Original article

IRAM: virus capsid database and analysis resource

Iman Almansour[1],*, Mazen Alhagri[2], Rahaf Alfares[1], Manal Alshehri[1], Razan Bakhashwain[3] and Ahmed Maarouf[3]

[1]Epidemic Diseases Department, Institute for Research and Medical Consultations, Imam Abdulrahman Bin Faisal University, P.O.Box 1982, Dammam 31441 Saudi Arabia [2]Scientific and High Performance Computing Center, Deanship of Information and Communication Technology, Imam Abdulrahman Bin Faisal University, P.O.Box 1982, Dammam 31441 Saudi Arabia and [3]Department of Physics, Institute for Research and Medical Consultations, Imam Abdulrahman Bin Faisal University, P.O.Box 1982, Dammam 31441 Saudi Arabia

*Corresponding author: Phone: (+966)013-3330000, ext. 30892; Email: ikalmansour@iau.edu.sa

Abstract

IRAM is an online, open access, comprehensive database and analysis resource for virus capsids. The database includes over 200 000 hierarchically organized capsid-associated nucleotide and amino acid sequences, as well as 193 capsids structures of high resolution (1–5 Å). Each capsid's structure includes a data file for capsid domain (PDB), capsid symmetry unit (PDB) and capsid structure information (PSF); these contain capsid structural information that is necessary to run further computational studies. Physicochemical properties analysis is implemented for calculating capsid total charge at given radii and for calculating charge distributions. This resource includes BLASTn and BLASTp tools, which can be applied to compare nucleotide and amino acid sequences. The diverse functionality of IRAM is valuable to researchers because it integrates different aspects of virus capsids via a user-friendly interface. Such data are critical for studying capsid evolution and patterns of conservation. The IRAM database can also provide initial necessary information for the design of synthetic capsids for various biotechnological applications.

Database URL: https://iram.iau.edu.sa

Introduction

Capsids are monomeric protein shells that enclose viral nucleic acids and uniquely protect viruses from external conditions. A typical virus capsid comprises protein subunits that are grouped into morphological units called 'capsomers'; these self-assemble to form the complete structure (1–3). In certain viruses, capsids are encoded by a single gene (4, 5), while in others, capsids are more complex
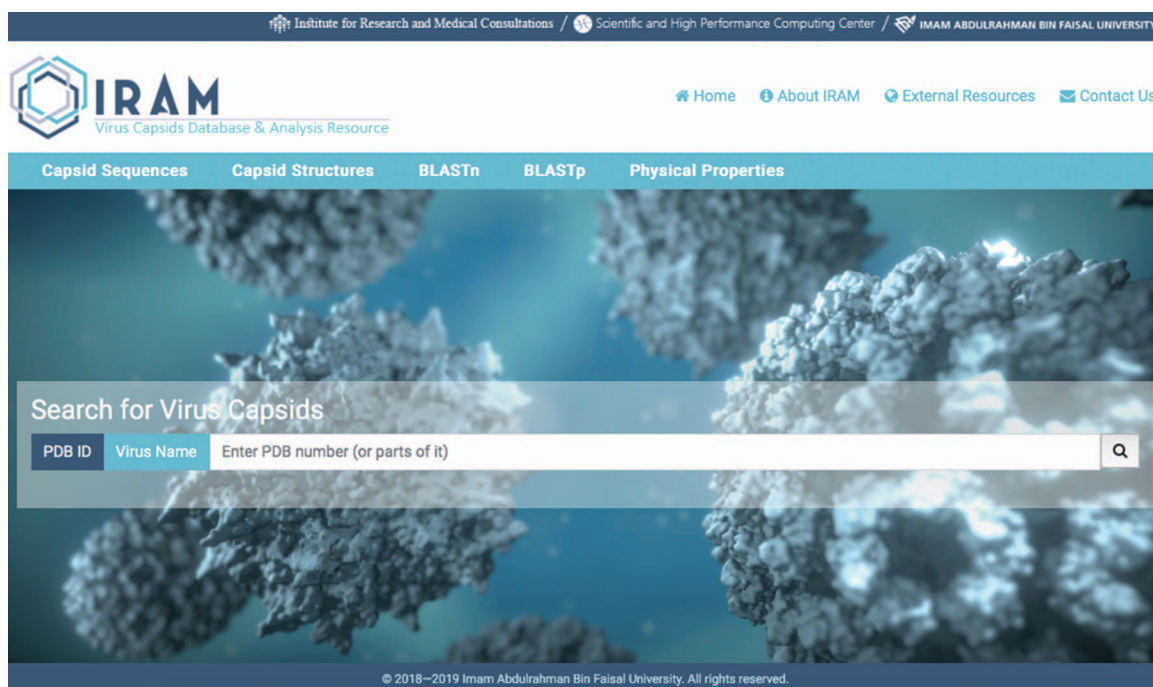
Figure 1. Snapshot of IRAM homepage.

and are generated from multiple polypeptide chains (6, 7). Capsids vary considerably in size, organization and symmetry (8–11).

Furthermore, virus capsids have naturally evolved to deliver their own genetic material into host cells with high efficiency. Studies have shown that capsid assembly is affected by the type of nucleic acid in the virus genome: in RNA viruses, capsids self-assemble around the viral nucleic acid; comparatively, in DNA viruses, capsid packaging occurs after capsid assembly (12–15). Due to their self-assembly properties, capsids have gained considerable attention in the gene delivery field as powerful carriers of nucleic acid vaccines and gene therapy (16–21). Similarly, capsids are being exploited for medical diagnostics, bioimaging and other bionanotechnological applications (22–25).

Despite remarkable genetic diversity among viruses, evolutionary studies on capsids sequences have been largely focused on homologous viruses (26–35). Genomics and evolutionary studies among diverse viruses' groups and families, however, have been limited (36–38). This is partially due to the unavailability of specialized databases on capsid sequences. Such data are crucial in understanding capsid evolution and patterns of conservation among diverse viral groups and host ranges. Furthermore, these data are necessary to help determine the relationship between sequence conservation and protein structure–function.

Since the discovery of virus structures by Casper and Klug (9, 39), x-ray crystallography and cryo-electron microscopy have provided an enormous body of information concerning capsid structure at or near the atomistic level. Despite these efforts, reports on the physical properties of capsids—such as charge distributions, assembly and disassembly—remain scarce, owing to their great complexity compared with other structural proteins (40–43). Thus, availability of specialized databases covering physical characteristics of individual capsid structure is crucial for accurate capsid modeling and for informing further biophysical simulation studies (Figure 1).

Presently, data on capsids are dispersed in different databases. The National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov) is a comprehensive database for all genes and protein sequences; the RCSB Protein Data Bank (https://www.rcsb.org) is a repository database for experimentally verified three-dimensional structural data of biological macromolecules; and the Virus Particle Explorer Database (http://viperdb.scripps.edu) is a specialized database for capsid PDB structures of diverse resolutions stored uniformly in $z(2)$-3–5-$x(2)$ conventions (44). A comprehensive database for virus capsids nucleotide and amino acids sequences organized into taxonomic classification of viruses, as well as database for high-resolution capsids structural information, has yet to be developed. To address these needs, we present IRAM (https://iram.iau.edu.sa/), an open access online database of virus
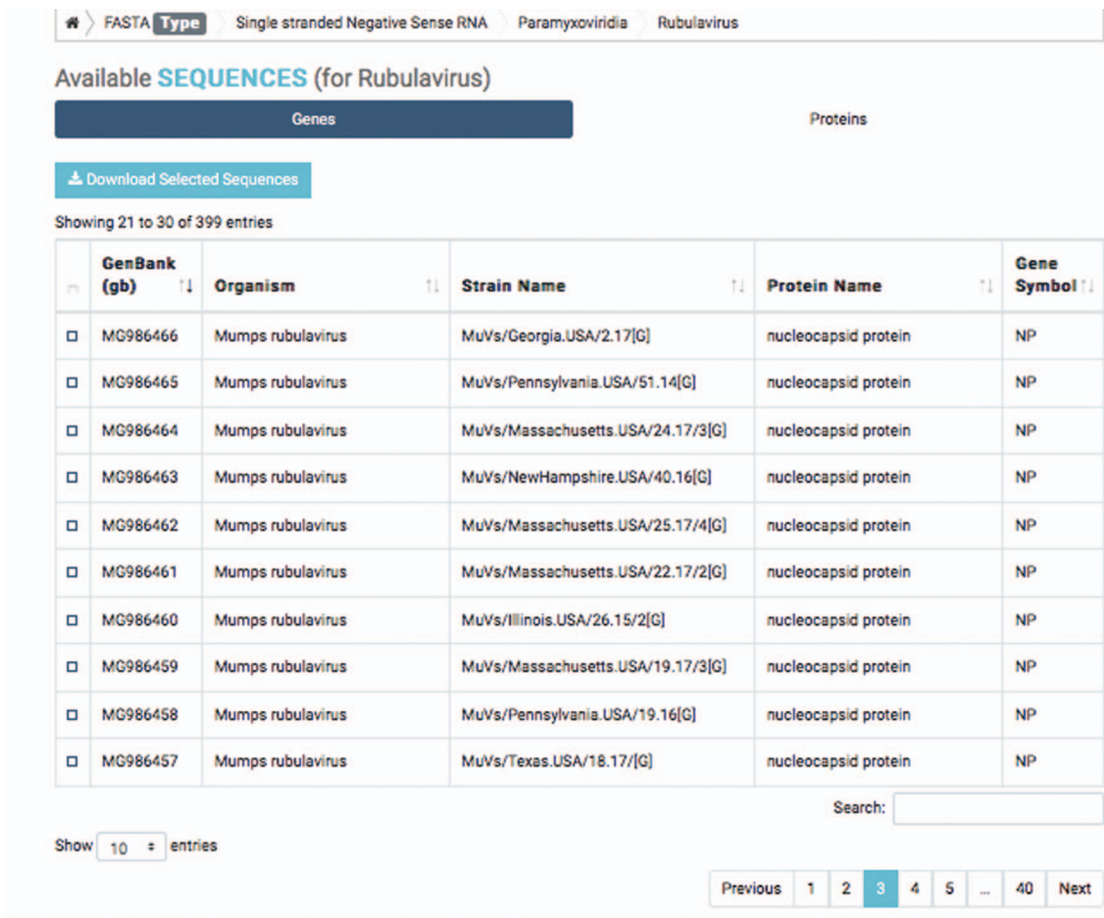
Figure 2. Detailed result of capsid nucleotide and amino acid sequences of a selected virus.

capsid information coupled with sequence- and structure-analytic capabilities. IRAM offers five major features:

(i) capsid sequences representing 15 different virus families;
(ii) capsid structural data files (PDB) at high resolution (1–5 Å);
(iii) capsid structural data files (PSF) containing physical attributes of capsids atoms;
(iv) capsid primary sequence alignments generated from BLASTn and BLASTp searches; and
(v) physicochemical properties calculator.
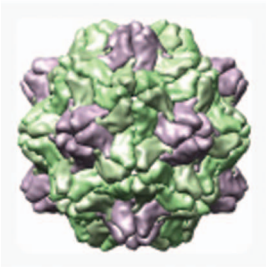
Database architecture

Capsid sequences

The first database includes over 200 000 capsid nucleotide and amino acid sequences, which have been manually curated from the NCBI and the National Institute of Allergy and Infectious Diseases' Virus Pathogen Database and Analysis Resource (45). Database organization is based on the type of viral nucleic acid the capsids carry: single-stranded (ss) DNA, ssRNA, double-stranded (ds)

DNA and dsRNA. Sequence taxonomy is based on the genomic classification of 15 virus families: *Arenaviridae*, *Bromoviridae*, *Bunyaviridae*, *Caliciviridae*, *Coronaviridae*, *Flaviviridae*, *Hepeviridae*, *Herpesviridae*, *Paramyxoviridae*, *Picornaviridae*, *Poxviridae*, *Reoviridae*, *Rhabdoviridae*, *Togaviridae* and *Virgaviridae*. Sequences are subcategorized according to genus and species. Users can download selected nucleotide or amino acid sequences; batch sequences affiliated with a given virus name can also be downloaded (Figure 2).

Capsid structures

The second database includes structural data of 193 experimentally categorized capsid protein structures at high resolution (1–5 Å). Each capsid entry contains three capsid data files: capsid structural domains (PDB domain), retrieved from the PDB (46); capsid complete symmetry unit (Full capsid PDB), generated by Python-implemented Chimera (47); and capsid structure information (Full Capsid PSF), generated by CHARMM-implemented VMD (48) (Figure 3). Each PSF file generated includes topology and parameter subfiles containing data on the capsid's atoms,

Figure 3. Details of a selected capsids structure. Capsid files: capsid domain PDB, full capsids PDB, full capsid PSF and capsid FASTA.

bonds, angles, dihedrals, impropers (dihedral force terms used to maintain planarity) and cross terms. The PSF file contains necessary information for applying further molecular dynamics simulations—such as coarse-grained and all-atom molecular dynamics—to capsid structures.

We classified capsid structures based on resolution (1–5 Å). For each capsid structure entry, a link to a page was provided; the page contained the capsid's name, PDB ID, triangular number, protein symmetry, residue counts, atom counts, method used, capsid PDB link and reference link. The interface allows the user to query by virus name or PDB ID (Figure 3).

BLAST searches

BLASTn and BLASTp are provided as complementary tools for analyzing capsid nucleotide and amino acid sequences, respectively. The IRAM database BLAST module is built with NCBI BLAST+2.7.0, which allows users to compare sequences against the locally generated sequence database. The selection of specific BLASTn searches—such as blastn-short, dc-megablast and megablast, as well as blastp searches (blastp-short and blastp-fast)—is customized, and an E value output option is available. Results are displayed on the webpage and can also be downloaded.

Physical properties calculator

Investigating the physicochemical properties of virus capsids may necessitate knowledge of their electronic charge

distributions ([49](#)) and atomistic modeling ([50](#),[51](#)). Charges of various atoms in the capsids of PDB-derived PSFs are treated as 'point' charges; however, because atomic charges are distributed throughout neighboring atomic positions, a point charge representation can lead to erroneous charge density calculations. This inaccuracy can be resolved by constructing a charge density function, ρ, defined at a general point r, where r is measured from the center of the capsid. Therefore, we adopted a continuous and realistic charge density model by assuming that atomic charge takes a Gaussian form and that it is centered on each atom:

$$\rho(r) = \frac{1}{(4\pi a)^{(3/2)}} \sum_{i=1}^{N} q_i e^{-\left(\frac{r-r_i}{2a}\right)^2},$$

where r and q_i are the location and charge of atom i, respectively; a is width (approximately 1.0 Å); and N is the number of atoms within the capsid. A reasonable estimate of a could be the van der Waals radius of an atom. Charges of various capsid atoms were taken from the PSFs, and auxiliary files were created for each capsid structure to facilitate various calculations. A charge density map can be plotted by using this density function, such as on the mid-plane of the capsid.

One advantage of using the constructed charge density function is that it involves an accurate calculation of the surface charge as a function of distance from the capsid's center, as well as the charge contained within a certain volume of the capsid. This calculation is accomplished by integrating the charge density function over the desired
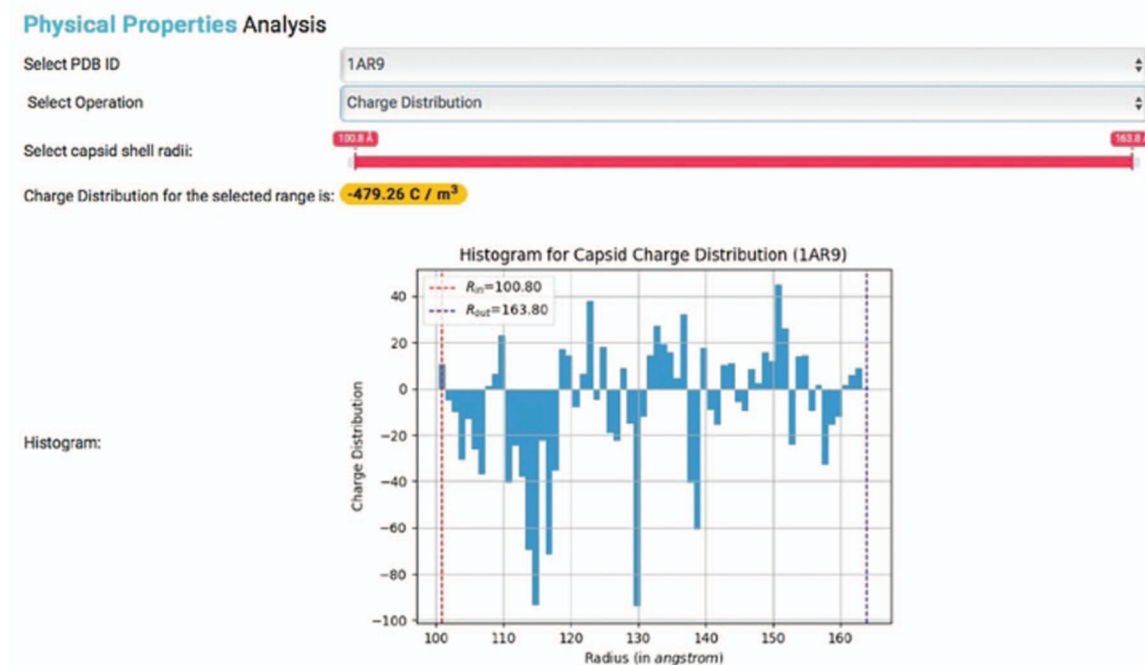
Figure 4. Physical properties analysis. Histogram showing the charge distribution of a selected capsid PDB ID, where R_{in} represents inner capsid radii and R_{out} represents outer capsid radii.

region. Alternatively, this charge can also be obtained by a simple sum of the charges within the volume concerned. For example, to determine the charge contained between shells at 7 Å and at 10 Å, we use

$$Q_{7 \to 10} = \int_7^{10} dr \int_0^{\pi} d\theta \, \sin(\theta) \int_0^{2\pi} d\phi \, r^2 \rho(r, \theta, \phi).$$

The surface charge density $\sigma(r)$ is obtained from

$$\sigma(r) = r^2 \int_0^{\pi} d\theta \, \sin(\theta) \int_0^{2\pi} d\phi \, \rho(r, \theta, \phi).$$

These built-in functions allow for the generation of various types of charge density plots for capsids in IRAM. Users with PDB IDs can also obtain corresponding PSFs and calculate capsid physical properties (Figure 4).

Implementation

The web interface of IRAM was written in standard HTML/JavaScript/CSS using the Vue.JS framework at the front end. The back end was written in GO, building on the Go-Swagger OpenAPI framework. MongoDB was used for data storage. Docker was utilized to package and deploy all IRAM web application components. Data analytics and physical property computations of IRAM were implemented in Python and distributed using C++.

Discussion and Conclusion

Here we introduced IRAM, an integrative platform and repository of over 200 000 capsid nucleotide and amino acid sequences and nearly 200 high-resolution capsid structures. The uniquely generated capsid structural information from PSF files can be used to study various biophysical molecular dynamics of capsids, such as assembly, disassembly and mechanical properties. We also implemented sequence analysis tools to aid researchers in exploring the evolutionary aspects of viral capsids and characterizing diverse host–virus interactions. Furthermore, we included a physical properties calculator of capsid charge distributions, which enables the study of capsids that are structurally similar but genetically divergent (42).

Engineering virus capsids is a vibrant area in synthetic biology, whereby capsids are exploited as drug and gene carriers. In this context, a specialized database, such as IRAM, is valuable for selecting capsid candidates based on sequence similarities and structural/physicochemical properties. For example, determining the inner and outer charges of capsids may contribute to the selection of capsids that are ideal for particular drug encapsulation. Furthermore, by integrating sequence and structural data into a single capsid database, researchers can effectively apply evolution-guided design of synthetic virus capsids.

We continue to add available capsid sequences to cover wider range of viruses for integration into the IRAM database. In addition, capsid structures will be updated

as the relevant data become available. Future plans include capsid immune epitopes as well as expansion of physical properties analysis. Furthermore, we are developing a tool for introducing point mutations in wild-type capsids to create engineered virus models with unique properties; these can be studied for targeted delivery, altered tropism and evasion from antibody neutralization.

References

1. Fraenkel-Conrat,H. and Williams,R.C. (1955) Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc. Natl. Acad. Sci. USA*, **41**, 690–698.
2. Crick,F.H. and Watson,J.D. (1956) Structure of small viruses. *Nature*, **177**, 473–475.
3. Crick,F.H. and Watson,J.D. (1957) Virus structure: general principles. *Nat Viruses*, **5**, 5–18.
4. Grimes,J.M., Burroughs,J.N., Gouet,P. *et al*. (1998) The atomic structure of the bluetongue virus core. *Nature*, **395**, 470.
5. Reinisch,K.M., Nibert,M.L. and Harrison,S.C. (2000) Structure of the reovirus core at 3.6 A resolution. *Nature*, **404**, 960.
6. Luque,D., Mata,C.P., Gonzalez-Camacho,F. *et al*. (2016) Heterodimers as the structural unit of the T=1 capsid of the fungal double-stranded RNA Rosellinia necatrix Quadrivirus 1. *J. Virol*., **90**, 11220–11230.
7. Grzesik,P., MacMath,D., Henson,B. *et al*. (2017) Incorporation of the Kaposi's sarcoma-associated herpesvirus capsid vertex-specific component (CVSC) into self-assembled capsids. *Virus Res*., **236**, 9–13.
8. Klug,A. and Caspar,D.L.D. (1961) The structure of small viruses. *Adv. Virus Res*., **7**, 225–325.
9. Caspar,D.L. and Klug,A. (1962) Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol*., **27**, 1–24.
10. Johnson,J.E. and Speir,J.A. (1997) Quasi-equivalent viruses: a paradigm for protein assemblies. *J. Mol. Biol*., **269**, 665–675.
11. Harrison,S.C., Knipe,D.M. and Howley,P.M.(Eds.) (2001) *Principles of Virus structure*, Vol. **1**. Philadelphia: Lippincott Williams & Wilkins, p. 59–85.
12. Bancroft,J.B., Bracker,C.E. and Wagner,G.W. (1969) Structures derived from cowpea chlorotic mottle and brome mosaic virus protein. *Virology*, **38**, 324–335.
13. Lee,J.Y., Irmiere,A. and Gibson,W. (1988) Primate cytomegalovirus assembly: evidence that DNA packaging occurs subsequent to B capsid assembly. *Virology*, **167**, 87–96.
14. Homa,F.L. and Brown,J.C. (1997) Capsid assembly and DNA packaging in herpes simplex virus. *Rev. Med. Virol*., **7**, 107–122.
15. Bruinsma,R.F. (2006) Physics of RNA and viral assembly. *Eur. Phys. J. E Soft Mater*, **19**, 303–310.
16. Choi,K.M., Choi,S.H., Jeon,H. *et al*. (2011) Chimeric capsid protein as a nanocarrier for siRNA delivery: stability and cellular uptake of encapsulated siRNA. *ACS Nano*, **5**, 8690–8699.
17. Choi,K.M., Kim,K., Kwon,I.C. *et al*. (2012) Systemic delivery of siRNA by chimeric capsid protein: tumor targeting and RNAi activity in vivo. *Mol. Pharm*., **10**, 18–25.
18. Dalkara,D., Byrne,L.C., Lee,T. *et al*. (2012) Enhanced gene delivery to the neonatal retina through systemic administration of tyrosine-mutated AAV9. *Gene Ther*., **19**, 176.
19. Freire,J.M., Veiga,A.S., Conceição,T.M. *et al*. (2013) Intracellular nucleic acid delivery by the supercharged dengue virus capsid protein. *PLoS One*, **8**, e81450.
20. Kay,C.N., Ryals,R.C., Aslanidi,G.V. *et al*. (2013) Targeting photoreceptors via intravitreal delivery using novel, capsid-mutated AAV vectors. *PLoS One*, **8**, e62097.
21. Medina-Kauwe,L.K. (2013) Development of adenovirus capsid proteins for targeted therapeutic delivery. *Ther. Deliv*., **4**, 267–277.
22. Czapar,A.E. and Steinmetz,N.F. (2017) Plant viruses and bacteriophages for drug delivery in medicine and biotechnology. *Curr. Opin. Chem. Biol*., **38**, 108–116.
23. Li,K., Nguyen,H.G., Lu,X. *et al*. (2010) Viruses and their potential in bioimaging and biosensing applications. *Analyst*, **135**, 21–27.
24. Chen,W., Cao,Y., Liu,M. *et al*. (2012) Rotavirus capsid surface protein VP4-coated Fe_3O_4 nanoparticles as a theranostic platform for cellular imaging and drug delivery. *Biomaterials*, **33**, 7895–7902.
25. Zeng,Q., Wen,H., Wen,Q. *et al*. (2013) Cucumber mosaic virus as drug delivery vehicle for doxorubicin. *Biomaterials*, **34**, 4632–4642.
26. Martinez,M.A., Dopazo,J., Hernandez,J. *et al*. (1992) Evolution of the capsid protein genes of foot-and-mouth disease virus: antigenic variation without accumulation of amino acid substitutions over six decades. *J. Virol*., **66**, 3557–3565.
27. Oberste,M.S., Maher,K., Kilpatrick,D.R. *et al*. (1999) Molecular evolution of the human enteroviruses: correlation of serotype with VP1 sequence and application to picornavirus classification. *J. Virol*., **73**, 1941–1948.
28. Nilsson,M., Hedlund,K.O., Thorhagen,M. *et al*. (2003) Evolution of human calicivirus RNA in vivo: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. *J. Virol*., **77**, 13117–13124.
29. Siebenga,J.J., Vennema,H., Renckens,B. *et al*. (2007) Epochal evolution of GGII. 4 norovirus capsid proteins from 1995 to 2006. *J. Virol*., **81**, 9932–9941.
30. Bull,R.A., Eden,J.S., Rawlinson,W.D. *et al*. (2010) Rapid evolution of pandemic noroviruses of the GII. 4 lineage. *PLoS Pathog*., **6**, e1000831.
31. Streck,A.F., Bonatto,S.L., Homeier,T. *et al*. (2011) High rate of viral evolution in the capsid protein of porcine parvovirus. *J. Gen. Virol*., **92**, 2628–2636.

32. Mushegian,A., Karin,E.L. and Pupko,T. (2018) Sequence analysis of malacoherpesvirus proteins: pan-herpesvirus capsid module and replication enzymes with an ancient connection to "Megavirales". *Virology*, **513**, 114–128.

33. Shaw,J., Jorba,J., Zhao,K. *et al.* (2018) Dynamics of evolution of poliovirus neutralizing antigenic sites and other capsid functional domains during a large and prolonged outbreak. *J. Virol.*, **92**, e01949-17.

34. De Grazia,S., Lanave,G., Bonura,F. *et al.* (2018) Molecular evolutionary analysis of type-1 human astroviruses identifies putative sites under selection pressure on the capsid protein. *Infect. Genet. Evol.*, **58**, 199–208.

35. Zhang,S., Qu,C., Wang,Y. *et al.* (2018) Conservation and variation of the hepatitis E virus ORF2 capsid protein. *Gene*, **675**, 157–164.

36. Dolja,V.V., Boyko,V.P., Agranovsky,A.A. *et al.* (1991) Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: two families with distinct patterns of sequence and probably structure conservation. *Virology*, **184**, 79–86.

37. Chare,E.R. and Holmes,E.C. (2004) Selection pressures in the capsid genes of plant RNA viruses reflect mode of transmission. *J. Gen. Virol.*, **85**, 3149–3157.

38. Chang,C.M., Huang,Y.W., Lee,C.W. *et al.* (2015) Sequence conservation, radial distance and packing density in spherical viral capsids. *PLoS One*, **10**, e0132234.

39. Klug,A. (1999) The tobacco mosaic virus particle: structure and assembly. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **354**, 531–535.

40. Lucas,W. (2010) Viral capsids and envelopes: structure and function. *In: eLS, John Wiley & Sons Ltd, Chichester*.

41. Rossmann,M.G. and Johnson,J.E. (1989) Icosahedral RNA virus structure. *Annu. Rev. Biochem.*, **58**, 533–569.

42. Dokland,T. (2000) Freedom and restraint: themes in virus capsid assembly. *Structure*, **8**, R157–R162.

43. Bamford,D.H., Grimes,J.M. and Stuart,D.I. (2005) What does structure tell us about virus evolution? *Curr. Opin. Struct. Biol.*, **15**, 655–663.

44. Reddy,V.S., Natarajan,P., Okerberg,B. *et al.* (2001) Virus particle explorer (VIPER), a website for virus capsid structures and their computational analyses. *J. Virol.*, **75**, 11943–11947.

45. Pickett,B.E., Sadat,E.L., Zhang,Y. *et al.* (2011) ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.*, **40**, D593–D598.

46. Berman,H.M., Westbrook,J., Feng,Z. *et al.* (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

47. Pettersen,E.F., Goddard,T.D., Huang,C.C. *et al.* (2004) UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

48. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.

49. Božič,A.L., Šiber,A. and Podgornik,R. (2012) How simple can a model of an empty viral capsid be? Charge distributions in viral capsids. *J. Biol. Phys.*, **38**, 657–671.

50. May,E.R. (2014) Recent developments in molecular simulation approaches to study spherical virus capsids. *Mol. Simul.*, **40**, 878–888.

51. Tarasova,E., Farafonov,V., Taiji,M. *et al.* (2018) Details of charge distribution in stable viral capsid. *J. Mol. Liq.*, **265**, 585–591.