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. Strukturanalyse: Validierung
7. Automatisierung der NOE-Zuordnung und Strukturberechnung

Konformations-daten 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 ^{1}H-^{1}H Distanzen, *d*
Für isoliertes Spinpaar im starren Molekül:

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

Eigenschaften:
- nur kurze Distanzen < 5 Å messbar
- dichtes Netzwerk bzgl. der Sequenz kurz- und langreichweitiger Distanzschranken
- viele ^{1}H Atome im Molekül → "Spindiffusion"
- interne Bewegungen → nicht-lineare Mittelung
- Bestimmung der Konstanten *C*?
- Überlapp → mehrdeutige Zuordnung, verfälschte Integrale

➔ Verwendung als obere Distanzschranken

NOE Calibration

Volume of NOESY cross peak

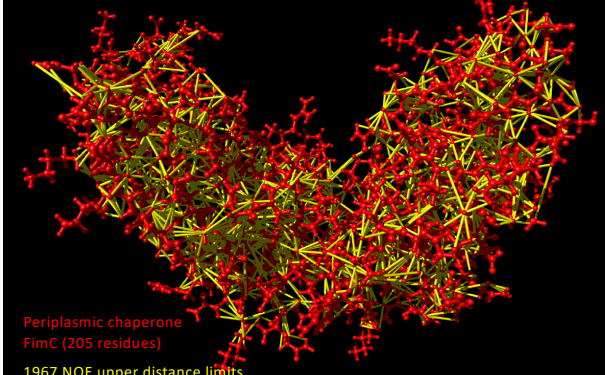
$$V = C / d^6$$

"Calibration constant"

Distance (upper distance bound)

How to set the calibration constant?
- Known distances (intraresidual or in standard secondary structures)
- Preliminary structure, if available
- User-defined value for the average (median) upper distance limit

NOE distance restraints → Protein structure



Periplasmic chaperone
FimC (205 residues)

1967 NOE upper distance limits

Pellecchia, M., Güntert, P., Glockshuber, R., Wüthrich, K. *Nature Struct. Biol.* 5, 885-890 (1998)

Problems when interpreting NOEs

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

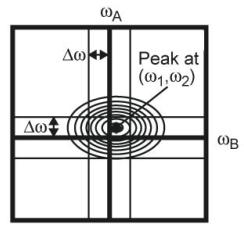
NMR resonance assignment is like solving a puzzle…

…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



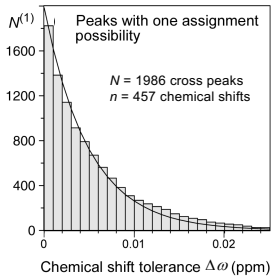
$|\omega_1 - \omega_A| < \Delta\omega \quad |\omega_2 - \omega_B| < \Delta\omega$

In general, several different ^{1}H 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!

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

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

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

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

- d_{eff} is never longer than any of the individual distances d_k:

  $d_{eff} \leq d_k$ for all k

- d_{eff} is close to the smallest individual distance:

  $d_{eff} \approx d_1$ if $d_1 << d_2, d_3, \ldots$

- Examples: $d_1 = 3$ Å, $d_2 = 10$ Å → $d_{eff} = 2.9996$ Å
  $d_1 = 3$ Å, $d_2 = \ldots = d_{10} = 10$ Å → $d_{eff} = 2.9967$ Å

^3J skalare Kopplungen

NMR Daten: Aufspaltung eines Signals

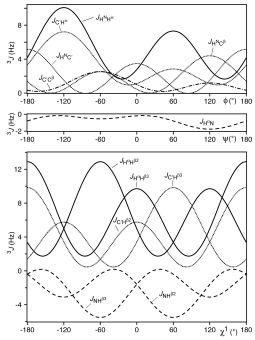
Konformationsdaten: Einschränkungen von Torsionswinkeln, θ

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

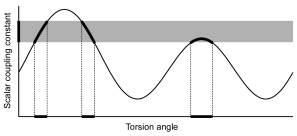
Zum Beispiel: $^3J_{HNH\alpha}(\phi)$, $^3J_{H\alpha H\beta}(\chi^1)$

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

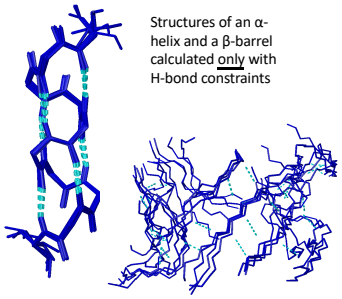
NMR Daten: langsamer ^{1}H → ^{2}H Austausch + NOEs

Konformationsdaten: Donor-Akzeptor Distanz

Typische H-Brücken: -N-H · · · 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



Structures of an α-helix and a β-barrel calculated only with H-bond constraints

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

Chemische Verschiebungen

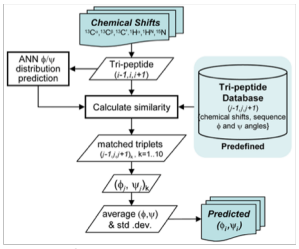
NMR Daten: chem. Verschiebungen, δ

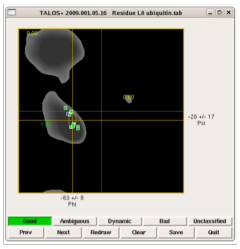
Konformationsdaten: (ϕ, ψ) Torsionswinkelbereiche

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

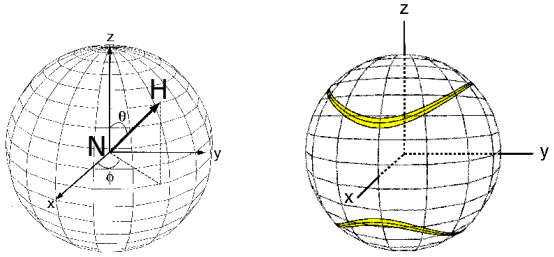
Residuelle dipolare Kopplung: $\boldsymbol{D(\theta,\phi) = A\,[(3\cos^2\theta - 1) + 3/2\,R\,\sin^2\theta\cos2\phi\,]}$

- 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 → Doppelkegel von Richtungen
- Bestimmung des Ausrichtungstensors (A, R)?

Residuelle dipolare Kopplungen



$$D(\theta,\phi) = A\,[(3\cos^2\theta - 1) + 3/2\,R\,\sin^2\theta\cos2\phi\,]$$

Strukturberechnungs-algorithmen

Ist NMR Strukturberechnung möglich?

- Grundsätzlich:
  - NOEs messen nur kurze Distanzen < 5 Å
  - ungenaue obere Schranken
  - Kann damit die globale Struktur eines 30 Å 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

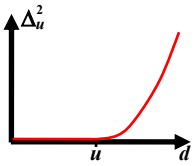
CYANA target function

$$T = \sum \Delta_u^2 + \sum \Delta_l^2 + \sum \Delta_a^2 + \cdots$$

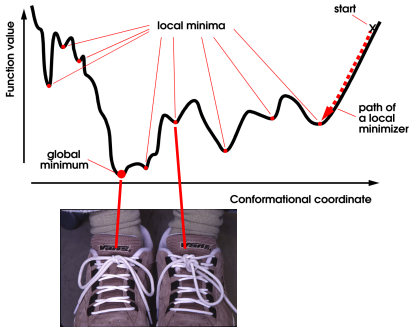
upper distance limits (NOEs) | lower distance limits (steric) | torsion angle restraints

Δ_u, Δ_l, Δ_a: restraint violations,

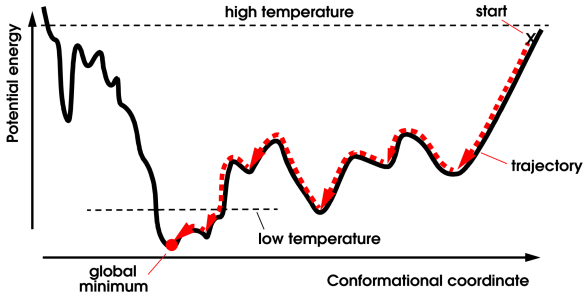
e. g., $\Delta_u = \begin{cases} d-u & \text{if } d > u \\ 0 & \text{otherwise} \end{cases}$



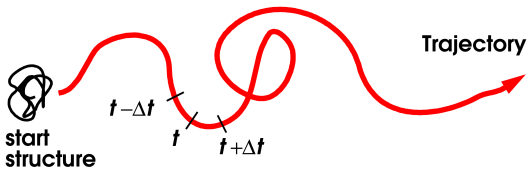
Target function = potential energy



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

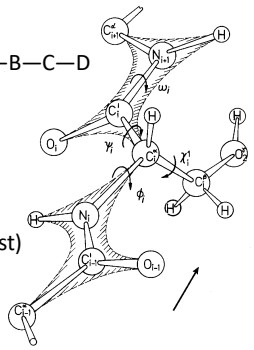
Strukturbeschreibung

Atomkoordinaten (kartesische Koordinaten):
- 3 Freiheitsgrade pro Atom
- abhängig von der Wahl des Koordinatensystems
- beinhalten auch "unwichtige" Freiheitsgrade
- einfach

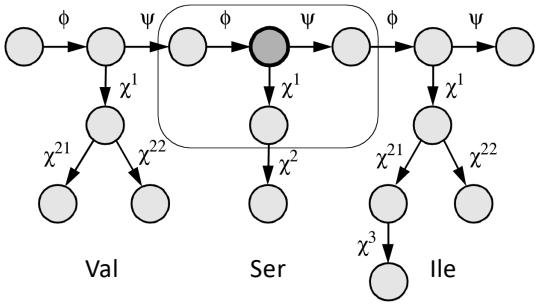
Torsionswinkel (= Diederwinkel, Dihedralwinkel):
- Drehungen um Einfachbindungen
- interne Koordinaten
- essentielle Freiheitsgrade
- Bindungslängen, Bindungswinkel fest
- kompliziertere aber effizientere Algorithmen

Torsionswinkel

- Definiert durch 4 Atome: A—B—C—D
  - Drehung um Bindung B—C
  - Werte von -180° bis +180°
- Torsionswinkel von AS *i*:
  ϕ_i: C'_{i-1}—N_i—C^α_i—C'_i
  ψ_i: N_i—C^α_i—C'_i—N_{i+1}
  ω_i: C^α_i—C'_i—N_{i+1}—C^α_{i+1} (fest)
  $\chi^1{}_i$: N_i—C^α_i—C^β_i—C^γ_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, Vaidehi, Rodriguez, *J. Comp. Phys.* 106, 258–268 (1993)
Güntert, Mumenthaler, Wüthrich, *J. Mol. Biol.* 273, 283–298 (1997)

Equations of motion

Cartesian coordinates: $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$

$$m_i \ddot{\boldsymbol{x}}_i = -\frac{\partial E_{\text{pot}}}{\partial \boldsymbol{x}_i} \qquad \textbf{(Newton)}$$

Generalized coordinates: $q_1, \ldots, q_n$

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

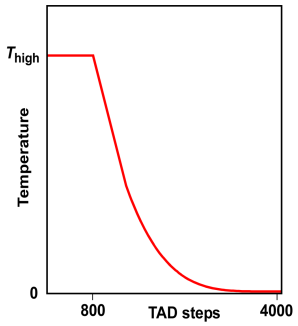
with $L = E_{\text{kin}} - E_{\text{pot}}$

Molecular Dynamics

Cartesian space		Torsion angle space
$E_{\text{kin}} = \frac{1}{2}\sum_{i=1}^{N} m_i \dot{\boldsymbol{x}}_i^2$	Kinetic energy	$E_{\text{kin}} = \frac{1}{2}\sum_{k,l=1}^{n} M(\theta)_{kl}\dot{\theta}_k\dot{\theta}_l$
diagonal, constant (elements m_i)	Mass matrix M	non-diagonal, non-constant, $n \times n$
$\ddot{\boldsymbol{x}}_i = -\frac{1}{m_i}\frac{\partial E_{\text{pot}}}{\partial \boldsymbol{x}_i}$	Accelerations	$M(\theta)\ddot{\theta} = C(\theta, \dot{\theta})$ (n linear equations)
proportional to N	Computational complexity	solving linear system of equations: ~n^3 exploiting tree structure of the molecule: ~n

Simulated annealing protocol

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



Temperature control

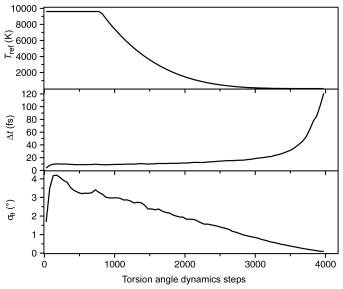
Weak coupling to a heat bath is used to control the temperature:

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

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

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

Simulated annealing mit Torsionswinkeldynamik



Temperatur

Zeitschritt

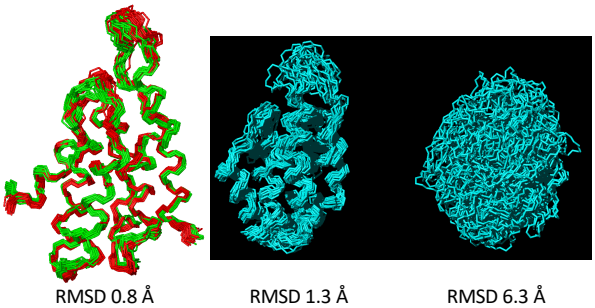
Torsionswinkel-änderung

Strukturbündel RMSDs

Strukturbündel

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

Strukturbündel



RMSD 0.8 Å RMSD 1.3 Å RMSD 6.3 Å

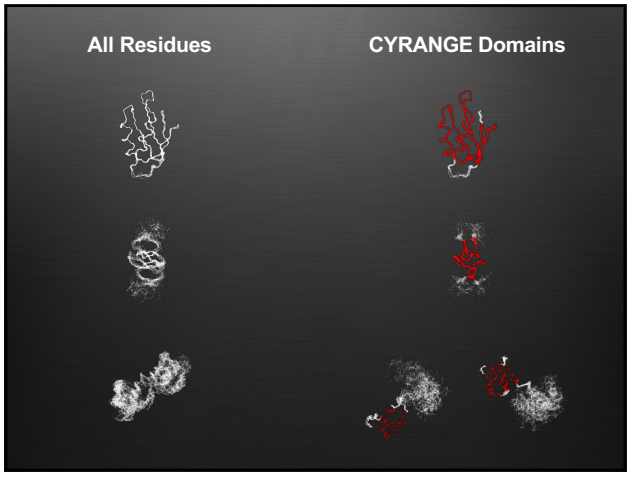
ENTH-VHS domain At3g16270

RMSD (root-mean-square deviation)

- Zwei Strukturen mit *n* Atomen und Koordinaten $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ und $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$

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

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

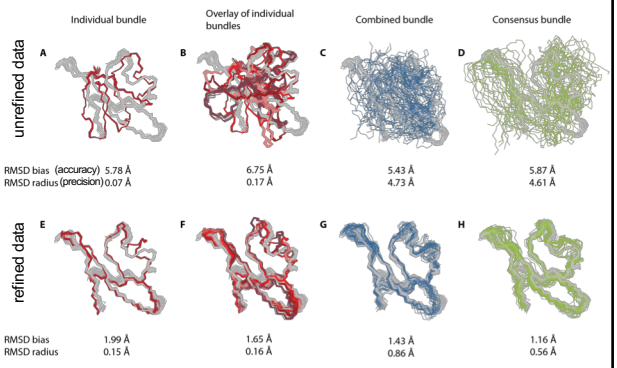


Consensus structure bundles

Structure accuracy vs. precision

- NMR structures are represented by bundles of conformers calculated from different randomized initial structures using identical experimental input data.
- The spread among these conformers indicates the **precision** of the atomic coordinates.
- However, there is as yet no reliable measure of structural **accuracy**, i.e. how close NMR conformers are to the "true" structure.
- Instead, the precision of structure bundles is widely (mis)interpreted as a measure of structural quality.
- Attempts to increase precision often overestimate accuracy by tight bundles of high precision but much lower accuracy.

Individual/combined/consensus bundles

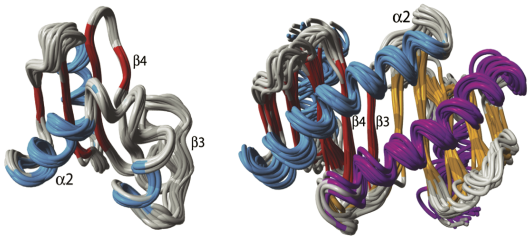


CASD-NMR protein StT322 (PDB 2LOJ; B. Wu & C. Arrowsmith)

Strukturanalyse Validierung

Correct and wrong structure: Dynein light chain 2A

Wrong structure (1TGQ) Correct structure (1Y4O): Homodimer



Nabuurs, S. B., Spronk, C. A. E. M., Vuister, G. W. & Vriend, G. (2006). Traditional biomolecular structure determination by NMR spectroscopy allows for major errors. *PLoS Comp. Biol.* 2, 71–79.

Validation principles

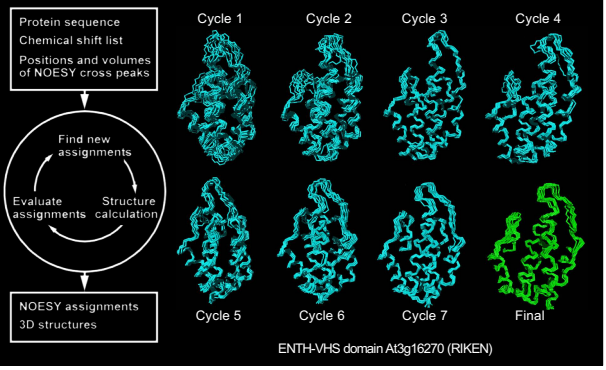
Agreement of the three-dimensional structure with
- Experimental data
- Unused experimental data: cross-validation
- Physical principles
- Empirical knowledge about protein structures

Validation of the
- Local structure
- Global structure

Absolute/relative validation:
- Is my structure correct? (“absolute”)
- Is structure *A* more likely to be correct than structure *B*? (“relative”)

Automatische NOE Zuordnung

Automated NOESY assignment and structure calculation with CYANA



ENTH-VHS domain At3g16270 (RIKEN)

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	3591	3872	3950	4115	4195	4194	
with short-range assignment |i-j|<=1	:	3295	3208	3165	3154	3120	3102	3089	
with medium-range assignment 1<|i-j|<5	:	1020	925	921	914	904	884	893	
with long-range assignment |i-j|>=5	:	773	661	644	669	676	680	681	
Upper distance limits:									
total	:	3786	2996	2832	2789	2707	2643	2683	2731
short-range, |i-j|<=1	:	2007	1586	1486	1440	1388	1348	1273	1304
medium-range, 1<|i-j|<5	:	1220	959	787	775	751	726	760	765
long-range, |i-j|>=5	:	559	451	559	574	568	569	650	662
Average assignments/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 × 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**

Unterlagen zur Vorlesung

http://www.bpc.uni-frankfurt.de/guentert/wiki/index.php/Teaching