NMR protein conformers described by network parameters: a preliminary study.

Lisa Beatrice Caruso [1]. Cesare Manetti [2] Alessandro Giuliani[3] Alfredo Colosimo[1]
Dept. of Anatomy. Histology. Forensic Medicine and Orthopaedics
Sapienza. University of Rome [1]
Dept. of Chemistry Sapienza. University of Rome[2]
Dept. of Environment and Primary Prevention - Ist. Sup. Sanita' - Roma[3]

Keywords: Protein NMR spectroscopy, Protein structural networks, Sperm Whale Myoglobin

Abstract: NMR spectroscopy is one of the techniques of choice for the determination of protein structures. Its use has a number of positive aspects, among which the possibility to observe the influence of the solvent on the molecular structure, as well as the local movement of small molecular domains.
However, due to the intrinsic flexibility of protein tertiary structures in solution, the NMR information does not lead to a single structure but to a set of conformers. Using the topological representation of such conformers we analyzed the corresponding network parameters, to enlight their association with some specific molecular feature. In this frame we showed that: i) the *node degree* parameter positively correlates with molecular 'compactness', ii) the *average shortest path length* parameter positively correlates with molecular flexibility, and iii) as expected, the two parameters are anticorrelated between each other.

1 INTRODUCTION

Both high resolution NMR spectroscopy and x-ray crystal diffraction are able to provide structural information on proteins at the atomic level. However, NMR has the unique ability to enlight the internal dynamics of the protein molecule in solution over a wide range of time scales. In particular, NMR applications include investigations of the dynamic features of molecular structures as well as of the kinetics of interaction between proteins and other molecules. Thus, by coupling structural and dynamics aspects, NMR spectroscopy may in principle provide a complete picture of proteins' behaviour [1].

Among the factors playing a role in the study of protein macromolecules by NMR, it is difficult to overstimate the importance of the Nuclear Overhauser Effects (NOEs) which, besides the solvent viscosity, depends upon structural (proton-proton distances, molecular size and shape) and dynamical (effective rotational correlation times and intramolecular motions) factors. [2]. Unlike x-ray diffraction, NMR spectroscopy does not provide a single molecular structure but a set of conformers which may, in principle, interconvert between each other and thus reflect the internal molecular dynamics. Figure 1 (left panel) shows the 12 different models submitted in the PDB Data Bank (ID:f1MY) for the sperm whale myoglobin (SW MbCO) structure. The same Figure 1(right panel) shows the 3D structure of the same molecule at high resolution (1.15 $\AA$) as obtained by X-ray crystal diffraction data. The message to be taken from the inspection of the two pictures, over and above the obvious similarity, is the extra amount of dynamic information included in the left panel: blurring of the overall picture is not due to experimental uncertainty of a single structure estimate, but to the partial overlapping of different conformers each representing - in principle - the molecular structure in a single time shot.

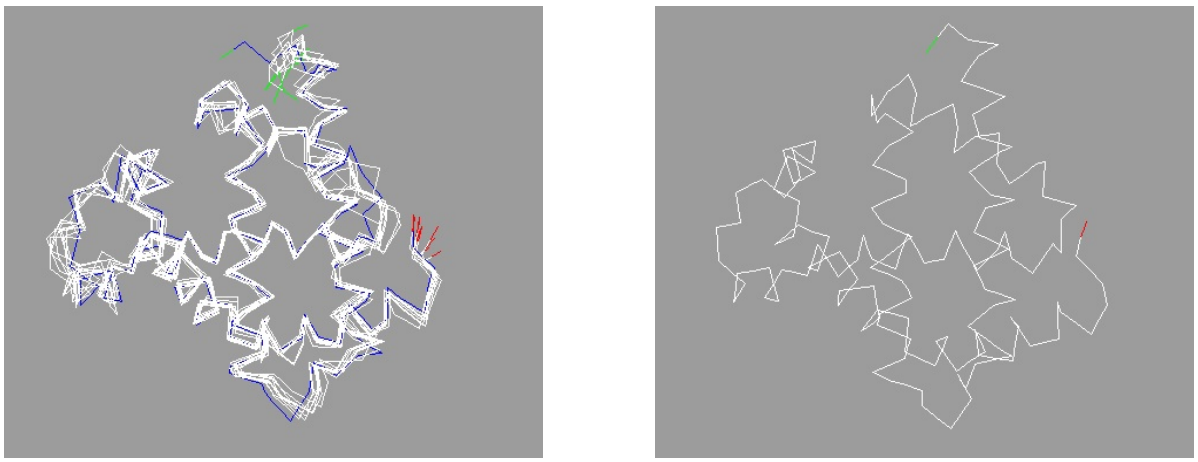

Figure 1: ***3****d structures of SpermWhale Mb(CO) backbone.*
LEFT: Conformers from NMR data. Among the 12 SW MbCO conformers (PDB ID:1MYF, NMR data) depicted in the figure, the blue one refers to the conformer endowed with the highest node degree value (# 6 in Table 1). Red and green colors indicate, respectively, the carboxy- and the amino- terminal. RIGHT: Static structure at 1.15 $\AA$ resolution (PDB ID:1A6G, X-rays diffraction data)

It would be useful to define a time-related ordering of different conformers, considered as frames of a single dynamical interconversion. To that purpose a network-inspired approach may be of help since it allows studying each conformer's network parameters and highlighting even small topological differences between conformers.

To test the approach we took advantage of a well known and relatively simple protein: the SW MbCO, whose structure in solution as determined by NMR and X-ray techniques is shown in the left and right panel of Figure 1, respectively. The structural data sets were obtained from the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do), and the number of NMR conformers submitted for this structure is 12 [3].

2 Materials and methods

Displaying 3D structures from PDB files can be obtained by several graphics programs. Among these PDB Viewer allows, in the case of proteins obtained from NMR spectroscopy, to single out each single conformer and study it in isolation and/or in connection with the others.

In the case of Sperm Whale Mb(CO) NMR data (PDB ID:1MYF), we set up a topological network for each of the 12 conformers using RING (Residue Interaction Network Generator) [4], a software tool useful to draw the network of chemical interactions (links) among amino acid residues. From the 3D coordinates in the PDB file a link between a couple of residues is drawn if the distance between any pair of atoms each pertaining to one member of the couple is lower than a predefined threshold. In our work we fixed the threshold to 5 $\AA$ since we point to the analysis of 2D NMR contact maps which essentially reflect covalent bonds, more than weaker and longer intramolecular interactions [5]. Once the whole grid of links is calculated, the network can be characterized through the evaluation of quantitative parameters and drawn using Cytoscape, an open source software platform visualizing molecular interaction and biological pathways [6–8]. The network parameters considered here are the *node degree* (ND) and the *average shortest path length* (ASPL). For details, see Appendix A.

3 Results

For each of the 12 conformers shown in Figure 1 a topological representation (network) can be obtained, characterized by a specific number of nodes and edges as well as by the average values of ND and ASPL (Table 1). It is noticeable that for all but one conformer (# 5) the associated number of nodes is lower than the number of residues in the protein (153). The obvious explanation is that the initial and terminal residues in the polypeptide chain are not recognized as network nodes since, due to their flexibility, their distance from their first neighbors may exceed 5 $\AA$.

More interesting is that the variability of both topological parameters, expressed by Percent Variation Coefficient (PVC)[1], namely 2.6 % and 1.8 %, is significantly lower than the corresponding value for the geometrical parameter (RMSD), namely 8.4 %. This could be taken as an indication of the higher sensitivity of network topological parameters to the protein dynamics, as reflected by the conformers emerging from NMR spectroscopy. In the panels of Figure 2 the numerical profiles of the two network parameters at hand are shown along the SW Mb(CO) sequence. The top panel refers to the numerical ND profile along the sequence, averaged over the 12 conformers for each residue. The maximal variability of this parameter is mostly localized at the protein-solvent interface and in the Xenon cavities (Figure 4), which are considered important for a possible role in allosteric phenomena [10, 11]. The bottom panel of Figure 2 shows the profile of the ASPL: also in this case high variance values correspond to residues close to the Xenon cavities (Figure 4).

An interesting observation emerging from figure 2 is the striking negative correlation between the two parameters along the sequence (Pearson correlation coefficient = -0.62). This could be taken as an effect of the essentially different structural information carried out by the two parameters, linked to the molecular compactness and flexibility, in the case of average ND and of ASPL, respectively.

Figure 3 provides more details in the above context, with reference to the connection between topological and geometrical features. The red and blue ND profiles depicted in the figure correspond to conformers #6 and #11, showing the maximal difference in terms of average ND. The clear association of the secondary structure elements in the bottom part of the figure with the increasing/decreasing trend of the profiles in the upper part, points to a phasic relation between α-elices and ND values. As a matter of fact, #6 and #11 conformers in Table 1 also show almost the smallest difference in ASPL, indicating the opposite physical interpretation to be assigned to such parameters, as possible indexes of molecular compacteness (ND) and molecular flexibility (ASPL).

As for ASPL, in Figure 4 the location the residues associated by our analysis to the highest values of ASPL variability are indicated (in blue) on the protein backbone, together with residues (in red) proposed by Del Sol and collaborators as responsible for the putative allosteric properties of SW Mb [12]. Only in the case of Leu 69 there is agreement between the two sets of data, which seems to point to a substantial difference between the underlying dynamical phenomena, occurring at quite different time scales.

[1] PVC = (Std.Dev./Mean)*100, as computed over the entire set of conformers.

Conformer	*# Nodes*	*# Edges*	*ASPL*	*Average ND*
1	149	800	3.39	10.37
2	151	832	3.34	10.80
3	151	808	3.46	10.73
4	141	808	3.46	10.52
5	153	822	3.47	10.75
6	152	807	3.43	11.05
7	149	789	3.42	10.31
8	152	821	3.40	10.58
9	150	804	3.42	10.39
10	151	799	3.44	10.41
11	151	797	3.48	10.09
12	149	798	3.28	10.27

Table 1: ***N****etwork parameters of the 12 NMR conformers.*
Columns 4 and 5 contain, respectively, the values of Node Degree (ND) and Average Shortest Path Length (ASPL) averaged through all nodes of the corresponding conformer. The parameters' values averaged over the whole set of conformers are 10.52 ± 0.27 (2.6%) and 3.42 ± 0.06 (1.8%) for Average Node Degree and Average Shortest Path Length, respectively. The corresponding value for the 'geometrical' parameter, Root Mean Squared Deviation, calculated for each possible couple of conformers, is 0.92 ± 0.08 (8.4 %).

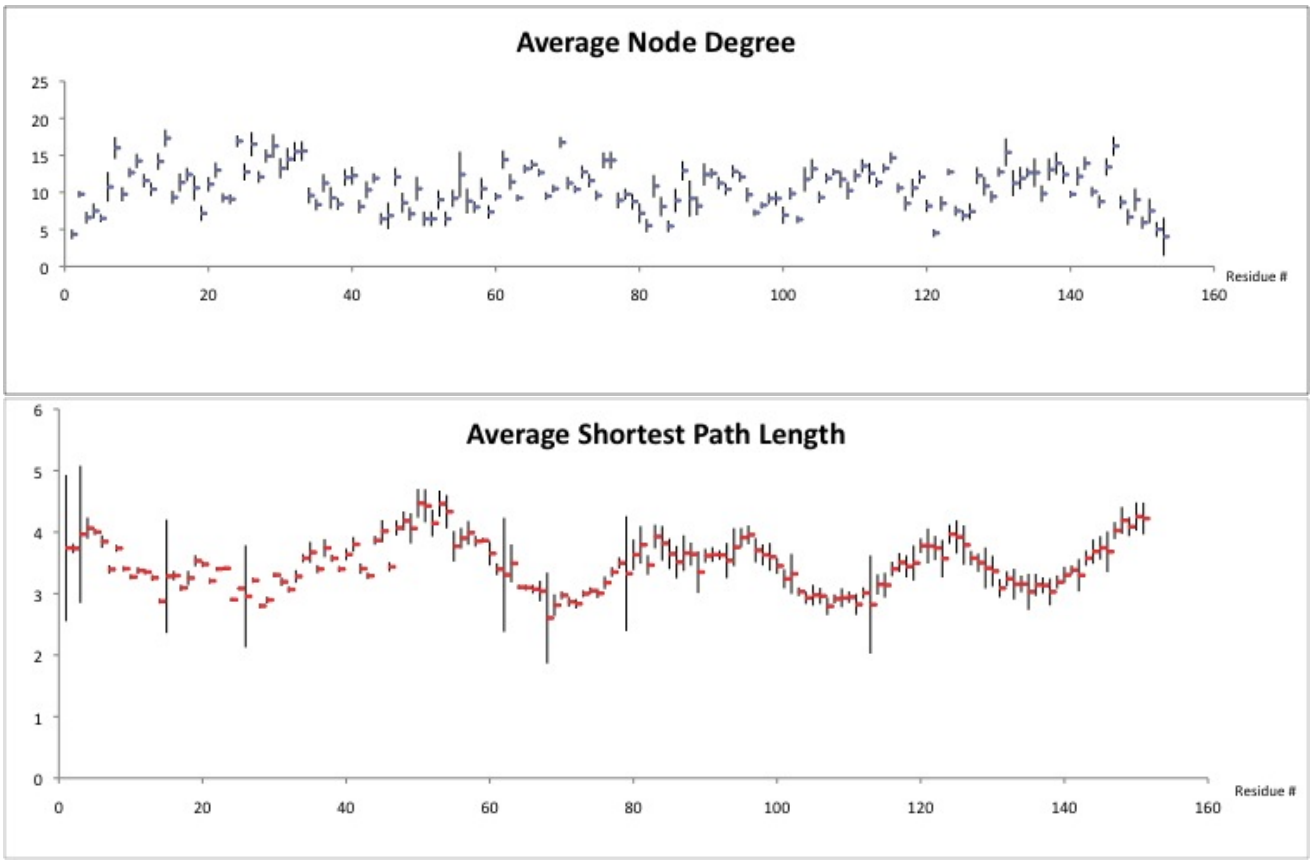

Figure 2: ***D****istribution of the topological parameters at hand (average values) along the SW Mb(CO) primary structure.*
In both the upper (Average Node Degree) and bottom (Shortest Path Length) panel the parameter value averaged over the 12 conformers and the related standard deviation, is indicated for each node, that is for each residue along the primary structure.

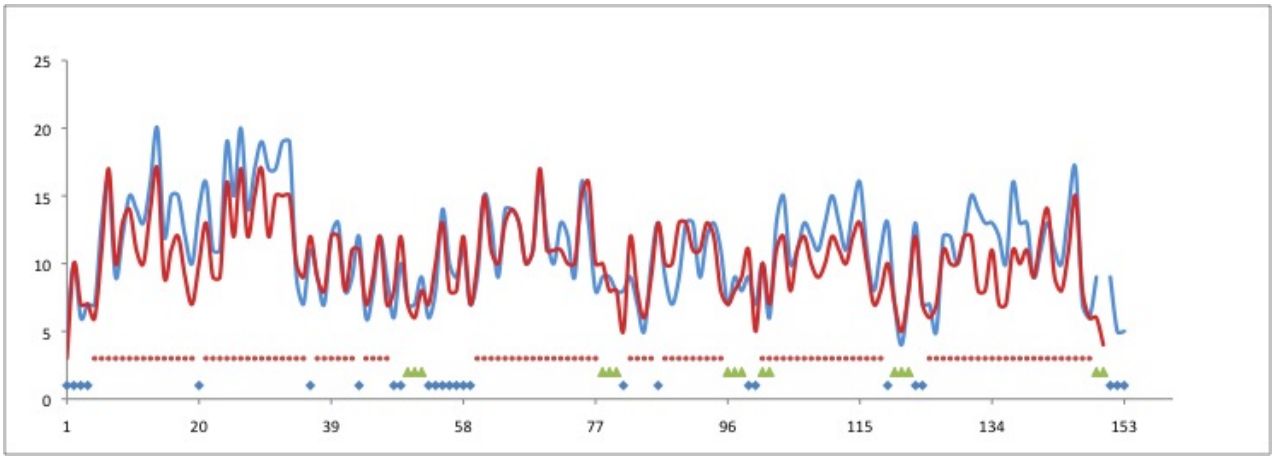

Figure 3: ***D**ifferences in the Node Degree profile along the primary structure between two SW MbCO conformers.*
The blue and red colors refer, respectively, to conformer 6 and 11 in Table 1. In the bottom part of the graph red, green and blue dots indicate the location of the secondary structure elements, i.e. α-elix, β-sheet and Random, respectively, along the primary structure.

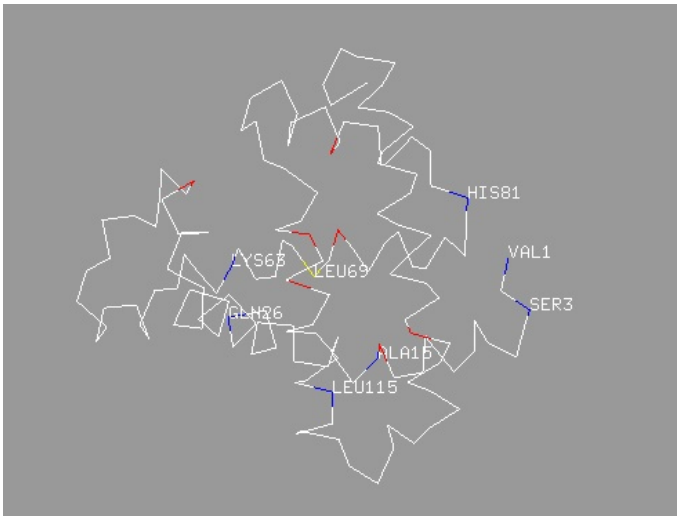

Figure 4: ***M**olecular flexibity in SW Mb(CO)*
On the background of the 3D X-rays protein backbone (ID PDB:1Myf), (Leu 69 (marked in yellow) appears endowed with maximal flexibility according to both this paper and to [12] (see the text for details); blue and red colors, however, are residues for which, respectively, maximal flexibility (this paper, Fig. 2, bottom panel) and possible involvement in allosteric conformational changes [12] have been proposed.

4 Discussion/Conclusions

1. Linking the hyperfine structural details provided by 2D NMR spectroscopy and the euristic power of topological representation of protein structures is made possible since: i) several structural conformers are often equally compatible with the same spectroscopic data set, and ii) each conformer can be associated to a specific topological network. Taking advantage of that, we look for correlations between topological and structural differences among NMR conformers to be possibly associated to a time-dependent sequential order.

2. In order to test the above considerations on the background of a well established set of structural data [3], our attention focused on the specific case of Sperm Whale MbCO. Besides confirming the opportunity provided by the network analysis to get complementary and independent structural information, we tested the role that ”Node Degree” and ”Average Shortest Path Length” network parameters may play in the design of reliable

structure-function mechanistic relationships of general application.

3. In particular, the ND indicates the more compact SW Mb(CO) conformers and, within each conformer, the more flexible stretches along the polypeptide chain. Understanding a specific geometrical counterpart of ASPL is not equally easy. A possible clue is that a low variability of such a parameter for given residues among different conformers could indicate a role of those residues in functional, and even allosteric mechanisms.

4. In the perspective of an innovative analysis of 2D NMR data, however, ASPL values appeared endowed with information useful to design networks compatible with a given set of NMR distance and chemical shift constraints.

5 Appendix A

The *node degree* (ND) or *connectivity* (k) is the most elementary characteristic of a node, namely the number of links connecting that node to any other node, and its average is referred to all nodes in the network. In direct graphs, the in-degree of a node is the number of incoming edges and the out-degree is the number of outgoing edges [9]. We only considered undirected graphs, so the ND simply indicates the number of links in which a node is involved.

The shortest path length between node i and node j ($d_{i,j}$) is the minimum number of edges that must be traversed to go from node i to node j. Thus, the average of $d_{i,j}$ over the whole network, *average shortest path length* (ASPL), is defined as:

$$ASPL = \frac{\sum_{i,j \in V} d_{i,j}}{N(N-1)} \tag{1}$$

where V is the set of nodes in the network (G), and N is the number of nodes in G. The ASPL gives the expected distance between connected nodes. This parameter can be computed on a node-by-node basis or averaged over the entire network so to get a general inverse measure of among nodes connection efficiency.

REFERENCES

[1] McGreevy K.S. Bertini I. and Parigi G. *NMR of Biomolecules*. Wiley-Blackwell, 2012.

[2] Wüthrich K. Nmr studies of structure and function of biological macromolecules (nobel lecture). *Angew.Chem.Int.Ed.42, 33403363*, 2003.

[3] Wright PE Osapay K, Theriault Y and Case DA. Solution structure of carbonmonoxy myoglobin determined from nuclear magnetic resonance distance and chemical shift constraints. *J Mol Biol.*, 244:183–97, 1994.

[4] Martin A.J.M., Vidotto M., Boscariol F., Di Domenico T., Walsh I., and Tosatto S.C.E.. Ring: Networking interacting residues, evolutionary information and energetics in protein structures. *Bioinformatics,*, 2011.

[5] Oliva G., Di Paola L., Pascucci F. Giuliani A., and Setola R. Resilience of protein contact networks. *Electronic Letters*, In Press, 2013.

[6] Smoot M., Keiichiro O., Ruscheinski J., Wang P.L., and Ideker T.. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics,*, 27(3):431–432, 2011.

[7] Cline M.S., Smoot M., Cerami E., Kuchinsky A., Landys N., Workman C., Christmas R., Avila-Campilo I., Creech M., and Gross B. et al. Integration of biological networks and gene expression data using cytoscape. *Nature Protocols*, 2:2366–2382, September 2007.

[8] Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B., and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–504, November 2003.

[9] Barabási A.L. and Oltvai Z.N. Network biology: understanding the cell's functional organization. . *Nat Rev Genet*, 5:101–113, 2004.

[10] Bellelli A. and Brunori M. Hemoglobin allostery: Variations on the theme. *Biochemica et Biophysica Acta*, 1807:1262–1272, 2011.

[11] Eaton W.A., Henry E.R., Hofrichter J., Bettati S., and Viappiani C.and Mozzarelli A. Evolution of allosteric models for hemoglobin: a critical review. *UBMB Life*, 59(8 9):586 599, August September 2007.

[12] del Sol A., Fujihashi H., Amoros D., and Nussinov R.. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Molecular Systems Biology*, page 112, 2006.

[13] De Ruvo M., Giuliani A., Santori D., and Di Paola L. Shedding light on protein-ligand binding by graph theory: the topological nature of allostery. *Biophysical Chemistry*, 2012.