Evolutionary and functional studies on the novel Hepatitis C virus core+1/ARF protein

Ioly Kotta-Loizou

Division of Cell & Molecular Biology, Imperial College London
i.kotta-loizou13@imperial.ac.uk

Abstract. Hepatitis C virus (HCV) infection is a major cause of chronic liver disease and hepatocellular carcinoma (HCC). HCV is an enveloped positive-stranded RNA virus, belonging to the *Flaviviridae* family. The HCV genome produces a polyprotein precursor which is processed by proteases and yields at least 10 proteins. HCV possesses a second open reading frame (ORF) within the core gene, encoding an intrinsically disordered protein known as core+1/ARFP. Core+1/ARFP has been implicated in HCC and there is evidence that it acts as a modulator; however, its function in the context of the HCV life cycle still remains elusive. The aim of this study was to gain insight into the biological role of core+1/ARFP. To this end, a dataset of 4428 core+1/ARF amino acid sequences from 18 confirmed HCV subtypes was constructed. As shown before, core+1/ARF ORF is present in all subtypes and its length is subtype-dependent. Furthermore, core+1/ARFP is less conserved than core but at least as conserved as the HCV non-structural (NS) 2 protein. Subsequently, 24 exemplar core+1/ARF amino acid sequences, one or two from each subtype, were used for further analysis. Compositional profiling of core+1/ARFP revealed an enrichment in both disorder-promoting and high-degeneracy residues; the latter a tendency of overlapping genes. Phylogenetic analysis of the *Flaviviridae* family supports the notion that core+1/ARF ORF originated *de novo* by overprinting within the ancestral core gene and estimation of selection constraint indicates that core+1/ARF ORF is under weak purifying selection. In addition, Hidden Markov Models (HMMs) revealed an interesting similarity between core+1/ARFP and the RNA-binding amino-terminal domain of core, which suggests that core+1/ARFP may also be an RNA-binding protein. Subsequently, core+1/ARF amino acid sequences were assessed using the Eukaryotic Linear Motif (ELM) database for the presence of protein binding and modification motifs conserved in all HCV subtypes. Finally, both protein-protein interaction and transcriptional regulation networks were generated using Cytoscape, based on data derived from the literature, and their analysis revealed that core+1/ARFP is implicated in regulation and response to external stimuli and stress, as well as in major signaling pathways known to play a role in HCC. In conclusion, this study provides insight into the molecular evolution of core+1/ARFP. Moreover, the modulatory role of core+1/ARFP is confirmed and expanded as new putative targets are identified – to be experimentally verified in the future – and core+1/ARFP's association with HCC is elaborated.