Experiment and Modelling in Structural NMR

November 28th – December 2nd 2011

INSTN – CEA Saclay, France

Michael Nilges
Institut Pasteur, France

Molecular modelling
applied to
NMR structure determination

[01003]

Organized by

Thibault Charpentier
Patrick Berthault
Constantin Meis

thibault.chapentier@cea.fr
patrick.berthault@cea.fr
constantin.meis@cea.fr

Molecular Modelling

Applied to NMR Structure Determination

Michael Nilges; Structural Bioinformatics Unit
Department of Structural Biology and Chemistry; Institut Pasteur
michael.nilges@pasteur.fr
033 1 45 68 82 30

Monday, November 28, 11

<http://aria.pasteur.fr/documentation/courses/saclay-november-2011/presentation>

pdf file of this talk

Monday, November 28, 11

- Introduction: the hybrid energy function
- NMR data: distances, angles, orientations, and noise
- Minimization algorithms
- Relation to probability theory

- The hybrid energy function concept
- NMR data: distances, angles, orientation
- Minimization algorithms
- Relation to probability theory

NMR structure determination steps

5

- NMR experiment
- Resonance assignment
- Structural restraints
 - distances
 - NOE assignment
 - torsion angles, orientation
- Structure calculation
- Structure validation

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

INSTITUT PASTEUR

Monday, November 28, 11

Structure calculation

6

- Have molecule and some data on conformation...
- Objectives:
 - find conformation(s) satisfying experimental data
 - maintain likely (local) conformation

NOE peak size

molecular conformation

$$NOE_{ij} \propto r_{ij}^{-6}$$

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

INSTITUT PASTEUR

Monday, November 28, 11

“Theory” ; “Forward model”

7

- Basis: model to calculate data from structure
 - model (e.g.): Isolated Spin Pair Approximation for NOE
 - calculate measurement (NOE) from structure (a distance)

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Hybrid energy function

8

- weighted sum of data and force field contributions
- each data term carries its own weight
- weights determined by
 - empirical means (trial and error)
 - statistical means (cross-validation)
 - Bayesian probability

$$E_{\text{hybrid}} = E_{\text{phys}} + w_{\text{data}} E_{\text{data}}$$

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Role of molecular modelling

9

- force field: supplements experimental data by previously known information
- penalty function: provides means to restrain / constrain molecular model to data (e.g., flat-bottom potential)
- minimization algorithm: move structure to minimize energy and satisfy data

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

10

- Introduction: the hybrid energy function
- NMR data: distances, angles, orientation, noise
- Minimization algorithms
- Relation to probability theory

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Experimental data from liquid state NMR

11

- chemical shift
 - local electronic environment; distances
 - torsion angles
- scalar coupling constants
 - torsion angles
 - distances (hydrogen bonds)
- NOE, ROE
 - interproton distances
- paramagnetic atoms
 - distances, orientation
- residual dipolar couplings etc.
 - bond orientation

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

INSTITUT PASTEUR

Monday, November 28, 11

Measurement of distances: NOE

12

- size of NOESY peak depends on
- crossrelaxation rate
 - the distance between the two protons
 - overall rotational motion
 - internal motion
- mixing time
- ... presence of other protons

$$\sigma_{ij} \propto r_{ij}^{-6} f(\tau_c)$$

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

INSTITUT PASTEUR

Monday, November 28, 11

Buildup curves to measure σ

- Crossrelaxation rate from slope of buildup curve
- measure several NOESY at different mixing time
- distinguishes between direct (a) and indirect (b) NOEs

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Distance from crossrelaxation rate/ NOE

- Interproton distance from
 - crossrelaxation rate (neglects internal dynamics)
- $$r_{ij} = (\sigma_{ij} f(\tau_c))^{-\frac{1}{6}}$$
- Crossrelaxation rate from:
 - NOE measurements at several mixing times (buildup curve)
 - NOE measurement and relaxation matrix calculation
 - Approximately: use NOE volume/ intensity (neglects spin diffusion)

$$r_{ij} \approx (C_{cal} V_{ij})^{-\frac{1}{6}}$$

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Typical interproton distances in proteins

methylene group	dH β 2-H β 3	1.8 Å	very strong NOE
aromatic	dH δ -H ϵ	2.4 Å	strong NOE
antiparallel β sheet	dH α -H α	2.2 Å	strong NOE
parallel β sheet	dH α -H α	2.7 Å	strong NOE
α helix	dHN-HN(i,i+1)	2.7 Å	strong NOE
α helix	dH α -HN(i,i+3)	3.3 Å	medium NOE

1 Å = 0.1 nm

- Known distances can be used to approximately convert NOE peak volumes into distance ranges

Monday, November 28, 11

NOE summary

- richest source of structural information; most important data for structure determination by NMR
- very powerful for qualitative analysis of structures (assignment of secondary structure)
- interactions between residues far apart in sequence
- potentially large errors due to approximate theory to convert NOEs to distances: approximate distance ranges

Monday, November 28, 11

Angular information from coupling constants

- 3-bond coupling depends on torsion angle

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Coupling constant and angle ϕ

- General dependency of J on angle:
- Karplus-relationship

$$J = A + B\cos(\theta) + C\cos(2\theta)$$

$$\phi = \theta - 60$$

- The parameters A, B, C need to be parametrized with known (X-ray) structures

A	B	C
6.51	-1.76	1.60
6.41	-1.46	1.90
6.98	-1.38	1.72

Monday, November 28, 11

Residual dipolar couplings from partial alignment

19

- Partial alignment due to:
 - bicelles
 - purple membranes
 - phages...
- magnetic interactions (DNA)
- paramagnetic tags...

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

INSTITUT PASTEUR

Monday, November 28, 11

Residual dipolar couplings

20

- Proteins:
 - direction of bond vectors (e.g., N-H) can be determined
 - relative to coordinate system attached to molecule

$$D^{res} \propto \frac{\gamma_i \gamma_j}{r_{ij}^3} \left[D_{ax} (3 \cos^2(\theta) - 1) + \frac{3}{2} D_{rh} \sin^2(\theta) \cos(2\phi) \right]$$

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

INSTITUT PASTEUR

Monday, November 28, 11

Noise in Data

- All data contain errors (experimental noise)
- All forward models contain approximations
- No ideal agreement between calculated and measured data possible
- Penalty function for data needs to contain way to include noise
- Automated methods to detect “noise peaks” (violation analysis, network anchoring)

Distance measurements contain errors

- NOEs only give approximate measure of distance
- measurement errors
 - evaluation of peak volumes
 - experimental parameters
- errors in conversion to distance
 - how to measure crossrelaxation rate?
 - spin diffusion
 - internal dynamics
 - peak broadening

Distance ranges

NOE	lower bound	upper bound
very strong	1.8 Å	2.5 Å
strong	1.8 Å	2.8 Å
medium	1.8 Å	3.6 Å
weak	1.8 Å	5.0 Å
very weak	1.8 Å	6.0 Å

- Error in measurement:
- derive consistent bounds on distance
- set bounds based on statistical analysis of known structures

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Standard NOE distance restraint potential

$$E_{data} \propto \sum_i^{N_{noe}} \begin{cases} (r_i(\mathbf{X}) - L_i)^2 & \text{if } r(\mathbf{X}) < L_i \\ 0 & \text{if } L_i \leq r(\mathbf{X}) \leq U_i \\ (r_i(\mathbf{X}) - U_i)^2 & \text{if } r(\mathbf{X}) > U_i \end{cases}$$

- sources of error
 - measurement, spin diffusion, internal dynamics
- loose upper and lower bounds
- FBWH potential
 - flat bottom harmonic walls
 - no force between L and U

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Consequence of bounds

- Bounds have to be large enough for cumulative error
- Precise value not (too) important:
 - even loose bounds restrict conformational space
- May affect
 - precision of structure
 - validation
 - noise peak recognition (see below)

Other ways to treat noise

- potential form
- weight in hybrid energy function

$$E_{\text{hybrid}} = E_{\text{phys}} + w_{\text{data}} E_{\text{data}}$$

Data summary

- Forward models contain non-measurable parameters that are necessary for the modelling
 - calibration factor
 - Karplus parameters for scalar couplings
 - tensor parameters
- Data potential needs to include parameter to treat (unknown) “noise”
- The weight in the hybrid energy needs to be set by empirical means

...

- influenced by internal dynamics:
- relaxation times
- NOE, ROE
- most data describe
 - the local environment of the protons
 - ...relative to each other
 - not the global conformation of the molecule

- Introduction: the hybrid energy function
- NMR data: distances, angles, orientation
- **Minimization algorithms**
- Relation to probability theory

3D structure calculation

- convert data (1D, 2D)
- + forcefield
- into 3D model

Structure calculation methods: minimize hybrid energy

31

- Metric matrix distance geometry (DISGEO, DG2)
- (Energy) minimization ("buildup method", DIANA)
- Simulated annealing (molecular dynamics) from random structures (X-PLOR, CNS)
- Simulated annealing (torsion angle dynamics) from random structures (X-PLOR, CNS, DYANA)

Monday, November 28, 11

Data from structure

32

- Basis of structure calculation: calculate data
 - NOE
 - approximate: distances
 - NOE from relaxation matrix calculations
 - Coupling constants
 - approximate: torsion angles/ Karplus relations;
 - QM calculations
 - RDCs
 - bond orientations in alignment tensor
 - Chemical shifts
 - empirical relations; QM calculations
- include error due to measurement/ approximations

Monday, November 28, 11

- Two principles:
- Minimization
 - make random proposal for structure
 - calculate data from structure
 - compare with experiment
 - modify structure to improve agreement
- Sampling
 - make random proposal for structure
 - calculate data from structure
 - make new random proposal

- Two principles:
- Minimization
 - make random proposal for structure
 - calculate data from structure
 - compare with experiment
 - modify structure to improve agreement
- Sampling
 - make random proposal for structure
 - calculate data from structure
 - make new random proposal

Experimental distances insufficient

- Data are incomplete:
 - NOE distance ranges only for protons
 - torsion angle ranges for some atoms
 - (orientation for some bonds)
- for most atoms no direct experimental observation

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Additional data: prior information

- Need prior information for building blocks: amino acids or nucleic acids
- topology (which atoms are connected)
- parameters
 - bond lengths
 - bond angles
 - planarity
 - chirality
 - atomic radii

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

- Metric matrix distance geometry (DISGEO, DG2)
- (Energy) minimization ("buildup method", DIANA)
- Simulated annealing (molecular dynamics) from random structures (X-PLOR, CNS)
- Simulated annealing (torsion angle dynamics) from random structures (X-PLOR, CNS, DYANA)

Structure calculation: Simulated Annealing

Multiple minimum problem

38

High energy
barriers
to fold protein

Standard
minimization
only "downhill"

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

INSTITUT PASTEUR

Monday, November 28, 11

Minimization by molecular dynamics

39

$$\frac{d^2 \mathbf{r}_i}{dt^2} = - \frac{c}{m_i} \frac{\partial}{\partial \mathbf{r}_i} E_{\text{hybrid}}$$

- Molecular dynamics solves Newton's equations of motion
- Molecular dynamics can overcome local energy barriers

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

INSTITUT PASTEUR

Monday, November 28, 11

Newton dynamics

40

- Direction of motion depends on
- force (derived from force field and experimental restraints)
- momentum

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Temperature control and variation: "MD-simulated annealing"

41

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Energy scaling

42

- more flexible annealing schemes
- different variation of different energy terms
- e.g.:
 - E_{phys} / E_{exp}
 - $E_{covalent} / E_{exp} / E_{nonbond}$
- equivalence:
- mass/ energy/ temperature scaling

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

43

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Structure calculation with MD

- NMR data: distances
- Start: random structure
- Difficult search problem: many degrees of freedom

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Structure calculation with MD

46

1988:

48 hours per structure on
mainframe (DISGEO, Havel)

2001:

seconds per structure on PC

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Torsion angle dynamics

47

- dynamics time step dictated by bond stretching: waste of CPU time
- important motions are around torsions
- ~ 3 degrees of freedom per AA
- (cf 3Natom for Newton dynamics)
- Available in DYANA, X-PLOR, CNS, X-PLOR-NIH

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

- calculation with simplified force field, torsion angle dynamics
 - no electrostatics, simplified van der Waals
- refinement with Cartesian dynamics
- very short final refinement with “full” force field in water

Final refinement in H₂O

Calculation of structure ensembles

50

- with identical data/ restraints:
- repeat calculation (20-100 times)
- random variation of initial conditions (starting structure/ velocities)
- obtain information on
 - uniqueness / different folds
 - "dynamics"

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

High quality structure ensembles

51

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Meaning of structure ensembles

- Simple way to assess uniqueness of solution
- This has very little to do with dynamics
- Distribution depends on
 - data
 - data representation
 - algorithm
 - forcefield
 - algorithm parameters
 - ...

Evaluation of structures

- Energetic criteria
 - E_{phys}
 - RMS from ideal values for covalent interactions
 - number of large deviations
- Comparison to other structures, “knowledge-based”
 - e.g., WhatIf
- Satisfaction of experimental data
 - restraint violations
 - E_{data}
 - RMS from data / bounds
- Statistical criteria for data
 - crossvalidation
- <http://proteins.dyndns.org/cing>

-
- Introduction: the hybrid energy function
 - NMR data: distances, angles, orientation
 - Minimization algorithms
 - Relation to probability theory

Minimisation and probability

- Where do potential forms come from
- Where do all the parameters come from
 - bounds
 - weights
 - any parameter required by theory

Probability and energy

56

$$E_{hybrid} = E_{phys}(\mathbf{X}) + w_{data} E_{data}(D, \mathbf{X})$$

Probability and energy

56

$$E_{hybrid} = E_{phys}(\mathbf{X}) + w_{data} E_{data}(D, \mathbf{X})$$

- force field $E_{phys} \Leftrightarrow$ probability (Boltzmann)

Probability and energy

$$E_{hybrid} = E_{phys}(\mathbf{X}) + w_{data} E_{data}(D, \mathbf{X})$$

- force field $E_{phys} \Leftrightarrow$ probability (Boltzmann)
- probability of distortion of molecule

Probability and energy

$$E_{hybrid} = E_{phys}(\mathbf{X}) + w_{data} E_{data}(D, \mathbf{X})$$

- force field $E_{phys} \Leftrightarrow$ probability (Boltzmann)
- probability of distortion of molecule
- force field: background information I

Probability and energy

$$E_{hybrid} = E_{phys}(\mathbf{X}) + w_{data} E_{data}(D, \mathbf{X})$$

- force field $E_{phys} \Leftrightarrow$ probability (Boltzmann)
- probability of distortion of molecule
- force field: background information I
- prior probability

Probability and energy

$$E_{hybrid} = E_{phys}(\mathbf{X}) + w_{data} E_{data}(D, \mathbf{X})$$

- force field $E_{phys} \Leftrightarrow$ probability (Boltzmann)
- probability of distortion of molecule
- force field: background information I
- prior probability

$$P(\mathbf{X}|I) = \exp \left[-\frac{E_{phys}(\mathbf{X})}{kT} \right]$$

Probability and energy

57

$$E_{hybrid} = E_{phys}(\mathbf{X}) + w_{data} E_{data}(D, \mathbf{X})$$

Probability and energy

57

$$E_{hybrid} = E_{phys}(\mathbf{X}) + w_{data} E_{data}(D, \mathbf{X})$$

- similar: $E_{data} \Leftrightarrow$ probability

Probability and energy

$$E_{hybrid} = E_{phys}(\mathbf{X}) + w_{data} E_{data}(D, \mathbf{X})$$

- similar: $E_{data} \Leftrightarrow$ probability
- probability that data is correct, given structure $\mathbf{X}$:

Probability and energy

$$E_{hybrid} = E_{phys}(\mathbf{X}) + w_{data} E_{data}(D, \mathbf{X})$$

- similar: $E_{data} \Leftrightarrow$ probability
- probability that data is correct, given structure $\mathbf{X}$:
- “likelihood”

- Example:
- Gaussian distribution of error for r ,
- standard deviation σ ,
- $\Rightarrow$ probability is

$$P(D|\mathbf{X}, \sigma) \propto \exp \left[\frac{-(r - r(\mathbf{X}))^2}{2\sigma^2} \right]$$

Likelihood and restraint potential

- Inversely, if we know probability distribution, we can derive potential

$$E_{data} \propto -\log [P(D|\mathbf{X}, \sigma)]$$

- For Gaussian error, harmonic potential ("least squares")

$$E_{data} \propto \frac{1}{2\sigma^2} (r - r(\mathbf{X}))^2$$

- The weight is related to the error in the data

Distances (NOEs) do not follow Gaussian

Gaussian distribution of
logarithms

Gaussian distribution

Rieping, Habeck, Nilges, JACS 2005

Monday, November 28, 11

Log-normal distribution

- Log-normal distributions
- and derived potentials

$$\text{LN}(x_0, x, \sigma) \equiv \frac{1}{\sqrt{2\pi\sigma^2}x_0} \exp\left[-\frac{1}{2\sigma^2} (\log[x_0] - \log[x])^2\right]$$

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

...Life is LogNormal

- ... there are a lot of data with only positive values
- examples on
- <http://stat.ethz.ch/~stahel/lognormal/>
 - no theoretical derivation

Disciplines		mu*	sigma*
Medicine	Onset of Alzheimer disease	~ 60 years	1.2
	Latent periods of infectious diseases	Hours to months	1.5
	Survival time after diagnosis of cancer	Months to years	3
Environment	Air pollution in the U.S.A.	40-110 PSI	1.5-1.9
	Rainfall	80-200 m3 (x103)	4-5
	Species abundance in ecology	-	6-30
Social sciences and linguistics	Income of employed persons	6.700sFr	1.5
	Lengths of spoken words	3-5 letters	1.5

Monday, November 28, 11

Joint probability from prior and likelihood

- To calculate joint probability from single probabilities, multiply:

Probability of a structure:
Posterior Probability

likelihood

$$P(\mathbf{X}|D, I) \propto P(\mathbf{X}|I)P(D|\mathbf{X}, \sigma, I)$$

prior distribution

Monday, November 28, 11

Hybrid energy revisited

64

$$E_{hybrid} = E_{phys}(\mathbf{X}) + w_{data}E_{data}(D, \mathbf{X})$$

- The hybrid energy function is negative logarithm of joint probability
- Minimum energy corresponds to maximum probability
- Relative weight “should” depend on data quality
- story is incomplete (what about w_{data} ?)

Probability of a structure

likelihood

$$P(\mathbf{X}|D, I) \propto P(\mathbf{X}|I)P(D|\mathbf{X}, \sigma, I) \dots$$

prior distribution

Monday, November 28, 11

Bayesian determination of data weight

65

$$E_{hybrid} = E_{phys} + w_{data}E_{data}$$

- Data weight has influence on structure quality

- Bayesian analysis:

$$w_{data} = \frac{k_B T}{2RMS^2}$$

- Update iteratively during structure calculation
- weight $\Leftrightarrow$ overall data quality
- only possible for “least squares”-type potential

Habeck M, Rieping W, Nilges M (2006). PNAS 103:1756

Monday, November 28, 11

- Minimizing hybrid energy corresponds to maximizing the probability of a structure, given data and force field
- ...if one knows the data quality, scale factors, ...
- Relative weights
 - usually set empirically (trial and error, experience, cross validation)
 - Bayesian determination of weight possible
- Relationship of error distribution and restraint potentials

-
1. Introduction: relating data to structure
 2. Hybrid energy and treatment of errors
 3. Minimisation of hybrid energy
 4. Relation to probability theory

Problems inherent in minimisation

68

data are incomplete: solution is degenerate
data are inconsistent: strictly speaking, no solution exists
many unknown parameters are necessary (“nuisance parameters”)
no objective figures of merit for structures
no consistent concepts of data quality evaluation

Example NOE

69

- incompleteness: assignments, NMR-“visibility”
- inconsistency: approximate theory, noise
- unknown quantities: calibration, data consistency
- basic question: how well do my data determine the structure remains unanswered, need of heuristics:
 - cross validation
 - independent validation

- standard approach works in practice with sufficient data of good quality
- sparse data:
 - problems with determining auxiliary parameters
 - structure calculation difficult
- no estimation of uncertainties in coordinates or data
 - RMSDs and R-factors depend on all auxiliary parameters
 - few restraints can change result drastically
 - no concept to evaluate data quality ("don't overfit"... "use data not used in structure calculation" ...)

Monday, November 28, 11

Inference instead of deduction

- *inference*:
 - assign a *probability* to each molecular conformation
- use probability theory:
 - prior probability from physical model (force field)
 - likelihood from forward model

$$P(\theta, \sigma, \gamma | D, I) \propto P(\theta, \sigma, \gamma | I) P(D | \theta, \sigma, \gamma, I)$$

Monday, November 28, 11

Sampling

72

- Posterior $P(X|D)$ is extremely complex for realistic problem
 - too many degrees of freedom to do “integration”
- Take representative samples (Markov Chain Monte Carlo)

Monday, November 28, 11

Sampling probability distributions

73

- Sampling is computationally much more complex than structure calculation by minimization
- cf calculating partition function in statistical mechanics
- Algorithm uses
 - hybrid Monte Carlo
 - torsion angle dynamics
 - replica exchange
 - Tsallis distribution

Monday, November 28, 11

Sampling probability distributions

73

- Sampling is computationally much more complex than structure calculation by minimization
- cf calculating partition function in statistical mechanics
- Algorithm uses
 - hybrid Monte Carlo
 - torsion angle dynamics
 - replica exchange
 - Tsallis distribution

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

INSTITUT PASTEUR

Monday, November 28, 11

Hybrid Monte Carlo Algorithm

74

- Monte Carlo is inefficient for polypeptides (polymers in general):
 - most moves either high non-bonded or covalent energy
 - many correlated degrees of freedom
- Combination of Molecular Dynamics and Monte Carlo:
 - assign random momenta
 - run short NVE MD to get new proposal state (e.g., 200 steps)
 - evaluate with Metropolis criterion on total energy

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

INSTITUT PASTEUR

Monday, November 28, 11

Torsion angle dynamics

75

- Important degrees of freedom: torsion angles
- True torsion angle dynamics: equations of motion with a complicated structure (time-dependent masses, non-diagonal mass matrix)

$$M(\phi) \frac{d^2 \phi}{dt^2} + C\left(\frac{d\phi}{dt}, \phi\right) = 0$$

- for *sampling* sufficient:

$$m \frac{d^2 \phi}{dt^2} = - \frac{\partial}{\partial \vec{r}} E_{\text{hybrid}}$$

Monday, November 28, 11

Replica hybrid MC algorithm

76

- start 25-50 hybrid Monte Carlo trajectories in parallel: **replicae**
- replicae run at different constant conditions (temperatures, weights)
- every 50 hybrid Monte Carlo steps:
exchange conformations between replicae;
- preserve "detailed balance"

Monday, November 28, 11

“Temperatures” and Tsallis distribution

$$E(\theta; q) = \frac{q}{\beta(q-1)} \log[1 + \beta(q-1)[E(\theta) - E_{min}]]$$

- prior (force field):
high T -> non-Boltzmann statistics (q; Tsallis)
- likelihood (data): high T -> exponent λ

Monday, November 28, 11

structures at different "temperatures"

hybrid MC with replica exchange

random starting structure

Monday, November 28, 11

Sampling over nuisance parameters

79

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

INSTITUT PASTEUR

Monday, November 28, 11

Monday, November 28, 11

Typical trace (SH3 domain)

81

replica exchanges

“energy”

data variance

calibration

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Program ISD

82

SH3 (Campbell):

- 150 NOEs from perdeuterated domain
- sparse data set; standard structure calculation does not produce unique fold

Rieping, Habeck, Nilges, Science (2005)

Monday, November 28, 11

Comparison to standard result

83

Monday, November 28, 11

Distribution of σ in Ubiquitin and SH3

84

- Distributions for all parameters
- No fixed “weight” but distribution
 - “marginalization”: integration over all other parameters
 - coordinates
 - scale factor

Monday, November 28, 11

Distribution of σ in Ubiquitin and SH3

84

- Distributions for all parameters
- No fixed “weight” but distribution
 - “marginalization”: integration over all other parameters
 - coordinates
 - scale factor

$$P(\sigma | D, I) = \int d\theta d\gamma P(\sigma | \theta, \gamma) P(\theta, \gamma | D, I)$$

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Computational requirements

85

- a few days on 50 Linux PCs
 - every “supertransition” is 50 short dynamics trajectories
 - in total, > 25000000 hybrid Monte Carlo steps
 - convergence of *distribution*, not only structures

Monday, November 28, 11

- CNS (refinement program for crystallography and NMR):
- <http://cns-online.org/v1.3/>
- ARIA (automated NOESY assignment and structure calculation):
- <http://aria.pasteur.fr/>
- ISD (Bayesian structure calculation and analysis):
- <http://www.isd.bio.cam.ac.uk/isd/>

Monday, November 28, 11

Literature: modelling, x-plor, CNS

- Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983). CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* 4, 187–217.
- Brunger, A.T. (1992). X-PLOR. A System for X-ray Crystallography and NMR. New Haven: Yale University Press.
- Brunger A.T., Clore, G.M., Gronenborn, A.M. & Karplus, M. (1986). Three-Dimensional Structure of Proteins Determined by Molecular Dynamics with Interproton Distance Restraints: Application to Crambin. *Proc. Natl. Acad. Sci. U.S.A.* 83, 3801–3805.
- Brunger, A.T., Kuriyan, J. & Karplus, M. (1987a). Crystallographic R Factor Refinement by Molecular Dynamics. *Science* 235, 458–460.
- A.T. Brunger, P.D. Adams, G.M. Clore, P. Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography & NMR System (CNS), A new software suite for macromolecular structure determination, *Acta Cryst.* D54, 905–921 (1998)

Monday, November 28, 11

Literature: reviews, NMR calculations

88

- Braun, W. Distance geometry and related methods for protein structure determination from NMR data. *Quart. Rev. Biophys.* 19:115-157, 1987.
- Brünger, A. T. and Nilges, M. Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR spectroscopy. *Quart. Rev. Biophys.* 26:49-125, 1993.
- Güntert, P. Structure calculation of biological macromolecules from NMR data. *Quart Rev Biophys* 31:145-237, 1998.
- Nilges, M. and O'Donoghue, S. I. Ambiguous noes and automated noe assignment. *Progr. Nucl. Magn. Reson. Spectrosc.* 32:107-139, 1998.
- Güntert P. Automated NMR structure calculation with CYANA. *Methods Mol Biol.* 2004;278:353-78

Monday, November 28, 11

Literature: Bayesian

89

- Rieping W, Habeck M, Nilges M. (2005) Inferential structure determination. *Science*, 309:303-306
- Habeck M, Nilges M, Rieping W. (2005) Bayesian inference applied to macromolecular structure determination. *Physical Reviews E*
- Habeck M, Rieping W, Nilges M. (2005) Bayesian Estimation of Karplus Parameters and Torsion Angles from Three-bond Scalar Couplings Constants. *J Magn Reson*
- Habeck M, Nilges M, Rieping W. Replica-exchange Monte Carlo scheme for bayesian data analysis. *Phys Rev Lett.* 2005 Jan 14;94(1):018105.
- Rieping W, Habeck, M, Nilges, M (2006). Refinement against NOE intensities using a lognormal distribution improves the quality of NMR structures. *JACS*,
- Nicastro G, Habeck M, Masino L, Svergun, D, Pastore, A. *J Biomol NMR* 2006, 36, 267–277.
- Bayrhuber M, ..., Habeck M ... et al. (2008). Structure of the human voltage-dependent anion channel. *Proc. Natl. Acad. Sci. USA*, 105: 15370–5.

Monday, November 28, 11

- <http://aria.pasteur.fr/documentation/courses/saclay-november-2011/unfolding.tar.gz>
- <http://aria.pasteur.fr/documentation/courses/saclay-november-2011/nmrcalc.tar.gz>

Monday, November 28, 11

Introduction to CNS

Monday, November 28, 11

CNS overview

- X-PLOR minimizes the hybrid energy function

$$E_{\text{hybrid}} = E_{\text{phys}} + w_{\text{exp}} E_{\text{exp}}$$
- where E_{phys} could be
 - a molecular dynamics force field (CHARMM, AMBER, OPLS/ AMBER)
 - a modified/ geometric force field (Engh/Huber, PROLSQ, PARALLHDG)
 - a distance geometry target function
- and E_{exp} would be derived from:
 - X-ray data
 - NMR data (distances, NOE volumes, torsion angles)
 - other (e.g. positional restraints)
 - ...

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Crystallography & NMR System

Main Menu

- ▶ About CNS
- ▶ Download
- ▶ Installation
- ▶ Getting started
- ▶ Input files
- ▶ Modules
- ▶ Libraries
- ▶ Utilities
- ▶ Tutorial
- ▶ Syntax Manual
- ▶ CNS wiki

Version: 1.3

Patch level: 0

Status: general release

Authors:

A.T.Brunger, P.D.Adams, G.M.Clore, W.L.Delano, P.Gros, R.W.Grosse-Kunstleve, J.-S.Jiang, J.M.Krahn, J.Kuszewski, M.Nilges, N.S.Pannu, R.J.Read, L.M.Rice, G.F.Schroeder, T.Simonson, G.L.Warren

Copyright (c) 1997-2010 Yale University License

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

About CNS

Description

Crystallography & NMR System (CNS) is the result of an international collaborative effort among several research groups. The program has been designed to provide a flexible multi-level hierarchical approach for the most commonly used algorithms in macromolecular structure determination. Highlights include heavy atom searching, experimental phasing (including MAD and MIR), density modification, crystallographic refinement with maximum likelihood targets, and NMR structure calculation using NOEs, J-coupling, chemical shift, and dipolar coupling data.

Please cite the following two references when using CNS:

1. A.T. Brünger, P.D. Adams, G.M. Clore, P. Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, N. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography & NMR System (CNS), A new software suite for macromolecular structure determination, *Acta Cryst. D* **54**, 905-921 (1998).

2. A.T. Brünger, Version 1.2 of the Crystallography and NMR System, *Nature Protocols* **2**, 2728-2733 (2007).

◀ Main page

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Input files

- ▶ Edit file
- ▶ General
- ▶ Auxiliary
- ▶ X-ray
- Heavy atom
- Phasing
- Mol Rep
- Refinement
- Utilities
- Twinning
- PDB submission
- mmCIF deposit
- ▶ NMR
- Structure calc

Defaults - Off

- ▶ Set
-
- ◀ Main page

Edit file

Enter the name of your CNSsolve input file:

no file selected

General

alternate.inp

Generate alternate conformations or multiple conformers

average_coordinates.inp

Computes average coordinates from an ensemble of coordinates.

bfactor_plot.inp

Calculate B-factor distribution for a structure

buried_surface.inp

Lee & Richards buried surface accessibility calculation

cis_peptide.inp

Identify cis-peptides in a protein structure

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

- Minimization:
 - Powell's conjugate gradient minimization
- Molecular dynamics:
 - (numerical solution of Newton's equations of motion) with temperature variation (simulated annealing)
- Torsion angle dynamics
- Rigid body minimization
 - (with Powell's method)
- Grid search
 - through command language
- Monte Carlo simulated annealing through command language

- the coordinates
- some other properties
 - occupancies
 - temperature factors

- conformational energy
- deviations from ideal geometry
- deviations from experimental data: R-values
- Crossvalidated R-values (free R-factor)

- With E_{phys} alone, CNS can be used for
 - Energy minimization
 - MD calculations
 - MD analysis (correlation functions)
 - Free energy calculations
(perturbation method, thermodynamic integration)

CNS is an interactive program

- typing “cns” generates the prompt `cns>`
- there is some online help
`cns> help`
 produces a list of available commands
- CNS has a powerful interpreted command language
 - variables (real and string)
 - if-statements
 - for- and while-loops
 - “vector” manipulations (e.g. coordinates) – data manipulations
 - mathematical functions
 - example:


```
evaluate ($count = $count + 1)
if ($count > 5) then
  do (x = ran()*y^2)
end if
```

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

- commands are in general not case sensitive:
`HELP = hElp`
- commands can go over several lines
- some commands have fixed number of parameters:
`ASSI (resi 1 and name ha) (resi 1 and name hn) 3.0 1.0 1.0`
- others end with “end”
`noe scale dist 50 end`
- usually, 4 characters (sometimes 3 or 5) are sufficient:
`ASSI = ASSIG = ASSIGN`

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

- 3 types of comment lines:
- exclamation mark !
the rest of the line is ignored example

```
coor    ! this is a comment...
```
- curly brackets
the contents of the { } is ignored; example

```
dynamics verlet init{ial}t = 1000 end
```
- REMARK
the line beginning with REMARK is ignored, but stored and written to the next output file (especially coordinate file)

Opening Files: @ and @@

- In general, files are opened with @ or @@.
- Both switch the “command stream” to the file.
- @-files are stored on internal command buffer (for loops or if-statements) and are only opened once in a loop
- @@-files are only parsed and cannot contain loops or if-statements
- “@ for command files – @@ for data files”
- Warning: some commands expect only the file name:

```
read trajectory  
    input = coords.crd  
end
```

- Filenames are case sensitive (UNIX), and can be specified with absolute or relative path:
 - `@@../../parallhdg.pro`
 - `@@/data4/Users/nilges/toppar2/parallhdg.pro`
- Environment variables can be used
 - `@@TOPPAR:parallhdg.pro`
- where TOPPAR has been defined by
 - `setenv TOPPAR /data4/Users/nilges/toppar2`

Topology and the PSF file

- Topology file: residue library that defines standard molecule components (amino acids, nucleotides):
- atom definitions (masses)
residue definitions (covalent topology) • charges
patches (“presidue”) for modifications:
 - peptide bond
 - disulfide bridge
 - N– and C–termini
- no coordinates!
no bond lengths, force constants etc!
- The topology of a specific molecule is stored in the PSF:
 - Sequence + Topology → PSF

Example of a residue in TOPALLHDG.PRO

106

```
residue ALA
  group
    atom N    type=NH1 charge=-0.36 end
    atom HN   type=H   charge= 0.26 end
    atom CA   type=CT   charge= 0.00 end
    atom HA   type=HA   charge= 0.10 end
    atom CB   type=CT   charge=-0.30 end
    atom HB1  type=HA   charge= 0.10 end
    atom HB2  type=HA   charge= 0.10 end
    atom HB3  type=HA   charge= 0.10 end
    atom C    type=C    charge= 0.48 end
    atom O    type=O    charge=-0.48 end
    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
    improper HA N C CB !chirality CA
    improper HB1 HB2 CA HB3 !methyl group CB
  end
```

Monday, November 28, 11

generate.inp

107

```
topology
  @@TOPPAR:topallhdg.pro
end
segment
  name="      "
  chain
    @@TOPPAR:toph19.pep
    sequence
      Ala Ala end
    end end
  REMARK ALA dipeptide
  write structure
    output=INPUT:diala.psf
  end
stop
```

Monday, November 28, 11

- “segment” defines a new segment of the molecular structure.
- several segments possible (e.g. in complexes or multimers)
– protein – DNA – water
- the name of the segment corresponds to the PDB coordinate file (last 4 characters before card number)
- “chain” concatenates residues, with definitions in file tophl9.pep
- “sequence” specifies the sequence
- “sequence ... end” can be replaced by “coor @@example.pdb” (note: “end” in coor file!)
- the REMARKs will be written to the PSF file
- PSF file is written by `WRITE PSF ... END`

Monday, November 28, 11

Example of a patch in TOPALLHDG.PRO

```

presidue PEPT
  add bond -C +N
  add angle -CA -C +N
  add angle -O -C +N
  add angle -C +N +CA
  add angle -C +N +HN
  add improper -O -C +N +CA
  add improper +HN +N -C -CA
  add improper -CA -C +N +CA
end

```

Monday, November 28, 11

Description of PSF file

- Note: usually no need to look at the file – do not modify
- Header: REMARK records
Filename,date etc are generated by WRITE

PSF

```

      3 !NTITLE
REMARKS  FILENAME="diala.psf"
REMARKS  ALA dipeptide
REMARKS  DATE:07-Sep-95  10:16:16   created by ...

```

- list of all atoms

```

      23 !NATOM
          1      1      ALA  CA   CT      0.220000      1
          2      1      ALA  HA   HA      0.100000      1
      ...
          22      2      ALA  OT1  OC      -0.570000      1
          23      2      ALA  OT2  OC      -0.570000

```

Monday, November 28, 11

- list of all bonds

```

      22 !NBOND: bonds
      9 1 1 2 1 3
      3 5 3 6 1 7
      ...
      21 22 21 23

```

- same for
 - bond angles
 - dihedrals
 - impropers
 - hydrogen bond donors and acceptors – non-bonded groups

Monday, November 28, 11

A simple energy minimization

- To minimize, we need
- PSF file
- energy parameters
- starting coordinates (X-PLOR PDB format)

```
structure @@diala.psf end
parameter @@TOPPAR:parallhdg.pro end
coord @@diala.pdb end
mini powell nstep= 50 end
REMARK after 50 steps powell
write coord output=diala_min.pdb end
stop
```

Monday, November 28, 11

CNS scripting language

Monday, November 28, 11

- variables (symbols)
- if-statements
- loops
- atom selection
- data structure manipulations
- many application statements
- mathematical functions
for variable and data structure manipulations

Symbol definitions and the EVALuate statement

- recognized by \$ sign
- symbols are defined and manipulated by EVALuate

```
evaluate ($count = 0)
evaluate ($filename = "dg.pdb")
```
- Symbols can be
 - real numbers
 - strings
 - logical
- the type definition is implicit by usage
- type conversion by encode and decode

```
evaluate ($name = encode($count))
evaluate ($number = decode($name))
```
- \$? produces list of all defined symbols

Arithmetic operations

- Standard operations

`+ - * / ** ^ ( )`

```
evaluate($number = (5*$count)^(3+$count))
```

- mathematical functions

`cos sin ran ...`

```
evaluate($number = sin( $count*ran() ))
```

Special symbols

- Fundamental constants `$pi` `$boltz`

- Results of certain operations (incomplete list) – `PRINT` statements define `$result`

```
print angle
```

```
evaluate ($rms_angle = $result)
```

- `SHOW` statements define `$result`

```
show average (x) (all)
```

```
evaluate ($x_ave = $result)
```

- `ENERgy`, `MINImiz` and `DYNAmics` define energy terms

```
energy end
```

```
display $ener $bond $angl
```

- basic structures:
 - IF (condition) THEN commands END IF
 - IF (condition) THEN commands ELSE commands END IF
 - “case” statement
 - IF (condition)
 - THEN commands
 - ELSEIF (condition)
 - THEN commands ...
 - END IF
- can be nested
- note: ELSEIF is not ELSE IF

- two “end if” necessary

```
if ( $count eq 1 )
  then
    coor copy end
  else
    if ($count eq 2)
      then
        coor fit end
      end if
    end if
```
- one “end if” necessary

```
if ( $count eq 1 )
  then coor copy end
elseif ($count eq 2)
  then coor fit end
end if
```

- WHILE loop:
WHILE (condition) LOOP loop-name
- commands
END LOOP loop-name

```
evaluate ($count = 1)
while ($count le 10) loop main
    evaluate ($count = $count + 1)
end loop main
```

- FOR loop 1:
FOR variable IN (set)

```
for $filename in ( "sa_1.pdb" "sa_3.pdb" )
    loop main
        coor @@$filename
    end loop main
```
- FOR loop 2:
FOR variable IN ID (selection)

```
for $loopid in id (all) loop main
    vector show element (x) (id $loopid)
end loop main
```

- elect atoms for certain operations
- selection by atom name
- wildcards and ranges
- selection by atom property
- different “queries” can be connected by AND / OR
- “queries” can be negated by NOT
- parantheses necessary for combinations of AND, OR, NOT

Selection by atom name

- The atom name consists of
 - – SEGId, segment name defined by SEGMENT
 - – RESId, residue “number” (also 48b etc!)
 - – RESName, residue name (ALA, VAL...)
 - – NAME, atom name (N, CA...)
- ```
coor select (resid 5 and name hn) ... end
coor select ((resid 5 or resid 7) and name hn) ..
coor select (resid 5:7 and not name h*) ... end
```
- Atoms can also be selected by
    - CHEM (atom type defined in topology)
    - ID (internal number)

## Wildcards and ranges

---

- wildcards and ranges can be used for
  - SEGI
  - RESI
  - RESName – NAME
  - CHEM
- ranges are lexicographical order indicated by ":"
  - `coor sele (name ha:hg#) ... end`
  - ... selects ha, hb1, hb2, hg1, hg2
- wildcard hierarchy
  - "\*" any string (abcd, 78, 8u)
  - "#" any number (2, 43, 39987)
  - "%" any character (a, 6, j)
  - "+" any digit (0, 1, ... 9)

## Selection by atom property

---

- ATTRibute selects on any atom property
  - coordinates, derivatives, mass, charge, ...)
  - `coor sele (attribute charge > 0) ... end`
- AROUnd, SAROund select atoms within cutoff of specified atoms
  - `coor sele ((resi 1 and name ca) around 5.0) ... end`
- SAROund selects atoms also in symmetry mates
- POINt ... CUT selects atoms around point
  - `coor sele (point (3.0 4.0 5.0) cut 5.0) ... end`

## Selection by residue etc

---

- **BYResidue (selection)** selects all atoms in a residue  

```
coor sele= (byres(point (0 0 0) cut 5.0))
```

 selects all atoms in residues that have at least one atom in a sphere around the origin


```
coor sele= (bygrp(resid 1 and name ca))
```

STOREi and RECALLi

- **Atom selections can be stored and used later**

```
iden (store1) (name ca)
```

```
coor sele= (store1) ...
```

```
coor sele= (recall1) ...
```

Vector manipulations

- **SHOW** and **DO** allow analysis and manipulation of atom properties and names.
  - **SHOW ELEMent** (AtomArray) (selection)
lists elements and defines \$result
  - **SHOW AVERage** (AtomArray) (selection)
  - **SHOW RMS** (AtomArray) (selection)
  - **SHOW SUM** (AtomArray) (selection)
  - **SHOW NORM** (AtomArray) (selection)
show element (resid)
(name ca and (resid 5 and name ca) around 5.0)
show average (x) (name ca)
- **DO** (expression) (selection)
vector do (b = b + x^2 + y^2 + z^2) (all)
- **IDEN** (STOREi) (selection)
defines a **STORE** to be used in atom selection later

Monday, November 28, 11

3D vectors and matrices

- **3D vectors** can be defined explicitly, or through atom selections

```

coor translate vector= (1 0 0) end
coor translate
 vector= (head=(resid 1 and name cb)
 tail=(resid 1 and name ca))
 distance= 5.0 end

```
- **3x3 matrices** can be defined by rotation center, axis, angle

```

coor rotate
 center= (0 0 0)
 matrix= AXIS (head=(resid 1 and name cb)
 tail=(resid 1 and name ca)) 90.0

```

  - or by Euler angles, Lattman angles, Quaternions, Spherical angles
  - or explicitly

```

coor rotate
 center= (0 0 0)
 matrix= (1 0 0) (0 1 0) (0 0 1) end

```

Monday, November 28, 11

- **DISPLAY files**
  - for DISPlay statements
  - open with SET DISPlay filename END
- **PRINT files**
  - for info from PRINt statements (e.g. PRINt ANGLeS)
  - open with SET PRINt filename END
- **coordinate, structure, parameter files**
  - with WRITE COOR (structure...) OUTPut= file- name end
- **trajectory files**

Monday, November 28, 11

Examples

```
set display rmsd.disp end
evaluate ($maxcount = 10)
evaluate ($count = 1)
for $filename in (@@file.list) loop fit
 coor @@$filename
 if ($count eq 1) then
 coor copy end
 end if
 coor sele (name ca) fit end
 coor sele (name ca) rms end
 display $count $filename rms $result A
 if ($count ge $maxcount) then
 exit loop fit
 end if
 evaluate ($count = $count + 1)
end loop fit
```

Monday, November 28, 11

- The file “file.list” contains a list of files, for example ordered by energy
  - "sa_1.pdb"
  - "sa_6.pdb"
  - ...
  - "sa_67.pdb"
- The display file rmsd.disp will look like this


```
1 sa_1.pdb rms 0 A
2 sa_6.pdb rms 1.245 A
...
10 sa_67.pdb rms 1.87 A
```

Monday, November 28, 11

```
set display rmsfluc.disp end
for $loopid in id (name ca) loop rms
 vector show element (resid) (id $loopid)
 evaluate ($resid = $result)
 vector show element (resn) (id $loopid)
 evaluate ($resn = $result)
 vector show norm (b) (byresidue(id $id) and not
 hydrogen)
 evaluate ($rmsfluc = $result)
 display $resn $resid $rmsfluc
end loop rms
```

Monday, November 28, 11

Energy minimization and molecular dynamics

Energy minimization (conjugate gradient)

135

- Conjugate gradient minimization (Powell method)
  - uses gradient information
  - a “complete” minimization is a series of one– dimensional minimizations (one for each degree of freedom)

- General syntax of minimization command
- started by MINIMIZE POWELL
- Minimization is performed until one convergence criterion is met.
  - NSTEP: maximum number of steps
  - TOLGradient: target norm of gradient
- Other parameters:
  - DROP: expected initial drop in energy (default 0.001, optimal value 10...100)
  - NPRINT: Information is printed every NPRINT steps
- Minimization defines variables \$ener, \$grad, \$bond... of energy terms that are turned on with FLAG statement
- Minimization often terminates with “Line search abandoned”.

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

An energy minimization script

```
structure @@protein.psf end
coord @@protein.pdb
parameters @@TOPPAR:parallhdg.pro
 nbonds repel= 0.78 rcon = 5.0 end
end
flags exclude elec include harm end
evaluate ($kharm = 10)
while ($kharm ge 0) loop mini
 vector do (harm = $kharm) (all)
 mini powell
 drop 10 nstep 100 nprint 10
end
 evaluate ($kharm = $kharm - 1)
end loop mini
write coord output = protein_m.pdb end
stop
```

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

Rigid body minimization

- started by MINImize RIGID
- Minimization is performed until one convergence criterion is met.
- same parameters as POWELL:
  - NSTEp, TOLGradient, DROP, NPRInt
  - rigid groups are defined by
group = <selection>, for example

```
mini rigid
 nstep 100
 group (segid A) group (segid B)
end
```

Example script using rigid body minimization

```
structure @@protein.psf end
structure @@DNA.psf end
coord @@protein.pdb
coord @@dna.pdb
NOE
 @@dock.tbl
end
flags exclude * include NOE end
constraints fix (segid "PROT") end
minimize rigid
 group (segid "1BNA")
 nstep 100
end
write coord
 sele= (segid "1BNA") output = dna_dock.pdb
end
```

Cartesian molecular dynamics

140

- Invoked by
  - dynamics cartesian ... end
- Important parameters:
  - nstep: number of steps
  - timestep: time step in ps
  - tcoup: switch on Berendsen's method? true/false
  - tbath: temperature of heat bath
  - nprint: print frequency
  - cmremove: remove COM movement? true/false
  - cmperiodic: period of COM movement removal
- initial velocities defined with "do (vx = maxwell(300)) (all)"
- coupling parameter fbeta defined with "do (fbeta = 10) (all)"

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11

A slow cooling script

141

```
evaluate ($bath = 1000)
do (fbeta = 10) (all)
do (vx = $bath) (all)
do (vy = $bath) (all)
do (vz = $bath) (all)
while ($bath > 50) loop cool
 evaluate ($bath = $bath - $tempstep)
 dynamics verlet
 nstep=1000 time=0.005
 tcoup=true temperature=$bath
 nprint=$nstep
 cmremove=true cmperiodic=0
end
end loop cool
```

Michael Nilges. Molecular Modelling Applied to NMR Structure Determination. Saclay 11/2011

Monday, November 28, 11