

# Geometry of the Structure of Viruses

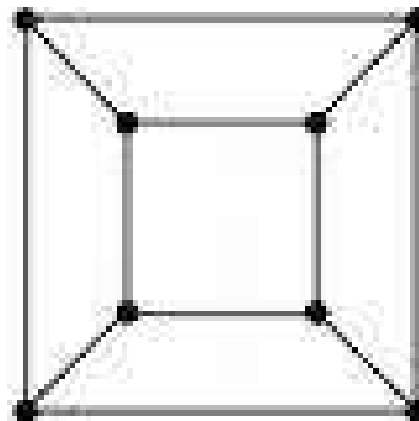
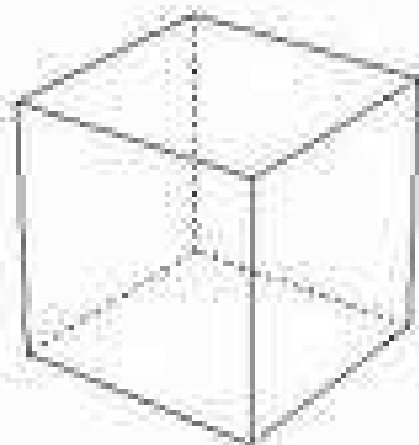
Michel Deza

Ecole Normale Supérieure, Paris

# 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  $\Gamma$ , 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 ..., 1812, ... 214127713, ... **isomers**  $F_n$ , for  $n = 20, 24, 26, 28, 30, \dots, 60, \dots, 200, \dots$ .
- **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) in **Physics**; maximizers, in **Tammes problem**, of minimum distance between  $n$  points on sphere.

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

Conjecture: **FULLERENES**

# Extremal volumes, minimal energy

- **Grace, 1963**, conjectured: all  $f$ -faces ( $f = 12$  or  $f \geq 14$ ) polyhedra of maximal volume, circumscribed about the unit sphere  $\mathbb{S}^2$ , are **fullerenes**.
- **Goldberg, 1969**, conjectured: all  $v'$ -vertices ( $v' = 12$  or  $v' \geq 14$ ) polyhedra of minimal volume, inscribed in  $\mathbb{S}^2$ , are **dual** to above **fullerenes**.
- Given  $n$  points  $\{r_1, \dots, r_n\}$  on  $\mathbb{S}^2$ , the resulting (spherical  $v$ -vertex) Voronoi polyhedron is called **minimal**, if its **energy**  $\sum_{i < j} |r_i - r_j|^{-2}$  is minimal.  
**Pisanski et al., 1996**, conjectured: only a finite number of non-fullerene minimal (spherical) Voronoi polyhedra exist. These cases include  $2 \leq \frac{v}{2} \leq 9, v = 22$  (i.e., when there are no fullerenes) and  $v = 32, 38$ .

# Fullerenes - Isoperimetric Problem

Lhuillier 1782, Steiner 1842, Lindelöf 1869, Steinitz 1927,  
Goldberg 1934, Fejes Tóth 1948, Pólya 1954

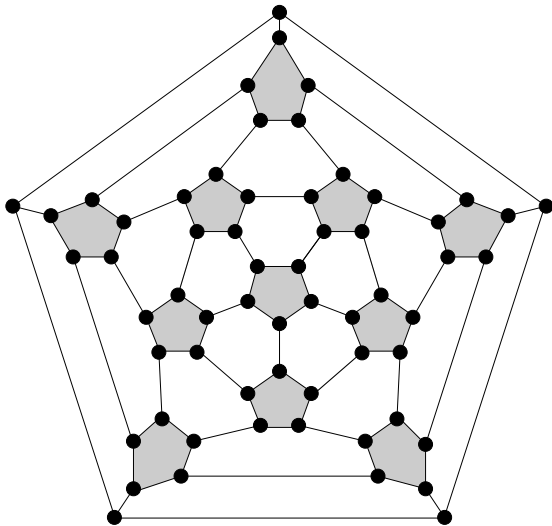
- **Schwarz, 1890**: for **IQ** (Isoperimetric Quotient)  $36\pi \frac{V^2}{S^3}$  of a solid, it holds  $IQ \leq 1$  (with equality only for sphere).
- **Goldberg, 1934**, conjectured: polyhedra with  $f = 12$ ,  $\geq 14$  faces and maximal  $IQ$  are **fullerenes**. For  $f = 4, \dots, 10, 12$ , they are duals of 8 convex deltahedra.
- In **Biology**: ratio  $\frac{V}{S}$  ( $=\frac{r}{3}$  for spherical animal of radius  $r$ ) affects heat gain/loss, nutrient/gas transport into body cells and organism support on its legs.

# II. Icosahedral fullerenes and their duals

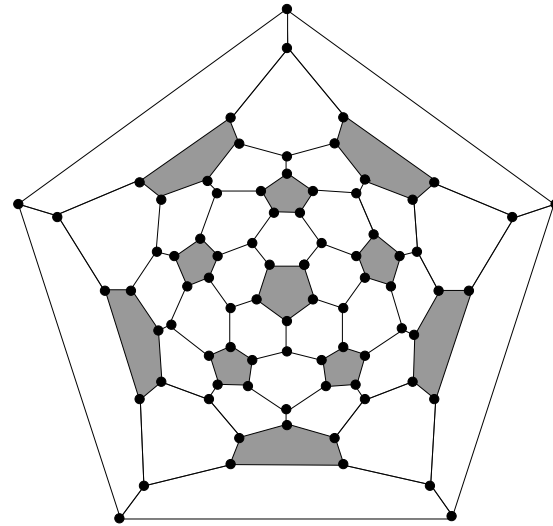
# Icosahedral fullerenes

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.



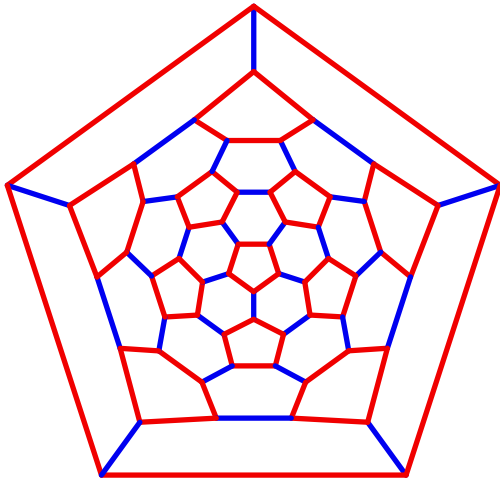
$C_{60}(I_h)=(1, 1)$ -dodecahedron  
**truncated icosahedron**



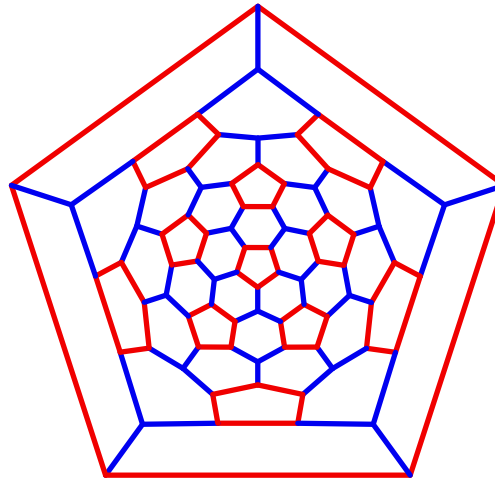
$C_{80}(I_h)=(2, 0)$ -dodecahedron  
**chamfered dodecahedron**

# Small examples

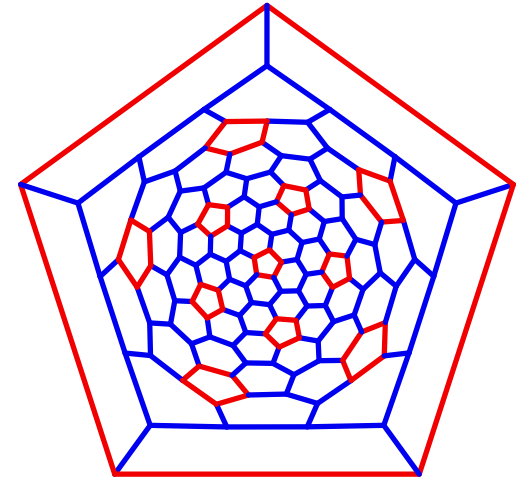
Besides  $F_{20}(I_h)$  with  $T = 1$ , the next smallest examples are:



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



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

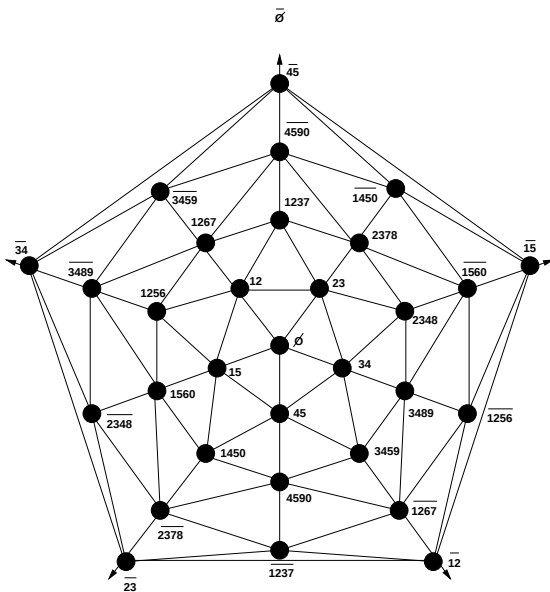


$C_{140}(I)$ : (2, 1)-  
dodecahedron,  
 $T=7$  (*laevo*)

Icosadeltahedron  $C_{20T}^*$ : dual of an icosahedral fullerene.

Icosadeltahedron  $C_{20T}^*$ : dual of an icosahedral fullerene.

- **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**



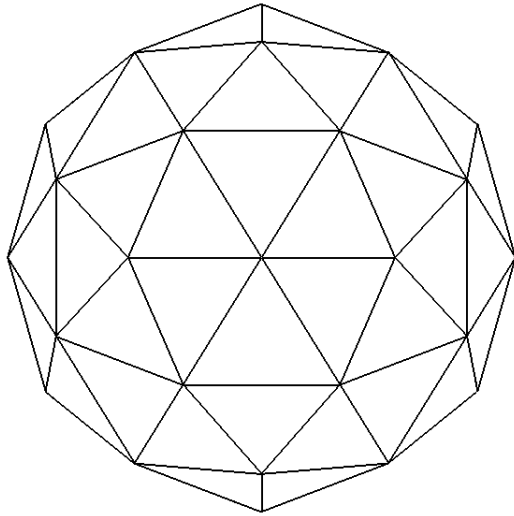
pentakis-dodecahedron



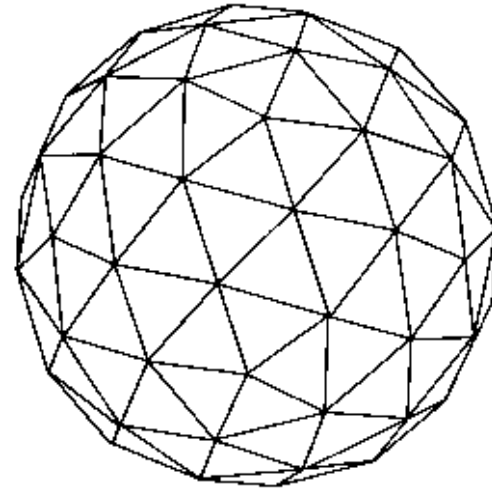
# GRAVIATION (Esher, 1952)

## omnicapped dodecahedron

# Small icosadeltahedra with $a=2$



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



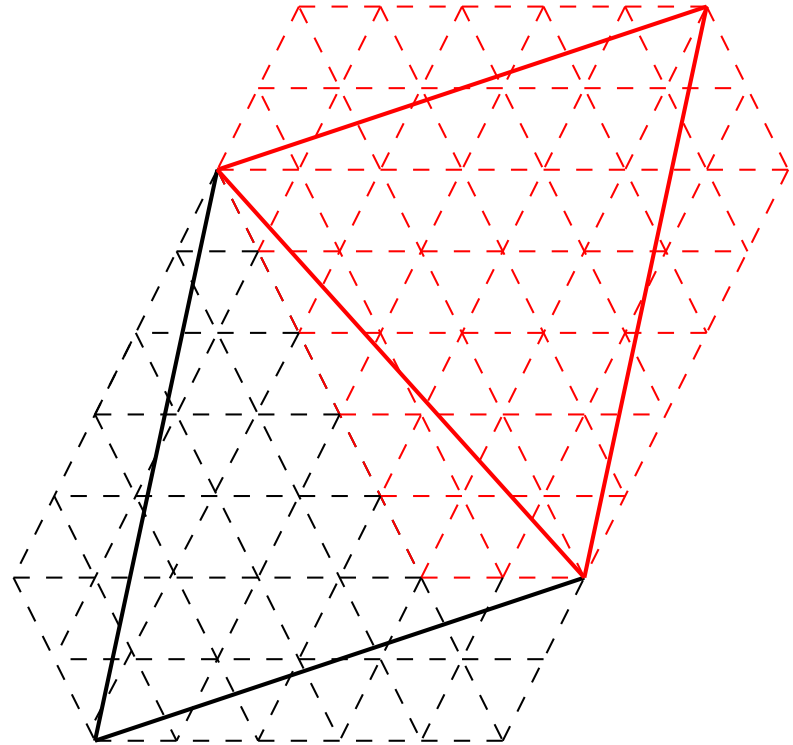
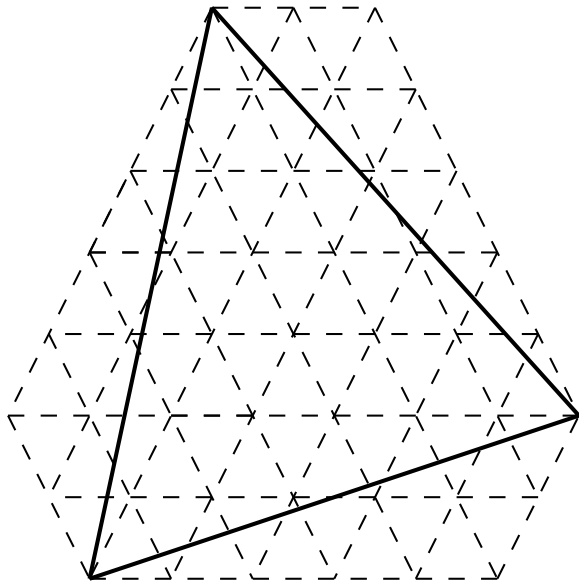
$$C_{140}^*(I), (a, b)=(2, 1)$$

$(a, b)$  (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**.

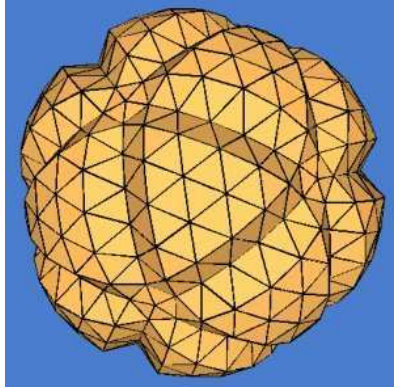
In general,  $T$  not define  $(a, b)$  but  $T=a^2, 3a^2$  imply  $(a, 0), (a, a)$ .  
Caspar-Klug: **classes**  $P=1, P=3$ ; Fuller: **Alternate, Triacon**.

# Goldberg-Coxeter construction

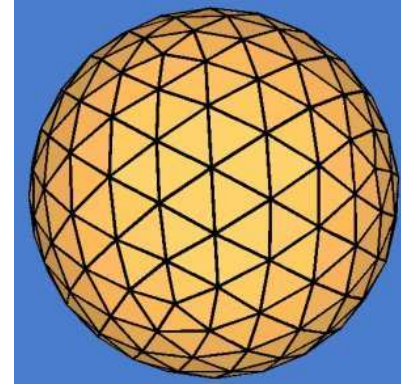
Given  $(a, b)$  ((5, 2) below), put lattice triangle on  $p6$  net  $\{3^6\}$ .  
Gluing pieces coherently gives other triangulation of plane.



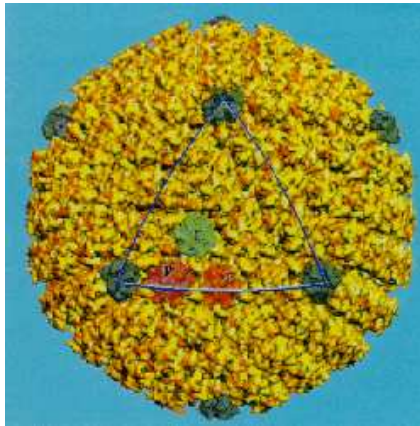
# Icosadeltahedra with $T = a^2$



$(7, 0)$



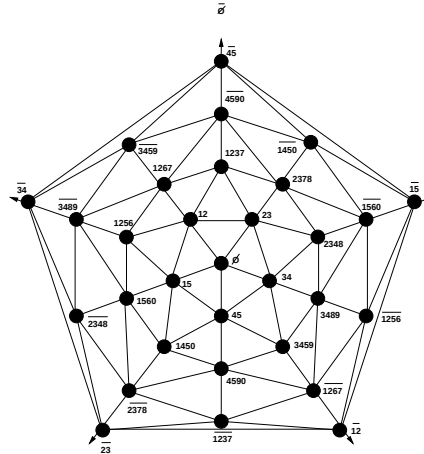
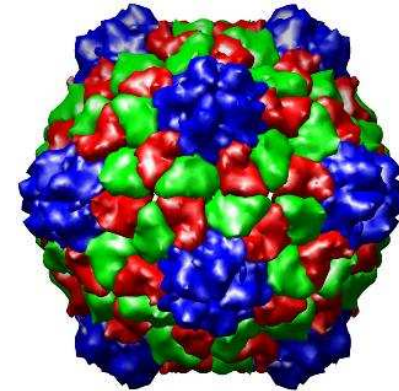
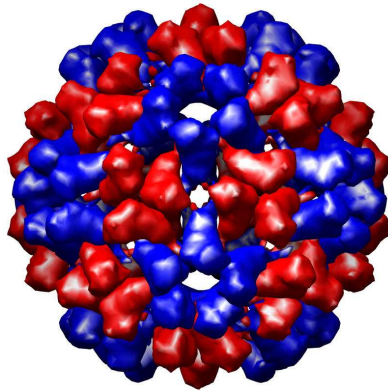
$(5, 0)$



$(4, 0)$ , herpes

# Cowpea mosaic virus CPCM: $T = 3$

Plant *comovirus* infecting cowpea leaf; 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

# Replicons as life fractions

- **life**: DNA and RNA (cells)
- $\frac{1}{2}$ -**life**: DNA or RNA cell parasites: **viruses**, satellites
- **?** Subviral (capsidless, “orphan”) genetic: RNA (viroids) or DNA (plasmids, transposons) or nanobiotech
- **??** Subviral protein replicons (no DNA/RNA): prions

|      | Atom    | DNA        | Cryo-EM    | Prion | Virus capsides       |
|------|---------|------------|------------|-------|----------------------|
| size | 0.2-0.3 | $\simeq 2$ | $\simeq 5$ | 11    | 18 – 90 – 400 – 1500 |
| nm   |         |            |            |       | HIV, Mimi            |

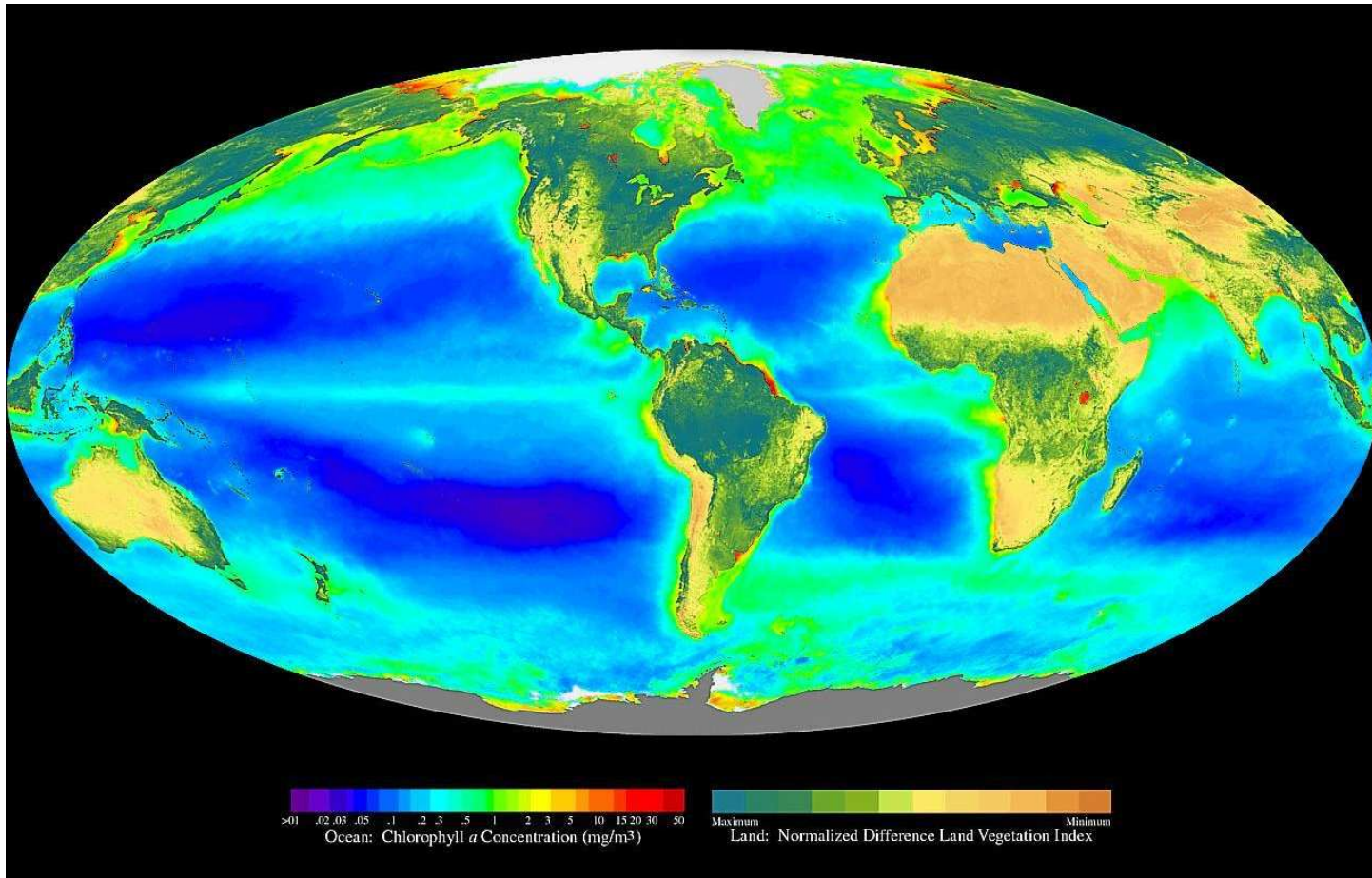
- **Virus**: from Latin *poison*. It has 2 stages: free (dormant transmissible particulate **virion** released by infected cell) and, if lucky, intracellular, active **infectious** stage. Bottle in ocean (extracellular space), then packet of bad news.

# Viruses in numbers

- $\approx 1.8 \times 10^6$  species are known; each has many viruses
- Diameter: 18 nm (Porcine circovirus PCV2) - 1500 nm (Pithovirus sibericum); helical ones: up to 3,000 nm.
- One  $\text{cm}^3$  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 =  $2.5 \times 10^8$  tons ( $4 \times 10^{30}$  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.
- Viruses/cells ratio:  $\approx 10/1$ . Phages kill 20 – 30% bacteria per day.  $\approx 10^{25}$  cells are virus-invaded per second.
- Global population:  $\approx 10^{32}$  viruses in  $\approx 10^8$  genotypes.

# Global primary biosphere 1997-2000

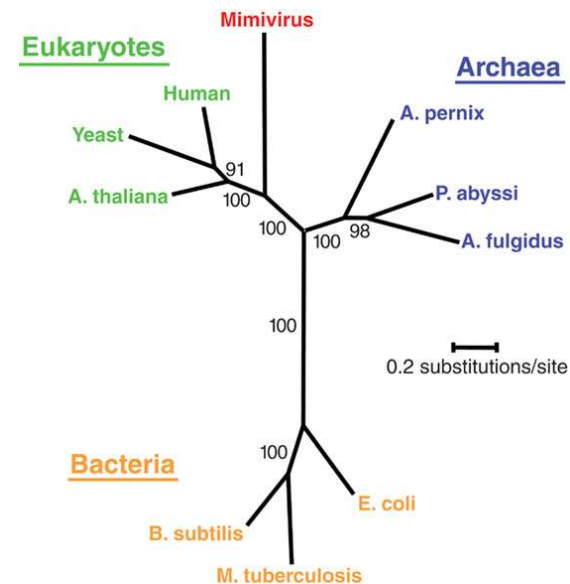
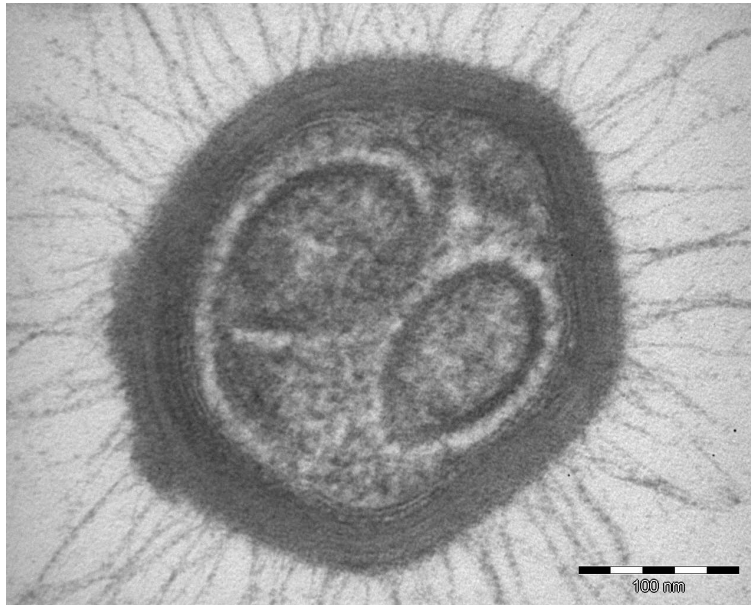
SeaWiFS Project, NASA, satellite-estimated terrestrial vegetation and sea-surface chlorophyll indicating their **primary production** (photosynthesis of organic compounds from atmospheric/aquatic  $CO_2$ ): 56.4 and 48.5 Gt C yearly.



# Viruses in numbers: genome

- The range of their **genome size** (DNA/RNA content) is 3.2 kb DNA (3.5 kb RNA) - 2.5 Mb (Pandoravirus). Also, 3-2556 protein-coding genes. 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 archaea 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 archaeal ones: no homologs for 90% of genes of viruses isolated from boiling acid water.

# Mimivirus and smallest non-viruses



Large capsid: 400 nm,  $\frac{1}{30}$  of its host amoeba, record is  $\frac{1}{10}$ .

Large genome:  $\approx 1.2$  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 theta* 0.55 Mb, *Nanoarchaeum equitans* 0.49 Mb and the record: bacterium *Carsonella ruddii* with genome 0.16 Mb and 182 genes.

# Mimivirus as an infected girus

- Mimivirus is conjectured to be **icosahedral**,  $T \approx 1180$ .
- **Mamavirus** (larger strain of Mimivirus) has a parasite ( $\frac{1}{100}$  of its host) - first **virophage** (harmful satellite) **Sputnik** (also **icosahedral**). Among its 21 genes, 3 are from Mimivirus (is it agent of lateral gene transfer for viruses?), 1 is a protein homologue of an archaeal virus, others: of bacterial and eukaryotic viruses.
- Other virophages are expected on giant viruses.
- **Giruses** (giant viruses) are viruses with size  $\geq 250$  nm and  $\geq 300$  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.
- $\approx 140$  *viral replication cycles* (from virion release till its "progeny") occur yearly. HIV-1 genome size is  $10^4$  bp; its mutation rate is  $3.4 \times 10^{-5}$  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  $\approx (3-30) \times 10^3$  nucleotid bits and high **mutation rate**  $10^{-3} - 10^{-5}$  substitutions per bit, i.e.  $\geq 1$  per replication cycle. **Diversity** (genome divergence) of HIV virus 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 mutations. **Pleiotropy**  $\rightarrow$  complex organisms are much slower.
- Newly arising mutants form a **quasispecies**, i.e., a swarm, cloud of related genomes acting a whole. 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?
- Prions: also swarms evolving via folding "mutations".

# 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 killed 99.8% in 3 months.
- 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 virus strains of lowered virulence and immune rabbits. The same going on for HIV-1 and us (HLA-B gene)?
- 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.
- Problem: **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?

# Viral niche construction

- Viruses have rapid evolutionary change, with small genomes, massive population sizes and error-prone replication (esp. RNA viruses) allowing for fast and relatively wide exploration of sequence space.
- The organismal virome includes weakly pathogenic, commensal (no harm/help), or even mutualistic viruses.
- Exp.: some parvoviruses contribute to host aphid survival by creating winged morphs that aid in dispersal to other plants. GB virus C inhibits HIV's replication.
- Niche construction (say, soil's turnover/consumption by earthworms) generalises standard evolutionary theory.
- Exp.: envelope proteins of endogenous retroviruses help to develop mammalian placentas to allow their vertical (parents-offspring) transmission. Seoul virus increases aggression in male Norway rats and is spread through bite wounds via saliva; same for rabies

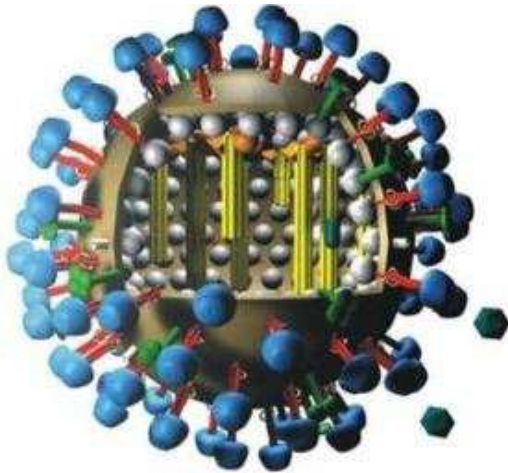
# Influenza A quasispecies

- 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**.

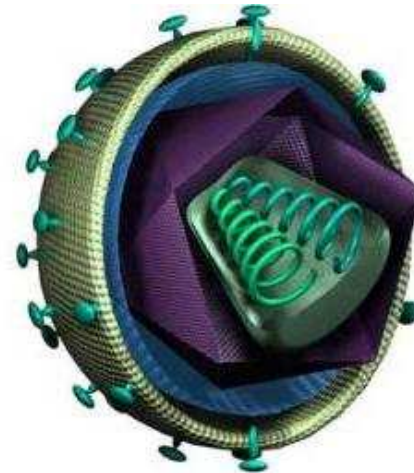
# Swine flu H1N1 2009-2010 pandemic

- 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 leukocytes into the lung tissue. People after 65 are victims:  $\approx 12\%$  while  $\approx 90\%$  for seasonal flu.
- $\approx 6$  times as many as normal ( $> \frac{1}{3}$  of population) are expected to get this flu. Safe vaccine exists.
- From November 2009 swine flu become predominant flu strain worldwide. Aboriginals in Australia are esp. hit.

# Swine flu H1N1 2009-2010 pandemic



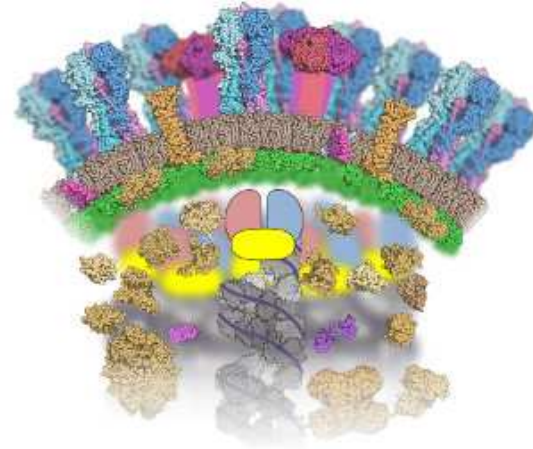
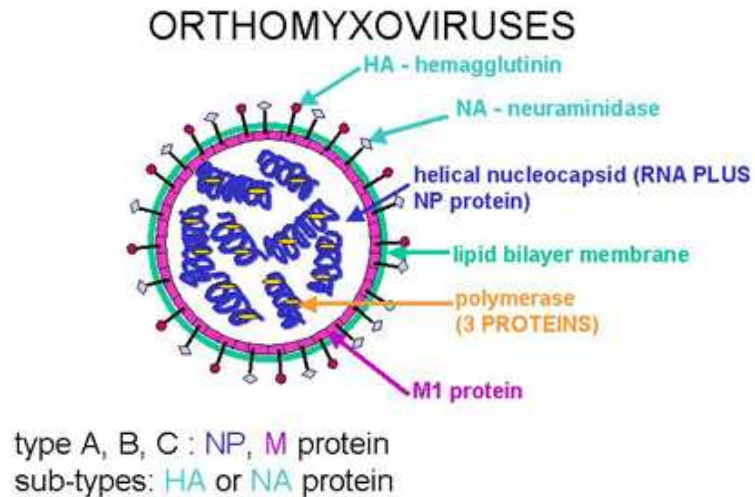
Swine flu



HIV

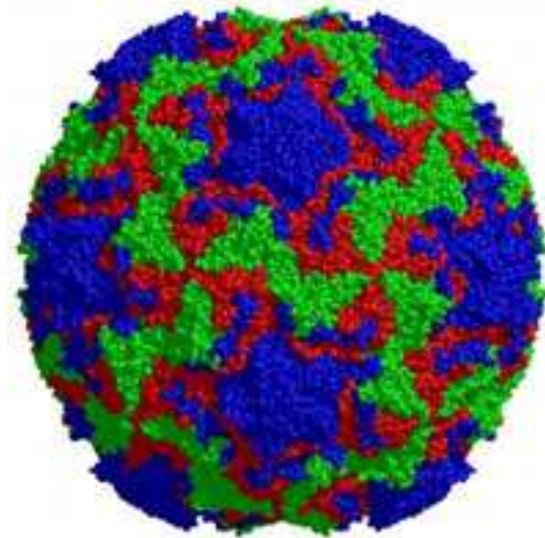
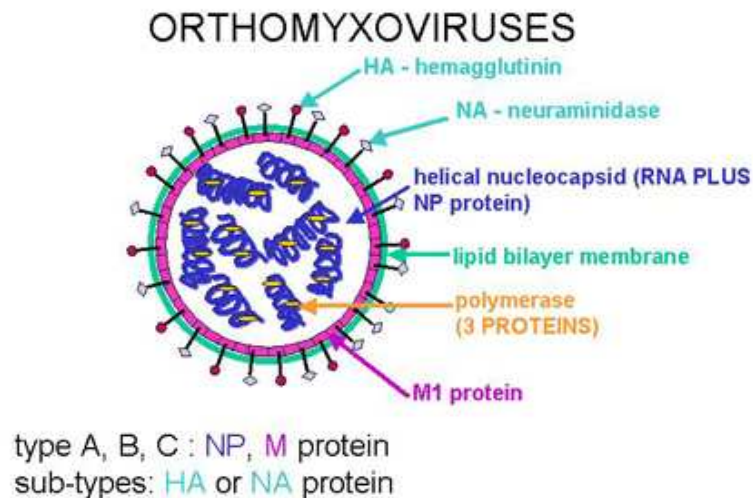
- Infection is deep in lungs (as killer H5N1), not in throat. Signs of severe form: shortness of breath, chest pain, blue lips. Only treatment: *oseltamivir* (Tamiflu).
- Most fatal cases: co-infection with bacteria and asthma, obesity. Biggest danger: reassortement with H5N1.
- In Europe, this flu come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

# Anatomy of a virion



Hutchinson: above cross-section of an influenza virion shows the locations and relative abundance of viral proteins (brightly colored) and host membrane and proteins (brown). Even uninfected host cells produce virion-like particles.

# 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.

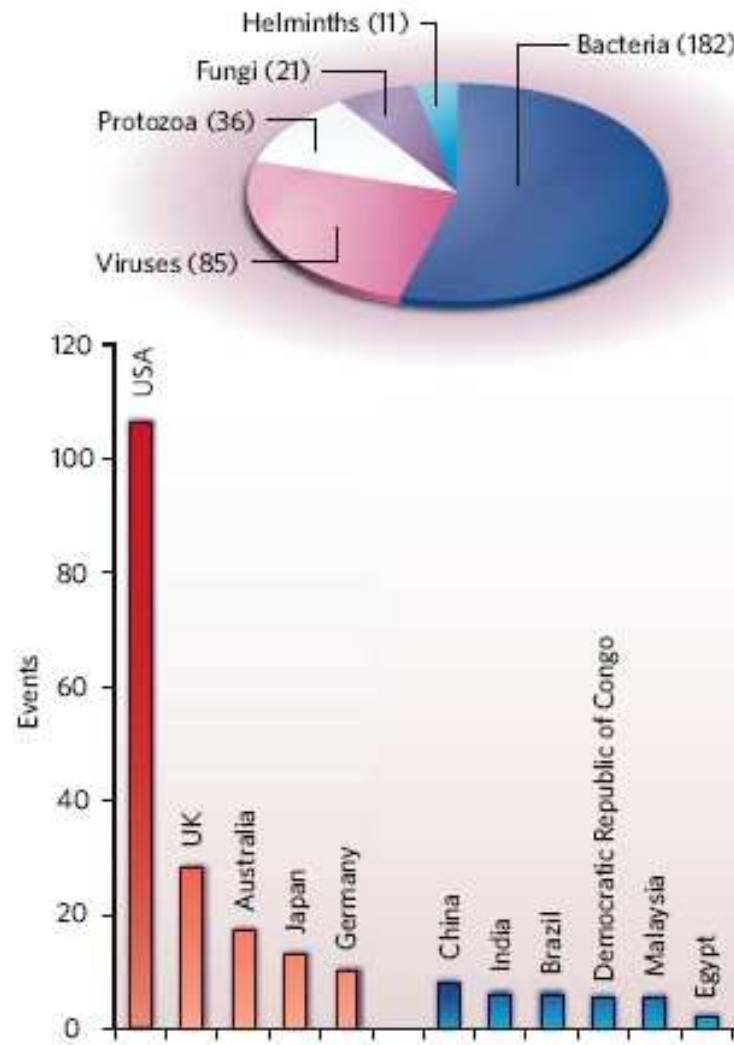
**Ligget et al., 2009**, found genetic codes for all 189 known strains and organized them into 15 sub-groups coming from distant ancestors. It explains cold uncurability at present.

# 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 mainly) cancer cells.  
Exp.: a natural **picornavirus** Seneca Valley Virus-001. infects only cancer cells in small cell lung cancer.
- **Virotherapy**: engineering oncolytic viruses by genetic modification of some natural viruses.  
OncoVEX GM-CSF is now in phase III for advanced melanoma, and likely to become 1st approved oncolytic agent in the western world. Poxvirus JX-594 is now in phases II. Next to modify: poliovirus, reovirus.
- **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 (lose many genes but remainder become permanent functional addition to host DNA). 10 – 20% 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.

# Origin of viruses: three hypotheses

First were RNA molecules: self-replicators with protein production and enzymic activity. Evolution worked via HZT (horizontal gene transfert) only. Viroids - survivors of it.

- **Virus-first**: capsids arose before LUCA (Last Universal Cellular Ancestor) and were not parasites. In fact, capsid protein found in most viruses, but not in cells. 50–90% of any virus genes are not known elsewhere.
- **Vagrancy, cell-first**: small viruses (as human hepatitis  $\delta$ ) derived from cell's escaping RNA or/and DNA transferrable fragments (as plasmids and transposons).
- **Degeneracy, cell-first**: larger ones (as Mimivirus) come by reductive evolution from cells-parasiting cells.

All 3 ways are complementary, esp. in pre-LUCA RNA world. Halo of viral “lianes” wrap around cellular tree of 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 upgrades cell's photosystem to keep it alive;  $\approx 5\%$  of world's oxygen results.

# 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.
- Viruses react on their environment (but not on stimuli), have no metabolism and homeostasis but mutate and reproduce (by self-assembly as crystals) inside a cell.
- They can be seen alive, when transforming a **ribocell**, encoding *ribosomes* (cell's protein builders) and dividing, into a **virocell**, producing virions.  
Major domains of life are 3 ribosome-encoding (cells) and some capsid-encoding (viruses).
- Virus-like entity could mediate Earth transition to a biotic world. They promoted the isolation of protocells. 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

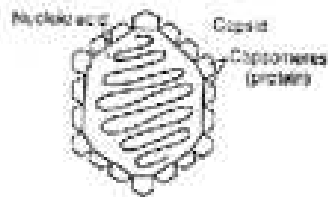
# Symmetry of viruses

- **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.

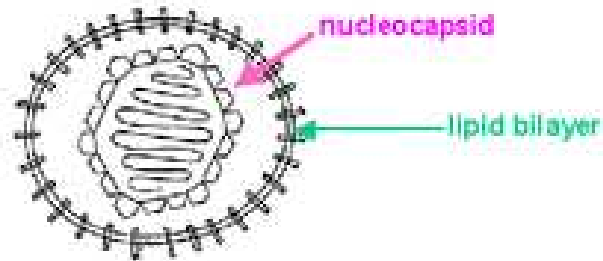
# Symmetry of viruses

## 5 BASIC TYPES OF VIRAL SYMMETRY

icosahedral nucleocapsid



ICOSAHERAL

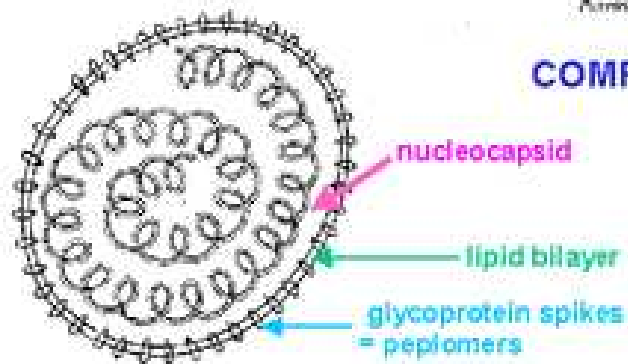


ENVELOPED ICOSAHERAL

helical nucleocapsid



HELICAL



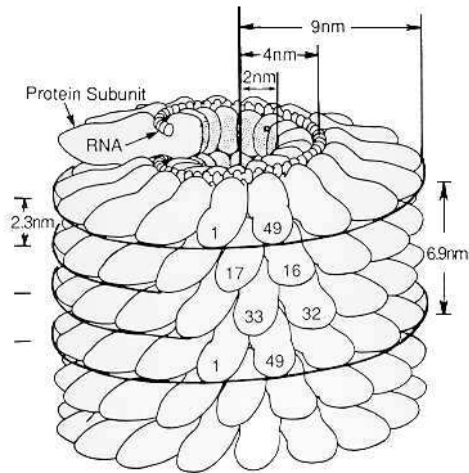
ENVELOPED HELICAL



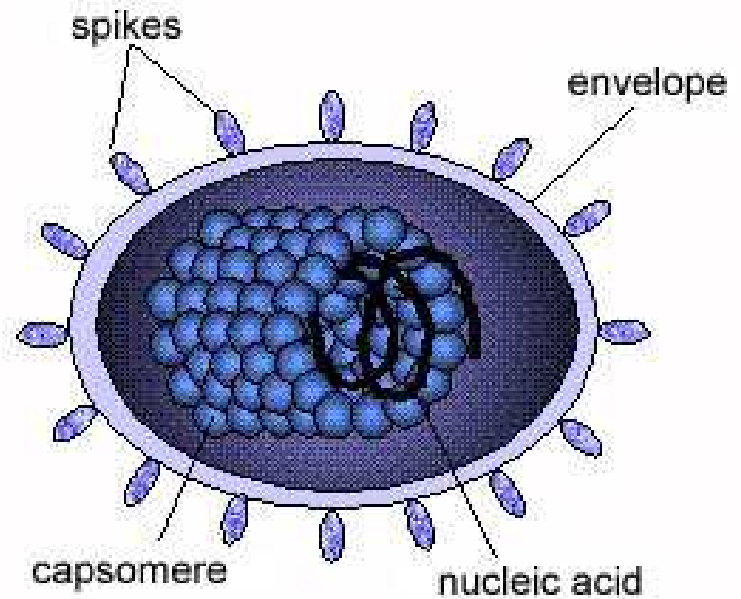
COMPLEX

Adapted from Schaeter et al., Mechanisms of Microbial Disease

# Helical viruses



Tobacco mosaic virus

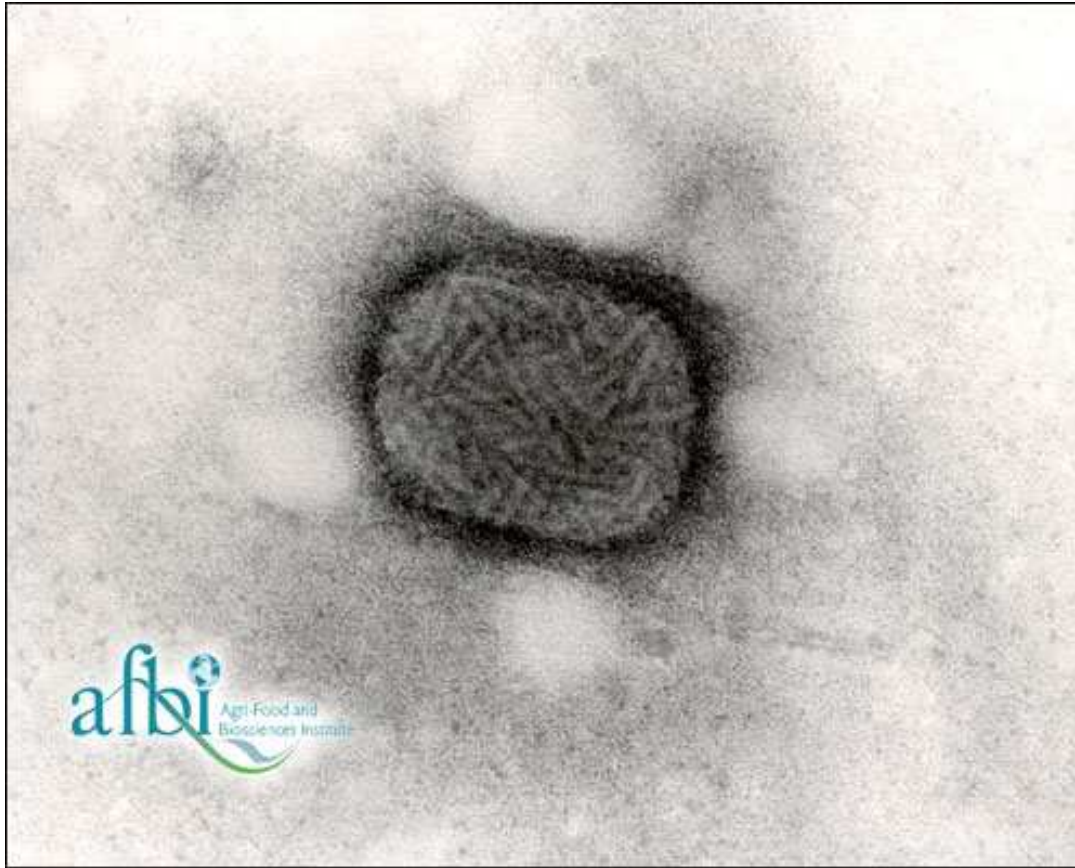


Enveloped helical virus

- Helical viruses are defined by their **amplitude** (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

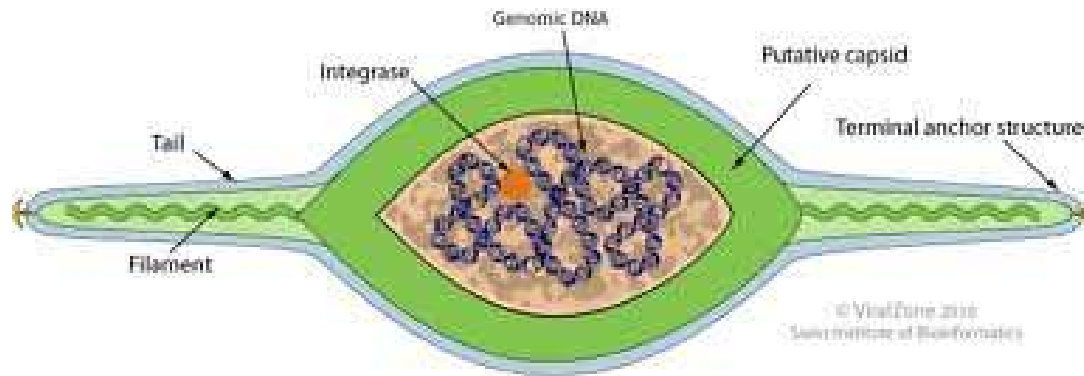


Avian poxvirus

Poxviruses are most large and complex DNA viruses; they also survive years outside the host.

# Complex structure: bicaudaviridae

Their shape is a spindle (lemon), a morphology unique to the archaeal virions.



Above *acidianus two-tailed virus* of a hyperthermophilic archaeon growing in acidic hot springs, is unique known virus having a structural transition **outside of the host cell**: at  $75^{\circ} - 90^{\circ}C$  it develop long tails at each end of “spingle”. These tail will help to infect another host.

# 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  $\beta$ -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.

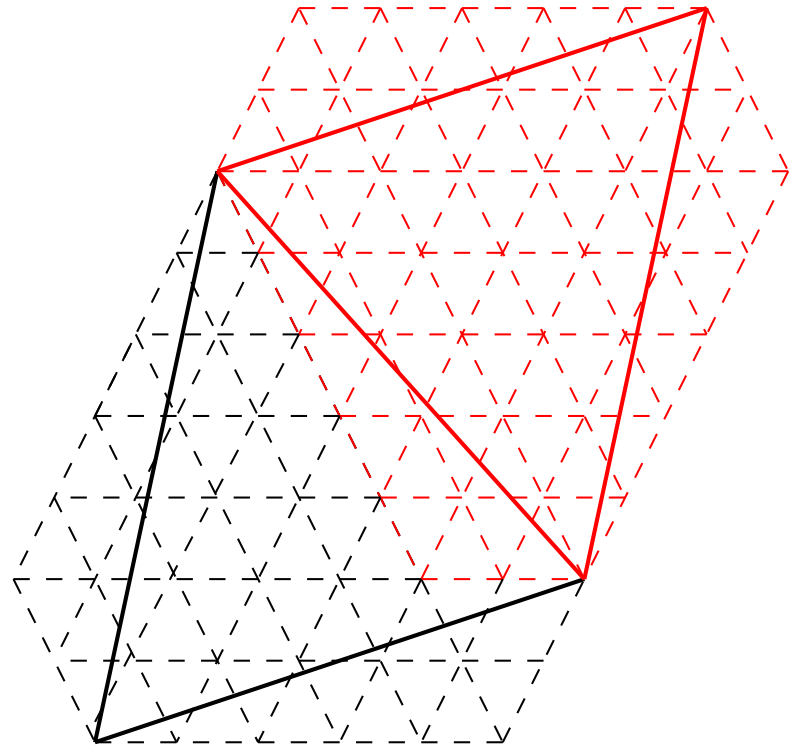
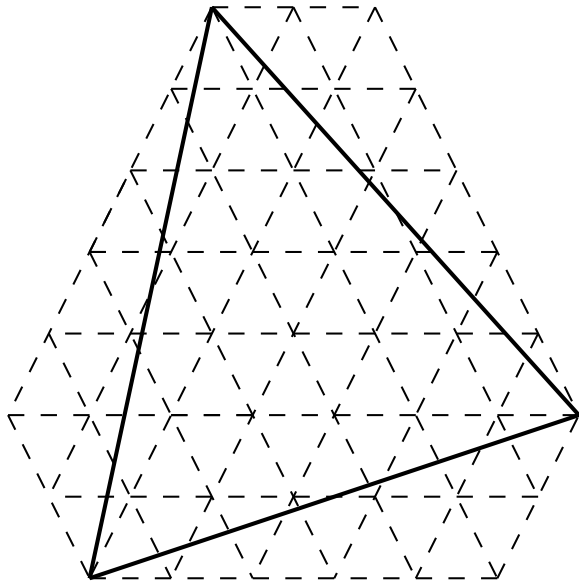
# VI. Icosahedral viruses

# 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  $\leq 4$  nm joined by *bonds* (protein/protein interaction) to form virus shell.
- Bonds are flexible:  $\simeq 5^\circ$  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 identical environment) subunits.
- But origin, energy, thermodynamics 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 lattice triangle on  $p6$  net  $\{3^6\}$  and replace its 6-valent vertices by pentamers. Gluing pieces coherently 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  $60 \times 3$  protomers clustered as: 12 pentamers plus  $(10 \times 3)-10$  hexamers (**Turnip Yellow Mosaic virus**);  $20 \times 3$  trimers (**Poliovirus**), and  $30 \times 3$  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-equivalent 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.

# Jalasvuori T-number formula

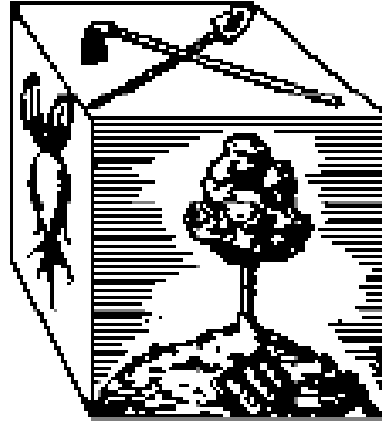
- **Jalasvuori, 2010:** if two icosahedral viruses have similar major proteins and their genomes are packed into the same density, then  $\frac{T_1}{T_2} = \left(\frac{G_1}{G_2}\right)^{\frac{2}{3}}$ , where  $T_1, T_2$  and  $G_1, G_2$  are their triangulation numbers and genome lengths.
- He claims: icosahedral viruses (with unique double  $\beta$ -folded protein and those of thermophilic/halophilic archea) form viral lineages from pre-LUCA life.
- Above *Crenarchea* viruses have unusually diverse morphology. Because, perhaps, better survival of their hosts during global catastrophes in early biosphere.
- He claims also: majority of mutations, incl. isolation of proto-cells, are to fight virus infections.  $\approx 75\%$  of gene diversity is located outside of cells. The cellular life is surrounded by rapidly evolving/attacking viral halo.

# “Most diseases come from icosahedra”

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



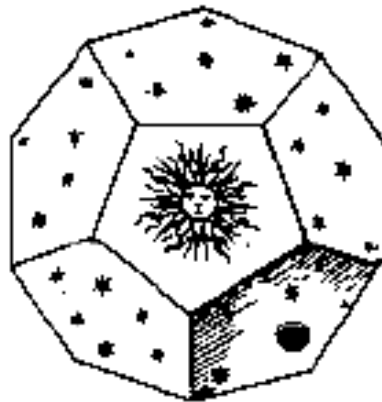
OCTAHEDRON  
Air



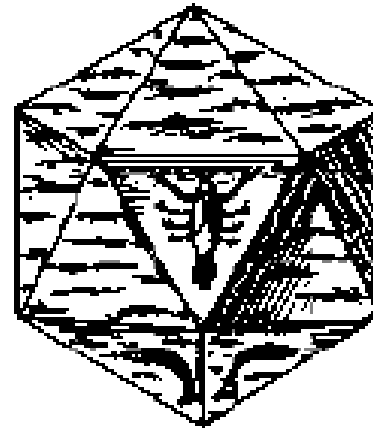
CUBE  
Earth



TETRAHEDRON  
Fire



DODECAHEDRON  
the Universe

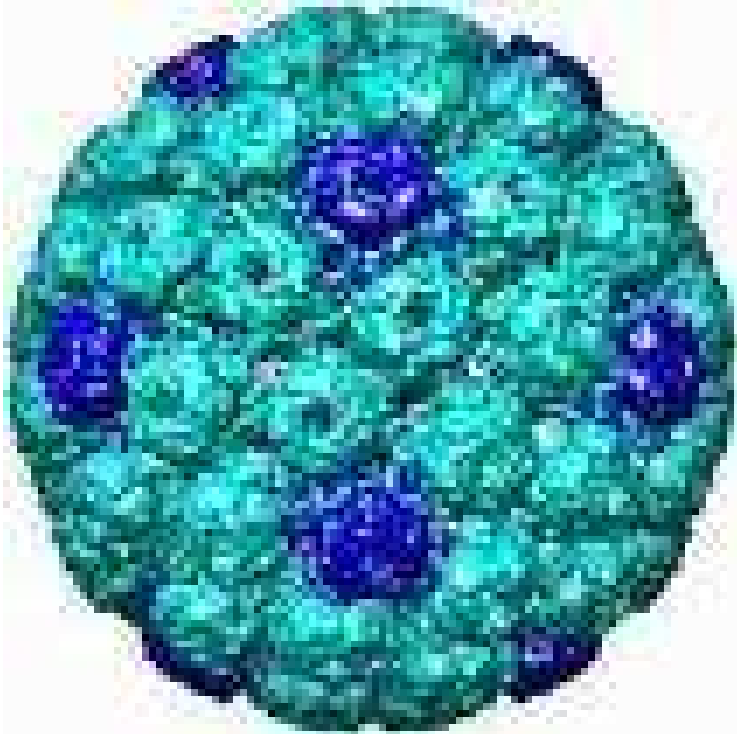


ICOSAHDEDRON  
Water

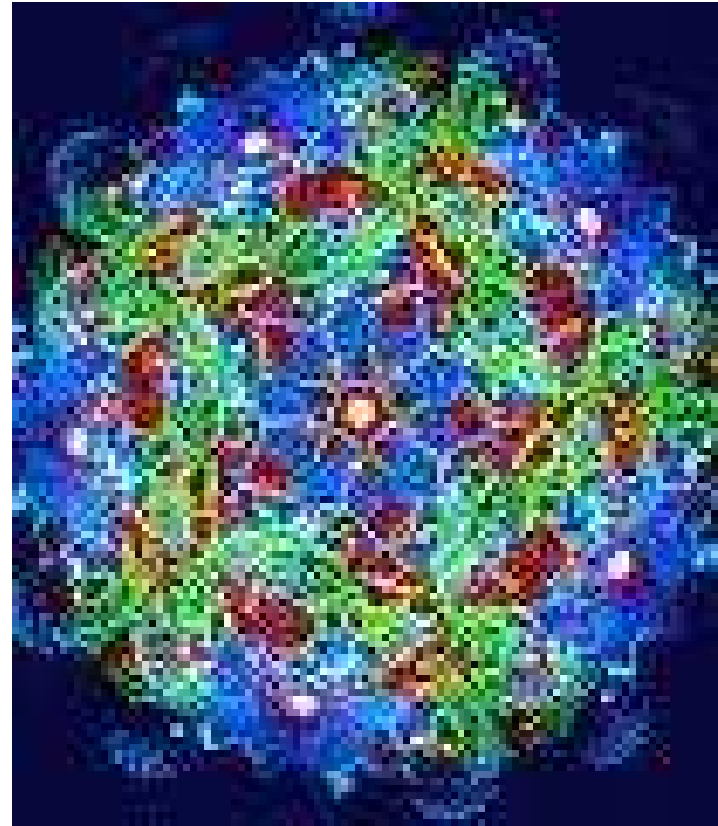
# Capsids of icosahedral viruses

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

# Examples



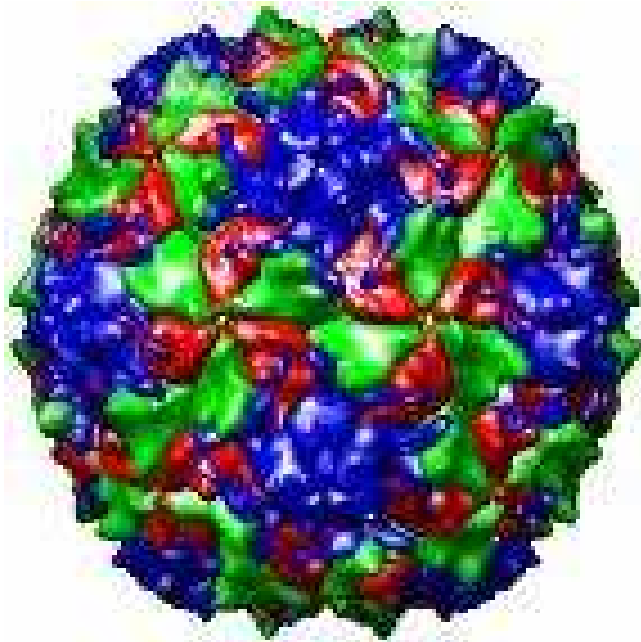
**Satellite** STMV,  $T=1$ , of TMV,  
helical Tobacco Mosaic virus  
1st found (**Ivanovski, 1892**,  
and **Beijerinck, 1899**),  
1st seen (**Stanley, 1931**) EM



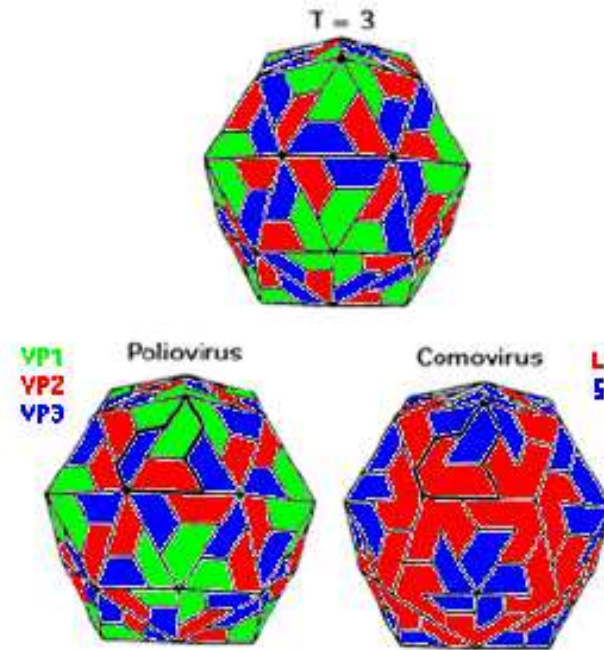
Foot-and-Mouth virus,  
 $T = 3$

# 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.

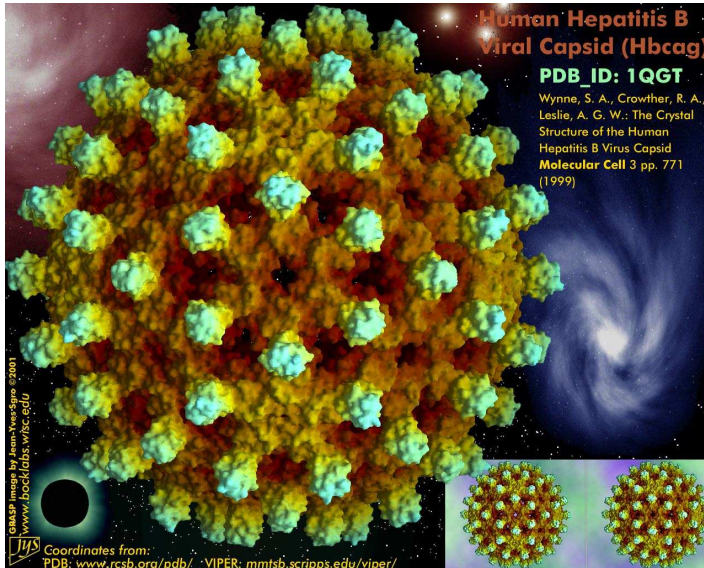


Poliovirus

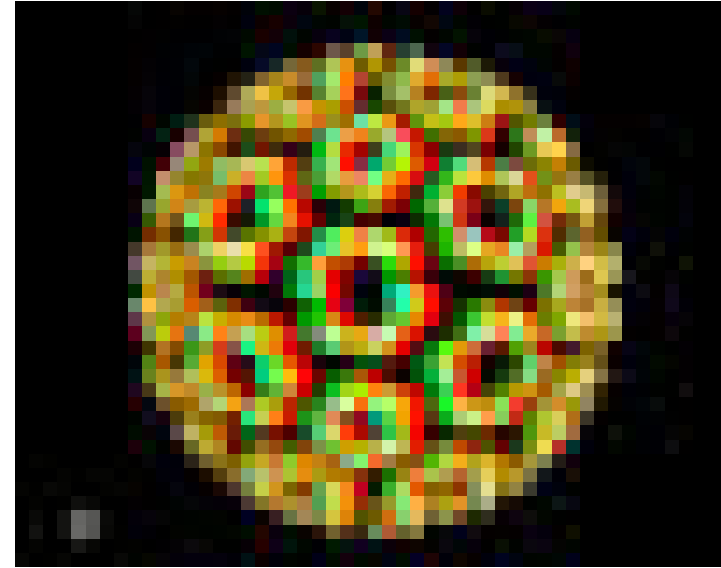


Polio and BPMV viruses

# Viruses with $T = 4$



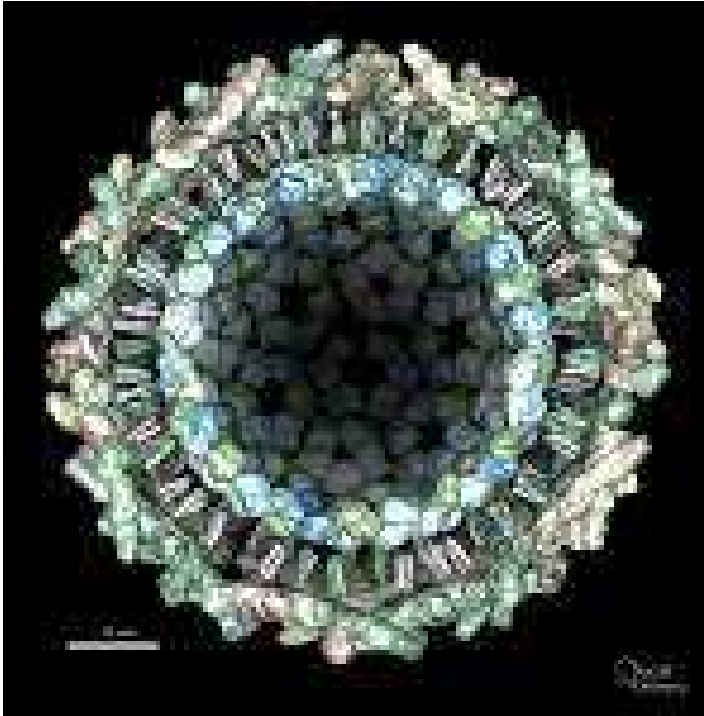
Human hepatitis B



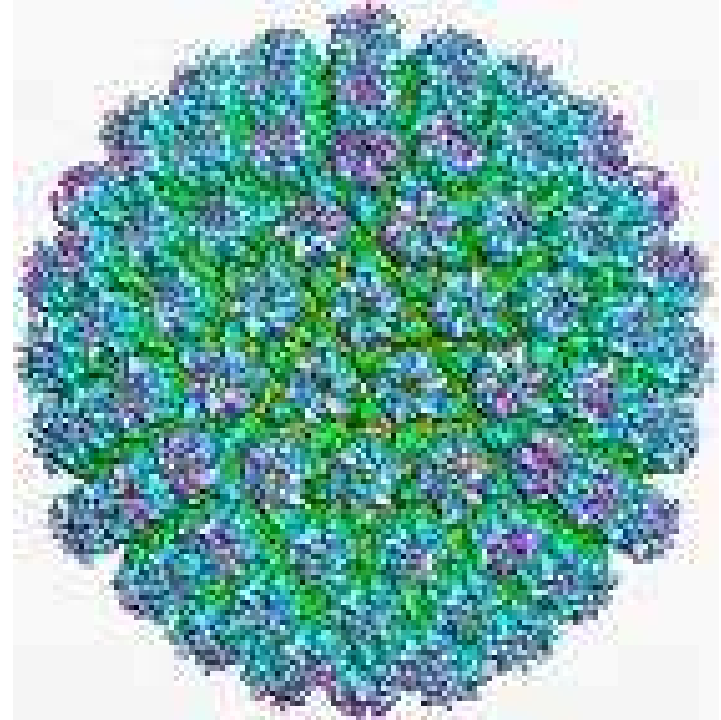
Semliki Forest virus

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,  
 $T = 4$

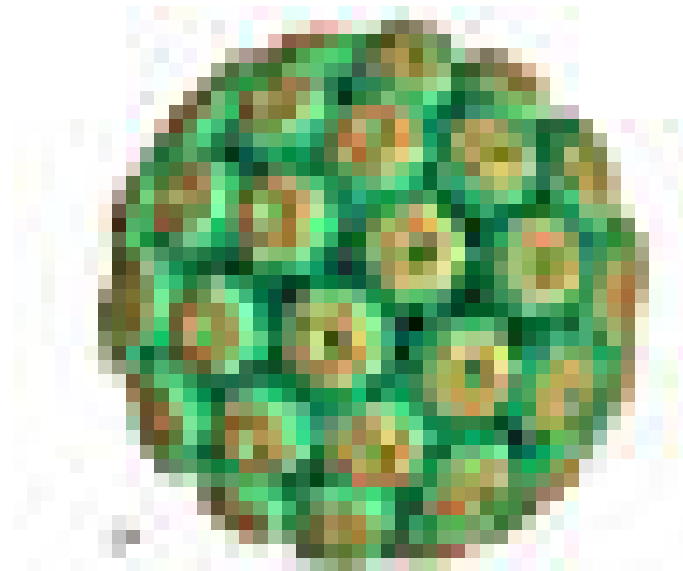


Herpes virus,  
 $T = 16$

# 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.

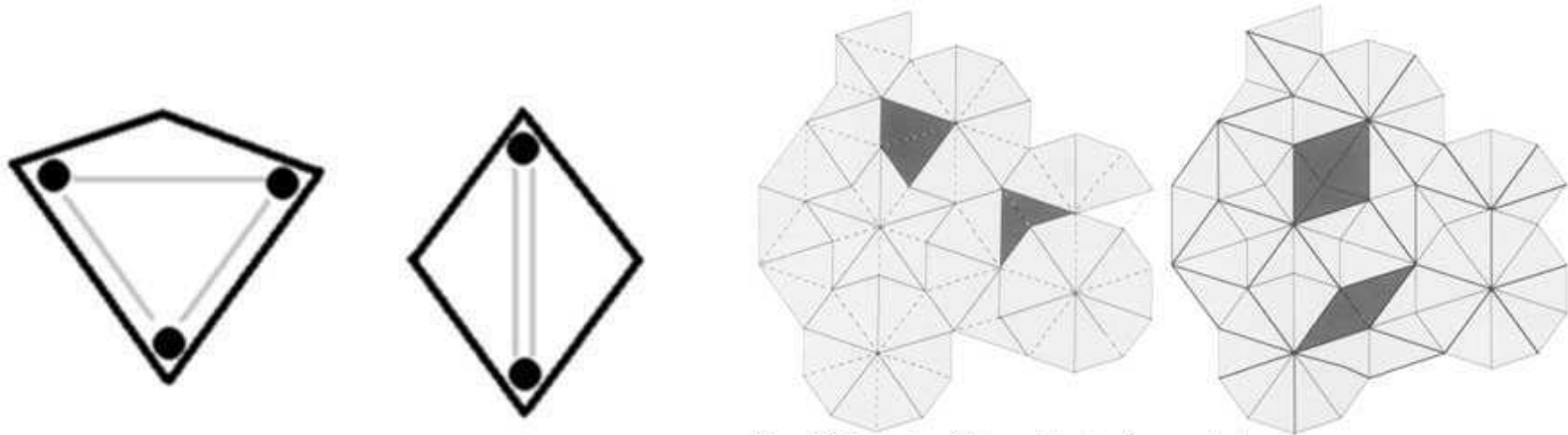
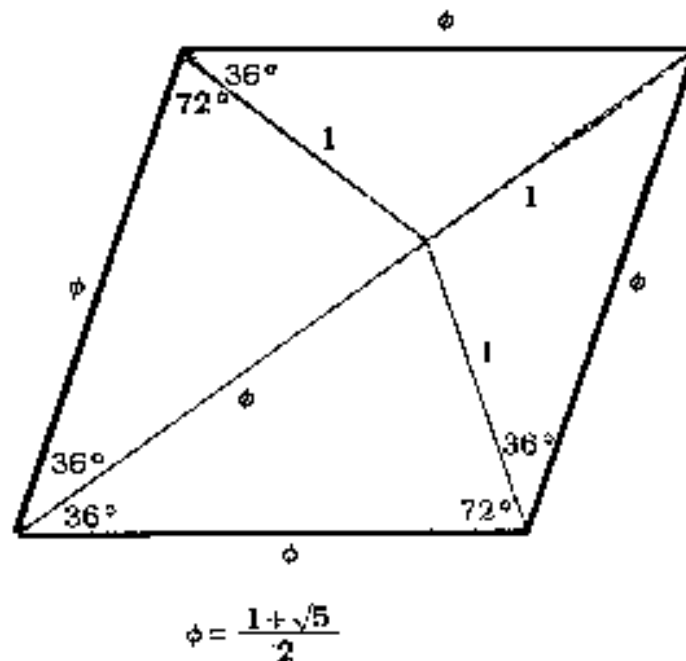
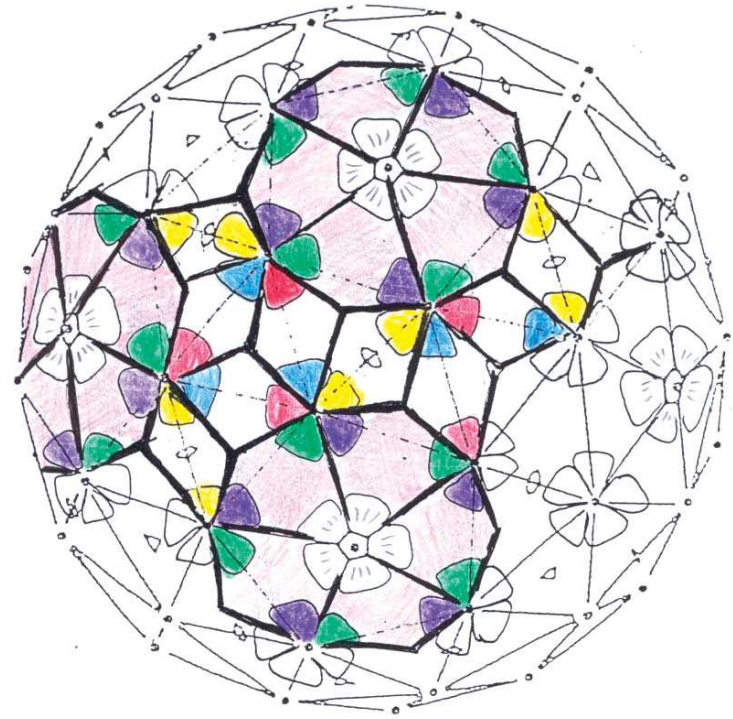


Figure 3. Translation of kites and darts to Penrose rhombs



# 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  $H_3=I_h$ .

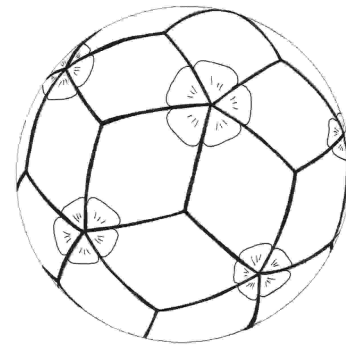
Tiles are rhombs and kites. All corners of the tiles meeting at 5-valent vertices mark the locations of protein subunits, i.e., exactly at tile corners subtending the same angle  $72^\circ$ .

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

- Checkerboard root lattice  $\mathbb{D}_n = \{x \in \mathbb{Z}^n : \sum_{i=1}^n x_i \text{ is even}\}$ .
- 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  $\alpha$  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  $\alpha$ )  $r_\alpha : x \rightarrow x - \frac{2\langle x, \alpha \rangle}{\langle \alpha, \alpha \rangle}$  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)  $\alpha$ . 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: simple cubic  $\mathbb{Z}_6$ , its half, face-centered cubic  $\mathbb{D}_6$ , and body-centered cubic  $\{x \in \mathbb{Z}^6 : x_i \equiv x_j \pmod{2} \text{ for all } 1 \leq i, j \leq 6\}$ .
- Embedding of noncrystallographic Coxeter groups into crystallographic ones:  $H_2 \rightarrow A_4$ ,  $H_3 \rightarrow D_6$ ,  $H_4 \rightarrow E_8$ .

# 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  $a_1, \dots, a_6$  are linearly independent over  $\mathbb{Q}$ ; so, the coordinates  $n_1(x), \dots, n_6(x)$  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  $\tau = \frac{1+\sqrt{5}}{2}$ .

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  $R_5^5 = R_3^3 = (R_5 R_3)^2 = 1$ .

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 ( $\approx 100$  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  $(n_1, \dots, n_6)$ .

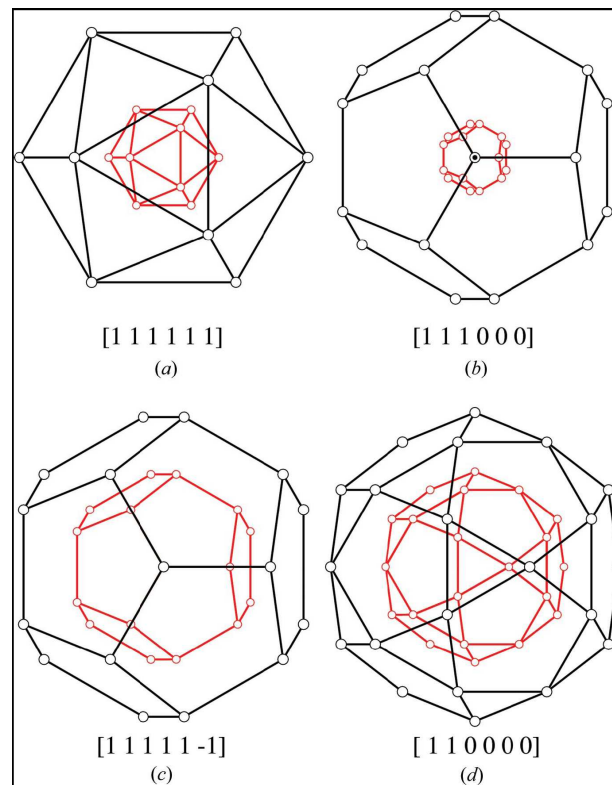
This radial scaling is given by matrices  $\frac{1}{\tau^2}I_3$  and  $\frac{1}{2}(4I_6 - J_6)$ , in terms of bases  $\{e_1, e_2, e_3\}$  and  $\{a_1, \dots, a_6\}$ , respectively.

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  $\approx 9$  nm but it depends on serotype and proteins.

# Janner's general model

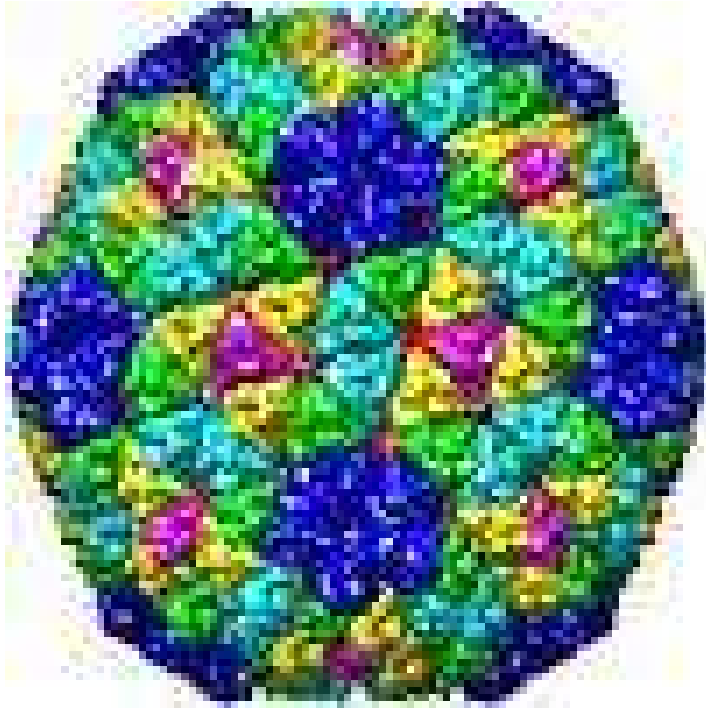
**Janner, 2007:** icosahedral capsid surface, delimited by external and internal form, can be characterized as icos. cluster of 2 polyhedra in  $3D \mathbb{Z}$ -module, generated from a single point in  $\mathbb{Z}_6$ . See below point orbits by the action of  $I_h \times C_4$ :  $\tau^2$ -scaled icosahedra,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



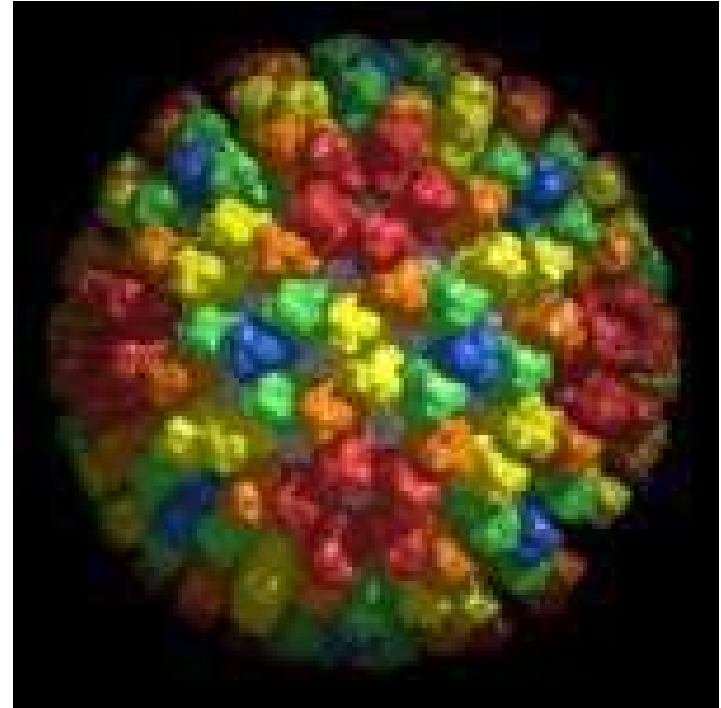
# Virus dynamics and group representation

- 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



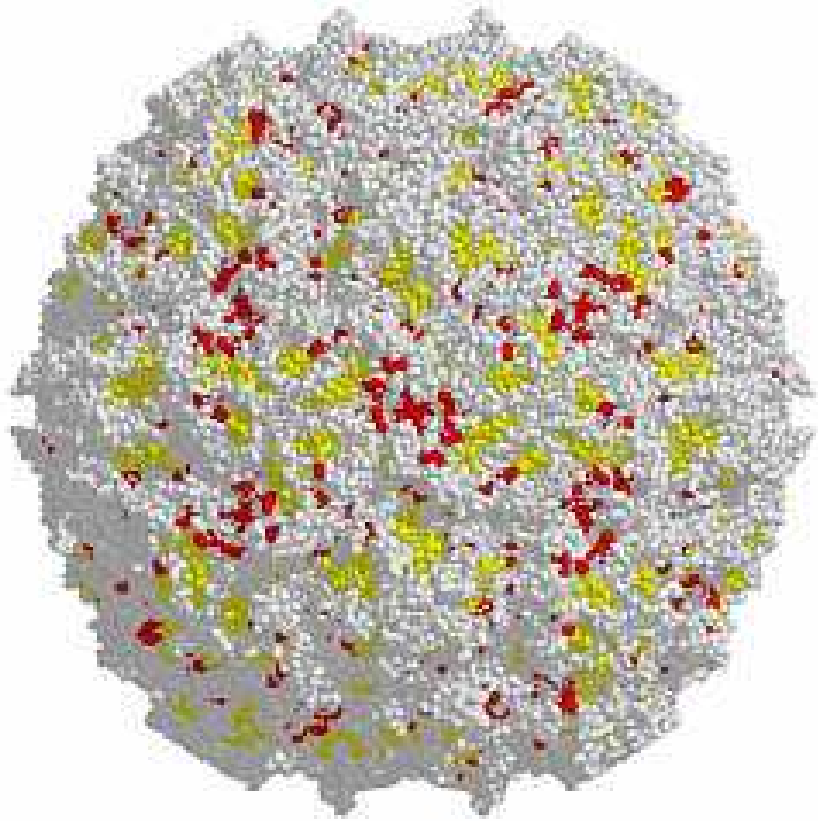
Bluetongue virus

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

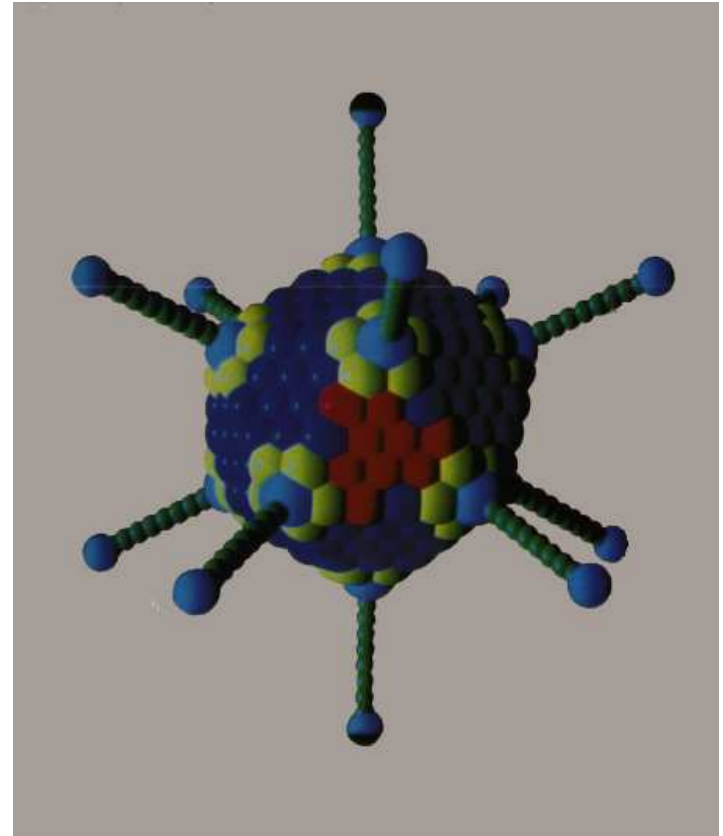
Bursal disease virus is an example of ones with  $13$  *dextro*.

# Viruses with $T = 25$

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

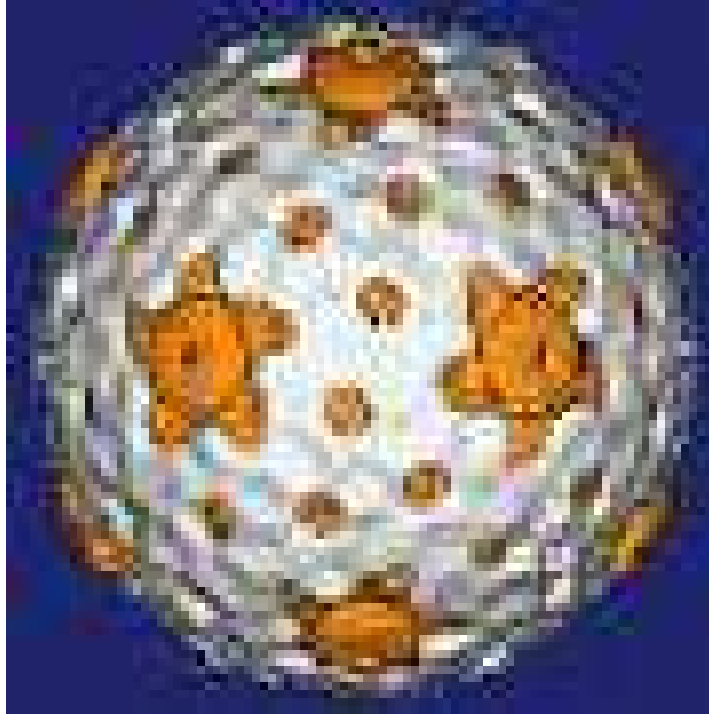


PRD1 virus

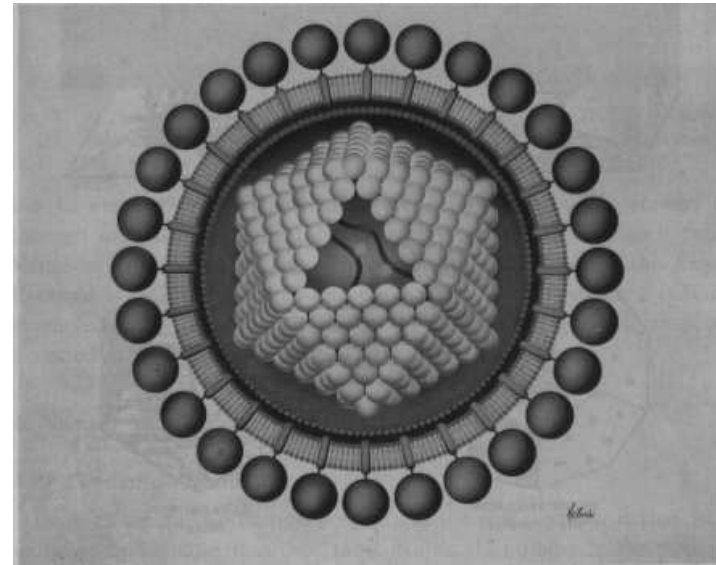


Adenovirus (with its spikes)

# More $I_h$ -viruses

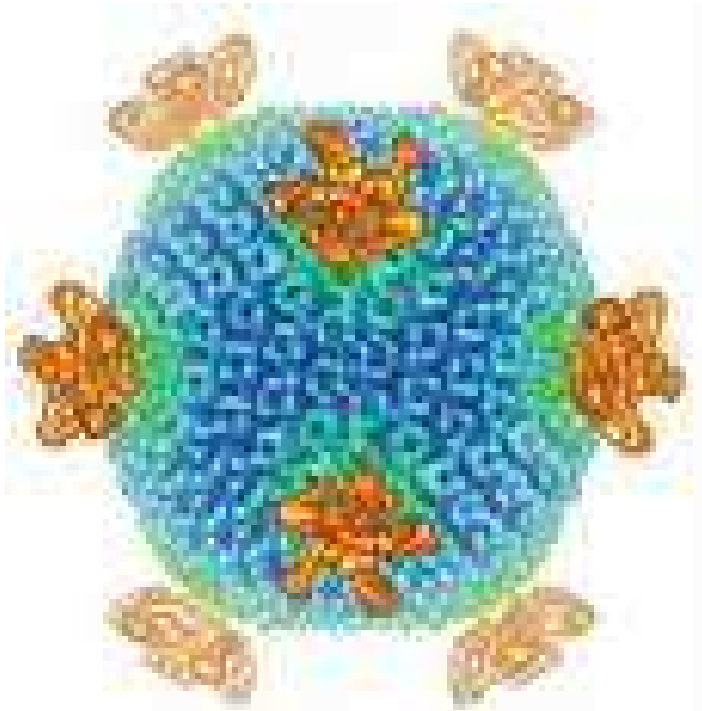


Pseudomonas phage phiKZ,  
 $T = 27$

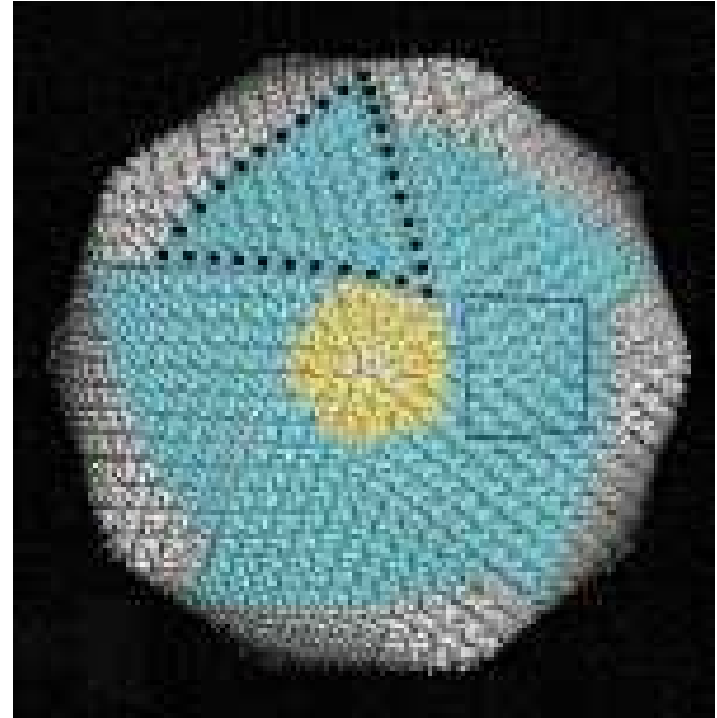


HTLV1 virus,  
 $T = 36$

# Large icosahedral viruses



**Archeal** virus STIV, (5, 1)

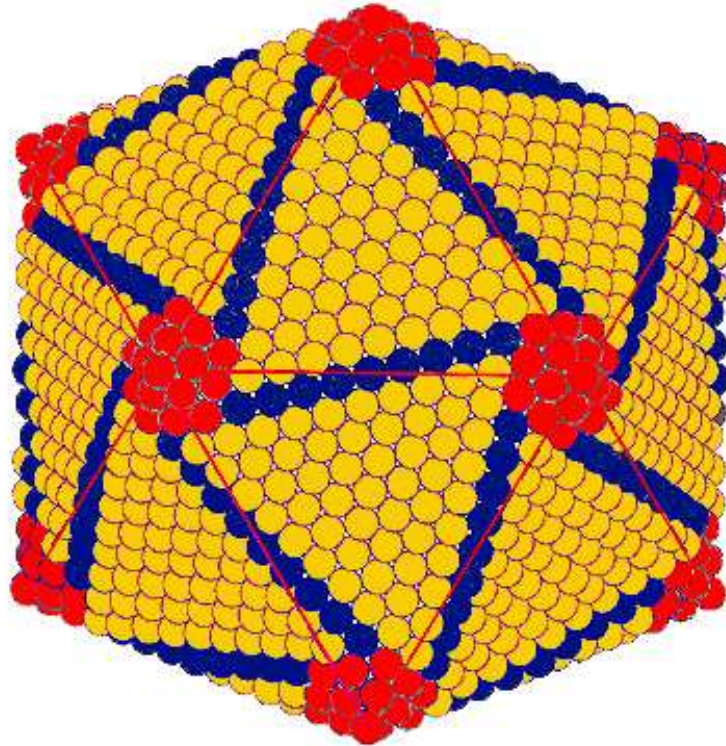


Algal chlorella virus PBCV1  
(4th:  $\simeq 331.000$  bp), (13, 0)

- 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?

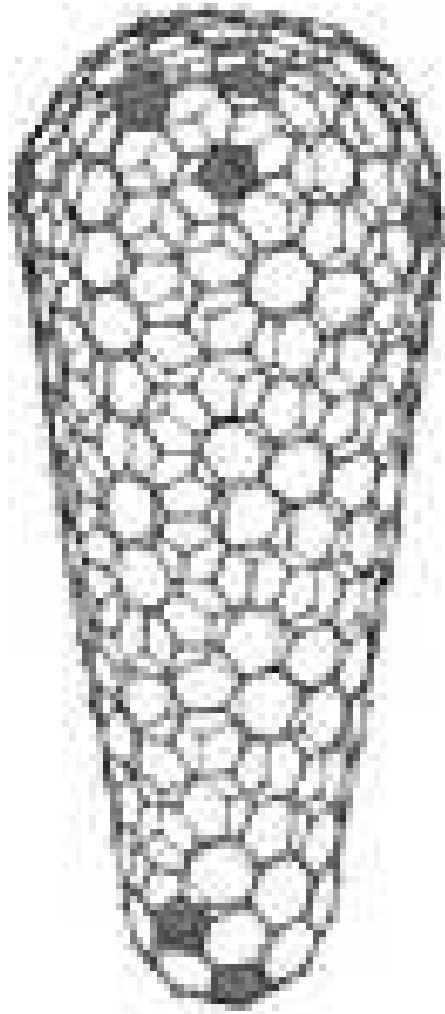
# Sericesthis irridescent virus (SIV)

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  $(20 \times 55) + (12 \times 16) + (30 \times 9) = 1562$  subunits, i.e. icosahedrality with  $T=156=10^2+(10 \times 4)+4^2$  (still, values 1472 or 1292 subunits are not excluded).

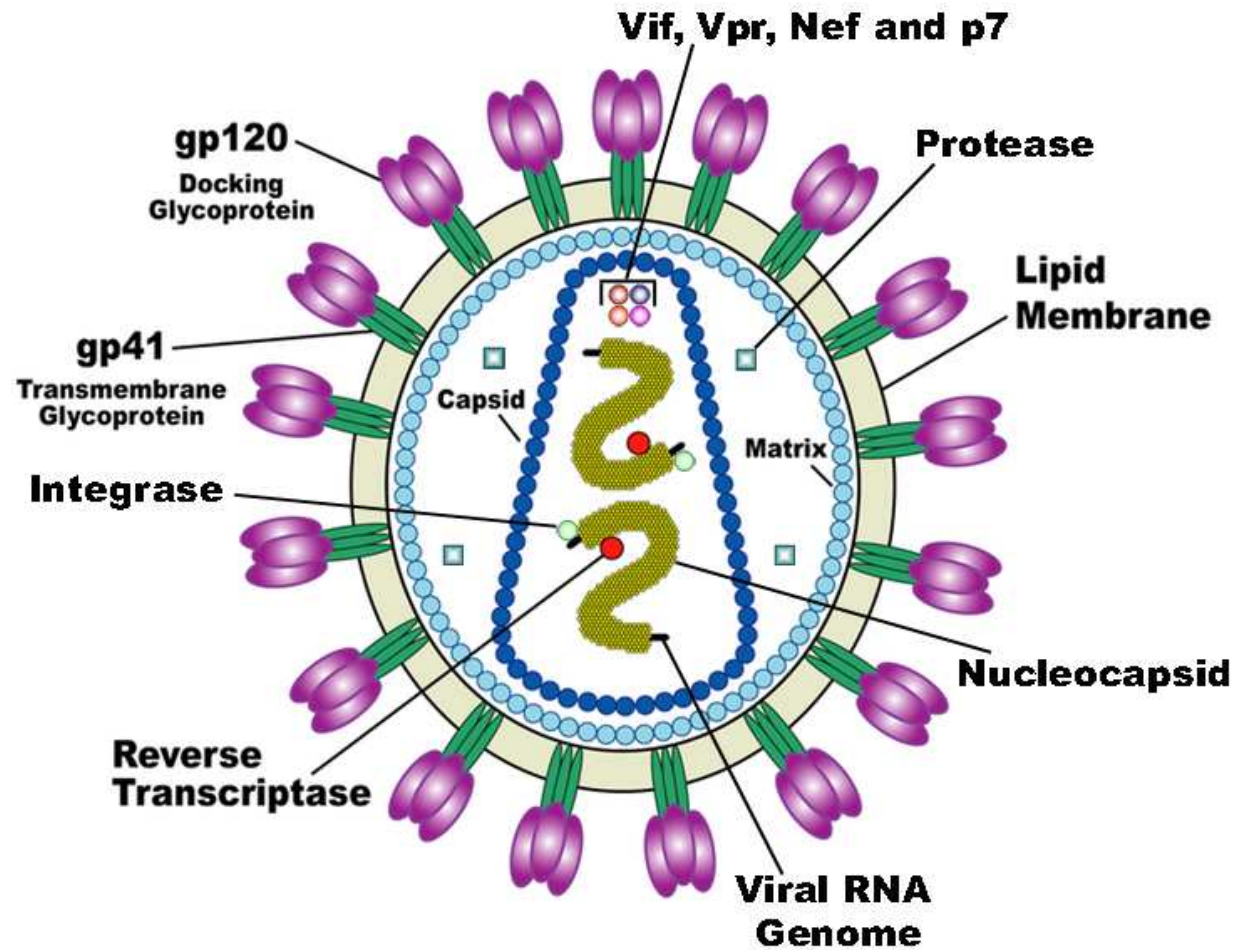


## VII. Other shapes

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



Capsid core  
(7, 5)



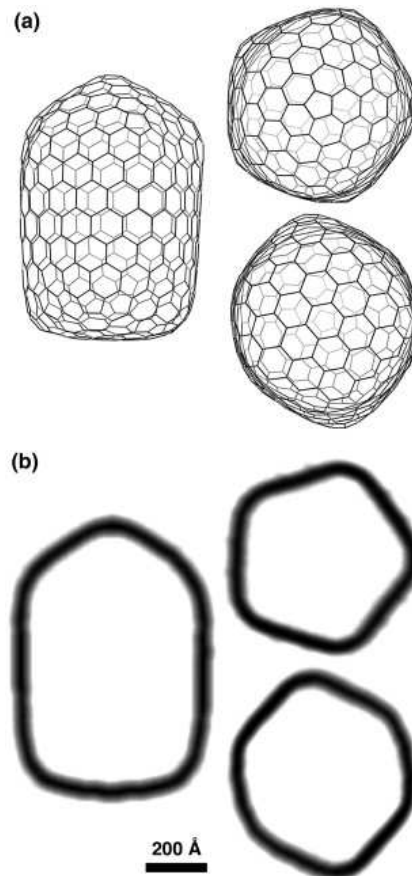
Icosahedral shape (spikes):  $T \simeq 71$ ?

# Fullerene quasi-equivalence variations

**Mellema et al., 1979:** Alfalfa Mosaic Virus ALMV is cylindric nanotube of hexagons with two icosahedral ( $T = 1$ ) caps.

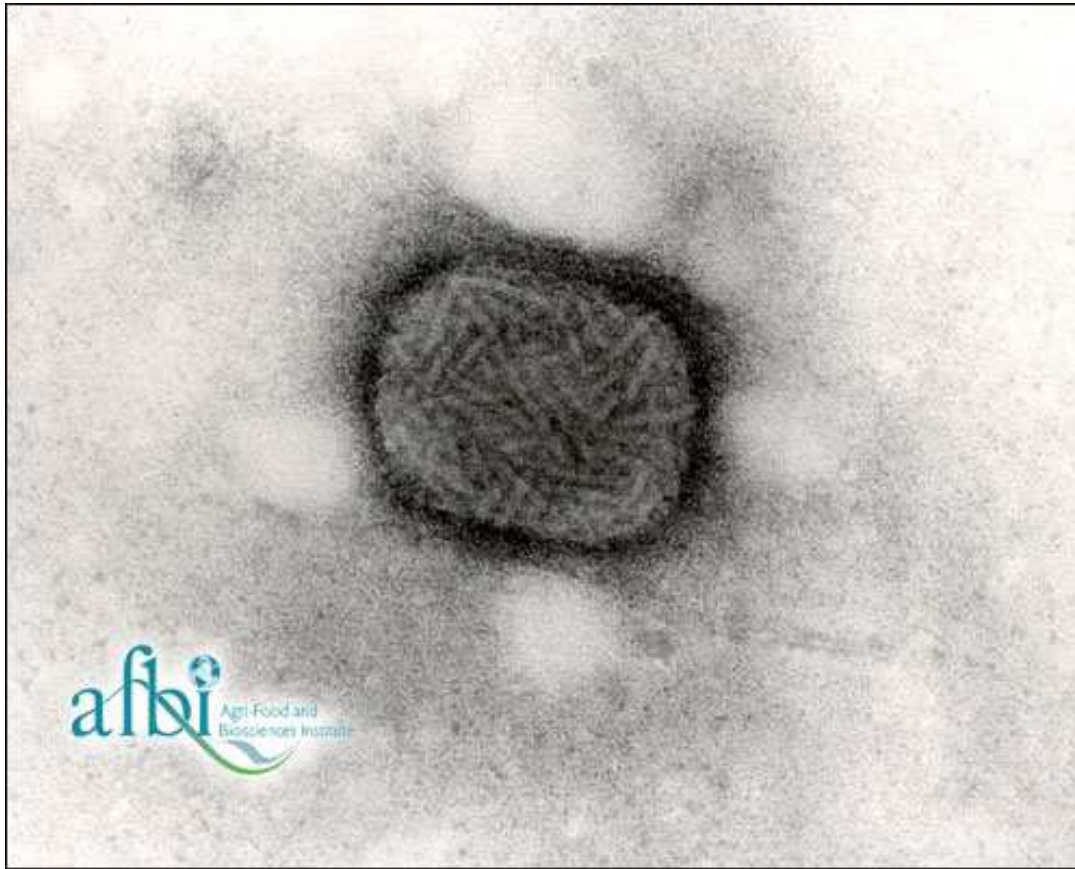
**Ganser et al., 1999:** HIV capsids are **conic (5, 7)-fullerenes** mainly (but still not visualized at high resolution EM).

**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



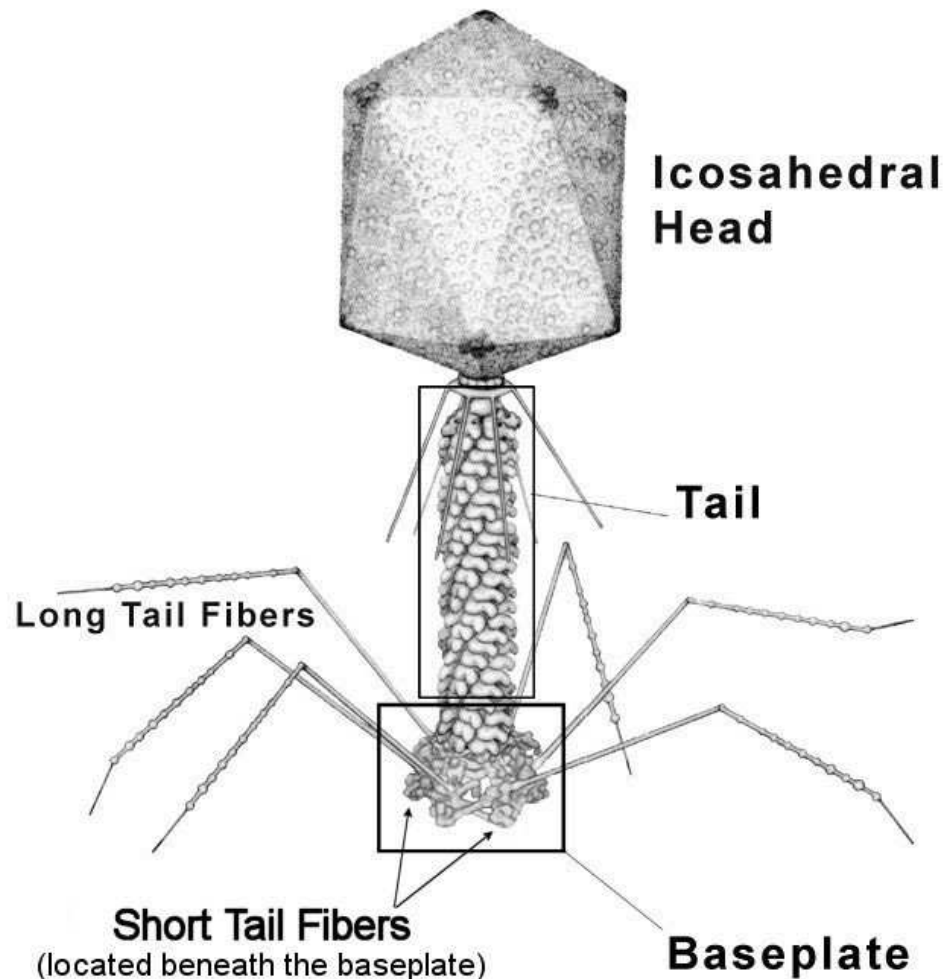
Avian poxvirus

Poxviruses are unique among DNA viruses: they replicate in cell's cytoplasm rather than in its nucleus.

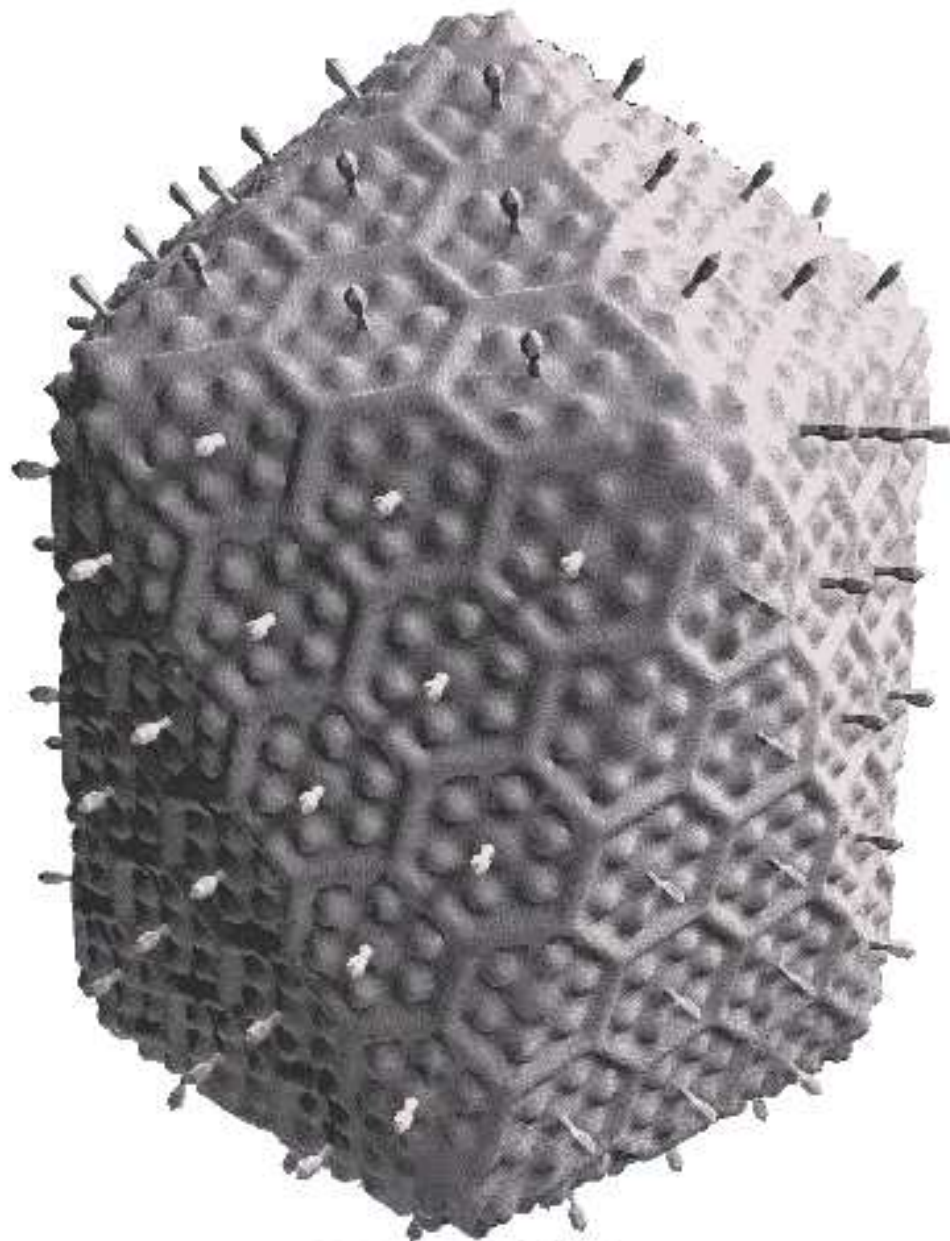
# T4: most complex known structure

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

**Schematic of T4 Bacteriophage**



# Prolate-icosahedral capsid head of T4



# Geometry of bacteriophage T4

This icosahedral head + helical tail architecture is known only in (but in almost all) phages, as this best-studied 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  $T_{end}=a_1^2+a_1b_1+b_1^2$  basic triangles; and 10 midsection facets, each defined by vectors from 0 to  $(a_1, b_1)$ ,  $(a_2, b_2)$  and having  $T_{mid}=a_1a_2+a_1b_2+b_1b_2$  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:  $T_{mid}=20$ .

# Elasticity and shape transition

- **Toan-Bruinsma1-Gelbart, 2005** explain shape transition of icosahedral capsids from spherical to tubular, as their radius increase, by stretching and bending energies competition of a closed 2D elastic shell network.
- Phages  $T4$ ,  $\phi CbK$ ,  $\phi 29$  have prolate Icosahedron tubular shape, i.e.  $F_{24+12t}$  ( $D_{6h/6d}$  if  $t$  is odd/even). Such capsid consist of two half-icosahedral caps connected by cylindrical, mid-portion of rings of 6-gons.
- By mutations or polymorphism (HIV, AMV), some viruses can be spherical, tubular or cone-like.
- If energy cost pentamer/hexamer ratio is large, then tubular structure have lower energy, since tube has lower pentamers/hexamers ratio.

# 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,  $\alpha_3$ -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:

$$\begin{aligned} F_{20}(I_h)(5, 0) &\rightarrow \frac{1}{2}H_{10} & F_{20}^*(I_h)(5, 0) &\rightarrow \frac{1}{2}H_6 \\ F_{60}^*(I_h)(0, 3) &\rightarrow \frac{1}{2}H_{10} & F_{80}(I_h)(0, 4) &\rightarrow \frac{1}{2}H_{22} \end{aligned}$$

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

$$\begin{aligned} F_{26}(D_{3h})(-, 0) &\rightarrow \frac{1}{2}H_{12} \\ F_{40}(T_d)(2, -) &\rightarrow \frac{1}{2}H_{15} & F_{44}(T)(2, 3) &\rightarrow \frac{1}{2}H_{16} \\ F_{28}^*(T_d)(3, 0) &\rightarrow \frac{1}{2}H_7 & F_{36}^*(D_{6h})(2, -) &\rightarrow \frac{1}{2}H_8 \end{aligned}$$

- Also, for graphite lattice (infinite fullerene), it holds:

$$(6^3)=F_\infty(0, 6) \rightarrow H_\infty, Z_3 \text{ and } (3^6)=F_\infty^*(0, 6) \rightarrow \frac{1}{2}H_\infty, \frac{1}{2}Z_3.$$

# 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 gonalties (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 Skymion (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 Skymion for any  $B = 5(a^2 + ab + b^2) + 2$  with integers  $0 \leq b < a$ ,  $\gcd(a, b) = 1$  (not any icosahedral Skymion has minimal energy).
- Skyrme, 1962 model is a Lagrangian approximating QCD (a gauge theory based on  $SU(3)$  group). Skymions are special topological solitons used to model baryons.