Inorganic Chemistry: An Indian Journal

Short Communication | Vol 16 Issue 2

Genus specific protein patterns of viruses

Sandeep Bansode

Dr D Y Patil Biotechnology and Bioinformatics Institute, India

E-mail: sandeep.bansode@dpu.edu.in

In the era of emerging and re-emerging viral infections, diagnostics and its allied fields have a significant role to play in combating the diseases. Enormous amount of the molecular sequence data available within the property right has the potential to contribute in a very major way within the development of novel diagnostic tools. One in every of the perquisites for such a study is that the identification of signature sequences i.e., small stretches of protein/nucleotide sequences that are unique to a given family/genus/organism. There exist several resources within the property right archiving signature sequences of proteins supported sequence identity/ similarity. However, these resources don't take under consideration the taxonomic information which encompasses a significant role to play in viral diagnostics. The current study is an endeavour to explicitly take under consideration the taxonomic information and thereby derive genus-specific signature sequences of viral proteins. The VirGen database provides early data for obtaining patterns, such as multiple sequence alignment (MSA). The patterns are extracted from the MSA using a perl script written in-house. The patterns are then confirmed by searching the NCBI non-redundant protein sequence database, allowing the sensitivity and specificity to be calculated. True-positives and true-negatives datasets are required for this validation. The total number of species belonging to a specific genus is retrieved using an Entrez query from the NCBI taxonomy database, yielding a true-positive dataset. Any protein sequence from a genus other than the one in question was included in the true-negative dataset. Patterns could be discovered for 125 of the 262 proteins in VirGen belonging to 19 families (RNA viruses), all of which clearly separated true-positives and false-positive sequences. These patterns are discovered to be part of crucial functional areas including the active site and dimerization interface when mapped onto their respective 3D structures (25 unique entries in Protein Data Bank). The resulting unique viral signature sequences/peptides can be used not only in detection tests and treatments, but also as prospective targets for viral vaccines.

The nucleocapsid (N) protein is important for genome functionality and is found in all four coronavirus genera: alpha, beta, gamma, and delta. Corona viral N sequences indicated two intrinsically disordered regions (IDRs) at the polypeptide's core, according to bioinformatic analysis. While both IDR structures were identified in alpha, beta, and gamma-coronaviruses, delta coronaviruses lacked the second IDR. The second IDR structure of two novel coronaviruses, currently assigned to the Gamma coronavirus genus, appeared intermediate in this regard, with a low probability of disorder. Interestingly, these two coronaviruses are the only ones known to have been isolated from marine mammals, notably the beluga whale and the bottlenose dolphin, two closely related species; the viruses' N proteins were almost similar, varying by only one amino acid. Because gamma coronaviruses are mostly found in birds, these two viruses remain phylogenetic outliers. Finally, regardless of the coronavirus genus in which they were found, both IDRs were high in Ser and Arg, which corresponded to their disordered structure The central IDRs are thought to play a key part in the nucleocapsid protein's multitasking activity, necessitating structural plasticity and perhaps affecting coronavirus host tropism and cross-species transmission.

ISSN: 0976-746X