

NMR Strukturbestimmung

Peter Güntert

Inhalt

1. Konformationsdaten aus NMR Messungen
2. Strukturbeschreibung mit Atomkoordinaten oder Torsionswinkeln
3. Strukturbestimmung: grundsätzliche Möglichkeit, praktische Schwierigkeiten
4. Strukturberechnungsalgorithmen: interaktiver Modellbau, Distanzgeometrie, Optimierung einer Zielfunktion, Simulated Annealing, Moleküldynamiksimulation, Torsionswinkeldynamik
5. Darstellung von NMR Strukturen: Strukturbündel, RMSDs
6. Automatisierung der NOE-Zuordnung und Strukturberechnung
7. Strukturbestimmung mit Festkörper-NMR: Alzheimerpeptid-Fibrillen
8. Informationsgehalt von Distanzeinschränkungen
9. Strukturanalyse: Übereinstimmung mit experimentellen Daten, Verletzungen von Konformationseinschränkungen, RMSD, Ramachandranplot, Validierung

Konformationsdaten aus NMR Messungen

Konformationsdaten aus NMR Messungen

1. NOEs
2. 3J skalare Kopplungen
3. H-Brücken
4. Chemische Verschiebungen
5. Residuelle dipolare Kopplungen (RDC)
- ...

NOE (Nuclear Overhauser Effect)

NMR Daten: Integral V von NOESY Kreuzsignalen
Konformationsdaten: obere Schranken für 1H - 1H Distanzen, d
Für isoliertes Spinpaar im starren Molekül:

$$V = C/d^6 \quad \text{mit } C = \text{konstant}$$

Eigenschaften:

- nur kurze Distanzen $< 5 \text{ \AA}$ messbar
- dichtes Netzwerk bzgl. der Sequenz kurz- und langreichweitiger Distanzschranken
- viele 1H Atome im Molekül $\rightarrow$ "Spindiffusion"
- interne Bewegungen $\rightarrow$ nicht-lineare Mittelung
- Bestimmung der Konstanten C ?
- Überlapp $\rightarrow$ mehrdeutige Zuordnung, verfälschte Integrale

$\rightarrow$ Verwendung als obere Distanzschranken

NOE Calibration

Volume of
NOESY
cross peak

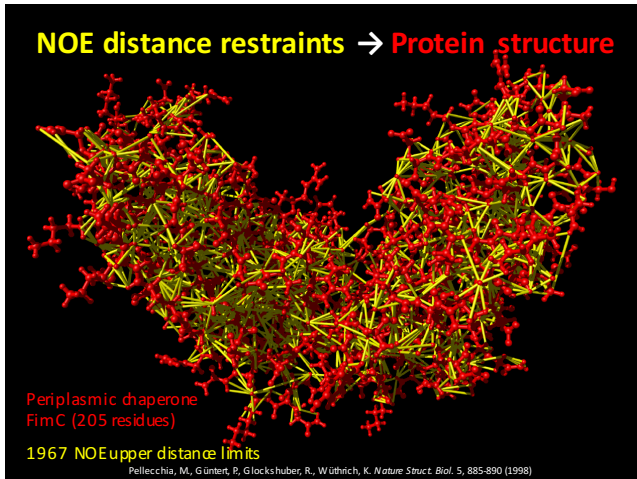
$$V = C / d^6$$

"Calibration
constant"

Distance
(upper
distance
bound)

How to set the calibration constant?

- Known distances (intrareidual or in standard secondary structures)
- Preliminary structure, if available
- User-defined value for the average (median) upper distance limit



Problems when interpreting NOEs

- Internal motion
- Spin diffusion
- Spectral overlap
- Chemical shift degeneracy
- Time consuming spectral analysis, if done manually → **automation**

NMR resonance assignment is like solving a puzzle...

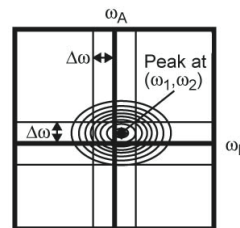
...with missing pieces
(incomplete signals)



...with additional pieces
(artifacts)

...in the mist
(low signal-to-noise, line-broadening)

Ambiguity of chemical shift based NOE assignment



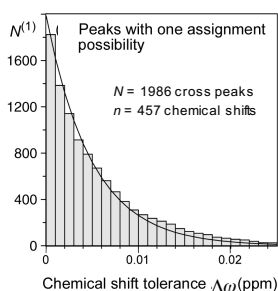
In general, several different ^1H chemical shifts ω_A, ω_B match the position of a NOESY peak within the experimental uncertainty $\Delta\omega$.

→ Assignment ambiguity

Manual assignment is very cumbersome!

$$|\omega_1 - \omega_A| < \Delta\omega \quad |\omega_2 - \omega_B| < \Delta\omega$$

NOEs with a unique chemical shift based assignment



2D NOESY:

$$N^{(1)} \approx N \exp(-4n \Delta\omega / \Delta\Omega)$$

3D NOESY:

$$N^{(1)} \approx N \exp(-2n \Delta\omega / \Delta\Omega)$$

$N^{(1)}$ Number of uniquely assigned peaks

N Number of cross peaks

n Number of chemical shifts

$\Delta\omega$ Chemical shift tolerance

$\Delta\Omega$ Spectrum width

Ambiguous distance restraints

$$d_{\text{eff}} = \left(\sum_k d_k^{-6} \right)^{-1/6} \leq b$$

upper distance bound
distance for assignment possibility k
sum over all assignment possibilities

- Restraint with multiple assignments
- If one assignment possibility leads to a sufficiently short distance, then the ambiguous distance restraint will be fulfilled.

→ The presence of wrong assignment possibilities has no (or little) influence on the structure, as long as the correct assignment possibility is present.

Nilges et al., *J. Mol. Biol.* 269, 408-422 (1997)

Properties of ambiguous distance restraints

$$d_{\text{eff}} = \left(\sum_k d_k^{-6} \right)^{-1/6}$$

- d_{eff} is never longer than any of the individual distances d_k :
 $d_{\text{eff}} \leq d_k$ for all k
- d_{eff} is close to the smallest individual distance:
 $d_{\text{eff}} \approx d_1$ if $d_1 \ll d_2, d_3, \dots$
- Examples: $d_1 = 3 \text{ \AA}, d_2 = 10 \text{ \AA} \rightarrow d_{\text{eff}} = 2.9996 \text{ \AA}$
 $d_1 = 3 \text{ \AA}, d_2 = \dots = d_{10} = 10 \text{ \AA} \rightarrow d_{\text{eff}} = 2.9967 \text{ \AA}$

3J skalare Kopplungen

NMR Daten: Aufspaltung eines Signals
 Konformationsdaten: Einschränkungen von Torsionswinkeln, θ

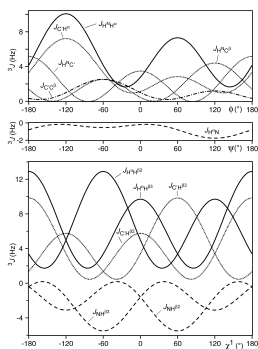
Karplus-Kurve: $^3J(\theta) = A \cos^2 \theta + B \cos \theta + C$
 mit empirischen Konstanten A, B, C

Zum Beispiel: $^3J_{\text{HNH}\alpha}(\phi), ^3J_{\text{H}\alpha\text{H}\beta}(\chi^1)$

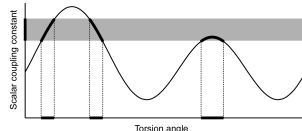
Eigenschaften:

- Information nur über lokale Konformation
- mehrdeutige Beziehung $^3J \leftrightarrow \theta$

3J scalar couplings



- $^3J(\theta) = A \cos^2 \theta + B \cos \theta + C$
- local information only
- ambiguous relation to torsion angle



H-Brücken

NMR Daten: langsamer $^1\text{H} \rightarrow ^2\text{H}$ Austausch + NOEs

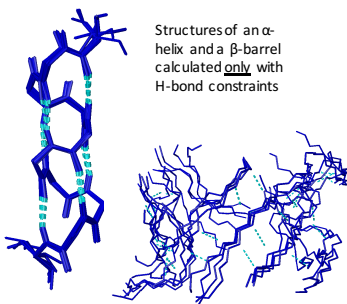
Konformationsdaten: Donor-Akzeptor Distanz

Typische H-Brücken: $\text{-N-H} \cdots \text{O=C-}$ in regulären Sekundärstrukturen (Helices, β -Blätter)

Eigenschaften:

- Bzgl. Sequenz mittel- und langreichweitig
- Donor (H) identifizierbar
- Akzeptor (O) i. A. nur indirekt bestimmbar (benachbarte NOEs + Annahmen über Sekundärstruktur)

Impact of hydrogen bond restraints



- Strong impact on structure
- Direct detection of H-bonds by NMR is possible, but not sensitive
- Without identification of acceptor atom $\approx$ assumption on secondary structure

Chemische Verschiebungen

NMR Daten: chem. Verschiebungen, δ

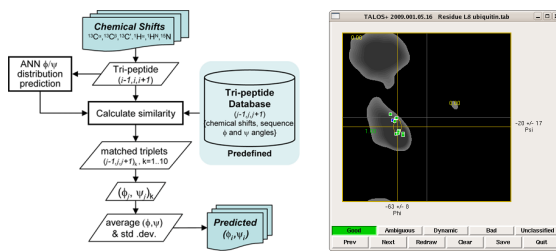
Konformationsdaten: (ϕ, ψ) Torsionswinkelbereiche

Komplexe Beziehung: $\delta \leftrightarrow (\phi, \psi)$

Eigenschaften:

- einfache Messung
- (ϕ, ψ) -Werte aus Datenbank von Proteinen mit bekannter Struktur und chem. Verschiebungen (TALOS)
- Information über lokale Konformation bzw. Sekundärstruktur

TALOS+: Torsion angle restraints from chemical shifts



Reliability of TALOS+ torsion angle predictions:

- On average, TALOS+ makes consistent predictions for about 88% of the residues.
- Over all 200 database proteins, about 2.5% of the unambiguous predictions made by TALOS+ were incorrect relative to the corresponding crystal structure. However, a substantial fraction of this 2.5% appears to reflect genuine differences relative to the crystalline state, and the true error rate therefore is believed to be below 2.5%.
- On average, the uncertainty as reported by TALOS+ for the consensus predictions was 12.6° for ϕ , and 12.3° for ψ .
- The actual RMSD of the "correct" predictions relative to the crystal structures was about 13.5° for ϕ , and 12.9° for ψ .

Residuelle dipolare Kopplungen (RDC)

NMR Daten: Zusätzliche Signalaufspaltung bei partieller Molekülausrichtung.
z.B. $^1J_{NH} \rightarrow ^1J_{NH} + D_{NH}$

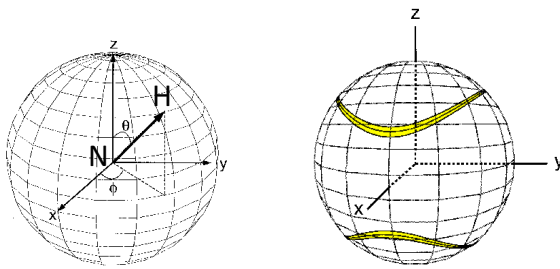
Konformationsdaten: Orientierung von Bindungen relativ zur Molekülausrichtung

Residuelle dipolare Kopplung: $D(\theta, \phi) = A [(3 \cos^2 \theta - 1) + 3/2 R \sin^2 \theta \cos 2\phi]$
A, R Amplitude (Betrag) und Rhombizität (Abweichung von Rotationssymmetrie) des Ausrichtungstensors
 θ, ϕ Richtung der Bindung relativ zum Ausrichtungstensor (Polarkoordinaten)

Eigenschaften:

- Proteinprobe in schwach ausrichtendem Medium (Flüssigkristalle/Bizellen, fadenförmige Phagen, komprimierte Gele)
- Information über globale Konformation, z.B. relative Ausrichtung von Domänen
- Entartung: 1 Messwert $\rightarrow$ Doppelkegel von Richtungen
- Bestimmung des Ausrichtungstensors (A, R)?

Residuelle dipolare Kopplungen



$$D(\theta, \phi) = A [(3 \cos^2 \theta - 1) + 3/2 R \sin^2 \theta \cos 2\phi]$$

Strukturberechnungs- algorithmen

Ist NMR Strukturberechnung möglich?

- Grundsätzlich:
 - NOEs messen nur kurze Distanzen $< 5 \text{ \AA}$
 - ungenau obere Schranken
 - Kann damit die globale Struktur eines 30 $\text{\AA}$ langen Proteins bestimmt werden?
JA, wenn genügend Daten vorhanden sind.
- Praktisch:
 - Zielfunktion hat viele lokale Minima
 - Kann eine (fast) optimale Struktur gefunden werden?
JA.

Strukturberechnungsalgorithmen

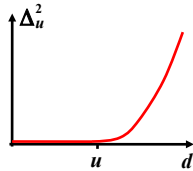
- Frühere Methoden:
 - Interaktiver Modellbau
 - Distanzgeometrie
 - Minimierung einer variablen Zielfunktion
- Simulated annealing:
 - Monte Carlo
 - Moleküldynamiksimulation im kartesischen Raum
 - Moleküldynamiksimulation im Torsionswinkelraum

CYANA target function

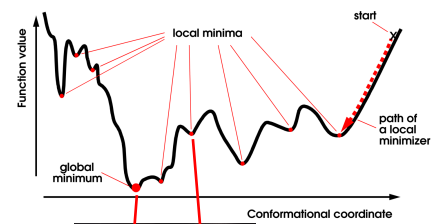
$$T = \sum_{\text{upper distance limits (NOEs)}} \Delta_u^2 + \sum_{\text{lower distance limits (steric)}} \Delta_l^2 + \sum_{\text{torsion angle restraints}} \Delta_a^2 + \dots$$

$\Delta_u, \Delta_l, \Delta_a$: restraint violations,

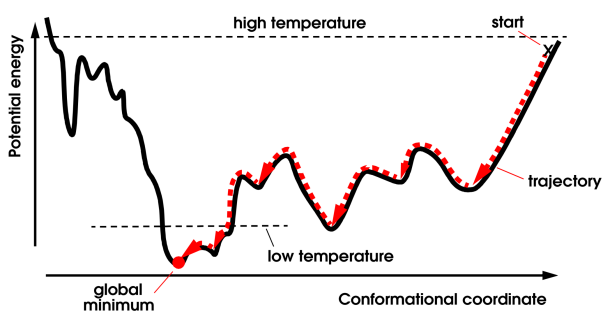
e. g., $\Delta_u = \begin{cases} d - u & \text{if } d > u \\ 0 & \text{otherwise} \end{cases}$



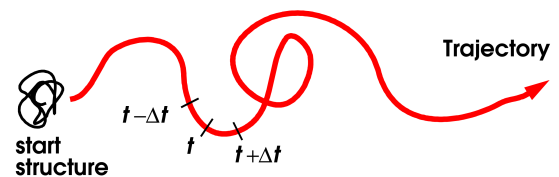
Target function = potential energy



Simulated annealing



Molecular Dynamics Simulation



Numerical integration of classical equations of motion

Integration of the equations of motion

e.g. "leap-frog" algorithm

$$q(t + \Delta t) = q(t) + \Delta t \dot{q}(t + \Delta t/2) + O(\Delta t^3)$$

$$\dot{q}(t + \Delta t/2) = \dot{q}(t - \Delta t/2) + \Delta t \ddot{q}(t) + O(\Delta t^3)$$

q coordinates (Cartesian or torsional)

$\dot{q} = \frac{dq}{dt}$ velocities

$\ddot{q} = \frac{d^2q}{dt^2}$ accelerations

Δt time step

Atomkoordinaten Torsionswinkel

Strukturbeschreibung

Atomkoordinaten (kartesische Koordinaten):

- 3 Freiheitsgrade pro Atom
- abhängig von der Wahl des Koordinatensystems
- beinhalten auch "unwichtige" Freiheitsgrade
- einfach

Torsionswinkel (= Diederwinkel, Dihedralwinkel):

- Drehungen um Einfachbindungen
- interne Koordinaten
- essentielle Freiheitsgrade
- Bindungslängen, Bindungswinkel fest
- kompliziertere aber effizientere Algorithmen

Torsionswinkel

- Definiert durch 4 Atome: A—B—C—D
- Drehung um Bindung B—C
- Werte von -180° bis $+180^\circ$

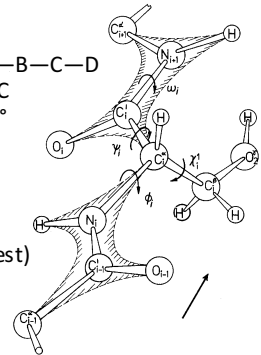
- Torsionswinkel von AS i :

$$\phi_i: C'_{i-1}-N_i-C^\alpha_i-C'_i$$

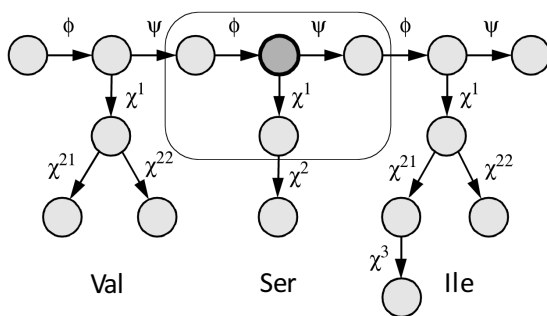
$$\psi_i: N_i-C^\alpha_i-C'_i-N_{i+1}$$

$$\omega_i: C^\alpha_i-C'_i-N_{i+1}-C^\alpha_{i+1} \text{ (fest)}$$

$$\chi^1_i: N_i-C^\alpha_i-C^\beta_i-C^\gamma_i$$



Torsionswinkel: Baumstruktur



MD Simulation im Torsionswinkelraum "Torsionswinkeldynamik"

- Klassische Mechanik
- N Torsionswinkeln als einzige Freiheitsgrade
- Etwa 10 Mal weniger Freiheitsgrade als im kartesischen Raum.
- Feste Bindungslängen und -winkel:
→ "Einfrieren" der schnellsten Bewegungen
→ Längere Zeitschritte

Jain, Valdehi, Rodriguez, *J. Comp. Phys.* 106, 258-268 (1993)
Güntert, Mumenthaler, Wüthrich, *J. Mol. Biol.* 273, 283-298 (1997)

Equations of motion

Cartesian coordinates: $x_1, \dots, x_N$

$$m_i \ddot{x}_i = - \frac{\partial E_{\text{pot}}}{\partial x_i} \quad \text{(Newton)}$$

Generalized coordinates: $q_1, \dots, q_n$

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) - \frac{\partial L}{\partial q_k} = 0 \quad \text{(Lagrange)}$$

with $L = E_{\text{kin}} - E_{\text{pot}}$

Molecular Dynamics

Cartesian space

$$E_{\text{kin}} = \frac{1}{2} \sum_{i=1}^N m_i \dot{x}_i^2$$

diagonal, constant
(elements m_i)

$$\ddot{x}_i = - \frac{1}{m_i} \frac{\partial E_{\text{pot}}}{\partial x_i}$$

proportional to N

Kinetic
energy

Mass matrix
 M

Accelerations

Computa-
tional
complexity

Torsion angle space

$$E_{\text{kin}} = \frac{1}{2} \sum_{k,l=1}^n M(\theta)_{kl} \dot{\theta}_k \dot{\theta}_l$$

non-diagonal,
non-constant, $n \times n$

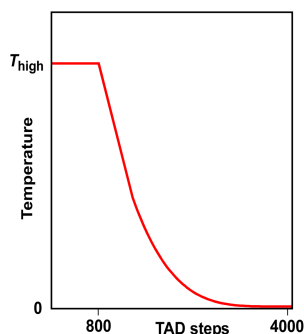
$$M(\theta) \ddot{\theta} = C(\theta, \dot{\theta})$$

solving linear system
of equations: $\sim n^3$

exploiting tree structure
of the molecule: $\sim n$

Simulated annealing protocol

- Start from random structure
- Use all restraints simultaneously
- Adjustable parameters:
 - start temperature, T_{high}
 - number of TAD steps



Temperature control

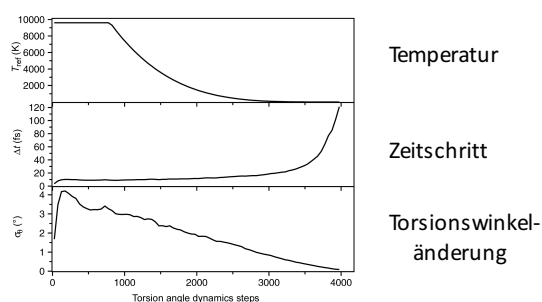
Weak coupling to a heat bath is used to control the temperature:

$$\dot{\theta} \leftarrow \dot{\theta} \sqrt{1 + \frac{T^{\text{ref}} - T}{\tau T}}$$

$\dot{\theta}$ torsional velocities
 T instantaneous temperature, $T = \frac{2E_{\text{kin}}}{nk_B}$
 τ coupling constant

(Berendsen et al., J. Chem. Phys. 81, 3684–3690, 1984)

Simulated annealing mit Torsionswinkeldynamik



Strukturbündel RMSDs

NMR Structure Calculation: Multiple Conformers



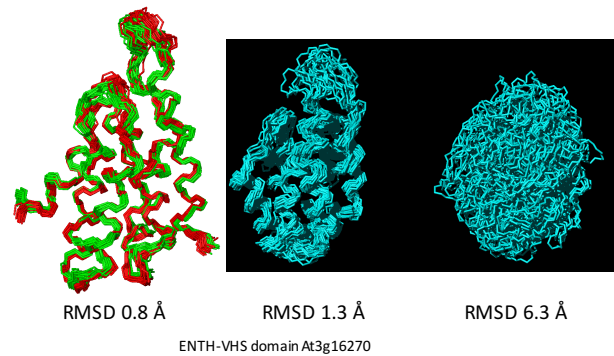
NMR Structure Calculation: Select Converged Conformers



Strukturbündel

- 100 Startstrukturen mit zufälligen Torsionswinkeln
- 100 unabhängige simulated annealing Läufe mit:
 - gleichen experimentellen Daten
 - unterschiedlichen Startstrukturen
- Auswahl der 20 "besten" Strukturen mit den tiefsten Zielfunktionswerten
- Sampling des Konformationsraums?

Strukturbündel



RMSD (root-mean-square deviation)

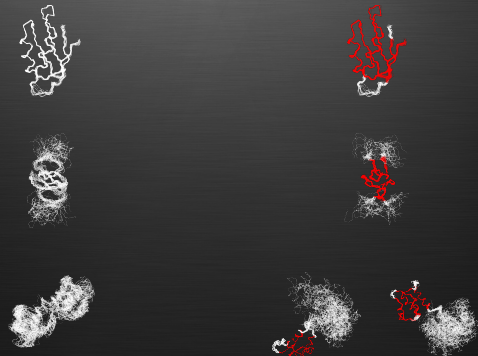
- Zwei Strukturen mit n Atomen und Koordinaten $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ und $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$

$$RMSD = \min_{R, \vec{t}} \sqrt{\frac{1}{n} \sum_{i=1}^n |\vec{x}_i - R\vec{y}_i - \vec{t}|^2}$$

- Minimum über alle Rotationen R und Translationen $\vec{t} \rightarrow$ optimale Überlagerung

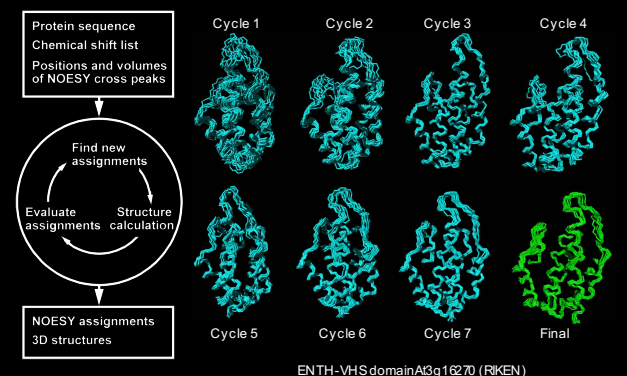
All Residues

CYRANGE Domains



Automatische NOE Zuordnung

Automated NOESY assignment and structure calculation with CYANA



Output overview table

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	391	3872	3950	4115	4195	4194	
with short-range assignment $ i-j < 5$	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	1303
medium-range, $1 < i-j < 5$	1220	959	977	75	75	76	76	76
long-range, $ i-j > 5$	559	451	529	574	569	650	650	652
Average assignments/restraint	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

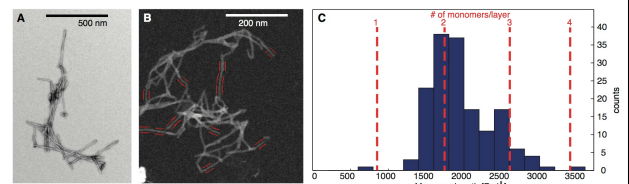
CYANA Computation Time

- Combined NOE assignment and structure calculation of a 114 amino acid residue protein with the program CYANA:
 - 8 cycles $\times$ 100 conformers = **800 structures**
 - 10000 torsion angle dynamics steps per conformer
- Linux cluster system with Quad-core Intel Xeon E5462 (2.8 GHz, 12 MB cache), 2 GB memory/core

Processors	Computation time (s)
100	147
50	217
25	354
10	769

Aβ1-40 E22Δ

Aβ1-40 E22Δ fibrils



Anne Schütz, Toni Vagt, Matthias Huber, Oxana Y. Ovchinnikova, Riccardo Cadalbert, Joseph Wall, Anja Böckmann, Rudi Glockshuber, Beat Meier

Aβ1-40 E22Δ fibrils: structure calculation

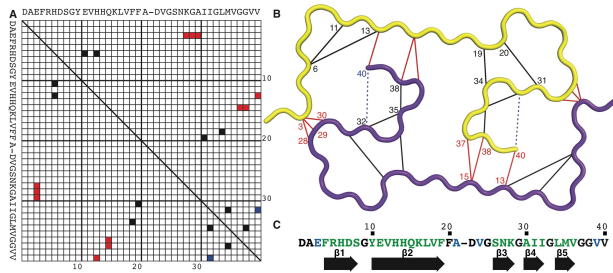
- 2 (or 3?) monomers per layer
- One set of resonances: symmetric arrangement
- System for calculation: 2 x 5 monomers
- β -sheet H-bonds along fibril axis
- NMR distance restraints intramolecular or intermolecular within one layer
- Symmetry restraints to keep all monomers identical and in symmetric relative orientation

**A β 1-40 E22 Δ fibrils:
manually assigned restraints**

DAR ¹			PAR ²			RDS ³		
α_1	α_2	α_3	α_1	α_2	α_3	α_1	α_2	α_3
R3-R28	R3	R28	R3-R28	R3	R28	A30-A30	A2	A30
R28	R3	R28	R28	R3	R28	A30	CA	A2
R3-R29	R3	R29	R3-R30	R3	R30	R3-R30	R3	R30
R29	R3	R29	A30	R3	CA	R3	A30	CA
R3-R30	R3	R30	HE-81	HE	HE	R3-R30	R3	R30
R30	R3	R30	R3	HE	HE	R3	A30	CA
R31-R40	R31	R40	P19-P34	P19	P34	R31-R40	R31	R40
R40	R31	R40	L34	P19	P19	R40	R31	R40
P19-P34	P19	P34	132-M35	132	M35	132-M35	M35	132
P34	P19	P34	132	M35	132	132	M35	132
HE-81	HE	HE	CHIC-7	α_1	α_2	P19-P34	P19	P34
HE	HE	HE	R3-R28	R3	R28	P19-P34	P19	P34
R28	HE	HE	HE-81	HE	HE	R3	R3	R28
P19-P34	P19	P34	R31-R40	R31	R40	P19-P34	P19	P34
P34	P19	P34	R3	HE	HE	R3	P19	P34
P20-P31	P20	P31	P19-P34	P19	P34	P20-P31	P20	P31
P31	P20	P31	132-M35	132	M35	132	M35	132
132-M35	132	M35	132-M35	M35	132	132-M35	M35	132
M35	132	M35	M35	132	M35	132	M35	132
R35-R38	R35	R38	PAIN⁴	α_1	α_2	134-G37	G37	G37
R38	R35	R38	R3-G29	R3	G29	G37	G37	G37
R35-R38	R35	R38	R3-R40	R3	R40	R3-R40	R3	R40
R38	R35	R38	R40	R3	R40	R40	R3	R40
R39-R40	R39	R40	R3-R40	R3	R40	R39-R40	R39	R40
R40	R39	R40	R3	R40	R40	R40	R39	R40

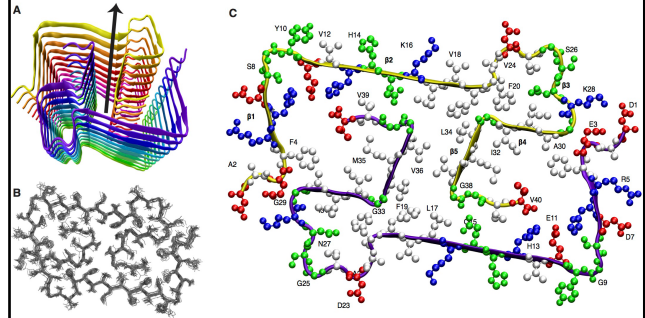
Anne Schütz, Toni Vagt, Matthias Huber, Oxana Y. Ovchinnikova, Riccardo Cadalbert,
Joseph Wall, Anja Böckmann, Rudi Glockshuber, Beat Meier

A β 1-40 E22 Δ fibrils: structure obtained with manually assigned restraints



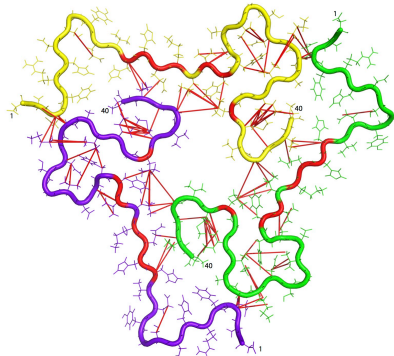
Anne Schütz, Toni Vagt, Matthias Huber, Oxana Y. Ovchinnikova, Riccardo Cadalbert, Joseph Wall, Anja Böckmann, Rudi Glockshuber, Beat Meier

A β 1-40 E22 Δ fibrils: structure obtained with manually and automatically assigned restraints



Anne Schütz, Toni Vagt, Matthias Huber, Oxana Y. Ovchinnikova, Riccardo Cadalbert, Joseph Wall, Anja Böckmann, Rudi Glockshuber, Beat Meier

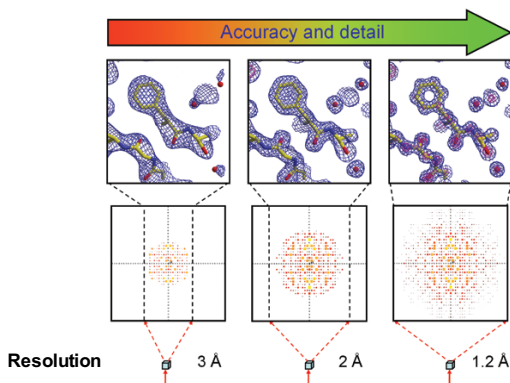
A β 1-40 E22 Δ fibrils: trial trimer structure



Anne Schütz, Toni Vagt, Matthias Huber, Oxana Y. Ovchinnikova, Riccardo Cadalbert, Joseph Wall, Anja Böckmann, Rudi Glockshuber, Beat Meier

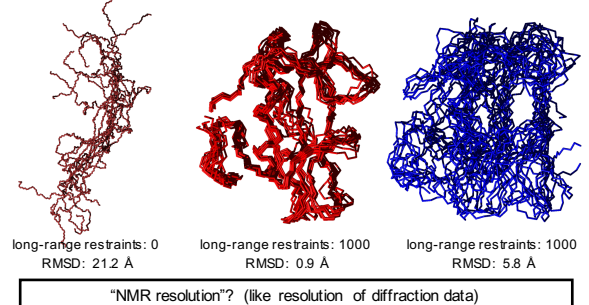
Information content

X-ray crystallography: Resolution → Amount of data → Structure quality



How to characterize conformational restraints for protein structure determination in a concise but meaningful way?

All these structures were calculated from 1000 distance restraints:



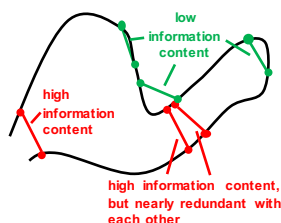
Information content of distance restraints

Table 1
Statistics of the 20 Final Solution Structures of the hPrp18 SFM Domain

Completeness of resonance assignments	96.8
Backbone (%)	99.2
Side chain (%)	99.2
Distance restraints	
Total NOE	1021
Intraresidue	320
Sequential ($ i - j = 1$)	278
Medium-range ($1 < i - j < 5$)	243
Long-range ($ i - j \geq 5$)	180
Other intermolecular restraints (HALLS)	
ϕ/ψ	28/27
CYANA target function ($\text{\AA}^2$)	0.019
Structure statistics	
NOE restraint violations	
Number > 0.10 $\text{\AA}$	0
Maximum ($\text{\AA}$)	0.12
Dihedral angle restraint violations	
Number > 2.5°	0
Maximum (°)	0.10
Energies (kcal/mol)	
Mean restraint violation energy	2.91
Mean AMBER energy	-3102.07
Ramachandran plot statistics (%)	
Residues in most favored regions	94.7
Residues in additional allowed regions	5.3
Residues in generously allowed regions	0
Residues in disallowed regions	0
RMSD from the average structure ($\text{\AA}$)	
Backbone atoms	0.39
Heavy atoms	1.02

He et al. *Proteins* 80, 968-974 (2012)

Counting the number of restraints is not very informative because individual restraints have widely different impact on defining the three-dimensional structure.



Information content of distance restraints

We define the information content of a set of conformational restraints as

$$\text{Information content} = -\log P(\text{restraints fulfilled by random structure})$$

For a set of distance restraints, $\{A_i\}$, the aforementioned probability can be defined as the conditional probability

$$P(\text{data set} | \text{random structure}) = P(\{A_i\} | 0) = \prod_i P(A_i | 0)^{1/R_i}$$

The information content $I(\{A_i\} | 0)$ of the restraint set $\{A_i\}$ in the context of the random structure, denoted by "0", is thus given by

$$I(\{A_i\} | 0) = -\log P(\{A_i\} | 0) = -\sum_i \frac{\log P(A_i | 0)}{R_i}$$

$P(A_i | 0)$ is the conditional probability that an individual restraint A_i is fulfilled by a random structure. R_i is the redundancy of the restraint A_i in the set $\{A_i\}$. For instance, for two identical restraints A_i and A_j , $R_i = R_j = 2$, leading to the same contribution $P(A_i | 0)^{1/2} P(A_j | 0)^{1/2} = P(A_i | 0)$ as a single restraint.

Information content of an isolated restraint

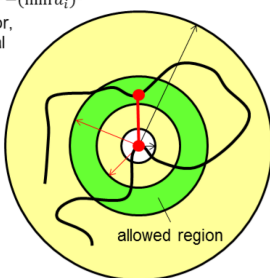
For a distance restraint A_i with an upper bound u_i and a lower bound l_i , the probability to be fulfilled by a random structure, represented by a bundle of conformers, can be computed as

$$P(A_i | 0) = \frac{u_i^3 - l_i^3}{(\max d_i)^3 - (\min d_i)^3}$$

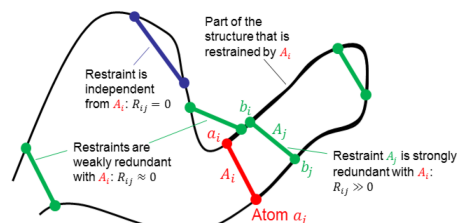
truncated to the range $[0, 1]$. In the denominator, $\max d_i$ and $\min d_i$ are the maximal and minimal values of the corresponding distance d_i in the bundle of random structures.

The expression for $P(A_i | 0)$ reflects the ratio between the volumes of two spherical shells with radii u_i and l_i , and $\max d_i$ and $\min d_i$, respectively.

In the absence of an upper bound u_i , $P(A_i | 0)$ is evaluated with $u_i = \max d_i$; in the absence of a lower bound l_i , with $l_i = \min d_i$.



Calculation of redundancy



Redundancy R_{ij} of restraint A_i with restraint A_j :

$$R_{ij} = \frac{u_i^3 - l_i^3}{(u_j + \Delta d_{ij})^3 - \max(l_j - \Delta d_{ij}, 0)^3}$$

with $\Delta d_{ij} = u_j + \min(\max d(a_i, a_j) + \max d(b_i, b_j), \max d(a_i, b_j) + \max d(b_i, a_j))$, where the maxima are taken over the random structures.

Redundancy R_i of restraint A_i with all restraints in the data set:

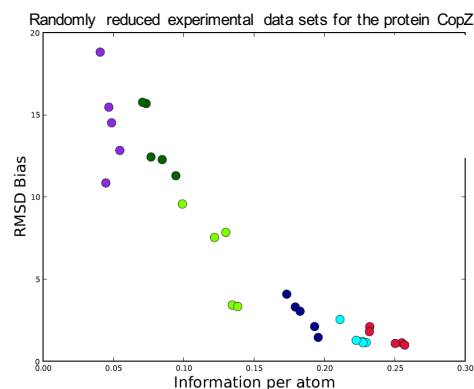
$$R_i = \sum_j R_{ij} \geq 1$$

Information content: Examples

Information content for modified experimental NMR data sets for the protein SH2:

Information content	Description
716	for the original data set with 1737 restraints
60	for its 419 intra-residual restraints
55	for its 423 sequential restraints
88	for its 251 medium-range restraints
512	for its 644 long-range restraints
571	for only the 644 long-range restraints
716	all restraints duplicated
170	all restraints with 2 assignments (instead of 1)
390	all upper bounds increased by 1 $\text{\AA}$
1385	all upper bounds decreased by 1 $\text{\AA}$
1598	lower bounds 0.5 $\text{\AA}$ below upper bound (eNOEs)
2	with respect to structure from the restraints

Information content vs. structure accuracy



Julia Weber, Lena Buchner