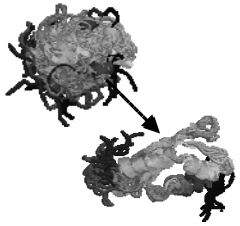
Introduction to ARIA

Structure calculations from NMR data using ARIA

(Ambiguous Restraints in Iterative Assignment)

Force Fields: topology and parameters
NMR data formats: NOEs, couplings, Dihedral angles, RDCs, Hydrogen bonds
Setup and Analysis



Henry Jonker, Biomolecular NMR spectroscopy
J.W. Goethe University, Frankfurt am Main (D)

From data to structure

- Lengthy process:
  - WET LAB: sample preparation
    - cloning, expression, purification — months
  - NMR Experiments
    - optimization of sample conditions
    - data acquisistion — months
  - Spectra Analysis
    - resonance assignent (backbone, sidechain) — weeks
  - Structure Determination
    - NOE analysis <-> Structure calculation — months

NMR Structure Information

- NMR observables that contain structural information:
  - Chemical Shift
    - secondary structure (backbone; CSI, PSSI, TALOS)
  - NOE (nuclear overhauser effect)
    - peak volume/intensities are correlated to atom distances
  - Scalar Coupling
    - correlate to (dihedral) angles
  - RDC (residual dipolar couplings)
    - orientational (long distance) restraints
  - Hydrogen bond

Structure calculation software

- XPLOR-NIH
  - Structure determination program build on the original X-PLOR. Includes tools developed at the NIH and an internal variable module (IVM) which allows efficient molecular dynamics and minimizations using internal coordinates. The IVM permits combined torsion angle, rigid body and cartesian coordinate dynamics.
- CNS (Crystallography and NMR System)
  - Provide algorithms in macromolecular structure determination. Includes several crystallographic refinement methods, as well as NMR structure calculation using NOEs, J-coupling, chemical shift, and dipolar coupling data. Basis engine for ARIA
- CNX (Crystallography & NMR eXplorer, Accelrys)
  - Commercial software package based on X-PLOR and CNS, with some extended functionality. Most of the functionalities of ARIA should be available using CNX.
- CYANA (Combined assignment and dynamics algorithm for NMR applications)
  - Automated structure calculation of biological macromolecules on the basis of conformational constraints from NMR. Automated NOESY cross peak assignment and structure calculation with a fast torsion angle dynamics algorithm.

ARIA

- ARIA (Ambiguous Restraints for Iterative Assignment)
  - Software for automated NOE assignment and NMR structure calculation. It speeds up and automates the NOE assignment process through the use of ambiguous distance restraints in an iterative structure calculation scheme.
  - ARIA does not perform itself the structure generation: it drives the structure generation (CNS) by analyzing the conformers obtained in the previous step to update the restraints and obtain a set of improved conformers. The final refinement step of the calculation is molecular dynamics in explicit solvent (water, dmso).
  - ARIA assess the quality of the final structure, using internal analyses (such as NOE violations) as well as outside tools (such as procheck).
- ARIA 1.0 - 1.2
  - Set of python scripts used to set-up, launch and analyse NMR structure calculation with CNS.
- ARIA 2.0 - 2.x
  - New graphical user interface (GUI).
  - Supports the Collaborative Computing Project for the NMR community (CCPN) data model
  - *Versions 2.0 - 2.1 lack the correction of spin diffusion....*

Protein Data Bank

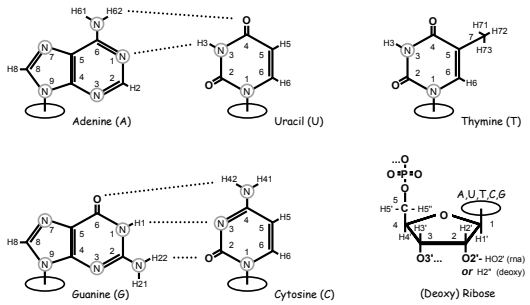
- The molecular structure of a protein, DNA, RNA and ligands are most commonly stored in the Protein Data Bank (PDB) format. Most calculation and visualisation programs can read and/or write PDB files.
- The PDB textfile contains:
  - Header, Title, Compound and Source information (chains, organisms, expression)
  - Journal references and Remarks (conditions, hardware, software, experimental data)
  - Related entries, Sequence,
  - Atomic coordinates (xyz) of one or more structures / models (bundle)
  - Atom connectivity, Secondary Structure
- Visualisation programs:
  - MOLMOL, Rasmol, Swiss PDB viewer
  - PyMOL, Chimera, VMD, ...
- Formats:
  - There is no real standard and even the atom nomenclature can be different for each programm

Forcefield

- For structural calculations a forcefield is needed with the topology and parameters (eg parallhdg, amber, charmm, opls) to define the atom bond lengths, angles, planarity and connectivity information for the (macro)molecule.
- ARIA and CNS contain several topology and parameter files for calculations with proteins, RNA and DNA. These include a macro to link the amino acids or nucleotides in order to generate the sequence.
- The parametrisation for small molecules as ligands can be obtained via the Dundee PRODRG server and/or the HIC-UP XPLO2D server.
  - proteins: allhdg5.3 forcefield
    includes PROLSQ, PARMALLH6, OPLS, PARAM19 nonbonded parameters
  - rna/dna: dna-rna-allatom forcefield
  - additional sets for carbohydrates, waterrefinement, etc

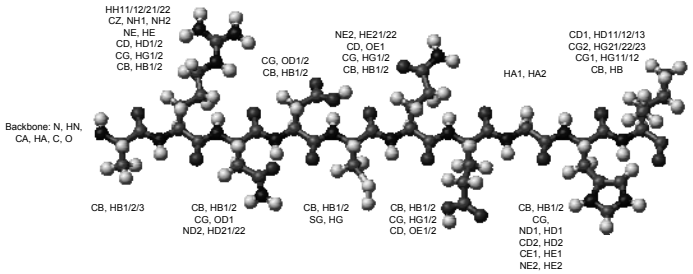
DNA & RNA nomenclature

- Nucleobases for RNA and DNA



Amino acid nomenclature

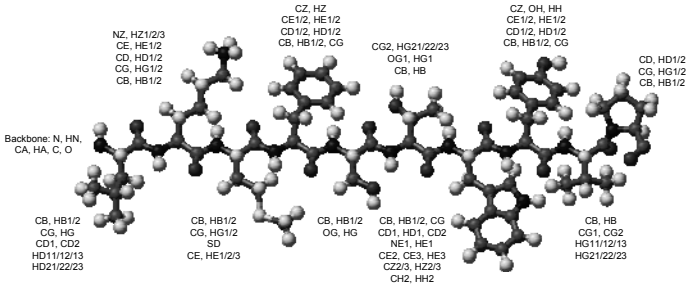
Arg (R) Asp (D) Gln (Q) Gly (G) Ile (I)



Ala (A) Asn (N) Cys (C) Glu (E) His (H)

Amino acid nomenclature

Lys (K) Phe (F) Thr (T) Tyr (Y) Pro (P)



Leu (L) Met (M) Ser (S) Trp (W) Val (V)

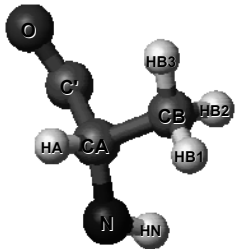
Topology

```
mass H      1.008
mass HA     1.008
mass C     12.011
mass CH1E  12.011
mass CH3E  12.011
mass NH1   14.007
mass O     15.999
```
atomic masses

```
residue ALA
  group
    atom N   type=NH1    charge=-0.570 end
    atom HN  type=H      charge= 0.370 end
    atom CA  type=CH1E   charge= 0.200 end
    atom HA  type=HA     charge= 0.000 end
    atom CB  type=CH3E   charge= 0.000 end
    atom HB1 type=HA     charge= 0.000 end
    atom HB2 type=HA     charge= 0.000 end
    atom HB3 type=HA     charge= 0.000 end
    atom C   type=C      charge= 0.500 end
    atom O   type=O      charge=-0.500 end
```
atoms, type and charge

```
  bond N   HN
  bond N   CA   bond CA  HA
  bond CA  CB   bond CB  HB1
  bond CB  HB2  bond CB  HB3
  bond CA  C
  bond C   O
```
bonds (connectivity)

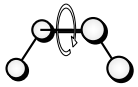
```
  improper HA  N   C   CB
  improper HB1 HB2 CA HB3

  dihedral HB1 CB  CA  C

  DONO HN   N
  ACCE O    C
end
```
improper torsion angles, dihedral angles, acceptors and donors



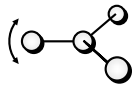
Alanine residue topology

Dihedral and improper angles

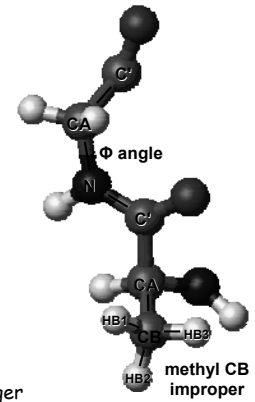
- Dihedral Angle:



- Improper angle:



- Defines
  - planarity (peptide bond, ring, base, ..)
  - chirality, stereochemistry, methyl stagger



Parameters

- Bonds (length in Ångstom)
- Angles (angle in degrees)
- Impropers (angle in degrees)
- Dihedrals (angle in degrees)
- Nonbonded (Lennard-Jones potential)

```
Term   Atoms   Force const.  value

BOND  C    CH1E          1000.00      1.525
BOND  C    NH1           1000.00      1.329
BOND  C    O             1000.00      1.231
BOND  CH1E HA            1000.00      1.080

ANGLe  C    CH1E CH1E     500.00     109.0754
ANGLe  C    CH1E HA       500.00     108.9914
ANGLe  C    NH1  CH1E     500.00     121.6541
ANGLe  C    NH1  H        500.00     119.2489

IMPRoper  C    CH1E HA   HA     500.00   0  -70.4072
IMPRoper  CH1E C    H    NH1    500.00   0    0.0000
IMPRoper  CH1E C    NH1  HA     500.00   0   66.2535
IMPRoper  HA   CH1E HA   HA     500.00   0  -66.5682

DIHEdral  C    CH2E CH2E CH1E   2.00     3    0.0000
DIHEdral  CH1E CH1E CH2E CH3E   2.00     3    0.0000
DIHEdral  CH2E CH1E CH1E NH1    2.00     3    0.0000
DIHEdral  HA   CH3E CH1E C      1.00     3    0.0000

NONBonded  C      0.105    3.750      0.013    3.750
NONBonded  H      0.050    0.500      0.004    0.500
NONBonded  N      0.170    3.250      0.021    3.250
NONBonded  O      0.210    2.960      0.021    2.960
```

multiplicity

epsilon sigma eps. (1:4) sig. (1:4)

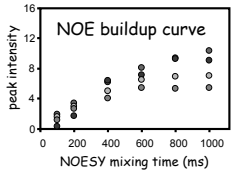
Distances from NOEs

- NOEs are due to through-space dipolar interactions
  - related to atom-atom distance: NOE ~ $1/r^6$
  - maximal 5 Ångstrom (usually put on 5.5 or 6.0 Å for calculations)
  - divided into strong/medium/weak classes or calibrated to known distances (automatically with relaxation matrix in ARIA).
  - spin-diffusion & exchange correction (for long mixing times)
- Derived from
  - 2D ^{1}H,^{1}H-NOESY

    ^{1}H ····· ^{1}H
  - 3D/4D (^{15}N,^{13}C) filtered NOESY

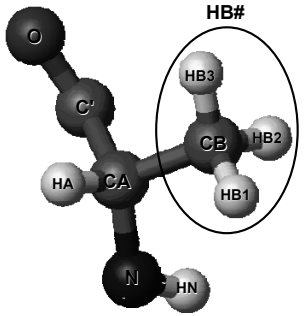
    ^{15}N-^{1}H ····· ^{1}H ^{13}C-^{1}H ····· ^{1}H

    ^{15}N-^{1}H ····· ^{1}H-^{15}N ^{13}C-^{1}H ····· ^{1}H-^{13}C ^{13}C-^{1}H ····· ^{1}H-^{15}N



Pseudo Atom Resonances

- Aria / CNS represents equivalent atom resonances (one resonance frequency) with a single pseudo atom resonance that can be represented by a wildcard " # "
- Ala: HB# → HB1, HB2, HB3
- Val: HG# → HG11, HG12, HG13 HG21, HG22, HG23
- Val: HG1# → HG11, HG12, HG13
- Val: HG2# → HG21, HG22, HG23



Alanine residue topology

ARIA/CNS format for NOEs

- SPARKY fully assigned peak:

```
K7HA-A8N-HN      3.978    120.847      8.059      9284700
```

assignment; peak position (ppm) & volume or intensity

- ARIA (.tbl):
  - Input data:

```
ASSI ( resid 8 and name HN ) ( resid 7 and name HA )
     6.0 0.1 0.1 peak 4 weight 1.1 volume 9284700 hpm1 120.847 ppm1 8.059 ppm2 3.978
```

weight (..w1)

  - Finaldata:

```
ASSI {  4}
  (( segid "    " and resid 8    and name HN  ))
  (( segid "    " and resid 7    and name HA  ))
     3.000     1.100     1.100 peak   4 spectrum    1 weight  1.1 volume ...
```

peak number

calibrated distance (Ångstrom; mean, -, +)

ARIA/CNS format for NOEs

- SPARKY (partially) unassigned peak:

```
?-E20N-HN     2.516    119.971     7.749     1876938
```

- ARIA (.tbl): ambiguous assignment

  - Input data:

```
ASSI ( resid 20 and name HN ) ( attr store1 < 2.546 and attr store1 > 2.486 )
     6.0 0.1 0.1 peak 58 volume 1876938 hpm1 119.971 ppm1 7.749 ppm2 2.516
```

  - Finaldata: weight (all, allw1)

```
ASSI {   58}
   (( segid "    " and resid 20   and name HN  ))
   (( segid "    " and resid 20   and name HB2 ))
     3.600     1.600     1.600 peak    58 spectrum    1 weight  1.0 volume  ...
OR {   58}
   (( segid "    " and resid 20   and name HN  ))
   (( segid "    " and resid 19   and name HB2 ))
```

ambiguous " OR " statement

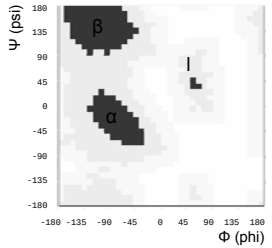
Dihedral angles restraints

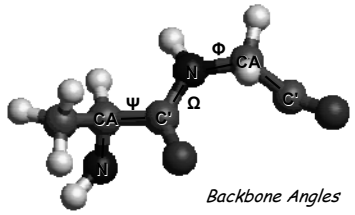
- Dihedral angles (as Φ and Ψ from chemical shifts using talos, csi, pssi)

```
assign (resid  2 and name C)  (resid  3 and name N)
       (resid  3 and name CA) (resid  3 and name C)   1.0  -66.0   20.0 2   } Φ (phi)

assign (resid  3 and name N)  (resid  3 and name CA)
       (resid  3 and name C)  (resid  4 and name N)   1.0  150.0   20.0 2   } Ψ (psi)
```

energy constant, angle (degrees), range, exponent of energy function

- Ramachandran Plot





Backbone Angles

^3J scalar couplings

- ^3J couplings (eg. $^3J_{HNHA}$)

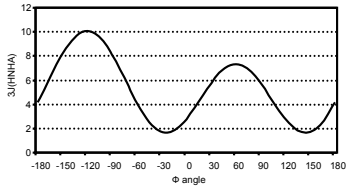
```
assign  (resid   8 and name C ) (resid   9 and name N )
        (resid   9 and name CA) (resid   9 and name C )   5.590    2
```

coupling (Hz), error

- Karplus equation:

  - $^3J = c_1 \times COS^2(\Phi+phase) + c_2 \times COS(\Phi+phase) + c_3$

  - for $^3J_{HNHA}$:
    - c1 = 6.98
    - c2 = -1.38
    - c3 = 1.72
    - Phase: -60



Residual Dipolar Couplings

- SANI (susceptibility anisotropy) restraints

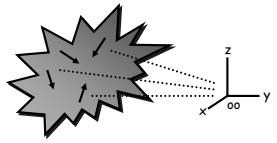
```
assign (resid 500 and name OO) (resid 500 and name Z)
       (resid 500 and name X) (resid 500 and name Y)
       (resid 9 and name N) (resid 9 and name HN) -11.12 2.000
```

rdc (Hz), error

- VEAN (vector angle) restraints

```
assign (resid 28 and name N)(resid 28 and name HN)   65.0  25.0 115.0  25.0
       (resid 50 and name N)(resid 50 and name HN)
```

angles and range



SANI type RDCs bond vector orientation with respect to an external alignment tensor
*use RAD values from PALES (R=Dr/Da, D=Da*DI) or MODULE*

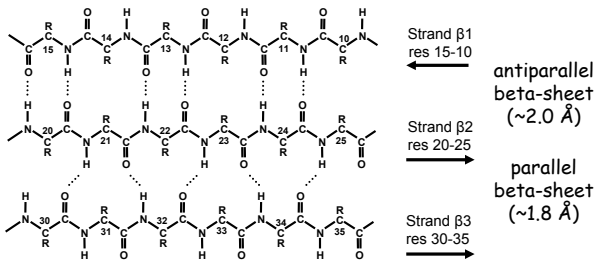
VANGLE type RDCs using purely intramolecular projection restraints
can be cacluated using RDCtools or Dipocoup

Hydrogen bond restraints

- Typically based on $^{3hb}J_{NC'}$ scalar coupling (HNCO) or deuterium exchange experiments
- Beta sheets

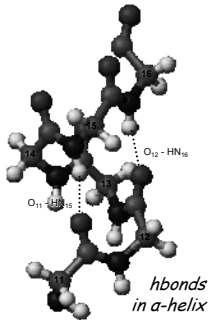
distance,-,+

```
assign ( resid 91 and name O )( resid 49 and name N  ) 2.8 0.3 0.5
assign ( resid 91 and name O )( resid 49 and name HN ) 1.8 0.3 0.5
```



Hydrogen bond restraints

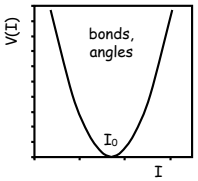
- Alpha helix and 3_{10} helix
  - α-helix hbonds between O_i and HN_{i+4} (2.08 Å)
  - 3_{10}-helix hbonds between O_i and HN_{i+3} (1.83 Å)
- DNA/RNA base pairs
  - AU / AT and GC Watson-Crick base pairs
  - planarity restraints for WC base pairs
  - Hoogsteen and non-canonical base pairs
- Ambiguous Hydrogen bonds

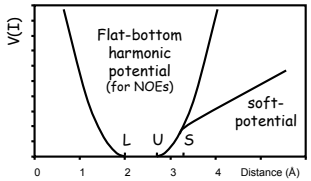



```
assign (resid  13  and name N)
       ((name O* and not (name O or name OT#)) or
       (not (resid  12 : 14 ) and (name O or name OT#)))  2.9 0.3 0.5
assign (resid  13  and name HN)
       ((name O* and not (name O or name OT#)) or
       (not (resid  12 : 14 ) and (name O or name OT#)))  1.9 0.3 0.5
```

Structural calculations

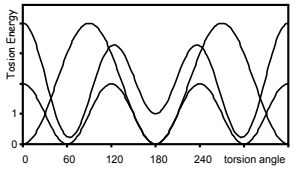
- 3D structure has to satisfy
  - experimental restraints (NOE distances, couplings, RDCs, ...)
  - chemical knowledge (empirical forcefield: bonds, angles, chirality)
- Find the minimum of a target energy function:
  - $V_{potential} = V_{bonds} + V_{angles} + V_{torsion} + V_{nonbonded} + V_{experimental}$
  - Harmonic Potential: $V(I) = \frac{1}{2}k \times (I-I_0)^2$



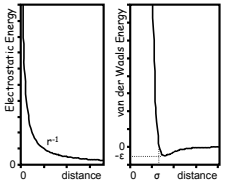


Structural calculations

- Torsional terms are usually expressed as cosine series expansions
  - $V(\omega) = \Sigma_n \frac{1}{2}V_n \times [\ 1 + \cos(n\omega-\gamma)\ ]$
  - multiple minima
  - not equal with multiple terms

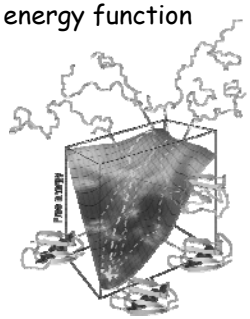


- Electrostatic interactions
  - Coulombs Law: $V_{elec} = \Sigma_i\Sigma_j\ q_iq_j\ /\ 4\pi\varepsilon_0 r_{ij}$
    - ε = dielectric constant, q = partial point charges
- Van der Waals interactions
  - Lennard-Jones: $V_{L-J} = 4\varepsilon\ [\ (\sigma/r)^{12} - (\sigma/r)^6\ ]$
    - ε = well depth, σ = collision diameter



Structure calculation methods

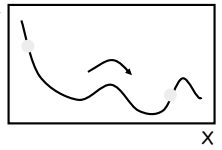
- Computational methods are needed to find the local energy minimum of the potential energy function
- Energy Minimization (EM)
  - simplex, steepest descent, ...
  - will only locate closest minimimum
  - cannot cross energy barriers
- Distance Geometry
  - direct conversion
  - optimization necessary in real space

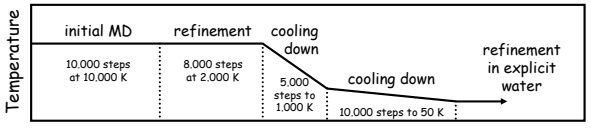


- Molecular Dynamics with Simulated Annealing
  - Follows Newton's laws of motion (inertia, F=ma, action=reaction)
  - Cartesian Angle Dynamics an Torsion Angle Dynamics

Molecular Dynamics (MD)

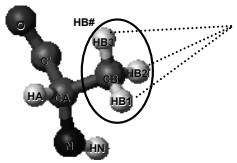
- The direction of motion depends on
  - forces (restraints and forcefield) and momentum
- MD can overcome local energy barriers
  - temperature is related to the kinetic energy and thus the velocity
  - conventional timestep for NMR: 20-50 fs (flexible bonds and vibrations: 2-5 fs)
  - Energy terms can be scaled differently (use of NOE term during cool-down)
  - Simulated Annealing (SA) protocol





Assignment and Validation

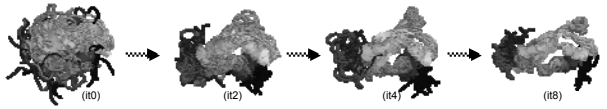
- Ambiguous assignments
  - resonace overlap causes multiple possibilities for assignment
  - trial and error and choosen to be consistent with structure
  - ambiguous NOEs correspond to the sum of individual contributions
    - $NOE = \Sigma_\alpha NOE_\alpha$
  - ARIA can use unassigned and ambigously assigned NOE peaks for direct refinement as ambiguous distance retraints (ADRs)
  - ADRs are also used for equivalent protons (eg aromatics and methyl groups) and are not treated as one fixed 'pseudo atom'.
  - The ambiguouty cut-off is reduced over the various iterations
- Validation
  - Structures should satisfy the restraints
  - check geometry, energies, violations, rmsd
  - quality check: procheck, whatif, ...
  - new assignments?



ARIA procedure

- Input NMR data (as NOESY peaks, resonance frequencies, rdcs, dihedral angle restraints, coupling constants, hydrogen bonds, talos, csi, ...) and the protein/dna/rna sequence (or pdb)
- ARIA generates an extended structure and restraints lists

- Calibrate NOESY(s), assign NOEs and calculate structures
- Analyse and repeat (iterations 1-7) with tighter restraints



- Calculate final structures (iteration 8) and analyse
- Refinement in explicit solvent (water) and analyse