Crystal structure of the soluble domain of RC1339/APRc from *Rickettsia conorii,* a retropepsin-like aspartic protease

Mi Li[1,2], Alla Gustchina[1], Rui Cruz[3,4], Marisa Simões[3,4], Pedro Curto[3,5], Juan Martinez[5], Carlos Faro[3,4], Isaura Simões[3,4] and Alexander Wlodawer[1]

[1]Protein Structure Section, Macromolecular Crystallography Laboratory, National Cancer Institute, Frederick, MD 21702, USA; [2]Basic Science Program, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA; [3]CNC-Center for Neuroscience and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal; [4]Biocant, Biotechnology Innovation Center, 3060-197 Cantanhede, Portugal; [5]Vector-Borne Diseases Laboratories, Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Baton Rouge, LA 70803, USA

The common secondary structure template among domains/monomers of pepsin-like aspartic proteases and retropepsins supports the view that these proteases are evolutionarily related, and that pepsins may have arisen by gene duplication and fusion of an ancestral form of retropepsins. The nature of this primordial single-lobed aspartic protease has been the matter of debate over the years, mostly due to the lack of compelling evidence for the presence of these enzymes in prokaryotes. However, this argument has been first challenged with finding pepsin homologs in a restricted number of bacteria and the observation that at least one of these genes encodes an active enzyme. More recently, we reported the identification of a gene coding for a membrane-embedded, single-lobed aspartic protease, highly conserved in the genomes of 55 species of *Rickettsia*. Using *Rickettsia conorii* gene homolog rc1339, we provided evidence that the encoded product (APRc), indeed shares several enzymatic properties with viral retropepsins (Cruz, *et al*., PLoS Pathog., 2014).

These resemblance of enzymatic features suggested that APRc might indeed represent a more primordial form of retropepsins. In this work, crystal structures of two constructs of the soluble domain of RC1339/APRc from *Rickettsia conorii* have been determined in three different crystal forms. The results clearly show that the fold of APRc monomer resembles that of retropepsins, but the quaternary structure of the dimer differed from the canonical retropepsins. Whereas the observed dimer is most likely an artefact of expression and/or crystallization since it cannot support the previously reported enzymatic activity of APRc, the fold of the core of each monomer is very closely related to the fold of retropepsins.

Overall, our results support the concepts that APRc may indeed represent a putative common ancestor of monomeric and dimeric aspartic proteases, as well as possible existence of a different evolutionary pathway for these enzymes.