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
6. Automatisierung der NOE-Zuordnung und Strukturberechnung
7. Strukturanalyse: Übereinstimmung mit experimentellen Daten, Verletzungen von Konformationseinschränkungen, RMSD, Ramachandranplot, Validierung

Konformationsdaten aus NMR Messungen

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

...

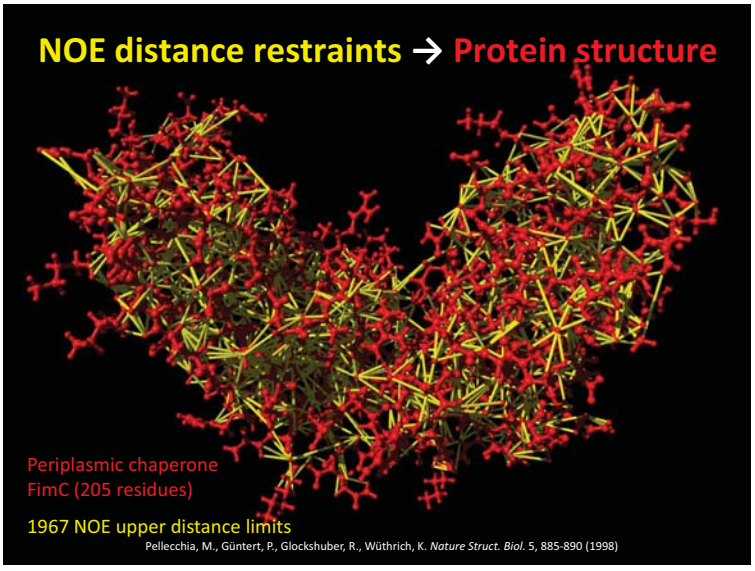
NOE (Nuclear Overhauser Effect)

NMR Daten: Integral V von NOESY Kreuzsignalen

Konformationsdaten: obere Schranken für ^{1}H-^{1}H Distanzen, d

Fuer isoliertes Spinpaar im starren Molekül:
$V = C/d^{6}$ mit C = konstant

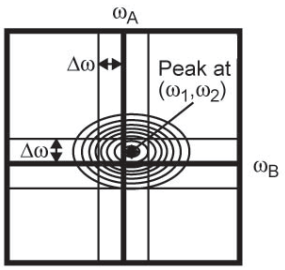
Eigenschaften:
- nur kurze Distanzen < 5 Å messbar
- dichtes Netzwerk bzgl. der Sequenz kurz- und langreichweitiger Distanzschranken
- viele ^{1}H Atome im Molekül → "Spindiffusion"
- interne Bewegungen → nicht-lineare Mittelung
- Bestimmung von C?
- Überlapp → mehrdeutige Zuordnung, verfälschte Integrale



Problems when interpreting NOEs

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

Ambiguity of chemical shift based NOE assignment



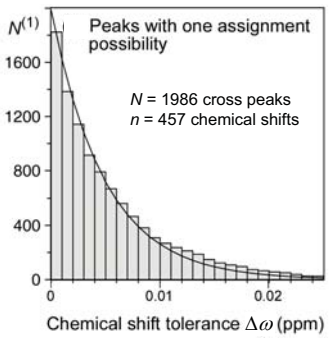
$|\omega_1 - \omega_A| < \Delta\omega \quad |\omega_2 - \omega_B| < \Delta\omega$

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

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

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)

^3J skalare Kopplungen

NMR Daten: Aufspaltung eines Signals

Konformationsdaten: Einschränkungen von Torsionswinkeln, θ

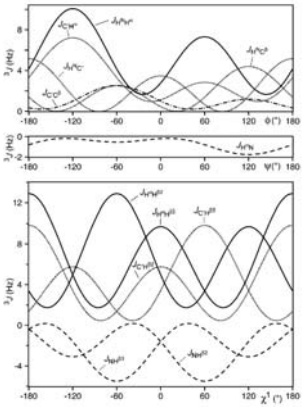
Karplus-Kurve: $^3J(\theta) = A\cos^2\theta + B\cos\theta + C$ mit emprischen Konstanten A, B, C

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

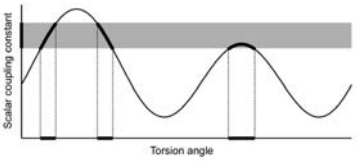
Eigenschaften:

- 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

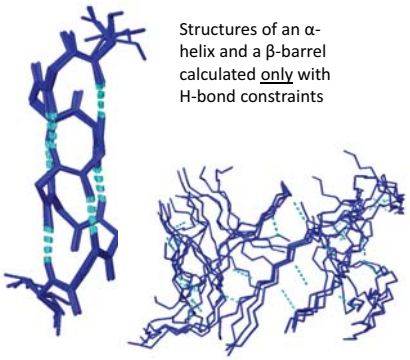
NMR Daten: langsamer $^1H \rightarrow {}^2H$ Austausch + NOEs

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) 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 nur über lokale Konformation

Residuelle dipolare Kopplungen (RDC)

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

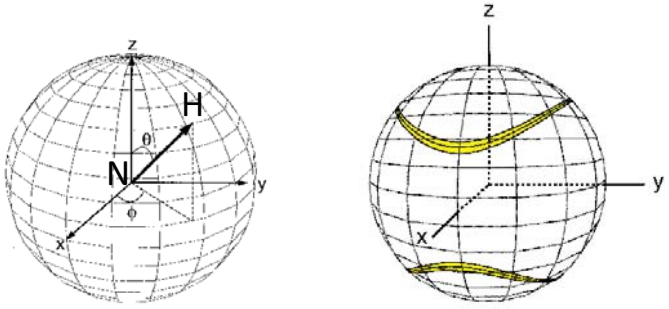
Konformationsdaten: Orientierung von Bindungen relativ zur Molekülausrichtung

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

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



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

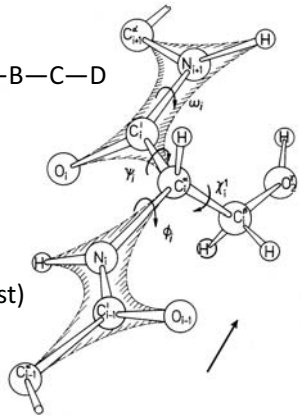
Strukturbeschreibung

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

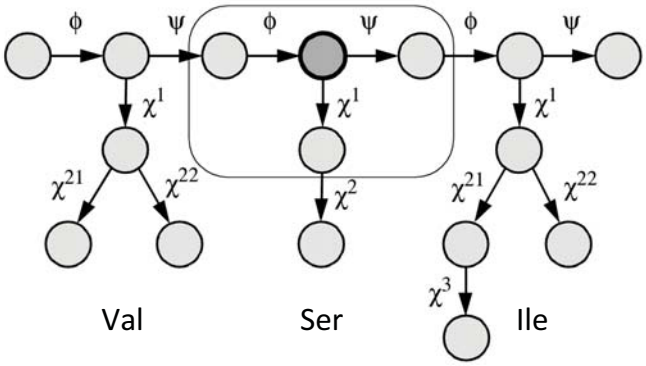
Torsionswinkel (= Diederwinkel, Dihedralwinkel):
- Drehungen um Einfachbindungen
- interne Koordinaten
- essentielle Freiheitsgrade
- Bindungslängen, Bindungswinkel fest
- kompliziertere aber effizientere Algorithmen

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: $C'_{i-1}—N_i—C^{\alpha}{}_i—C'_i$
  ψ_i: $N_i—C^{\alpha}{}_i—C'_i—N_{i+1}$
  ω_i: $C^{\alpha}{}_i—C'_i—N_{i+1}—C^{\alpha}{}_{i+1}$ (fest)
  $\chi^1{}_i$: $N_i—C^{\alpha}{}_i—C^{\beta}{}_i—C^{\gamma}{}_i$



Torsionswinkel: Baumstruktur



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 da sind.
- Praktisch:
  - Zielfunktion hat viele lokale Minima
  - Kann die optimale Struktur gefunden werden?
    JA.

Strukturberechnungsalgorithmen

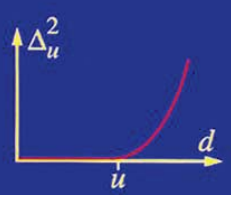
- Frühere Methoden:
  - Interaktiver Modellbau
  - Distanzgeometrie
  - Minimierung einer variablen Zielfunktion
- Simulated annealing:
  - Monte Carlo
  - Moleküldynamiksimulation im kartesischen Raum
  - Moleküldynamiksimulation im Torsionswinkelraum

Target function (CYANA)

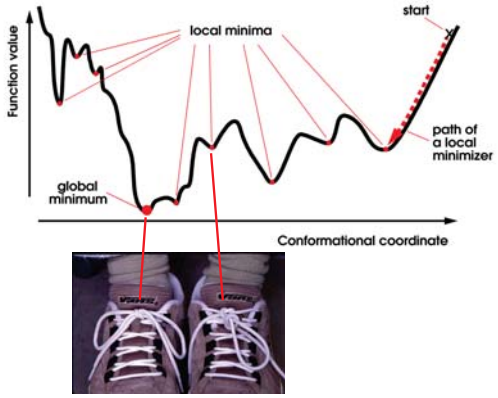
$$T = \sum_{\text{upper distance limits (NOEs)}} \Delta_u^2 + \sum_{\text{lower distance limits (steric)}} \Delta_l^2 + \sum_{\text{torsion angle restraints}} \Delta_a^2$$

Δ_u, Δ_l, Δ_a: restraint violations,

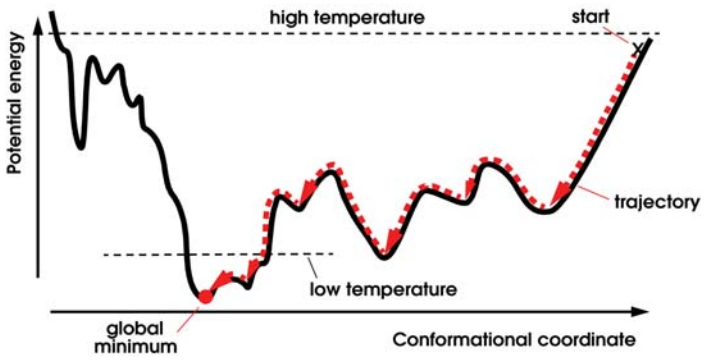
e. g., $\Delta_u = \begin{cases} d-u & \text{if } d > u \\ 0 & \text{otherwise} \end{cases}$



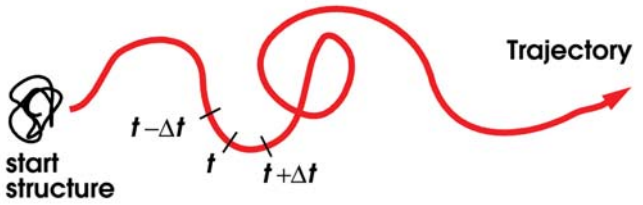
Target function = potential energy



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

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: $x_1, \ldots, x_N$

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

Generalized coordinates: $q_1, \ldots, q_n$

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

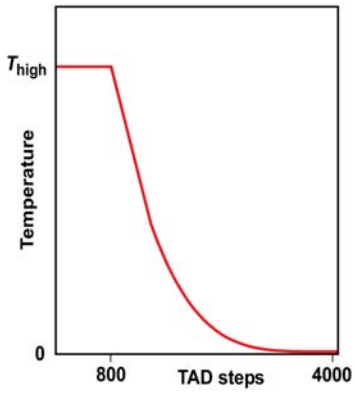
with $L = E_{\text{kin}} - E_{\text{pot}}$

Molecular Dynamics

Cartesian space		Torsion angle space
$E_{\text{kin}} = \frac{1}{2}\sum_{i=1}^{N} m_i \dot{x}_i^2$	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, $n \times n$
$\ddot{x}_i = -\frac{1}{m_i}\frac{\partial E_{\text{pot}}}{\partial x_i}$	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

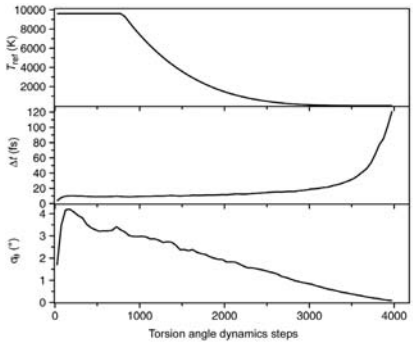
Weak coupling to a heat bath is used to control the temperature:

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

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

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

Simulated annealing mit Torsionswinkeldynamik



Temperatur

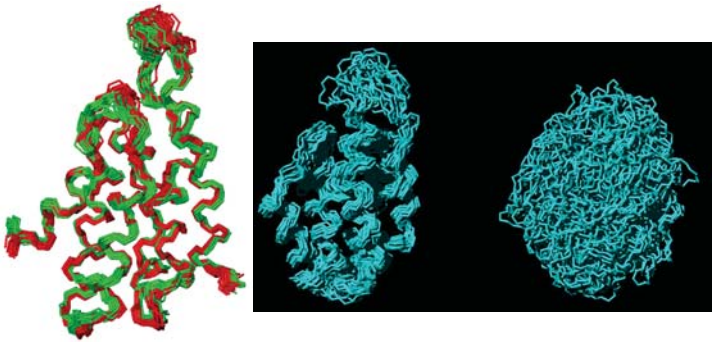
Zeitschritt

Torsionswinkel-änderung

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

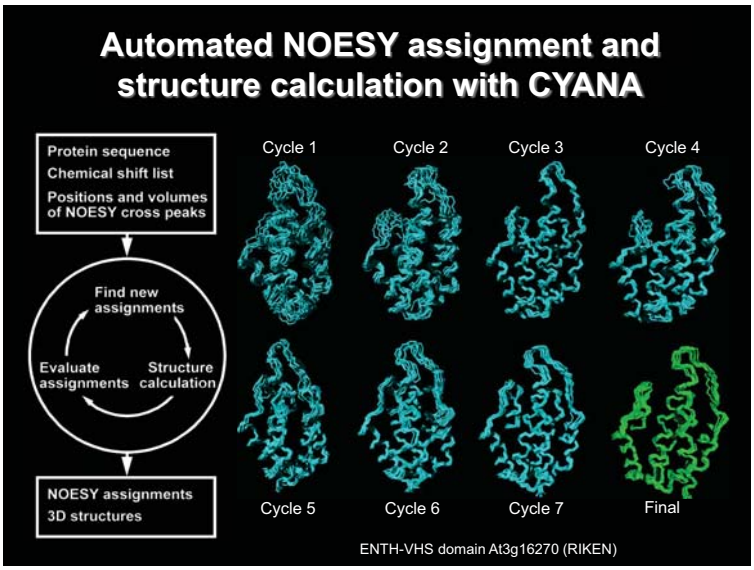
ENTH-VHS domain At3g16270

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 $\boldsymbol{t}$ → optimale Überlagerung



Output overview table

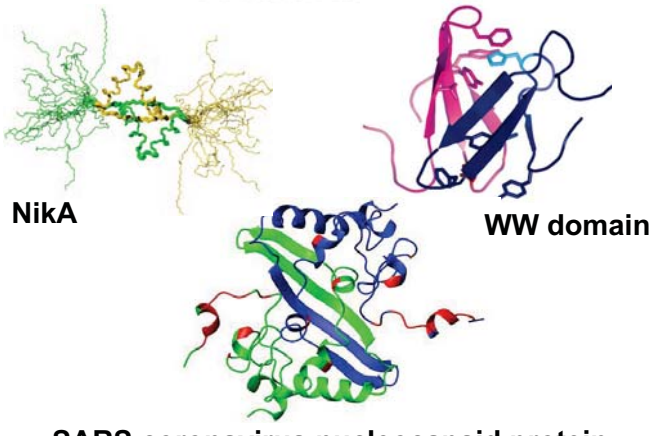
Cycle	:	1	2	3	4	5	6	7	final
Peaks:									
selected	:	5439	5439	5439	5439	5439	5439	5439	
with assignment	:	5100	4806	4742	4749	4712	4678	4675	
without assignment	:	339	633	697	690	727	761	764	
with diagonal assignment	:	12	12	12	12	12	12	12	
Cross peaks:									
with off-diagonal assignment	:	5088	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)	
	Total	per structure
100	**147**	**0.011**
50	**217**	**0.019**
25	**354**	**0.036**
10	**769**	**0.088**

Homodimers

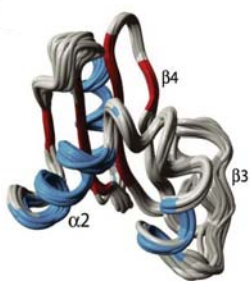


NikA **WW domain**

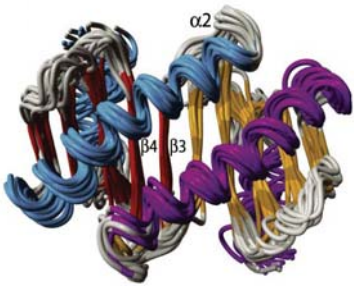
SARS coronavirus nucleocapsid protein

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”)

X-ray crystallography: *R*-factor

- Measures agreement between measured data (reflections) and 3D structure
- Definition: Relative difference between structure factors, $F(hkl)$, that were observed (F_{obs}) and back-calculated from the 3D structure (F_{calc}):

$$R = \frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|} \quad \text{with} \quad I_{hkl} \propto |F(hkl)|^2$$

  I_{hkl} = intensity of reflection (hkl)
- Perfect agreement: $R = 0$
  Good protein X-ray structure: $R < 0.2$
  Random structure: $R \approx 0.6$

X-ray: Free *R*-factor

- Use, say, 90% of the data (reflections) for the structure determination
- Use the remaining 10% to compute the *R* value → “free” *R* value, obtained from independent data
- Detects errors better than conventional *R*-factor
- Each reflection influences whole electron density
- Many reflections → No problem to omit 10% of the reflections from the structure determination

Brünger, A. T. (1992). Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472-475.

R-factor in NMR

- NMR restraints (NOEs) are not raw data but require assignments, calibration, etc.
- Back-calculation of NOEs from 3D structures needs data or assumptions on dynamics and consideration of spin diffusion → “Relaxation matrix calculations”
- Agreement between measured and back-calculated NOESY peak volumes:
  - dominated by strong short-range NOEs
  - absence/presence of a weak (but structurally important!) long-range NOE has negligible influence on the *R*-factor
- Agreement of distances?

Free *R*-factor using RDCs

- Use NOE distance restraints to determine structure
- Use residual dipolar couplings to validate
- Quality factor (*R*-factor):

$$Q = \mathrm{rms}(D^{calc} - D^{obs})/\mathrm{rms}(D^{obs}),$$

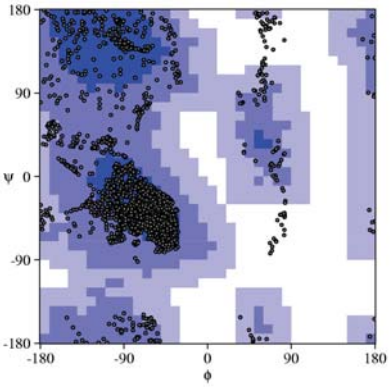
where D^{obs} and D^{calc} are observed and calculated one-bond dipolar couplings.

Simon, K., Xu, J., Kim, C. & Skrynnikov, N. (2005). Estimating the accuracy of protein structures using residual dipolar couplings. *J. Biomol. NMR* 33, 83-93.

Validation without experimental data

- Stereochemical quality
- “Normality” of the structure with respect to the existing structures in Protein Data Bank
- Parameters:
  - Bond lengths, bond angles
  - Ramachandran plot
  - Steric overlap (“bumps”)
- Conformational energy
- **3D structure (molecular graphics!)**

Ramachandran-Plot



Example:
Each black dot =
1 residue in 1 conformer

- 73% in most favored regions (dark blue)
- 21% in additionally allowed regions (light blue)
- 4% in generously allowed regions (blue-grey)
- 2% in disallowed regions (white)

(Programm PROCHECK)

WHAT_CHECK validation checks

- **Administrative checks:** nomenclature, missing atoms
- **Geometry:** chirality, bond lengths, bond angles, torsion angles (evaluation, Ramachandran plot, omega, χ^1/χ^2), rings and planarity, proline puckering
- **Structure:** inside/outside profile, bumps, packing, backbone (number of hits, backbone normality, peptide flips), sidechain rotamers
- **Hydrogen bonds:** unsatisfied, flip check, His assignments
- **Summary:** overall *Z*-scores and RMS *Z*-scores

$$Z = \frac{X_i - \langle X \rangle}{\sigma(X)} \qquad \text{RMS-}Z = \sqrt{\langle Z^2 \rangle}$$

WHAT_IF/WHAT_CHECK output

- Structure *Z*-scores, positive is better than average:

```
1st generation packing quality :  0.891
2nd generation packing quality :  1.444
Ramachandran plot appearance   : -0.105
chi-1/chi-2 rotamer normality  : -0.431
Backbone conformation          :  0.551
```

- RMS *Z*-scores, should be close to 1.0:

```
Bond lengths                   :  0.887
Bond angles                    :  1.143
Omega angle restraints         :  0.437 (tight)
Side chain planarity           :  0.934
Improper dihedral distribution :  1.053
B-factor distribution          :  3.497 (loose)
Inside/Outside distribution    :  0.930
```

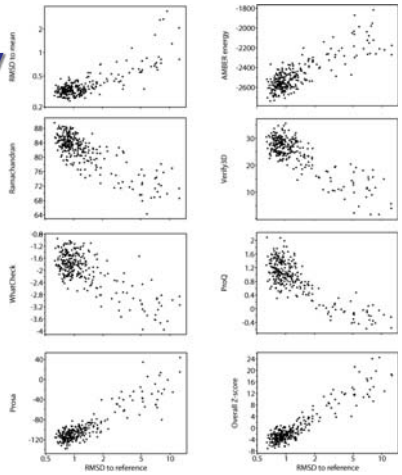
Hooft, R. W. W., Vriend, G., Sander, C., Abola, E. E. (1996) Errors in protein structures. *Nature* 381, 272.

Correlation between validation parameters and structure accuracy

- 252 ubiquitin structures
- Accuracy = RMSD from reference structure
- 7 quality parameters, S_i
- Overall *Z*-score:

$$Z = \sum_{i=1}^{7} \frac{S_i - \overline{S}_i}{\sigma(S_i)}$$

Correlation coefficient 93%



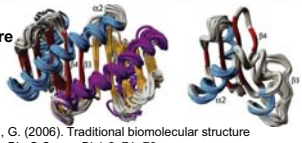
Teppei Ikeya

Quality indicators for correct and wrong structures of DLC2A

Table 1. Average Quality Indicators of the 1Y4O and 1TGQ Structure Ensembles before and after Refinement in Explicit Solvent

Criteria	Characteristic	1Y4O (Original)	1Y4O (Refined)	1TGQ (Original)	1TGQ (Refined)
Agreement with experimental data	RMS violation 1Y4O distance restraints (Å)	0.0129	0.0097	0.607	0.0284
	Violations >0.5 Å 1Y4O distance restraints	0	0	63	0
	RMS violation $1TGQ_{sym}$ restraints (Å)	12.8	12.6	0.521	0.0231
	Violations >0.5 Å $1TGQ_{sym}$ restraints	32	32	4	0
	RMS violation 1Y4O dihedral restraints (°)	0.497	0.336	25.0	1.59
	Violations >5° 1Y4O dihedral restraints	0	0	34	4
PROCHECK validation results[a]	Most favored regions	91.2	90.5	67.7	85.8
	Additionally allowed regions	8.4	9.0	27.3	12.8
	Generously allowed regions	0.2	0.2	4.7	0.5
	Disallowed regions	0.2	0.3	0.2	0.9
WHAT IF structure *Z*-scores[b]	Packing quality	−0.4	0.1	−2.1	−1.5
	Ramachandran plot appearance	−3.6	−3.3	−6.6	−4.6
	χ_1/χ_2 rotamer normality	−0.3	−0.7	−5.8	−3.0
	Backbone conformation	−0.8	−1.1	−5.4	−5.4

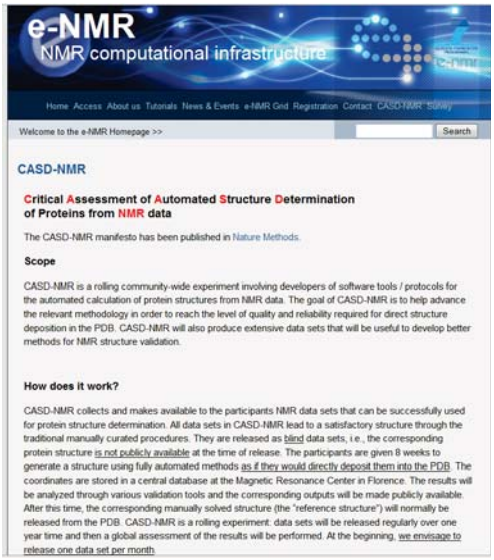
- **Better quality indicators for correct structure**
- **But difficult to detect wrong structure without knowledge of correct structure**



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.

www.enmr.eu/CASD-NMR

A. Rosato et al.
CASD-NMR: a rolling experiment for the critical assessment of automated structure determination of proteins from NMR data.
Nature Methods 6, 625–626 (2009)

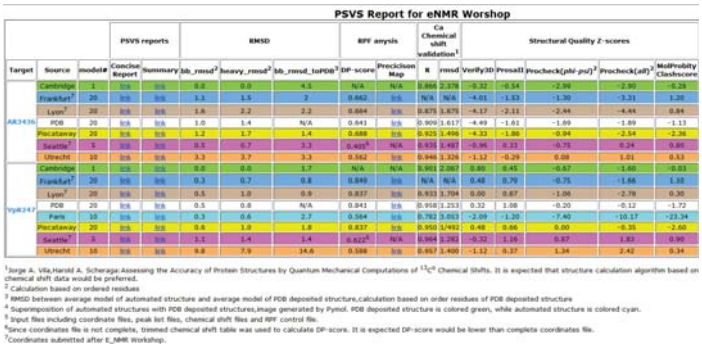


CASD-NMR data sets

This page provides access to the data sets and corresponding validation results for CASD-NMR. Disclosed data sets are flagged.

Target (download)	Oligomeric state	Date of release	PDB deposition date	PDB ID	PDB release date	Blind (Y/N)	LACS CA/CB Offset
VpR247	Monomer	16/04/2009	03/05/2009	2KIF	30/06/2009	N	-0.11
AR3436A	Monomer	16/04/2009	22/05/2009	2KJ6	17/07/2009	N	-0.19
HR5537A	Monomer	30/06/2009	14/06/2009	2KK1	09/08/2009	N	0.03
ET109A	Monomer	30/06/2009	29/06/2009	2KKX 2KKY	24/08/2009	N	-0.25/ -0.35
AtT13	Monomer	14/09/2009	02/09/2009	2KNR	28/10/2009	N	na
PGR122A	Monomer	17/09/2009	31/07/2009	2KMM	09/11/2009	Y	na
NeR103A	Monomer	22/10/2009	20/10/2009	2KPM	11/12/2009	Y	-0.19
CGR26A	Monomer	22/10/2009	20/10/2009	2KPT	14/12/2009	Y	0.13

Validation report for CASD-NMR structures (PSVS)


PSVS Report for eNMR Worshop

Target	Source	model#	PSVS reports: Concise Report	PSVS reports: Summary	RMSD: bb_rmsd[2]	RMSD: heavy_rmsd[2]	RMSD: bb_rmsd_toPDB[3]	RPF anysis: DP-score	RPF anysis: Precision Map	Cα Chemical shift validation[1]: R	Cα Chemical shift validation[1]: rmsd	Structural Quality Z-scores: Verify3D	Structural Quality Z-scores: Prosall	Structural Quality Z-scores: Procheck(phi-psi)[2]	Structural Quality Z-scores: Procheck(all)[2]	Structural Quality Z-scores: MolProbity Clashscore
AR3436	Cambridge	1	link	link	0.8	0.8	4.5	N/A	N/A	0.866	2.376	-0.32	-0.54	-2.99	-2.90	-0.28
	Frankfurt[5]	20	link	link	1.1	1.5	2	0.662	link	N/A	N/A	-4.01	-1.53	-1.30	-3.31	1.20
	Lyon[5]	20	link	link	1.6	2.2	2.2	0.684	link	0.875	1.675	-4.17	-2.11	-2.44	-4.44	0.84
	PDB	20	link	link	1.0	1.4	N/A	0.641	link	0.909	1.617	-4.49	-1.61	-1.69	-1.89	-1.13
	Piscataway	20	link	link	1.2	1.7	1.4	0.688	link	0.925	1.496	-4.33	-1.86	-0.84	-2.54	-2.36
	Seattle[5]	5	link	link	0.5	0.7	3.3	0.405[6]	N/A	0.935	1.487	-0.96	0.33	-0.75	0.24	0.66
	Utrecht	10	link	link	3.3	3.7	3.3	0.562	link	0.946	1.326	-1.12	-0.29	0.08	1.01	0.53
VpR247	Cambridge	1	link	link	0.0	0.0	1.7	N/A	N/A	0.981	2.067	0.80	0.41	-0.67	-1.60	-0.03
	Frankfurt[5]	20	link	link	0.3	0.7	0.8	0.849	link	N/A	N/A	0.48	0.79	-0.75	-1.66	1.10
	Lyon[5]	20	link	link	0.5	1.0	0.9	0.837	link	0.933	3.704	0.00	0.87	-1.06	-2.78	0.30
	PDB	20	link	link	0.5	0.8	N/A	0.841	link	0.958	1.253	0.32	1.08	-0.20	-0.12	-1.72
	Paris	10	link	link	0.3	0.6	2.7	0.564	link	0.782	3.063	-2.09	-1.20	-7.40	-10.17	-23.34
	Piscataway	20	link	link	0.6	1.0	1.8	0.837	link	0.950	1.492	0.48	0.66	0.00	-0.35	-2.60
	Seattle[5]	5	link	link	1.1	1.4	1.4	0.621[6]	N/A	0.964	1.281	-0.32	1.16	0.97	1.83	0.90
	Utrecht	10	link	link	9.8	7.9	14.6	0.588	link	0.957	1.400	-1.12	0.37	1.34	2.42	0.34

[1] Jorge A. Vila,Harold A. Scheraga:Assessing the Accuracy of Protein Structures by Quantum Mechanical Computations of $^{13}C^{\alpha}$ Chemical Shifts. It is expected that structure calculation algorithm based on chemical shift data would be preferred.
[2] Calculation based on ordered residues
[3] RMSD between average model of automated structure and average model of PDB deposited structure,calculation based on order residues of PDB deposited structure
[4] Superimposition of automated structures with PDB deposited structures,image generated by Pymol. PDB deposited structure is colored green, while automated structure is colored cyan.
[5] Input files including coordinate files, peak list files, chemical shift files and RPF control file.
[6] Since coordinates file is not complete, trimmed chemical shift table was used to calculate DP-scors. It is expected DP-score would be lower than complete coordinates file.
[7] Coordinates submitted after E_NMR Workshop.

Validation – Literature

- Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486.
- Nederveen, A. J., et al. (2005). RECOORD: A recalculated coordinate database of 500+proteins from the PDB using restraints from the BioMagResBank. *Proteins* 59, 662–672.
- Spronk, C., Nabuurs, S. B., Krieger, E., Vriend, G. & Vuister, G. W. (2004). Validation of protein structures derived by NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* 45, 315–337.
- 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.
- Sippl, M. J. (1993). Recognition of errors in 3-dimensional structures of proteins. *Proteins* 17, 355–362.