Protein Structure Prediction Using Residual Dipolar Couplings

Sotirios I. Pantos
Department of Biology
National Kapodistrian University of Athens - Greece

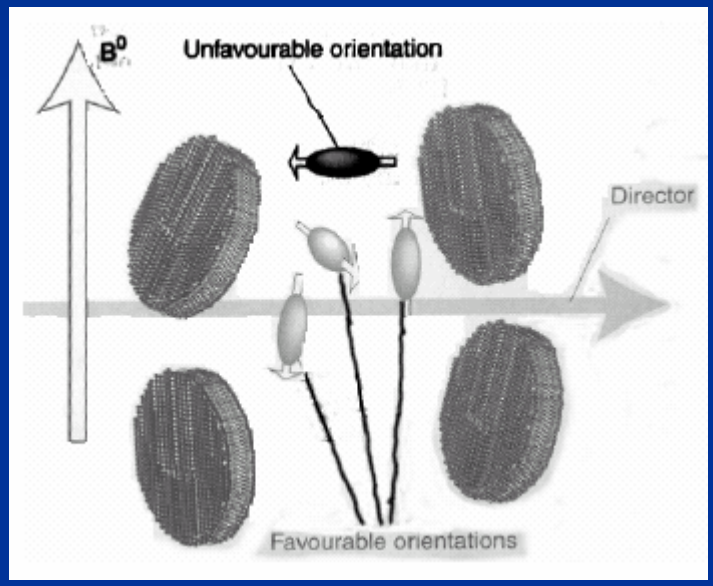
Ioannis Z. Emiris
Department of Informatics & Telecommunications
National Kapodistrian University of Athens - Greece

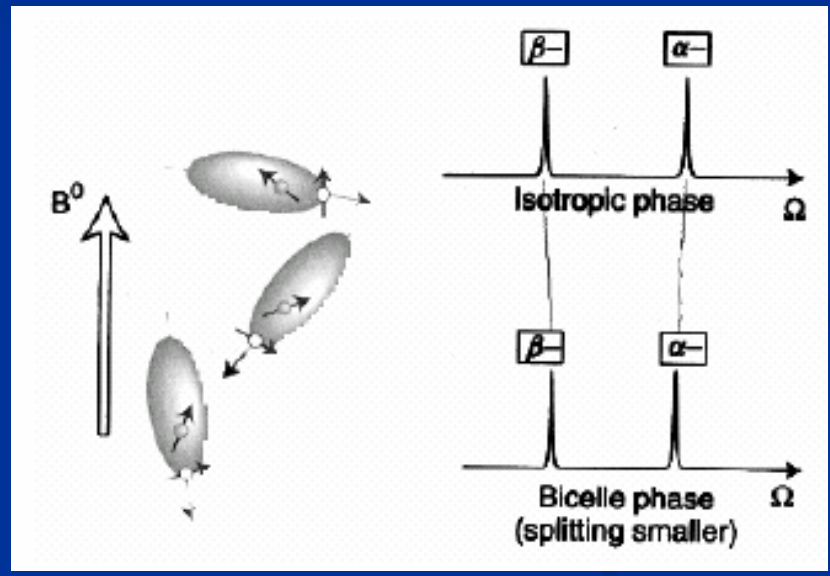
Athens
October 2007

Contents

Introduction

- Nuclear Overhauser Effect (NOE):
  - Proton dipolar coupling through space (NOESY spectra).
  - Short range distance restraints (< 6 Å).





- Residual Dipolar Coupling (RDC) [Saupe 63],[Tjandra & Bax 97]:
  - Nuclear spin coupling in anisotropic media (HSQC spectra).
  - Long range orientation restraints.

The Fundamental Equation of RDC

$$D = r^T S r$$

D: RDC measurement (Hz).

r: Internuclear unit vector relative to a molecular frame (MF).

S: Saupe matrix, 3x3, symmetric, traceless [Losonczi 99].

Saupe Matrix Determination

► A linear system determines the 5 independent Saupe elements (s).

$$\boldsymbol{As = d}$$

► ***A:*** Matrix of direction cosines of internuclear vectors.

► ***d***: Vector of RDC measurements.

Exact Algebraic Methods Using RDC

Protein Structure Validation

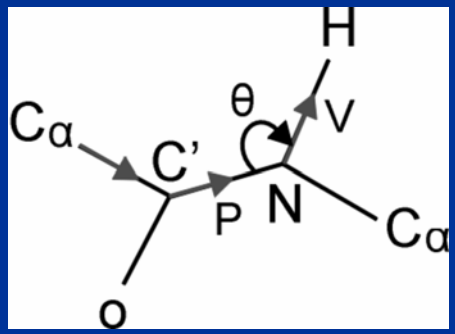
Protein Backbone Conformation

Protein Fold Determination

Bond Vector Determination from RDC Equation

$$D = r^T S r$$

Saupe diagonalization simplifies RDC equation.



$$D = P_x^2 S_{xx} + P_y^2 S_{yy} + P_z^2 S_{zz}$$

P: C-N unit vector in Principle Order Frame.

Parameterize solutions with an ellipse.

$$P_y = \xi_p \cos\tau_p$$

$$P_z = \lambda_p \sin\tau_p$$

$$P_x = \pm\sqrt{1 - p_y^2 - p_z^2}$$

$$P \cdot V = \cos(180 - \theta)$$

Constrain solutions with angle θ

$$104.3326u^4 - 70.8993u^3 - 31.3688u^2 + 13.9179u + 3.6976 = 0$$

This quartic equation is solved by MAPLE

C-N Discrete Solutions from RDC Equation

- The calculated vectors:

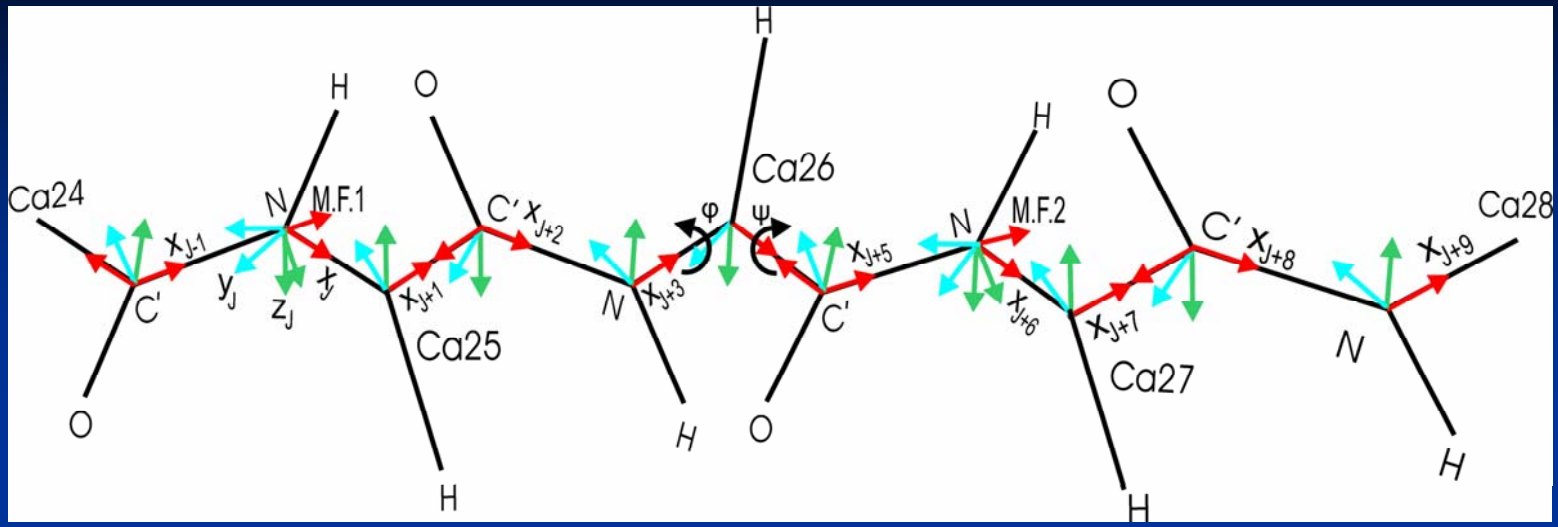
$$\text{CN25Calc=}\left[\begin{bmatrix} 0.9848434360732049 \\ 0.1388170093041705 \\ -.1039867545071002 \end{bmatrix}, \begin{bmatrix} 0.2173337871200098 \\ -.4707876292564832 \\ 0.8550584970693126 \end{bmatrix}, \begin{bmatrix} -.3291767546873323 \\ -0.02089263523876389 \\ 0.9440371631155303 \end{bmatrix}, \begin{bmatrix} 0.4284065825420640 \\ 0.8803000951946887 \\ -.2038125010267301 \end{bmatrix}\right]$$

- The experimentally observed vector:

$$\text{"CN25Obs="}, \begin{bmatrix} 0.2173383296403011 \\ -.4707743749301929 \\ 0.8550646394498999 \end{bmatrix}$$

Protein Backbone Conformation

Protein Backbone as a Robotic Arm



$$ {}^{MF\,1}\boldsymbol{T}_{MF\,2} = {}^{LAB}\boldsymbol{T}^{T}_{MF\,1} \cdot {}^{LAB}\boldsymbol{T}_{MF\,2} $$

Inverse kinematics problem

Solve for (φ, ψ)

4 candidate solutions.

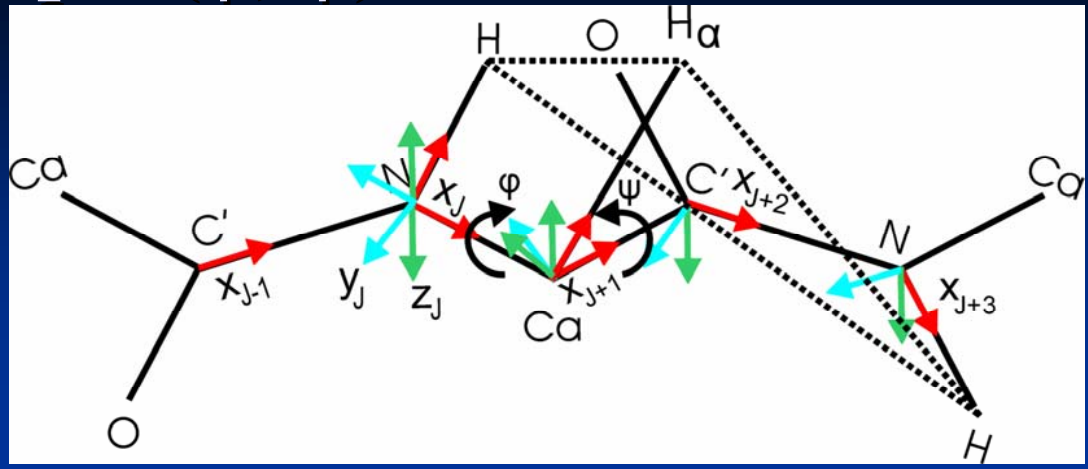
$$ (\varphi,\psi)_{CALC} = \{(\mathbf{-76.34^o,-38.08^o}),(-71.22,61.72^o),(26.07^o,-62.93^o),(37.52^o,31.29^o)\} $$

Check consistency with NOE.

Deviation from experimental values.

$$ (\varphi,\psi)_{OBS} = \{(\mathbf{-62.94^o,-43.73^o})\} $$

Unique (φ, ψ) from Exact NOE Distances



$$\left\| H_i^N - {}^J T_{J+1} H_i^\alpha \right\|^2 = dist^2(H_i^N, H_i^\alpha)$$

Solve for φ

$$\phi = \{-168.43^o, -62.89^o\}$$

$$\left\| H_i^\alpha - {}^{J+1} T_{J+3} H_{i+1}^N \right\|^2 = dist^2(H_i^\alpha, H_{i+1}^N)$$

Solve for ψ

$$\psi = \{-80.9^o, -43{,}08^o\}$$

$$\phi \times \psi = \{(-62.89^o, 43.08^o)\}$$

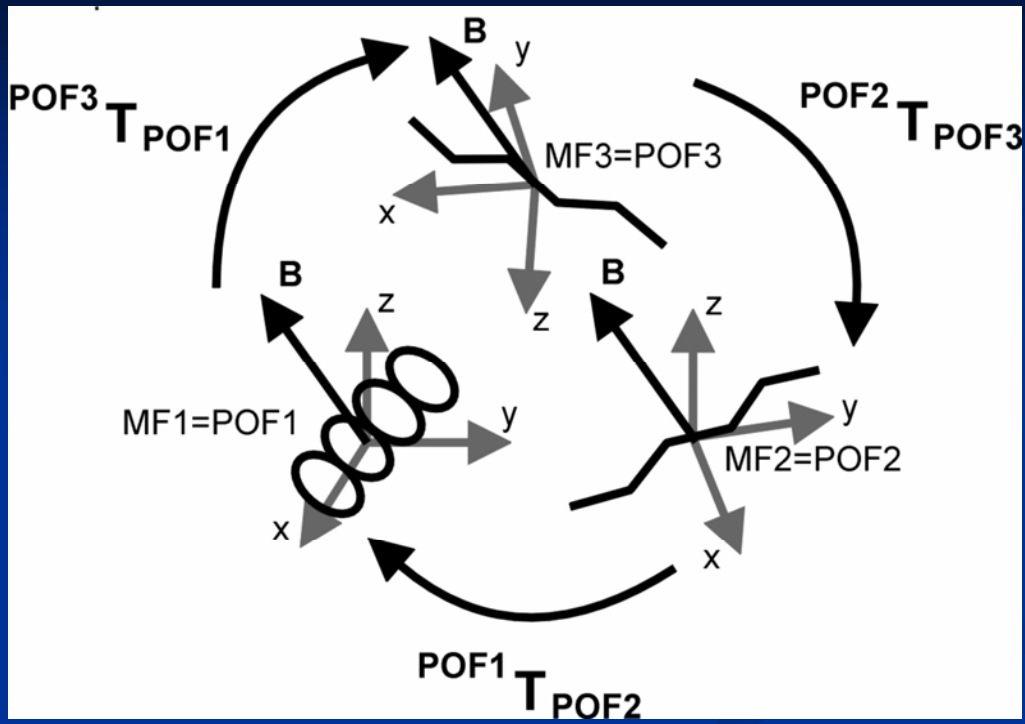
Consistent with NOE distance.

$$dist(H_i^N, H_{i+1}^N)$$

Protein Fold Determination

One exact solution using RDCs in 2 media

Relative Orientation of Secondary Structures



$$B_1 = {}^{POF\,1}T_{POF\,2} \cdot B_2$$

⇓

Magnetic vector coordinates transformation

⇓

Secondary structure orientation

Magnetic Vector Equation

Principle Alignment coordinate System defined - S diagonal

$$D = u^T S u$$

$$u = {}^{PAS}T_{POF} \cdot v$$

Transform from PAS to POF.

$$D = v^T ({}^{POF}T_{PAS} \cdot S \cdot {}^{PAS}T_{POF}) v$$

$$D = v^T (s) v$$

Relate POF Saupe with PAS Saupe.

$${}^{POF}T^T{}_{PAS} \cdot s \cdot {}^{POF}T_{PAS} = S$$

Specify S_{ZZ} from Da [Clore 98].

$$x^2 s_{xx} + y^2 s_{yy} + z^2 s_{zz} = S_{ZZ}$$

Magnetic unit vector B= (x,y,z) relative to Principle Order Frame

Magnetic Vector Determination from 2 Media

$$x^2 s_{xx} + y^2 s_{yy} + z^2 s_{zz} = S_{ZZ}$$ 1st medium

$$x'^2 s'_{xx} + y'^2 s'_{yy} + z'^2 s'_{zz} = S'_{ZZ}$$ 2nd medium

$$x^2 + y^2 + z^2 = 1$$

$$B' = {}^{POF'}T_{POF} \cdot B$$ Transform from 1st to 2nd medium.

At most 8 magnetic vector solutions

Relative orientation from magnetic vector transformation

Protein Fold from Magnetic Vector Transformation

$$B_{POF\,1} = {}^{POF\,1}T_{POF\,2} \cdot B_{POF\,2}$$

Inverse Kinematics

Parameterize T with Euler angles α, β, γ.

$$B_1 = R_z(\alpha)R_y(\beta)R_z(\gamma)B_2$$

$$\min\left\| d^{OBS} - A^{CALC} \cdot s_{CALC} \right\|$$

Filter orientations with min RDC error.

$${}^{POF\,1}T_{POF\,2} \qquad {}^{POF\,2}T_{POF\,3} \qquad {}^{POF\,3}T_{POF\,1}$$

$$\min\left\| {}^{POF\,2}T_{POF\,1} - {}^{POF\,2}T_{POF\,3} \cdot {}^{POF\,3}T_{POF\,1} \right\|$$

Consistency with orientations from a 3rd secondary structure.

One exact protein fold solution

One Exact Orientation Solution

$$
{}^{POF\,1}T_{POF\,2}{}^{CALC} = \begin{pmatrix} -.8324 & .5501 & .067 \\ -.5542 & -.8263 & -.1006 \\ 0 & -.1209 & .9927 \end{pmatrix}, \quad {}^{POF\,1}T_{POF\,2}{}^{OBS} = \begin{pmatrix} -.8605 & .5090 & -.0194 \\ -.5071 & -.8597 & -.0617 \\ -.0481 & -.0433 & .9979 \end{pmatrix}
$$

$$
{}^{POF\,2}T_{POF\,3}{}^{CALC} = \begin{pmatrix} .9999 & -.0037 & .0001 \\ -.0037 & -.9996 & .029 \\ 0 & -.029 & -.9996 \end{pmatrix}, \quad {}^{POF\,2}T_{POF\,3}{}^{OBS} = \begin{pmatrix} .9869 & .1512 & -.0558 \\ .1521 & -.9882 & .0121 \\ -.0533 & -.0205 & -.9984 \end{pmatrix}
$$

$$
{}^{POF\,3}T_{POF\,1}{}^{CALC} = \begin{pmatrix} -.8708 & -.4873 & -.0646 \\ -.4835 & .8728 & -.0662 \\ .0887 & -.0264 & -.9957 \end{pmatrix}, \quad {}^{POF\,3}T_{POF\,1}{}^{OBS} = \begin{pmatrix} -.7708 & -.6278 & -.1073 \\ -.6328 & .7742 & .015 \\ .0736 & .0795 & -.9941 \end{pmatrix}
$$

Very small deviation from experimental orientations

$$
{}^{1}T_{2}{}^{T} \cdot {}^{1}T_{2}{}^{OBS} = \begin{pmatrix} 0.997 & 0.052 & 0.050 \\ -0.048 & 0.995 & -0.080 \\ -0.054 & 0.077 & 0.995 \end{pmatrix},
$$

$$
{}^{2}T_{3}{}^{T} \cdot {}^{2}T_{3}{}^{OBS} = \begin{pmatrix} 0.986 & 0.154 & -0.055 \\ -0.154 & 0.987 & 0.016 \\ 0.057 & -0.008 & 0.998 \end{pmatrix},
$$

$$
{}^{3}T_{1}{}^{T} \cdot {}^{3}T_{1}{}^{OBS} = \begin{pmatrix} 0.983 & 0.179 & -0.001 \\ -0.178 & 0.979 & 0.091 \\ 0.018 & -0.089 & 0.995 \end{pmatrix},
$$

Conclusions I

A. Protein Structure Validation

- Discrete RDC substructure solutions are derived by RDC equation.
- The RDC structure solutions confirm the validity of NOE structure.

B. Protein Backbone Conformation

- Protein backbone is modeled as a robotic arm and an inverse kinematics problem is solved using RDCs.
- One solution (φ,ψ) is distinguished by at most 3 NOE constraints.

Conclusions II

C. Protein Fold

- A novel algebraic method solves protein fold efficiently.
- No exponential, stochastic or qualitative graphics methods are used [Fowler 00], [Al-Hashimi 00].
- One exact solution by solving low-degree polynomials.
- A general method for α-helices, β-strands and remote protein domains.
- No experimental NOE constraints are used [Fowler 00], [Wang 04].
- Only minimal RDC experimental measurements of (N-H) are used.
- Reinforce the importance of RDC experiment for structural genomics.

Bibliography

[1] I. Z. Emiris and S. I. Pantos: "Protein Structure Prediction Using Residual Dipolar Couplings" Proceedings Second International Conference Algebraic Biology, LNCS 4545, pp. 217-231, (2007).

[2] Cornilescu, G., Marquardt, J., Ottiger, M., Bax, A:. Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. *J. Am. Chem. Soc. 120, (1998), 6836-6837.*

[3] Fowler, C.A., Tian, F., Al-Hashimi, H.M., Prestegard, J.H.: *Rapid Determination of Protein Folds Using Residual Dipolar Couplings.* J. Mol. Biol., 304, (2000), 447-460.

[4] Al-Hashimi, H.M., Valafar, H., Terrell, M., Zartler, E.R., Eidsness, M.K., Prestegard, J.H.: Variation of Molecular Alignment as a Means of Resolving Orientational Ambiguities in Protein Structures from Dipolar Couplings. J. Magnetic Resonance, 143, (2000), 402-406.

[5] Losonczi, J., Andrec, M., Fischer, M., Prestegard, J.: Order Matrix Analysis of Residual Dipolar Couplings Using Singular Value Decomposition. *Journal of Magnetic Resonance 138, (1999) 334–342.*

[6] Tjandra, N., Bax, A.: Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium. Science, 278, (1997), 1111-1114.

[7] Clore, G.M., Gronenborn, A.M., Bax, A.: A Robust Method for Determining the Magnitude of the Fully Asymmetric Alignment Tensor of Oriented Macromolecules in the Absence of Structural Information. J. Magnetic Resonance, 133, (1998), 216-221.

[8] Wedemeyer, W., Rohlb, C., Scheraga, H.: Exact solutions for chemical bond orientations from residual dipolar Couplings. *Journal of Biomolecular NMR, 22 (2002) 137–151.*

[9] Wang, L., Donald, B.R.: *Exact solutions for internuclear vectors and backbone dihedral angles from NH residual dipolar couplings in two media, and their application in a systematic search algorithm for determining protein backbone structure.* J. Biomolecular NMR, 29, (2004), 223-242.

[10] Hus, J., Peti, W., Griesinger, C., BrÄuschweiler, R.: *Self-Consistency Analysis of Dipolar Couplings in Multiple Alignments of Ubiquitin.* J. Am. Chem. Soc., 125, (2003), 5596-5597.