Are Proteins composed of little structural bricks?

(And can we play with them?)

Alexandre G. de Brevern[1,2,3,4]
(alexandre.debrevern@univ-paris-diderot.fr, http://www.dsimb.inserm.fr/~debrevern/)
[1]INSERM UMR-S 1134, DSIMB, [2]Univ Paris, [3]Institut National de Transfusion Sanguine, [4]GR-Ex laboratoire d'excellence
6, rue Alexandre Cabanel, 75739 Paris cedex 15, France.

The protein structures are classically described as composed of two regular states, the α-helices and the β-strands and one non-regular and variable state, the coil. Nonetheless, the representation of only two repetitive states hides other interesting repetitive structures, e.g., PolyProline II helix [1], or super secondary motifs. The definition of secondary structures is often considered as fixed and ideal. In fact, the rules for secondary structure assignments are complex and can also bias our analyses [2]. It is so interesting to look at other complementary description such as sets of small prototypes or "structural alphabets", able to analyze local protein structures and to approximate every part of the protein backbone. The principle of a structural alphabet is simple (see Figure 1a). A set of average local protein structures is firstly designed. They approximate (efficiently) every part of the structures. As one residue is associated to one of these prototypes, we can translate the 3D information of the protein structures as a series of prototypes (letters) in 1D, as the amino acid sequence.

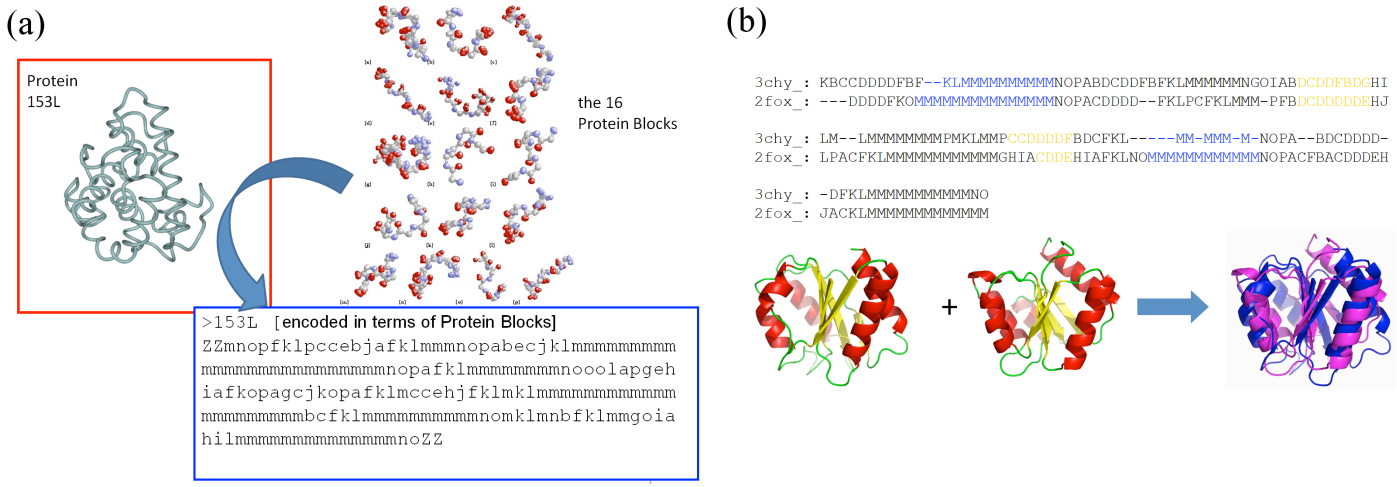

Figure 1: (a) Principle of the coding of protein structures (3D) into a series of letters (1D); (b) use of the Protein Blocks to superimpose protein structures.

Structural alphabets have also been used to predict the protein backbone conformation and in *ab initio* / *de novo* methods. Our structural alphabet is composed of 16 mean protein fragments of 5 residues in length, called Protein Blocks (PBs) [3]. They have been used both to describe the 3D protein backbones and to perform a local structure prediction. PBs have been cited in more than 350 publications worldwide from the prediction of long fragments, to definition of binding site [4]. We have used this approach to compare / superimpose protein structures (see Figure 1b). The assessment of a simple approach done on the classical benchmark sets was surprisingly excellent. It is equivalent or better than the best actual approaches [5] and still is. Moreover, PBs can be used to assess protein flexibility with efficiency [6].

[1] Mansiaux Y., Joseph A.P., Gelly J.-C., de Brevern A.G. (2011) Assignment of PolyProline II conformation and analysis of sequence - structure relationship *Plos One* 6(3): e18401.

[2] Tyagi M., Bornot A., Offmann B., de Brevern A.G. (2009) Analysis of loop boundaries using different local structure assignment methods, *Protein Science* 18(9):1869-81.

[3] de Brevern A.G., Etchebest C. & Hazout S. (2000) Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks, *Proteins*, 41:271-87.

[4] Joseph A.P., Agarwal G., Mahajan S., Gelly J.-C., Swapna L.S., Offmann B., Cadet F., Bornot A., Tyagi M., Valadié H., Schneider B., Etchebest C., Srinivasan N., de Brevern A.G. (2010) A short survey on Protein Blocks, *Biophysical Reviews* 2(3):137-145.

[5] Joseph A.P., Srinivasan N., de Brevern A.G. (2012) Progressive structure-based alignment of homologous proteins: Adopting sequence comparison strategies, *Biochimie* 94:2025-34.

[6] Narwani T.J., Craveur P., Shinada N.K., Floch A., Santuz H., Vattekatte A.M., Srinivasan N., Rebehmed J., Gelly J.-C., Etchebest C., de Brevern A.G. (2019) Discrete analyses of protein dynamics, *JBSD* (2019) in press.