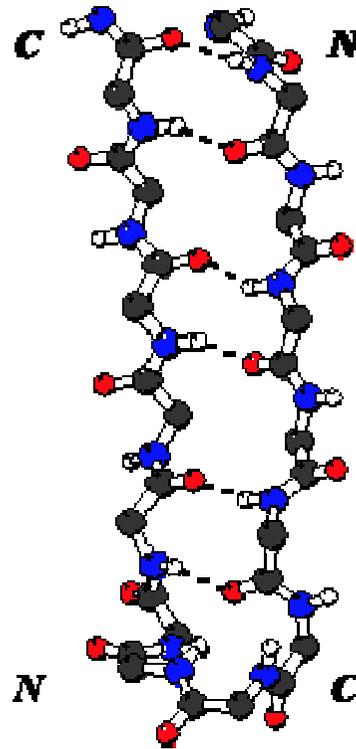


***$\alpha$ -helix***

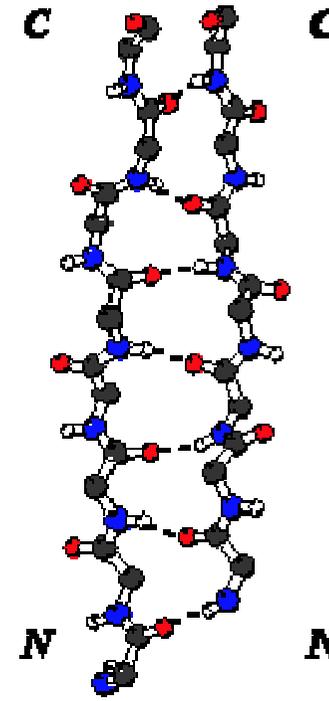
right handed helix

$\phi, \psi, -60^\circ, -40^\circ$

$(\text{C}=\text{O})_i \dots (\text{H}-\text{N})_{i+4}$  H-bonds



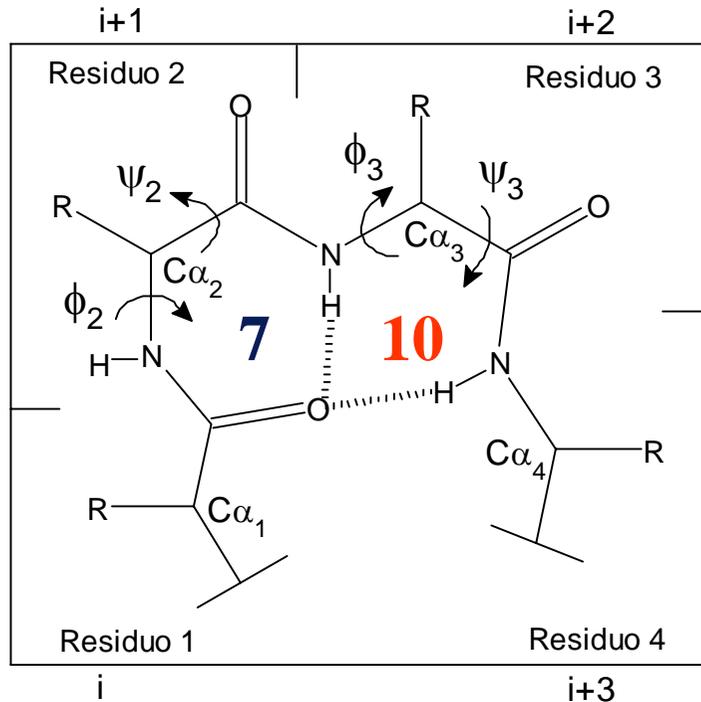
***anti-parallel  
 $\beta$ -sheet***



***parallel  
 $\beta$ -sheet***

# Secondary structure elements of peptides and proteins

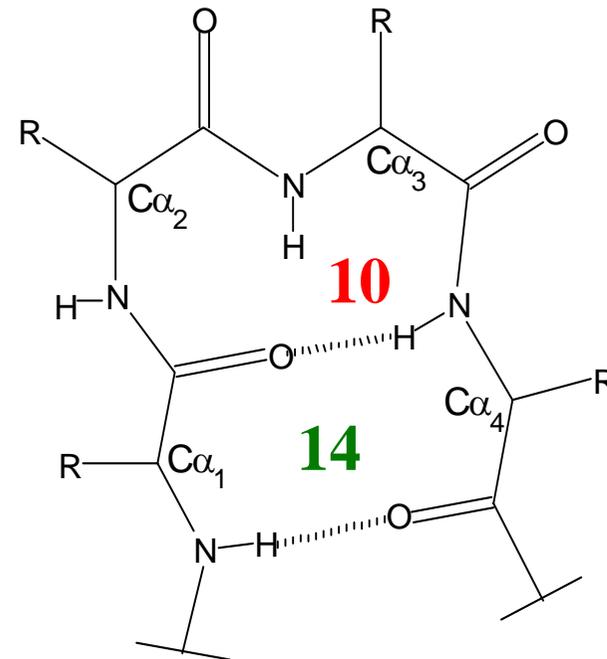
## REVERSE-TURN



$\beta$ -turn: 4 consecutive residues  
 $(C=O)_i \dots (H-N)_{i+3} \rightarrow 10$ -memb. ring H-bond

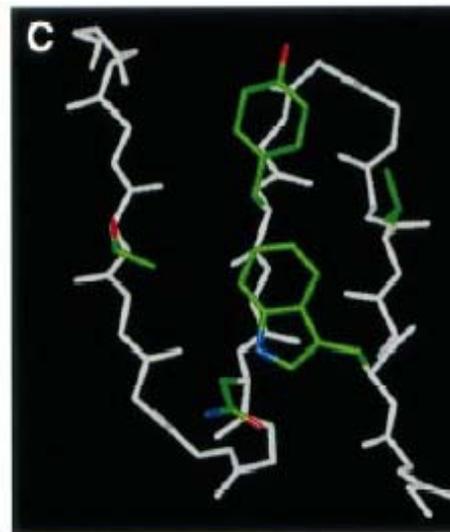
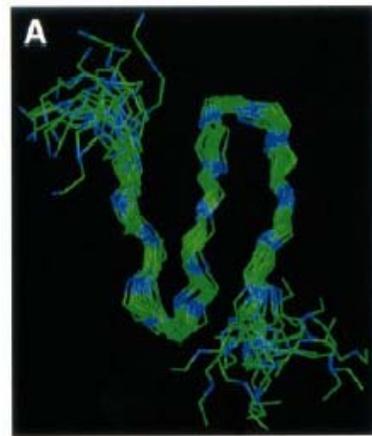
$\gamma$ -turn: 3 consecutive residues  
 $(C=O)_i \dots (H-N)_{i+2} \rightarrow 7$ -memb. ring H-bond

## $\beta$ -HAIRPIN



minimal  $\beta$ -hairpin:  
(10+14) H-bonds

The Asn-Gly (NG) dipeptide is known to induce  $\beta$ -turns in linear peptides as shown by Kortemme et al. in the three-stranded antiparallel  $\beta$ -sheet 20-residue designed peptide Betanova

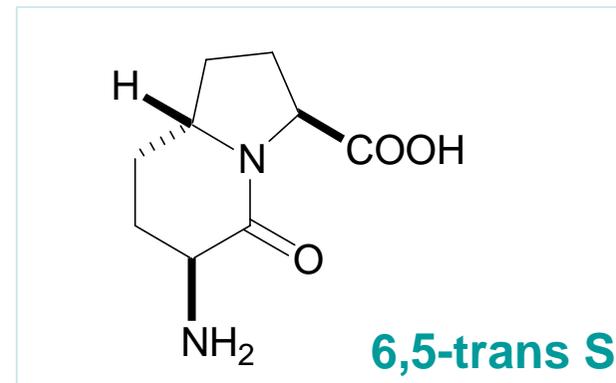


**Betanova**



# DESIGN OF $\beta$ -HAIRPIN-LIKE PEPTIDES CONTAINING REVERSE-TURN MIMICS

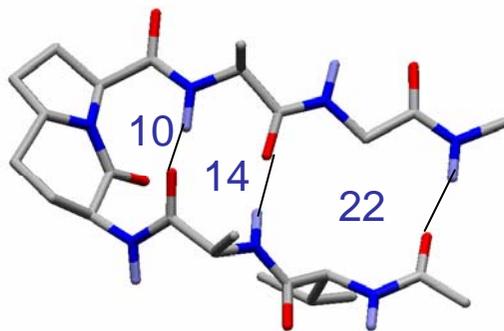
6,5-fused bicyclic lactam showing the best turn-inducing properties within a library of stereoisomeric azabicycloalkane amino acids (Synlett, 2004, 9, 1449-1471)



Constrained dipeptide mimic

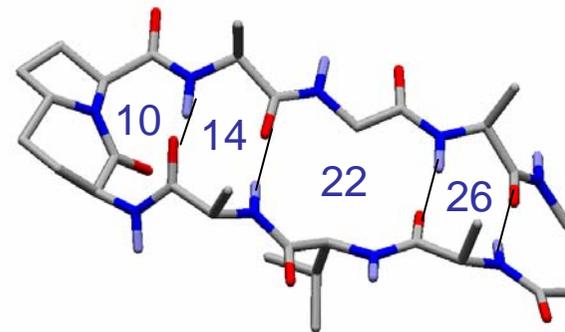
Hexapeptide mimic

Ac-Val-Ala-6,5transS-Ala-Gly-NHMe



Octapeptide mimic

Ac-Ala-Val-Ala-6,5transS-Ala-Gly-Ala-NHMe



Hairpin 10+14+22+26

# EXERCISE

- 1) Start up and use a molecular modeling program (Macromodel/Maestro, [www.schrodinger.com](http://www.schrodinger.com))
- 2) Input a molecular structure for the N-acetyl-N'-methanamide tetra-peptide Ala-Asn-Gly-Ala
- 3) Set up and run molecular mechanics calculations (energy calculation, energy minimization)
- 4) Set up and run a MCMM conformational search
  - ↪ What conformation or conformations are important in this system?
  - ↪ find a set of minimum energy conformers
- 5) Analyze the results
  - ↪ analyze the conformers in terms of geometrical properties, (formation of intramolecular H-bonds or secondary structure elements) and relative population according to the calculated energy function

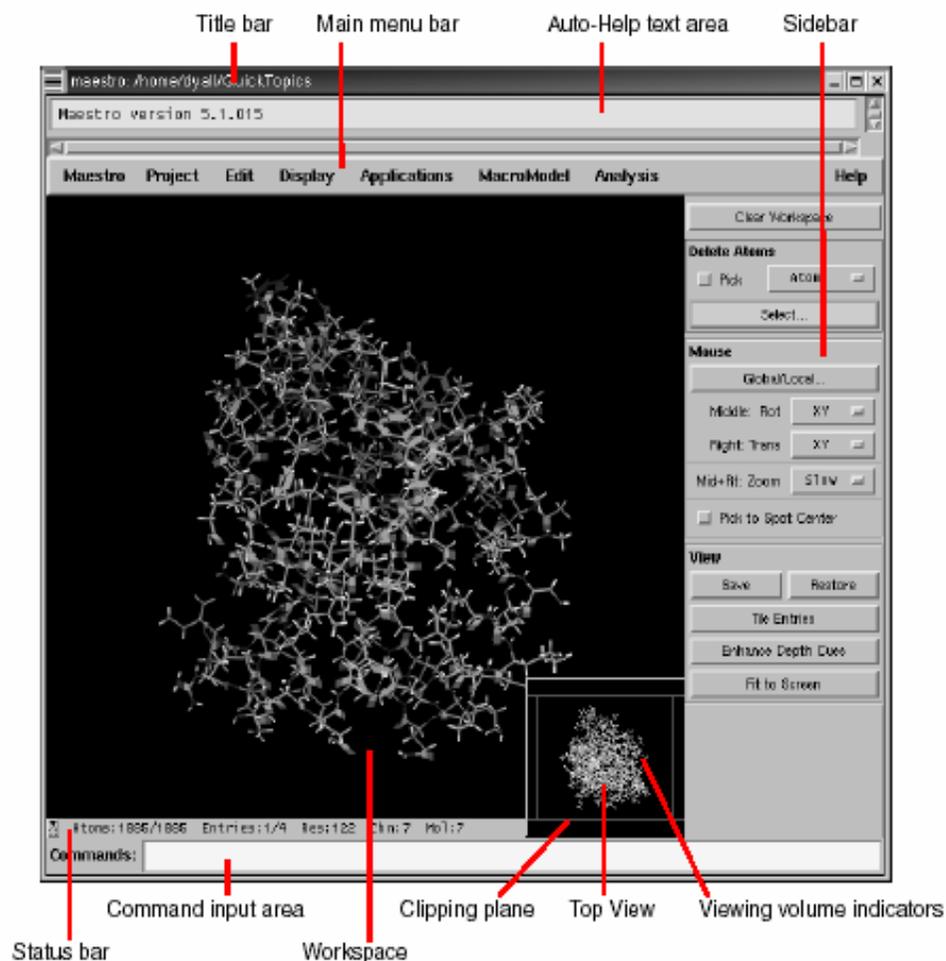
# 1) Start up and use a molecular modeling program

In this exercise we will use

the MacroModel molecular mechanics program (on the Linux cluster)

the Maestro graphical user interface (on the Windows PC)

Maestro main application window



## 2) Building a molecular structure

We will build an input structure for the linear tetrapeptide Ala-Asn-Gly-Ala using

Maestro main application window

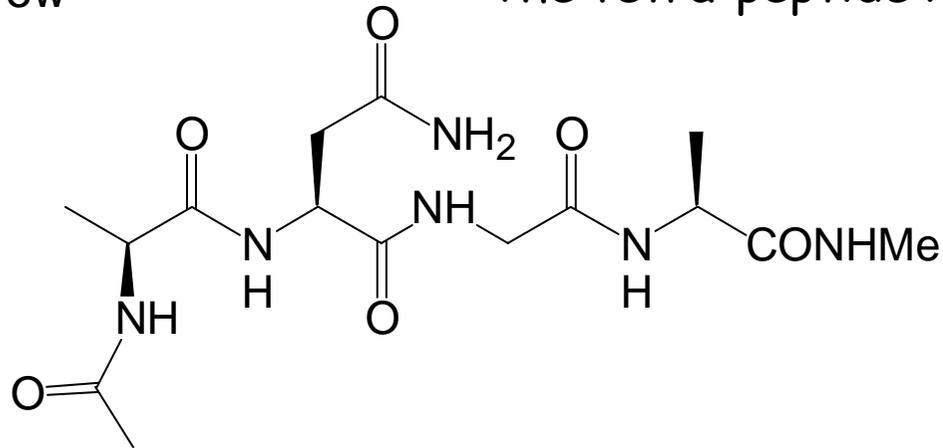
*Edit* menu

*Build* panel

*Fragments* Amino acids

*Method* grow

The tetra-peptide Ala-Asn-Gly-Ala (ANGA)



*Project* menu

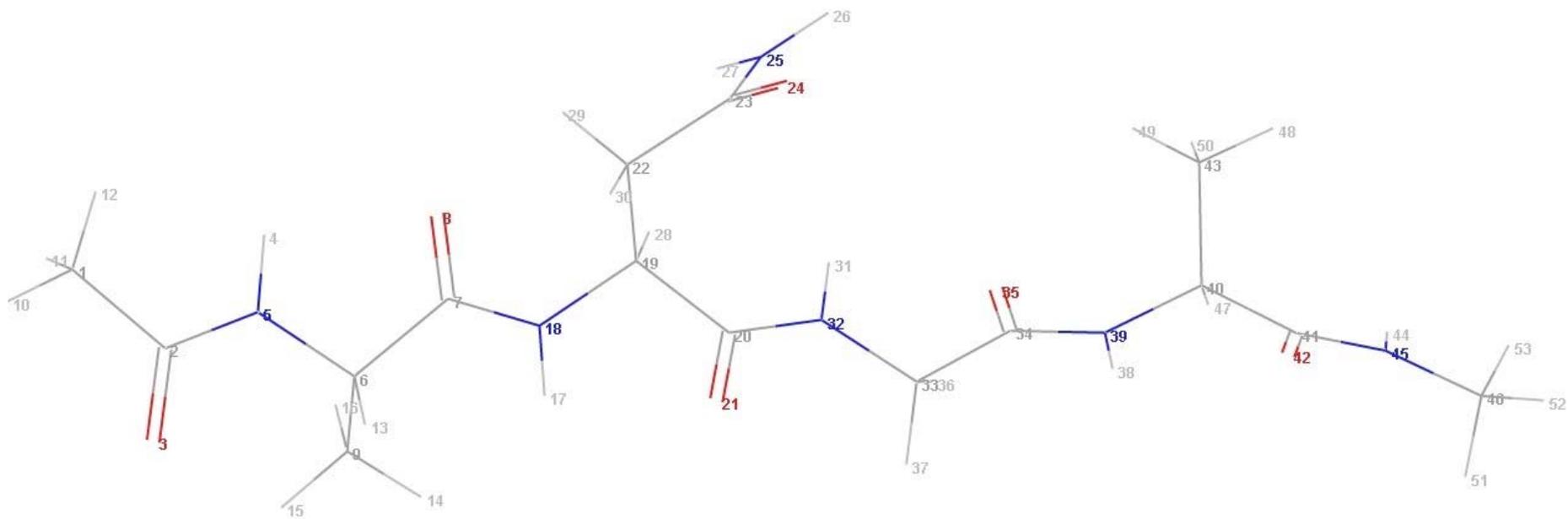
Create entry from workspace

and/or

Export Structures (to save the structure in a file, input.mae)

53 atoms

Display Atom Labels



### 3) Set up and run molecular mechanics calculations energy calculation

Maestro main application window  
*Applications MacroModel*  
*Current Energy* panel

**Potential** folder appears on all of the  
Maestro MacroModel energy panels  
We will specify  
-the AMBER\* force field  
-the water GB/SA solvent treatment  
-the extended cutoff  
for non-bonded interactions

**Ecalc** folder  
We will specify complete energy listing

**Write** job files



MacroModel command file  
Input molecular structure file

jobname.com  
jobname.mae

run job on the Linux cluster

Report file  
Output structure file

jobname.log  
jobname-out.mae

```
jobname.mae  
jobname-out.mae  
OPCD ||||| ||||| ||||| ||||| FFFFF.FFFF FFFFF.FFFF FFFFF.FFFF FFFFF.FFFF  
OPCD ||||| ||||| ||||| ||||| FFFFF.FFFF FFFFF.FFFF FFFFF.FFFF FFFFF.FFFF  
OPCD ||||| ||||| ||||| ||||| FFFFF.FFFF FFFFF.FFFF FFFFF.FFFF FFFFF.FFFF
```

In the current energy calculation the number of low quality stretch, bend, and torsional parameters in use are listed in the MacroModel *jobname.log* file.

## FORCE FIELD

Parameters in the force field files are labelled as to their origin as

O = original from the force field authors

M = modified from the original values

A = added from some other source where no original parameter exists

They are also labelled by quality as

1 = high quality final value

2 = tentative value based on more than one experimental  
or quantum calculation

3 = crude low quality parameter

In addition to calculating the total energy for the structure, MacroModel can be instructed to produce a *.mmo file* during a current energy calculation.

This file contains detailed information on the various contributions to the potential energy and can lead to a number of insights.

(FFVIEW)

### 3) Set up and run molecular mechanics calculations energy minimization

Maestro main application window  
*Applications MacroModel*  
*Minimization* panel

#### Potential folder

We will specify

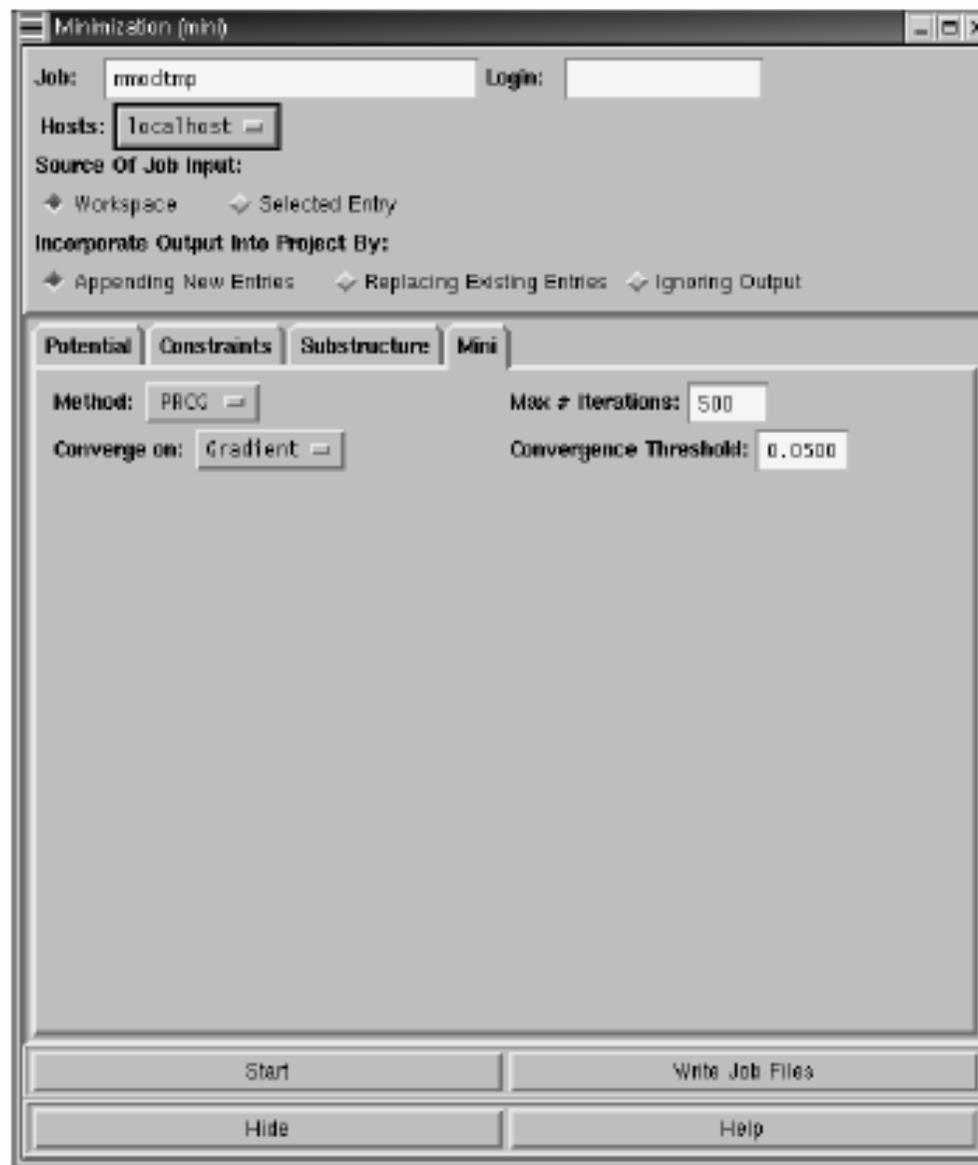
- the AMBER\* force field
- the water GB/SA solvent treatment
- the extended cutoff  
for non-bonded interactions

#### Mini folder

We will specify

- TNCG method
- gradient convergence
- convergence threshold
- number of iterations

**Write** job files



## 4) Set up and run a MCMM conformational search (Monte Carlo Multiple Minimum)

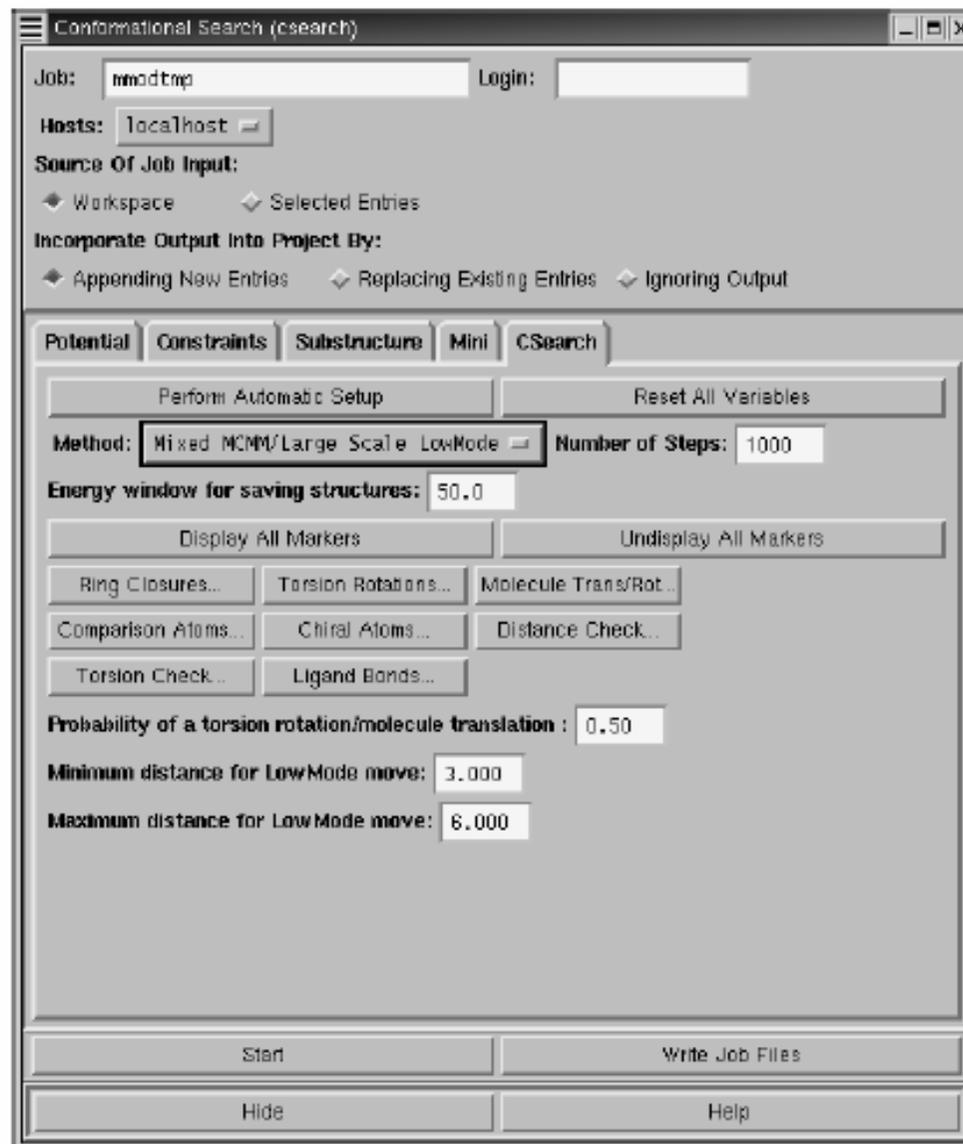
Maestro main application window  
*Applications MacroModel*  
*Conformational search panel*

**Potential** folder  
AMBER\* force field, water GB/SA  
extended cutoff

**Mini** folder  
TNCG, gradient convergence

**Csearch** folder  
search method, energy window,  
set up torsion, chiral center,  
elimination of duplicates

**Write** job files



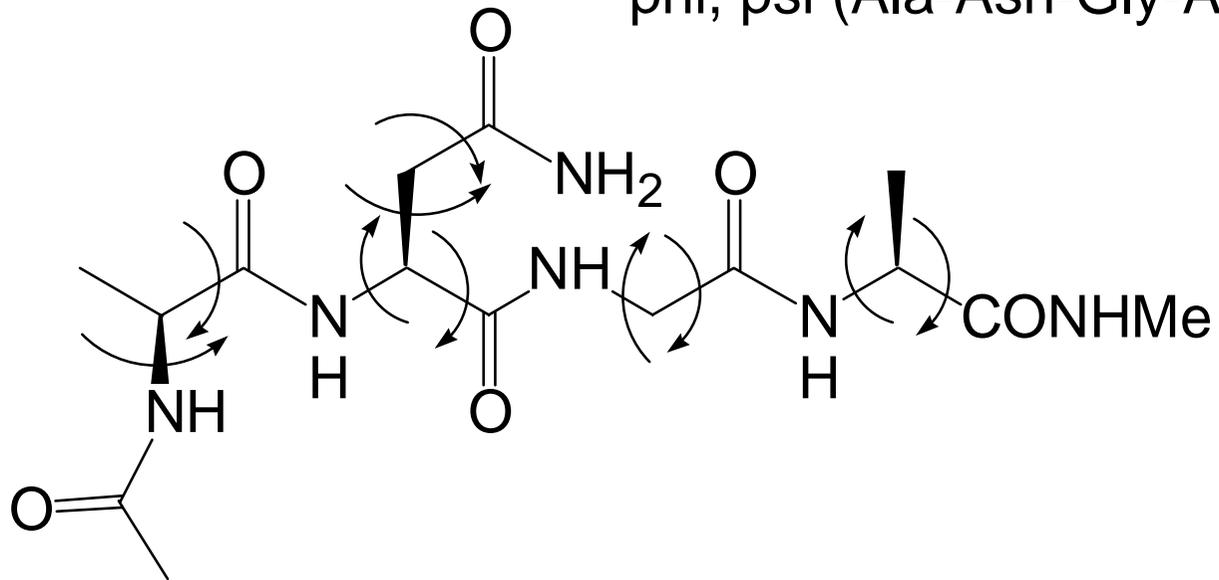
**Method:** Monte Carlo torsional sampling (Restricted, Automatic setup)

**Number of steps:** 20000

**Energy window:** 25 kJ/mol

**Conformational search variables:** 10 torsion angles

phi, psi (Ala-Asn-Gly-Ala); chi1, chi2 (Asn)



**MC starting structure selection:** usage-directed

the least used structures are used as starting geometries, as long as they are within 25 kJ/mol of the lowest energy conformer found in the search

**Comparison atoms:** heavy atoms

structures with equivalent atoms separated by more than 0.50 Å upon superimposition are considered different

## MCMM conformational search .log file

Final report:

238 unique conformations found

238 minimized with good convergence

Found 3 confs within 1.00 kcal/mol ( 4.18 kJ/mol) of glob. min.

Found 9 confs within 2.00 kcal/mol ( 8.37 kJ/mol) of glob. min.

Found 25 confs within 3.00 kcal/mol (12.55 kJ/mol) of glob. min.

Found 128 confs within 5.00 kcal/mol (20.92 kJ/mol) of glob. min.

Found 238 confs within 10.00 kcal/mol (41.84 kJ/mol) of glob. min.

Global minimum E = -706.67 found 364 times.

Total number of structures processed = 20000

Conformations with poor convergence marked with a \*

Conformation	1 (	-706.6652	kJ/mol)	was found	364 times
Conformation	2 (	-706.0395	kJ/mol)	was found	284 times
Conformation	3 (	-705.9930	kJ/mol)	was found	111 times
Conformation	4 (	-702.0687	kJ/mol)	was found	245 times
Conformation	5 (	-701.5905	kJ/mol)	was found	162 times
Conformation	6 (	-701.0535	kJ/mol)	was found	28 times
Conformation	7 (	-700.9714	kJ/mol)	was found	24 times
Conformation	8 (	-700.8899	kJ/mol)	was found	102 times
Conformation	9 (	-700.8743	kJ/mol)	was found	117 times
Conformation	10 (	-697.7496	kJ/mol)	was found	36 times

## Convergence of conformational searching

**There is no way to prove that the search was fully convergent, but there are several methods which can at least indicate that it was not.**

- Conduct a search of approximately 1000-2000 MC steps for every variable degree of freedom
- Check if the number of new unique conformations which are found begin to approach zero
- Ensure all the structures saved are minimized to low gradients and found many times
- Run a second search (using a different sequence of random numbers or a different starting geometry)

## 5) Analyzing and interpreting the results by the Maestro graphical interface

Maestro main application window

*Tools* menu

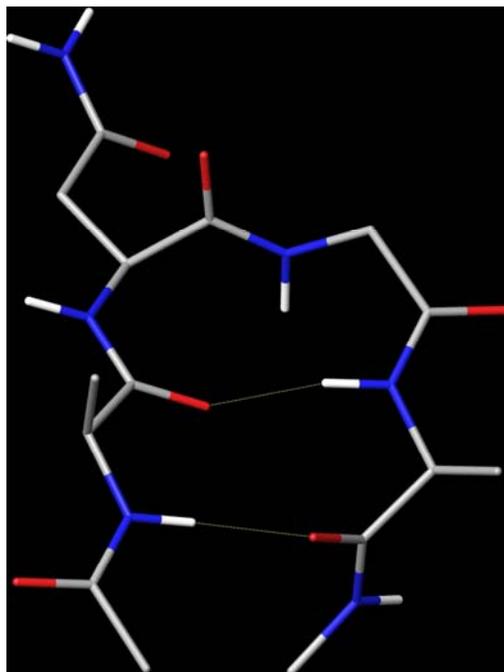
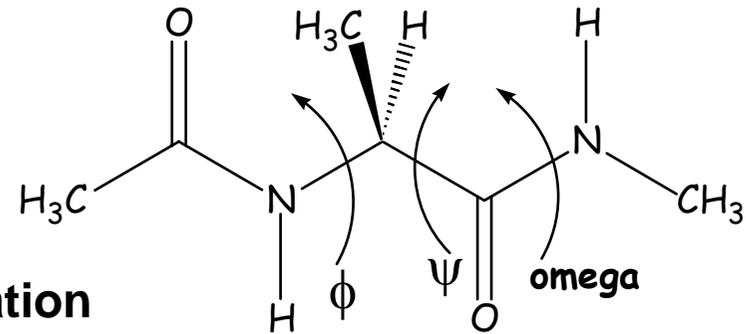
*Measurements* panel

(distance, dihedrals, H-bonds)

### Secondary structure and backbone conformation

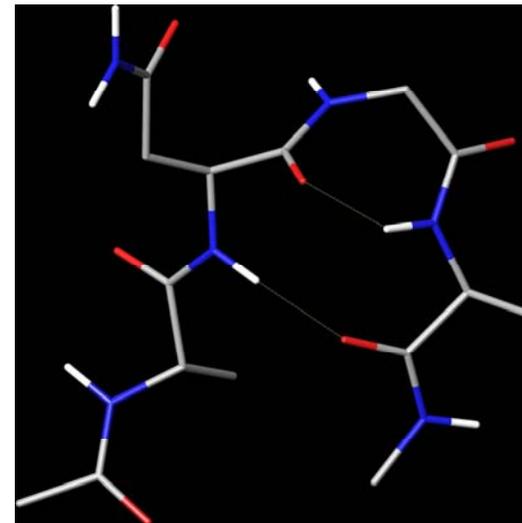
Main chain torsion angles of a polypeptide

phi ( $\Phi = \text{C-N-C}\alpha\text{-C}$ ) psi ( $\Psi = \text{N-C}\alpha\text{-C-N}$ ) omega ( $\Omega$ )



Minimum energy  
conformation 8  
 $\Delta E = 1.4$  kcal/mol

Minimal  $\beta$ -hairpin  
 $\beta$ -II turn Asn-Gly  
(10+14 H-bonds)



Minimum energy  
conformation 9  
 $\Delta E = 1.4$  kcal/mol

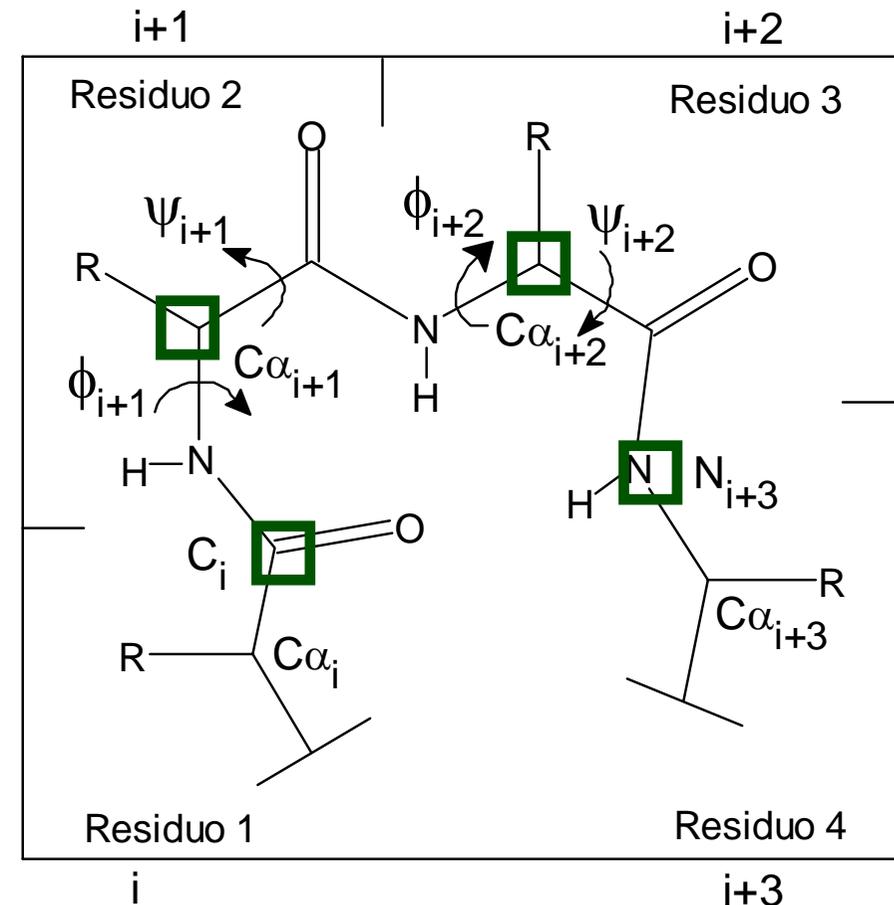
$\gamma$ -turn Gly  
(7+11 H-bonds)

# REVERSE-TURN

$\beta$ -turn	$\phi_{i+1}$	$\psi_{i+1}$	$\phi_{i+2}$	$\psi_{i+2}$
I	-60	-30	-90	0
I'	60	30	90	0
II	-60	120	80	0
II'	60	-120	-80	0
III	-60	-30	-60	-30
III'	60	30	60	30
$\gamma$ -turn	70/ 85	-60/ -70		
inverse $\gamma$ -turn	-70/ -85	60/ 70		

- \* distanza  $C\alpha_i - C\alpha_{i+3} < 7 \text{ \AA}$
- \* angolo diedro improprio  $\beta$   
 $-30^\circ < C_i - C\alpha_{i+1} - C\alpha_{i+2} - N_{i+3} < 30^\circ$

## Criteri di classificazione



- \* parametri legami H intramolecolari:  
distanza  $NH \cdots O=C < 2.5 \text{ \AA}$   
angolo legame  $N-H \cdots O > 120^\circ$   
angolo legame  $H \cdots O=C > 90^\circ$