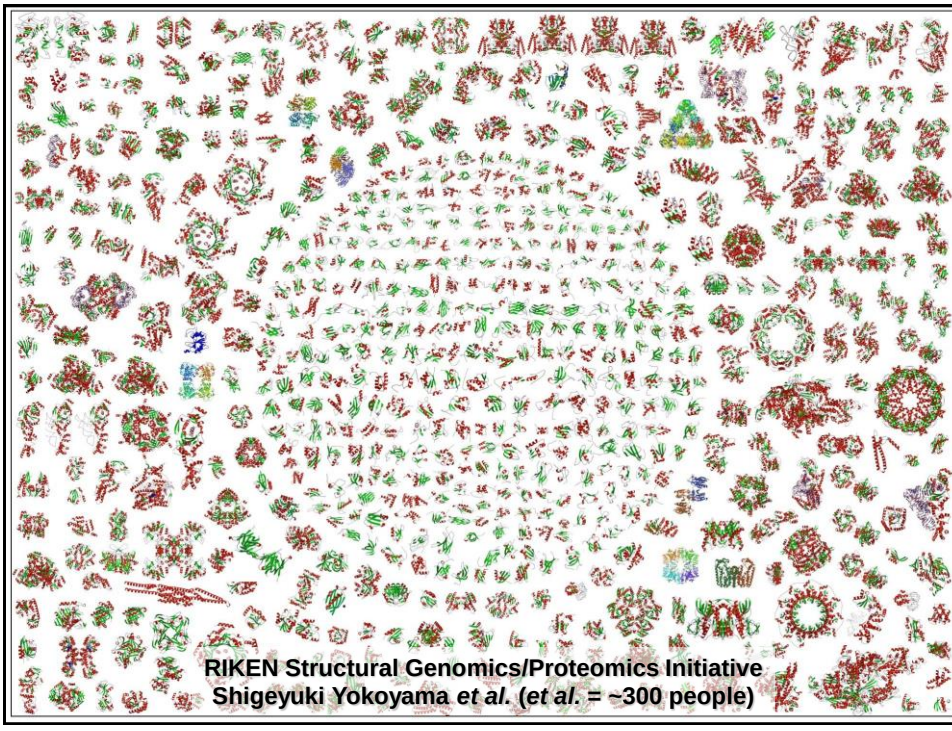
Strukturelle Bioinformatik
(M.Sc. Bioinformatik/Biochemie)

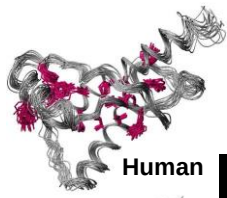
Strukturbestimmung mit NMR Spektroskopie

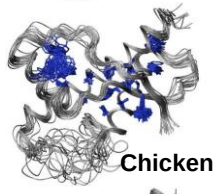
Sommersemester 2014

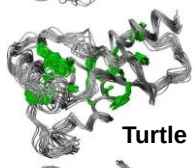
Peter Güntert

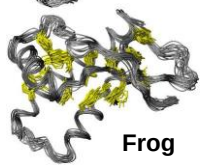


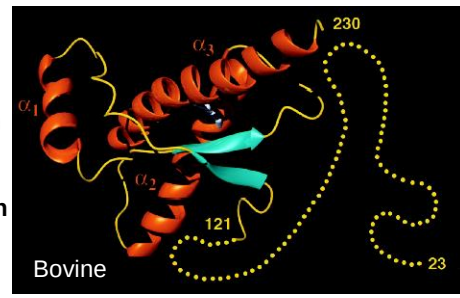
Prion proteins

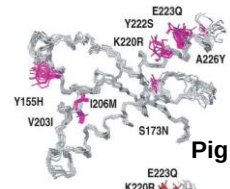

Human


Chicken

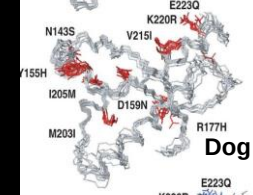

Turtle


Frog

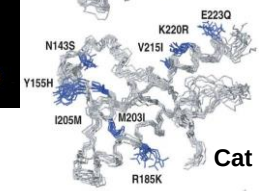




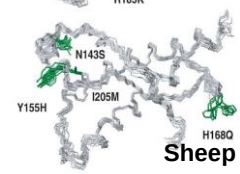
Pig



Dog



Cat



Sheep

Calzolai, L., Lysek, D. A., Pérez, D. R., Güntert, P., Wüthrich, K. *PNAS* 102, 651-655 (2005).

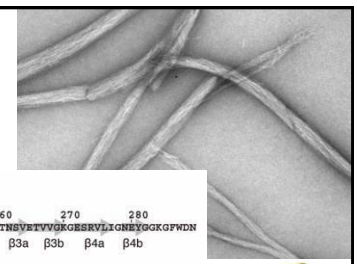
Lysek, D. A., Schorn, C., Nivon, L. G., Esteve-Moya, V., Christen, B., Calzolai, L., von Schroetter, C., Fiorito, F., Herrmann, T., Güntert, P., Wüthrich, K. *PNAS* 102, 640-645 (2005).

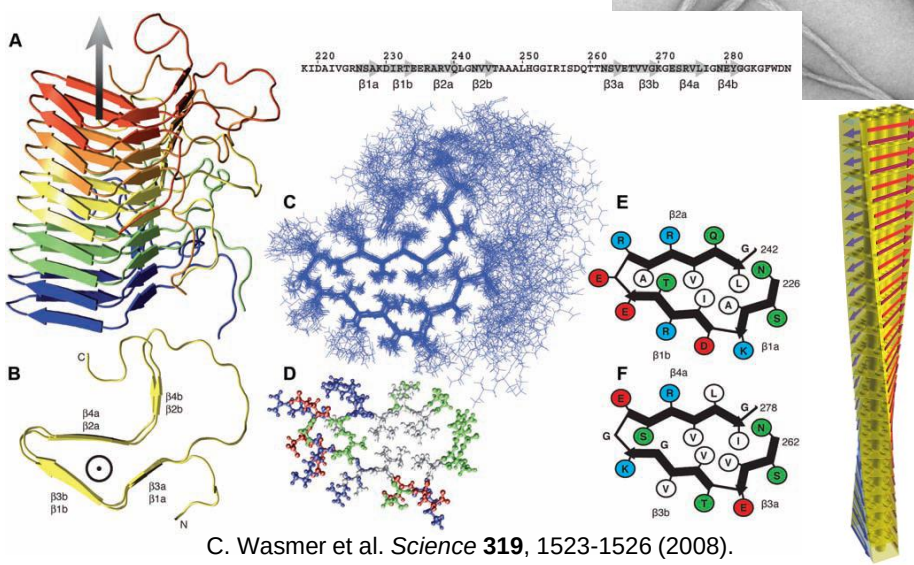
Lührs, T., Riek, R., Güntert, P., Wüthrich, K. *JMB* 326, 1549-1557 (2003).

Zahn, R., Güntert, P., von Schroetter, C., Wüthrich, K. *JMB* 326, 225-234 (2003).

Calzolai, L., Lysek, D. A., Güntert, P., von Schroetter, C., Riek, R., Zahn, R., Wüthrich, K. *PNAS* 97, 8340-8345 (2000).

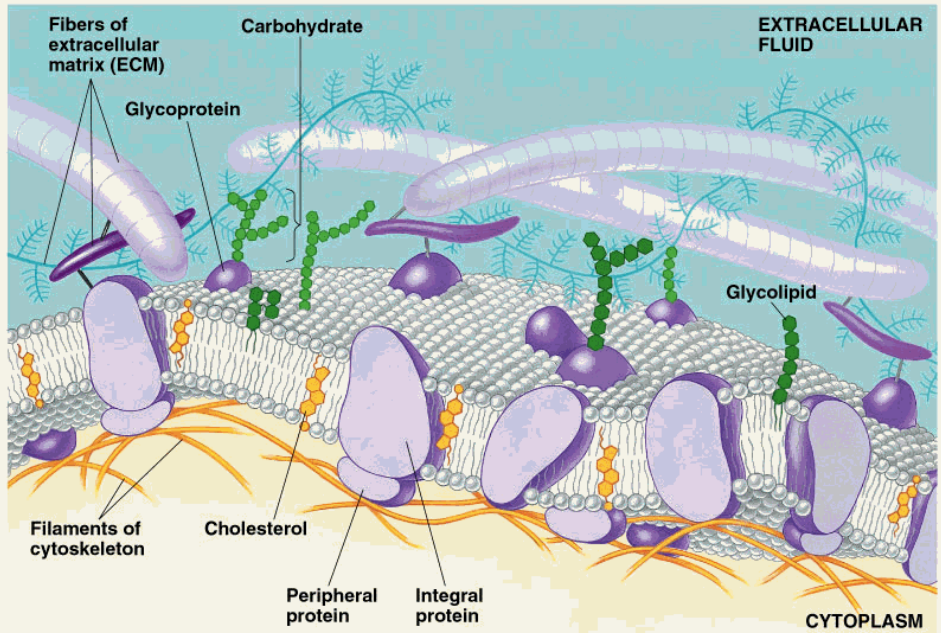
Structure of HET-s prion amyloid fibrils





C. Wasmer et al. *Science* **319**, 1523-1526 (2008).

Membrane proteins



Membrane protein structure determination: Proteorhodopsin NMR structure

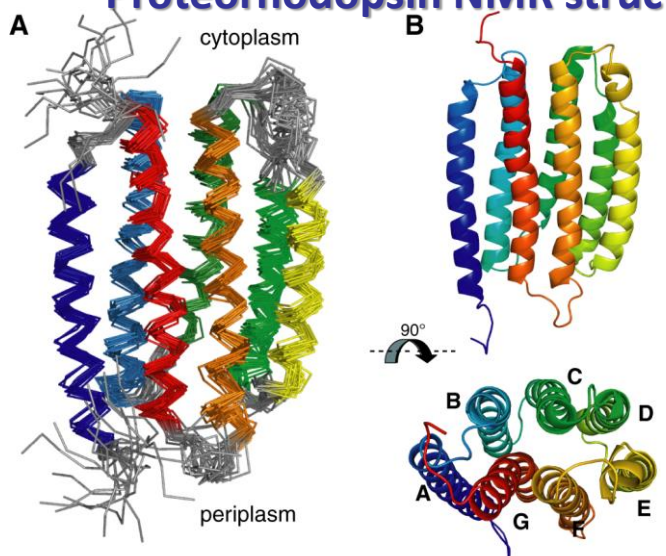
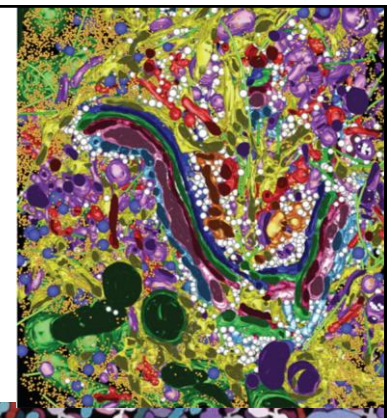


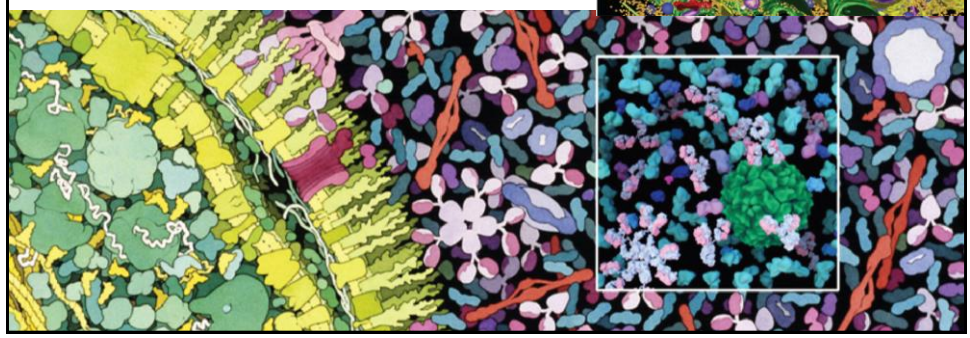
Fig. 2. Structure of PR. (A) Bundle of the 20 conformers with lowest CYANA target function obtained from structure calculation. Helices are color-coded from helix A in dark blue to helix G in red. (B) Cartoon representation of the conformer with the lowest CYANA target function seen from the side and from the top. In the lower panel helices are additionally labeled A-G.

Reckel, S., Gottstein, D., Stehle, J., Löhr, D., Verhoefen, M. K., Takeda, M., Silvers, R., Kainosho, M., Glaubitz, C., Wachtveitl, J., Bernhard, F., Schwalbe, H., Güntert, P. & Dötsch, V., Angew. Chem. (2011).

Cellular interior



3D model of a cryo-electron tomography image of the Golgi region of an insuline-secreting HIT-T15 cell. The Golgi complex with its cisternae is shown in the center.



Artistic representation of an *E. coli* cell (cellular interior in light green, cell membrane in yellow) in blood serum (pink to violet). The inset is a 3D model created from experimentally determined protein structures. Serum albumin is shown in turquoise. Y-shaped molecules and the large complex at lower left are antibodies. A poliovirus particle is depicted in green.
Y. Ito & P. Selenko. Cellular structural biology. Curr. Opin. Struct. Biol. 20, 640–648 (2010)

In-cell NMR structure determination

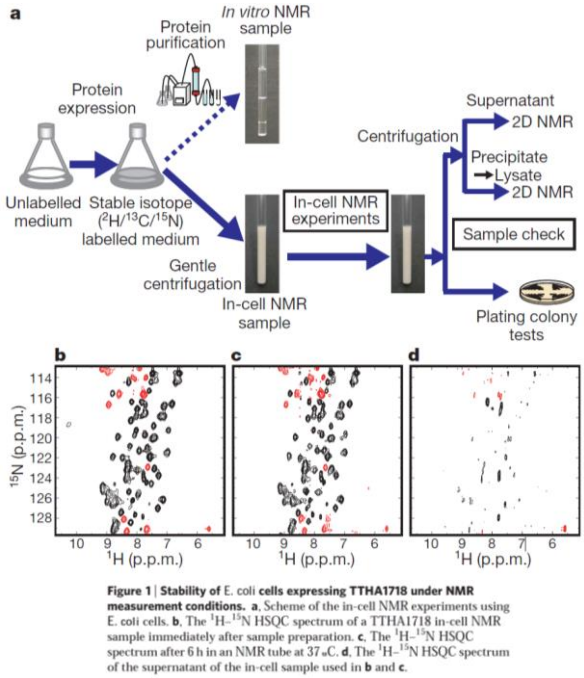


Figure 1 | Stability of E. coli **cells expressing TTHA1718 under NMR measurement conditions.** **a**, Scheme of the in-cell NMR experiments using E. coli cells. **b**, The ^{1}H–^{15}N HSQC spectrum of a TTHA1718 in-cell NMR sample immediately after sample preparation. **c**, The ^{1}H–^{15}N HSQC spectrum after 6 h in an NMR tube at 37 °C. **d**, The ^{1}H–^{15}N HSQC spectrum of the supernatant of the in-cell sample used in **b** and **c**.

Yutaka Ito
Tokyo Metropolitan University

Sakakibara *et al.*,
Nature 458, 102-105 (2009)

In-cell NMR structure of TTHA1718

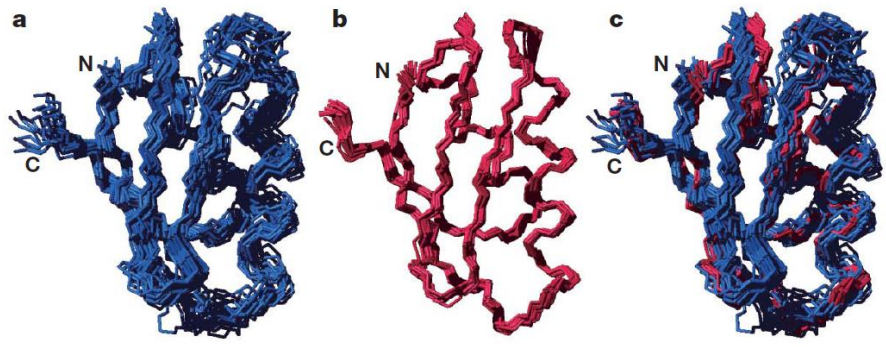


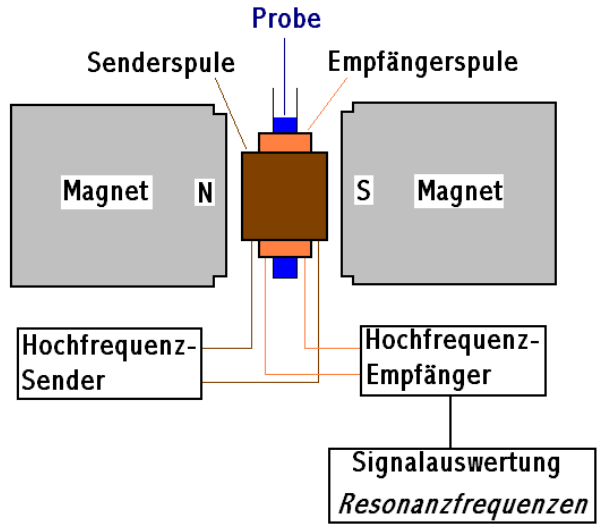
Figure 4 | NMR solution structure of TTHA1718 in living *E. coli* cells. **a**, A superposition of the 20 final structures of TTHA1718 in living *E. coli* cells, showing the backbone (N, Cα, C′) atoms. **b**, A superposition of the 20 final structures of purified TTHA1718 *in vitro*. **c**, A comparison of TTHA1718 structures in living *E. coli* cells and *in vitro*. The best fit superposition of backbone (N, Cα, C′) atoms of the two conformational ensembles are shown

Sakakibara *et al., Nature* 458, 102-105 (2009)

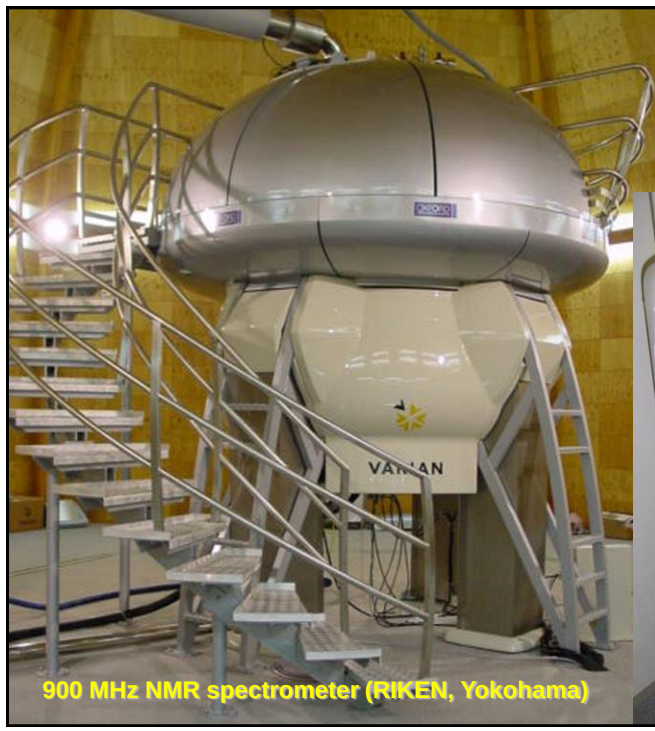
NMR Spektroskopie: Geschichte

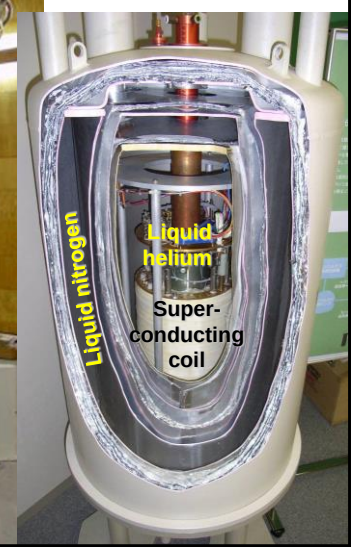
1924, Wolfgang Pauli: Vorhersage des Kernspins
1933, Isidor Rabi: Molekularstrahlmagnetresonanzdetektion
1945: Edward Purcell, Felix Bloch: Kernspinresonanz (NMR)
1953: A. Overhauser, I. Solomon: Nuclear Overhauser Effekt
1966, Richard Ernst: Fouriertransformations-NMR
1971, Jean Jeener: 2D NMR Spektren
1981, Kurt Wüthrich et al.: Resonanzzuordnung in Proteinen
1984, Kurt Wüthrich et al.: 3D Proteinstruktur in Lösung
1991, Ad Bax et al.: Tripelresonanzspektren (^{13}C, ^{15}N, ^{1}H)
1997: TROSY, NMR Spektroskopie von großen Proteinen
2014: ~10400 NMR Strukturen in der Protein Data Bank

NMR Spektrometer



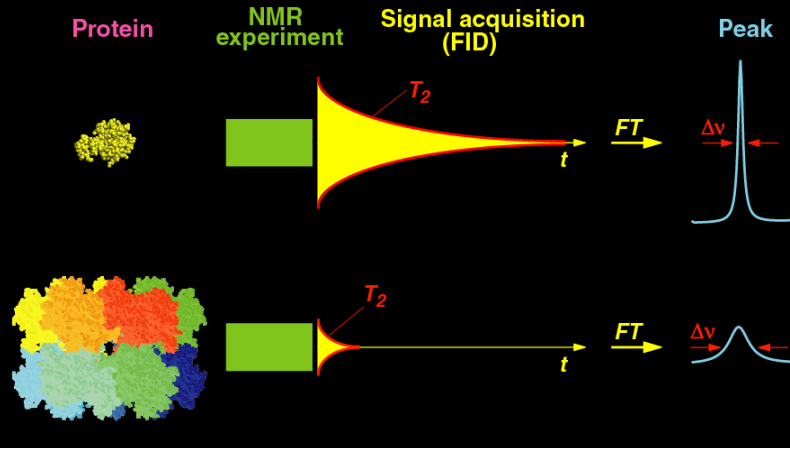
NMR Spectrometer



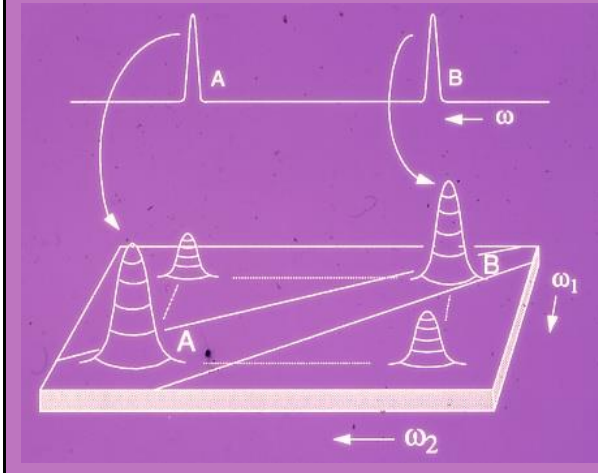


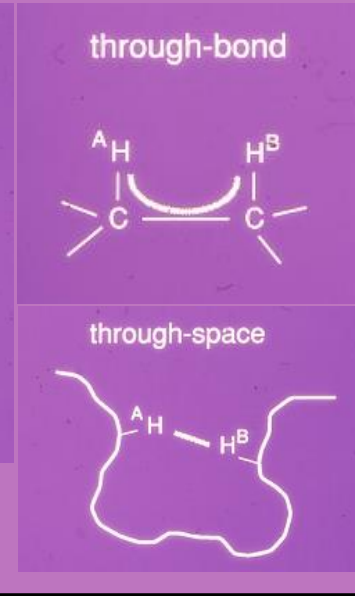
NMR with large proteins

- Large number of signals → crowded spectra
- Fast transverse relaxation → broadened signals

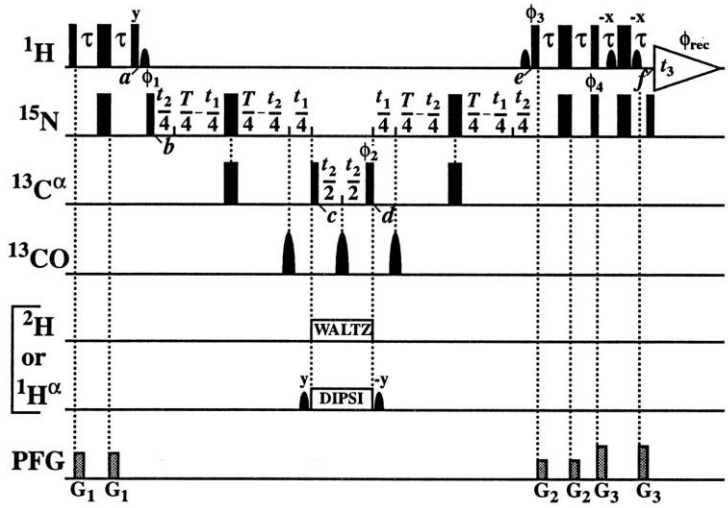


2D NMR Spectra





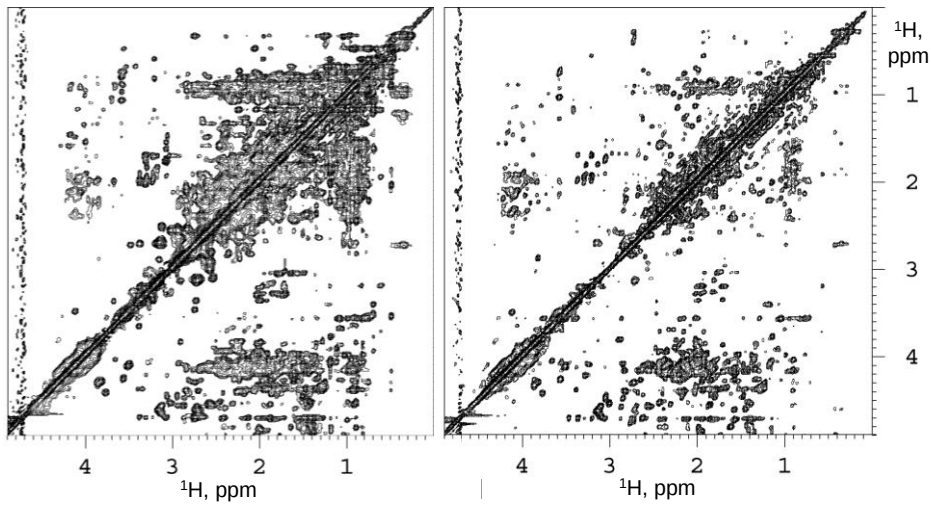
Experimental scheme for the [^{15}N,^{1}H]-TROSY-HNCA experiment

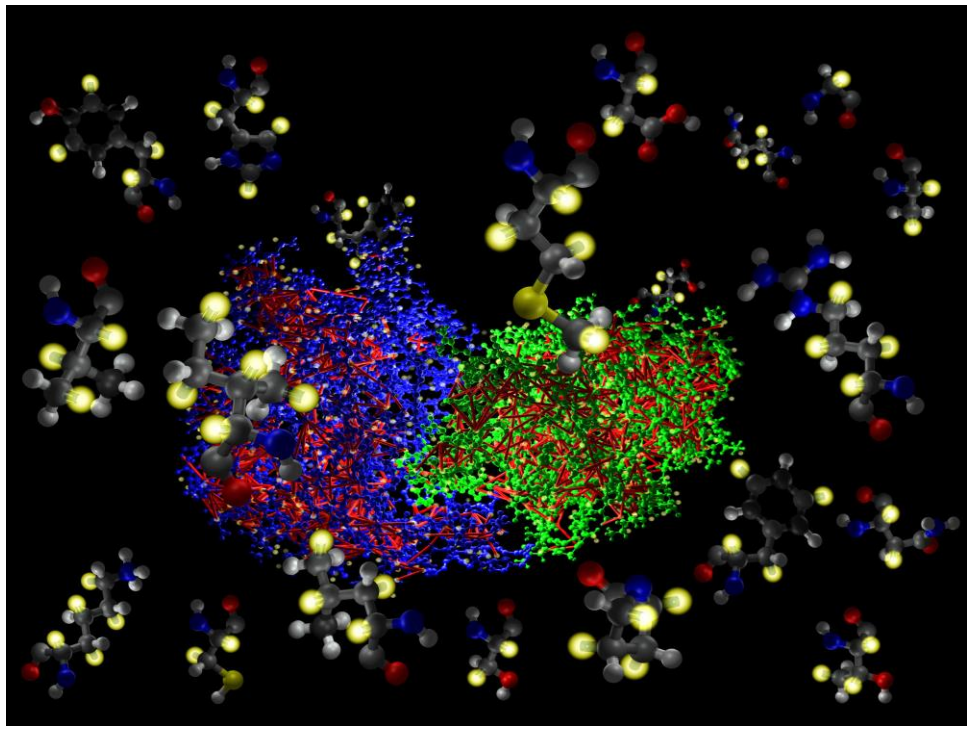


Salzmann M et al. PNAS 1998;95:13585-13590

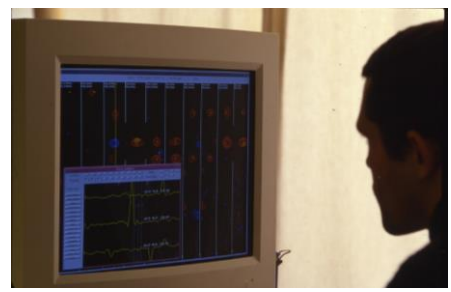
Calmodulin NOESY spectra

uniformly labelled **SAIL**





NMR Spektrenauswertung

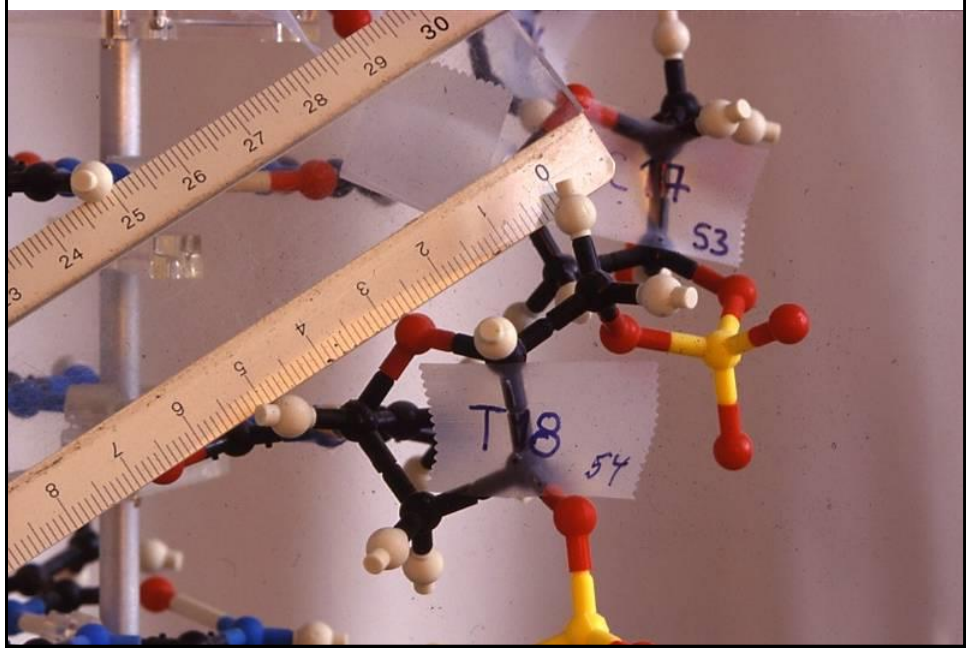


Manuell

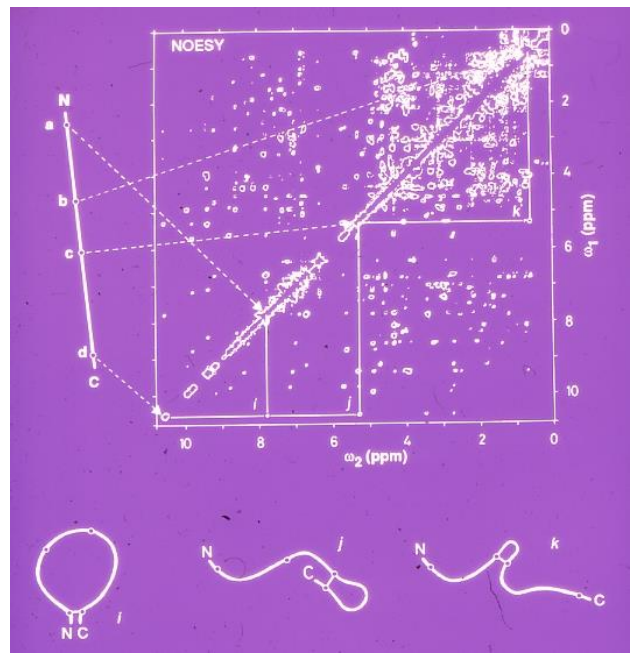
Interaktiv

Automatisch

NMR measures distances between atoms



NOESY Spektrum



Konformationsdaten aus NMR Messungen

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

...

Experimental data Systems

- NOEs
  Hydrogen bonds
  Paramagnetic relaxation enhancement
  ambiguous NOEs; docking (HADDOCK)
  "exact" NOEs (eNOEs)
- Chemical shifts (TALOS)
  Scalar coupling constants
  Ramachandran plot; rotamers
- ^3J scalar coupling constants
- Partially aligned proteins
- Paramagnetic proteins
- Partially aligned proteins
- Known size, shape
- Symmetric multimers; fibrils
- Symmetric multimers; fibrils
- Energy refinement

Conformational restraints in CYANA

- Distance restraints
  - exact distances
  - upper bounds, lower bounds
  - ambiguous distance restraints
  - ensemble-averaged restraints
- Torsion angle restraints
  - single torsion angles
  - multiple torsion angles
- ^3J scalar coupling constants
- Residual dipolar couplings (RDC)
- Pseudocontact shifts (PCS)
- Chemical shift anisotropy (CSA)
- Radius of gyration restraints
- Multimer identity restraints
- Multimer symmetry restraints
- AMBER force field

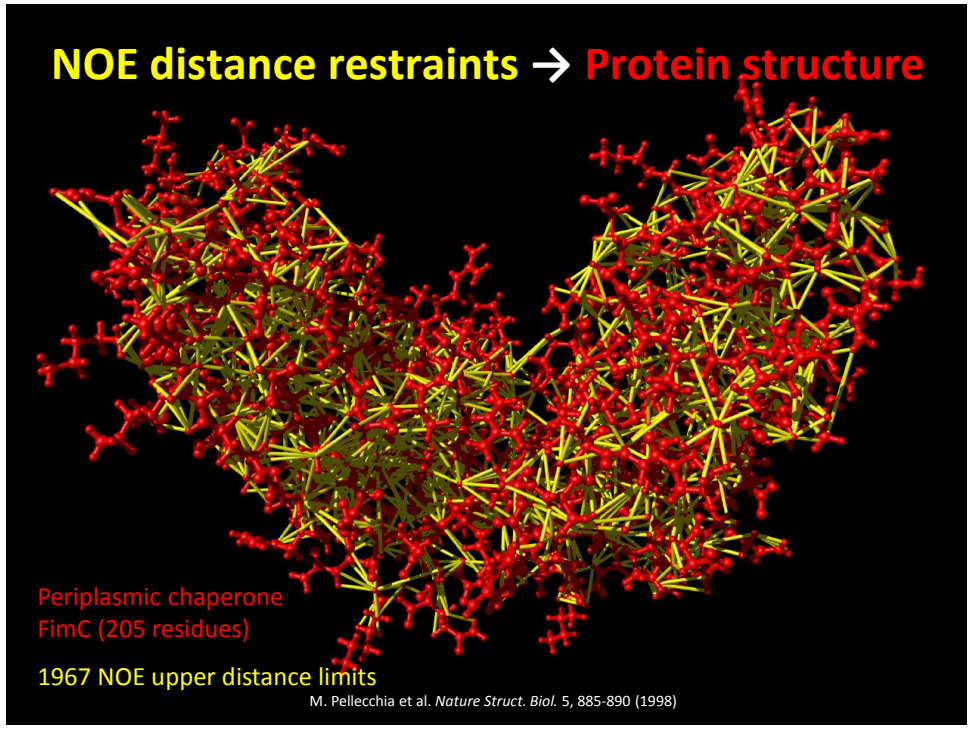
NOE (Nuclear Overhauser Effect)

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



^3J skalare Kopplungen

NMR Daten: Aufspaltung eines Signals

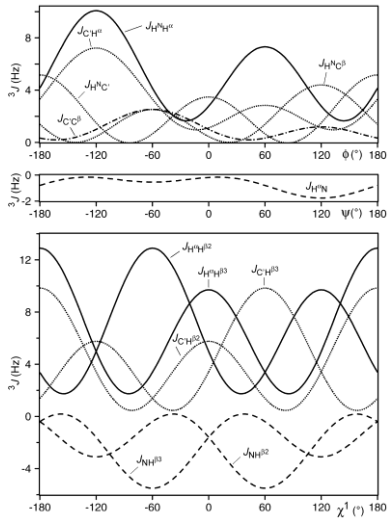
Konformationsdaten: Einschränkungen von Torsionswinkeln, θ

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

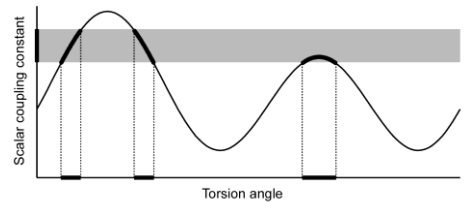
Zum Beispiel: $^3J_{HNH\alpha}(\phi)$, $^3J_{H\alpha H\beta}(\chi^1)$

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

^3J skalare Kopplungen



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



H-Brücken

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

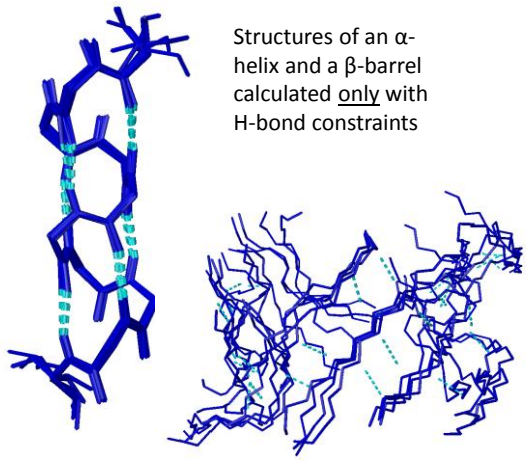
Konformationsdaten: Donor-Akzeptor Distanz

Typische H-Brücken: -N-H · · · O=C- in regulären Sekundärstrukturen (Helices, β-Blätter)

Eigenschaften:

- Bzgl. Sequenz mittel- und langreichweitig
- Donor (H) identifizierbar
- Akzeptor (O) 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

Three principal challenges of NMR protein structure analysis

1. **Efficiency**
   Spectrum analysis requires (too) much time and expertise.
2. **Size limitation**
   Structures of proteins > 30 kDa are very difficult to solve.
3. **Objectivity**
   Agreement between structure and raw NMR data?

Computational tasks in NMR structure determination

Peak picking	→	Signal frequencies
Shift assignments	→	Spin frequencies
NOESY assignment	→	Structural restraints
Structure calculation	→	3D structure
Refinement, validation	→	Final structure

Use of automation for different stages of PDB NMR structures

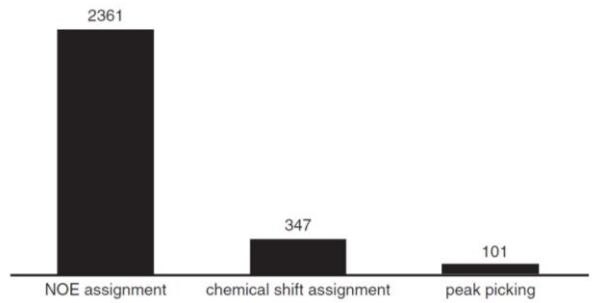


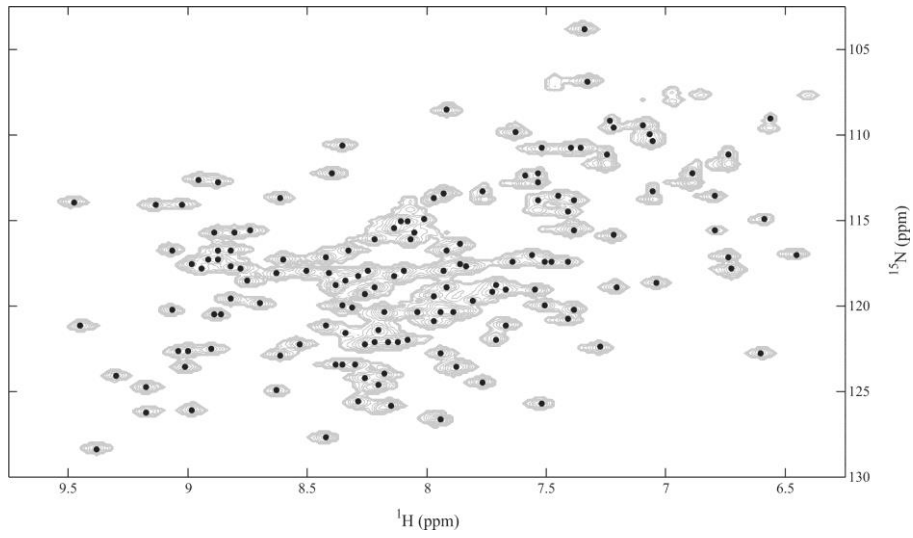
Fig. 4. The use of automation – in terms of PDB depositions – for the different stages of the traditional protocol for NMR protein structure determination. The histograms represent the number of structures returned when searching the PDB for one of the programs published for the respective stages. Exact search strings can be found in the Appendix (Tables A1, A2 and A3).

Guerry, P. & Herrmann, T. *Q. Rev. Biophys.* **44**, 257-309 (2011).

Computational tasks in NMR structure determination

Peak picking	→	**Signal frequencies**
Shift assignments	→	Spin frequencies
NOESY assignment	→	Structural restraints
Structure calculation	→	3D structure
Refinement, validation	→	Final structure

Peak picking



Alipanahi et al. Bioinformatics 25:i268-i275 (2009)

Automatically picked peaks for the protein ENTH

Spectrum	Expected peaks	Measured peaks [%]	Missing peaks [%]	Artifact peaks [%]	Deviation
^{15}N-HSQC	164	138	14	58	0.138
^{13}C-HSQC	685	113	12	51	0.434
HNCO	134	150	12	63	0.308
HN(CA)CO	269	74	35	16	0.449
HNCA	274	116	18	39	0.331
HN(CO)CA	134	150	10	61	0.395
CBCANH	529	112	29	47	0.458
CBCA(CO)NH	270	149	13	63	0.405
HBHA(CO)NH	365	134	35	75	0.510
(H)CC(CO)NH	451	88	34	25	0.530
H(CCCO)NH	664	56	57	21	0.673
HCCH-COSY	2469	97	66	70	0.609
(H)CCH-TOCSY	2449	136	45	93	0.568
HCCH-TOCSY	3574	44	66	20	0.632
^{15}N-edited NOESY	1776	120	47	74	0.486
^{13}C-edited NOESY	5958	144	48	103	0.495
Total	**20165**	**99**	**49**	**69**	**0.524**

Missing peaks: Percentage of expected peaks that cannot be mapped to a measured peak using the manually determined reference chemical shifts. **Artifact peaks:** Percentage of measured peaks to which no expected peak can be mapped. All percentages are relative to the number of expected peaks. **Deviation:** Root-mean-square deviation between the chemical shift position coordinates of the measured peaks to which an expected peak can be mapped and the corresponding reference chemical shift value, normalized by the chemical shift tolerances of 0.03 ppm for ^{1}H and 0.4 ppm for ^{13}C and ^{15}N.

Computational tasks in NMR structure determination

Peak picking	→	Signal frequencies
Shift assignments	→	**Spin frequencies**
NOESY assignment	→	Structural restraints
Structure calculation	→	3D structure
Refinement, validation	→	Final structure

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)

Chemical shift assignment software used for PDB NMR structures

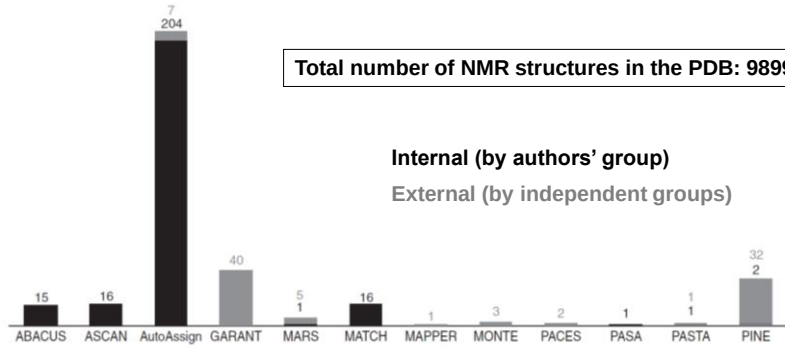


Fig. 8. Histogram plot of the number of citations from PDB depositions for selected chemical shift assignment programs as of 27 August 2010. Only those programs are listed for which at least one citation was found. Citations from internal depositions are represented in black, whereas those from external use of the program are shown in grey. In this context, we define internal depositions as those for which one or more of the developers appear as a structure author, or which originate from the same structural genomics project in which the program was developed. Note that (Bermejo & Llinás, 2010) count 31 depositions for ABACUS. Exact search strings can be found in the Appendix, Table A1.

Guerry, P. & Herrmann, T. *Q. Rev. Biophys.* **44**, 257-309 (2011).

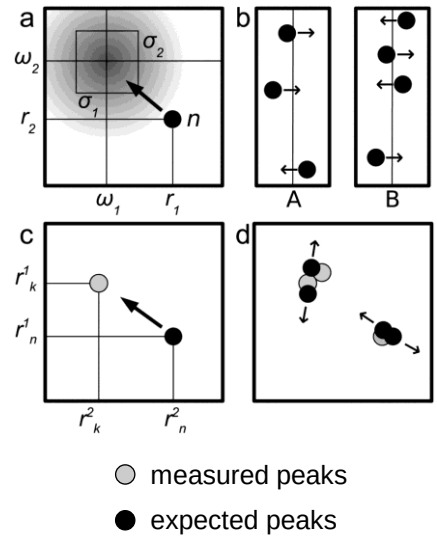
Characteristics of a correct assignment

a) Shift normality:
Chemical shifts are consistent with general chemical shift statistics.

b) Alignment:
Peaks assigned to the same atom are aligned.

c) Completeness:
As many peaks as possible are assigned.

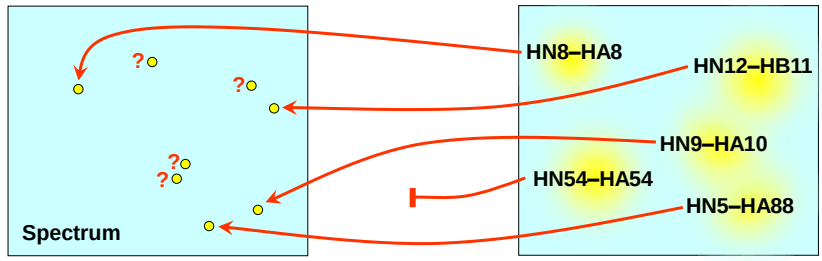
d) Low degeneracy:
The number of degenerate peaks is small.



FLYA Automated Assignment Algorithm

Observed peaks
Position known
Assignment unknown

Expected peaks
Assignment known
Position known only approximately



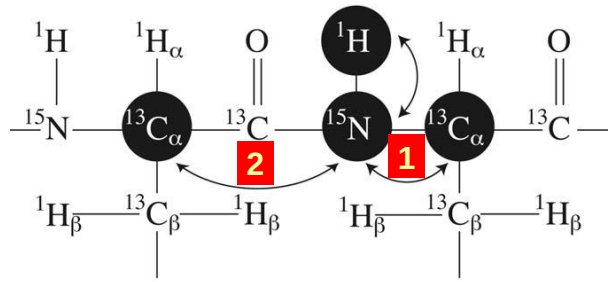
Assignment = Find **mapping** between expected and observed peaks.

Score for assignment
Presence of expected peaks
Alignment of peaks assigned to the same atom
Normality of assigned resonance frequencies

Optimization of assignment
Evolutionary algorithm combined with local optimization

Elena Schmidt
J. Am. Chem. Soc. 134, 12817-12829 (2012)
Christian Bartels *et al.*
J. Comp. Chem. 18, 139–149 (1997)
J. Biomol. NMR 7, 207–213 (1996)

Generation of expected peaks
Example: HNCA experiment

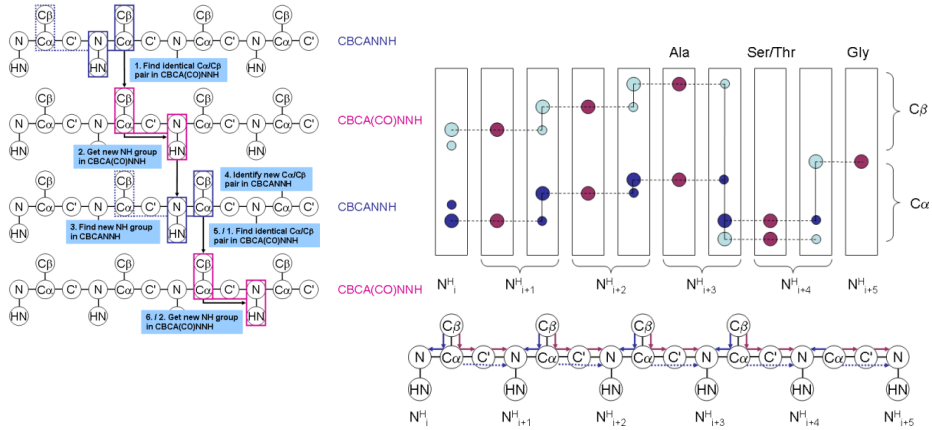


Magnetization path entries in CYANA library:

```
SPECTRUM HNCA
1  0.98 H_AMI N_AMI C_ALI
2  0.80 H_AMI N_AMI C_BYL C_ALI
```

Observation probability

Sequential assignment
with triple resonance spectra



FLYA: Spectra types

Triple resonance (backbone assignment)	Through-bond (2D & side-chains)	Through-space (NOESY)	Solid-state NMR
• H_CA_NH	• COSY	• NOESY	• NCACB
• HNCA	• TOCSY	• D2ONOESY	• NCACALI
• iHNCA	• D2OCOSY	• N15NOESY	• NCOCACB
• HN_CO_CA	• D2OTOCSY	• C13NOESY	• CANCOCA
• HN_CA_CO	• C13H1 HSQC	• C13NOED2O	• CANCO
• HNCO	• N15H1 HSQC	• CCNOESY	• NCACO
• HCACO	• CB_HARO	• CNNOESY	• CCC
• HCA_CO_N	• N15TOCSY	• NNNOESY	• NCACX
• CBCANH	• HCCH TOCSY		• NCOCA
• CBCACONH	• HCCH COSY		• NCOCA
• HBHACONH	• CCH	2D	• NCOCX
• HNHB	• C_CO_NH	3D	• DARR
• HNHA	• HC_CO_NH	4D	• DREAM
	• HC_CO_NH_4	*n*D	• PAIN
	• APSY		• NHHC

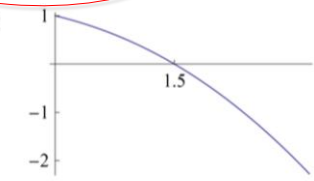
FLYA: Global assignment score

assigned atoms; shift normality; mapped peaks; peak alignment; degeneracy

$$G = \frac{\sum_{a \epsilon A}\left[w_1(a)Q_1(a) + \sum_{n \epsilon N'_a} w_2(a,n)Q_2(a,n)/b(n)\right]}{\sum_{a \epsilon A_0}\left[w_1(a) + \sum_{n \epsilon N_a} w_2(a,n)\right]}$$

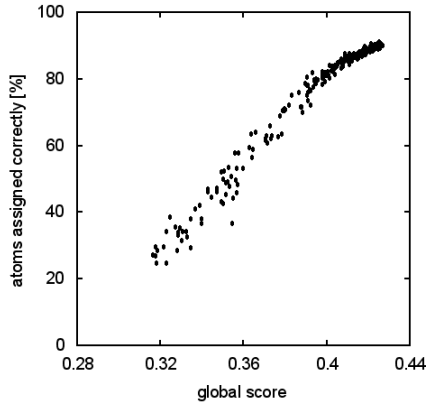
atoms with expected peaks; weight; expected peaks for atom *a*; weight

- Quality measures Q are designed such that
  - $Q = 1$ for a perfect match
  - $Q < 1$ in all other cases
  - $Q = 0$ for a deviation considered "as bad as no assignment"
  - $Q = -\infty$ for an infinitely large deviation

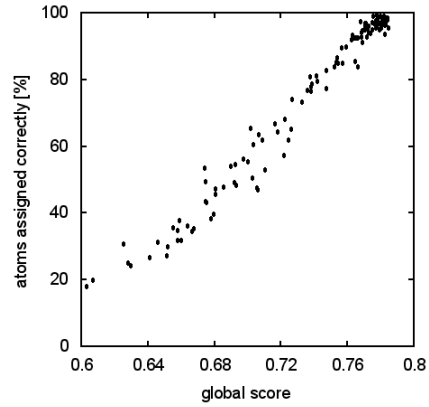


- Hence, the global score G is normalized such that
  - $G = 1$ for a perfect assignment of all atoms
  - $G < 1$ in all other cases
  - $G = 0$ if, for instance, there are either no assignments at all or if all assignments have deviations "as bad as no assignment"
  - $G < 0$ is in principle possible for (very) bad assignments.

Correlation between global score and percentage of correctly assigned atoms



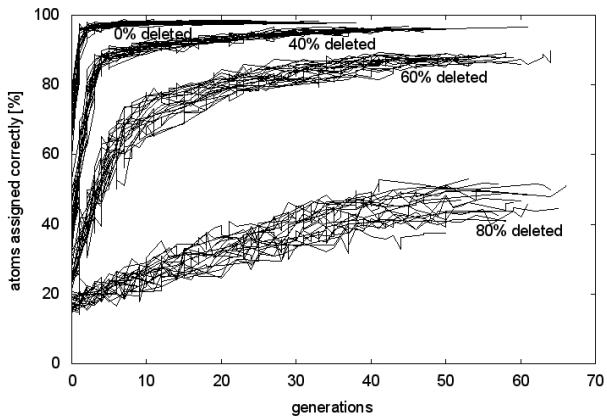
Standard calculation with the full set of 15 peak lists for SH2



Calculation with 7 experiments for the backbone assignment

Data points refer to the current best scored solutions, which were saved during the calculation.

FLYA: Evolutionary optimization



Higher quality input data

↓

More correct assignments

Faster convergence

Less divergence among individual runs

20 calculations each, using simulated data for SH2 (15 spectra) with chemical shift tolerance 0.04 ppm for ^{1}H, 0.4 ppm for $^{13}C/^{15}N$, 0–80% missing peaks, and no additional artifact peaks.

FLYA: Consensus chemical shifts

- Ensemble of *n* independently calculated chemical shift values $\omega_1,\ldots,\ \omega_n$ for each nucleus:

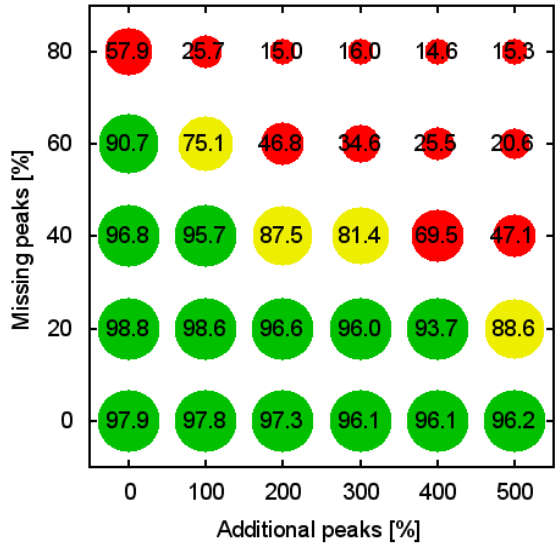
- **Consensus chemical shift:** Value ω that maximizes the function

$$\mu(\omega) = \frac{1}{n}\sum_{j=1}^{n} \exp\left(-\frac{1}{2}\left(\frac{\omega-\omega_j}{\Delta\omega}\right)^2\right)$$

$\Delta\omega$ = chemical shift tolerance, e.g. 0.03 ppm for ^{1}H, 0.4 ppm for ^{13}C/^{15}N

- Most individual shifts $\omega_1,\ldots,\ \omega_n$ near consensus value
  → **"strong" (self-consistent) assignment**
  Otherwise → "weak" (tentative) assignment

FLYA: Assignment accuracy *vs.* quality of input data



>90% correct
70–90% correct
<70% correct

Calculations using simulated data for SH2 (15 spectra) with 0–80% missing peaks and 0–500% additional artifact peaks.

Chemical shift tolerance:
0.04 ppm for ^{1}H
0.4 ppm for ^{13}C/^{15}N

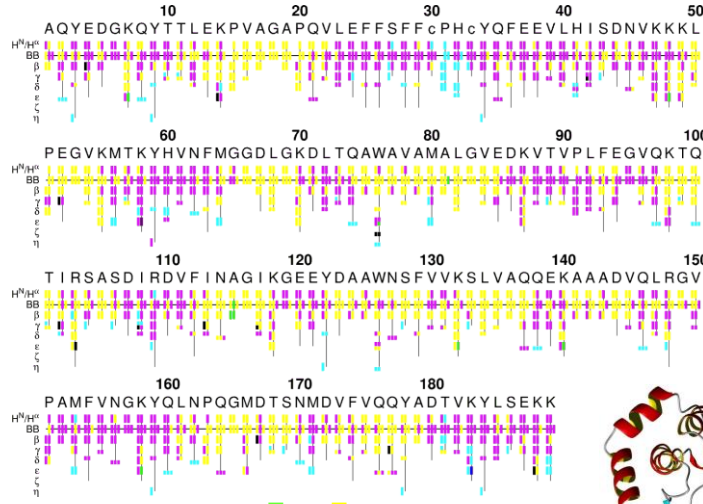
Elena Schmidt
JACS 134, 12817-12829 (2012)

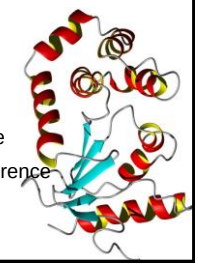
DsbA automated assignment with FLYA

189 aa; SAIL

Spectra:

CBCACONH
CBCANH
C_CO_NH
HBHACONH
HC_CO_NH
HN_CA_CO
HNCO
N15NOESY
C13NOESY





safe unsafe: equal to reference
safe unsafe: different from reference
safe unsafe: no reference
only reference

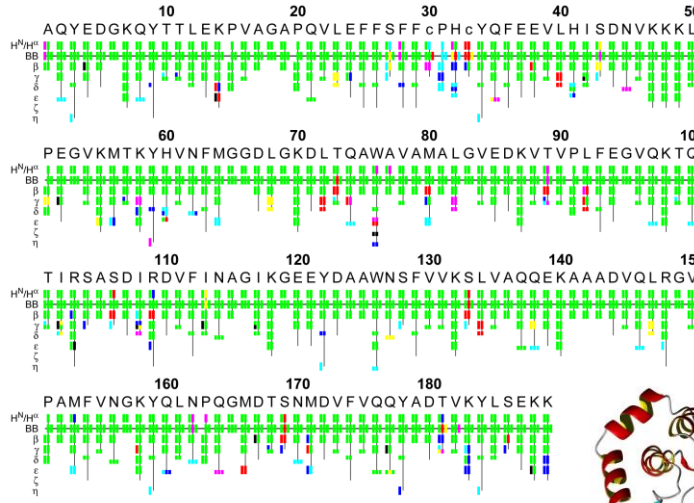
E. Schmidt
T. Ikeya, M. Takeda, M. Kainosho

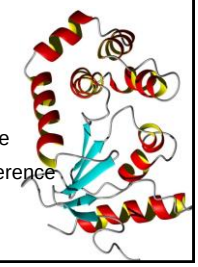
DsbA automated assignment with FLYA

189 aa; SAIL

Spectra:

CBCACONH
CBCANH
C_CO_NH
HBHACONH
HC_CO_NH
HN_CA_CO
HNCO
N15NOESY
C13NOESY





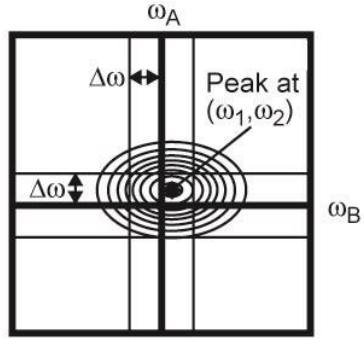
Automatically prepared peak lists
94% of all shifts correct
96% of bb + HA/HB shifts correct

strong weak: equal to reference
strong weak: different from reference
strong weak: no reference
only reference

Computational tasks in NMR structure determination

Peak picking	→	Signal frequencies
Shift assignments	→	Spin frequencies
NOESY assignment	→	**Structural restraints**
Structure calculation	→	3D structure
Refinement, validation	→	Final structure

Ambiguity of chemical shift based NOE assignment



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

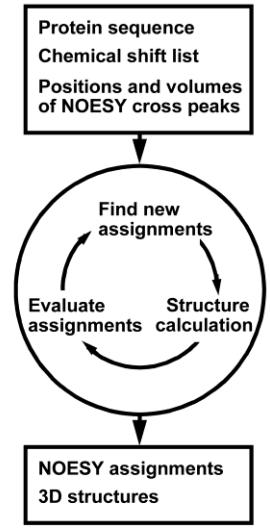
In general, several different ^{1}H chemical shifts ω_A, ω_B match the position of a NOESY peak within the experimental uncertainty $\Delta\omega$.

→ Assignment ambiguity

Manual assignment is very cumbersome!

Automated NOESY assignment and structure calculation

- Automated methods are
  - much faster
  - more objective
- Problems may arise because of
  - imperfect input data
  - limitations of the algorithms used
- Iterative process: All but the first cycle use the structure from the preceding cycle.
- The first cycle is important for the reliability of the method.



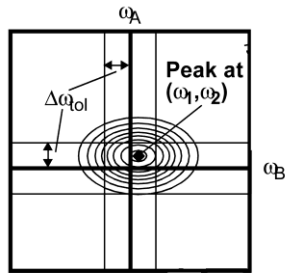
Automated NOE Assignment and Structure Calculation

- Distance restraints from not uniquely assigned NOEs:
  → Ambiguous distance restraints
- Reduction of assignment ambiguity prior to the structure calculation:
  → Network-anchored assignment
- Robustness against erroneous assignments:
  → Constraint combination

T. Herrmann, P. Güntert, K. Wüthrich. *J. Mol. Biol.* **319**, 209-227 (2002)
P. Güntert. *Prog. NMR Spectrosc.* **43**, 105-125 (2003)

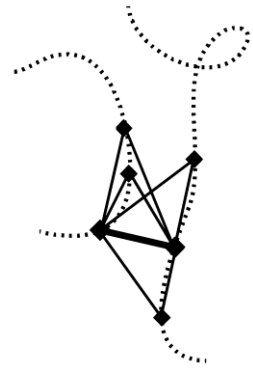
Conditions for valid NOESY assignments

Chemical shift agreement

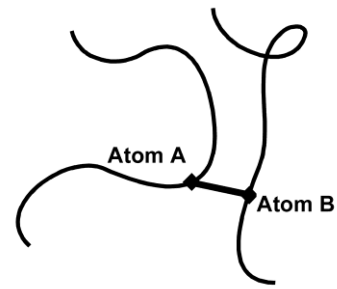


$$|\omega_1 - \omega_A| < \Delta\omega_{tol}$$
$$|\omega_2 - \omega_B| < \Delta\omega_{tol}$$

Network-Anchoring



Consistency with preliminary structure



$d_{AB} < d_{max}$

NOE assignment probability

(CYANA 2.1, 3.0)

Probability(*assignment to atoms A-B is correct*) =

Probability(*chemical shifts match*) x
Probability(*distance A-B < upper limit*) x
Probability(*other assignments predict NOE A-B*)

$$P_{tot} = P_{shift} \cdot P_{structure} \cdot P_{network}$$

Accept assignments with $P_{tot} > P_{min}$ (= 20%)

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

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

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

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

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

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

- 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

Information content of NOEs

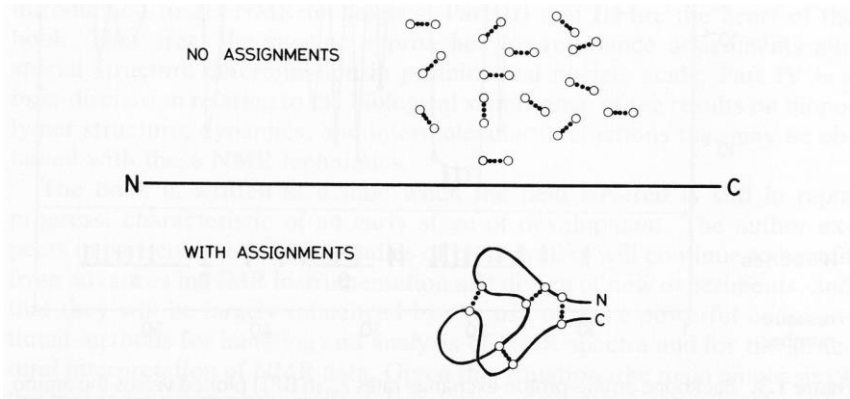
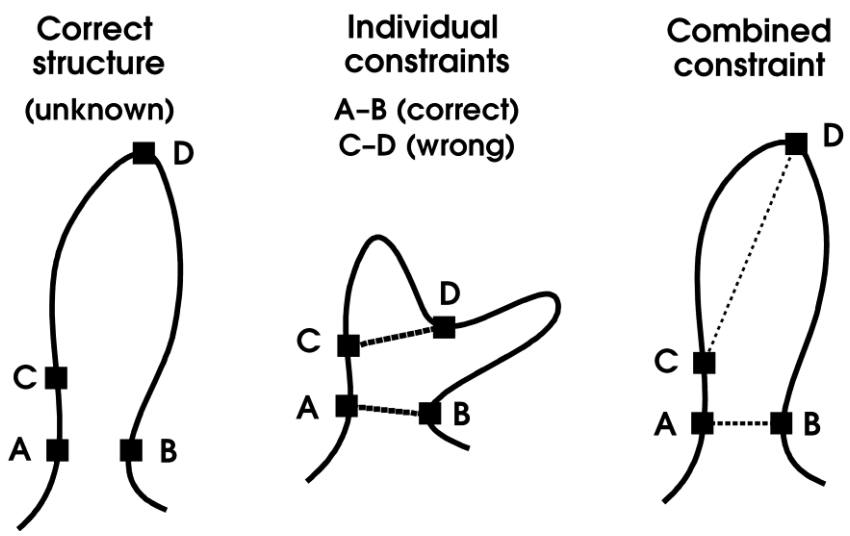


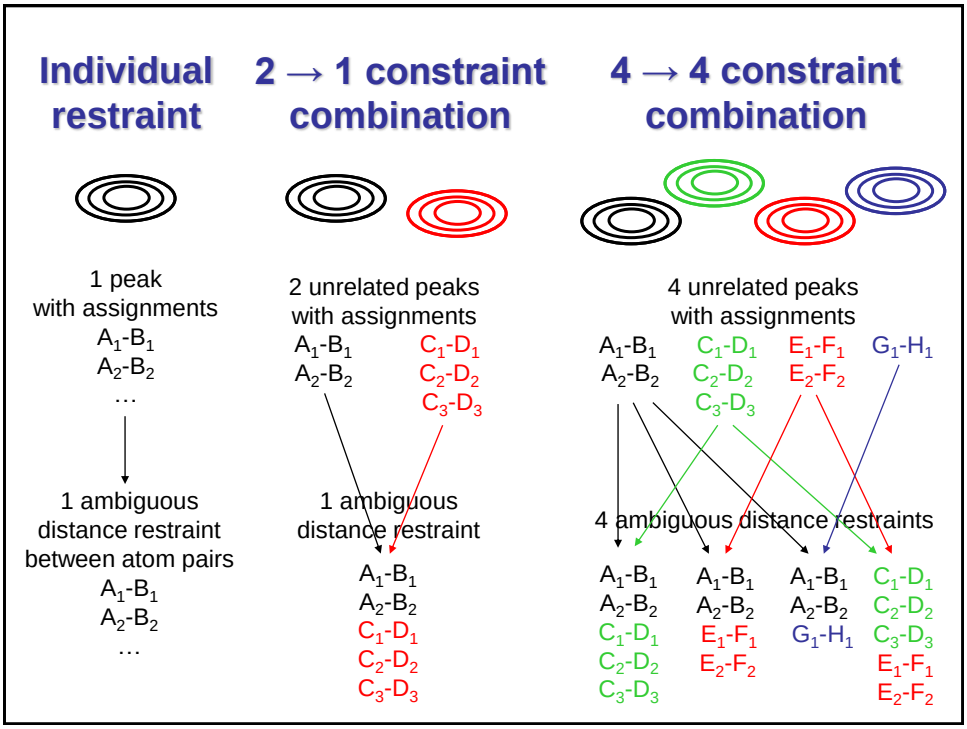
Figure 1.1. Information content of ^1H–^1H NOE's in a polypeptide chain with and without sequence-specific resonance assignments. Open circles represent hydrogen atoms of the polypeptide. The polypeptide chain is represented by the horizontal line in the center.

Constraint Combination



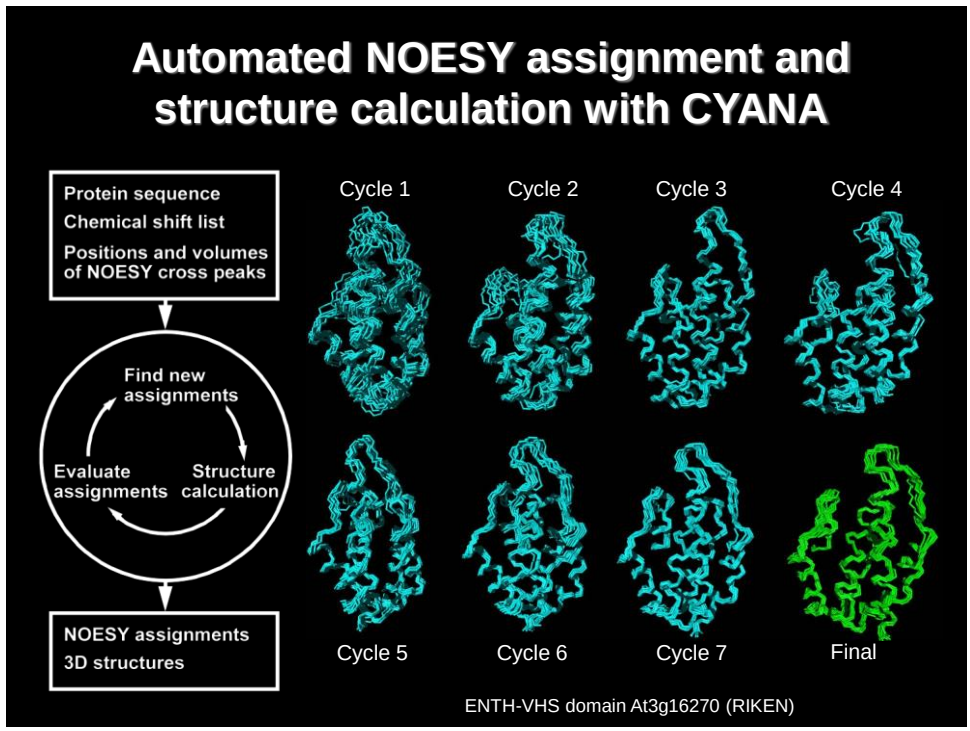
Constraint combination

- **Problem:** Peaks with wrong medium- or long-range assignments may severely distort the structure, especially in the first cycles of automated NOE assignment and structure calculation, and may lead to convergence to a wrong structure.
- **Idea:** From two long-range peaks each, combine the assignments into a single distance restraint.
  → Occurrence of erroneous restraints is reduced.



Effect of constraint combination

- Example: 1000 long-range peaks, 10% of which would lead to erroneous restraints.
- Individual restraints:
  1000 constraints, 1000 x 0.1 = 100 wrong (10 %)
- 2 → 1 constraint combination:
  500 restraints, ~500 x 0.1^2 = 5 wrong (~1%)
- 4 → 1 constraint combination:
  1000 restraints, ~1000 x 0.1^2 = 10 wrong (~1%)



Computational tasks in NMR structure determination

Peak picking	→	Signal frequencies
Shift assignments	→	Spin frequencies
NOESY assignment	→	Structural restraints
Structure calculation	→	**3D structure**
Refinement, validation	→	Final structure

Structure calculations

- Structure calculation programs try to fold a protein into a three-dimensional structure that agrees with the measured data.
- Differences between measured data and the structure are manifested as violations of conformational restraints.
- Violations cause forces that act on the molecule, driving it towards minimal (pseudo)energy and optimal agreement with the measured data.
- The target function (pseudoenergy) is the sum of squares of the violations.
- The energy landscape of this target function is complex and has many local minima.

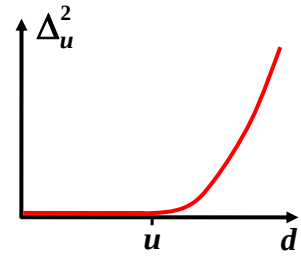
CYANA target function

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

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

Δ_u, Δ_l, Δ_a: restraint violations,

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

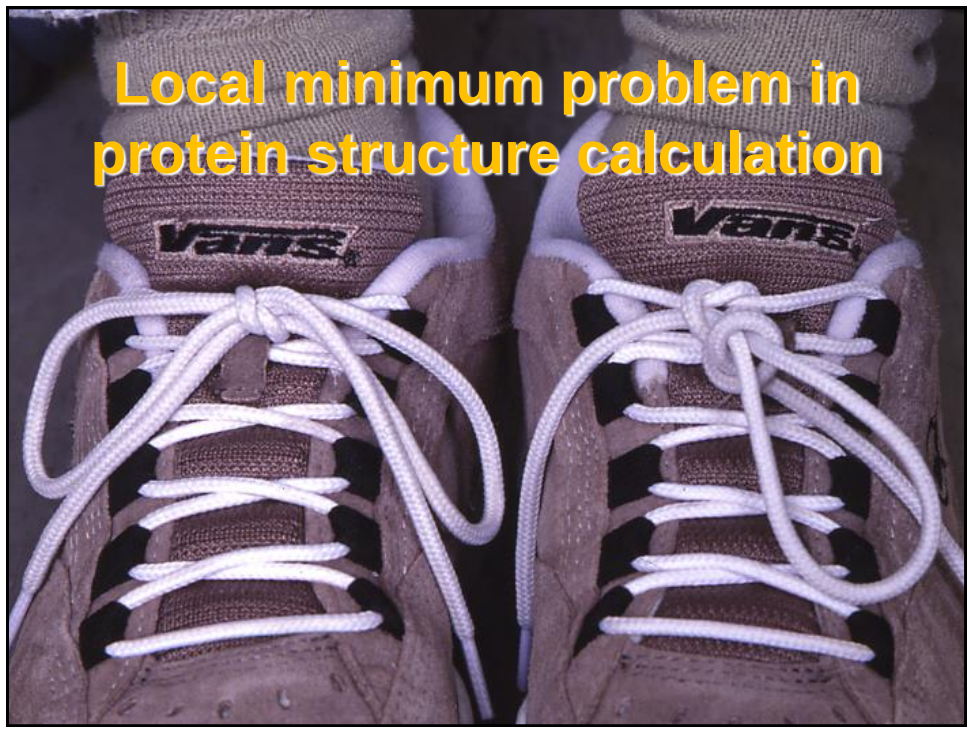


Strukturberechnungsalgorithmen

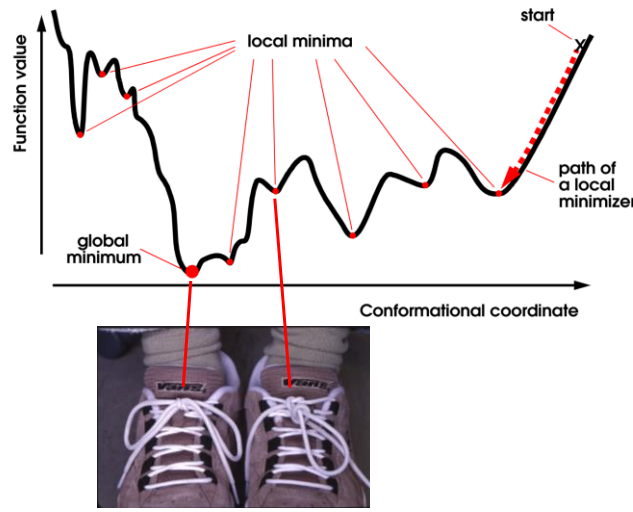
- Frühere Methoden:
  - Interaktiver Modellbau
  - Distanzgeometrie
  - Minimierung einer variablen Zielfunktion
- Simulated annealing:
  - Monte Carlo
  - Molekküldynamiksimulation im kartesischen Raum
  - Molekküldynamiksimulation im Torsionswinkelraum

Ist NMR Strukturberechnung möglich?

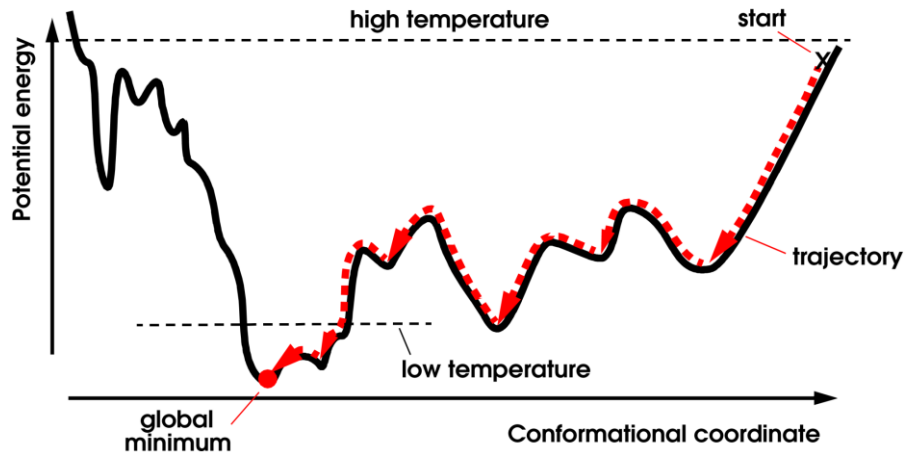
- Grundsätzlich:
  - NOEs messen nur kurze Distanzen < 5 Å
  - ungenaue obere Schranken
  - Kann damit die globale Struktur eines 30 Å großen Proteins bestimmt werden?
    JA, wenn genügend Daten da sind.
- Praktisch:
  - Zielfunktion hat viele lokale Minima
  - Kann eine (fast) optimale Struktur gefunden werden?
    JA.



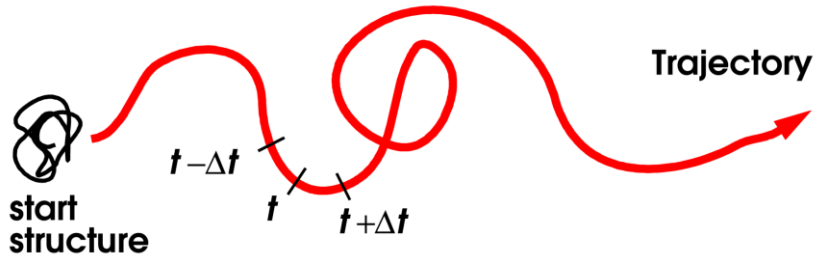
Target function = potential energy



Simulated annealing



Molecular Dynamics Simulation



Numerical integration of classical equations of motion

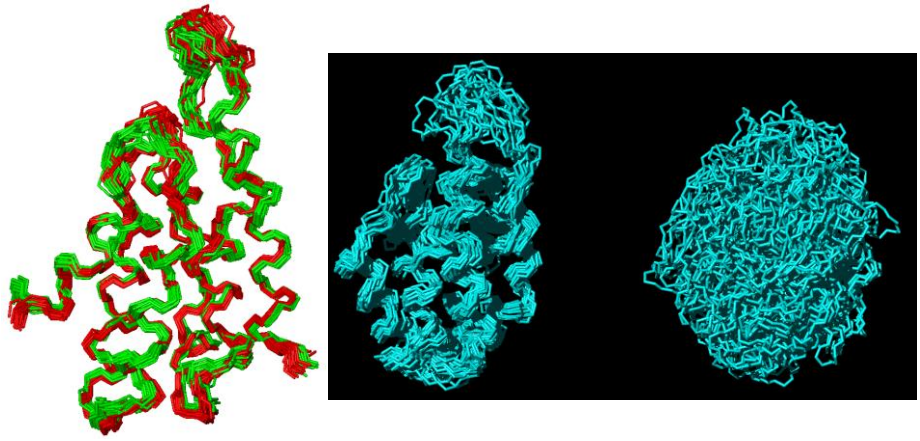
Strukturbündel

- 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?


NMR Structure Calculation:
Multiple Conformers


NMR Structure Calculation:
Select Converged Conformers

Strukturbündel



RMSD 0.8 Å RMSD 1.3 Å RMSD 6.3 Å

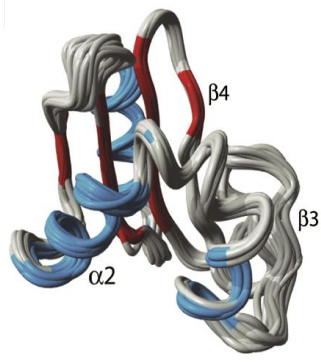
ENTH-VHS domain At3g16270

Computational tasks in NMR structure determination

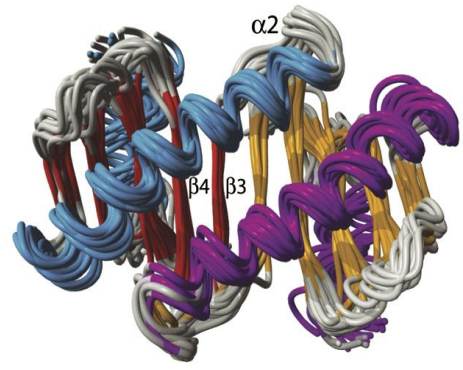
Peak picking	→	Signal frequencies
Shift assignments	→	Spin frequencies
NOESY assignment	→	Structural restraints
Structure calculation	→	3D structure
Refinement/validation	→	**Final structure**

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 $\rightarrow$ "free" *R* value, obtained from independent data
- Detects errors better than conventional *R*-factor
- Each reflection influences whole electron density
- Many reflections $\rightarrow$ 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):

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

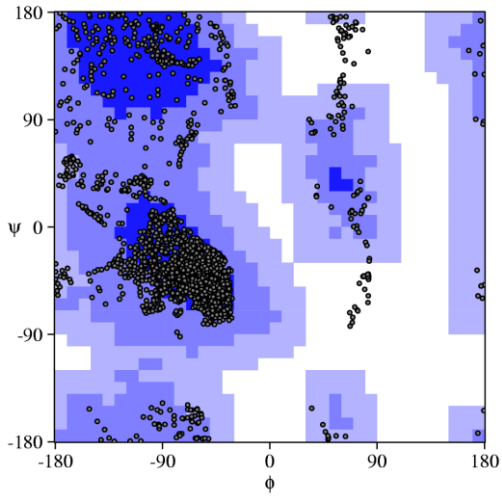
where D^{obs} and D^{calc} are observed and calculated one-bond dipolar couplings.

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.

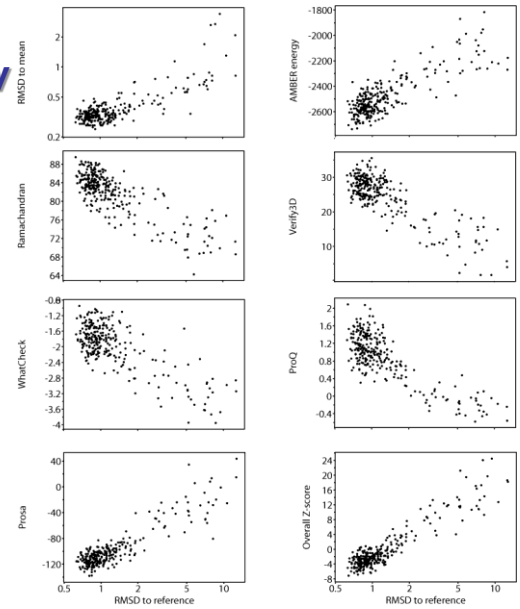
Correlation between validation parameters and structure accuracy

- 252 ubiquitin structure bundles calculated with CYANA (FLYA)
- 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

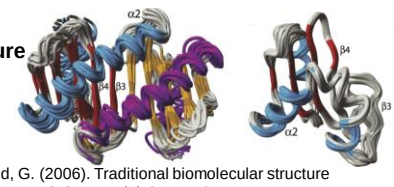


Quality indicators for correct and wrong structures of DLC2A

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 $1TGQ_{sim}$ restraints (Å)	12.8	12.6	0.521	0.0231
	Violations >0.5 Å $1TGQ_{sim}$ 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.

CASD-NMR: Critical Assessment of Structure Determination by NMR

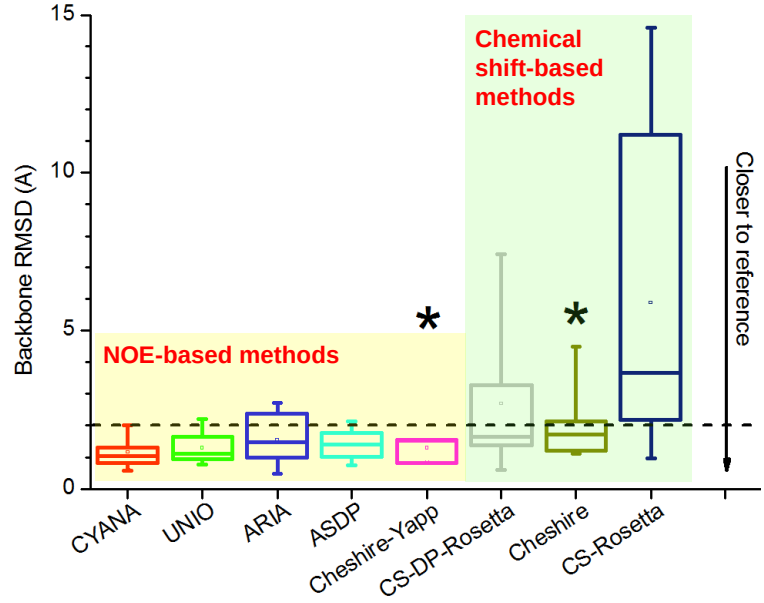
- **Evaluation of current algorithms for automated NOESY assignment and structure calculation**
- **Blind test (analogous to CASP):**
  - *NMR data are provided 8 weeks before the release of the structure by the PDB.*
  - *Structures obtained by different algorithms are collected before the original PDB structure is released.*
- **Open to anybody for providing data and for calculating structures by automated methods**
  - *In 1st round: 10 protein NMR data sets, 7 algorithms.*

http://www.wenmr.eu/wenmr/casd-nmr

Rosato, A. *et al.*, *Nature Methods* 6, 625–626 (2009)

Rosato, A. *et al.*, *Structure* 20, 227–236 (2012)

CASD-NMR results: Structure accuracy



CASD-NMR results: Correlation between accuracy and validation scores

	DP-score	Verify3D	ProsaII	Procheck (phi-psi)	Procheck (all)	MolProbity Clashscore
RMSD	-0.66	-0.14	-0.16	0.11	0.26	0.07

Literatur über über NMR Proteinstrukturbestimmung

- K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, 1986.
- J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, N. J. Skelton & M. Rance, M. *Protein NMR Spectroscopy. Principles and Practice*, Academic Press, ^{2}2006.
- M. Williamson, *How Proteins Work*, Garland, 2012.

Skript

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