Computational analysis of structural properties of classical and novel non ribosomal aminoacyladenylate forming domains.

Leonardo Di Vincenzo [a], Ingeborg Grgurina [a] and Stefano Pascarella [a,b]

[a]*Dipartimento di Scienze Biochimiche "A. Rossi Fanelli" Università "La Sapienza", 00185 Roma, Italy*

[b]*Centro Interdipartimentale di Ricerca per la Analisi dei Modelli e dell'Informazione nei Sistemi Biomedici (CISB), Università "La Sapienza", 00185 Roma, Italy.*

Keywords. nonribosomal peptide synthetases; homology modelling; computational analysis

Nonribosomal peptide synthetases (NRPSs) are multidomain, multifunctional enzymes involved in the biosynthesis of many bioactive microbial peptides such as phytotoxins, siderophores, biosurfactants, and anticancer agents. The minimal module required for a single monomer addition consists of a condensation domain (C), an adenylation domain (A) and a peptidyl carrier protein (PCP) domain also denoted as thiolation (T) domain [1]. Systematic comparative analyses identified 8 or 10 sequence positions lining the active site pocket which are held responsible for substrate recognition and selection in A domain [2]. Recently, it has been pointed out [3] that several enzymes possibly involved in lysine metabolism in eucaryotes display a 3-domain architecture where the two *N*-terminal domains are homologous to the A and T domains from NRPS systems. The third *C*-terminal section may contain a PQQ, a NADPH or a functionally uncharacterized domain.

Our work is aimed at the structural characterization and the study of common molecular features of the family of the aminoacyladenylate-forming enzymes from NRPS and from the recently discovered homologous enzymes.

Psi-BLAST [4] searches were applied over the GeAll and Non-Redundant databanks using query sequences Ebony (gi:3286766) from *Drosophila melanogaster*, 5-aminoadipic acid synthase (gi:30348962) from *Mus musculus* and aminoadipate-semialdehyde dehydrogenase from yeast (swissprot:LYS2_YEAST). Thirty-two sequences were identified from different eucaryotic species and the domain assignments were confirmed by CDD [4] and Pfam [5] queries. The sequence subsets containing the A-T domains were aligned utilizing the HMMER package[6].

On the basis of the structural homology encoded in this multiple alignment, the potential occurrence of a "specificity code" similar to that described for the NRPS systems has been tested. The residues which interact with the α-amino and α-carboxy groups of the amino acid substrates [2], Asp235 and Lys517 respectively, are conserved, the only exceptions being Ebony protein (gi:3286766) from *Drosophila melanogaster* and (gi:21291643) from *Anopheles gambiae* where the Asp235 is replaced by valine.

Homology molecular modelling has been utilized to map the conserved residues onto a hypothetical active site structure of the 5-aminoadipic acid synthase from *Homo sapiens* (gi:32261239) and Ebony (gi:3286766) from *Drosophila melanogaster* to understand the role of the conserved residues and to predict their interaction with the putative substrates. In case of Ebony proteins, the Asp235 is replaced by Val, while Pro236, conserved in all 5-aminoadipic acid synthase and aminoadipate-semialdehyde dehydrogenase, is substituted by Asp which can form hydrogen bond with the β-amino group of the β-alanine substrate. The β-amino group interacts via hydrogen bonds also with Ser301 and Asp331 (Fig. 1). The other residues line and shape the active site pocket.

Characterization of the α-aminoadipate synthase is under way.

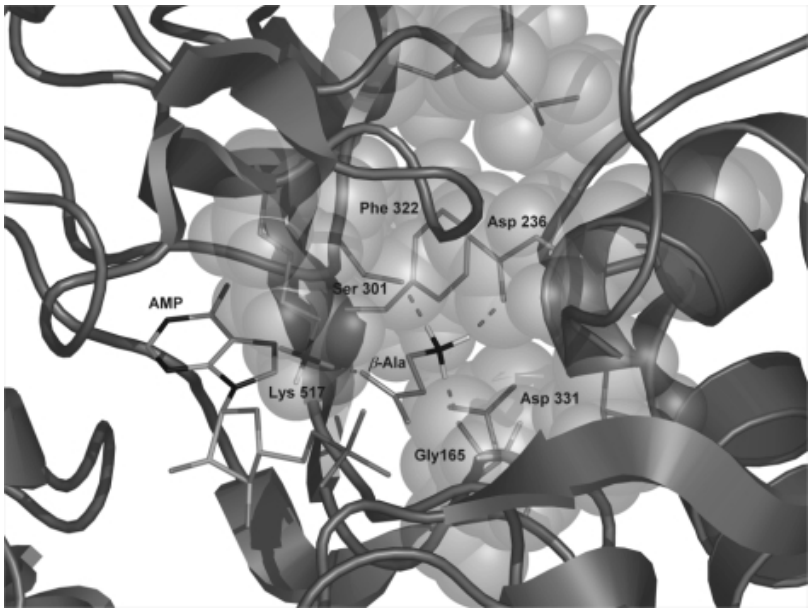

Fig. 1 Model structure of the active site of *Ebony* from *Drosophila melanogaster Ebony* is represented in ribbons; AMP is shown as stick, β-alanine is represented as stick and the specifity code residues are shown as sticks and CPK models.

Acknowledgements

This research was supported in part by a grant from the italian Ministero dell'Istruzione, Università e Ricerca (MIUR) e from the University of Roma "La Sapienza". Part of this work will be submitted by LDV in partial fulfillment of the requirements of the degree of Dottorato di Ricerca at Università di Roma "La Sapienza". Authors are grateful to Prof. Francesco Bossa for his encouraging support and advice.

References

[1] M. A. Marahiel, T. Stachelhaus and H. D. Mootz, Modular Peptide Synthetases Involved in Noribosomal Peptide Synthesis, *Chem. Rev.,* 97:2651-2673, 1997.
[2] T. Stachelhaus, H. D. Mootz and M. A. Marahiel, The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.*, 6:493-505, 1999.
[2] T. Kasahara & T. Kato, A new redox-cofactor vitamin for mammals, *Nature*, 422: 832, 2003.
[3] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller & D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res*., 25:3389-3402, 1997.
[4] A. Marchler-Bauer, J. B. Anderson, C. DeWeese-Scott, N. D. Fedorova, L.Y. Geer, S. He, D. I. Hurwitz, J. D. Jackson, A. R. Jacobs, C.J. Lanczycki, C. A. Liebert, C. Liu, T. Madej, G. H. Marchler, R. Mazumder, A. N. Nikolskaya, A. R. Panchenko, B. S. Rao, B. A. Shoemaker, V. Simonyan, J. S. Song, P. A. Thiessen, S. Vasudevan, Y. Wang, R. A. Yamashita, J. J. Yin, S. H. Bryant, CDD: a curated Entrez database of conserved domain alignments, *Nucleic Acids Res.,* 31:383-387, 2003.
[5] A. Bateman, L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, D. J. Studholme, C. Yeats, S. R. Eddy, The Pfam protein families database. *Nucleic Acids Res.,* 1:32, 2004.
[6] S. R. Eddy, Profile hidden Markov models. *Bioinformatics*, 14:755-63, 1998.