

Biologic constraints on modelling virus assembly

Robert L. Garcea*

Department of Pediatric Hematology/Oncology, University of Colorado School of Medicine,
Aurora, CO, USA

(Received 20 January 2008; final version received 29 April 2008)

The mathematic modelling of icosahedral virus assembly has drawn increasing interest because of the symmetric geometry of the outer shell structures. Many models involve equilibrium expressions of subunit binding, with reversible subunit additions forming various intermediate structures. The underlying assumption is that a final lowest energy state drives the equilibrium toward assembly. In their simplest forms, these models have explained why high subunit protein concentrations and strong subunit association constants can result in kinetic traps forming off pathway partial and aberrant structures. However, the cell biology of virus assembly is exceedingly complex. The biochemistry and biology of polyoma and papillomavirus assembly described here illustrates many of these specific issues. Variables include the use of cellular ‘chaperone’ proteins as mediators of assembly fidelity, the coupling of assembly to encapsidation of a specific nucleic acid genome, the use of cellular structures as ‘workbenches’ upon which assembly occurs, and the underlying problem of making a capsid structure that is metastable and capable of rapid disassembly upon infection. Although formidable to model, incorporating these considerations could advance the relevance of mathematical models of virus assembly to the real world.

Keywords: polyomavirus; papillomavirus; virus assembly; chaperone proteins

The mathematic modelling of virus assembly has drawn increasing interest over the past decade. In particular, icosahedral viruses have been attractive targets because of their symmetric geometry. Assembly models have ranged from ‘coarse’ to ‘fine grain’ approximations, using geometric shapes interacting by a fixed set of predetermined rules, to molecular dynamic calculations predicting the lowest energy state of interacting proteins. All of these models attempt to simplify assembly pathways in tractable ways. However, the *in vivo* biology of virus assembly reveals many variables as yet unaccounted for by these reductionist models that may confound their predictive value. Polyoma and papillomavirus capsid assembly highlight several of these possible complicating factors.

The polyoma and papillomaviruses were once grouped into the same virus family, the *Papovaviridae*, based upon their similar appearance and structure [12]. However, they have been separated recently into their own families (*Polyomaviridae*, *Papillomaviridae*) because of their distinctive genomic organizations. They are small (50–60 nm diameter), non-enveloped viruses with a double-stranded DNA genome (Figure 1). Their capsids (*i.e.* shells or coats) are comprised of 72 pentamers, termed capsomeres, of a major capsid protein (VP1 for polyoma, and L1 for papilloma) arranged on a $T = 7$ icosahedral lattice [1,2,20]. This geometry is non-quasi-equivalent, in that Caspar–Klug theory would predict 60 hexamers and 12 pentamers in such an icosahedron [6]. Instead, 60 pentamers are in ‘hexavalent’ positions surrounded by 6 pentamers,

*Email: bob.garcea@uchsc.edu

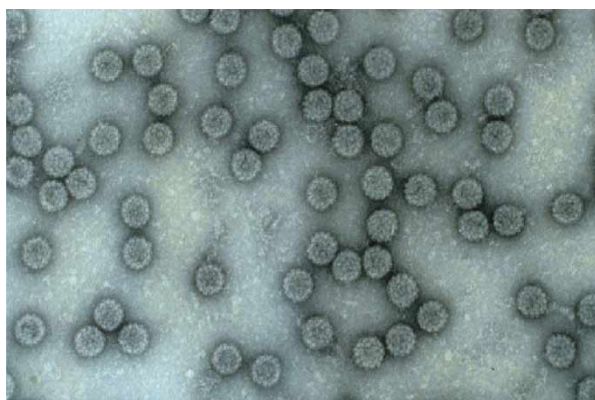


Figure 1. Electron micrograph of purified papillomavirus virions. These are non-enveloped (no lipid coat) viruses with a diameter of approximately 55 nm. The subunits, or capsomeres, can be seen. Available in colour online.

and 12 pentamers (at the icosahedral five-fold axes) are in pentavalent positions. The hexavalent pentamers thus must assume distinct bonding contacts with their different neighbouring pentamers, and require ‘bond switching’ as new pentamers enter the assembly process (Figure 2). Atomic structures have been determined for both papilloma and polyoma pentamers, and these structures define how this bond switching occurs [7,15,18]. Moreover, these structures support the general notion that inter-capsomere bonding involves ‘invasion’ of pentamers by carboxy-terminal arms of the neighbouring capsid proteins, and this invasion point is where bond switching occurs.

The papilloma/polyoma capsids have been particularly attractive for studying assembly not only because of their relative simplicity (*e.g.* no envelope, one major protein) and geometry but also the ability of their capsid proteins to self-assemble *in vitro* into capsid-like structures. VP1 or L1 proteins expressed recombinantly in bacteria yield pentameric capsomeres after purification [13,21]. These purified pentamers can be induced to assemble either by addition of calcium or high ionic strength (for VP1), or by oxidation of disulfide bonds (for L1) (Figure 3). This amazing property reflects the ‘intrinsic self-assembly potential’ of these proteins,

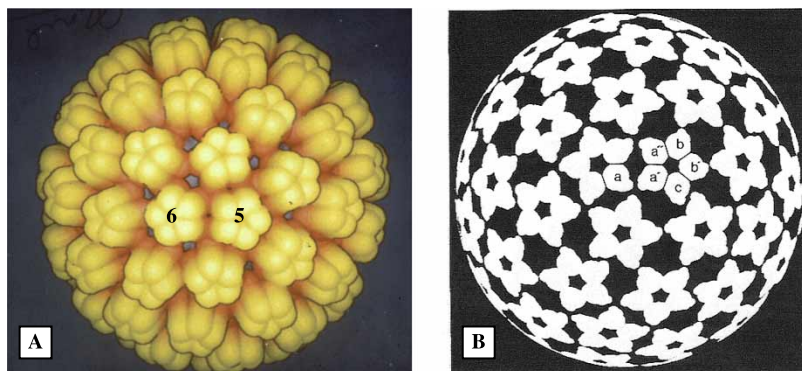


Figure 2. A computer model of the $T = 7$ capsid of polyoma and papillomaviruses. Seventy-two pentamers (capsomeres) of the major coat protein (either VP1 or L1) form the shell. Pentamers occupy both ‘hexavalent’ [6] and ‘pentavalent’ [5] positions on the lattice (panel A), giving rise to five different bonding interactions between protein monomers (panel B) (adapted from Ref. [21]). Available in colour online.

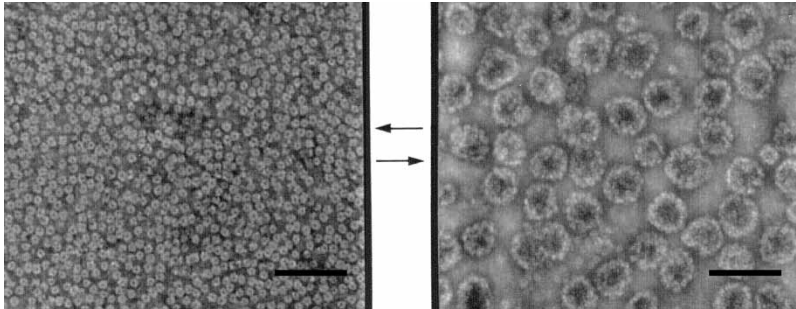


Figure 3. *In vitro* self-assembly of polyomavirus VP1 protein. The left panel is an electron micrograph of VP1 pentamers (capsomeres) purified after recombinant expression of the VP1 protein in *E. coli*. Upon addition of calcium to a solution of these capsomeres, virus-like particles (right panel) spontaneously assemble. Magnification is the same in both panels and fiducial bar is 110 nm (adapted from Ref. [21]).

undoubtedly conferred because they are evolutionarily designed to assemble quickly and efficiently during virus infection. The ability to induce capsid assembly *in vitro* has allowed both kinetic and biochemical study of the assembly process for these viruses, but also illustrates many of the complications in mathematical modelling.

One of the main observations for *in vitro* self-assembly is that even under optimized conditions the assembled capsid products, although grossly correct, are slightly irregular [22]. Indeed, the assembly buffer conditions can be modified to obtain a variety of polymorphic assemblies, from $T = 1$ (12 pentamer icosahedra) to ribbons to tubes. For example, the crystals of the HPV16 L1 protein used for atomic structure determination were formed from $T = 1$ assemblies of L1 capsomeres, because this polymorphic variant was produced almost exclusively under the specific crystallographic buffer conditions [7]. A challenge for mathematic modelling is predicting these off pathway assemblies.

Despite potential *in vivo* confounding factors, *in vitro* VP1 and L1 assembly have been used to test the 'local rules' model of capsid formation, as well as obtain experimentally derived kinetic values for rate constants that can be used to validate other models. Local rules utilizes the atomic structures of SV40 and murine polyomavirus to define six types of inter-pentamer bonding contacts that can be used when pentamers enter the assembly pathway [3]. By using only two of the rules, small $T = 1$ icosahedral structures can be modelled [24]. Other alterations of the rule set give rise to spiral structures that have been observed experimentally with different mutant forms of VP1. However, ribbon-like structures and octahedral VP1 assemblies are not easily explained by simple variations on the rule set, highlighting limitations of this simplified and static geometric model.

In order to derive experimental values for assembly rate constants and nucleation size, *in vitro* assembly of L1 analysed by multi-angle light scattering has been used [5]. Light scattering measures the average size of the entire protein population, and for $T = 7$ assembly the size of L1 particles approaches a plateau at 20 megaDa. L1 pentamers were found to assemble *in vitro* at a rate that is a function of the starting protein concentration, with an observed lag time. The nucleation structure was determined by relating the slope of the log of the capsid concentration (*i.e.* log of the extent of assembly) versus the log of the free capsomere concentration for single time points. This value was 2.02, suggesting that dimers of pentamers were the nucleus for assembly. Pentavalent pentamers have the most stable, energy favourable contacts with their five neighbours and thus the potential of such bonds was thought to be a nucleating species. However, this experimental finding is in contrast to the hypothesized

'5 around 1' nucleation structure proposed by protein energetic considerations [25]. Although subject to the caveats of 'in vivo' factors, in this case the data does not support this hypothesis based upon energetic speculation. Furthermore, elongation kinetics were determined to be second order, consistent with rapid sequential addition of single pentamers to the growing shell.

In contrast to the relative 'low fidelity' seen for capsid assembly *in vitro* from pentamer subunits, the recombinant expression of VP1 or L1 in eukaryotic cells (*e.g.* yeast or insect cells) yields quite uniform assembled capsids of the correct size [19] (Figure 4). Moreover, these capsids are found in the nucleus of the recombinant host cell (*i.e.* as a general rule DNA animal viruses assemble in the cell nucleus where their genomes are replicating), suggesting that specific intracellular conditions are required for high fidelity capsid assembly. The complete lack of capsids in the cytoplasm emphasizes the compartmentalization and special requirements of the assembly process *in vivo*. These assemblies have been termed 'virus-like particles', or VLPs, although the same nomenclature may also be used for *in vitro* assembled capsomeres that appear as $T = 7$ sized capsids.

The lack of capsid assembly in the cytoplasm presents an obvious biologic question – what is preventing assembly of these subunits that have intrinsic assembly potential? Tentative answers point to the role of cellular chaperone proteins, often termed heat-shock proteins because they are induced by stress [17]. Simplistically, chaperone proteins aid in the folding of disordered domains of proteins, using ATP to oscillate the protein through different conformational states until a lowest energy, 'correctly folded' final structure is attained. For example, heat-stressed cellular proteins may be partially denatured, and binding of heat-shock chaperones prevents them from further denaturation and aids in their proper re-folding. During infection of mouse cells polyomavirus VP1 associates with the cellular heat-shock protein 70 (hsc70) immediately after its translation, and hsc70 appears to accompany VP1 (as pentamers) into the nucleus [10]. Furthermore, the chaperone protein binds VP1 at its carboxy terminus, the domain where assembly through arm invasion is mediated, likely inhibiting assembly through steric interference

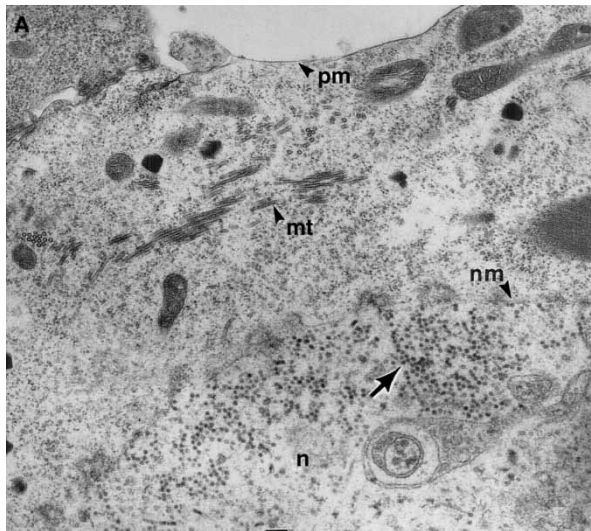


Figure 4. Electron micrograph of a eukaryotic cell expressing the VP1 protein. Pentamers of VP1 are transported into the nucleus of the cell where virus-like particles (arrow) spontaneously assemble. Plasma membrane (pm); microtubules (mt); nuclear membrane (nm); nucleus (n) (adapted from Ref. [19]).

[8]. Thus, one control on the intracellular assembly of VP1 may be hsc70 association preventing assembly until the capsomere reaches the nucleus. The role of chaperone proteins in L1 and VP1 assembly has been explored *in vitro* using purified protein components. The *Escherichia coli* chaperone proteins DnaK and DnaJ (analogous to the hsc70 and hsc 40 proteins in eukaryotic cells) have been shown to assemble VP1 and L1 into capsids *in vitro* under otherwise non-assembling conditions [8]. This assembly requires ATP, and yields capsids of greater fidelity (size, morphology) than seen by buffer-induced assembly triggers (Figure 5). Thus, cellular chaperone proteins may serve not only to prevent premature assembly in the cytoplasm but also serve a catalytic and/or proofreading function, which may be possible to model mathematically.

As yet, mathematical models of virus assembly have been confined to capsid formation, and have not considered the contribution of the genome being packaged. There are two basic strategies for genome encapsidation: assemble capsid proteins around the genome or 'inject' the genome onto a pre-assembled capsid structure. In general, larger viruses have chosen the latter pathway while polyoma and papilloma appear to assemble their capsid proteins onto the genome. In addition, a significant tactical problem for encapsidation is identifying the correct nucleic acid (the virus) for packaging among a large excess of cellular targets. For most viruses, both the genomic recognition signal and the triggering event for encapsidation are unknown. A reasonable assumption is that a viral genome specific binding protein rather than a host cell protein serves to label the genome for encapsidation. For polyoma, this signal may be the virally encoded large T-antigen, a multifunctional protein that binds specifically to the genome at the origin of viral DNA replication and regulates both viral transcription and replication [26]. T-antigen has a prototypic 'J-domain' at its amino terminus, which recognizes the cell hsc70 protein. Thus, incoming VP1 pentamers bound by hsc70 could be specifically targeted to the viral genome by T-antigen and the hsc70/J interaction might then initiate encapsidation.

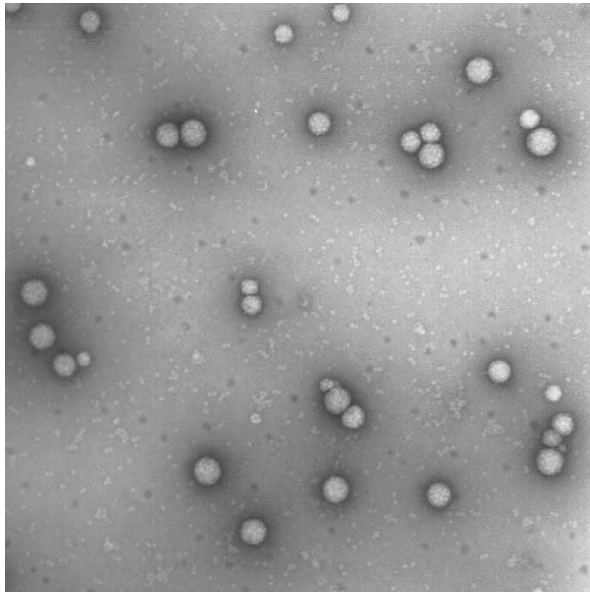


Figure 5. Electron micrograph of virus-like particles of the polyomavirus VP1 protein formed by the *in vitro* assembly using chaperone proteins. These particles are more regular than those formed by chemical manipulation of assembly buffer (as seen in Figure 3 (adapted from Ref. [8])).

Importantly for modelling is that the nucleic acid provides an overall stability to the capsid and may provide a thermodynamically irreversible step in the assembly process. For example, papilloma capsids by themselves are readily dissociated by disulfide bond reducing agents into their component capsomere subunits. However, when papilloma virions, with genomic DNA inside, are treated with reducing agents the capsid shell expands (by $\approx 10\%$ in diameter) but remains intact and upon oxidation the shell returns to its former state [14].

In contrast to *in vitro* assembly reactions involving recombinant capsid proteins, virus assembly *in vivo* undoubtedly does not occur in solution. Although the specifics have not been delineated, the emerging concept is that assembly in the cell occurs within virus 'factories' [28]. These subcellular domains are specific gathering points for genome replication and capsid protein delivery. For polyoma and papilloma viruses, these factories are likely a subset of 'nuclear dots', operationally defined as spots in the nucleus seen when viral proteins or DNA are localized by fluorescent-labelled antibodies. Although several variants of nuclear dots have been described, the nuclear domain 10 (ND 10) or promyelocytic leukemia protein bodies seem to be associated with polyoma/papilloma assembly and replication processes [4,11,16,27]. These factories might be conceptualized as assembly lines or workbenches, where viral replication is tethered to a cellular protein matrix juxtaposed with capsid proteins that are delivered/transported specifically to these locations, thus controlling local subunit concentrations. Such workbenches could have a significant kinetic and thermodynamic impact on capsid assembly.

A paradox of virus structures is that they must be exceptionally stable in the environment but then readily dissociate upon infection and cell entry. Virions have thus been termed 'metastable' intermediates, poised at an energy minimum with conformational triggers that allow the structure to jump an energy barrier to complete disassembly. For polyoma capsomeres, cellular chaperone proteins appear to facilitate proper high fidelity assembly *in vivo*. Upon virus entry during infection, these same virions are also susceptible to chaperone disassembly [9,23]. The 'ying/yang' of this equilibrium must therefore be controlled by additional, as yet unknown, cellular factors that have such a duality because they are compartmentalized (*e.g.* endoplasmic reticulum for disassembly of entering virions; nucleus for assembly of virions at the end of infection) and thus the kinetic/thermodynamic pathways are isolated and distinct. Such compartmentalization is not a feature of current mathematical models, and the disassembly of viruses may be as relevant as the assembly pathway in understanding the energy of the virion.

Summary

Virus assembly is a complex biological process that is still poorly understood. Treating assembly as a soluble, thermodynamically driven chemical reaction is an extreme simplification. The beauty of the assembly process derives from Nature's ingenious tricks that ensure efficiency and fidelity in the production of virions. These tricks include the use of chaperone proteins and cellular workbenches. Furthermore, the assumption that the final structure is the thermodynamic energy minimum may be only partially correct, since virions are metastable and must disassemble rapidly. Incorporation of these considerations into new models may yield more robust predictability, and potentially reveal new biologic paradigms.

Acknowledgements

This work was supported by Public Health Service grant CA37667 to RLG from the NIH/NCI.

References

- [1] T.S. Baker, J. Drak, and M. Bina, *The capsid of small papova viruses contains 72 pentameric capsomeres: Direct evidence from cryo-electron-microscopy of simian virus 40*, *Biophys. J.* 55 (1989), pp. 243–253.
- [2] T.S. Baker, W.W. Newcomb, N.H. Olson, L.M. Cowser, C. Olson, and J.C. Brown, *Structures of bovine and human papillomaviruses: Analysis by cryoelectron microscopy and three-dimensional image reconstruction*, *Biophys. J.* 60 (1991), pp. 1445–1456.
- [3] B. Berger, P.W. Shor, L. Tucker-Kellogg, and J. King, *Local rule-based theory of virus shell assembly*, *Proc. Natl Acad. Sci. (USA)* 91 (1994), pp. 7732–7736.
- [4] R. Bernardi and P.P. Pandolfi, *Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies*, *Nat. Rev. Mol. Cell Biol.* 8 (2007), pp. 1006–1016.
- [5] G.L. Casini, D. Graham, D. Heine, R.L. Garcea, and D.T. Wu, *In vitro papillomavirus capsid assembly analyzed by light scattering*, *Virology* 325 (2004), pp. 320–327.
- [6] D.L.D. Caspar and A. Klug, *Physical principles in the construction of regular viruses*, *Cold Spring Harbor Sym. Quant. Biol.* 27 (1962), pp. 1–24.
- [7] X.S. Chen, R.L. Garcea, I. Goldberg, G. Casini, and S.C. Harrison, *Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16*, *Mol. Cell* 5 (2000), pp. 557–567.
- [8] L.R. Chromy, J.M. Pipas, and R.L. Garcea, *Chaperone-mediated in vitro assembly of polyomavirus capsids*, *Proc. Natl Acad. Sci.* 100 (2003), pp. 10477–10482.
- [9] L.R. Chromy, A. Oltman, P.A. Estes, and R.L. Garcea, *Chaperone-mediated in vitro disassembly of polyoma- and papillomaviruses*, *J. Virol.* 80 (2006), pp. 5086–5091.
- [10] T.P. Cripe, S.E. Delos, P.A. Estes, and R.L. Garcea, *In vivo and in vitro association of hsc70 with the polyomavirus capsid proteins*, *J. Virol.* 69 (1995), pp. 7807–7813.
- [11] P.M. Day, C.C. Baker, D.R. Lowy, and J.T. Schiller, *Establishment of papillomavirus infection is enhanced by promyelocytic leukemia protein (PML) expression*, *Proc. Natl Acad. Sci. USA* 101 (2004), pp. 14252–14257.
- [12] A.J. Klug, *Structure of viruses of the papilloma-polyoma type II. Comments on other work*, *J. Mol. Biol.* 11 (1965), pp. 424–431.
- [13] M. Li, T.P. Cripe, P.A. Estes, M.K. Lyon, R.C. Rose, and R.L. Garcea, *Expression of the human papillomavirus type-11 L1 capsid protein in Escherichia coli: Characterization of protein domains involved in DNA-binding and capsid assembly*, *J. Virol.* 71 (1997), pp. 2988–2995.
- [14] M. Li, P. Beard, P.A. Estes, M.K. Lyon, and R.L. Garcea, *Intercapsomeric disulfide bonds in papillomavirus assembly and disassembly*, *J. Virol.* 72 (1998), pp. 2160–2167.
- [15] R.C. Liddington, Y. Yan, J. Moulai, R. Sahli, T.L. Benjamin, and S.C. Harrison, *Structure of simian virus 40 at 3.8-Å resolution*, *Nature* 354 (1991), pp. 278–284.
- [16] G.G. Maul, *Nuclear domain 10, the site of DNA virus transcription and replication*, *BioEssays* 20 (1998), pp. 660–667.
- [17] M.P. Mayer, *Recruitment of hsp70 chaperones: A crucial part of viral survival strategies*, *Rev. Physiol. Biochem. Pharmacol.* 153 (2005), pp. 1–46.
- [18] Y. Modis, B.L. Trus, and S.C. Harrison, *Atomic model of the papillomavirus capsid*, *EMBO J.* 21 (2002), pp. 4754–4762.
- [19] L. Montross, S. Watkins, R.B. Moreland, H. Mamon, D.L.D. Caspar, and R.L. Garcea, *Nuclear assembly of polyomavirus capsids in insect cells expressing the major capsid protein VP1*, *J. Virol.* 65 (1991), pp. 4991–4998.
- [20] I. Rayment, T.S. Baker, D.L.D. Caspar, and W.T. Murakami, *Polyoma virus capsid structure at 22.5 Å resolution*, *Nature (London)* 295 (1982), pp. 110–115.
- [21] D.M. Salunke, D.L.D. Caspar, and R.L. Garcea, *Self-assembly of purified polyomavirus capsid protein VP1*, *Cell* 46 (1986), pp. 895–904.
- [22] ———, *Polymorphism in the assembly of polyomavirus capsid protein VP1*, *Biophys. J.* 56 (1989), pp. 887–900.
- [23] M. Schelhaas, J. Malmström, L. Pelkmans, J. Haugstetter, L. Ellgaard, K. Grünwald, and A. Helenius, *Simian virus 40 depends on ER protein folding and quality control factors for entry into host cells*, *Cell* 131 (2007), pp. 516–529.
- [24] R. Schwartz, R.L. Garcea, and B. Berger, *“Local rules” theory applied to polyomavirus polymorphic capsid assemblies*, *Virology* 268 (2000), pp. 461–470.
- [25] T. Stehle, S.J. Gamblin, Y. Yan, and S.C. Harrison, *The structure of simian virus 40 refined at 3.1 Å resolution*, *Structure* 4 (1996), pp. 165–182.

- [26] C.S. Sullivan and J.M. Pipas, *T antigens of simian virus 40: Molecular chaperones for viral replication and tumorigenesis*, *Microbiol. Mol. Biol. Rev.* 66 (2002), pp. 179–202.
- [27] Q. Tang, P. Bell, P. Tegtmeyer, and G.G. Maul, *Replication but not transcription of simian virus 40 DNA is dependent on nuclear domain 10*, *J. Virol.* 74 (2000), pp. 9694–9700.
- [28] T. Wileman, *Aggresomes and pericentriolar sites of virus assembly: Cellular defense or viral design?*, *Annu. Rev. Microbiol.* 61 (2007), pp. 149–167.