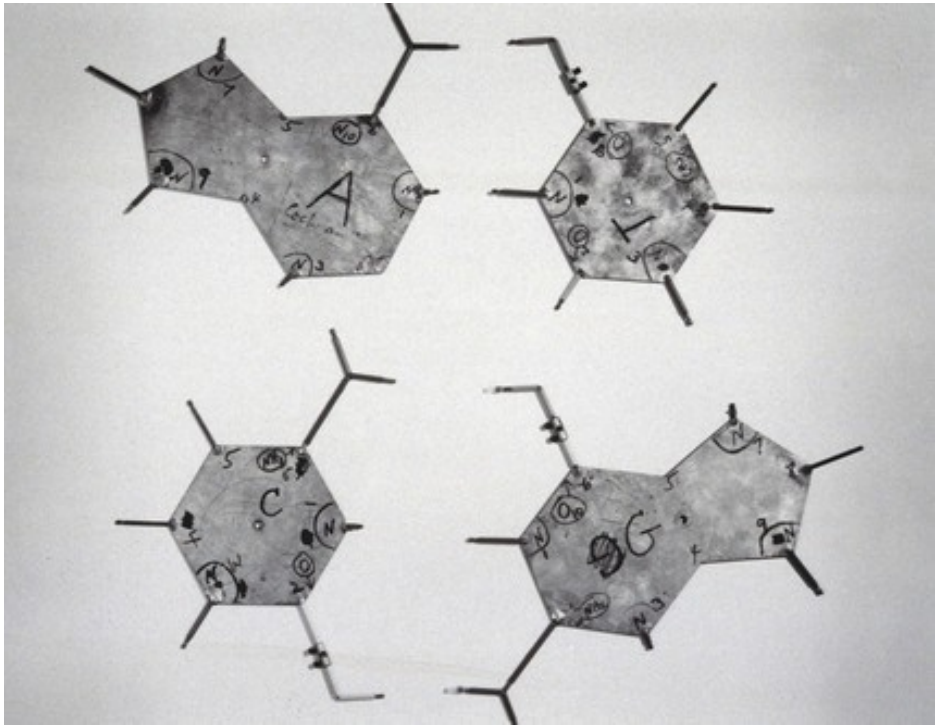
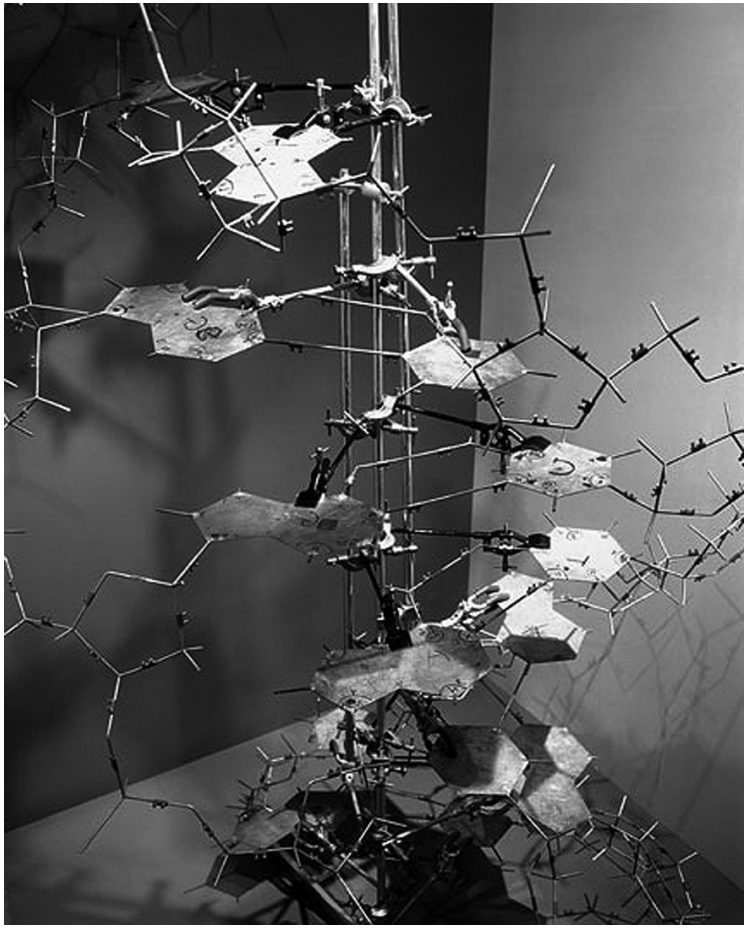
Peptide/Protein Structure Determination Using NMR Restraints and CYANA

CHEM526

Watson and Crick DNA Model





Given what was known about molecular geometry, hydrogen bonding *etc* Watson and Crick could build a model to satisfy the experimental Xray data.

Some Available Programs for Automated Structures Determination

CNS - Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W.L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.;Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren G. *L. Acta Cryst. D* 1998, 54, 905.

X-PLOR-NIH - Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J.Magn.Reson.* 2003, 160, 65.

DYANA/CYANA - Güntert, P.; Mumenthaler, C.; Wüthrich, K. *J. Mol. Biol*. 1997, 273, 283.
-Güntert, P. Prog. NMR Spectrosc. 2003, 43, 105-125.

ARIA - Linge, J. P.; Habeck, M.; Rieping, W., et al. *Bioinformatics* 2003, 19, 315-316.

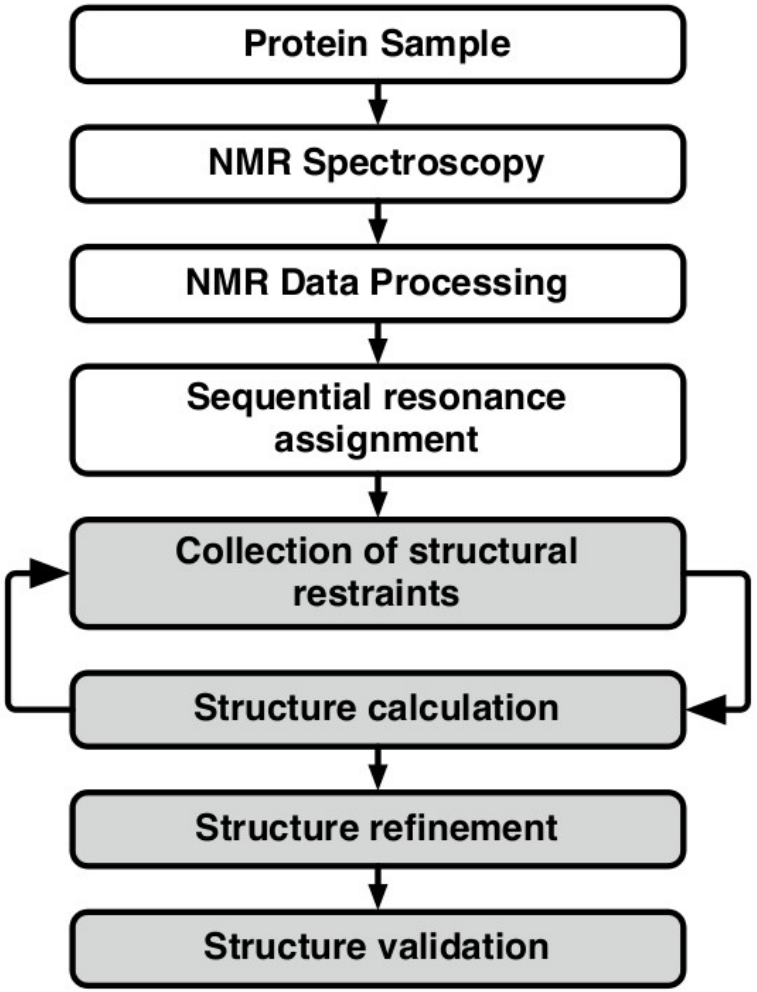
CS-Rosetta - Yang Shen; Oliver Lange; Frank Delaglio; Paolo Rossi; James M. Aramini; Gaohua Liu; Alexander Eletsky; Yibing Wu; Kiran K. Singarapu; Alexander Lemak; Alexandr Ignatchenko; Cheryl H. Arrowsmith; Thomas Szyperski; Gaetano T. Montelione; David Baker; Ad Bax; *Proceedings of the National Academy of Sciences,* 2008, 105(12), 4685-4690.

*CYANA is my favorite, but requires a license.
*CS-Rosetta is very powerful and free but does have a learning curve. Very much worth while learning.
*XPLOR-NIH is free and seems to be flexible in annealing methods and modifiable wrt the force field

Web Portals and Downloads

- https://csrosetta.bmrb.wisc.edu/csrosetta
- http://cns-online.org/v1.3/
- http://aria.pasteur.fr/
- http://www.las.jp/english/products/cyana.html
- http://nmr.cit.nih.gov/xplor-nih/
- http://www-nmr.cabm.rutgers.edu/NMRsoftware/nmr_software.html

Typical Scheme for NMR Structure Determination



- Protein prep in the wet-lab is a critical step. If possible degas the samples and flame seal. Argon degassing buffers and purging nmr samples helps too.
- NMR data acquisition should be routine and shoot for as high as field as possible as the information content is greatest.
- There are a number of resonance interfaces (semi-automated and automated). Always manually check assignments carefully.
- As already seen there are a number of programs for generating structures. We will focus on CYANA with torsional angle dynamics.

Molecular Mechanics Molecule

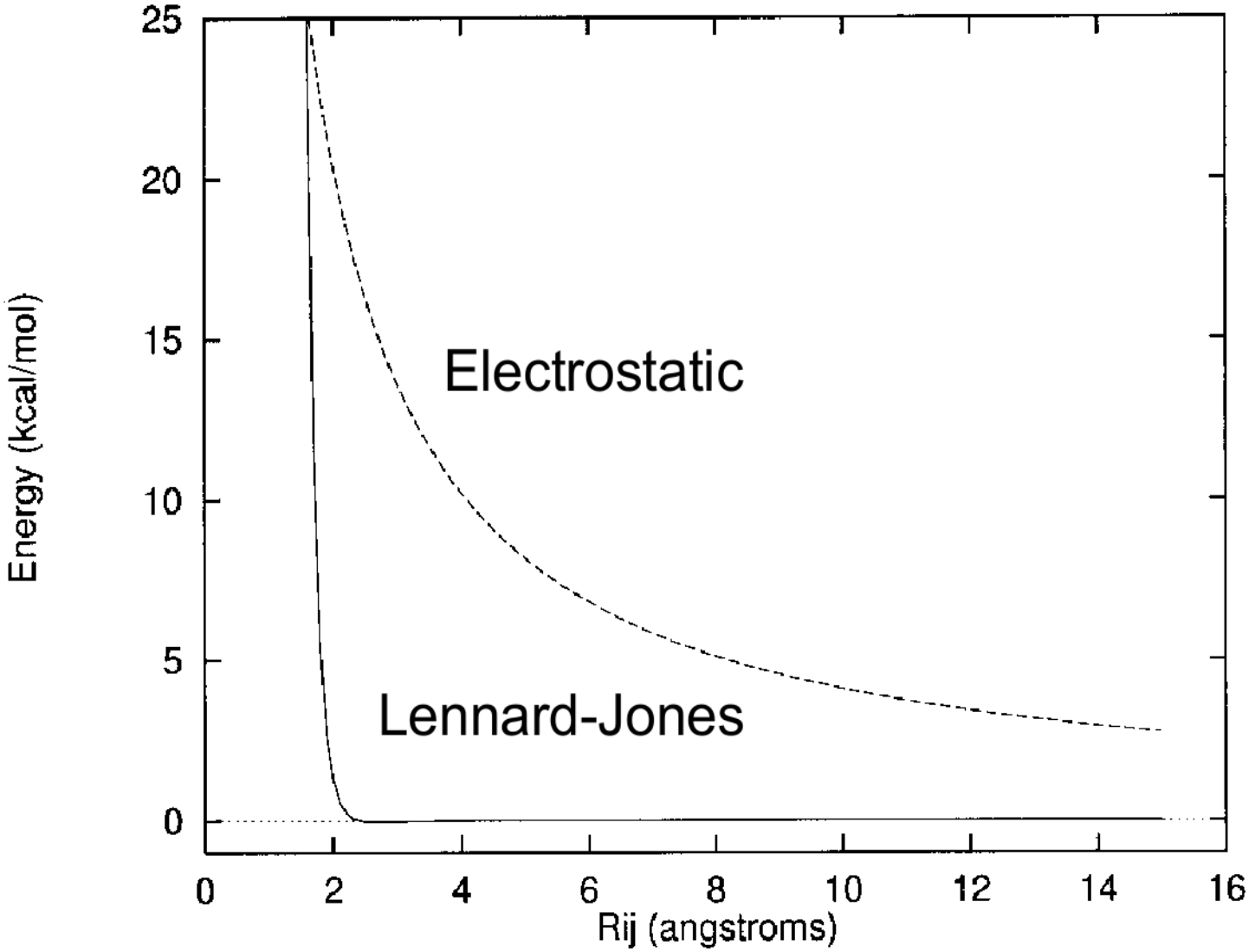
- No explicit electrons, net atomic charges
- No polarization, electron transfer, or correlation
- Conformational energies for ground state
- No chemistry (covalent bond breaking/forming)
- Semi-empirical force field
- Water and counter-ion representations
- ~1000 to 100,000 atoms
- Dynamics up to ~100ns typical now

The AMBER FF

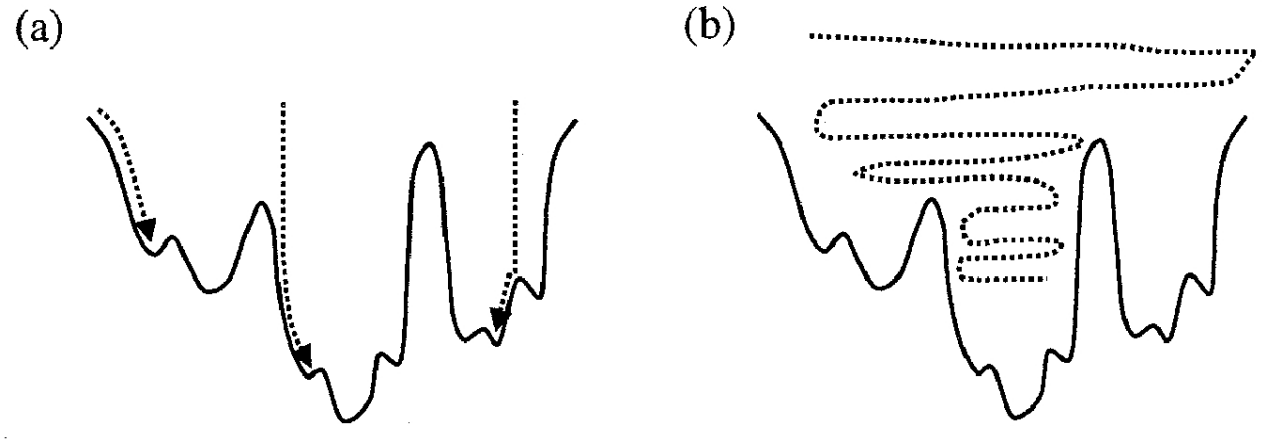
$$U = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{impropers} K_w w^2$$
$$+ \sum_{torsions} K_\phi [1 + \cos(n\phi)] + \sum_{nonbonded\ pairs} \left\{ \left[\frac{B_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r} \right\}$$

- Top line from X-Ray structures, quantum calculations and vibrational spectroscopy.
- Partial charges from fitting electrostatic potential from HF/6-31G*
- van der Waals ε, σ from neat liquids (not water/solute simulations)
- torsional parameters from quantum calculations

Lennard-Jones and Electrostatic Repulsion Behavior



Energy Minimization(a) and Simulated Annealing(b)



- (a) Small changes are only needed during the energy minimization routine. Consequently a computationally expensive force field, such as the AMBER FF, would be adequate given a structure close to its native fold.
- However, we need to search a large conformation space with (b) simulated annealing, when starting a structure determination *de novo*. High temperature ramping is used usually in the SA routine. Unfortunately there are a number of draw backs to the AMBER FF here and simplifications are needed.

AMBER vs CYANA FF

$$U = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{impropers} K_w w^2$$
$$+ \sum_{torsions} K_\phi [1 + \cos(n\phi)] + \sum_{nonbonded\ pairs} \left\{ \left[\frac{B_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r} \right\}$$

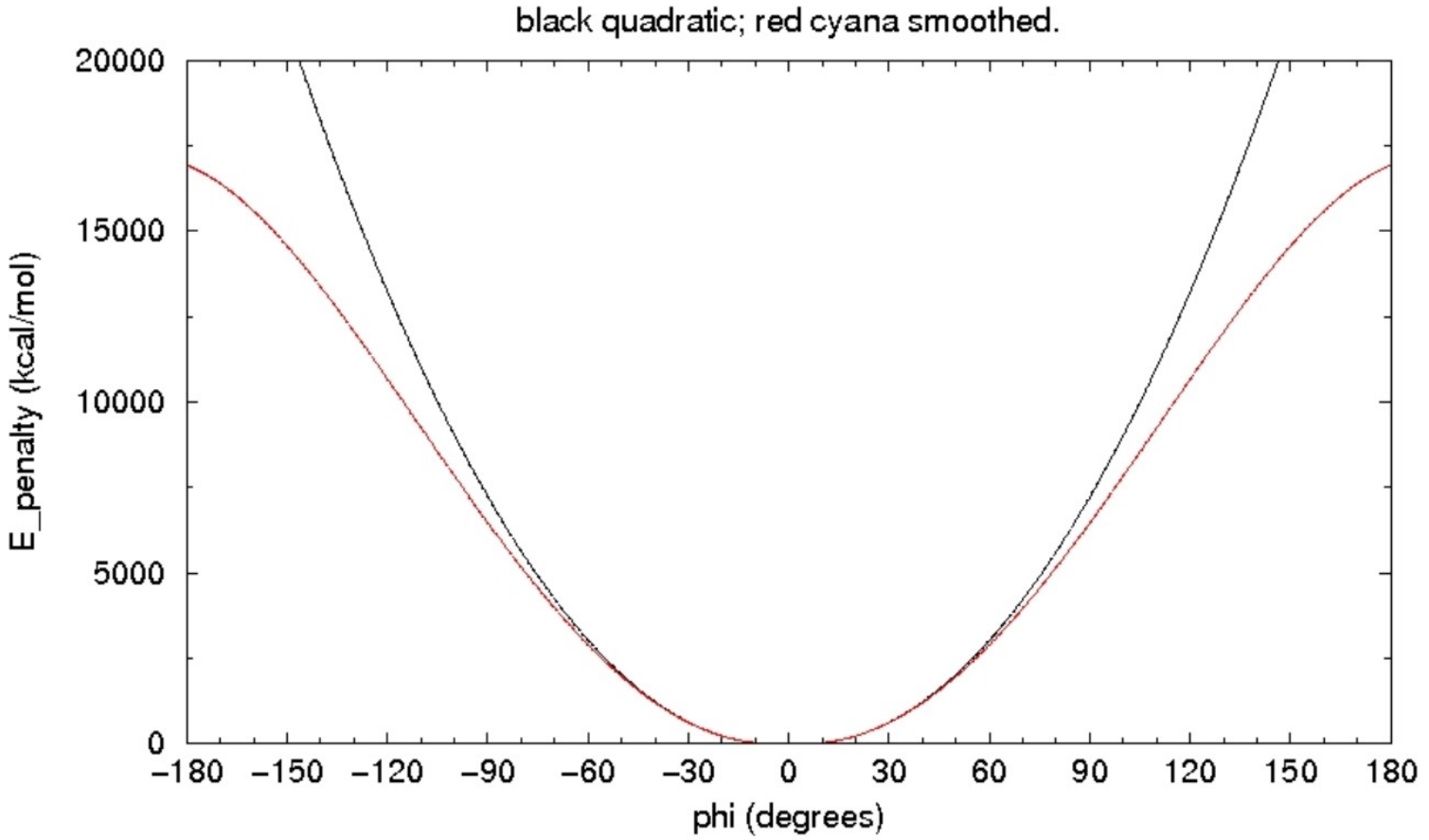
→ CYANA FF

$$V = \sum_{\alpha,\beta} K_d \left(r_\alpha - r_\beta\right)^2 + \sum_{\omega} K_\omega \left(1 - \left(\frac{\delta}{2\,\Gamma}\right)^2\right) * \delta^2$$

K = force constant
r = distance (vdW, NMR expt)

δ = phi/psi violation in degrees
Γ = ½ forbidden phi or psi space

$$V = \sum_{\alpha,\beta} K_d (r_\alpha - r_\beta)^2 + \sum_{\omega} K_\omega \left(1 - \left(\frac{\delta}{2\Gamma}\right)^2\right) * \delta^2$$

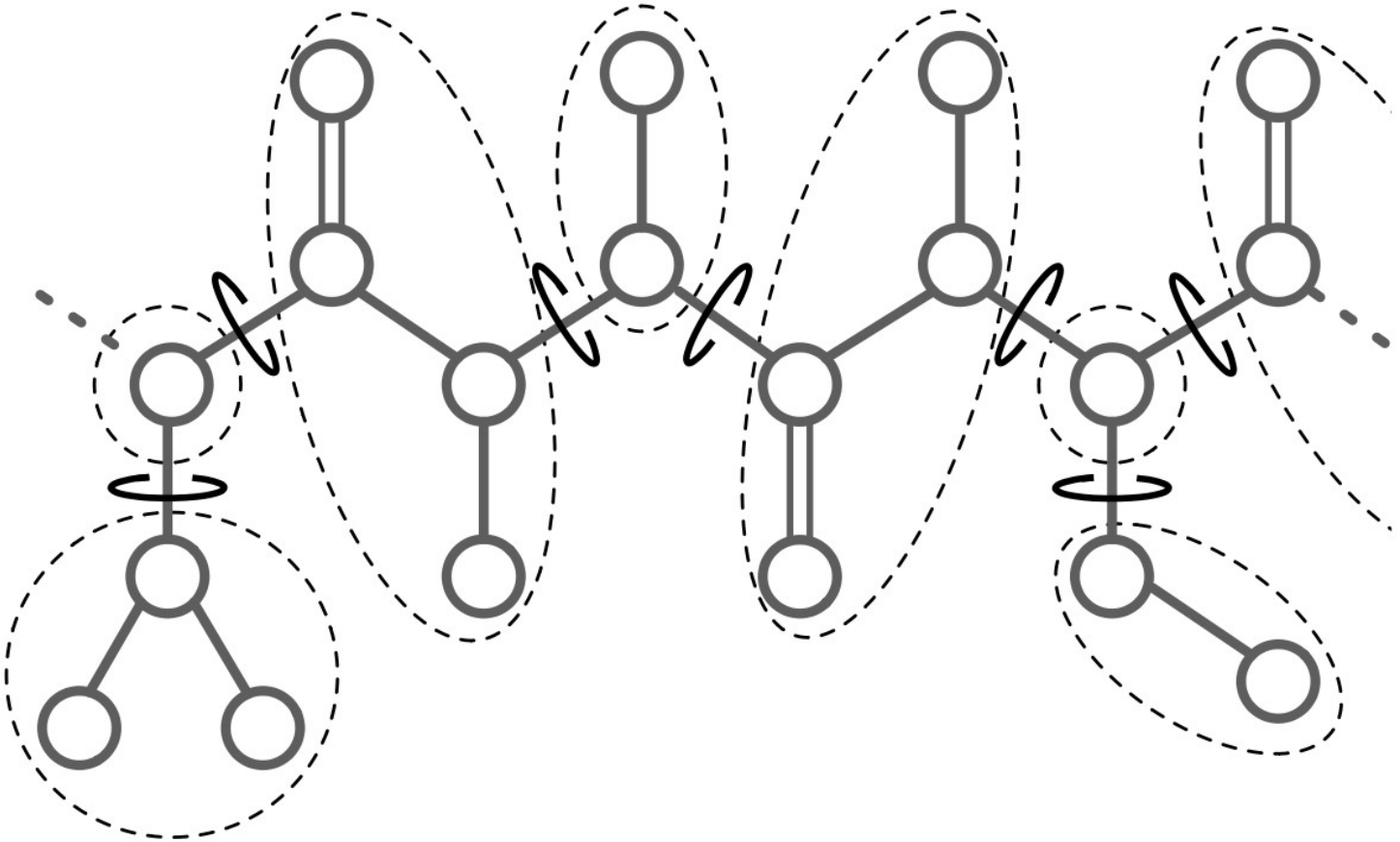


- Both terms above are quadratic in nature actually.
- The angular potential is smoothed so as not too restrictive on phi/psi limits.

Comments on Torsional Angle Dynamics

- In TAD the bond lengths, angles, chiralities and planarities are kept fixed at their optimal values.
- This reduces the number of degrees of freedom by ~10, and allows for significant speedup in calculation.
- During Cartesian simulations (AMBER) a strong force constant is needed to preserve the covalent structure (ie, chirality flipping, bond distortions)
- Strong force constants lead to high frequency modulations especially at elevated temperatures and Cartesian simulations often have velocity limit errors.
- The fixed covalent structures in CYANA allows for large time steps during the MD simulation and isn't plagued with high frequency bond motions during annealing.
- CYANA doesn't represent the system as “realistically” as the full AMBER FF and couldn't be used in traditional MD simulations at room temperature for example with long range electrostatic interactions and vdW considerations.

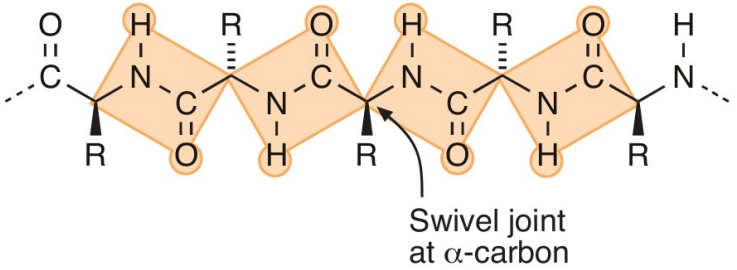
Torsional Angular Dynamics Tree

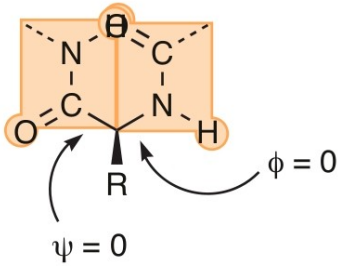


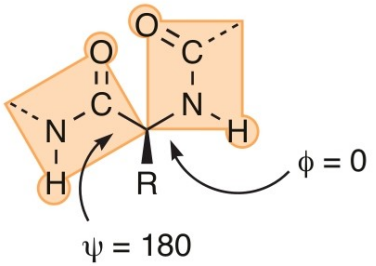
The fixed peptide bonds and side chain elements reduce the degrees of freedom needed in the simulation.

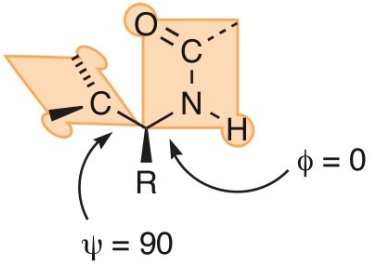
Allowable Torsion Space

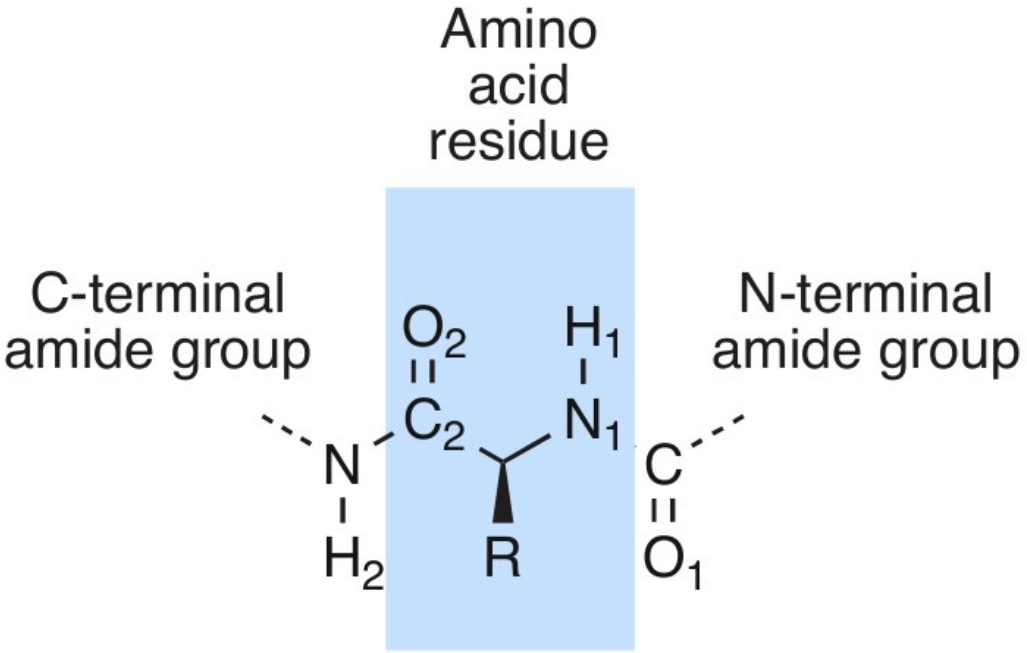
- Covalent geometry and steric clash limits the allowable phi/psi space that a peptide/protein can adopt.







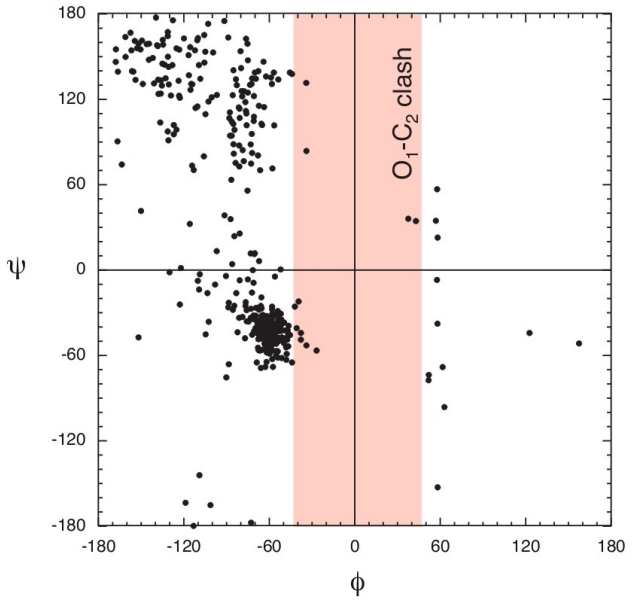




Consider the naming convention to the left. We can vary all phi/psi from -180 to 180 and consider the steric clash encountered by the atoms.

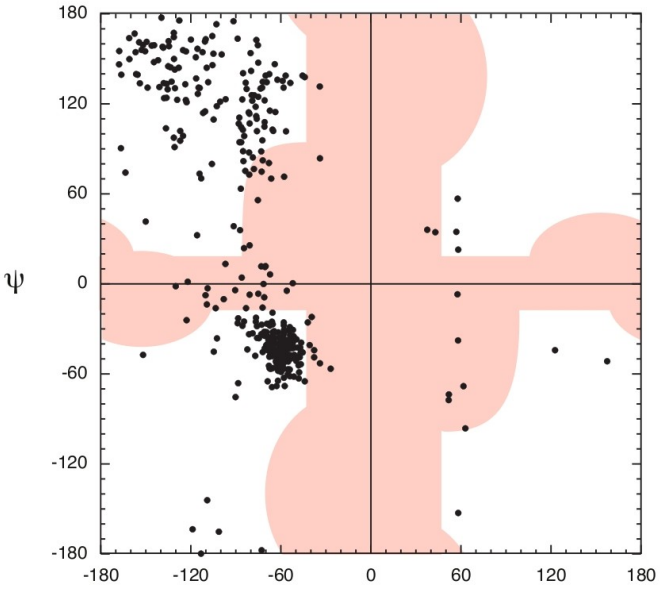
Allowable Torsion Space

O1-C2 clash



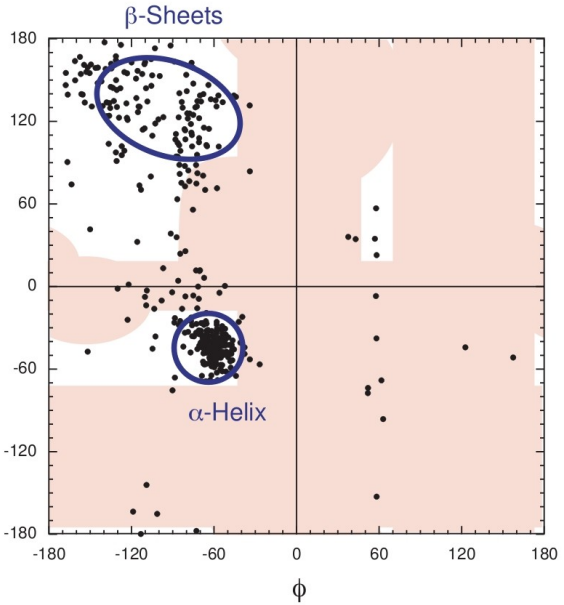
+

All backbone clash



+ etc

=

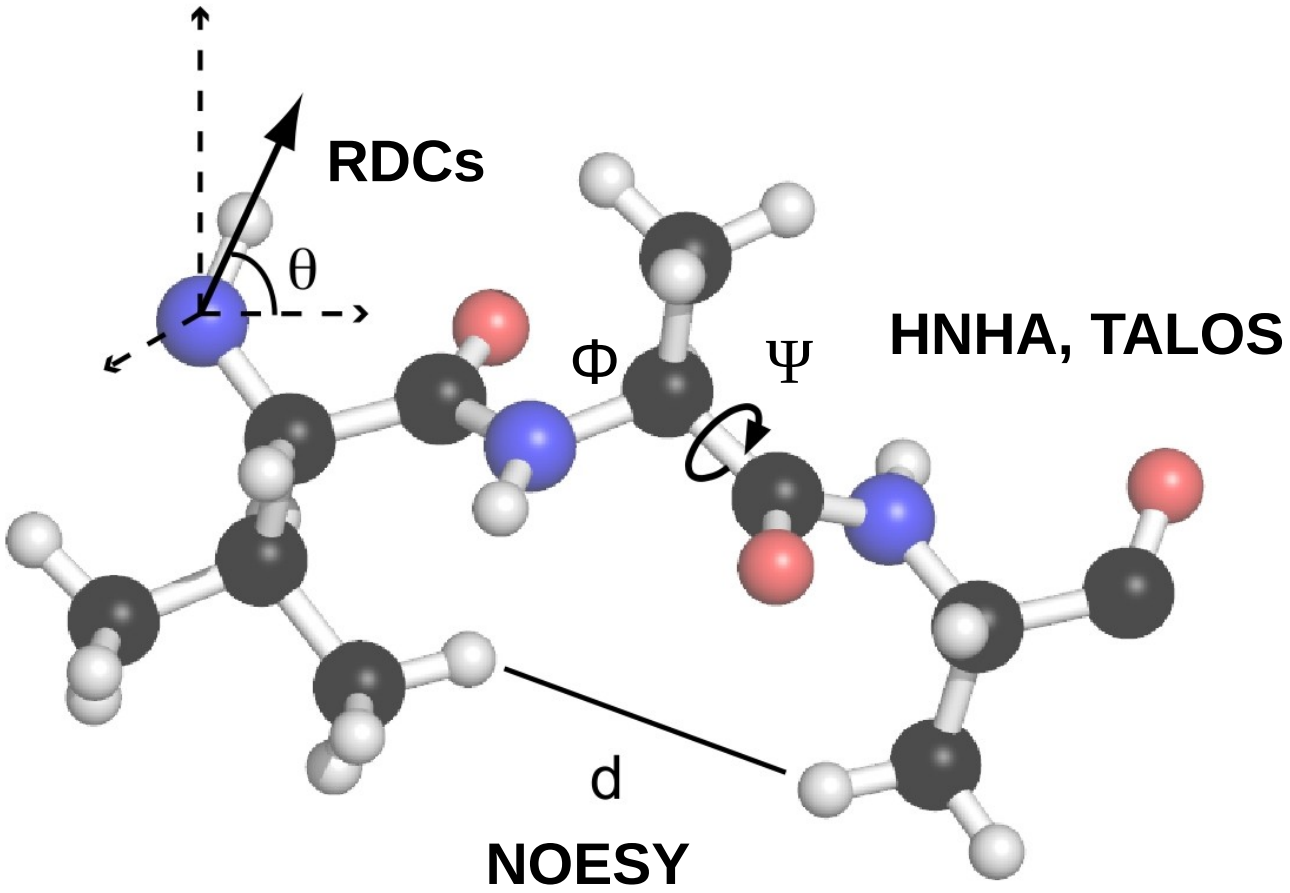


As you can see the space is further limited by the peptide geometry and atom clashes.

The secondary structures fall into particular regions in the plots. Residues which fall out of these allowed regions are often suspect and the restraints should be checked and the structure recalculated.

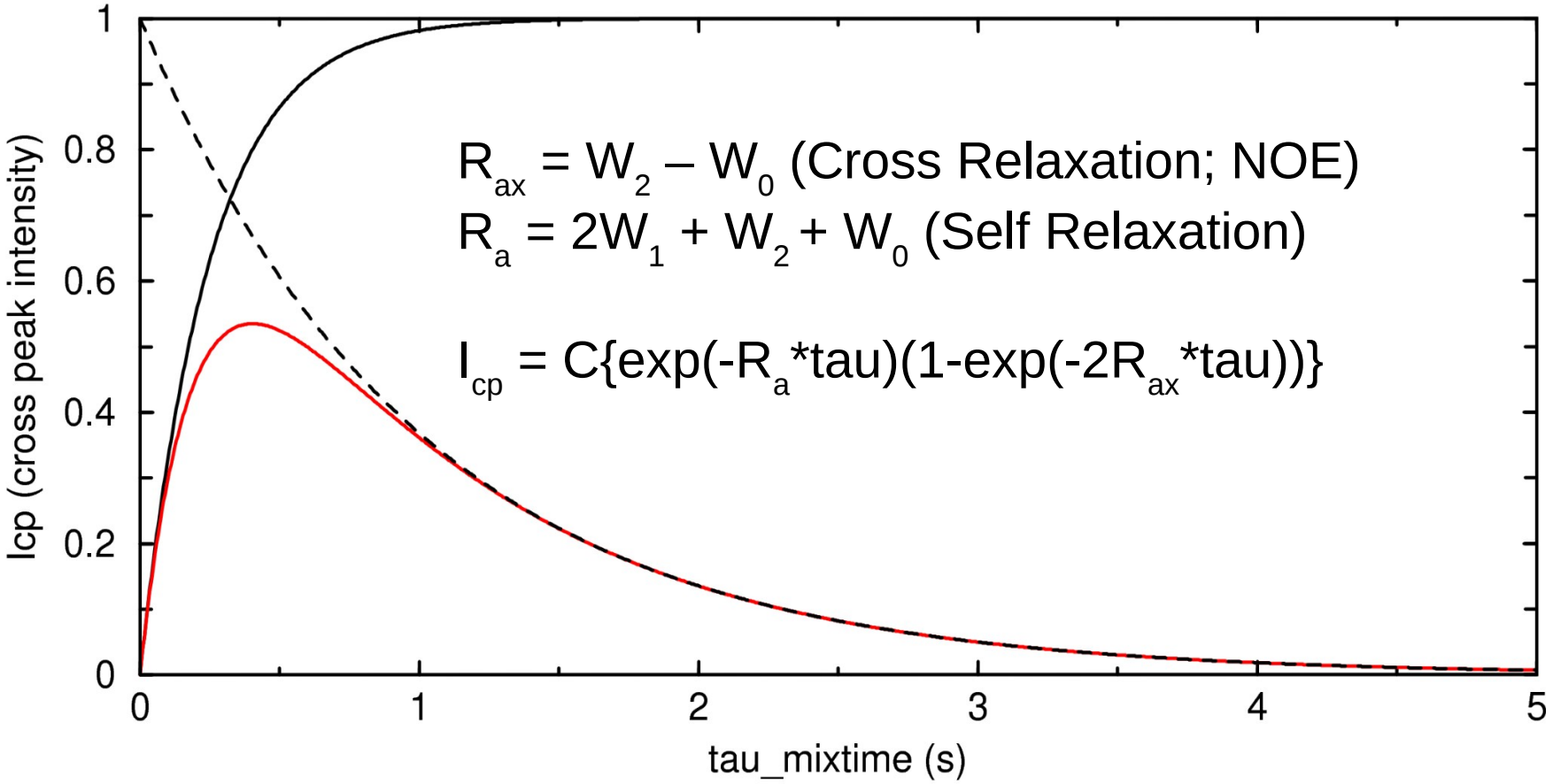
www.greeley.org/~hod/papers/Unsorted/Ramachandran.doc.pdf

Measurable NMR Restraints

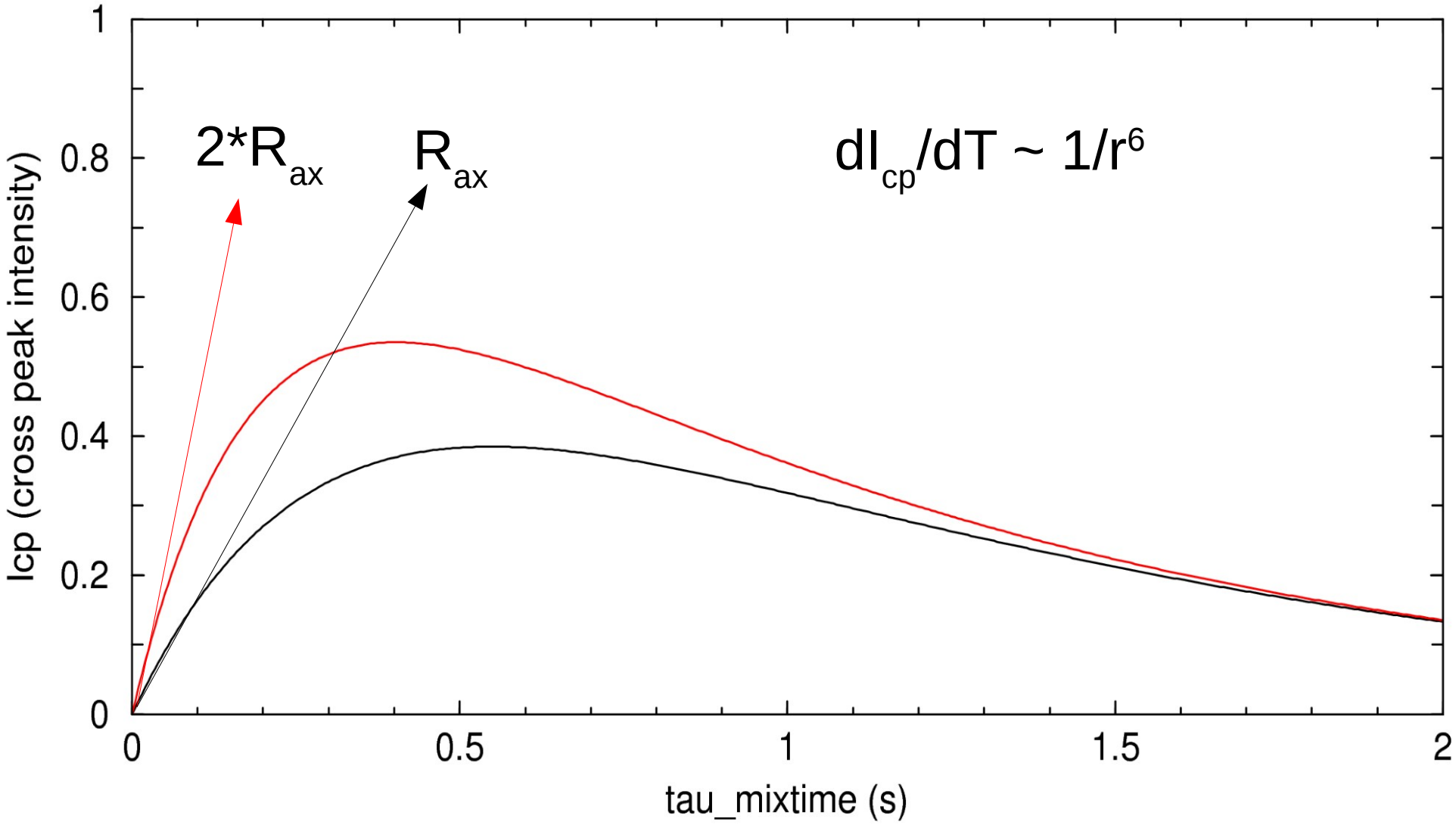


NOE Restraints

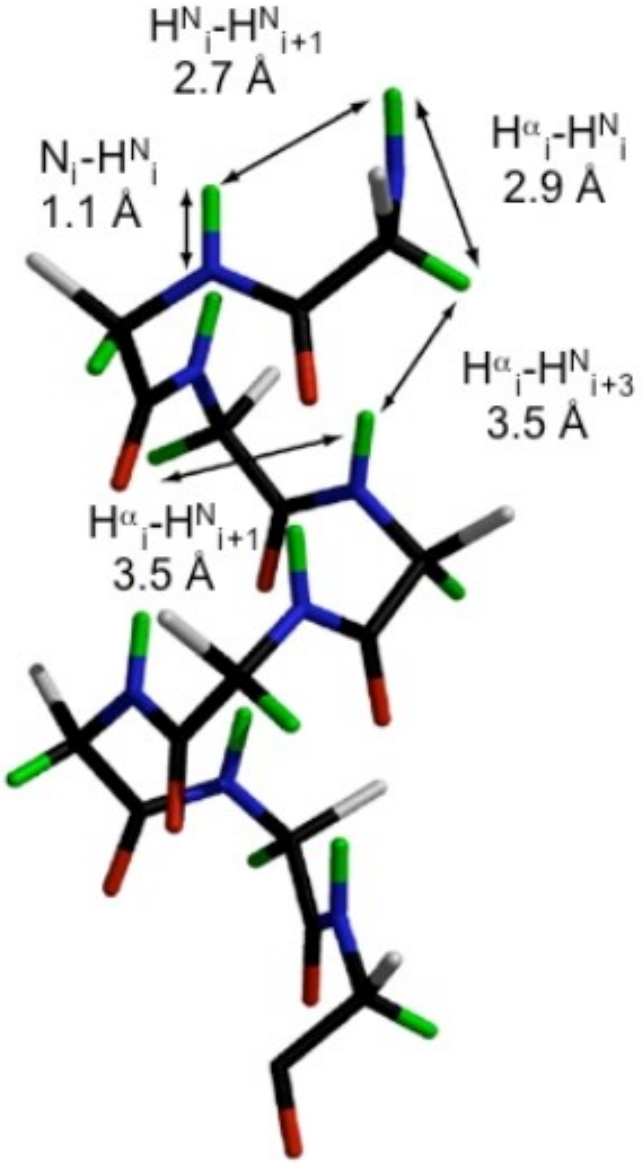
- Experimental distance restraints are key to a successful structure. They are usually obtained from 2D/3D NOESY NMR.
- It's interesting to note that Albert Overhauser developed this theory while a post-doctoral student at the University of Illinois during 1951-1953. He was about 27 years old.



-The initial part of the buildup curve contains distance information.
-At longer times self relaxation takes over and the analysis becomes convoluted.



NOE Restraints

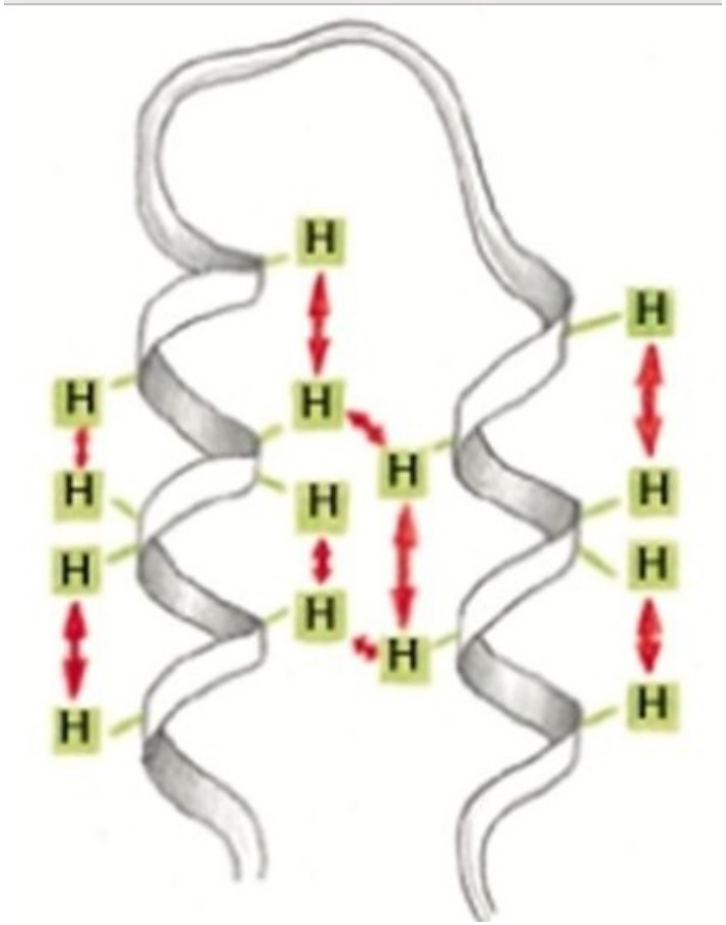


Sequential NOEs can be a finger print for a secondary structure like an alpha helix (on left).

We often “bin” NOE distances into categories:

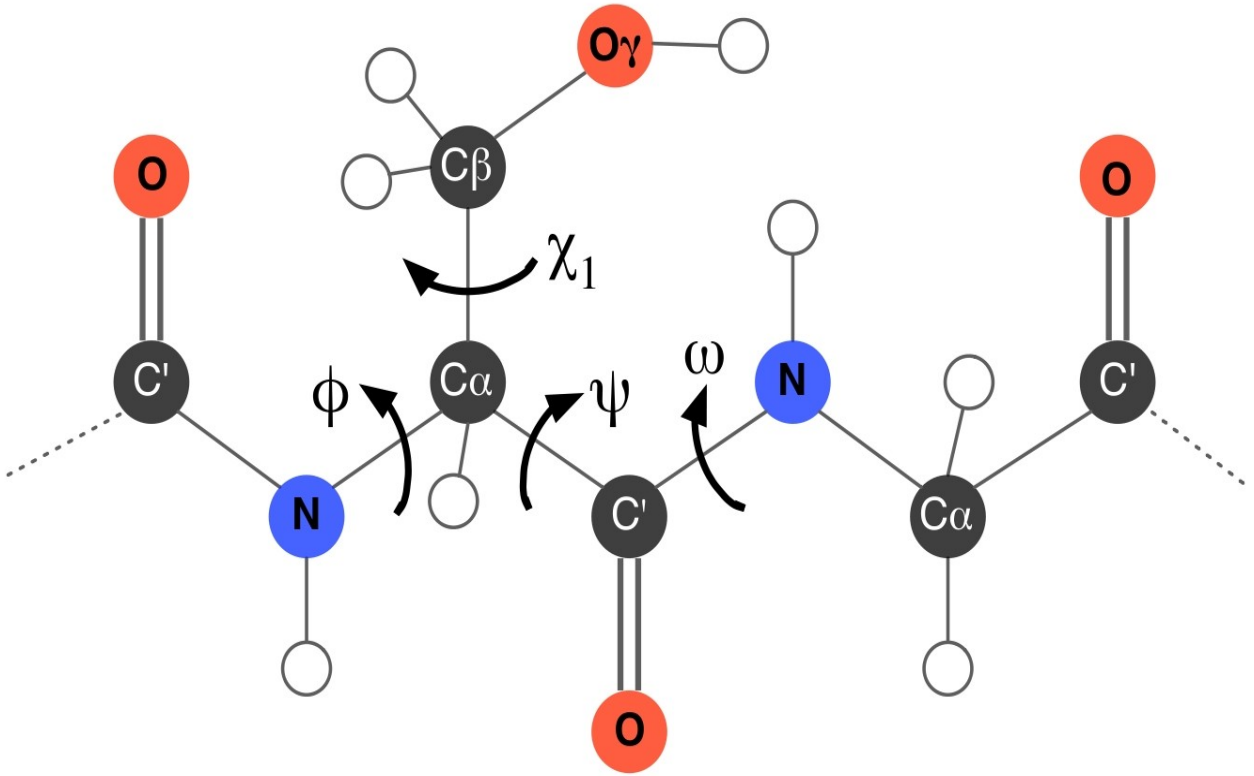
- “strong” 1.8-2.7 Å
- “medium” 1.8-3.3 Å
- “weak” 1.8-5.0 Å
- “very weak” 1.8-6.0 Å
- (lower bound is sum of van der Waals radii for two protons)

NOE Restraints



- Long range NOEs are particularly useful in a structures determination.
- The through space behavior of the NOE allows structural elements to be precisely arranged.
- Long range can be defined as >5 residues apart in the primary sequence.
- It is important that they not be misassigned as they can lead to severe distortions in the protein fold.

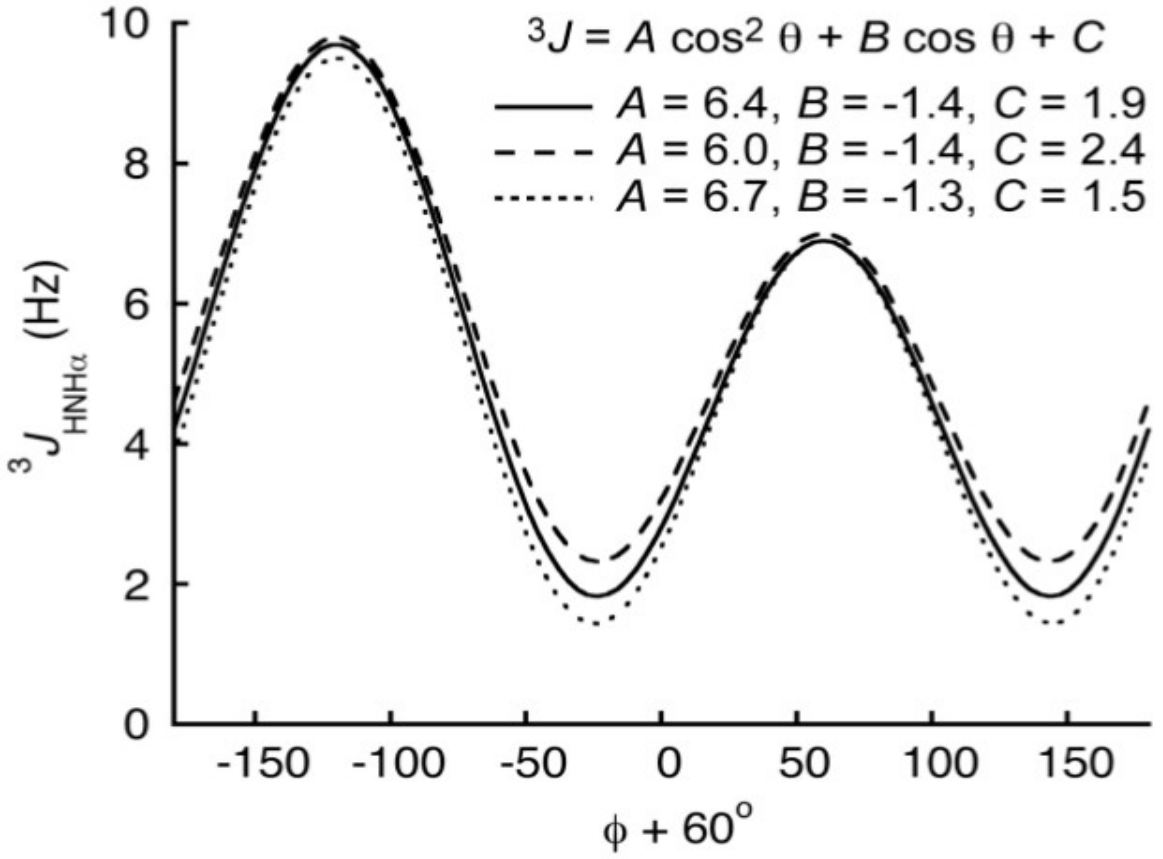
Torsional Restraints



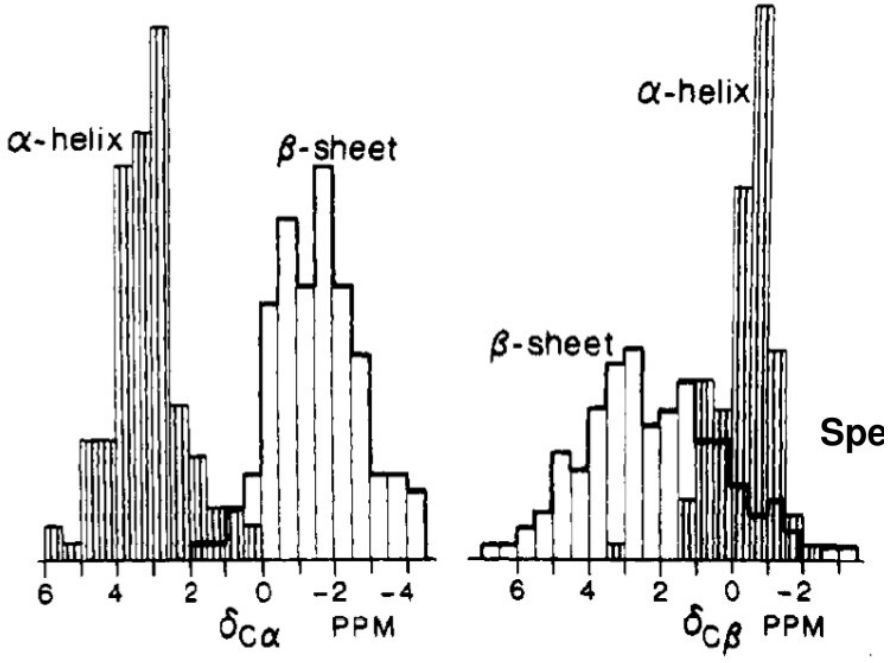
- Local back bone and torsional restraints can be found through J-coupling experiments and chemical shifts. NOEs and RDCs also can restrict the backbone structure.

Torsional Restraints

- Coupling constants can be described by the Karplus equations.
- Values for phi/psi/chi1/chi2 can often be measured directly.
- eg) HNHA is a routine expt for phi determination
- These expts are particularly useful when J>6Hz; Beta sheets



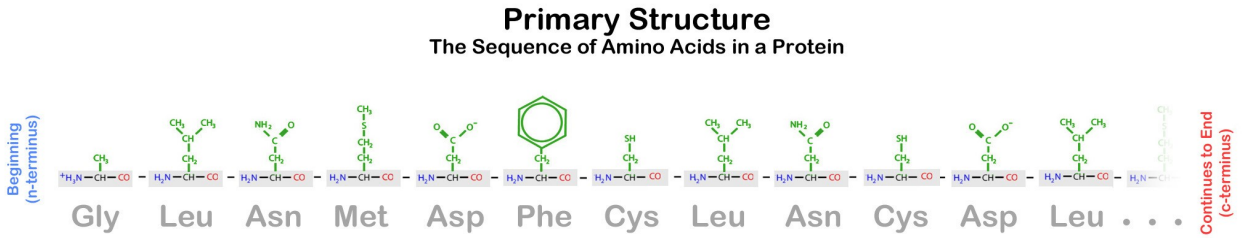
Torsional Restraints



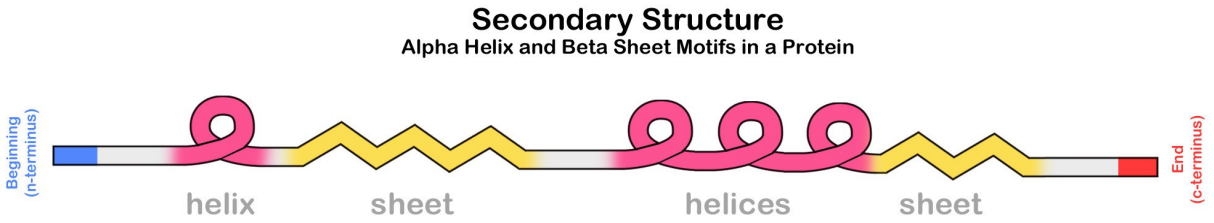
- Observed shift deviations from random coils values provide secondary structure information. This is know as “chemical shift indexing”.
- The TALOS program is almost always used in modern NMR structure determinations. TALOS takes advantage of a data base of observed shifts in the BMRB and elsewhere to help refine phi/psi predictions including errors.
- Recent versions of TALOS also provide side chain CHI1/CHI2 in addition to PHI/PSI.

“TALOS”: Cornilescu et al (1999) J. Bio. NMR, 13, 289-302.

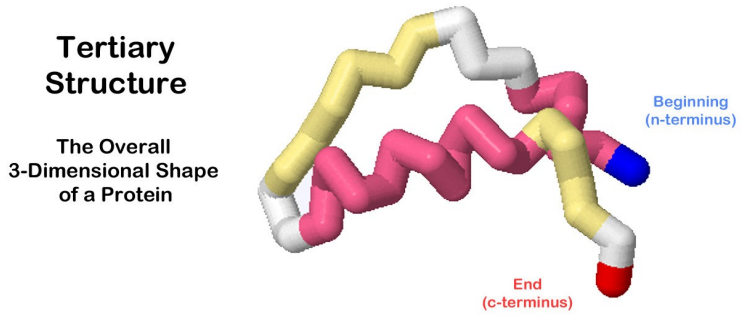
Overview of structure and experimental restraints relations



Mass Spec
Chemical Shifts assign



Chemical Shift Indexing
Short range NOEs
RDCs



Long range NOEs
RDCs

Note: For visual clarity, the R-groups (sidechains) are not shown in the secondary and tertiary structure illustrations.

Brief Summary

CYANA:

- Fixed bond lengths, geometries, chiralities.
- Simplified FF (actually called 'target potential')
- Not suitable for standard MD runs (ie no restraints included)
- vdW lower bounds included as a distance restraint (internally)

Needs restraint inputs:

- Distance (NOE)
- Torsional (TALOS, 3J-modulated type expts, RDC's)
- SSNMR restraints also feasible (13C-13C, 13C-1H, 13C-15N distances)
- Lower bound repulsive terms (particular contact not seen in expt)

CYANA File Naming

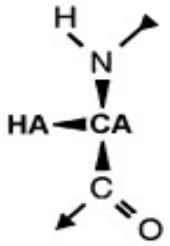
- A table of the naming convention used
- It is important to use these extensions so as the program can keep track of whats in the files. eg) data.aco would be seen as torsion data

file type	format	default extension
dihedral angle constraints	DIANA	.aco
dihedral angles	DIANA	.ang
coupling constants	HABAS	.cco
Cartesian coordinates	DG	.cor
residue library	DYANA	.lib
lower limit distance constraints	DIANA	.lol
orientation constraints	DYANA	.ori
Cartesian coordinates	PDB	.pdb
peak list	XEASY	.peaks
chemical shift list	XEASY	.prot
residue sequence	DIANA	.seq
upper limit distance constraints	DIANA	.upl
XPLOR distance and angle constraints	XPLOR	.xplor

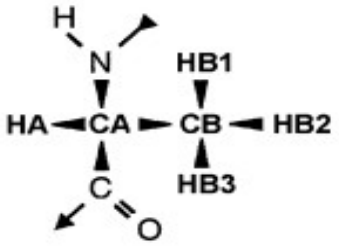
Formats of most files are those of already existing programs: DIANA

CYANA 2.1 atom names (IUPAC nomenclature)

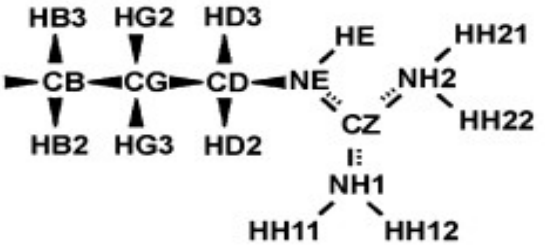
Backbone Atoms



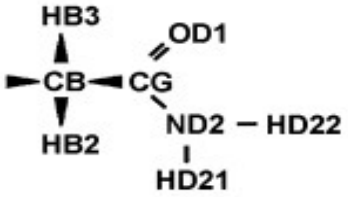
Ala



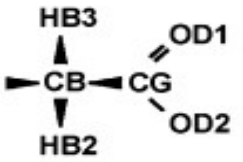
Arg



Asn

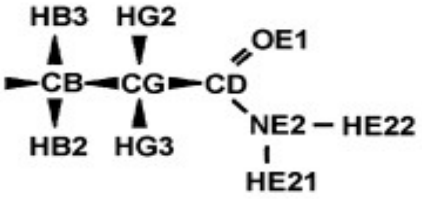


Asp

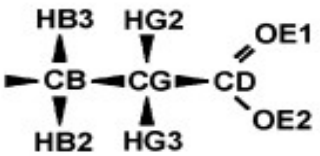


Cys

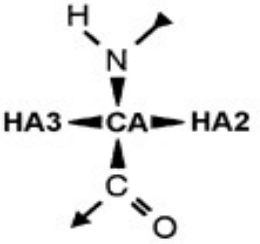
Gln



Glu

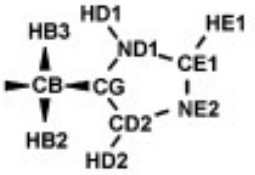


Gly

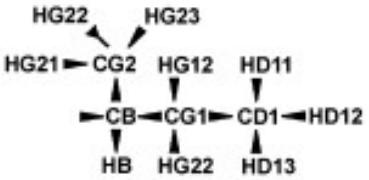


You'll often need to refer back to the atom names and make corrections if an error pops up during program initiation.

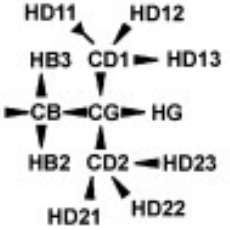
His



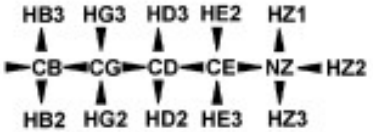
Ile



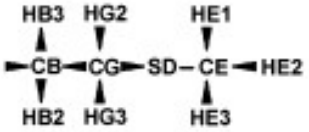
Leu



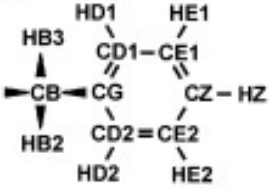
Lys



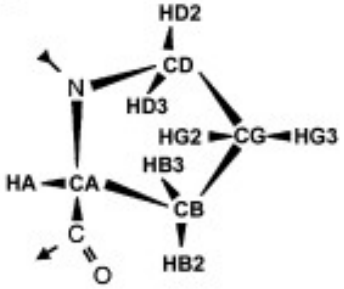
Met



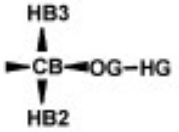
Phe



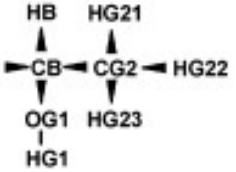
Pro



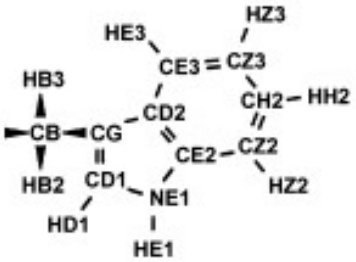
Ser



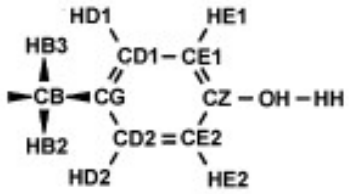
Thr



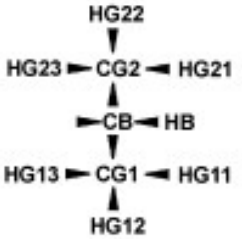
Trp



Tyr



Val



Degenerate Atoms

- A group of degenerate protons is represented in the chemical shift list by a “pseudo atom” (or the corresponding “ambiguity index” in BMRB files)
- CYANA expands distance constraints to pseudo atoms into ambiguous distance constraints with all the corresponding protons represented by the pseudo atom(s): “1/r6-summation”.(Fletcher et al., J. Biomol. NMR 8, 292 (1996))

  - Degenerate pairs of methylene protons:
    QB ($H^{\beta 2}/H^{\beta 3}$), …
  - Methyl groups:
    QB (Ala), QG1/QG2 (Val), QD1/QD2 (Leu), …
  - Degenerate pairs of methyl groups:
    QQG (Val), QQD (Leu)
  - Phe/Tyr aromatic ring protons:
    QD ($H^{\delta 1}/H^{\delta 2}$), QE, ($H^{\varepsilon 1}/H^{\varepsilon 2}$), QR (all ring protons)

Diastereotopic Protons

- Stereospecifically assigned:
  - 2 entries in the chemical shift list, e.g.: `HB2`/`HB3`
  - CYANA command to supress swapping:
    `atom stereo "HB2 23"`
- Not stereospecifically assigned, not degenerate:
  - 2 entries in the chemical shift list: `HB2`/`HB3`
  - During the calculation CYANA will periodically check for the optimal stereo-assignment and swap the two protons, if needed.
    (Folmer et al., *J. Biomol. NMR* **9**, 245 (1997))
- Degenerate:
  - 1 pseudo atom entry in the chemical shift list: `QB`

Input file formats:

R – 1st one

.seq

```
GLY 11 SER ILE PRO CYSS LEU LEU SER cPRO TRP SER GLU TRP CYSS
```

C E **.prot** R

```
 1 999.000   0.000 N       1
 2 999.000   0.000 H       1
 3 999.000   0.000 CA      1
 4 999.000   0.000 HA2     1
 5 999.000   0.000 HA3     1
 6 999.000   0.000 QA      1
 7 999.000   0.000 C       1
 8 999.000   0.000 O       1
 9 999.000   0.000 N       2
10 999.000   0.000 H       2
11 999.000   0.000 CA      2
12 999.000   0.000 HA      2
13 999.000   0.000 CB      2
14 999.000   0.000 HB2     2
15 999.000   0.000 HB3     2
16 999.000   0.000 QB      2
17 999.000   0.000 OG      2
18 999.000   0.000 HG      2
19 999.000   0.000 C       2
20 999.000   0.000 O       2
```

.aco

```
# column 1: residue number
# column 2: amino acid
# column 3: angle identifier
# column 4: the angle's lower limit
# column 5: the angle's upper limit
```

R L U

```
  6 SER   PHI    -200.0   -80.0
  6 SER   PSI      40.0   220.0
 10 GLN   PHI    -200.0   -80.0
 10 GLN   PSI      40.0   220.0
 11 LYS   PHI    -200.0   -80.0
 11 LYS   PSI      40.0   220.0
 15 GLU   PHI    -200.0   -80.0
 15 GLU   PSI      40.0   220.0
 16 GLN   PHI    -200.0   -80.0
 16 GLN   PSI      40.0   220.0
 18 TRP   PHI    -200.0   -80.0
 18 TRP   PSI      40.0   220.0
```

R = residue number
L = lower bound
U = upper bound
C = chemical shift
E = error

.lol

```
# Lower distance limit
```

L

R R

```
 81 ILE  O    85 ILE  H      1.80
 81 ILE  O    85 ILE  N      2.70
 82 PRO  O    86 ASP  H      1.80
 82 PRO  O    86 ASP  N      2.70
 83 LEU  O    87 HIS  H      1.80
 83 LEU  O    87 HIS  N      2.70
 84 LEU  O    88 LEU  H      1.80
 84 LEU  O    88 LEU  N      2.70
 85 ILE  O    89 LEU  H      1.80
 85 ILE  O    89 LEU  N      2.70
 86 ASP  O    90 SER  H      1.80
 86 ASP  O    90 SER  N      2.70
```

.upl

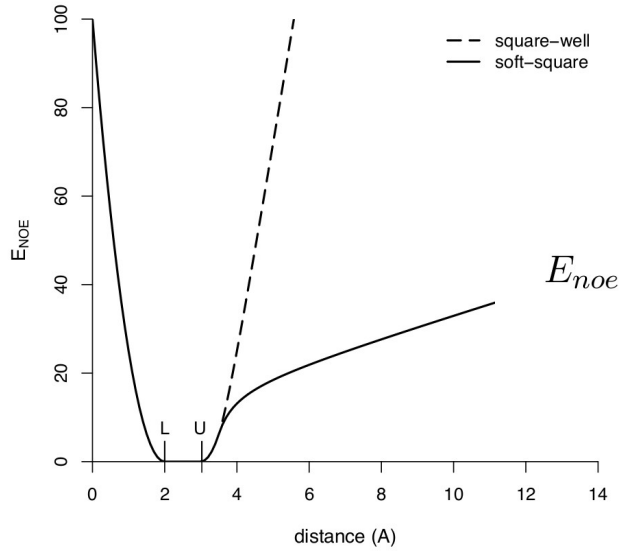
```
# Upper distance limit
```

U

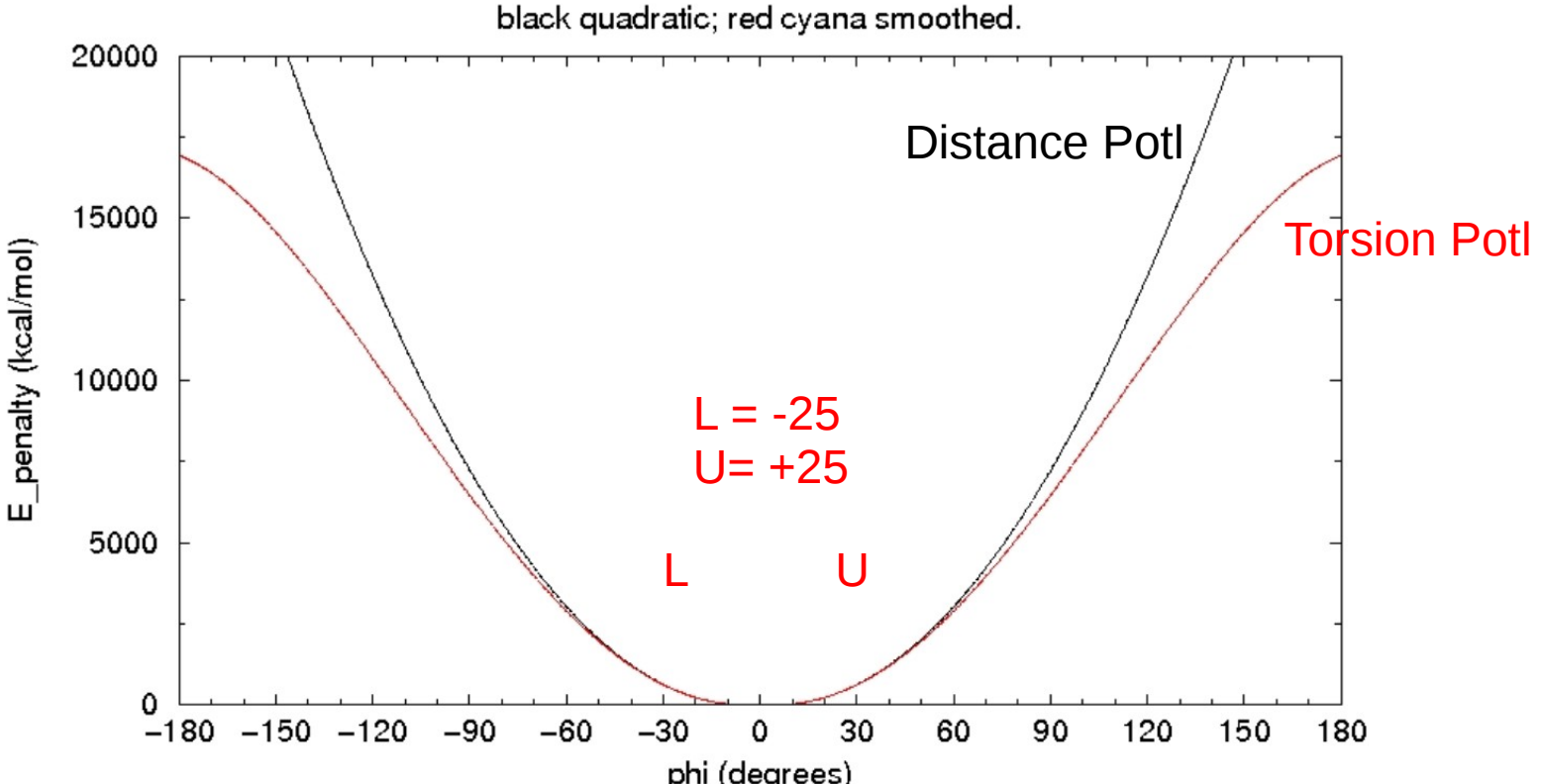
R R

```
 91 THR  HB   93 GLN  H      5.50
 81 ILE  HB   82 PRO  HA     5.19
 80 SER  HB3  82 PRO  HD2    3.86
 27 GLU  HB2  28 VAL  HA     5.50
 28 VAL  HA   32 LEU  QD1    5.50
 81 ILE  HA   84 LEU  HB3    3.74
 81 ILE  HA   81 ILE  QG2    3.46
 81 ILE  HA   81 ILE  HG12   3.77
 81 ILE  HA   81 ILE  HG13   3.77
 28 VAL  HA   31 LEU  HB2    4.12
 28 VAL  HA   39 LEU  QD1    4.59
 28 VAL  HA   31 LEU  QD1    3.38
 28 VAL  HA   39 LEU  QD2    4.93
 28 VAL  HA   31 LEU  HB3    5.30
```

Upper and lower limit bounds



$$E_{noe} = \begin{cases} (L-d)^2 & \text{if d} < \text{L} \\ 0 & \text{if L} \leq \text{d} \leq \text{U} \\ (d-U)^2 & \text{if U} < \text{d} < \text{U+}s \\ a + \frac{b}{(d-U)} + c(d-U) & \text{if d} > \text{U+}s \end{cases}$$



- NMRView NOE Peak List (.xxpk)

Pk box
Width/height

Res assign

Volume

Intensity

```
label dataset sw sf
HC H C
aro.nv
{9602.14 } {9613.04 } {7945.97 }
{800.0340 } {800.0340 } {201.1810 }
HC.L  HC.P  HC.W  HC.B  H.L  H.P  H.W  H.B  C.L  C.P  C.W  C.B  vol  int stat flag0
     0 {?}   7.815  0.041  0.177 {?}   7.815  0.061  0.157 {?} 136.593  0.886  1.080    3218.654    -52.635  0  0
     1 {?}   7.816  0.105  0.122 {?}   0.947  0.110  0.132 {?} 136.606  0.865  1.080     335.673     -2.701  0  0
     2 {?}   7.691  0.034  0.191 {?}   7.690  0.071  0.152 {?} 136.552  1.080  1.080    6282.021   -103.746  0  0
     3 {?}   7.690  0.058  0.058 {?}   2.260  0.120  0.120 {?} 136.538  0.107  0.107      20.434     -1.095  0  0
     4 {?}   7.689  0.082  0.082 {?}   1.445  0.120  0.120 {?} 136.551  0.712  0.712     141.973     -1.573  0  0
     5 {?}   7.755  0.039  0.227 {?}   7.754  0.068  0.150 {?} 136.102  0.921  1.080    6278.629    -97.126  0  0
     6 {?}   7.755  0.092  0.086 {?}   0.289  0.126  0.126 {?} 136.106  0.739  0.646      78.427     -1.582  0  0
     7 {?}   8.002  0.037  0.154 {?}   8.001  0.065  0.146 {?} 135.730  0.864  1.080    4813.683    -84.831  0  0
     8 {?}   8.002  0.102  0.094 {?}   0.494  0.171  0.157 {?} 135.726  0.837  1.080     191.780     -1.947  0  0
     9 {?}   7.247  0.043  0.162 {?}   7.246  0.062  0.153 {?} 134.222  1.080  1.080    3214.984    -40.696  0  0
    10 {?}   7.246  0.080  0.111 {?}   3.601  0.131  0.175 {?} 134.199  1.080  1.080     525.849     -3.608  0  0
    11 {?}   7.246  0.108  0.134 {?}   0.949  0.114  0.145 {?} 134.212  1.080  1.080     501.951     -2.921  0  0
    12 {?}   7.247  0.136  0.136 {?}   0.801  0.091  0.104 {?} 134.199  1.080  1.080     360.399     -2.160  0  0
    13 {?}   7.245  0.095  0.124 {?}   0.697  0.114  0.146 {?} 134.206  0.918  1.080     432.267     -3.234  0  0
    14 {?}   7.246  0.104  0.104 {?}   0.009  0.145  0.145 {?} 134.209  0.490  0.490     102.112     -1.302  0  0
    15 {?}   6.936  0.062  0.062 {?}   8.859  0.121  0.121 {?} 131.585  0.368  0.368      64.327     -1.183  0  0
    16 {?}   6.937  0.060  0.060 {?}   7.446  0.060  0.047 {?} 131.549  0.762  1.080      58.763     -1.631  0  0
    17 {?}   6.936  0.040  0.200 {?}   6.936  0.075  0.162 {?} 131.527  1.080  1.080    4477.247    -57.101  0  0
```

- These are cross-peaks in the NOESY spectrum.
- Their intensities/volumes are used as distance information

- XEASY Format
- ANSIG Format

Chemical Shifts and Peak Lists

- Several NOE expts are usually acquired each with their own peak list. eg) 13C or 15N filtered NOESY, Aromatic NOESY, 80ms mix NOESY, 300ms mix NOESY.
- NOE_C.xpk, NOE_N.xpk, NOE_Aro.xpk, NOE_80ms.xpk etc
- Invariably there are slight chemical shift differences between each expt.
- CYANA handles this by allowing an adjustable tolerance criteria:

  1H +/- 0.03 ppm; 13C +/- 0.5 ppm 15N +/- 0.5 ppm

  to be defined.
- Remember everything is referenced to the assignments in **.prot** file
- A trick I like to use is to transfer the assignments seen in the long NOESY expt (300ms) and use those shifts in the .prot file. This can be done by slightly adjusting the .xpk files in NMRView.

CYANA Master Files

init.cya

```
rmsdrange:=10-100                          # define residue range for RMSD calculations
cyanalib                                   # read standard CYANA residue library
read seq demo.seq                          # read amino acid sequence
```

CALC.cya

```
peaks       := c13.peaks,n15.peaks,aro.peaks  # NOESY peak lists in XEASY format
#peaks       := c13.xpk,n15.xpk,aro.xpk     # alternative peak lists in NMRView format
prot        := demo.prot                    # names of chemical shift lists
restraints  := demo.aco                     # additional (non-NOE) restraints
tolerance   := 0.040,0.030,0.45             # shift tolerances: H, H', C/N', C/N
#calibration_constant:=6.7E5,8.2E5,8.0E4   # calibration constants, automatic if empty
structures  := 100,20                       # number of initial, final structures
steps       := 10000                        # number of torsion angle dynamics steps
randomseed  := 434726                       # random number generator seed

noeassign peaks=$peaks prot=$prot autoaco  # perform NOESY assignment/structure calculation
```

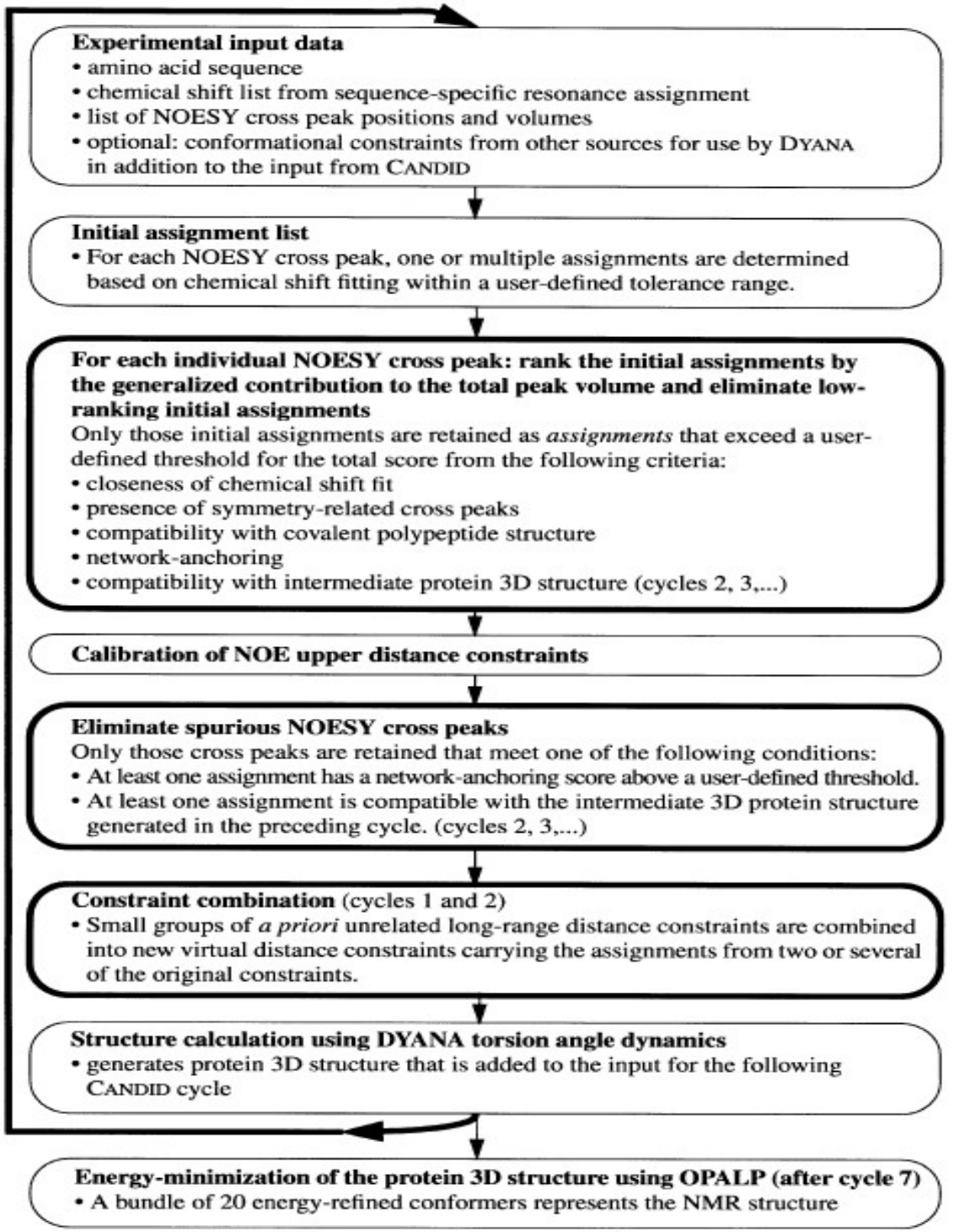
To launch CYANA on command line:

cyana < CALC.cya

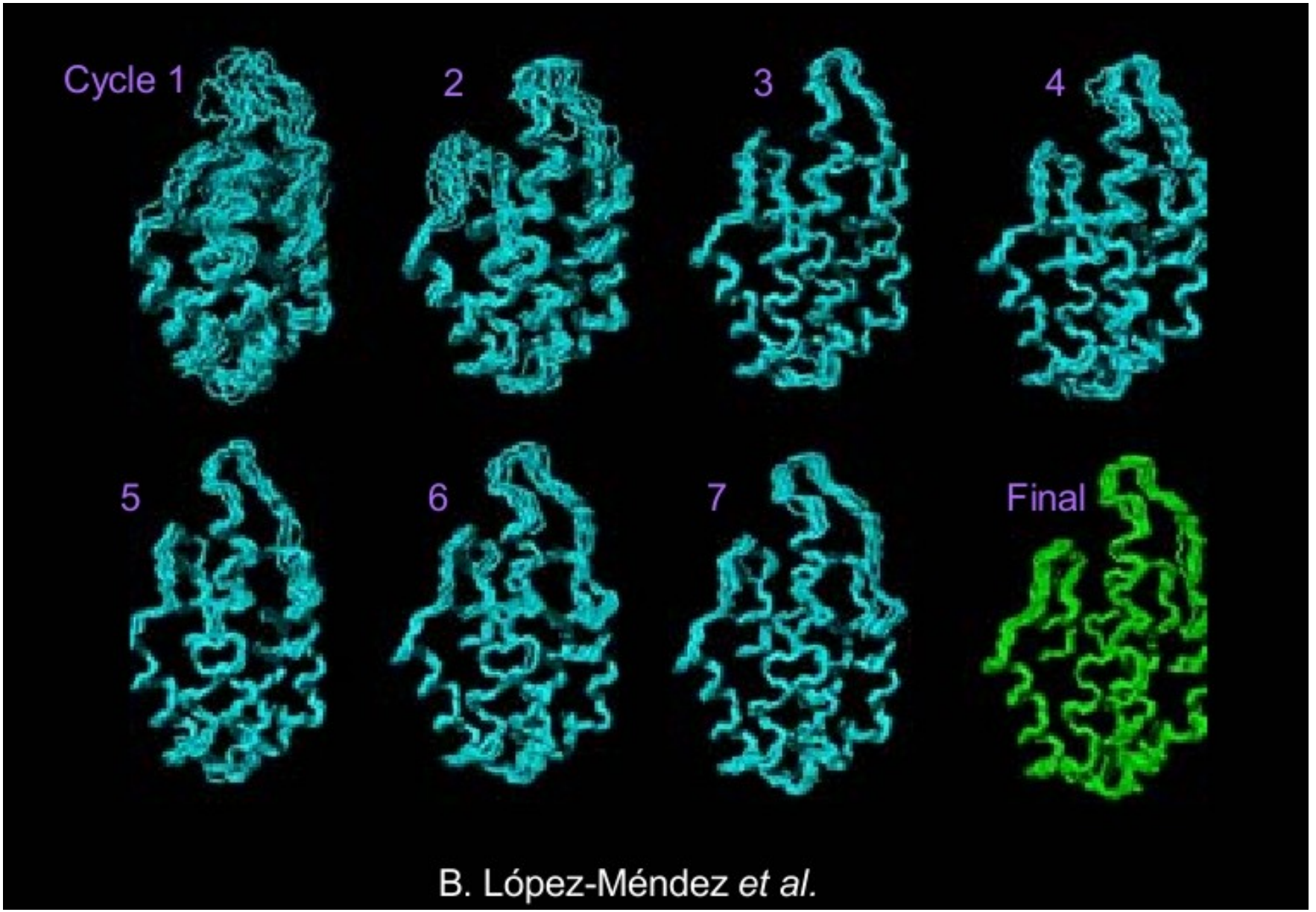
< = pipe in the CALC.cya file into CYANA program

This flow chart outlines what happens during the seven cycles of assignment and annealing of structures.

As you can see there are some additional NOE assignment filtering tricks that we need to consider like 'network anchoring' etc.



Progression of refinement for each cycle



Output files generated into working folder

- From each cycle (*N* = 1,…,7):
  - `cycle`*N*`.noa` NOE assignment details for each peak
  - `cycle`*N*`.upl` NOE upper distance limits
  - `cycle`*N*`.cor` Bundle of conformers
  - `cycle`*N*`.ovw` Target function/violation overview
- In addition for the last cycle (cycle 7)
  - *peaklist*`-cycle7.peaks`
    NOESY peak lists with assignments from CYANA
    (name of input peak list: *peaklist*`.peaks`)
- From the final structure calculation:
  - `final.upl` Final NOE upper distance limits
  - `final.cor` Final bundle of conformers
  - `final.ovw` Target function/violation overview

The .ovw file:

Cycle	1	2	3	4	5	6	7	final
Peaks:								
selected	5439	5439	5439	5439	5439	5439	5439	
with assignment	5100	4806	4742	4749	4712	4678	4675	
without assignment	339	633	697	690	727	761	764	
with diagonal assignment	12	12	12	12	12	12	12	
Cross peaks:								
with off-diagonal assignment	5088	4794	4730	4737	4700	4666	4663	
with unique assignment	675	3591	3872	3950	4115	4195	4194	
with short-range assignment |i-j|<=1	3295	3208	3165	3154	3120	3102	3089	
with medium-range assignment 1<|i-j|<5	1020	925	921	914	904	884	893	
with long-range assignment |i-j|>=5	773	661	644	669	676	680	681	
Upper distance limits:								
total	3786	2996	2832	2789	2707	2643	2683	2731
short-range, |i-j|<=1	2007	1586	1486	1440	1388	1348	1273	1304
medium-range, 1<|i-j|<5	1220	959	787	775	751	726	760	765
long-range, |i-j|>=5	559	451	559	574	568	569	650	662
Average assignments/constraint	4.81	1.73	1.27	1.25	1.18	1.14	1.00	1.00
Average target function value	230.84	69.79	68.20	9.22	3.99	2.98	1.70	0.43
RMSD (residues 15..130):								
Average backbone RMSD to mean	1.34	0.97	0.57	0.67	0.68	0.60	0.53	0.53
Average heavy atom RMSD to mean	1.76	1.44	1.09	1.19	1.20	1.07	0.98	1.01

Recent advances in NOE Assignment Automation

- It is often the case that NOESY cross-peaks can be simultaneously assigned to several different atom pairs. These are know as “ambiguous restraints” . Wrong assignments leads to a distorted/incorrect structure and are best left out.
- ARIA treats each NOESY peak as a superposition of all peaks with the same assignment according to:

$$\bar{D} = \left(\sum_{c=1}^{N_c} d_c^{-6} \right)^{-\frac{1}{6}}$$

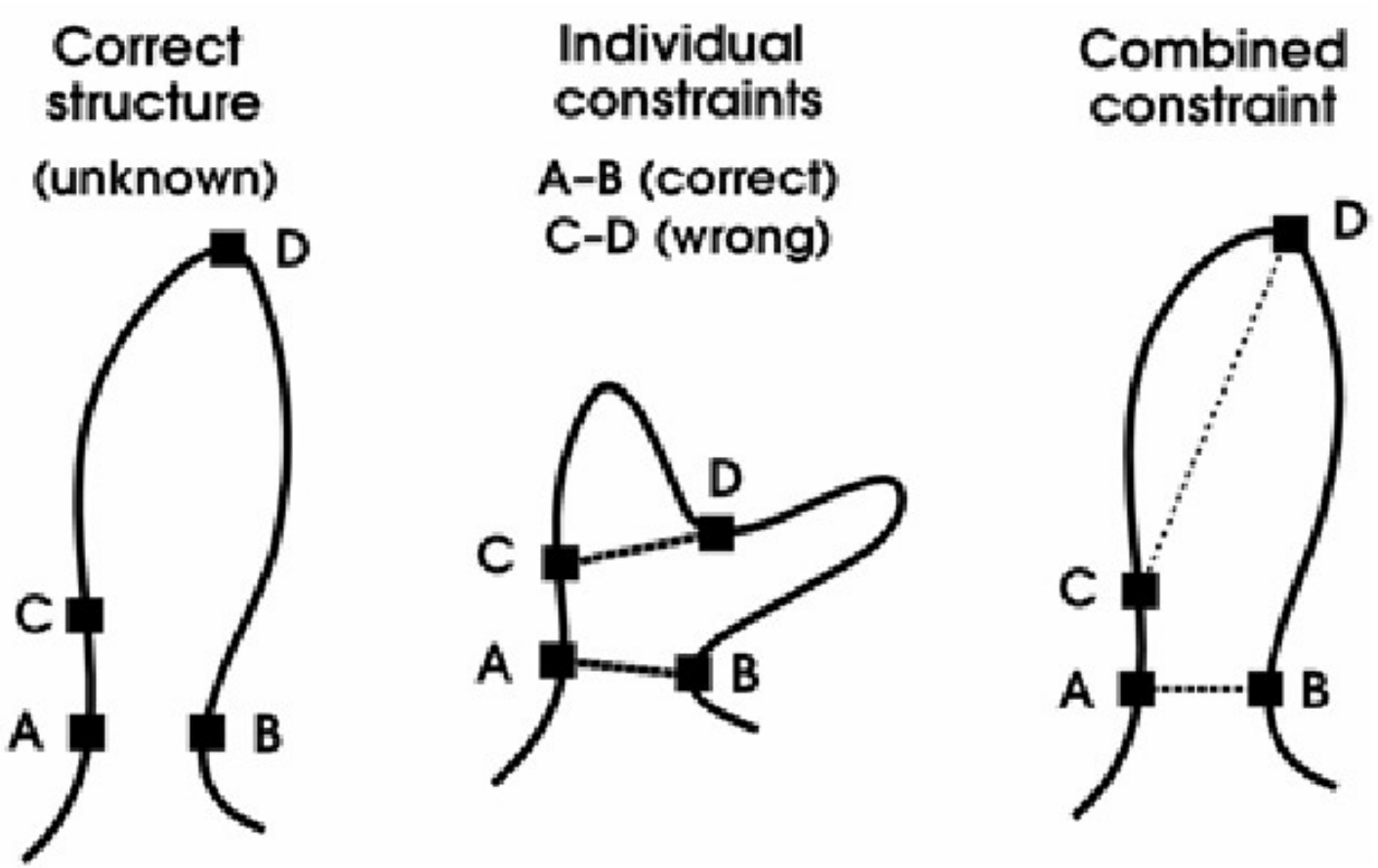
Distance effective A virtual restraint temporarily.

Individual ambiguous distances

d1	d2	d3	d4	D
2.00	-	-	-	2.00
1.00	2.00	3.00	4.00	0.997
2.00	3.00	4.00	5.00	1.97
5.00	5.00	5.00	5.00	3.97

Only uses this NOE Assignment if it exists

Effect of Constraint Combination

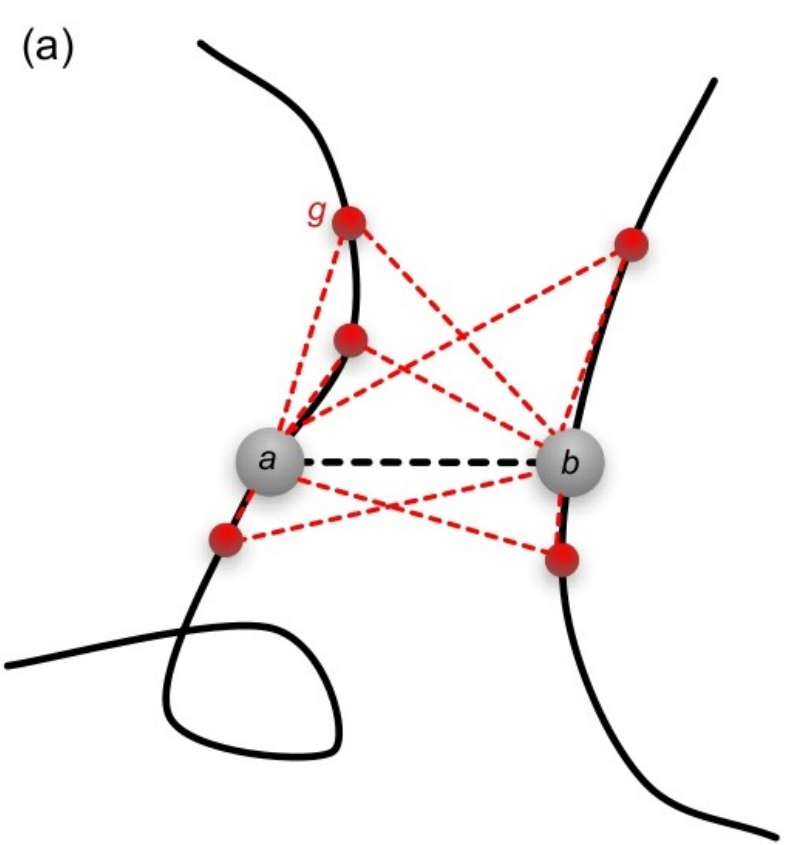


Effect of Constraint Combination

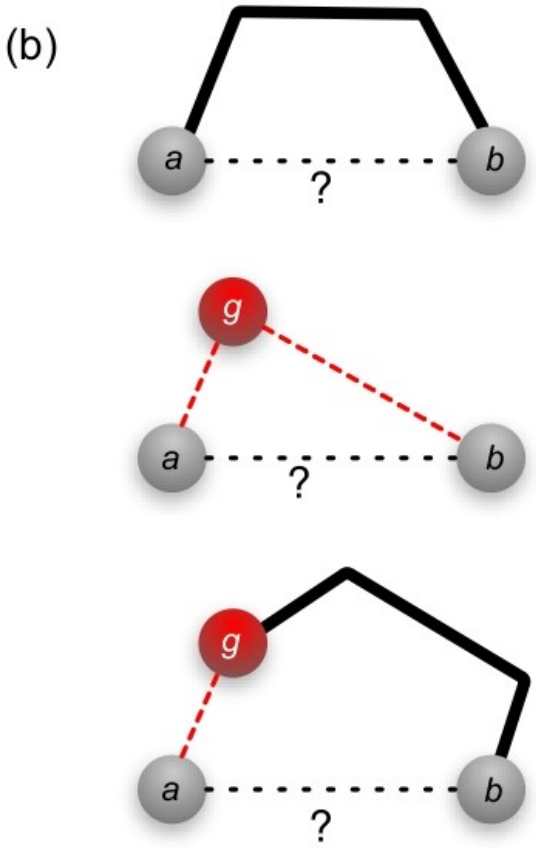
- There are some surprising statistical effects by combining restraints in such a way.
- Consider: 1000 long range restraints with 10% erroneous
- 1000 * 0.10 = 100 wrong restraints
- 2 → 1 reduction in restraints by combining gives:

  500 * 0.1*0.1 = 5 wrong restraints

Network Anchoring



a-b is a given NOE assignment possibility btw 2 protons. The program then searches all other possibilities according local geometry or covalent structure. The result should be a self-consistent subset of assignments like above.



a-b from covalent geometry there should be an NOE.

a-b-g exist to neighboring residue?

b-g covalent geometry now back check to -a

- Network Anchoring/constraint combination has a sizable effect in the first cycles of CYANA

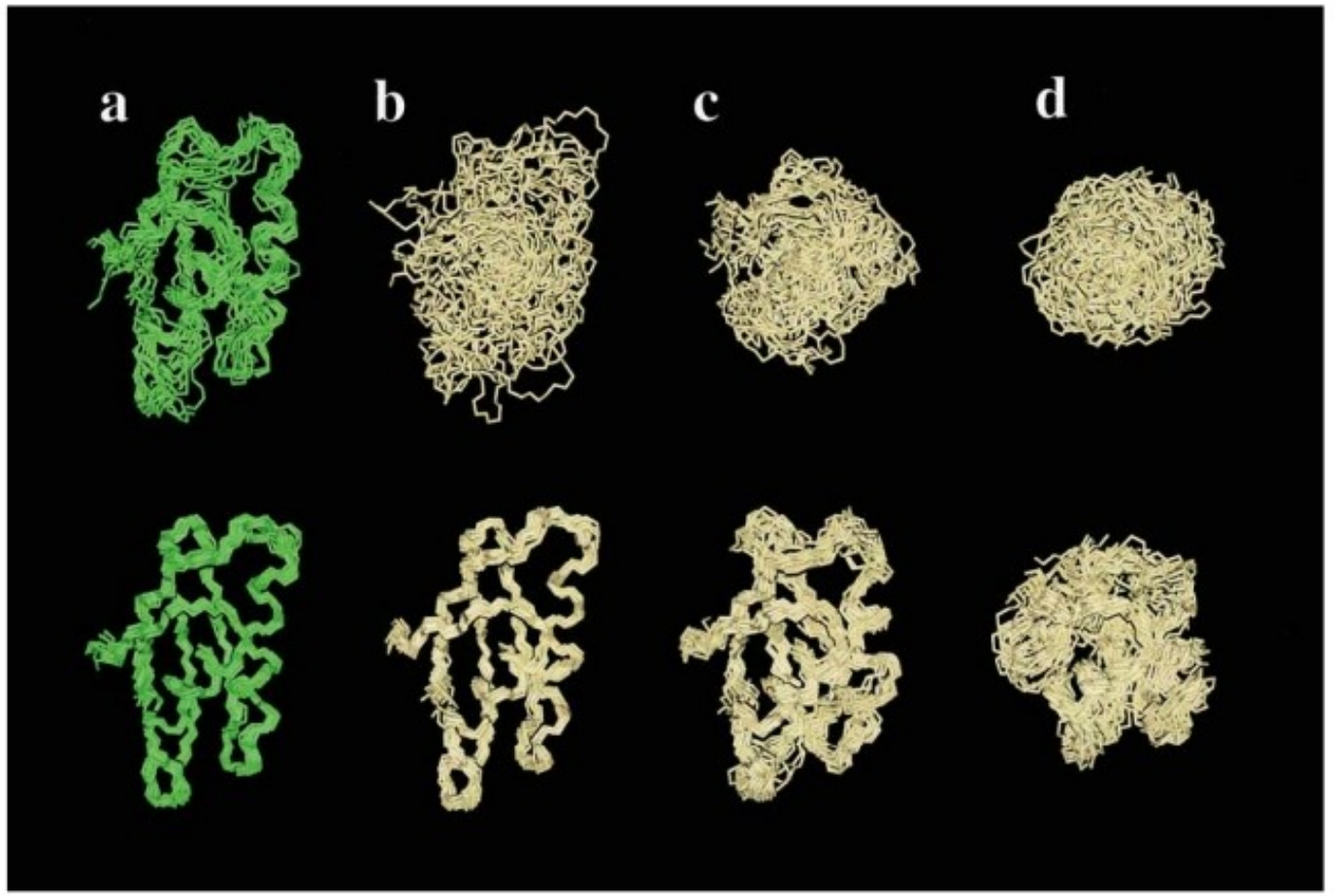


Figure 5. Bundles of the conformers with the lowest residual target function values for the protein CopZ after the CANDID cycles 1 (top) and 6 (bottom). The structures were obtained with the four CANDID calculations of Figure 4: (a) using network-anchoring and constraint-combination; (b) no network-anchoring, constraint-combination; (c) network-anchoring, no constraint-combination; (d) no network-anchoring, no constraint-combination.

Successful Structure?

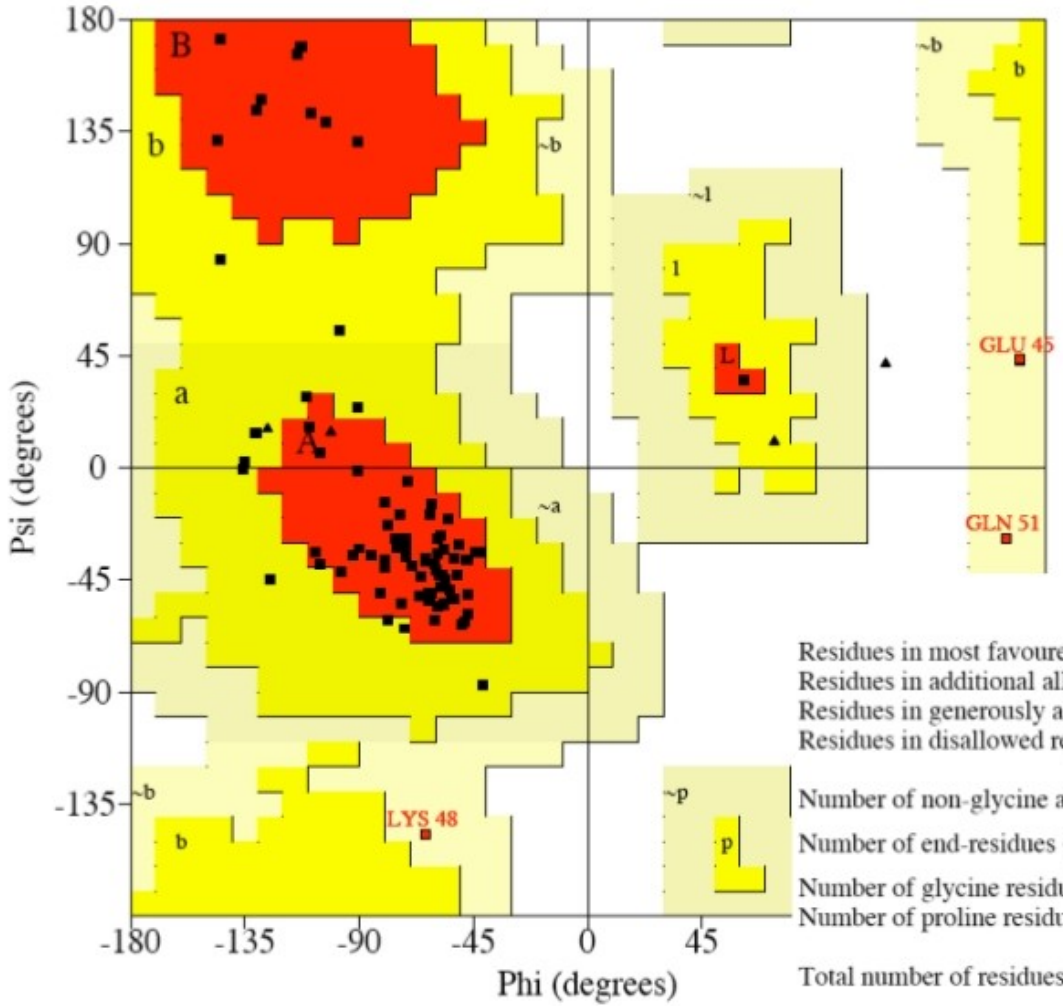
- 90% of non-labile and backbone protons should be assigned.
- The heavy backbone RMSD should be < 3 Angstroms in CYCLE 1. Exclude contributions from loops and disordered regions however.
- Less then 20% of Long range NOEs should have been discarded.
- Check Phi/Psi space quality *etc* with Procheck.

Procheck Assessment of PDB Ensemble

- Covalent geometry
- Dihedral angles
- Non-bonded interactions
- Main chain hydrogen bonds
- Stereochemical parameters
- Residue by residue analysis
- Planarity
- Chirality
- Trans-omegas

Procheck Example

-Distribution of phi-psi angles (Ramachandran plot)



- **Most ϕ,σ pairs should fall in favoured or allowed regions**

Plot statistics

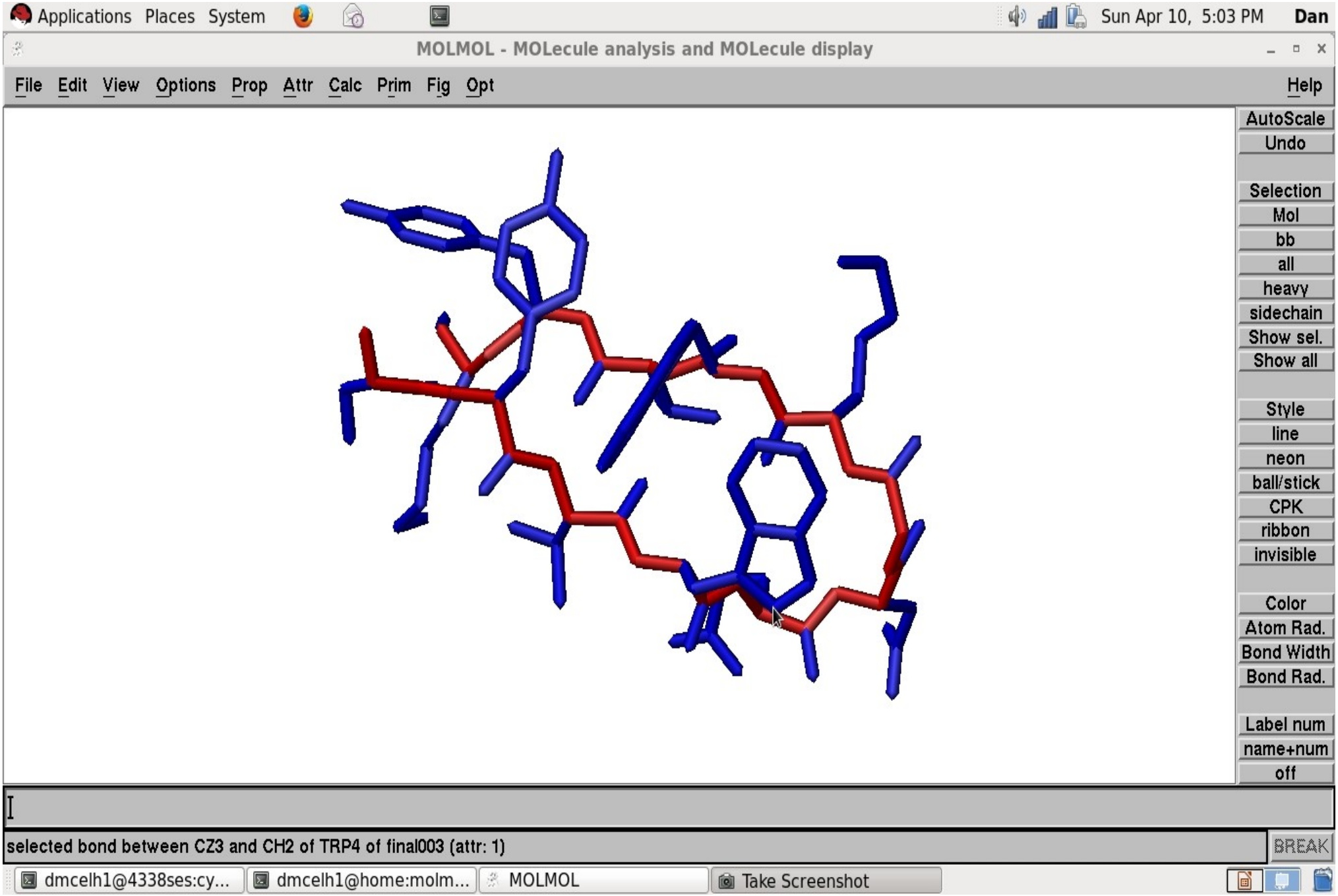
Residues in most favoured regions [A,B,L]	140	83.3%
Residues in additional allowed regions [a,b,l,p]	22	13.1%
Residues in generously allowed regions [~a,~b,~l,~p]	6	3.6%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	168	100.0%
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	8	
Number of proline residues	0	
Total number of residues	180	

Viewing in MOLMOL

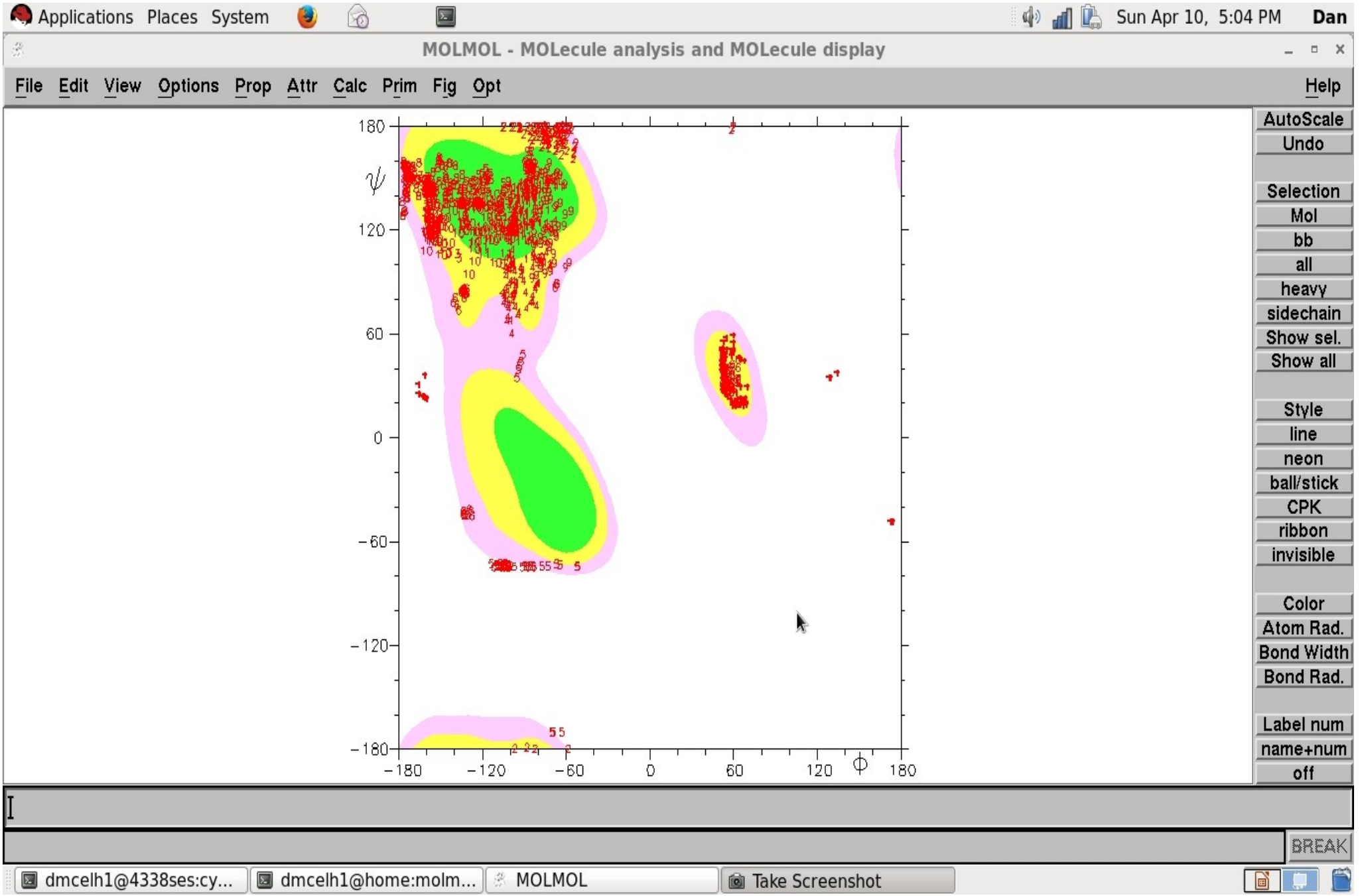
- MOLMOL is a lightweight yet powerful viewer which interfaces well with NMR studies.
- This is yet another program developed in the Wuthrich laboratory (in addition to CYANA).
- It is able to calculate RMSD and read in CYANA

  .upl and .lol as well as .aco files into the viewer.
- Publication quality figures can also be generated with practice, but people have switched to programs like 'pymol', 'vmd' and 'chimera' nowadays.

Viewing in MOLMOL



Viewing in MOLMOL



CYANA/MOLMOL Practice Sets

- We have CYANA and Molmol on the Varian PC in room 2210 SEL. You are welcome to use it to practice. It is the first computer straight ahead when you walk in the door. **BOLD** are typed commands in the linux terminal/shell
- Enter the cyana directory: **cd /home/vnmr1/chem526**
- **ls** – list dirs. You'll see several dir: ./test1 ./test2 etc. choose one as they're all the same.
- **cd cyana** ; so now your full working dir is : /home/vnmr1/chem526/test1

  you can test by **pwd** - print working dir
- **ls –** lists all the files
- **./clean –** a script for removing files before running cyana again. Have a look inside it if you like..
- **cyana < CALC.cya –** run the cyana program.

  You'll see it begin to cycle thru the 7 cycles of refinement. At the end it will give some statistics and save the final.aco fina.ovw final.pdb files and so on. Have a look inside them with 'vi' or more.
- **molmol final.pdb** - this will load the overlaid top10 pdbs generated within cyana for viewing.
- ##############################
- Go ahead and edit the CALC.cya tolerances or what ever and run the program again if you like.
- Remember to always use **./clean** before executing a new cyana run however or else it will crash.
- When finished there is no need to log out as this computer is always open to the public.

Useful Methods in Linux

- I highly recommend learning to work in the Linux environment as a scientist.
- Popular systems are: Ubuntu, CentOS and Debian. http://www.ee.surrey.ac.uk/Teaching/Unix/
- It is possible to create a 'dual boot' system having both Windows and Linux on one hard drive. So now when you reboot you can choose either Windows or Linux environment.
- Most free and useful programs are in Linux format.

Additional Methods

- Learn the editor “VI” pronounced “vee eye” . Having learned just several commands you should be able to move quickly thru text and edit programs with ease. I find it much more useful than Office. Useful commands are: dd, :wq!, 1G, d1G, dG, x, etc. please see: http://www.tutorialspoint.com/unix/unix-vi-editor.htm
- Learn how to “shell script” and the program PERL or maybe PYTHON. With these it is easy to handle large data sets and perform simple simulations. The shell scripts allow for semi-automation of programs.

References

- I borrowed many figures from these references below:

1) Dave Case, Biophysics http://casegroup.rutgers.edu/lnotes.html

2) Güntert P (2004) Automated NMR protein structure calculation with CYANA. Meth. Mol. Biol. 278:353–378

3) Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J. Mol. Biol. 273:283–298

4) James Prestegard, Chem 8190, Protein Structure Determination Using NMR Restraints.

5) Thesis, Benjamin Bardiaux (2009) "Structure calculation of proteins from solution and solid-state NMR data: Application to monomers and symmetric aggregates.