Sense-Antisense Homology Boxes in Proteins: Structural Motifs Enconded in the DNA?

Carlos A. Del Carpio [1]
carlos@translell.tutkie.tut.ac.jp

Valentin Gogonea [2]
gogonea@organik.uni-erlangen.de

Katsuhisa Yamaguchi [1]
yamaguti@translell.tutkie.tut.ac.jp

Makoto Takei [1]
takei@translell.tutkie.tut.ac.jp

Kouji Nishimura [1]
mame@translell.tutkie.tut.ac.jp

[1] Lab. for Informatics & AI in Molecular and Biological Sciences
Department of Ecological Engineering,
Toyohashi University of Technology. Tempaku. Toyohashi 441, Japan
[2] Computer Chemistry Center, Institute for Organic Chemistry
Erlangen-Nurnberg University, Nagelsbachstr. 25, D-91052 Erlangen, Germany

Abstract

Experimental evidence implying that complementary DNA strands encode amino acids which exhibit complementary hydrophobic characteristics has led us to the inspection of sense-antisense homology in several hundreds of proteins recorded in the PDB. We present here partial results of this analysis which relate localized peculiar structural characteristics of proteins to the sense-antisense homology boxes found in the primary sequences. A further analysis is performed in order to determine whether these sense-antisense homology boxes,if existent within the protein, are encoded by unique sequences of codons in the DNA. We also make here a progress report about the methodology and the results obtained so far.

1 Sense-Antisense Homology Boxes in Proteins

It is well known that local interactions between amino acids are the primary determinant of the three dimensional structure of a protein. Relative contributions of a given amino acid to the local hydrophobic character of a peptide sequence have been estimated by Kyte and Doolittle [1] who, combining empirical and theoretical approaches, computed values reflecting tendencies of interaction among natural amino acids. Based on these values, it was observed that complementary DNA strands encode amino acids which exhibit complementary hydrophobic characteristics.
Blalock and Smith [2], demonstrated that codons read in the 5' to 3' direction specify amino acids whose hydropathic properties tend to be opposite in sign to those coded in the 3' to 5'direction. Furthermore, based on these evidences, Markus et. al [3], suggested an interaction model for these type of sense and antisense peptides, proposing a new type of secondary structure for proteins. Moreover, Campbell et. al [4], and Baranyi et.al.[5] synthesized many antisense peptides and measured experimentally the degree of binding and interaction with their sense counterparts. In the present study,

we have written a program to identify all the subsequences bearing sense and antisense amino acids within a protein sequence and have analyzed the 3D structural characteristics based on the secondary structure types defined by Kabsch et. al.[6].
Motifs consisting of two arbitrary fragments of a determined number of amino acid residues are searched for within the protein and classified according to the distances among the α Carbons of corresponding amino acid residues. A classification of these pairs of substructures is performed by computation of the codebook vectors for classes of these double stranded motifs. This is performed by a classification algorithm based on a sort of artificial neural network called the Self Organizing Map (SOM) and firstly suggested by T. Kohonen [8].

This analysis was performed for many protein structures recorded in the Brookhaven Protein Data Bank (PDB), and we present some generalizations to which we arrived. The goal of the present study is the incorporation of this type of knowledege into our automatic system for protein 3D structure prediction[7].

The classes of double stranded motifs reflect motifs whose strands are close to each other within the protein and others where the distances among the corresponding amino acids are large. Analysis of the classes of double stranded motifs and their sense and antisense amino acid content alludes a correlation between the structural characteristics of the peptide and the content of the sense - antisense amino acids. A typical double stranded motif of this kind can be described as two extended structural elements separated by a reverse turn. This was also found by Campbell [4]. and others [5] in experimental studies. A further analysis has been performed on many other proteins recorded in protein data banks. Furthermore, we have performed an analysis of the frequency of codon usage to determine whether this type of structural characteristics of the protein are already encoded at the DNA level. The program is based on the new genetic programming paradigm, and we have carried out the analysis for the E. Coli Genome. Firsthand results of the system reveal patterns of codon usage heavily dependent on the neighboring amino acids (at the amino acid sequence level). This leads to postulate that sense-antisense homology can be a means of encoding not only structural characteristic but protein function at the DNA level.

References

[1] Kyte,J., Doolittle, R.F., J. Mol. Biol., 157,105-132(1982)

[2] Blalock,J.E., Smith, E.M., Biochem. Biophys. Res. Commun, 121,203-307(1984)

[3] Markus, G., Tritsch, G.L., Parthasarathy, R., Arch. of Biochem. and Biophys.,272-2,433-439(1989)

[4] Campbell, W., Okada,H., Biochem. Biophys. Res. Commun, 175-1,207-214(1991)

[5] Baranyi, L., Campbell, W., Ohshima, K., Fujimoto, S., Boros, M., Okada H.; Nature Medicine,

[6] 1-9, 894-901(1995) Kabsch, W., Sander, C., Biopolymers, 22, 2577-2637(1983)

[7] Del Carpio C., J. Chem. Inf. and Comput. Sci., 36, 258-269(1996)

[8] Kohonen T.,Proceedings of the IEEE, 78-9 (1990) .