Geometry of the Structure of Viruses

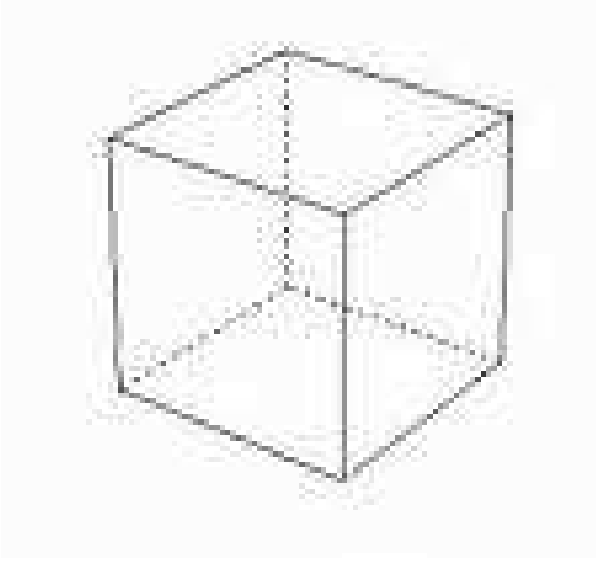
Michel Deza

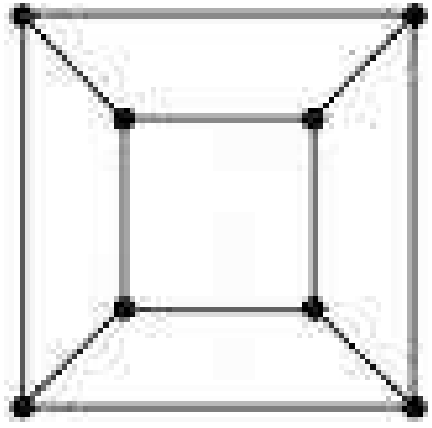
Ecole Normale Superieure, Paris, and JAIST, Ishikawa

I. Polyhedra and fullerenes

Polyhedra and their faces

- A polyhedron P is the convex hull of a finite set $X \subset \mathbb{R}^3$.
- An i-face of P is the i-dimensional ($1 \leq i \leq 3$) set $\{x \in P : f(x) = 0\}$ for a linear function $f \geq 0$ on P.
  A 0-, 1-, 2-face is called vertex, edge, face, respectively; their sets are $V(P)$, $E(P)$, $F(P)$, respectively.
- (Poincaré) dual polyhedra P, P^* on sphere: bijection of $V(P)$ with $F(P^*)$, $F(P)$ with $V(P^*)$ and $E(P)$ with $E(P^*)$.





Polyhedra: their skeletons and groups

- The skeleton of polyhedron P is the graph $G(P)$=(V, E).
- Steinitz, circa 1927: a graph is the skeleton of a polyhedron if and only if it is *planar* and 3-*connected*, i.e. removing any two edges keep it connected.
- Point group $Isom(P) \subset Aut(G(P))$, combinatorial group Mani, 1971: for any planar 3-connected graph Γ, there is a polyhedron P with $G(P) \simeq \Gamma$ and $Isom(P) \simeq Aut(G(P))$.
- All finite groups of isometries of $\mathbb{R}^3$ are known. In Schoenflies notations, they are: C_1, C_s, C_i, C_m, C_{mv}, C_{mh}, S_{2m}, D_m, D_{md}, D_{mh}, O_h, O, T_d, T_h, T, $I \simeq Alt_5$, the rotations group of regular Dodecahedron, and $I_h \simeq I \times C_2$ (or H_3), its isometries group, called proper and extended icosahedral group, respectively.

Definition of fullerene

A fullerene F_n is a simple (i.e., 3-valent) n-vertex polyhedron with 12 5-gonal and $(\frac{n}{2} - 10)$ 6-gonal faces.

- F_n exist for all even $n \geq 20$ except $n = 22$.
- $1, 1, 1, 2, 5 \ldots, 1812, \ldots 214127713, \ldots$ isomers F_n, for $n = 20, 24, 26, 28, 30 \ldots, 60, \ldots, 200, \ldots$.
- Thurston,1998, implies: number of F_n grows as n^9.
- IP, i.e. with isolated pentagones, F_n are denoted by C_n.
- $C_{60}(I_h)$, $C_{80}(I_h)$ are only icosahedral (i.e., with highest symmetry I_h or I) fullerenes with $n \leq 80$ vertices.

What nature wants?

Fullerenes or their duals are ubiquitous esp. in nanoworld:

- Biology: virus capsids and clathrine coated vesicles,
- Organic (i.e., carbon) Chemistry, even Architecture,
- also: (energy) minimizers in Thomson problem (for n unit charged particles on sphere) and Skyrme problem (for given baryonic number of nucleons); maximizers, in Tammes problem, of minimum distance between n points on sphere.

Which, among simple polyhedra with given number of faces, are the “best” approximation of sphere?

Conjecture: FULLERENES

Isoperimetric problem for polyhedra

Lhuilier 1782, Steiner 1842, Lindelöf 1869, Steinitz 1927, Goldberg 1933, Fejes Tóth 1948, Pólya 1954

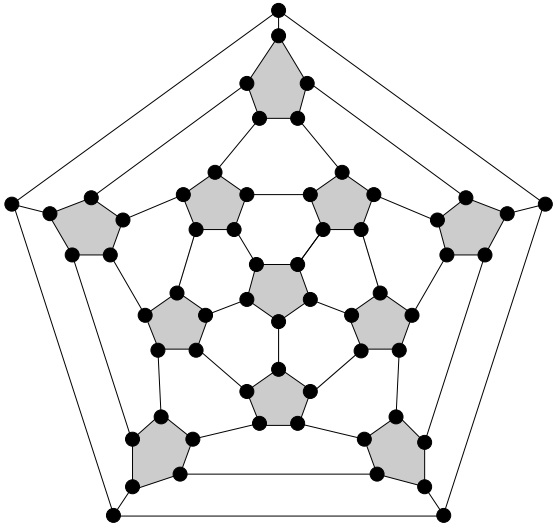
- For a polyhedron P with m faces, maximizing its volume V for given surface S is equivalent to minimizing V if P is circumscribed around the unit sphere.
- Schwarz,1890: for IQ (Isoperimetric Quotient) of a solid, it holds $IQ = 36\pi\frac{V^2}{S^3} \leq 1$ (with equality only for sphere).
- Goldberg, 1933, conjectured: polyhedra with $m \geq 12$ faces having maximal IQ are fullerenes. For $m \leq 12$ (i.e., $4, \ldots, 10$ and 12), it is duals of 8 convex deltahedra.
- In Biology: ratio $\frac{V}{S}$ (=$\frac{r}{3}$ for spherical animal of radius r) affects heat gain/loss, nutritient/gas transport into body cells and organism support on its legs.

II. Icosahedral fullerenes and their duals

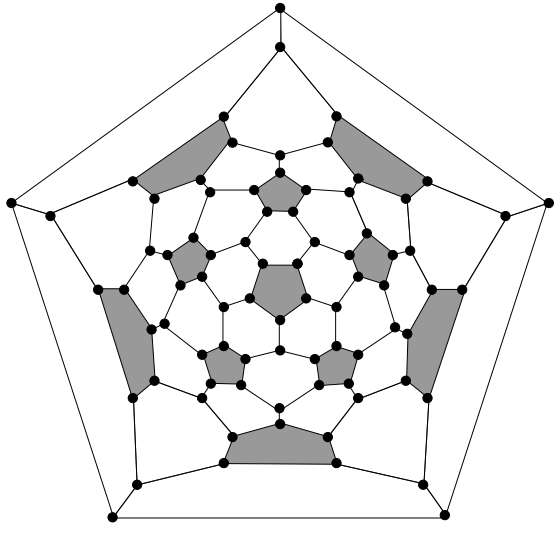
Icosahedral fullerenes

Call icosahedral any fullerene with symmetry I_h or I.

- n=$20T$ for T=a^2+ab+b^2 (triangulation number), $0 \leq b \leq a$.
- I for $0 < b < a$ and I_h for $a = b \neq 0$ or $b = 0$.
- Dodecahedron $F_{20}(I_h)$: smallest ((a,b)=$(1,0)$, T=1) and unique non-IP (with adjacent 5-gons) icosahedral one.



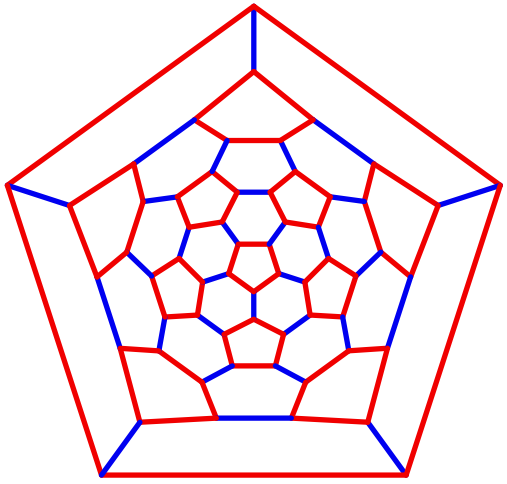
$C_{60}(I_h)$=$(1,1)$-dodecahedron
truncated icosahedron



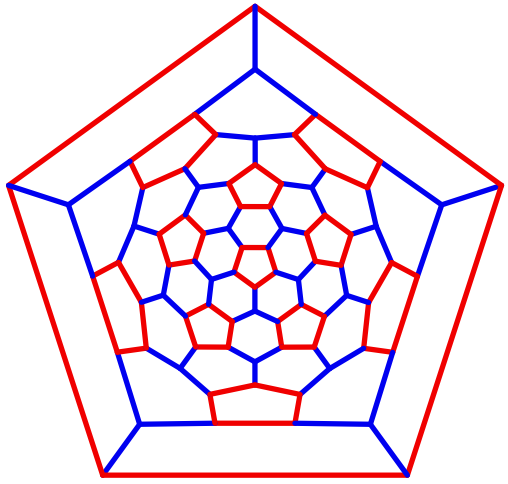
$C_{80}(I_h)$=$(2,0)$-dodecahedron
chamfered dodecahedron

Small examples

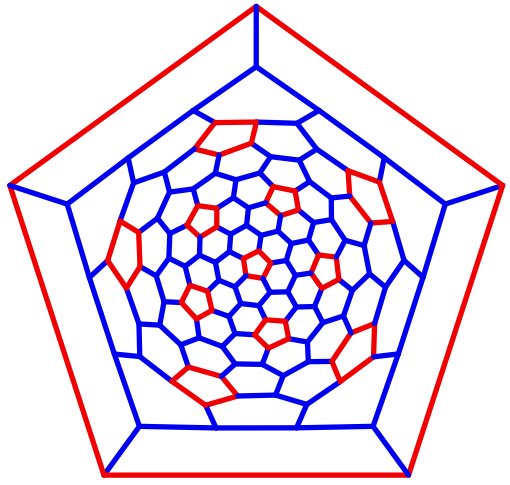
Besides $F_{20}(I_h)$ with $T = 1$, the next smallest examples are:



$C_{60}(I_h)$: $(1,1)$-dodecahedron, T=3



$C_{80}(I_h)$: $(2,0)$-dodecahedron, T=4

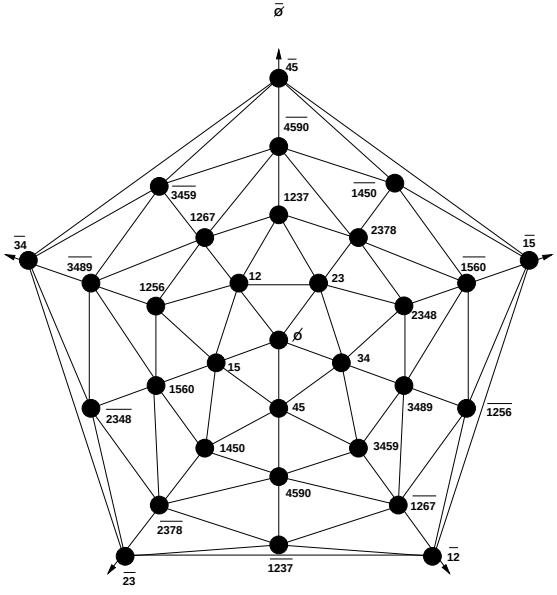


$C_{140}(I)$: $(2,1)$-dodecahedron, T=7 (*laevo*)

Icosadeltahedra

Icosadeltahedron C^*_{20T}: dual of an icosahedral fullerene.

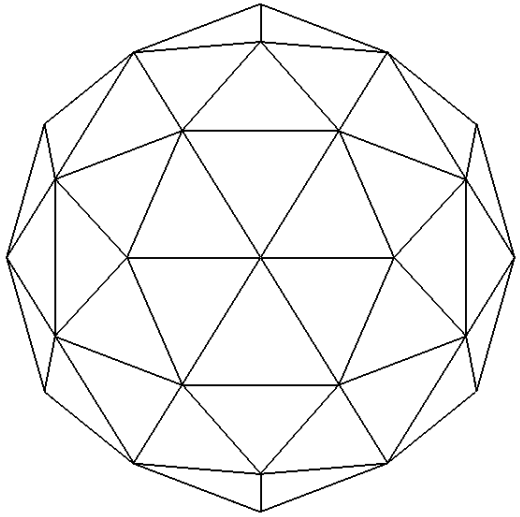
- *Geodesic domes*: Fuller, patent 1954
- Capsids of viruses: Caspar and Klug, Nobel prize 1982
- *Carbon* $C_{60}(I_h)$: Kroto-Curl-Smalley, Nobel prize 1996



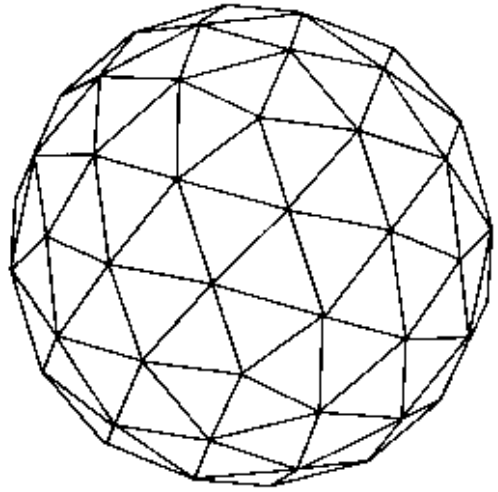
$C^*_{60}(I_h)$, $(a,b)=(1,1)$
pentakis-dodecahedron

GRAVIATION (Esher, 1952)
omnicapped dodecahedron

Small icosadeltahedra with a=2



$C^*_{80}(I_h)$, (a,b)=$(2,0)$



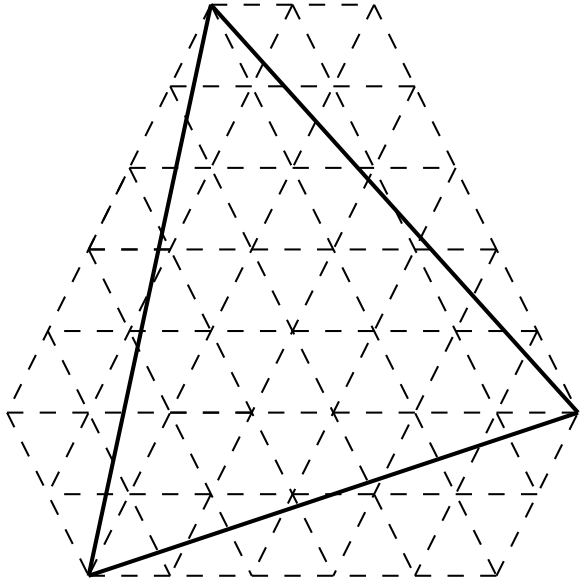
$C^*_{140}(I)$, (a,b)=$(2,1)$

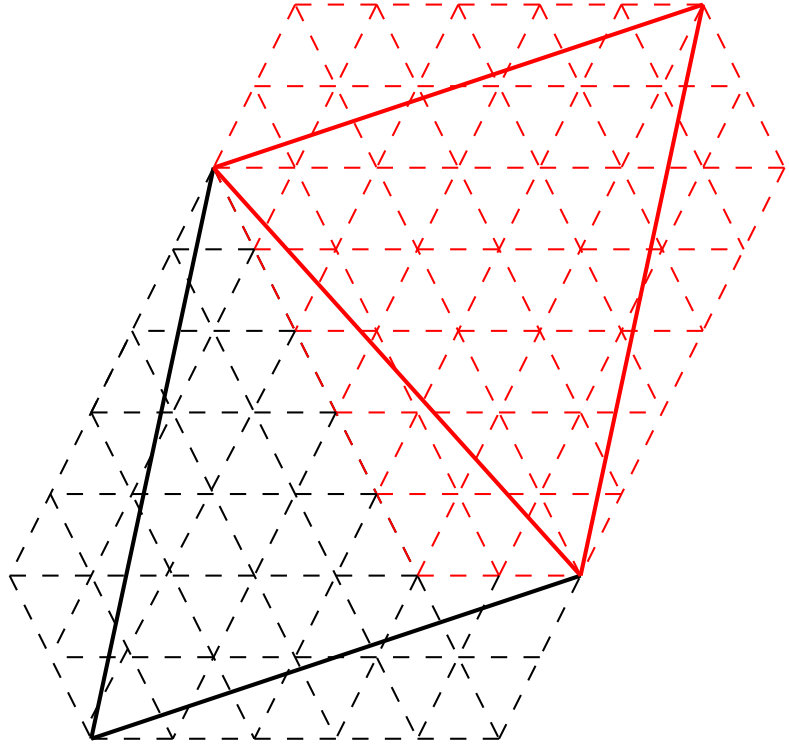
(a,b) (or (h,k) as Caspar-Klug) are the numbers of steps, in 2 directions, on the shortest way in the graph of C^*_n between two closest 5-valent vertices. Their distance $a+b$ is Fullers's frequency, while $T=a^2+ab+b^2$ is triangulation number.

In general, T not define (a,b) but T=$a^2, 3a^2$ imply $(a,0),(a,a)$. Caspar-Klug: classes P=1, P=3; Fuller: Alternate, Triacon.

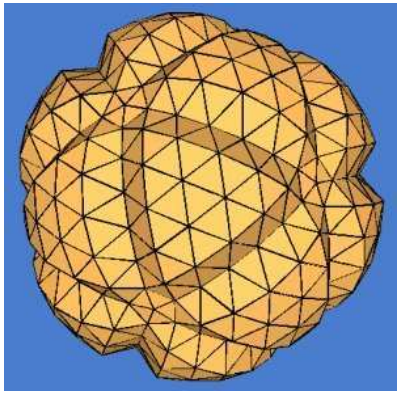
Goldberg-Coxeter construction

Given (a, b) ($(5, 2)$ below), put latiice triangle on $p6$ *net* $\{3^6\}$. Gluing pieces coherantly gives other triangulation of plane.

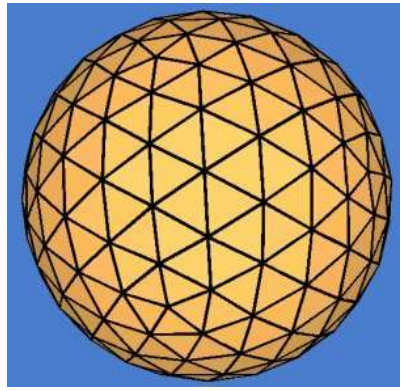




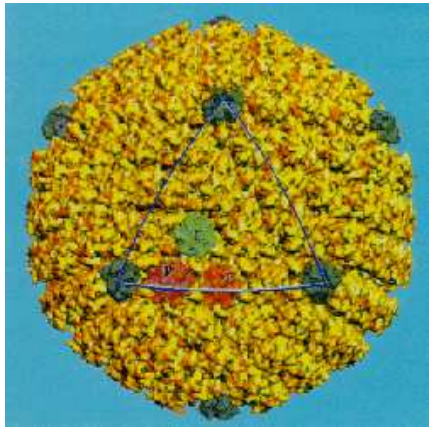
Icosadeltahedra with $T = a^2$



$(7, 0)$



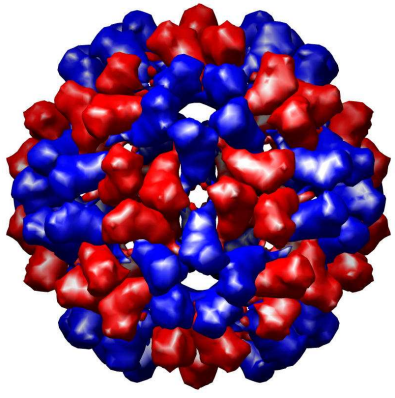
$(5, 0)$

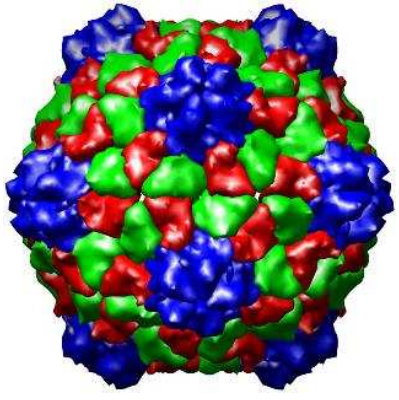


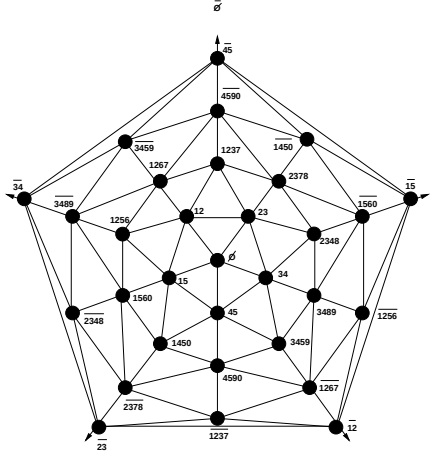
$(4, 0)$, herpes

Cowpea mosaic virus CPCM: $T = 3$

Plant *comovirus* infecting cowpea leafs; high yields 1-2 g/kg







$C_{60}^{*}(I_h)$, $(a, b) = (1, 1), T = 3$
pentakis-dodecahedron

Icosadeltahedra in Architecture

(a,b)	Fullerene	Geodesic dome
$(1,0)$	$F^*_{20}(I_h)$	One of Salvador Dali houses
$(1,1)$	$C^*_{60}(I_h)$	Artic Institute, Baffin Island
$(3,0)$	$C^*_{180}(I_h)$	Bachelor officers quarters, US Air Force, Korea
$(2,2)$	$C^*_{240}(I_h)$	U.S.S. Leyte
$(4,0)$	$C^*_{320}(I_h)$	Geodesic Sphere, Mt Washington, New Hampshire
$(5,0)$	$C^*_{500}(I_h)$	US pavilion, Kabul Afghanistan
$(6,0)$	$C^*_{720}(I_h)$	Radome, Artic dEW
$(8,8)$	$C^*_{3840}(I_h)$	Lawrence, Long Island
$(16,0)$	$C^*_{5120}(I_h)$	US pavilion, Expo 67, Montreal
$(18,0)$	$C^*_{6480}(I_h)$	Géode du Musée des Sciences, La Villete, Paris
$(18,0)$	$C^*_{6480}(I_h)$	Union Tank Car, Baton Rouge, Louisiana

$b = 0$ Alternate, $b = a$ Triacon and $a + b$ Frequency (distance of two 5-valent neighbors) are Buckminster Fullers's terms.

Geodesic Domes

US pavilion, World Expo 1967, Montreal

Spaceship Earth, Disney World's Epcot, Florida

In fact, the same structure of icosadeltahedron is adopted in the architecture of (virion capsid of) viruses.

III. Generalities

on viruses

Life fractions

- life: DNA and RNA (cells)
- $1/2$-life: DNA or RNA (cell parasites: viruses)
- **?** Subviral: satellite viruses, no protein RNA (viroids) or DNA: plasmids (extra-chromosomal replicable DNA), nanobiotech, and "junk" (i.e., non-coding) DNA.
- **??** Subviral (no DNA/RNA) self-replicating protein: prion

	Atom	DNA	Cryo-EM	Prion	Virus capsides
size	0.2-0.3	$\simeq 2$	$\simeq 5$	11	$20 - 50 - 100 - 400$
nm					SV40, HIV, Mimi

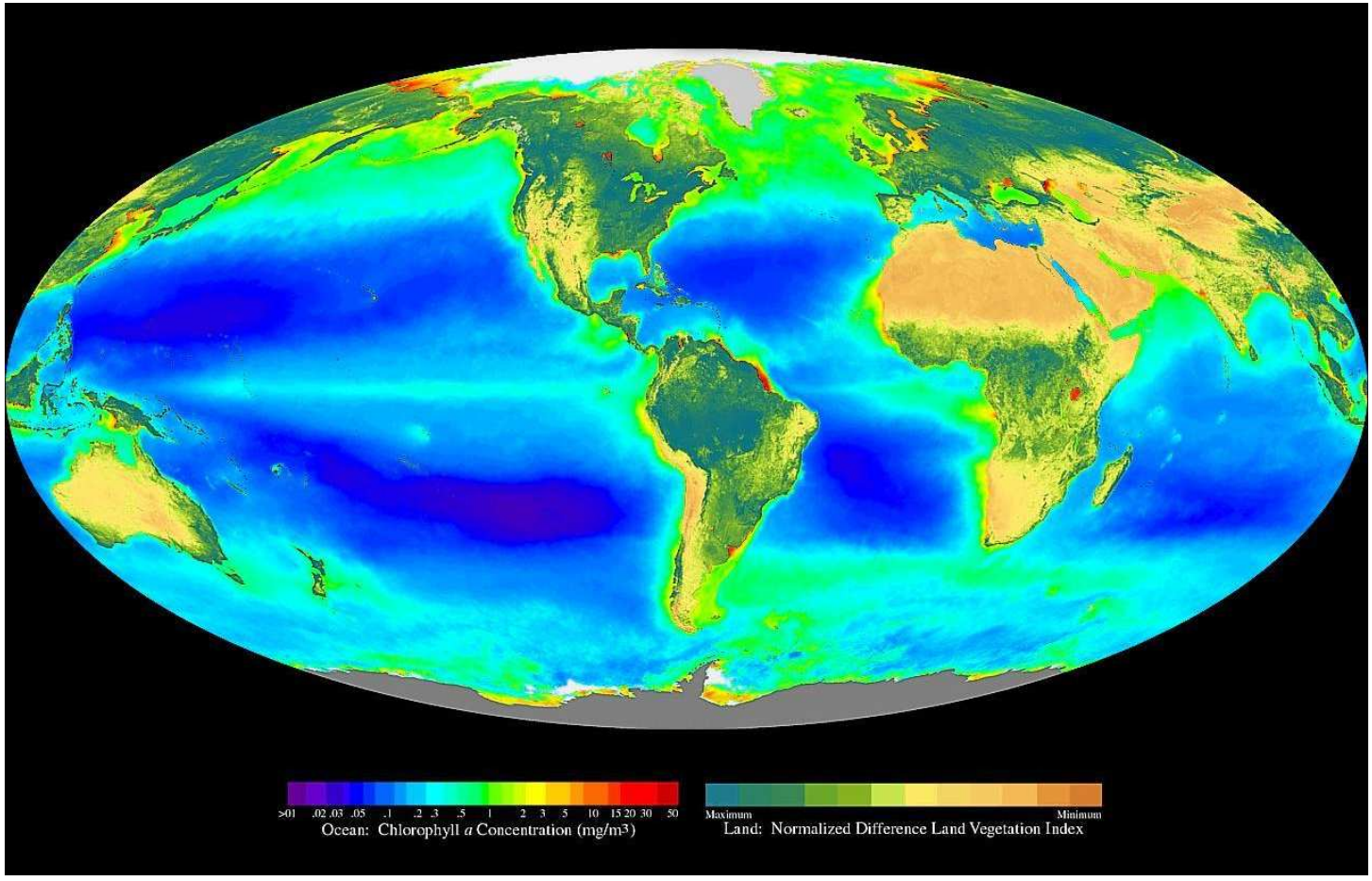
- Virus: from Latin *poison*. It has 2 stages: free (dormant transmissible particulate virion released by infected cell) and, if lucky, intracellular, active infectious stage. Bottle in ocean (extracellurar space), then packet of bad news.

Viruses in numbers

- $\approx 2,000$ virus species (but $\approx 30,000$ strains) are known.
- Diameter: 18 nm (Porcine circovirus PCV2) - 600 nm (Mimivirus). Length of helical ones: up to $3,000$ nm.
- One cm^3 in some ecosystems contains 10^8-10^{10} viruses. HIV-infected person releases > 10^{10} viruses daily.
- In water, phages float free and are most of the biomass: $\simeq \frac{1}{4}$ Gt=2.5×10^8 tons (4×10^{30} viruses at 0.2 fg of carbon and 100 nm each). Cf. human biomass: $\simeq 0.4$ gigaton.
- Biomass of virosphere exceeds those of all eukaryotes.
- Estimated global population: $\approx 10^{31}$ individuals of ≈ 100 million types. If stacked end to end, they would span the distance 10^{13} AU (Earth-Sun), i.e. 160 million light years.

Global primary biosphere 1997-2000

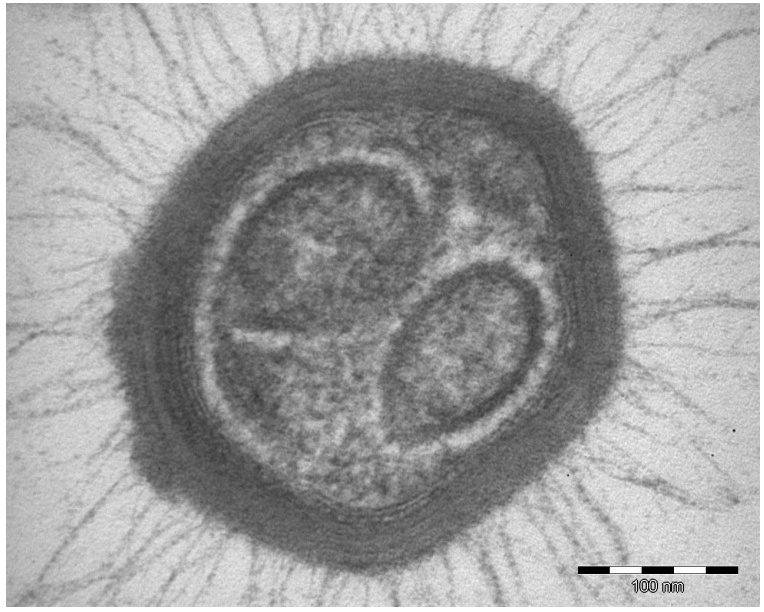
SeaWiFS Project, NASA, sattelite-estimated terrestrial vegetation and sea-surface chlorophyll indicating their primary production (photosynthesis of organic compounds from atmospheric/aquatic CO_2): 56.4 and 48.5 Gt C yearly.

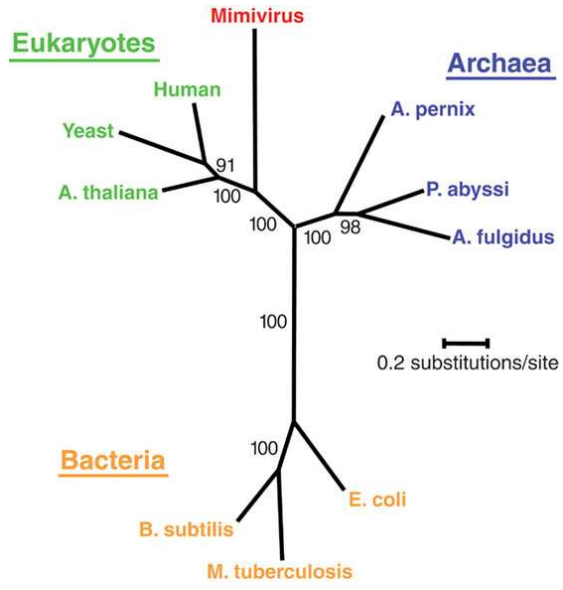


Viruses in numbers: genome

- About 1000 viral genomes are sequenced by now.
- The range of their genome size (DNA/RNA content) is 3.2 kb DNA (3.5 kb RNA) - 1.1814 Mb (Mimivirus). Also, 3-911 protein-coding genes and 4-200 proteins. Min. genome: $\simeq 300$ genes for 35-40 proteins. LUCA?
- Cf. human (3.3 Gb, $\approx 20,500$ genes) and highest: 670 Gb (Amoeba Dubia), 98000 genes (Trichomonas Vaginalis).
- By far most (esp. plant) viruses have RNA. All archea viruses and most (bacteria)phages have DNA. Most DNA (RNA) viruses replicate in nucleus (cytoplasm).
- 80% of viral genes not appear in another virus or a cell, their function is unknown. Esp. diverse (in morphology, DNA, proteins) are archeal ones: no homologs for 90% of genes of viruses isolated from boiling acid water.

Mimivirus and smallest non-viruses





Largest capsid: 400 nm, $\frac{1}{30}$ of its host amoeba, record is $\frac{1}{10}$.
Largest genome: ≈ 1.2 Mb and 911 protein-coding genes.

Cf. minimal free-living (bacterium *Mycoplasma genitalium*): 150 nm, 0.583 Mb, 521 genes. Conj. minimum: 256 genes.

Smallest endosymbionts are: eukaryot *Guillardia* θ 0.55 Mb, *Nanoarchaeum equitans* 0.49 Mb and the record: bacterium *Carsonella ruddii* with genome 0.16 Mb and 182 genes.

Mimivirus as infected girus

- Mimivirus is conjectured to be icosahedral, $T \approx 1180$.
- 2008: Mamavirus, even larger strain of Mimivirus, has a parasite ($\frac{1}{100}$ of its host) - first virophage (harmful satellite virus) Sputnik. It is also icosahedral.
  3 of its 21 genes are from Mimivirus; is it agent of lateral gene transfer between giant viruses?
  1 gene is a protein homologue of an archaeal virus; others 17, of bacteriophages and eukaryotic viruses.
- Other virophages are expected on giant viruses.
- Giruses (giant viruses) are viruses with size ≥ 250 nm and ≥ 300 genes. They have DNA and form large part of DNA virus population in marine environments.
  Are they ex-free-living as obligate parasitic/symbiotic bacteria, while small viruses are cell products?

Example: HIV-1 dynamics *in vivo*

- Viral genome load per ml of plasma is measured for virion RNA and viral DNA: pre-integrative, *proviral* (integrated into host cell DNA) and expressed one.
- Provirus infection is *latent* (replicated along with host's genome) or *productive* (transcribed into messenger RNA producing new virions infecting other cells).
- Mean life-span of cell-free virus, virus-producing and latently infected cells is 8 h, $16-48$ h, $4.6-44$ months.
- ≈ 140 *viral replication cycles* (from virion release till its "progeny") occur yearly. HIV-1 genome size is 10^4 bp; its mutation rate is 3.4×10^{-5} per bp per replication cycle. So, mutations occur often at each genome position daily
- Average infected person releases $> 10^{10}$ virions daily; most by productively infected $CD4^+$ T lymphocytes. No treatment give $<10^6$: latent reservoir in drug-free tissues

Viral quasispecies: multilevel selection

- RNA virus genome has $(3-30)\times 10^3$ nucleotid bits and high mutation rate $10^{-3}-10^{-5}$ substitutions per bit, i.e., ≥ 1 per replication cycle. Newly arising mutants form a quasispecies, i.e., a swarm, cloud of related genomes. Eigen-Shuster, 1977, introduced it to model biogenesis.
- Instead of single genotype fitness, here selection acts on their swarm: its genotypes tend to same fitness, to adapt better to changes (as sex preserves diversity). Are they also complementary, to colonize more niches?
- Diversity (genome divergence rate) of RNA viruses as HIV reach 15% per year in long-term survivors, or 5.5 mutations per essential genome (for proteins whose failure is deadly) per genomic replication cycle. "Speed limit" (evolutionary error threshold): 6 such mutations. Pleiotropy $\rightarrow$ complex organisms are 1000 times slower

Virus evolution

- There is more genetic diversity among viruses (since their genome duplicated in short time) than in all the rest of the Animal, Plant and Bacterial kingdoms, all of whose genomes consist of double-stranded DNA.
- Main mechanisms of viral genome change are: antigenic drift: small, gradual change by accumulating random point mutations of individual nucleotides, and antigenic shift: major genome change by recombination (join of broken DNA strand to the end of other molecule DNA) or reassortment (called *"viral sex"*): similar RNA viruses with *segmented* (into up to 12 parts in capsid) genomes can produce offspring by shuffling of genes.
- Antigenic drift can lead to resistance to antiviral drugs. Antigenic shift leads usually to a pandemic.

IV. Impact of viruses

Virus-host coevolution

- Myxoma was introduced to Australia in 1950 against rabbit plague. Virus was 99.8% lethal, in 6-10 days.
- But from 1953 it kill in 3-4 weeks, so that sick rabbits could be mosquito-bitten 3-5 times more. Now, after 7 epizootics, lethality during epidemics fell below 50% (attenuation of the virus and inherited immunity).
- Group selection for transmissibility let emerge/dominate strains of lowered virulence. Tick-transmitted viruses outcompete them but ticks bite also dead rabbits.
- Virus strains compete for resources, i.e., the host. There is a trade-off between strain traits: transmission rate, duration of infectiousness and case mortality.
- Problemes: Are viruses evolve toward benignity (not eradicate their hosts)? What is *virulence*: rate of virus replication (transmission) or host death rate (time to it)? What is virus-host mutual impact on genetic diversity?

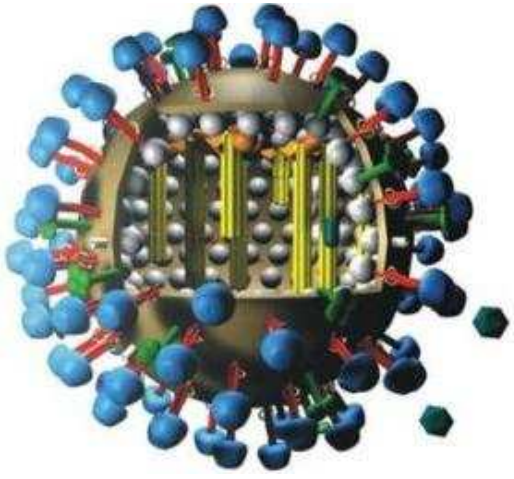
Influenza A quasispecies

- Main example of quasispecies created by reassortment: swarms of related influenza viruses genomes. They consist of 8 RNA segments acting as “chromosomes”, and each flu virus assembly requires one copy of each. Reassortment between an avian and a human virus caused H2N2, 1957, and H3N2, 1968, pandemic flu. Antigenic drift in gene H caused 1962, 1964 epidemics.
- Worst influenza virus is A (from wild aquatic birds) since it infects all mammals; B, C infect humans but milder. Its human pandemics/*serotypes* (antibody responses): known flu pandemics 1580, 1729, 1781, 1830, 1847;
  H2N2? "Russian flu" 1889-90: 1 million died;
  H1N1 "Spanish flu" 1918-20: 50 million;
  H2N2 "Asian flu" 1957-58: 1.5-2 million;
  H3N2 "Hong Kong flu" 1968-69: $700,000$;
  H5N1 Hong Kong $\geq$ 1997: fatality 50% but slow spread;
  H1N2: seasonal flu endemic in humans, pigs; H7N7.

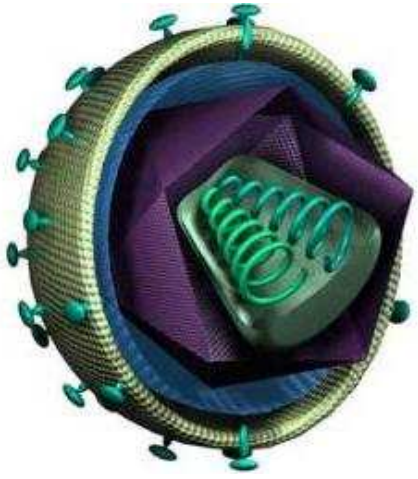
Swine flu H1N1 2009-2010 pandemic

- It started in February due large pig farms, 8.5 km to the north of the village *La Gloria* in Mexico's state Veracruz.
- This virus is unusually *mongrelised* (of mixed ancestry): possible reassortment of two (American and Eurasian) swine influenza virus strains. Six genes from American swine flu are mixtures of swine, bird and human flu.
- It kills esp. not elderly/children but people aged 25-50, producing a *cytokine storm*, immune (over)-reaction: release of too much *cytokines* (cell-signaling molecules) and then leikocytes into the lung tissue.
- ≈ 6 times as many as normal ($>\frac{1}{3}$ of population) are expected to get this flu. Safe vaccine exists and appear to work after just one dose, even in children under 10.
- From November 2009 swine flu become predominant flu strain worldwide. Aboriginals in Australia are esp. hit.

Swine flu H1N1 2009-2010 pandemic



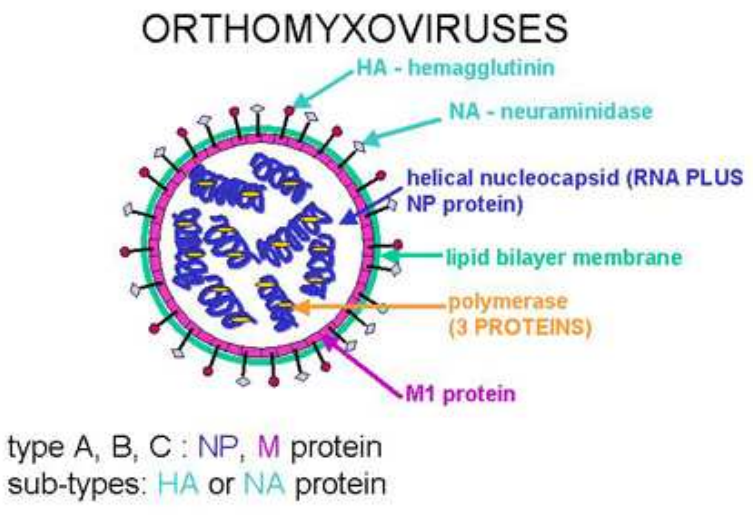
Swine flu

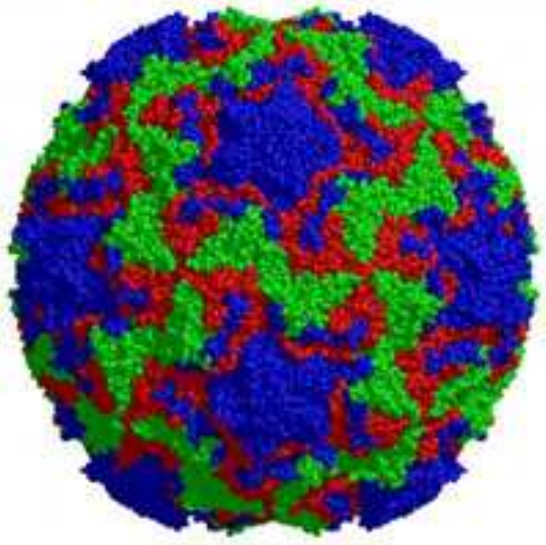


HIV

- Infection is deep in lungs (as killer H5N1), not in throat. Signs of severe form: shortness of breath, chest pain, blue lips. Only treatment: *oseltamivir* (Tamiflu).
- Most fatal cases: co-infection with bacteria and asthma, obesity. Biggest danger: reassortement with H5N1.
- In Europe, this flu come earlier than usual one, met rhinovirus (cold) and is slowed, perhaps, by interference with it and cross-immunity with seasonal endemic H1N1

Flu and common cold





Flu proteins are antigens M1, NP used to determine type (A, B or C) and HA, NA used to determine the particular sub-type (strain) of influenza A. The capsid is helical. Liddington et al., 2009, found human monoclonal antibodies inhibiting cell entry for many flu A viruses, including H5N1.

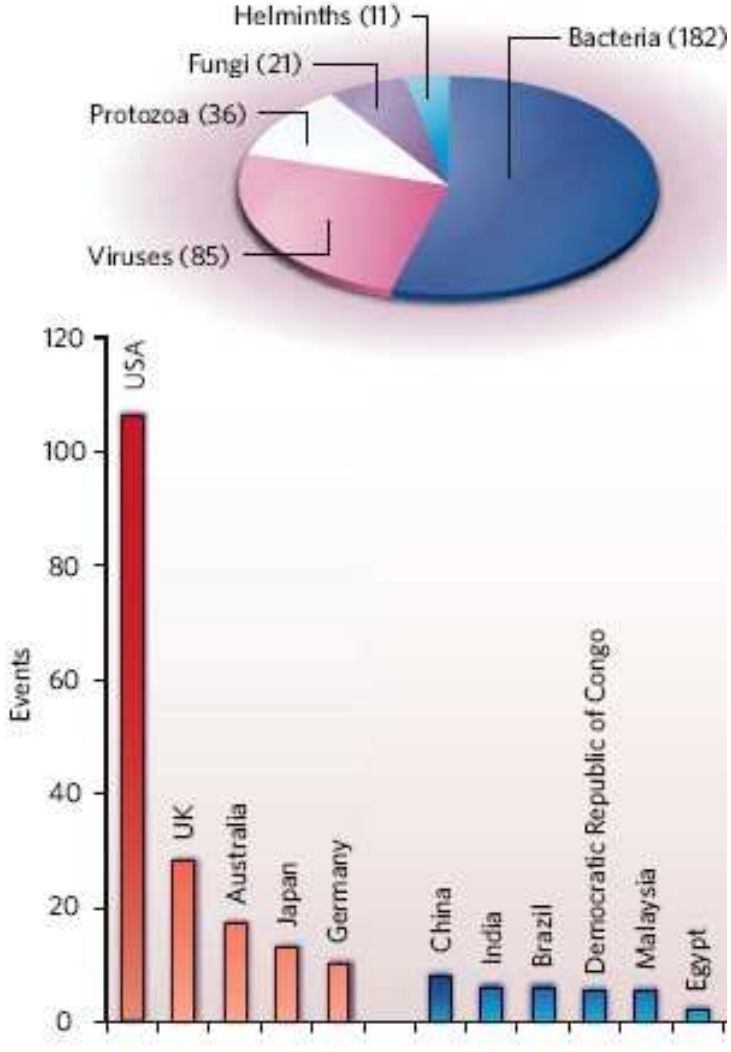
Cold (human rhinovirus, species A or B) is also RNA but its capsid is icosahedral (T=3) with 3 proteins on its surface. Ligget et al., 2009, found genetic codes for all 189 known strains and organized them into 15 sub-groups coming from distant ancestors. It explains cold uncurability at present.

Viruses versus cancer and stem cells

- Main human oncoviruses (i.e., inserting or enhancing oncogenic genes in the host DNA): Hepatitis B and C, Herpes 4 (Epstein-Barr) and 8 (Kaposi's sarcoma), Papilloma (HPV), T-cell leukaemia (HTLV-1).
- Oncolytic viruses kill (only or preferentially) cancer cells. A natural picornavirus Seneca Valley Virus-001 infects only cancer cells in small cell lung cancer.
  Tumor suppressing protein $p53$ mutates in $\frac{1}{2}$ of cancers. Genetically modified adenoviruses Onyx 015 and H101, using that, infect mainly $p53$ mutant (so, cancer) cells. Such virotherapy is permitted in China from 2005 in head and neck cancer. Added to chemotherapy, H101 injected in a tumor, doubled short-time response rates.
- Fueyo et al,2007, did virus killing brain tumor stem cells
- Retro-, lenti- and adenoviruses are used in research reprogramming of adult cells into pluripotent stem cells.

Statistics of EIDs in 1940-2004

Jones et al., in NATURE, 21-2-2008: distribution of all 335 events of origin of EIDs (emerging infectious diseases).



Impact of viruses as pathogenes

- Smallpox (major and minor variola viruses) imported by Spanish, killed $\approx 70\%$ of Native Americans.

  "Spanish flu" 1918 (variation of influenza A virus H1N1) killed 3-5% of world population; case fatality ratio: 2-20%. More than 25 million have died of HIV/AIDS from 1981. HIV, Ebola, Marburg, SARS, Avian flu are yet uncurable
- Make war: only <10 virus species are wiped out by now Only vaccine-eradicated disease: smallpox, from 1980
- Make love: sexual reproduction is not effective in cost, risk and speed. But it is the main antihazard defense of the multicellulars: arising mutations disrupt virus work.
- Shackelton and Holmes, 2008: variations of the genetic code (codon reassignments in mitochondria, ciliates, yeasts, etc.) were selected as an antiviral defense.

Impact of "domesticated" viruses

- Provirus: one latent in host cell DNA; it multiplies only when cell divides, or environment suitably changes. Some: domesticated (loose many genes but remainder become permanent functional addition to host DNA). $10-20\%$ of of bacterial DNA is prophage. In eukaryotes, such DNA come from retroviruses, i.e., RNA viruses which, infecting a cell, convert their RNA in DNA.
- $\geq 8\%$ of human genome are proven ERV (endogeneous ex-retroviruses) derived from ancient infections of germ cells. Next 40%: DNA imported horizontally by viruses as mammalian placenta and human immune system. 35 of vital human genes are of viral origin.
  In Nov. 2006, Phoenix, 5 Mya old ERV, was resurrected.
- Not family trees (vertical descent), but horizontal gene transfer by viruses acquiring genes from their hosts and esp. recombining in co-infected cells, shaped early life.

Impact of viruses on evolution

- Nakahara and Sagawa, 1968: virus theory of evolution (it was caused by viruses manipulating genetic code via transportation and mixing genes across species).
- Bell, 2001: viral eukaryogenesis hypothesis, that nucleus of eukaryotic cell evolved from endosymbiosis event: a girus (giant virus) took control of a *micoplasma* (i.e. without wall) archeal or bacterial cell but, instead of replicating and destroying it, became its ”nucleus”.
- Forterre et al., 2006: viruses evolved in "RNA world" (when RNA-organisms could produce proteins) and caused main events (DNA, cell nucleus) in early life. They combined pairs of ss RNA in ds DNA; then three takeovers of RNA chromosomes by more stable viral DNA, created 3 domains of cellular DNA-replicating life.
- Béjà, 2009: cyanophage infect *cyanobacteria* algae and upgrade cell’s photosystem to keep it alive; $\approx 5\%$ world’s oxigen result (cell takes electrons from more proteins).

Viruses: big picture

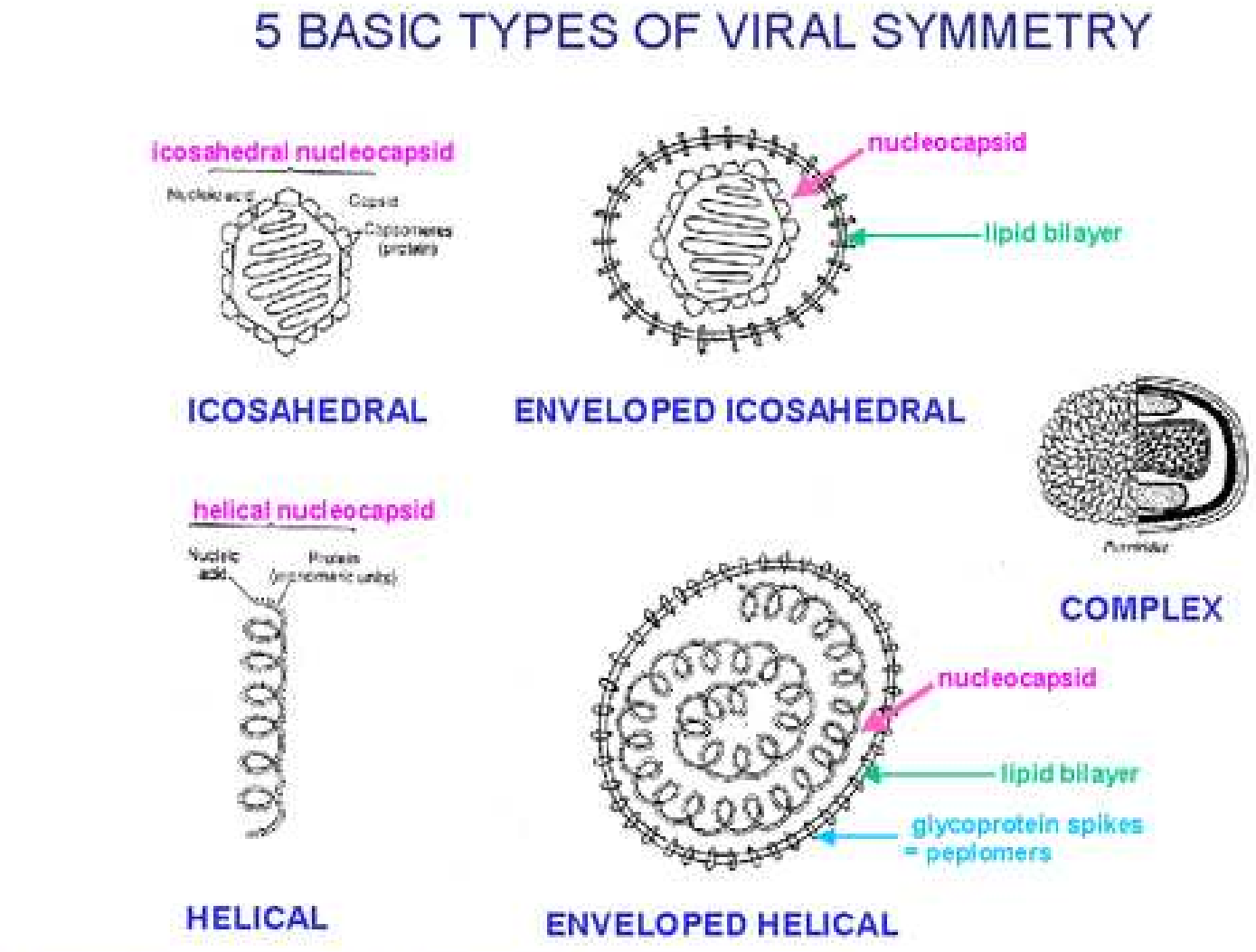
- Viruses are dominant (most abundant/diverse) Earth genetically reproducing entities. They are associated with all forms of life and kinds of environment.
- Are viruses only mobile genetic elements or they form 4-th domain (Acytota, i.e., without cells) of life? They react on their environment (but not on stimuli), have no homeostasis, mutate and reproduce spontaneously (by self-assembly as crystals, not cell division) inside a cell.
- Small viruses could evolve from plasmides (transferable pieces of DNA) as viroid $\rightarrow$ satellite $\rightarrow$ virus; larger ones, from cells-parasitizing cells.
- Virus-like entity could mediate Earth transition to a biotic world. But Jalasvuori et al., 2008: they almost not survive sporification and their diversity (>1 ancestral strain) points to local, not panspermic origin of life.

V. Symmetry of virus capsids

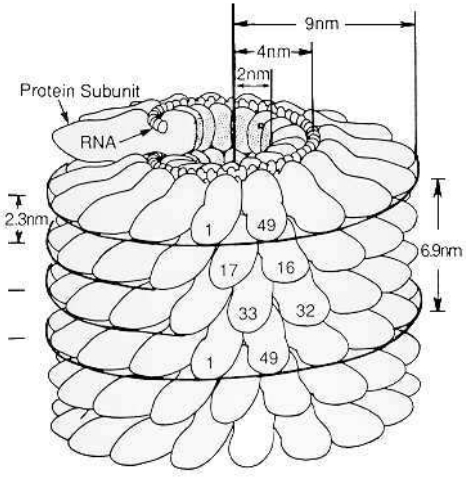
Symmetry of viruses

- Capsid of virion: protein shell enclosing genome. Some produce lipid (fat) envelope from membrane of host cell They exit cell by *budding* (continued release), not killing.
- Virus structure hints on where it initiate infection and on its *epitopes* (sites where antibodies can recognize it). It is studied by EM (electron microscopy) (showing virus-cell interaction at low and virus surface at high magnification) and cryo-EM (computer analysis of positions of x-rays diffracted by crystallized virus).
- Main virus morphologies: helical (linear), icosahedral, complex and naked or enveloped (for them shapes are of the core). Archeal viruses can be also lemon (with tails on both ends), droplet and bottle shaped.
- Icosahedral one is also called *isometric*, *polyhedral*, of *spherical shape*, of *cubic* or *rotational 2-3-5* symmetry.

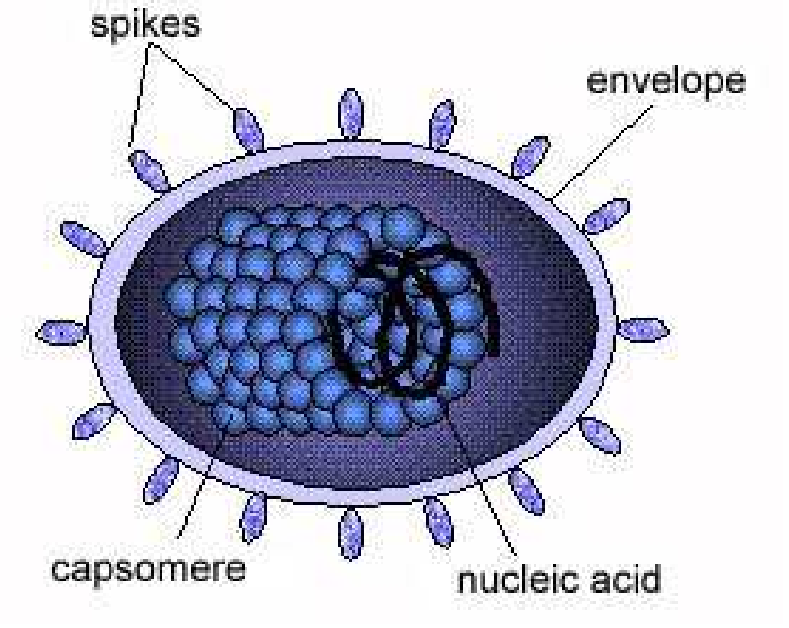
Symmetry of viruses



Helical viruses



Tobacco mosaic virus

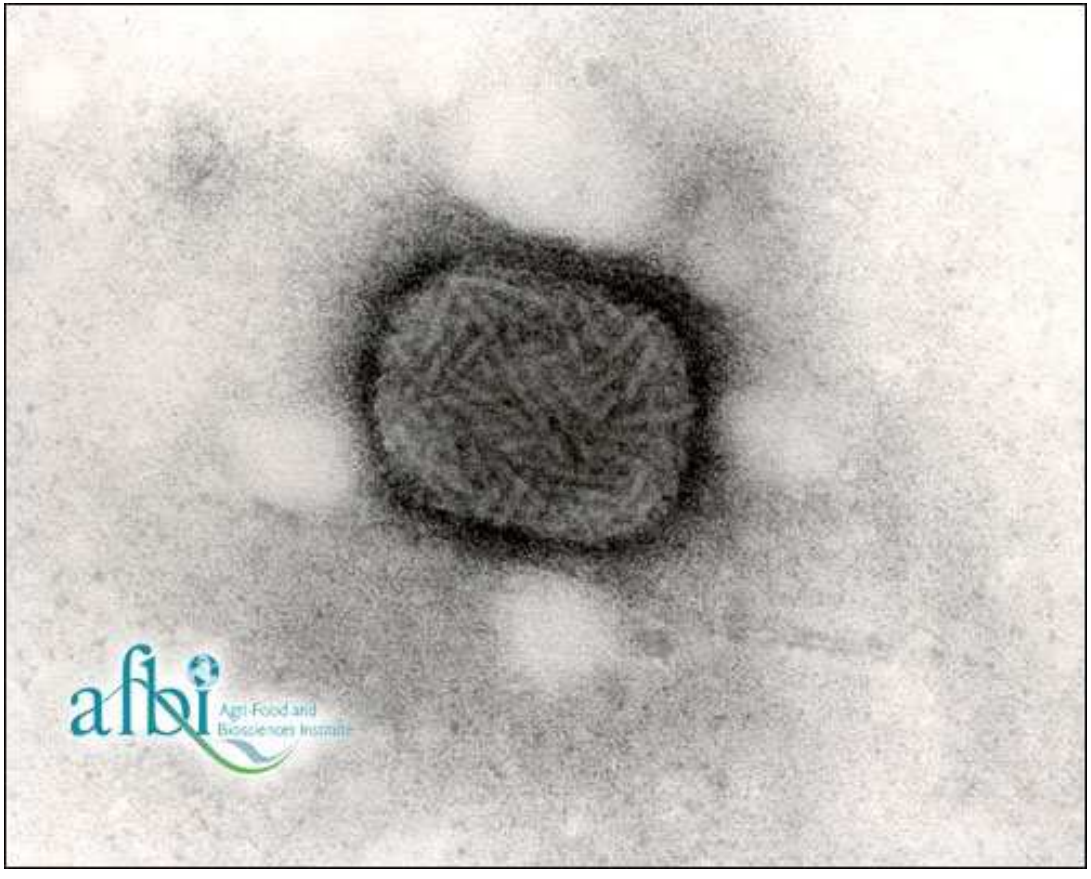


Enveloped helical virus

- Helical viruses are defined by their amlitude (diameter) and pitch (the distance covered by each complete turn of the helix). Longer ones are flexible: curved or bent.
- Until 1960, only known helical viruses were of plants. All helical animal viruses, known by now, have RNA.

Complex structure: poxviruses

Their large (visible in OM) capsids are brick-shaped or oval.



Orthopoxvirus

Avian poxvirus

Crick and Watson had a dream

- Crick and Watson's article in Nature, 10-3-1956, starts: "It is a striking fact that almost all small viruses are either rods or spheres." They suggested 3 Platonic groups. But Hodkin, 1949: small spherical viruses could be built up of subunits related by cubic symmetry.
- In fact, all virions, except complex ones (as poxviruses, tailed phages and β-like viruses) are helical or ($\approx \frac{1}{2}$ of all, most animal and almost all human) icosahedral.
- Icosahedral viruses are defined by their six 5-, ten 3-, 15 2-fold axes of symmetry at vertices, faces, edges.
- For weaker icosahedrality, see *pseudo-equivalence* and Twarock, 2004, Janner, 2006, Chen et al., 2007 below.

  For non-icosahedral fullerene forms, see retroviruses HIV, RSV and prolate shape of complex phages below.

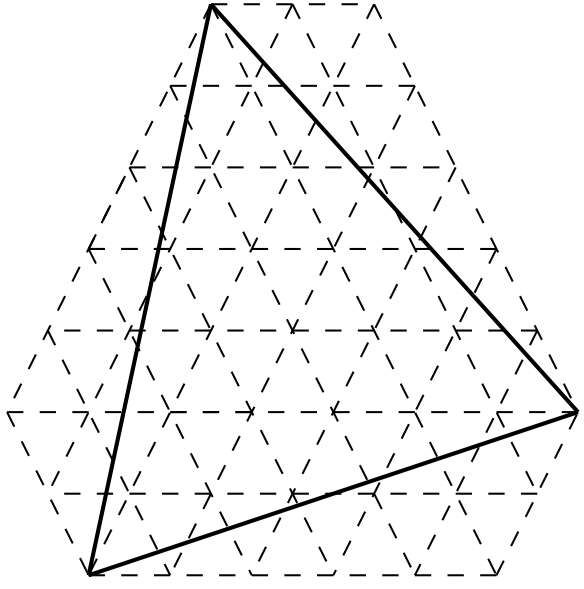
VI. Icosahedral viruses

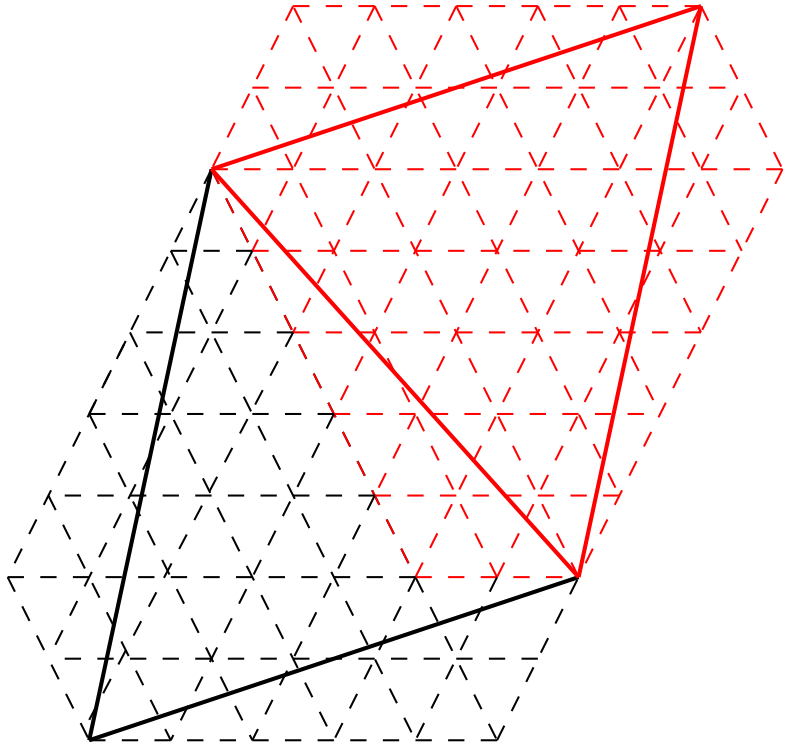
Icosahedral viruses

- Protomer: each protein subunit (1 or more polypeptide chains) in a capsid. Capsomers: polygonal rings of n protomers (n=$5, 6, 3, 2$) of diameter ≤ 4 nm joined by *bonds* (protein/protein interaction) to form virus shell.
- Bonds are flexible: $\simeq 5^0$ deviation from mean direction. *Self-assembly*: slight but regular changes in bonding. Its disruption gives non-infectious structures. Dengue virus (T=3) creates 1 viable particle in each $4,000$ tries.
- Caspar-Klug principle: virion minimizes free energy by organizing capsomers quasi-equivalently: icosahedral symmetry but with more than 60 (maximum if protomers have identic environment) subunits.
- But origin, energy, termodynamics and kinetics of this self-assembly is unclear. Modern computers cannot evaluate capsid free energy by all-atom simulations.

Goldberg-Coxeter construction

Given (a, b) ($(5, 2)$ below), put latiice triangle on $p6$ *net* $\{3^6\}$ and replace its 6-valent vertices by pentamers.
Gluing pieces coherantly gives other triangulation of plane.
Fold then 20 identic Icosahedron faces into *quasi-equivalent Icosahedron*, i.e., icosideltahedron.





Capsomer arrangement

- It is given, on virion surface, by the *triangulation number* T=a^2+ab+b^2 of capsid icosadeltahedron C^*_{20T}.
  Capsid has $60T$ protomers but EM resolves only some capsomers: around vertices (12 pentamers and $10T$-10 hexamers), on $20T$ triangular facets (trimers of three protomers, 1 in each corner), or on $30T$ edges (dimers).
  Some mutations produce $60T$ monomers, no clustering.
- Following viruses with T=3 have their 60×3 protomers clustered as: 12 pentamers plus (10×3)-10 hexamers (Turnip Yellow Mosaic virus); 20×3 timers (Poliovirus), and 30×3 dimers (Turnip Crinkle Virus).
- Clustering maximizes interactions; so, stabilizes virion.
  Capsomers bonds are weaker than between protomers.

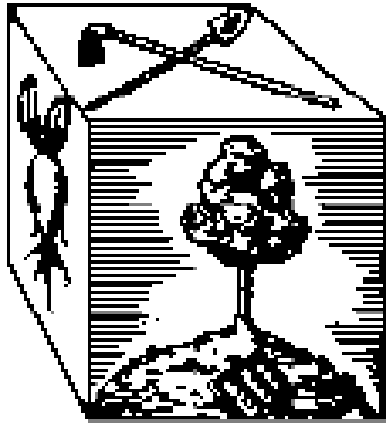
Quasi-equivalent bonding

- So, in contrast to crystal lattice, capsomers can now be *quasi-equivalent*, i.e. similar (on the same face of Icosahedron), but not symmetry equivalent. *Quasi-equivalent icosahedron* means *icosideltahedron*.
- Triangulation number T=a^2+ab+b^2 is the number of locations with non-equvalent bonding;
each face of Icosahedron correspond to T small triangles (subunit in each corner; so, $60T$ protomers). Those icosahedral asymmetric units are related by *quasi-equivalent symmetry axes*, i.e. symmetry elements holding only locally.
- General quasi-equivalence: any small non-random variation in a regular bonding pattern leading to a more stable structure than strictly equivalent bonding.

“Most diseases come from icosahedra”

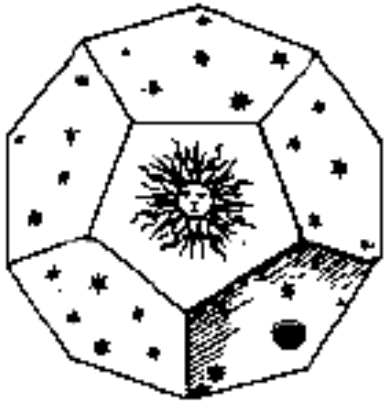
Hippocrates, circa 400 BC: disease is icosahedra (water) excess in the body.

OCTAHEDRON
Air

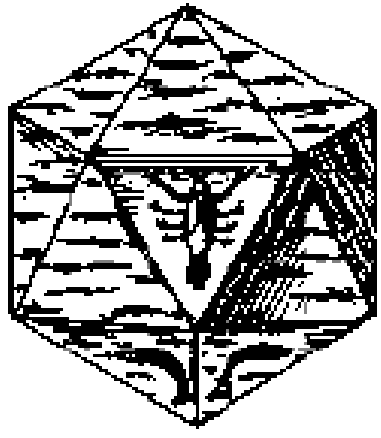


CUBE
Earth

TETRAHEDRON
Fire



DODECAHEDRON
the Universe

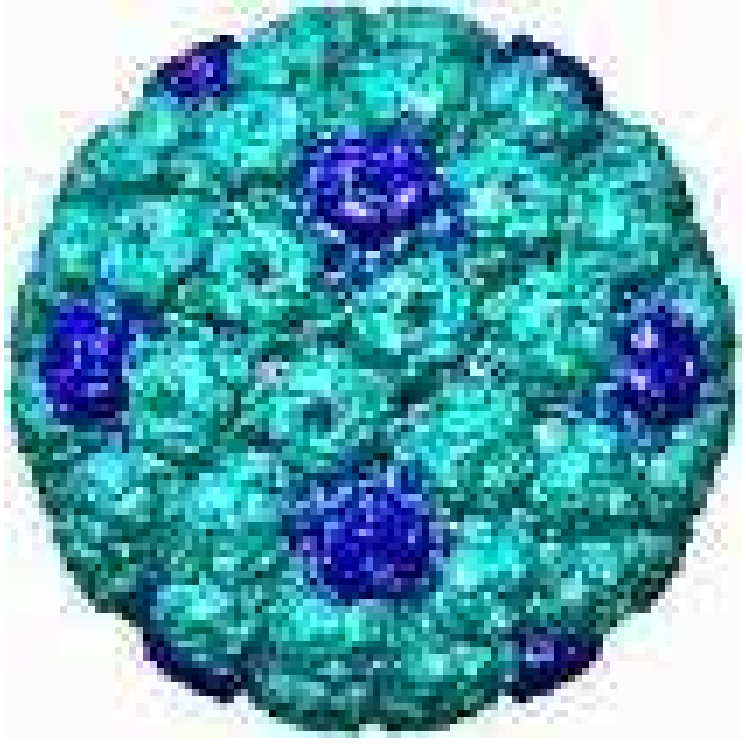


ICOSAHEDRON
Water

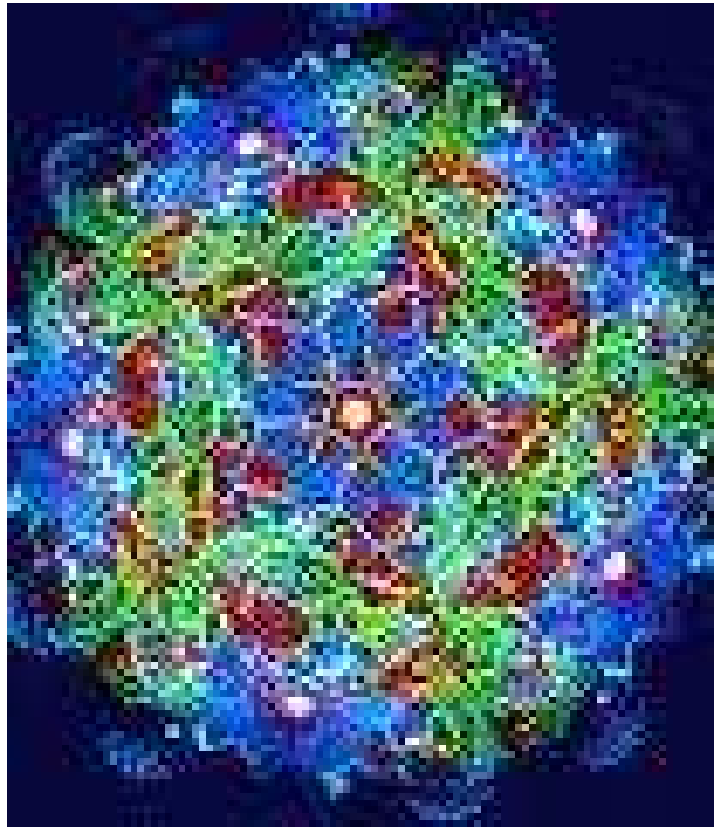
Capsids of icosahedral viruses

(a,b)	$T=a^2+ab+b^2$	Fullerene	Examples of viruses
$(1,0)$	1	$F^*_{20}(I_h)$	*B19 parvovirus, cowpea mosaic virus*
$(1,1)$	3	$C^*_{60}(I_h)$	*picornavirus, turnip yellow mosaic virus*
$(2,0)$	4	$C^*_{80}(I_h)$	*human hepatitis B, Semliki Forest virus*
$(2,1)$	$7l$	$C^*_{140}(I)_{laevo}$	*HK97, rabbit papilloma virus, Λ-like viruses*
$(1,2)$	$7d$	$C^*_{140}(I)_{dextro}$	*polyoma (human wart) virus, SV40*
$(3,1)$	$13l$	$C^*_{260}(I)_{laevo}$	*rotavirus*
$(1,3)$	$13d$	$C^*_{260}(I)_{dextro}$	*infectious bursal disease virus*
$(4,0)$	16	$C^*_{320}(I_h)$	*herpes virus, varicella*
$(5,0)$	25	$C^*_{500}(I_h)$	*adenovirus, phage PRD1*
$(3,3)$	27	$C^*_{540}(I_h)$	*pseudomonas phage ϕKZ*
$(6,0)$	36	$C^*_{720}(I_h)$	*infectious canine hepatitis virus, HTLV1*
$(7,7)$	147	$C^*_{2940}(I_h)$	*Chilo iridescent iridovirus* (outer shell)
$(7,8)$	$169d$	$C^*_{3380}(I)_{dextro}$	*Algal chlorella virus PBCV1* (outer shell)
$(7,10)$	219	$C^*_{4380}(I)_{dextro?}$	*Algal virus PpV01*

Examples



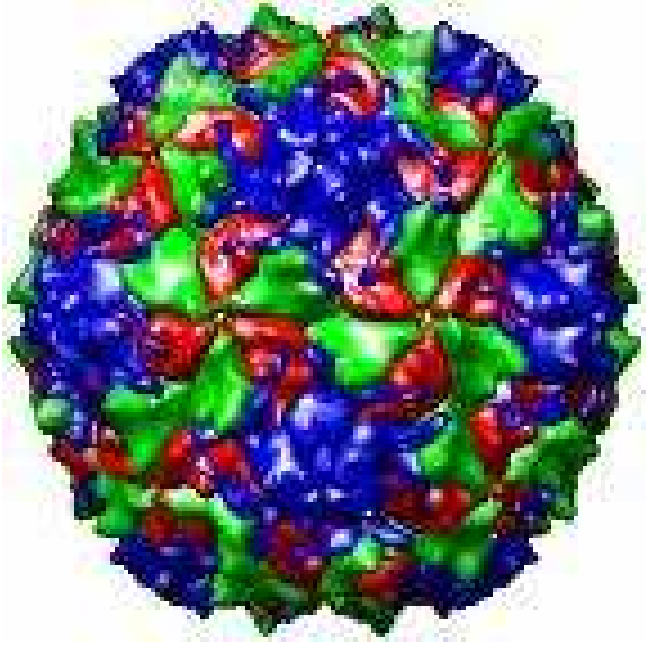
Satellite STMV, T=1, of TMV, helical Tobacco Mosaic virus 1st found (Ivanovski, 1892, and Beijerinck, 1899), 1st seen (Stanley, 1931) EM



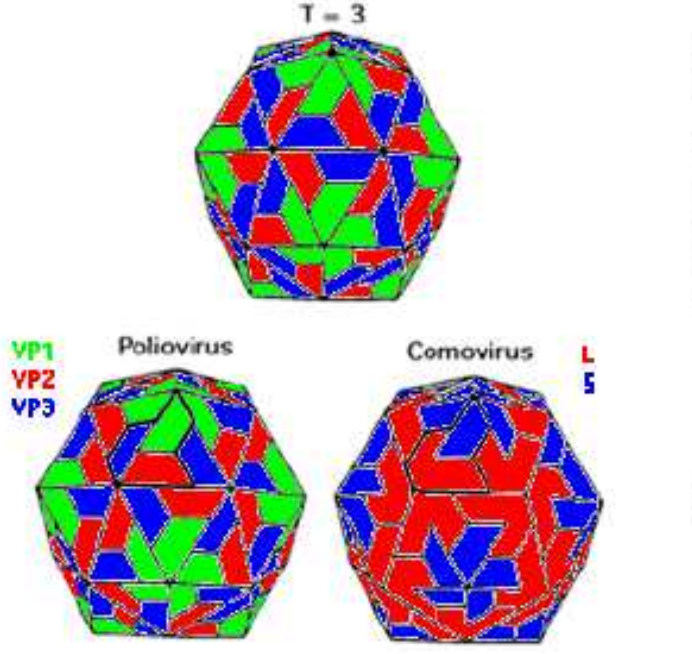
Foot-and-Mouth virus, $T = 3$

Pseudo-equivalence

Pseudo-equivalence is reduced quasi-equivalence, arising if $60T$ protomers are not chemically (by sequence) identical. Poliovirus and BPMV comovirus both have $T = 3$. Their subunits (trapezoids) are proteins VP1, VP2,VP3 and S, L. In Polio, around 5-fold, 5 VP1 (quasi-equiv.), but around 3-fold (triangle centers), 2 ways; so, pseudo 3-fold axes.

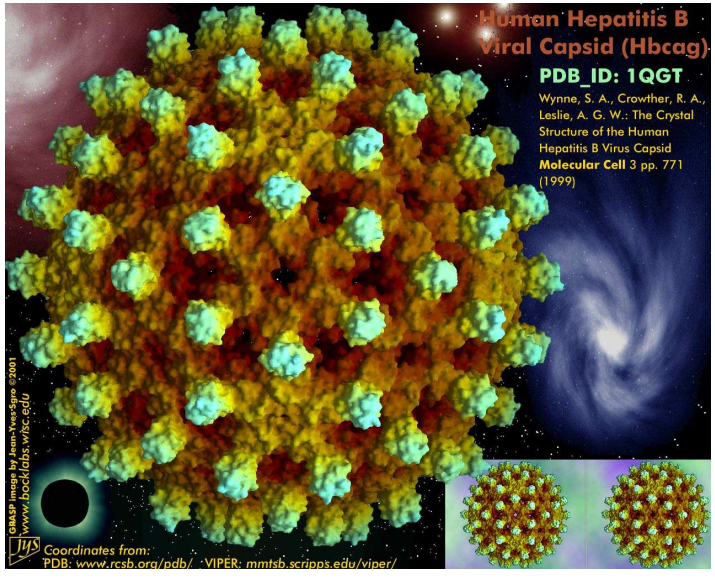


Poliovirus

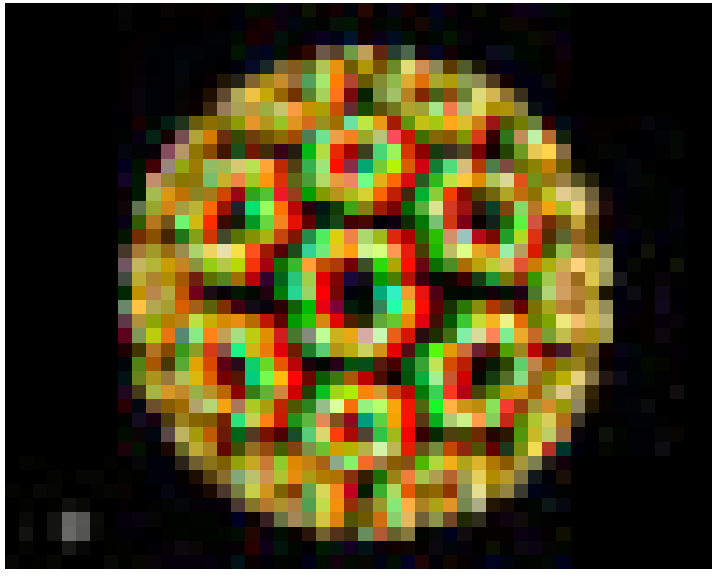


Polio and BPMV viruses

Viruses with $T = 4$



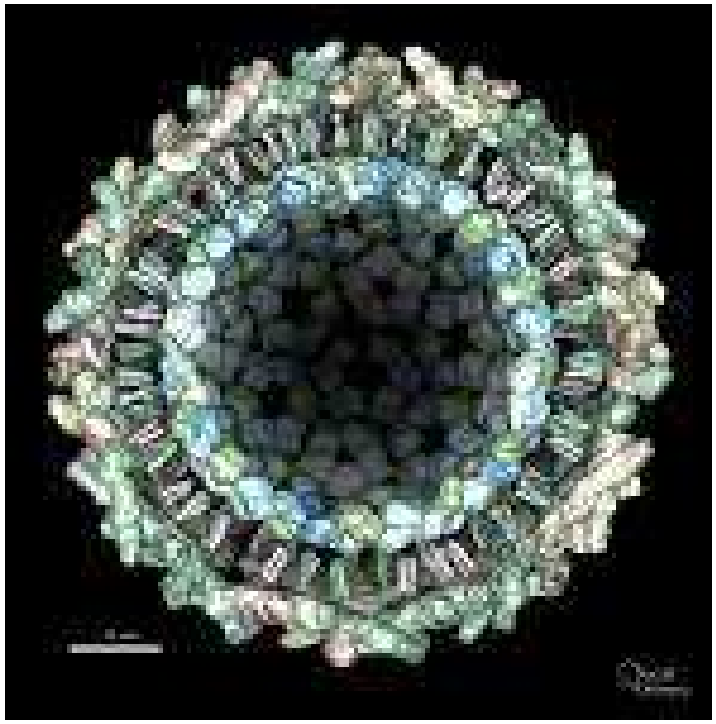
Human hepatitis B



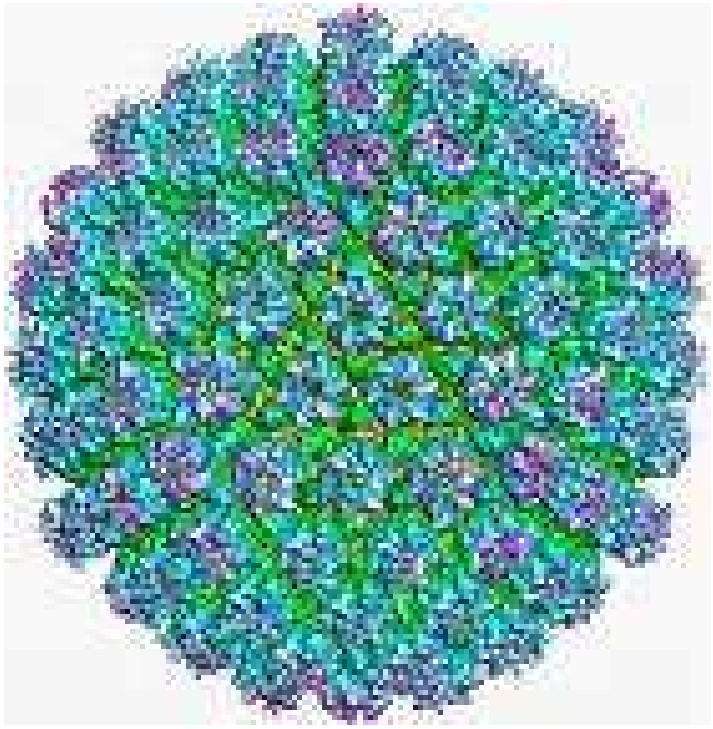
Semliki Forest virus

But all known viruses with $T > 7$ have more than one core capsid protein (subunit type) and so, pseudo-equivalent.

More $T = a^2$ viruses



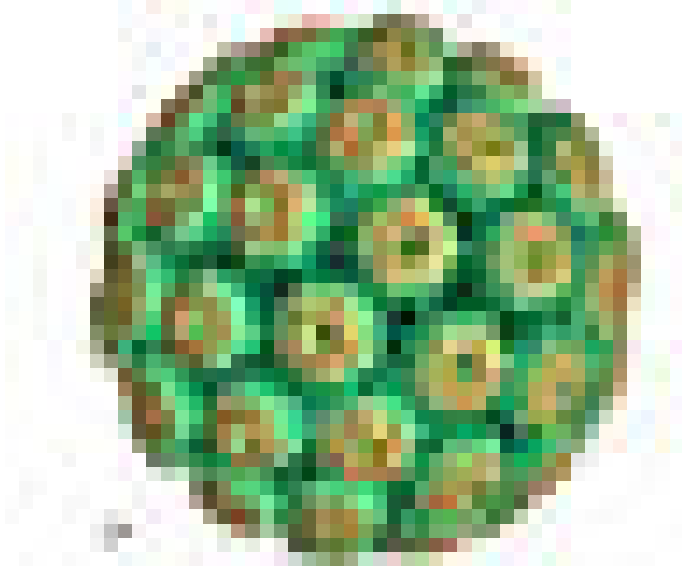
Sindbis virus,
$T = 4$



Herpes virus,
$T = 16$

72 pentamers puzzle: 2 papilloma viruses

Human polyoma, T=$7d$?



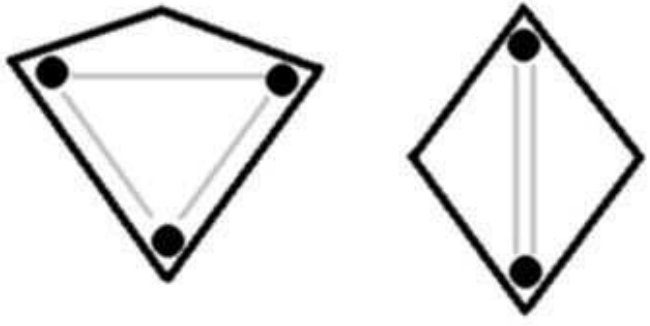
Simian virus $SV40$, T=$7d$?

They violate quasi-equivalence: there are 72 capsomers (so, 72=12+$10(T$-$1)$ imply T=7) but all (instead of only 12) are 5-mers (so, 360 subunits implying T=6, since Caspar-Klug tiling is by $20T$ regular triangles, 1 subunit in each corner).

Twarock, 2004, solved it proposing instead Penrose-like tilings of Icosahedron by rhombus and kite.

Twarock's model of protein interactions

Trimer and dimer seen as golden kite and thick rhombus.



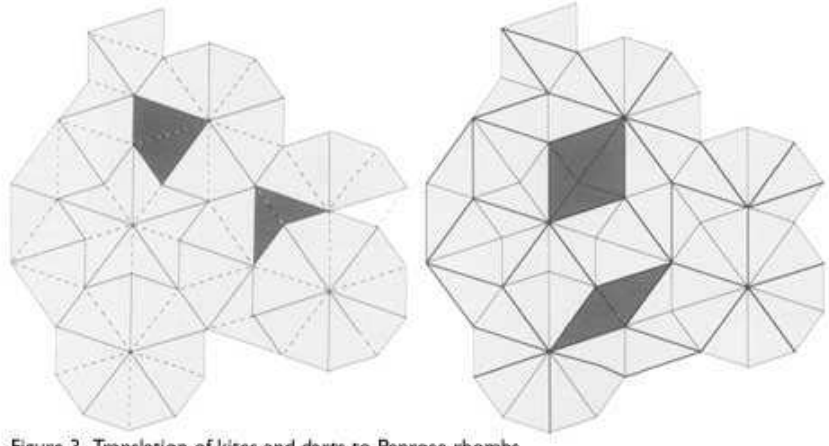
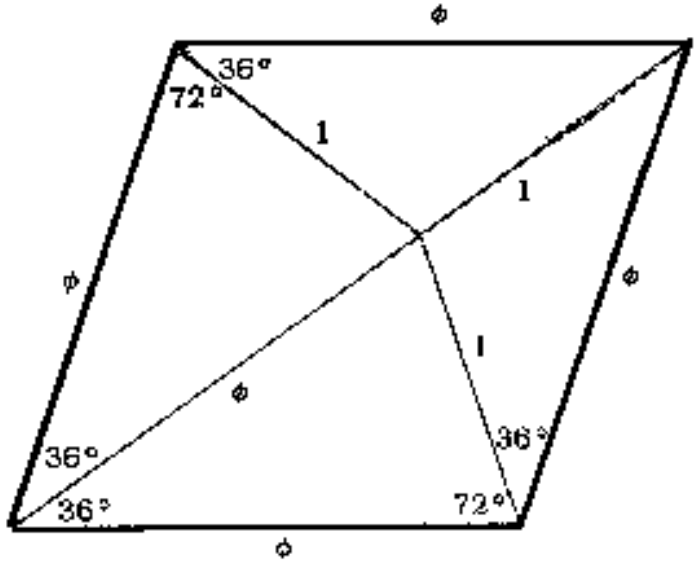
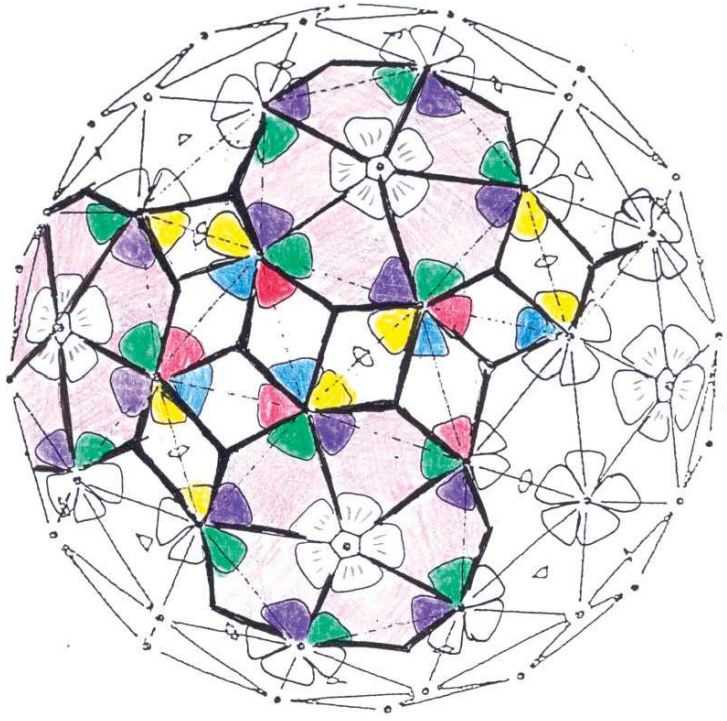


Figure 3. Translation of kites and darts to Penrose rhombs



$$\phi = \frac{1+\sqrt{5}}{2}$$

72 pentamers puzzle: Twarock solution



The locations of (12) 5-, (30) 2- and (20) 3-fold symmetry axes visualize the action of the icosahedral group H_3=I_h.

Tiles are rhombs and kites. All corners of the tiles meeting at 5-valent vertices mark the locations of protein subunits, i.e., exactly at tile corners subtending the same angle $72°$.

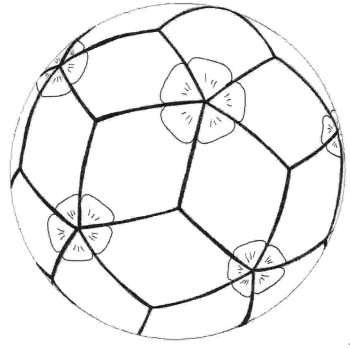
Quasi-crystal from $\mathbb{D}_6$

- Checkerboard root lattice $\mathbb{D}_n$= $\{x \in \mathbb{Z}^n : \sum_{i=1}^{n} x_i$ is even$\}$.
- Let $\{e_i\}$ be the unit vectors of $\mathbb{R}^6$ and $\{a_i\}$ be the *roots*: a_1=e_2-e_1; a_2=e_1-e_3; a_3=e_3-e_6; a_4=e_5+e_6; a_5=-e_4-e_5; a_6=e_4-e_5
- Map $\{e_i\}$ on vectors pointing to 6 non-aligned vertices of Icosahedron: $e_1 \rightarrow \frac{1}{2}(1, 0, \tau)$; $e_2 \rightarrow \frac{1}{2}(\tau, 1, 0)$; $e_3 \rightarrow \frac{1}{2}(0, \tau, 1)$; $e_4 \rightarrow \frac{1}{2}(-1, 0, \tau)$; $e_5 \rightarrow \frac{1}{2}(0, -\tau, 1)$; $e_6 \rightarrow \frac{1}{2}(\tau, -1, 0)$. $\tau = \frac{1+\sqrt{5}}{2}$.
- $\mathbb{D}_6$ is $\mathbb{Z}$-linear combination of a_i, $1 \leq i \leq 6$; its projection in $\mathbb{R}^3$ is ($\mathbb{Z}[\tau]$=$\{a + \tau b : a, b \in \mathbb{Z}\}$)-linear combination of a'_1=$\frac{1}{2}(\tau - 1, 1, -\tau)$, a'_2=$\frac{1}{2}(1, -\tau, 1 - \tau)$, a'_3=$\frac{1}{2}(-1, \tau, 1 - \tau)$.
- Vertices α of Icosidodecahedron (24 even permutations of $\frac{1}{2}(\pm 1, \pm\tau, \pm(1 - \tau))$ and 8 permutations of $(\pm 1, 0, 0)$) encode the generators of H_3 as reflections (in the plane orthogonal to the vector α) $r_\alpha : x \rightarrow x - \frac{2\langle x,\alpha\rangle}{\langle \alpha,\alpha\rangle}$ for $x \in \mathbb{R}^3$.

Twarock tilings

Their vertex-sets are on nested shells in S_m, the set of all $\mathbb{N}$-linear combinations of, up to $m \leq 32$ vectors (vertices) α. In S_3, it occurs Triacontahedron, a decorated Icosahedron.





- Penrose tilings of $\mathbb{R}^2$, $\mathbb{R}^3$, $\mathbb{R}^4$ have symmetry types H_2, H_3, H_4, respectively. Those tilings can be realized as irrational slices of root lattices E_6 or E_8.
- *Coxeter groups* H_2=$I_2(5)$, H_3=I_h and H_4 of orders $2! \cdot 5$, $3! \cdot 20$ and $4! \cdot 600$ are symmetry groups of 2-dim. 5-gon, 3-dim. 20-faced Icosahedron and 4-dim. 600-cell.

Digression: noncrystallographic groups

- A finite *Coxeter group* (not product of smaller ones) is noncrystallographic (not stabilizes a lattice) if and only if it is H_4, H_3 or $I_2(p)$ with $p \neq 2, 3, 4, 6$.
- Which their higher-dim. (so, reducible) representations can be viewed as the point group of a lattice?
- $I_2(p)$ is crystallographic in $\mathbb{R}^n$ iff $n \geq \varphi(p)$, where $\varphi(p)$ is the Euler's totient function $|\{1 \leq i \leq p\} : gcd(i,p) = 1|$. It is equal to $1, 2, 2, 4, 2, 6, 4, 6, 4, 10, 4$ for $2 \leq p \leq 12$.
- H_3 is crystallographic in $\mathbb{R}^n$ if and only if $n \geq 6$. All H_3-symmetric lattices in $\mathbb{R}^6$ are: simle cubic $\mathbb{Z}_6$, its half, face-centered cubic $\mathbb{D}_6$, and body-centered cubic $\{x \in \mathbb{Z}^6 : x_i \equiv x_j (mod\, 2)$ for all $1 \leq i, j \leq 6\}$.
- Embedding of noncrystallographic Coxeter groups into crystallographic ones: $H_2 \to A_4$, $H_3 \to D_6$, $H_4 \to E_8$.

Three most recent theories

- Reidun Twarock, 2004 (German/British mathematician): got icosahedral tilings by projection from the lattice D_6. Other way, by local rules, was in Berger-Shor, 1994.
- Aloysio Janner, 2006-2007 (Dutch crystallographer): construct icosahedral polyhedra in the icosahedral lattice generated by 6 non-aligned (and suitably scaled) vectors to vertices of Icosahedron from its center. It generalizes Caspar-Klug and Twarock tilings of Ico.
- Chen-Zhang-Glotzer, 2007 (US): molecular Monte Carlo simulation of the self-assembly of cone-shaped particles minimizing free energy subject to convexity constraint. Caspar-Klug and Twarock tilings come out for $N = 12, 32, 72, 132$, i.e. $\frac{N-2}{10} = T = 1, 3, 7, 13$. Bruinsma-Gelbart-Requera-Rudnick-Zandi, 2003 (US) got $N = 12, 32, 42, 72$ supposing, moreover, sphericity.

Icosahedral lattice IL

It is $\{\sum_{i=1}^{6} n_i a_i : n_i \in \mathbb{Z}\}$ (a projection of $\mathbb{Z}^6$ into $\mathbb{R}^3$), where $\{a_i\}$ (projections of $\{e_i \in \mathbb{Z}^6\}$) are vectors from Icosahedron center to its 6 non-aligned vertices, i.e., on 5-fold axes.
The vectors $a_1, \ldots, a_6$ are linearly independent over $\mathbb{Q}$; so, the coordinates $n_1(x), \ldots, n_6(x)$ are unique for any $x \in IL$.

For some number a_0, it holds: a_1=$a_0(e_1$+$\tau e_3)$,a_2=$a_0(e_2$+$\tau e_1)$, a_3=$a_0(e_3$+$\tau e_2)$, a_4=$a_0(-e_1$+$\tau e_3)$,a_5=$a_0(e_3$-$\tau e_2)$,a_6=$a_0(-e_2$+$\tau e_1)$ where $\{e_1, e_2, e_3\}$ is the orthonormal $\mathbb{R}^3$-basis and τ=$\frac{1+\sqrt{5}}{2}$.

The icosahedral group $I \simeq A_5$ is the group of proper rotations of Icosahedron, generated by 5- and 3-fold rotations R_5, R_3 (around a_1, a_1+a_2+a_3) with R_5^5=R_3^3=$(R_5R_3)^2$=1.
2-fold rotation around e_3 is given by $R_5^2 R_3 R_5^{-1}$.

Janner's general conjecture: capsid acts as resonator with nodes of wave-like eigenmodes at various IL lattice points.

Janner's model of rhinovirus

Cold (human rhinovirus) form largest (≈ 100 serotypes) genus of Picornaviruses having icosahedral (T=3) capsid. Caspar-Klug's model, C_{60}^*, explains 3 proteins on its surface but not $VP4$ at interface between capsid and RNA cavity. X-ray diffraction gave structure of serotypes 16,14,3,2,1A.

Janner, 2006: slight deviation, affine C_{60}^*, has vertices in IL: 12 5-valent (of Icosahedron) with even and 20 6-valent ones (of $\frac{1}{\tau^2}$-rescaled Dodecahedron) with odd values $(n_1, \ldots, n_6)$.

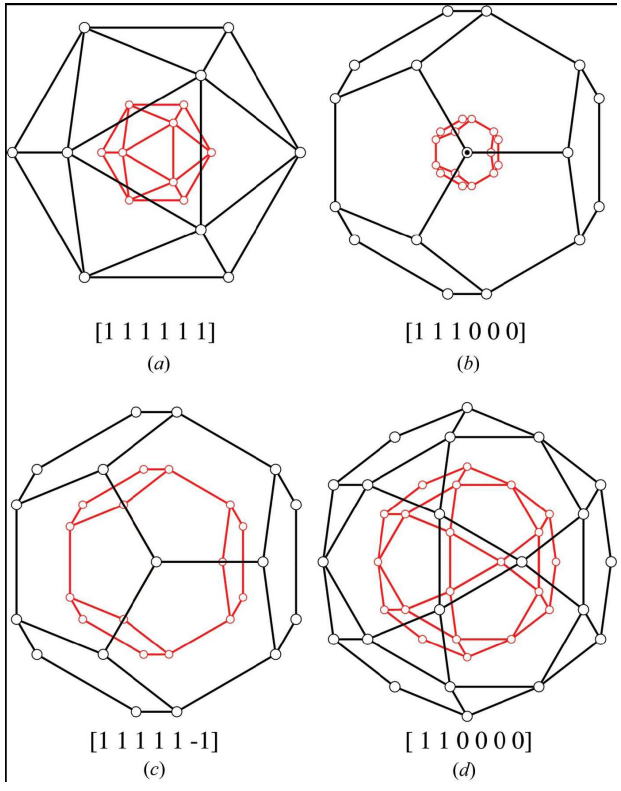
This radial scaling is given by matrices $\frac{1}{\tau^2} I_3$ and $\frac{1}{2}(4I_6 - J_6)$, in terms of basises $\{e_1, e_2, e_3\}$ and $\{a_1, \ldots, a_6\}$, respectively.

Capsid is encapsulated between 2 such polyhedra: internal, $\frac{1}{\tau}$ smaller, delimites core. Moreover, layers of each protein lie on polyhedra in IL with symmetry of subgroups of I.

Value a_0 is ≈ 9 nm but it depends on serotype and proteins.

Janner's general model

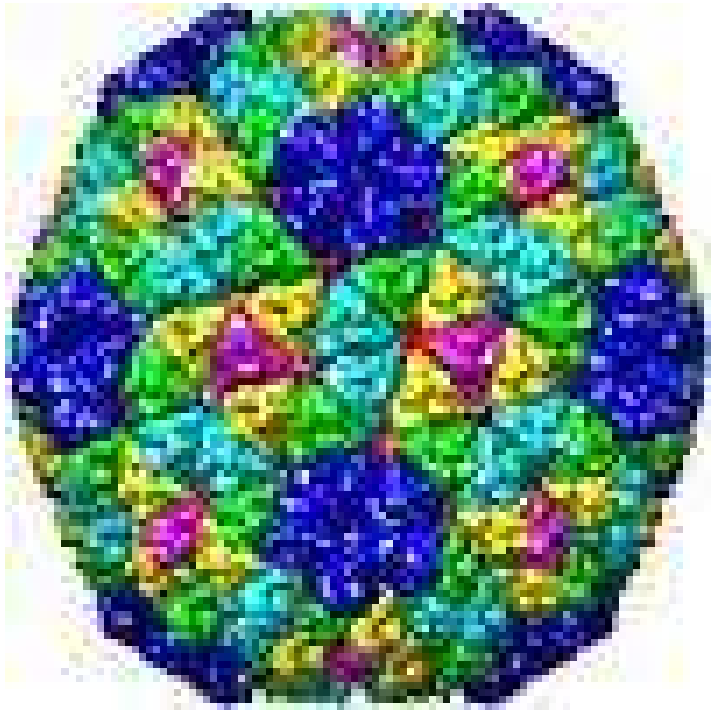
Janner, 2007: icosahedral capside surface, delimited by external and internal form, can be characterized as icos. cluster of 2 polyhedra in $3D$ $\mathbb{Z}$-module, generated from a single point in $\mathbb{Z}_6$. See below point orbits by the action of $I_h \times C_4$: τ^2-scaled icosahedra, τ^3-scaled dodecahedra, τ-scaled dodecahedra and τ-scaled icosidodecahedra.



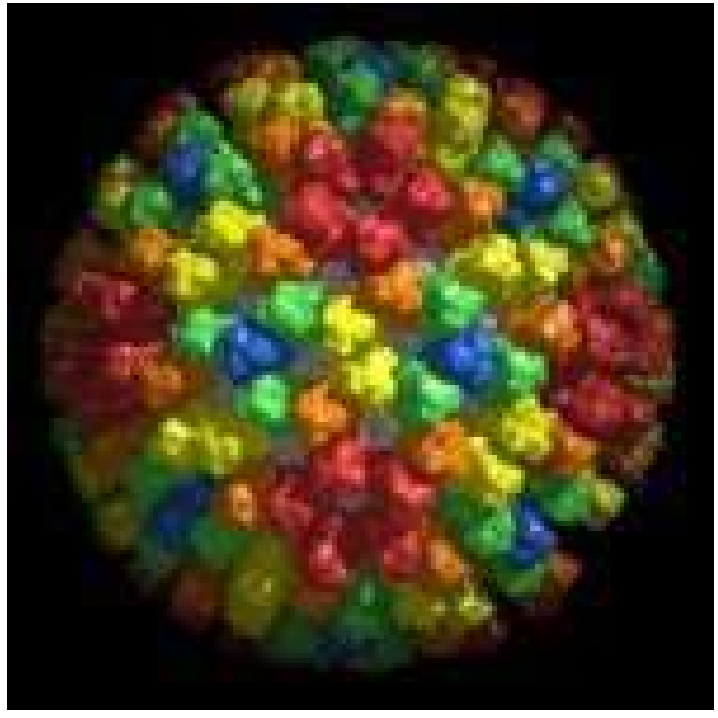
Virus dynamics and group representation

- If capsid A has symmetry $G \in \{I_h, I\}$ with triangulation number T, then A has $n{=}60T$ protein units (protomers).
- Peeters-Taormina, 2008, compute vibration spectrum of A approximating protomers as n point spring-masses.
- An non-linear n-atoms molecule has $3n-6$ degrees of vibrational freedom: $3n$ for translational motion minus 3 translations and 3 rotations of the molecule as a whole.
- (Reducible) displacement representation of G (and A) consists of $|G|$ $3n{\times}3n$ matrices $D_g{=}P_g{\otimes}R_g$, $g \in G$ where P_g are permutation $n{\times}n$ matrices and R_g are rotation $3{\times}3$ matrices forming an irreducible representation of G.
- Using decomposition of D_g in block diagonal form, they obtain, for example, for Rice Yellow Mottle Virus having $T{=}3$: 54 Raman active modes (including 45 degenerated 5-fold) and 25 degenerated 3-fold infrared active modes.

Viruses with $T = 13$ *laevo*, i.e., $(3, 1)$



Rice dwarf virus



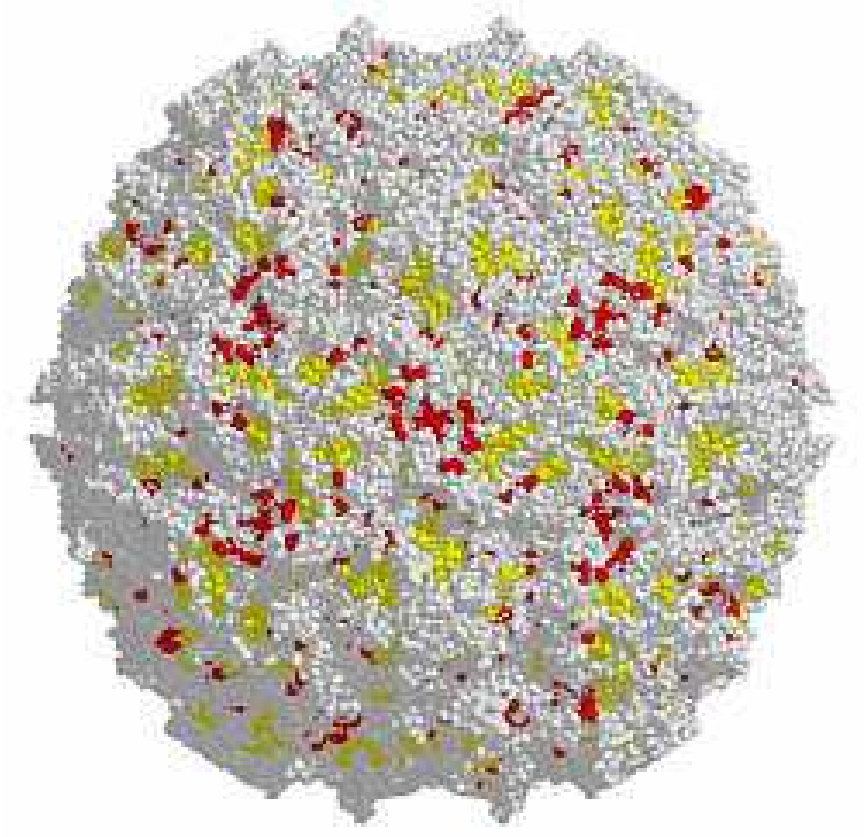
Bluetongue virus

Rhesus rotavirus have multiple layers with $13\,laevo$ and different organization of $13{\times}60$ subunits into capsomers.

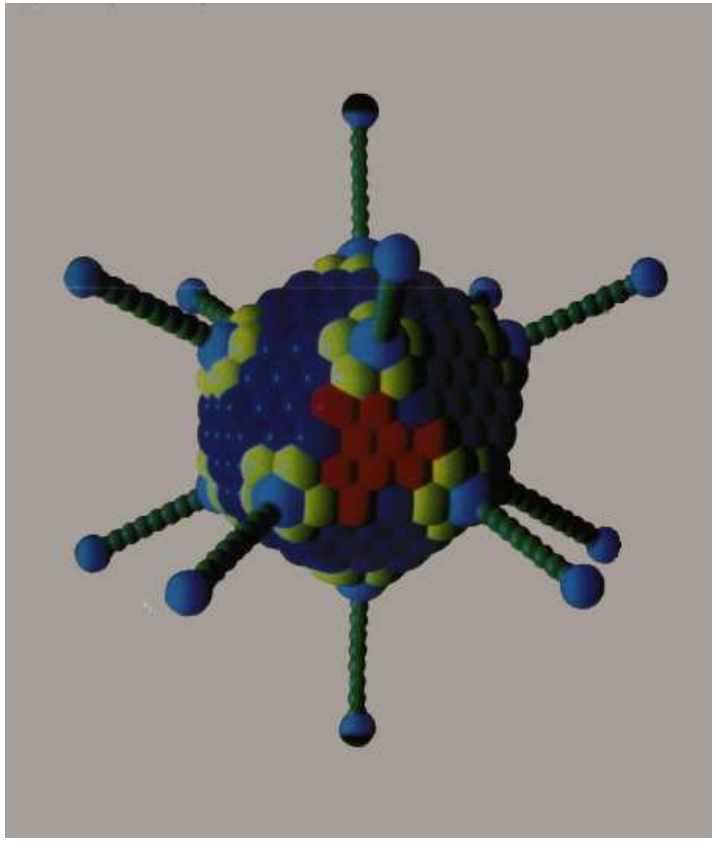
Bursal dicease virus is an example of ones with $13\,dextro$.

Viruses with $T = 25$

Smallest viruses observed directly (by EM) have $T = 25$.

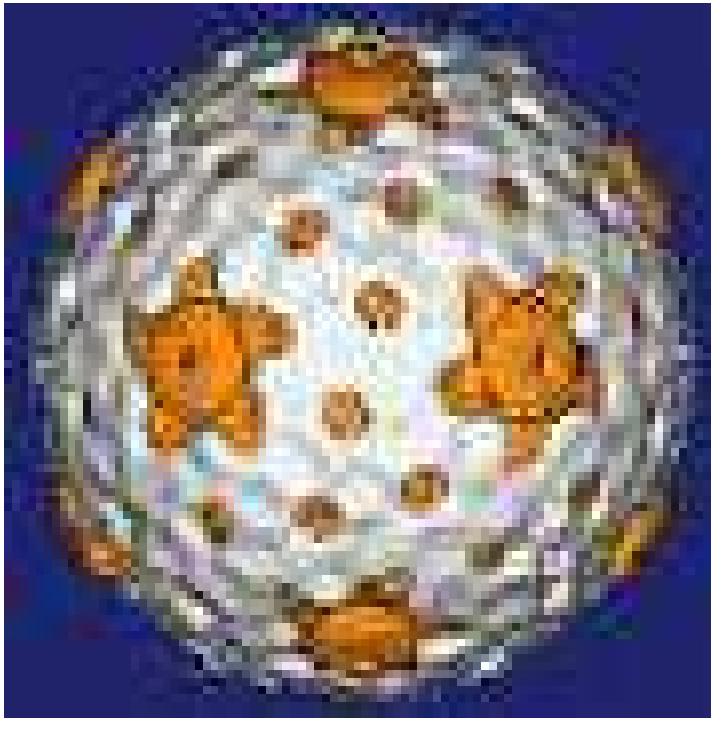


PRD1 virus

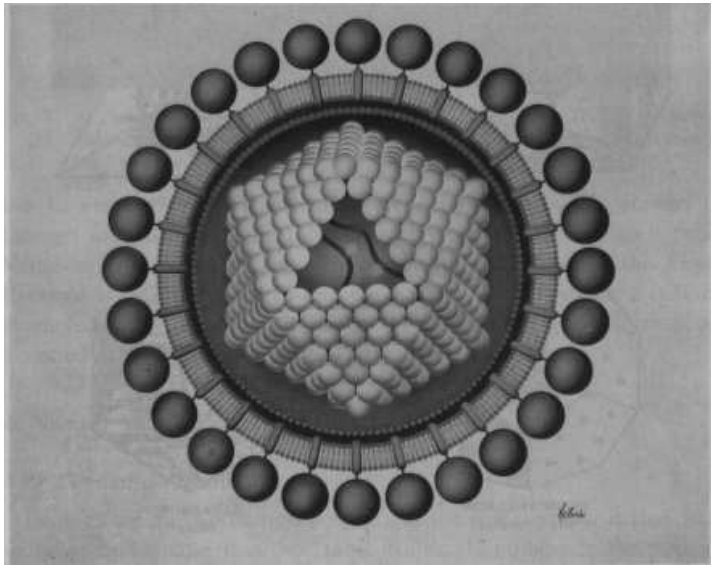


Adenovirus (with its spikes)

More I_h-viruses

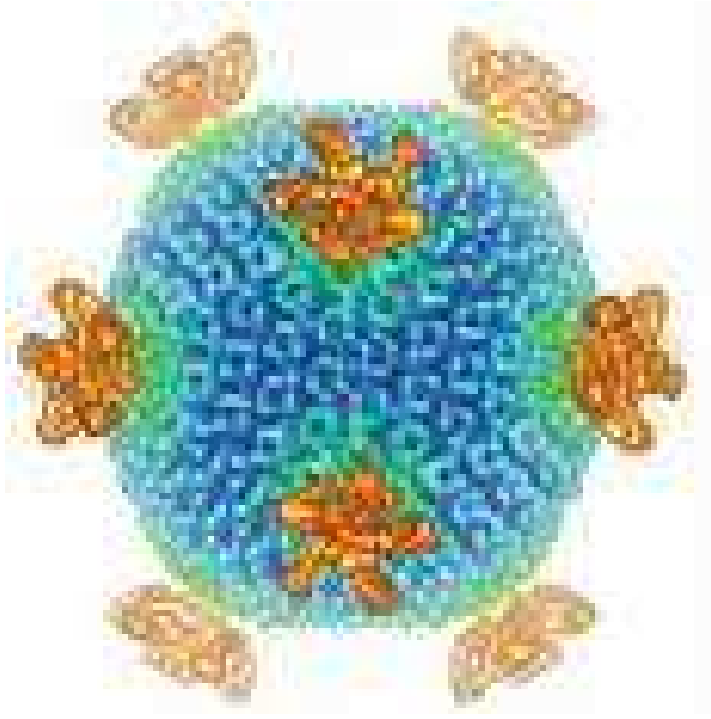


Pseudomonas phage phiKZ,
$T = 27$

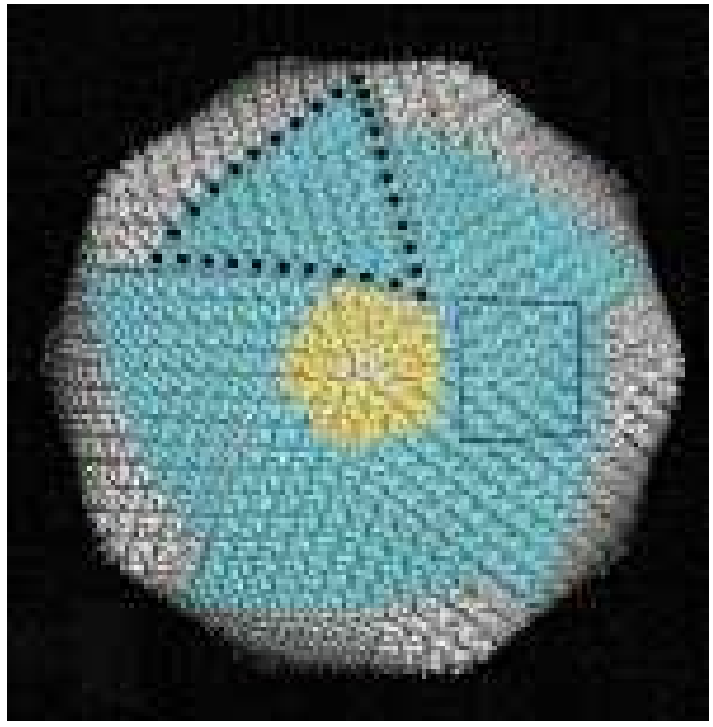


HTLV1 virus,
$T = 36$

Large icosahedral viruses



Archeal virus STIV, $(5, 1)$

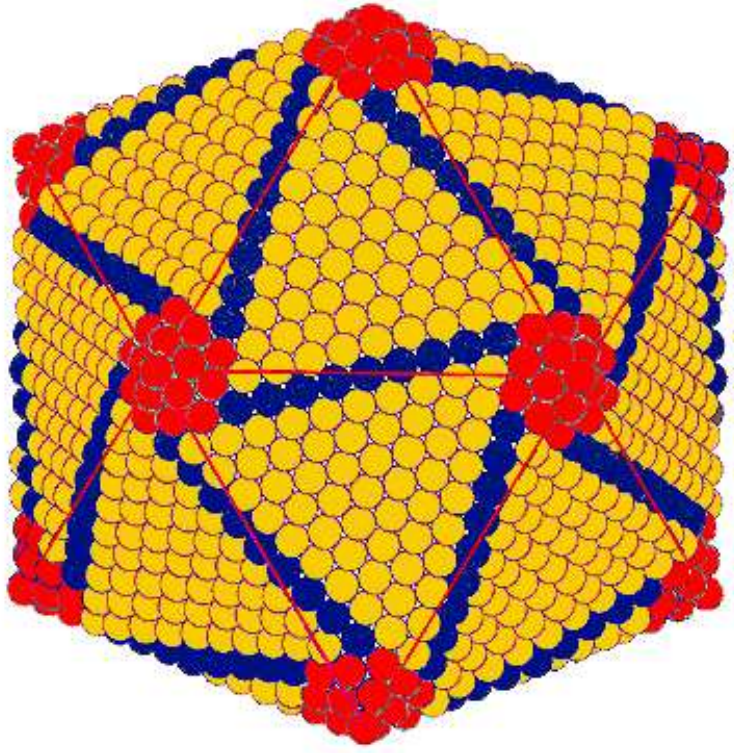


Algal chlorella virus PBCV1 (4th: $\simeq 331.000$ bp), $(13, 0)$

- Sericesthis and Tipula iridescent viruses: $(12, 1)$, $(7, 7)$?
- Phytoplankton virus PpV01: T=219, largest known T.
- Mimivirus (largest known virus): $1078 \leq T \leq 1371$; 1179?

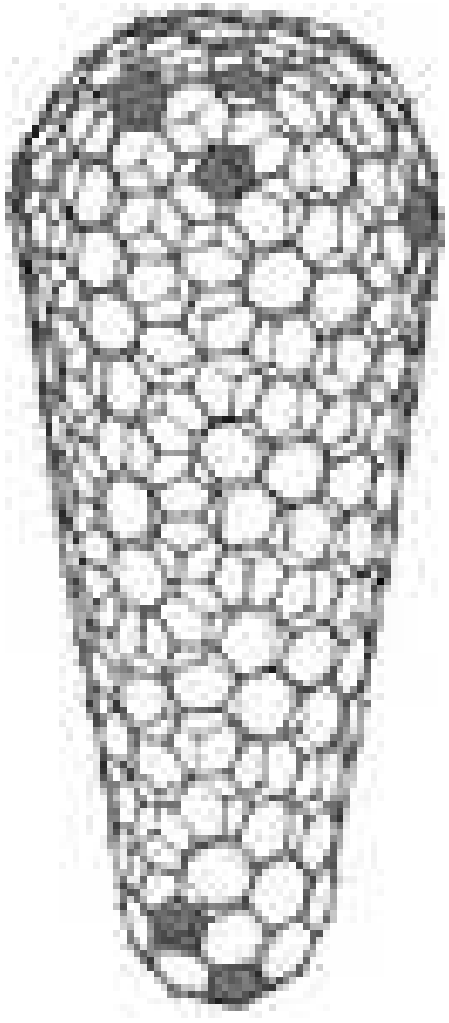
Sericesthis irridescent virus (SIV)

Prolonged storage of SIV in distilled water at 4°C led to the disintegration of virions into 3-, 5- and 2-gonal fragments consisting of 55, 16, and 9 subunits respectively. So, model (Wrigley, 1969) below gives (20×55)+(12×16)+(30×9)= 1562 subunits, i.e. icosahedrality with T=156=10^2+(10×4)+4^2 (still, values 1472 or 1292 subunits are not excluded).

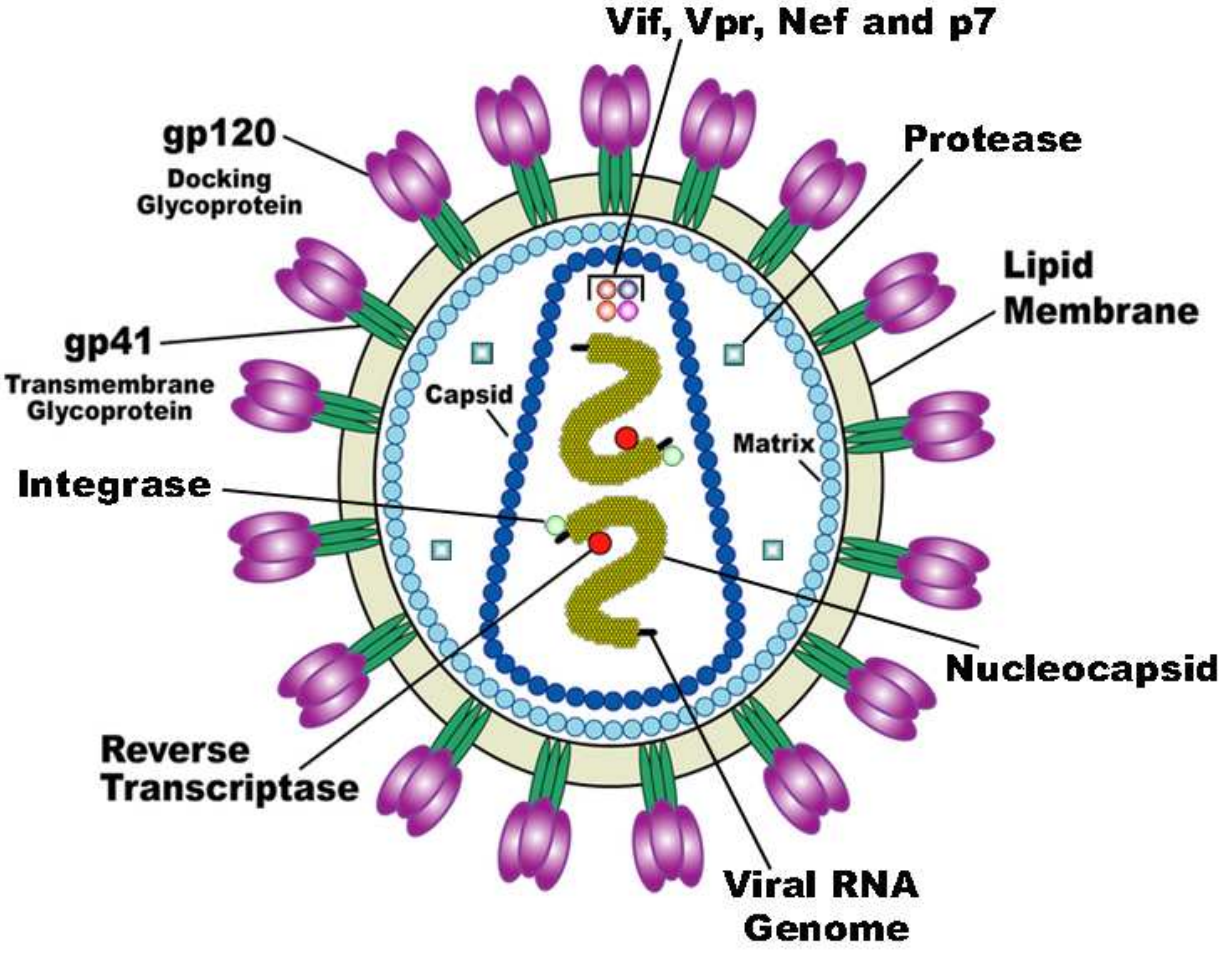


VII. Other shapes

HIV conic fullerene; which $F_n(G)$ it is?



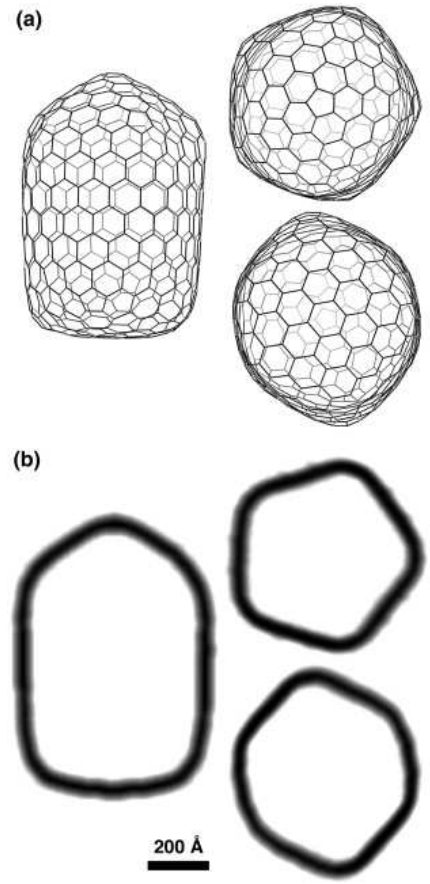
Capsid core
$(7, 5)$



Icosahedral shape (spikes): $T \simeq 71$?

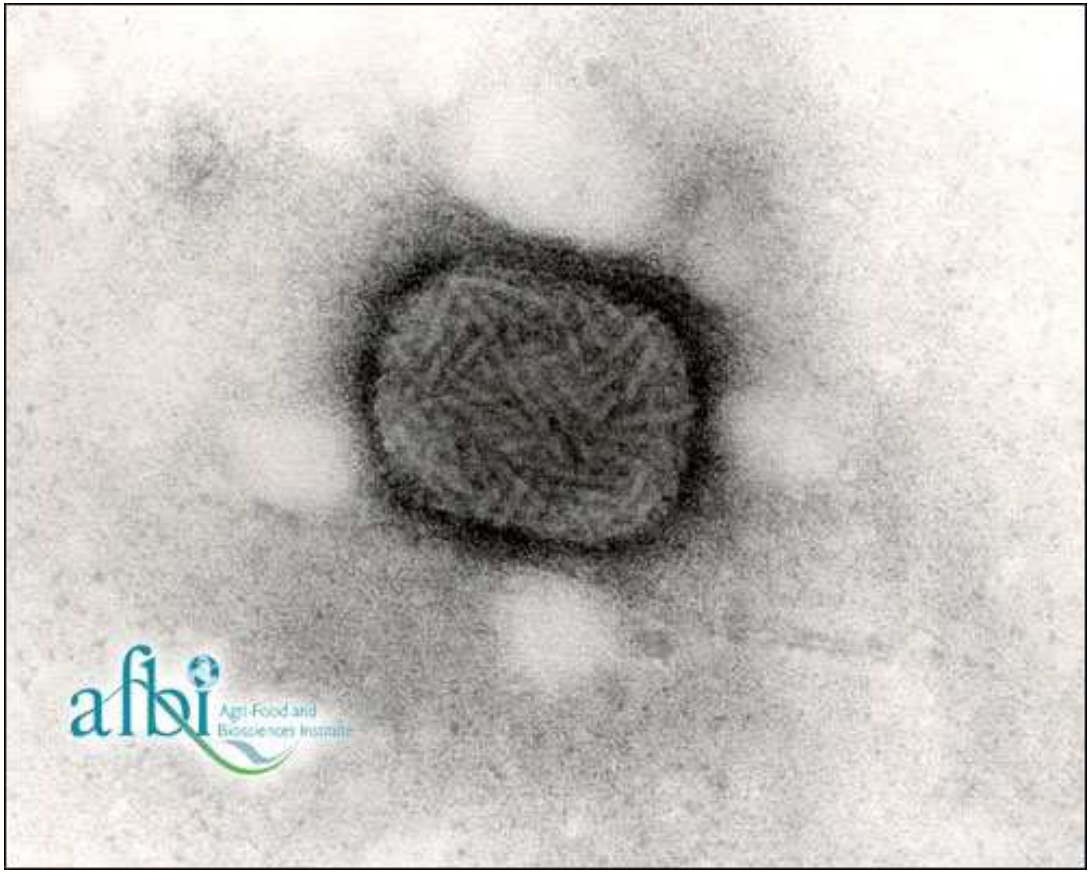
Fullerene quasi-equivalence variations

Mellema et al., 1979: Alfalfa Mosaic Virus ALMV is cylindric nanotube of hexagons with two icosahedral ($T = 1$) caps.
Ganser et al., 1999: HIV capsids are conic $(5, 7)$-fullerenes mainly (but still not visualized at high resolution EM).
Butan et al., 2007: other *retrovirus* (DNA-replicating RNA), avian Rous sarcoma, is in fullerene coffins (5+1, 6+0 caps).



Complex structure: poxviruses

Their large (visible in OM) capsids are brick-shaped or oval.

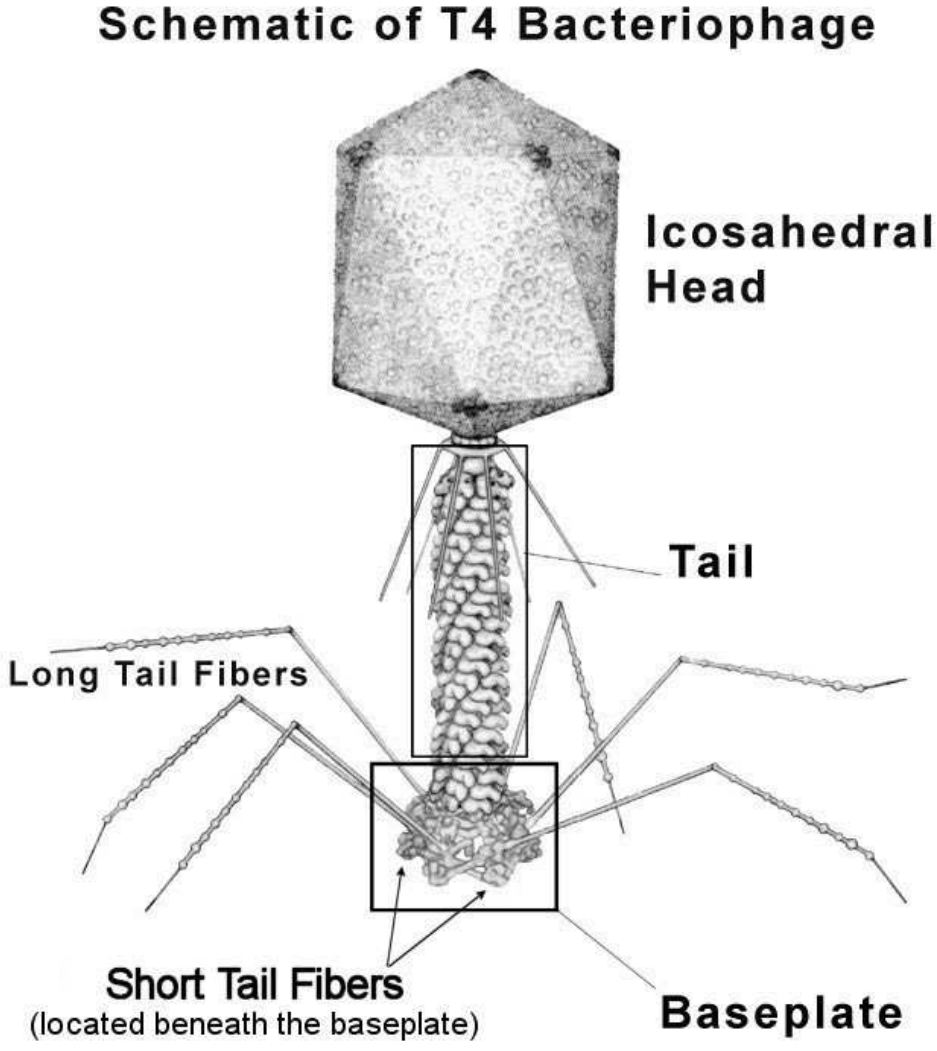


Orthopoxvirus

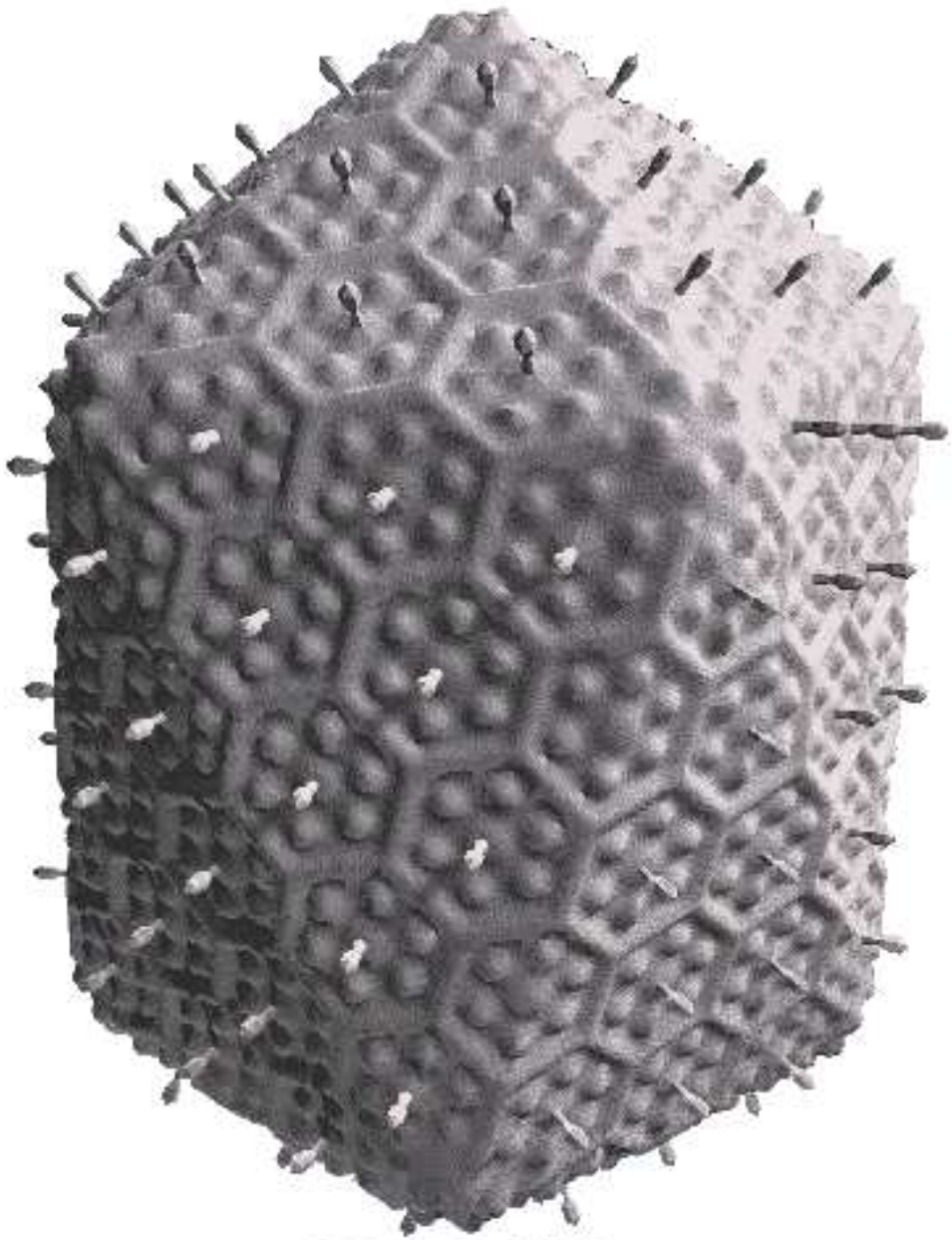
Avian poxvirus

T4: most complex known structure

Large DNA phage T4 has icosahedral (elongated) capsid and helical tail with hexagonal base plate and many fibers.



Prolate-icosahedral capsid head of T4



Geometry of bacteriophage T4

This icosahedral head $+$ helical tail architecture is known only (but common) in *phages* (viruses of bacteria) as the best-sudied tailed phage T4 of *Escherichia coli*.

Such prolate icosahedral capsid has 20 triangular facets:

5+5 equilateral facets on caps, each defined by vector from 0 to (a_1, b_1) and having T_{end}=a_1^2+$a_1 b_1$+b_1^2 basic triangles;

and 10 midsection facets, each defined by vectors from 0 to (a_1, b_1), (a_b, b_2) and having T_{mid}=$a_1 a_2$+$a_1 b_2$+$b_1 b_2$ triangles.

So, $10(T_{end}$+$T_{mid})$ triangles, $30(T_{end}$+$T_{mid})$ protein subunits.

For regular icosahedron, (a_1, b_1)=(a_2, b_2) and T_{end}=T_{mid}.

For phage T4, (a_1, b_1)=$(3, 1)$; so, T_{end}=13.
Fokin et al, 2004 by cryo-EM: T_{mid}=20.

VIII. More on fullerenes

Fullerenes in Organic Chemistry

Carbon C and, possibly, silicium Si are only 4-valent elements producing homoatomic long stable chains or nets

- Graphite sheet: "infinite fullerene", bi-lattice (6^3), Voronoi partition of the hexagonal lattice (A_2).
- Diamond packing: bi-lattice D-complex, α_3-centering of the lattice f.c.c.=A_3.
- Fullerenes: Kroto, Curl, Smalley, 1985 $C_{60}(I_h)$ (or tr. icosahedon, football, Cayley of A_5); Nobel prize 1996 but Ozawa, 1984 (in japanese). "Cheap" C_{60}: 1990. Iijima, 1991: nanotubes (coaxial cylinders). Also isolated chemically by now: C_{70}, C_{76}, C_{78}, C_{82}, C_{84}. If > 100 carbon atoms, they go in concentric layers; if < 20, cage opens for high temperature.
- Full. alloys, stereo org. chemistry, carbon: semi-metal.

Fullerenes as isom. subgraphs of $\frac{1}{2}$-cubes

- All isometric embeddings of skeletons (with $(5R_i, 6R_j)$ of F_n), for I_h- or I-fullerenes or their duals, are:

$$F_{20}(I_h)(5,0) \to \tfrac{1}{2}H_{10} \quad F^*_{20}(I_h)(5,0) \to \tfrac{1}{2}H_6$$
$$F^*_{60}(I_h)(0,3) \to \tfrac{1}{2}H_{10} \quad F_{80}(I_h)(0,4) \to \tfrac{1}{2}H_{22}$$

- (Shpectorov-Marcusani, 2007: all others isometric F_n are 3 below (and number of isometric F^*_n is finite):

$$F_{26}(D_{3h})(-,0) \to \tfrac{1}{2}H_{12}$$
$$F_{40}(T_d)(2,-) \to \tfrac{1}{2}H_{15} \quad F_{44}(T)(2,3) \to \tfrac{1}{2}H_{16}$$
$$F^*_{28}(T_d)(3,0) \to \tfrac{1}{2}H_7 \quad F^*_{36}(D_{6h})(2,-) \to \tfrac{1}{2}H_8$$

- Also, for graphite lattice (infinite fullerene), it holds: (6^3)=$F_\infty(0,6) \to H_\infty, Z_3$ and (3^6)=$F^*_\infty(0,6) \to \frac{1}{2}H_\infty, \frac{1}{2}Z_3$.

Embedding of duals and cell vesicles

The five above embeddable dual fullerenes F_n^* correspond exactly to five special (Katsura's "most uniform") partitions (5^3, $5^2.6$, 5.6^2, 6^3) of n vertices of F_n into 4 *types* by 3 gonalities (5- and 6-gonal) faces incident to each vertex.

- $F_{20}^*(I_h) \rightarrow \frac{1}{2}H_6$ corresponds to $(20,-,-,-)$
- $F_{28}^*(T_d) \rightarrow \frac{1}{2}H_7$ corresponds to $(4,24,-,-)$
- $F_{36}^*(D_{6h}) \rightarrow \frac{1}{2}H_8$ corresponds to $(-,24,12,-)$
- $F_{60}^*(I_h) \rightarrow \frac{1}{2}H_{10}$ corresponds to $(-,-,60,-)$
- $F_{\infty}^* \rightarrow \frac{1}{2}H_{\infty}$ corresponds to $(-,-,-,\infty)$

It turns out, that exactly above 5 fullerenes were identified as clatrin coated vesicles of eukaryote cells (the vitrified cell structures found during cryo-electronic microscopy).

Fullerenes as optimizers on sphere

- To find n unit charged particles on sphere,
  with minimal energy, is Thomson problem,
  with maximal minimum distance, is Tammes problem.
- Almost all optimizers for above problems, in the range $25 \leq n \leq 125$, are fullerenes.
  Then 7-gonal faces appear; if $n > 300$: almost always.
- But Graver, 2005: in all large optimizers, the 5- and 7-gonal faces occurs in 12 distinct clusters, corresponding to a unique underlying fullerene.

Skyrmions and fullerenes

- **Conjecture** (Battye-Sutcliffe, 1997): any minimal energy Skyrmion (baryonic density isosurface for single soliton solution) with *baryonic number* (the number of nucleons) $B \geq 7$ is a fullerene F_{4B-8}.
- **Conjecture** (true for $B < 107$; open from $(b, a) = (1, 4)$): there exist icosahedral fullerene as a minimal energy Skyrmion for any $B = 5(a^2 + ab + b^2) + 2$ with integers $0 \leq b < a$, $gcd(a, b) = 1$ (not any icosahedral Skyrmion has minimal energy).
- Skyrme, 1962 model is a Lagrangian approximating QCD (a gauge theory based on $SU(3)$ group). Skyrmions are special topological solitons used to model baryons.