CURRENT SCIENCE

Volume 119 Number 7 10 October 2020

GUEST EDITORIAL

COVID-19, structural biology and the march of science

Max Perutz and John Kendrew determined the structures of haemoglobin and myoglobin in 1962 and 1960 respectively. Their achievement was the result of more than two decades of sustained efforts. Their success was recognized by the award of the 1962 Nobel Prize in Chemistry and celebrated as the beginning of structural studies on proteins by X-ray diffraction techniques.

The first report on a novel virus infecting the upper respiratory tract and causing common cold-like symptoms in humans appeared in 1965 (Tyrrell, D. A. J. and Bynoe, M. L., *Br. Med. J.*, 1965, **1**, 1467–1470). The publication went without much fanfare. Two years later, electron micrographs of the virus were obtained (Almeida, J., *Nature*, 1968, **220**, 16). Virus particles were found to be distorted spheroids of 70–80 nm diameter. It was named coronavirus based on the appearance of a protein (spike protein) protruding outwards from the virus particles and resembling solar corona during total solar eclipse. Coronaviruses infect a large number of animals such as bats, rats, cats, dogs, monkeys, pangolin, cattle, etc. The protein capsid of coronaviruses that encapsulates and protects the genome of the virus is constructed from four proteins, the spike protein, the envelope protein, the membrane protein and the nucleic acid associated protein. The capsids are also surrounded by a bilayer lipid membrane. The positively charged nucleic acid associated protein interacts with the negatively charged genome. The size of coronavirus is large compared to common human viruses such as poliovirus and common cold virus, the particles of which are rigid and spherical (icosahedral) with diameters of ~30 nm. In the Guangdong Province of China, infection by a novel strain of coronavirus was recognized in 2003 to cause a severe infection of the human respiratory tract leading to breathing difficulties. It was thought to be a zoonotic virus transferred by animals to humans. The virus strain was named Severe Acute Respiratory Syndrome coronavirus (SARS; coronavirus). The infection affected more than 8000 individuals causing over 700 deaths in 30 different countries. Instantly, crystallographers (and to some extent NMR experts) around the world and later electron microscopists started working furiously towards determining the three-dimensional structures of coronavirus genome-encoded proteins. The pace of the work increased after 2012 when another strain of coronavirus was found to cause widespread infection in Saudi Arabia (Middle Eastern Respiratory Syndrome MERS-coronavirus). MERS coronavirus infection spread to many Arabian countries and caused several thousand deaths. The current (2019) pandemic of SARS coronavirus, COVID-19, is unprecedented in its impact and the infection has spread to almost all countries of the world with lightning speed causing inestimable damage, although the mortality rate of COVID-19 is less than those of SARS- and MERS-coronaviruses. This has resulted in an unbelievable volume of research on coronavirus-encoded proteins. The volume of work done on the function of proteins coded by coronavirus genome and the mechanism of its entry and stages of replication is even larger. Only the work on the structural biology of coronaviruses is briefly described here. Because of the heterogeneity in particle shape, it has not been possible to crystallize the virus and determine the structure by X-ray diffraction techniques.

Of all the known single-stranded RNA viruses, the genome of coronaviruses is the largest and consists of 27,000–30,000 nucleotides. The genome size of all other known single-stranded RNA viruses infecting animals varies between 4000 and 12,000 nucleotides. The genome of coronaviruses codes for four proteins (S for spike protein; E for envelope protein; N for nucleic acid associated protein and M for membrane protein) that are expressed through sub-genomic messenger RNAs. These four proteins get incorporated into mature virus particles. Apart from these ‘structural proteins’, the genome codes for another 16 non-structural proteins (nsp) that are essential for virus multiplication. These proteins are produced by post-translational cleavage of two long polyprotein chains resulting from the translation of 5′-terminal open reading frame of the genome by two viral encoded proteases. Any protein essential for virus life cycle could serve as a target for antiviral drugs, which is the reason why crystallographers and other structural biologists are busy elucidating coronavirus encoded protein structures.

The structure of coronavirus main protease (nsp5), essential for cleaving the polyprotein into functional protein domains (nsps), was determined in 2003

(Anand, K. *et al.*, *Science*, 2003, **300**, 1763–1767). Subsequently, several high affinity inhibitors that bind to the protease active site have been designed. Structure of another papain-like protease (nsp3) was elucidated in 2006 (Ratia, K. *et al.*, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 5717–5722), following which suitable inhibitors were designed (Lee, H. *et al.*, *ACS Chem. Biol.*, 2015, **10**, 1456–1465). The structure of non-structural protein 12 (nsp12) that functions as an RNA dependent RNA polymerase responsible for the synthesis of progeny RNA chains was determined by cryo-electron microscopy to a resolution of 2.9 Å (Robert, N. *et al.*, *Nature Commun.*, 2019; https://doi.org/10.1038/s41467-019-10280-3; Go, *et al.*, *Science*, 2020; 10.1126/science.abb7498). Mouse coronavirus spike protein structure was determined earlier by cryo-electron microscopy. The way it interacts with angiotensin-converting enzyme 2 (ACE2), the receptor present on the host cell membrane and the resulting conformational changes in the spike protein leading to the entry of virus into susceptible cells have subsequently been elucidated. The structure of COVID-19 spike protein was determined by cryo-electron microscopy in record time after its infection became pandemic (Alexandra, C. *et al.*, *Cell*, 2020, **180**, 281–292). Structures of other non-structural proteins determined include X-ray crystal structures of nsp1, nsp4, nsp9, nsp10, nsp15 and nsp16 at resolutions of 1.6 Å, 2.8 Å, 1.8 Å, 1.8 Å, 2.6 Å and 2.0 Å respectively. Structures of nsp7 and nsp8 have been determined by cryo-electron microscopy at 3.1 Å resolution. These structures may lead to the discovery of effective drugs against coronaviruses. When compared to the initial toil by pioneers like Perutz and Kendrew, the pace with which these structural studies have been carried out is astonishing and illustrates the grand march of science resulting from developments in experimental techniques and theoretical and computational methods.

In parallel with structural studies, biochemical investigations on the functions of coronavirus-encoded proteins, the mechanism of virus entry and replication have also been extensively investigated. Hundreds of laboratories are now engaged in the development of an effective vaccine against COVID-19. The international competition to bring out effective vaccines is being carried out at neck-breaking speed. It is likely that a few of these efforts will succeed. However, the fight against coronavirus will by no means end. There are other strains of coronaviruses that infect wild animals. These viruses might acquire the ability to infect humans, particularly when people spend long periods of time in close proximity of wild animals in live animal markets. Also, coronaviruses undergo recombination leading to new strains of the virus. The vaccine developed for COVID-19 may not be effective against emerging strains.

The story of another virus, Zika virus, reveals yet another aspect of scientific research. Zika virus was first identified in captive rhesus monkeys in the Zika valley of Uganda in 1947. It was also found in *Aedes africanus* mosquitoes, suggesting the identity of the vector responsible for infection. Sporadic human infection by Zika virus was noticed in several African countries in the 1960s, although the identity of the pathogen was not well established. The infection caused rashes on skin, conjunctivitis, muscle and joint pain. However, it did not raise alarm as the symptoms were mild and patients recovered from the infection within a week or ten days without treatment. Large scale human infection by Zika virus was first noticed in the Yap islands far away from Uganda in 2007. Yap Islands are a small group of islands in the Pacific Ocean and are situated to the east of Malaysia. Although it spread to several east Asian countries, the seriousness of infection was not realized. In 2015, Zika virus appeared in Brazil, initially infecting thousands of people. It quickly became an epidemic. The thousands of infected people included pregnant women. Retarded brain growth (microcephaly) was observed in some of the children born to infected mothers. Immediately structural studies were initiated on virus particles. Zika virus capsid has icosahedral symmetry. The uniform shaped capsids consist of 180 copies of the envelope glycoprotein and the membrane protein anchored to the lipid membrane. Cryo-electron microscopy, which has now developed into a tool rivalling X-ray diffraction, was used to obtain Zika virus structure in 2016 (Devika Sirohi, *et al.*, *Science*, 2016; doi:10.1126/science.aaf5316). The structure determination was accomplished in a few months' time. Although the resolution of the EM structure was only 3.9 Å, it was possible to get atomic resolution structure by fitting the known structure of the envelope protein into the cryo-EM map. The carbohydrate moiety attached to an asparagine residue of the glycoprotein on the outer surface of the capsid was proposed to function as an attachment site of the virus to host cells and hence may be a suitable site for drug development against Zika virus. In parallel, it was shown that Zika virus is closely related to flaviviruses, West Nile virus, Dengue virus and Japanese encephalitis virus. Although the virus spread to Mexico and southern region of the United States, the infection did not become a worldwide epidemic. Therefore, unlike in the case of coronavirus, Zika virus did not initiate frantic efforts to develop vaccines. The work towards understanding the structure, infection and replication mechanisms of these two viruses are examples of how quickly our understanding could be advanced by concerted, global efforts.

M. R. N. Murthy

Molecular Biophysics Unit,
Indian Institute of Science,
Bengaluru 560 012, India
e-mail: mrn@iisc.ac.in