Novel Classes of Eukaryotic Aspartic Proteases and the Identification of their Specificity Determining Residues

María Victoria Revuelta & Arjen ten Have. IIB-CONICET-UNMdP - mrevuelta@mdp.edu.ar

Eukaryotic Aspartic proteases (APs, A01, EC 3.4.23) constitute one of the seven families of proteolytic enzymes that are distinguished. They are classified into the A01A and A01B subfamilies (Rawlings et al., 2010). APs are found in animals, plants, fungi and many other eukaryotic lineages. A closely related subfamily with particular characteristics (retropepsin, A2) found in viruses is also classified as EC3.4.23. APs have recently been identified in seven bacterial genomes as well (Rawlings & Bateman, 2009). APs are made of two lobes that, although structurally similar, do not show significant sequence similarity. A2 APs are made of two identical polypeptides, each polypeptide corresponding to one lobe. The active site of APs is made of two catalytic aspartates (D32 and D215, all numbering according to pig pepsin sequence), associated in a D[TS]G motif that forms the active site in a binding cleft that is enclosed by two β-barrel domains (Davies, 1990). Another distinctive hallmark is the strictly conserved Y75, which is located in a loop called "flap region". D32 and D215 form a conserved loop (psi-loop) with two XaaXaaG, Xaa being any hydrophobic residue (Blundell et al., 1998).

Generally, APs have a preference for hydrophobic substrates due to the presence of the XaaXaaG in the psi-loop. Certain APs accept a wide range of substrates, while others have a high substrate specificity. Yapsins (Yeast Glycosylphosphatidylinositol anchored APs) belong to the latter kind mentioned (Monod et al., 2002). APs are synthesized as inactive zymogens. They either have a propeptide that directly interacts with the binding cleft, blocking the access to the catalytic site, or they have a propeptide that alters the tridimensional structure of the enzyme, avoiding the approach of the two D[TS]G necessary for the catalysis (Dunn, 2002). These propeptides are removed in the process of activation. APs have been shown to function in: 1) Haemoglobin cleavage by the malaria causing parasite Plasmodium falciparum (plasmepsins I-IV, Francis et al., 1997); 2) Blood pressure regulation (renin, Sielecki et al., 1989); 3) Protein digestion in the digestive system in mammals (pepsin and gatricsin, Beck, 1973); 4) Site-specific proteolysis in viral replication cycle (HIV protease, Wlodawer & Erickson, 1993); 5) Effector proteins targeting of P. falciparum (plasmepsin V, Russo et al., 2010); 6) Fungal host infection (Monod et al., 2002); 7) Fungal sexual reproduction (Alby et al., 2009); and 8) Fungal cell wall integrity (Gagnon-Arsenault et al., 2006). Recent studies have indicated that fungi have a relatively large number of APs (Amselem et al., 2011). All together, this indicates they have acquired a set of diverse paralogues. Therefore, we study fungal APs and their functional diversification and redundancy.

We performed a detailed phylogenetic analysis of AP sequences obtained by HMMer profiling of 107 completely sequenced eukaryotic genomes using profiles that were seeded with MEROPS A01A and A01B sequences as well as many sequences for which X-ray structures are available from the PDB database (http://pdb.org/pdb/home/home.do). The obtained phylogeny contains eight major subfamilies, six of which are exclusive to
fungal APs. The two other clades correspond to the A01A and A01B MEROPS subfamilies.

Phylogenies that comprise both ortologues and paralogues are difficult to interpret. The topology is the result of both genetic drift (taxonomic distance) and shift. The topology of the obtained tree has ample indication for the presence of genetic shift, hence suggests functional diversification. Sequences that share a functional aspect such as particular substrate specificity or the same cellular compartment will most likely have common features. We examined the protease alignments of the AP proteases to identify amino acid motifs that are conserved within and across clades, with particular emphasis on motifs that may relate proteases to their subfamily, and provide insights into protein function. For the purpose of identifying specificity determining positions, a number of available software such as SDPfox (Mazin et al., 2010) were used.

Interestingly, we found a pair of acid residues exclusive for the Yapsin subfamily that are likely to be involved in monobasic sequence recognition and substrate processing. These residues, D77 and E129 are located at the tip of two flap regions that cover the binding cleft. X-ray structures of the yapsin SAP1 from Candida albicans made in the presence or absence of the inhibitor pepstatin suggest these two residues are involved in substrate recognition.