

NMR Strukturbestimmung

Peter Güntert

1

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

2

Konformationsdaten aus NMR Messungen

3

Konformationsdaten aus NMR Messungen

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

4

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.

5

NOE Calibration

$$V = C / d^6$$

Volume of NOESY cross peak \rightarrow V

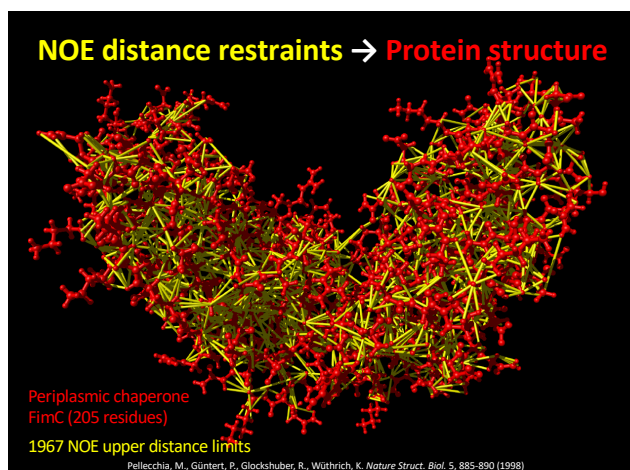
"Calibration constant" \rightarrow C

Distance (upper distance bound) \rightarrow d

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

6



7


Problems when interpreting NOEs

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


8

NMR resonance assignment is like solving a puzzle...

...with missing pieces
(incomplete signals)



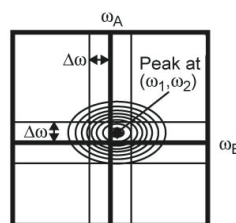
...with additional pieces
(artifacts)



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

9

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 $\Delta\omega$.

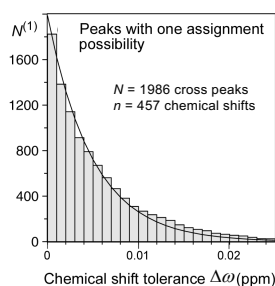
→ Assignment ambiguity

Manual assignment is very cumbersome!

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

10

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

11

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)

12

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: $d_1 = 3 \text{ \AA}, d_2 = 10 \text{ \AA} \rightarrow d_{\text{eff}} = 2.9996 \text{ \AA}$
 $d_1 = 3 \text{ \AA}, d_2 = \dots = d_{10} = 10 \text{ \AA} \rightarrow d_{\text{eff}} = 2.9967 \text{ \AA}$

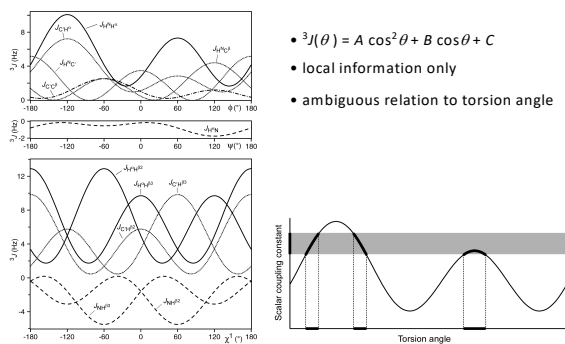
13

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 $^3J \leftrightarrow \theta$

14

3J scalar couplings



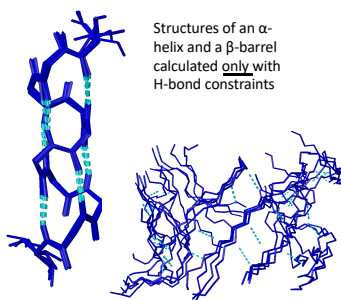
15

H-Brücken

- NMR Daten: langsamer $^1\text{H} \rightarrow ^2\text{H}$ Austausch + NOEs
Konformationsdaten: Donor-Akzeptor Distanz
Typische H-Brücken: $-\text{N}-\text{H} \cdots \text{O}=\text{C}-$ 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)

16

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

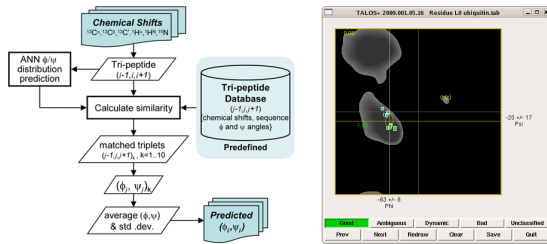
17

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

18

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 .

19

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

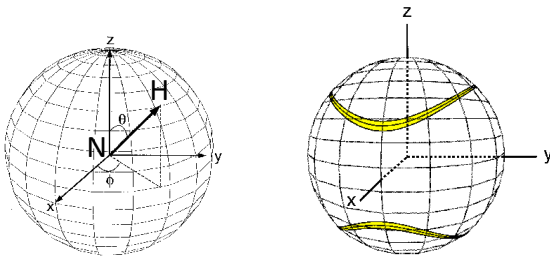
θ, ϕ Richtung der Bindung relativ zum Ausrichtungstensor (Polarkoordinaten)

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)?

20

Residuelle dipolare Kopplungen



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

21

Strukturberechnungs- algorithmen

22

Ist NMR Strukturberechnung möglich?

- Grundsätzlich:
 - NOEs messen nur kurze Distanzen $< 5 \text{ \AA}$
 - ungenaue obere Schranken
 - Kann damit die globale Struktur eines 30 \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.

23

Strukturberechnungsalgorithmen

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

24

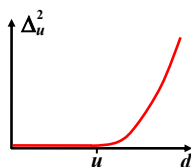
CYANA target function

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

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

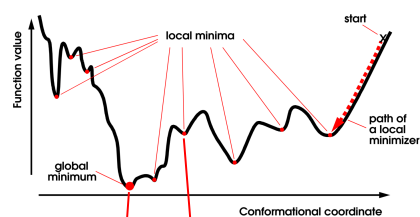
$\Delta_u, \Delta_l, \Delta_a$: restraint violations,

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



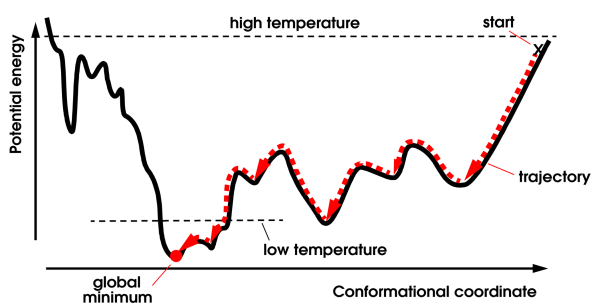
25

Target function = potential energy



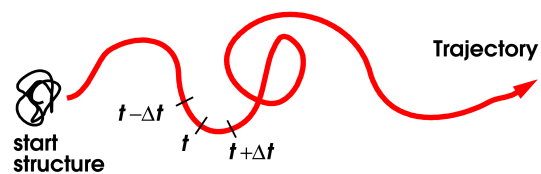
26

Simulated annealing



27

Molecular Dynamics Simulation



Numerical integration of classical equations of motion

28

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

29

Atomkoordinaten Torsionswinkel

30

Strukturbeschreibung

Atomkoordinaten (kartesische Koordinaten):

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

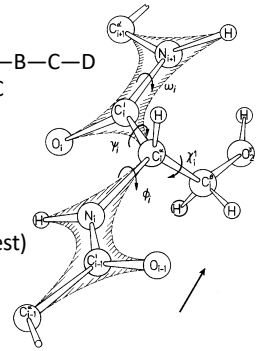
Torsionswinkel (= Diederwinkel, Dihedralwinkel):

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

31

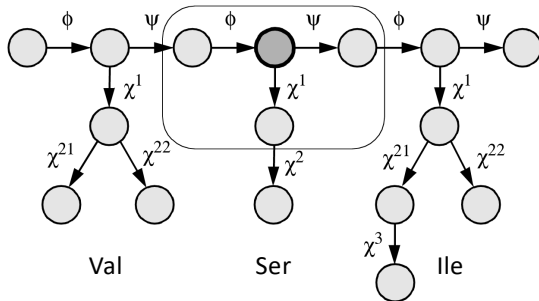
Torsionswinkel

- Definiert durch 4 Atome: A—B—C—D
 - Drehung um Bindung B—C
 - Werte von -180° bis +180°
- Torsionswinkel von AS i :
 - ϕ_i : $C'_{i-1}-N_i-C^{\alpha}_i-C'_i$
 - ψ_i : $N_i-C^{\alpha}_i-C'_{i-1}-N_{i+1}$
 - ω_i : $C^{\alpha}_i-C'_{i-1}-N_{i+1}-C^{\alpha}_{i+1}$ (fest)
 - χ^1_i : $N_i-C^{\alpha}_i-C^{\beta}_i-C'_i$



32

Torsionswinkel: Baumstruktur



33

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)

34

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

35

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

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, $n \times n$

$$M(\theta) \ddot{\theta} = C(\theta, \dot{\theta}) \quad (n \text{ linear equations})$$

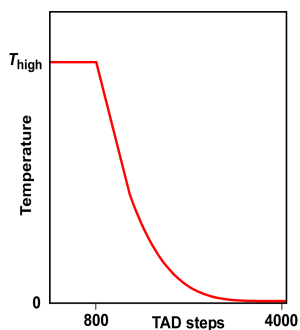
solving linear system of equations: $\sim n^3$

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

36

Simulated annealing protocol

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



37

Temperature control

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

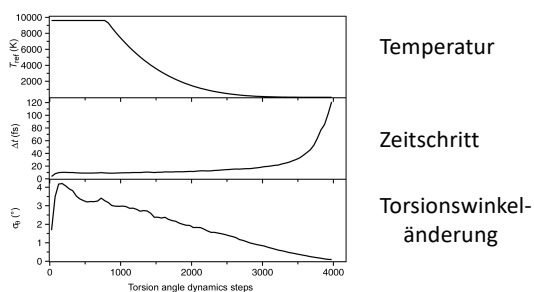
$$\dot{\theta} \leftarrow \dot{\theta} \sqrt{1 + \frac{T^{\text{ref}} - T}{\tau T}}$$

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

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

38

Simulated annealing mit Torsionswinkeldynamik



39

Strukturbündel RMSDs

40

NMR Structure Calculation: Multiple Conformers



41

NMR Structure Calculation: Select Converged Conformers



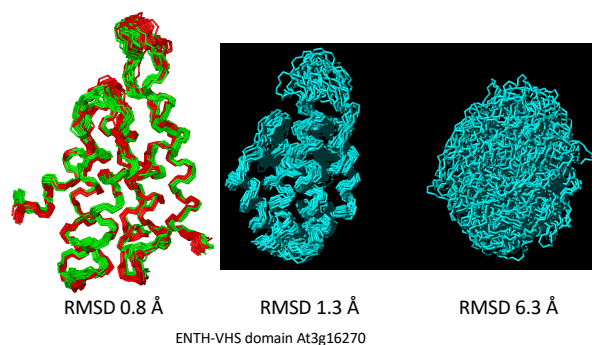
42

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?

43

Strukturbündel



44

RMSD (root-mean-square deviation)

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

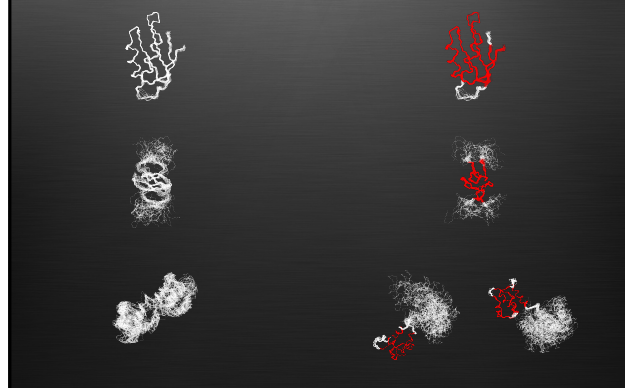
$$RMSD = \min_{R, \vec{t}} \sqrt{\frac{1}{n} \sum_{i=1}^n |\vec{x}_i - R\vec{y}_i - \vec{t}|^2}$$

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

45

All Residues

CYRANGE Domains



46

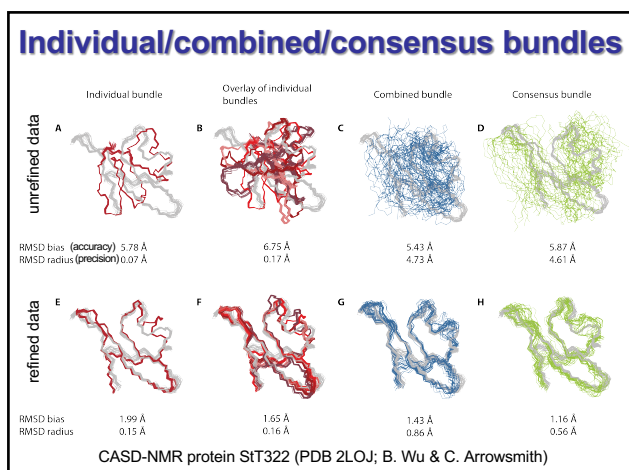
Consensus structure bundles

47

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.

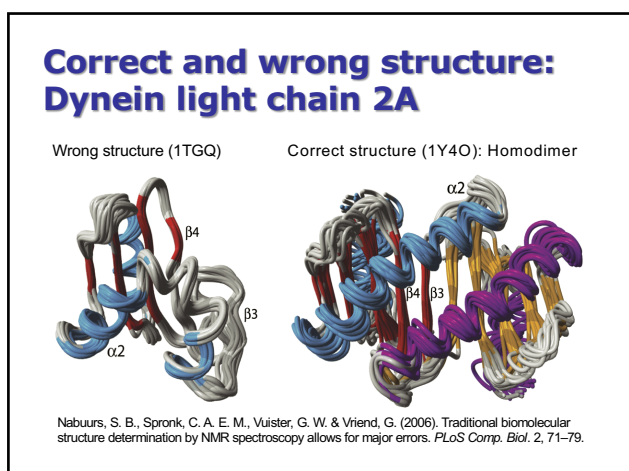
48



49

Strukturanalyse Validierung

50



51

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

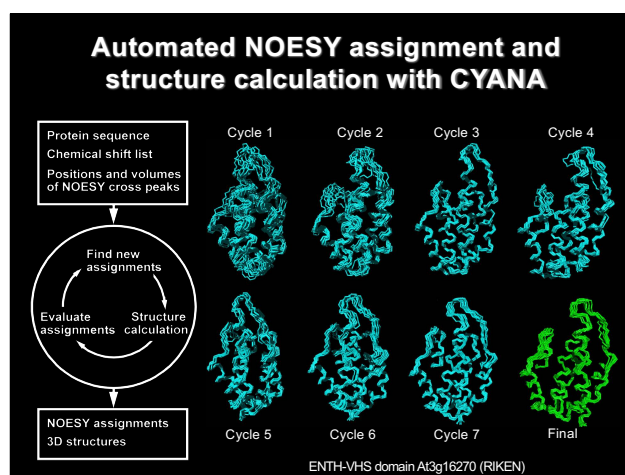
Absolute/relative validation:

- Is my structure correct? ("absolute")
- Is structure A more likely to be correct than structure B? ("relative")

52

Automatische NOE Zuordnung

53



54

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	4394	
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/zstraint	:	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

55

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

56

Unterlagen zur Vorlesung

<http://www.bpc.uni-frankfurt.de/guertert/wiki/index.php/Teaching>

Evaluation

Veranstaltung:	000* 1.3 Methoden zur Strukturbestimmung von Biomolekülen
Lehrperson:	Univ.-Prof. Dr. Peter Güntert
Evaluationstermin:	07.12.2020, 11:30 - 14:00 Uhr
URL:	http://r.sd.uni-frankfurt.de/743832f7
QR-Code:	

57