

Group Theory Methods in Virology: Landau Density Wave Approach

Vladimir L. Lorman and Sergei B. Rochal

Abstract Viruses are organized biological nanosystems which display high level of spatial organization. In the present work we focus on the group theory methods application to the problems of virus self-assembly and resulting viral structure formation. The approach is based on the successive application of methods of representation theory for continuous and discrete groups and invariant theory for the groups not-generated by reflections. It generalizes the Landau density wave theory of crystallization to the case of compact crystal-like manifold assembly. To compare the predictions of the theory with the available cryoelectronic microscopy data we use the calculated density distribution functions which generate the protein positions on a spherical surface of the cage protecting viral genome. We also discuss the relation between density distribution functions and viral infectivity.

Keywords Group theory · Virology · Density waves · Self-assembly · Icosahedral viruses · Crystallization

1 Introduction

Viruses occupy the “gray area” between living and non-living matter. In contrast with other forms of living matter viruses cannot replicate independently, they need a host cell and its biosynthetic machinery to reproduce new viral particles. Their genetic material is vulnerable to degradation before the infection into the cell can occur. Therefore, it is protected by the viral protein shell (capsid) which encloses the genome. Though virus capsid formation involves biologically specific events, some

V.L. Lorman (✉)

Laboratoire Charles Coulomb, UMR 5221, CNRS-Université de Montpellier,
Pl. E. Bataillon, 34095 Montpellier Cedex 5, France
e-mail: vladimir.lorman@univ-montp2.fr

S.B. Rochal

Southern Federal University, ul. Zorge 5, 344090 Rostov-on-Don, Russia
e-mail: rochal_s@yahoo.fr

© Springer Japan 2016

R.S. Anderssen et al. (eds.), *Applications + Practical Conceptualization + Mathematics = fruitful Innovation*, Mathematics for Industry 11,
DOI 10.1007/978-4-431-55342-7_2

steps of the self-assembly are similar to passive physical processes and show universal features. All viruses follow some common scenario: they deliver viral genomic material into host cell, subvert cell's biosynthetic mechanisms into producing viral genome and proteins, then, new viral particles self-assemble in the infected cell, and finally, new generation of virus leaves the cell [1]. Genomic material of viruses is strongly diversified: it can be DNA or RNA (both "+" and "-" type), single- or double-stranded, linear or circular, one or several copies. On the contrary, capsid structure is quite universal. Viral capsids are made of many copies of identical subunits (one or few proteins) self-assembling in a two-dimensional (2D) shell. The positions and orientations of subunits display high level of spatial organization. With a typical diameter of the order of 50 nm and the regular protein organization, capsids represent nano-systems well-suited to modern structural methods of study like synchrotron radiation diffraction or cryoelectron tomography [2]. Recent structural data obtained due to the progress of cryoelectronic technique rise a whole number of questions concerning unconventional positional order of subunits in the shell, thermodynamics and physical mechanisms of the self-assembly.

From the point of view of geometry, viral capsids are divided into two wide classes: (i) open cylindrical structures with helical rod-like protein arrangement; (ii) spherical ones. Several viral families constitute notable exception to this classification. For example, capsids of many retro-viruses including HIV are conical shells with continuously varying curvature. In the present work we focus on viruses with compact spherical topology, on successive steps of the spherical virus assembly process, and on the underlying physical mechanisms and mathematical formalism. We are interested mainly in the first step of the process. At this step a solid spherical protein shell self-assembles from the aqueous solution of individual proteins.

In their pioneer work Crick and Watson (CW) [3] stated that viruses with spherical topology should have the symmetry (but not necessarily the shape) of one of regular polyhedra. Using more detailed X-ray diffraction data Caspar and Klug (CK) precised that spherical capsids adopt icosahedral point symmetry [4]. Decades of experimental studies showed that lateral, strongly orientation-dependent type of capsid protein interaction, together with intrinsic curvature and, especially, asymmetry of capsid proteins define particular docking preferences. Consequently, protein interaction results in specific geometric arrangements and influences the capsid assembly phenomenon. Both CW and CK insisted on the fact that typical viruses have very small genome (with respect to any other biological system, i.e. bacteria). Thus, it can code only for a few proteins, and among them there is typically only one "coating" protein. Viral capsid is then constituted with multiple copies of the same coating protein. They also stressed that the interaction of identical proteins should lead to identical local environments, including local orientational and positional order and local chemical bonding, and proposed to construct the icosahedral shells possessing these properties.

Due to these advances the main problems in the field of physical structural virology could be formulated: (i) *how to construct a regular 2D shell with the icosahedral symmetry formed by multiple copies of identical asymmetric proteins in identical local*

environments, including local orientational and positional order and local chemical bonding; (ii) *how to relate the symmetry of individual proteins and the symmetry of the assembled shell to the free energy of the assembly process.*

2 Geometric Restrictions

A whole number of physical and geometric restrictions arise in this way and lead to the selection rules for the viral capsid structures. To understand the implementation of these rules it is necessary to take into account the fact that they are combined with the evolutionary optimization of viral genome. This remarkable feature of biological systems distinguishes them from the non-living ones. Evolutionary pressure pushes the virus to increase its genome, and, consequently, to increase the volume of the protective capsid shell. Taking into account WC and CK ideas about “identical sub-units in identical environments” one can easily formulate the simplest group theory problem: *what is the 3D irreducible symmetry group of the protein distribution forming a spherical shell of maximal volume and composed with a finite number of 2D proteins of fixed finite area?*

2.1 Protein Asymmetry and Rotational Icosahedral Symmetry of Viral Capsids

Since the proteins are asymmetric they can form the structures with *rotational* symmetry elements only, excluding inversion and mirror planes. Indeed, because of the asymmetry at least several first moments of their mass and charge distribution are different from zero. Let us take into account the simplest one, usually called chirality. Corresponding moment has the symmetry of a pseudo-scalar ε of the fixed sign. The sum of pseudo-scalars of the same sign over any distribution being non-zero $\sum_i \varepsilon_i \neq 0$, the average capsid chirality is also non-zero. Thus the capsid structure is chiral, and its symmetry group contains no spatial inversion nor mirror planes.

Identical asymmetric proteins can be put in identical environments if they are located in the positions which form one *regular orbit* of the 3D rotational point group G only. To occupy positions on a n -fold axis the molecules should have corresponding symmetry themselves. The number of positions of identical asymmetric proteins is equal to the dimension $\dim [\text{Orb } G]$ of a regular orbit of a discrete symmetry group G which is equal to the group order $|G|$. For the tetrahedral rotational group $G = T$ the number of proteins constituting the shell is $\dim [\text{Orb } T] = 12$, for the octahedral group $G = O$ this number is $\dim [\text{Orb } O] = 24$ while for the icosahedral rotational group $G = I$ it is $\dim [\text{Orb } I] = 60$. It is evident that the icosahedral symmetry of the protein distribution leads to the shell constituted with 60 copies of the same

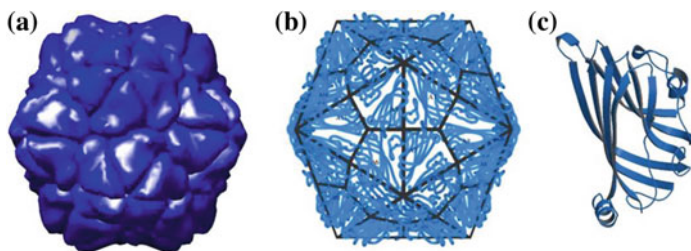
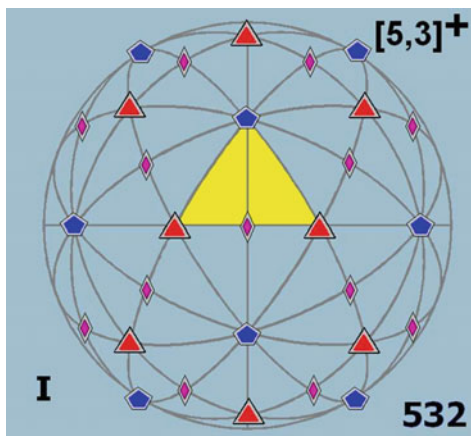


Fig. 1 Typical small spherical virus (Satellite Tobacco Necrosis Virus) with the icosahedral symmetry of protein organization: **a** Shaded Surface Diagram of the cryoelectronic density; **b** Ribbon Diagram with schematic protein presentation; **c** Asymmetric Protein

protein. Corresponding icosahedral shell has the volume much bigger than the shell volume in tetrahedral and octahedral cases.

The majority of spherical viruses are indeed icosahedral and the ordered distribution of proteins in their capsids has the symmetry group I (Fig. 1). The tessellation of a sphere by the action of its elements (called Möbius tessellation) is presented in Fig. 2 [5]. Note that the fundamental domain of the group I (shown in Fig. 2 the yellow triangle with one 5-fold and two 3-fold axes in the vertices) is usually called “asymmetric protein unit” in structural virology (though this group theory notion is not used explicitly in this field). Without taking into account evolution arguments, 60 would be the maximum possible number of proteins with identical local environments in an icosahedral shell. However, evolutionary pressure compels virus to increase more and more its genome size and the capsid volume. Experimentally, many viruses contain more than 60 identical proteins in the shell. In general, their number is equal to $60N$, where N is a positive integer. Their positions belong to different regular orbits of the group I and cannot be equivalent.

Fig. 2 Möbius tessellation of a sphere by the action of icosahedral symmetry elements [5]. Fundamental domain of the rotational icosahedral group I is shown in yellow, symmetry axes are given by corresponding polygons



2.2 Selection Rules: Caspar and Klug Geometric Model

A partial explanation of this discrepancy was given by CK [4]. The task of CK was to find a way to put identical proteins in different but nearly equivalent positions and to explain the origin of this “quasi-equivalence”. For that aim CK proposed to use the properties of plane periodic structures formed by the same type of asymmetric particles, and then to transfer these properties onto a sphere submitted to the action of the icosahedral group. Indeed, the above restriction on the group orbit size is proper to point groups. By contrast, the translational symmetry of a 2D plane lattice makes infinite the dimension of the regular orbit of the 2D space group $\dim[OrbG] = \infty$. Consequently, the number of asymmetric proteins which can be put in equivalent environments becomes infinite in a 2D planar crystal. Then CK were looking for an almost regular mapping of the periodic planar structure on the icosahedron surface. Mapping splits one regular orbit of the plane structure symmetry group into several different regular orbits of the icosahedron group I but maintains some “traces” of their former “equivalence” in the plane structure.

Specific properties of the purely geometric CK model impose the selection rules for the value of N , i.e. for the number of different regular orbits of the I group. Consequently, they select the number $60N$ of proteins constituting the shell. Only the values which satisfy the relation $N = h^2 + k^2 + hk$, where h and k are non-negative integers are allowed by the CK selection rules. Constructively, CK proposed to map a net of an icosahedron slitted in the 2D hexagonal lattice to the icosahedron surface (Fig. 3). The mapping is chosen in such a way that the vertices of regular triangular faces of the icosahedron coincide with the 6-fold axes of the plane hexagonal struc-

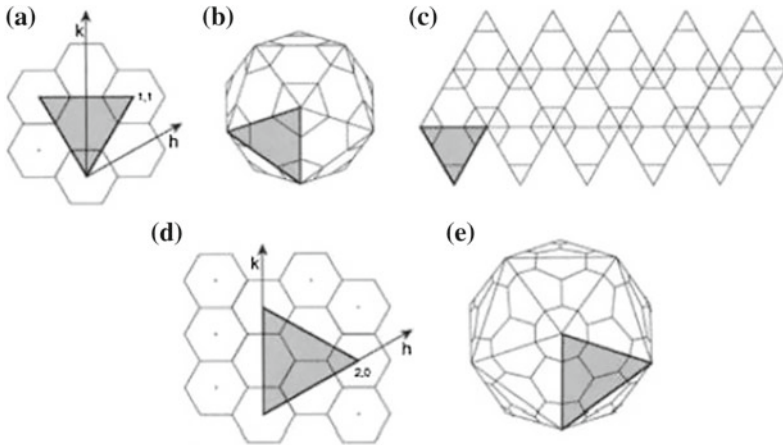


Fig. 3 Caspar and Klug Geometric Model: **a** triangular icosahedron face chosen in the hexagonal lattice with the edge vector $(h, k) = (1, 1)$; **b** folded icosahedron with $N = 3$ resulting from choice (a); **c** icosahedron net cut in the hexagonal lattice according to (a); **d** icosahedron face with $(h, k) = (2, 0)$; **e** folded icosahedron with $N = 4$ resulting from choice (d)

ture. By cutting out 60° -sector one can transform the 6-fold axis into the 5-fold one and then join the sector edges on the icosahedron surface. The edge length of the icosahedron face is determined by the vector joining vertices of the net triangles. Vertices being situated in the nodes of the hexagonal lattice, the vector length square is given by $N = h^2 + k^2 + hk$, where h and k are non-negative integers. This number is evidently equal to the number of the lattice nodes contained in two net triangles. Figure 3 shows first two non-trivial icosahedron nets with $N = 3$ and $N = 4$, respectively. In the hexagonal coordinates (h, k) the first net is indexed as $(1, 1)$, the index of the second one is $(2, 0)$. The procedure imposes also specific local arrangement of protein positions induced by the CK mapping. They are forced to form “pentamers” in the vicinity of 5-fold axes and “hexamers” elsewhere.

From the theorist’s point of view the procedure proposed by CK has several problems of arbitrariness. It concerns the mapping choice, which is not unique, and the interpretation of geometric properties of manifolds. However, they do not constitute the main problem of the CK geometric model application to the virus structures and self-assembly process. The main problem, as it is often the case, comes from the refined experimental data. Though a big number of virus capsid structures are in a good agreement with the CK scheme, there is a growing number of experimentally

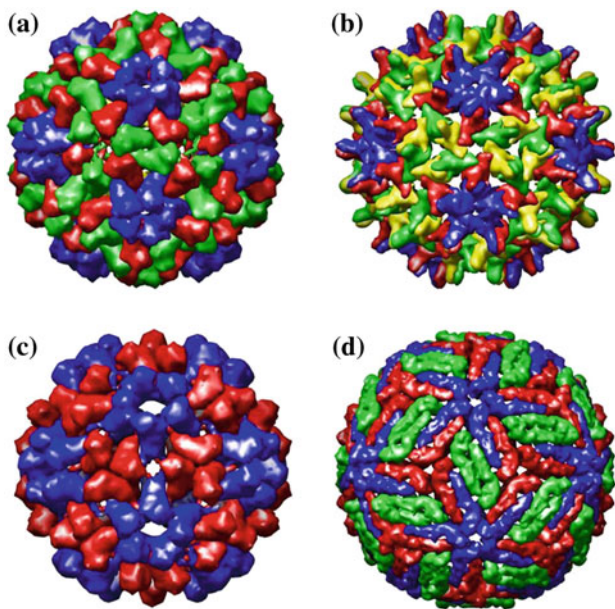


Fig. 4 Experimental cryoelectronic density for several viral capsids. Different positions of the same protein are given in different colors. Capsids (a) and (b) satisfy to the Caspar and Klug selection rules, while (c) and (d) are forbidden by them: **a** capsid of Cowpea Chlorotic Mottle Virus with $N = 3$ and “hexamer” organization of proteins; **b** capsid of Hepatitis B Virus with $N = 4$ and “hexamer” organization of proteins; **c** capsid of L-A virus with the forbidden $N = 2$; **d** capsid of Dengue Virus with $N = 3$ but without any “hexamer”

resolved structures which do not satisfy the CK selection rules nor their predictions about local proteins arrangement [6–8]. Figure 4 shows experimental cryoelectronic density distributions for several viruses. Among them Cowpea Chlorotic Mottle Virus (CCMV) is a typical example of virus with $N = 3$ verifying all predictions of the CK geometric model. The same remark concerns Hepatitis B Virus (HBV) with $N = 4$. On the other hand the capsid of L-A Virus (usually infecting yeast) is constituted by 120 proteins (i.e. $N = 2$), this number being forbidden by the CK selection rules. Capsid of Dengue Virus (DENV) shows local arrangement of proteins without any hexamer, also violating the CK geometric model. Thus, it becomes relevant to look for a theory which describes in a uniform way all experimentally observed small spherical viruses including those which can not be obtained using the CK geometric model.

3 Landau Density Wave Approach

In the present work we develop an approach which replaces the geometric notions by the notions of statistical physics. The positions of identical proteins in the shell are generated in the frame of the theory by one statistical protein density distribution function. To ensure “quasi-equivalence” of the positions the function is chosen irreducible under the action of the symmetry group I . The self-assembly thermodynamics is then described by the invariant free energy functional dependent on the distribution function. Similar approach was applied with success to describe crystallization process of atomic crystals [9–11] and quasicrystals [12–15]. Here we associate viral capsid formation with the unconventional crystallization process and propose to describe the capsid self-assembly using a generalization of the Landau theory of crystallization [9–11]. It is based on the successive application of methods of (a) representation theory of continuous and discrete groups; (b) invariant theory for the groups not-generated by reflections; (c) bifurcation theory of invariant functionals. To compare the predictions of the theory with the available cryoelectronic microscopy and AFM data we use the calculated irreducible density distribution functions which generate the protein positions on a spherical capsid.

3.1 Critical System of Density Waves

In the frame of the proposed theory the probability density ρ of protein distribution in the capsid in the vicinity of crystallization point is presented as:

$$\rho = \rho_0 + \Delta\rho, \quad (1)$$

where ρ_0 is an isotropic density in the solution of isolated proteins and $\Delta\rho$ corresponds to the density deviation induced by the ordering. The symmetry breaking

during the crystallization is associated with one critical order parameter which spans an irreducible representation of the symmetry group of the disordered state. For the crystallization process the order parameter represents a critical system of density waves (CSDW) [9–11]. For the assembly process on a sphere considered as an unconventional crystallization the critical part $\Delta\rho_l$ of the density is determined by a CSDW with the *fixed* wave number l . The spherical harmonics Y_{lm} constituting CSDW on a sphere span one irreducible representation (IR) of the $SO(3)$ symmetry group of the disordered state, thus $\Delta\rho_l$ is given by:

$$\Delta\rho_l(\theta, \phi) = \sum_{m=-l}^{m=l} A_{lm} Y_{lm}(\theta, \phi), \quad (2)$$

where l is the IR number, A_{lm} are the amplitudes of the spherical harmonics Y_{lm} and θ and ϕ are the angular variables of the spherical coordinate system.

3.2 Selection Rules: Representation Theory and Invariant Theory

The analysis based on the representation theory and the invariant theory shows that for any critical order parameter which drives the icosahedral assembly of asymmetric proteins the wave number l satisfies the relation:

$$l = 15 + 6i + 10j, \quad (3)$$

where i and j are positive integers or zero. Equation (3) defines the list of l numbers for which the restriction of an IR of the $SO(3)$ group on the icosahedral group I contains at least one totally symmetric representation. To verify selection rule (3) one can use the well-known subduction criterion [16]. Active IR of the symmetry group $G_0 = SO(3)$ of the disordered state must subduce the identity representation of the symmetry group $G = I$ of the ordered state. The critical density function $\Delta\rho_l(\theta, \phi)$ is given by the basis functions $f_l^i(\theta, \phi)$ ($i = 1, 2 \dots n_l$) of all n_l totally symmetric representations of the icosahedral group I in the restriction of the ‘active’ IR of the $SO(3)$. The CSDW is a linear combination of these functions invariant with respect to the I group:

$$\Delta\rho_l(\theta, \phi) = \sum_{i=1}^{n_l} B_i f_l^i(\theta, \phi), \quad (4)$$

where B_i are arbitrary coefficients.

The frequency of subduction is defined as

$$n_l = (1/|G|) \sum_G \xi(\hat{g}) \quad (5)$$

where the sum runs over the elements \hat{g} of the I group, $|G| = 60$ is the I group order, and $\xi(\hat{g})$ is the character of the $SO(3)$ group element which reads as [16]:

$$\xi(l, \alpha) = \frac{\sin((l + 1/2)\alpha)}{\sin(\alpha/2)},$$

where l is the IR number and the angle α is determined by the rotation \hat{g} . Then the explicit form of (5) becomes:

$$n_l = \frac{1}{60} (2l + 1 + 15\xi(l, \pi) + 20\xi(l, 2\pi/3) + 12\xi(l, 2\pi/5) + 12\xi(l, 4\pi/5)). \quad (6)$$

From successive application of (6) one can deduce that $n_l \neq 0$ only for wave numbers l satisfying (3).

Selection rule (3) is justified in the frame of the theory of invariants [16, 17]. Any scalar function $F(\mathbf{r})$ invariant with respect to the group I , $F(\hat{g}\mathbf{r}) = F(\mathbf{r})$ for all $\hat{g} \in I$, can be expanded in formal polynomial series of x , y and z components of \mathbf{r} . All the terms in the series are the monomials of four following basis invariants (full set of generators for the ring of invariant polynomials: they constitute *integrity basis* [16, 17] of the I symmetry group):

$$J_0 = x^2 + y^2 + z^2, \quad J_1 = \prod_{i=1}^6 \mathbf{r} \cdot \mathbf{n}_i, \quad J_2 = \prod_{i=1}^{10} \mathbf{r} \cdot \mathbf{p}_i, \quad J_3 = \prod_{i=1}^{15} \mathbf{r} \cdot \mathbf{q}_i, \quad (7)$$

expressed here in terms of scalar products of the radius-vector $\mathbf{r} = \langle x, y, z \rangle$ with the vectors \mathbf{n}_i , \mathbf{p}_i and \mathbf{q}_i parallel to the icosahedron rotational axes. Namely, to the six 5-fold axes, to the ten 3-fold axes, and to the fifteen 2-fold axes, respectively. To verify the fullness of integrity basis (5) we use Molien's theorem [16, 17]. It gives the generating function $H(t)$ for invariants of the vector representation of a finite discrete group G in the form:

$$H(t) = (1/|G|) \sum_M \frac{1}{\det(I - tM)} \quad (8)$$

where $M(\hat{g})$ is the representation matrix for the group element \hat{g} and I is the identity matrix, the sum running over all $\hat{g} \in G$. The resulting generating function $H(t)$ can be presented in the Molien-Weyl form [16, 17]. For the rotational icosahedral group I it reads:

$$H(t) = \frac{1 + t^{15}}{(1 - t^2)(1 - t^6)(1 - t^{10})} \quad (9)$$

This indicates that the integrity basis for the I group contains algebraically independent basis invariants of degree 2, 6 and 10, and an additional invariant of degree 15. In contrast with the case of groups generated by reflections, the number m of invariants in the basis of a group not generated by reflections: (i) is greater than the vector space dimension ($m > 3$ for the group I); (ii) the product of degrees of basis invariants is not equal to the number of group elements $|G|$ ($\prod_{k=0}^3 \deg(J_k) \neq 60$ for the group I); (iii) invariants of the basis form a syzygy, an algebraic relation. In the case of the group I syzygy is the relation of the form $J_3^2 = P(J_0, J_1, J_2)$, where P is the homogeneous polynomial of 30th degree.

On the other hand, CSDW (4) is a linear combination of spherical harmonics with a given odd l . Even values of l are excluded from the consideration because the CSDW invariant with respect to the rotational group I , is not symmetric under the spatial inversion operation. Thus, it is a homogeneous function of a given odd degree l in $\mathbf{r} = \langle x, y, z \rangle$. Any irreducible scalar function $B_l(x, y, z)$ of a given odd degree l and invariant with respect to the rotational group I can be presented as:

$$B_l(x, y, z) = \sum_{2k+6i+10j+15=l} J_3(A_{k,i,j} J_0^k J_1^i J_2^j). \quad (10)$$

k, i and j being the degrees of the J_0, J_1 and J_2 invariants, respectively. Syzygy implies that the terms containing J_3^2 in any degree can be expressed in function of J_0, J_1 , and J_2 . Thus, all J_3 -containing terms in any series invariant with respect to the group I are linear in J_3 . Due to homogeneity of function (10) its wave number l verifies the relation $l = 2k + 6i + 10j + 15$ where k, i and j are non-negative integers. On the unit sphere surface ($r = 1$) the invariant J_0 becomes constant: $J_0 = 1$. Consequently, the form of the density function changes. Let us introduce a radial unit vector $\mathbf{e}_r = \mathbf{r}/r = \langle \sin\theta \cos\phi, \sin\theta \sin\phi, \cos\theta \rangle$ depending on angular coordinates θ and ϕ in a standard way. Then, on the spherical surface the irreducible scalar function $B_l(\theta, \phi) = B_l(\mathbf{e}_r)$ has the form:

$$B_l(\mathbf{e}_r) = J_3(C + A_{1,0}J_1 + A_{0,1}J_2 + A_{2,0}J_1^2 + A_{1,1}J_1J_2 + \dots \sum_{6i+10j+15=l} A_{i,j}J_1^iJ_2^j) \quad (11)$$

where $C = A_{0,0}$ is constant. This gives, finally, that the wave number l for irreducible density functions (11) verifies the relation $l = 15 + 6i + 10j$, where i and j are non-negative integers.

3.3 Icosahedral Density Distribution Functions and Protein Arrangement in Viral Capsids

Selection rule (3) gives the possibility to obtain the explicit form of critical density (2). For small icosahedral capsids the practical construction of the protein density distribution is simplified because the CSDW contains only one function $f_l(\theta, \phi)$. Indeed, according to subduction criterion $n_l = 1$ for all $l \leq 43$. In this simplest case $\Delta\rho_l(\theta, \phi) = Bf_l(\theta, \phi)$, where B is a single arbitrary coefficient. The positions of maxima of the density function do not depend on the value of B . They are generated by a single universal function $f_l(\theta, \phi)$ which has no fitting parameter. In the following consideration the functions $f_l(\theta, \phi)$ possessing this properties are called irreducible icosahedral density functions. The explicit form of the irreducible density function $f_l(\theta, \phi)$ for a given value of l is obtained by averaging of $Y_{lm}(\theta, \phi)$ harmonics over the I symmetry group.

$$f_l(\theta, \phi) = \frac{1}{60} \sum_G Y_{l,m}(\hat{g}(\theta, \phi)). \quad (12)$$

The sequence of the wave number values l permitted by selection rule (3) is given by: $l = (15, 21, 25, 27, 31, 33, 35 \dots)$. This sequence determines possible capsid shell structures for small icosahedral viruses. Figure 5 presents irreducible icosahedral

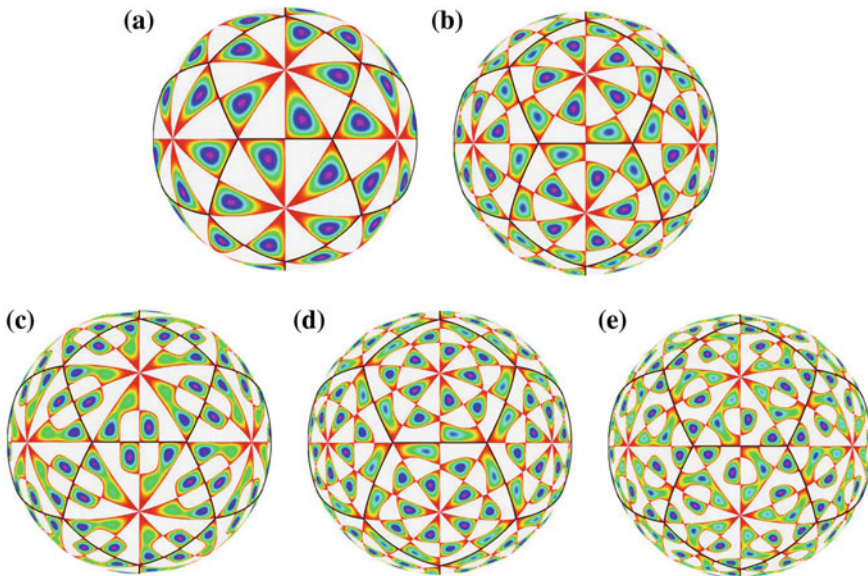


Fig. 5 Irreducible icosahedral density functions for the first five permitted values of the wave number l : **a** $l = 15$ resulting in $N = 1$ orbit of maxima; **b** $l = 21$ with $N = 2$ orbits; **c** $l = 25$ with $N = 3$; **d** $l = 27$ with $N = 3$; **e** $l = 31$ with $N = 4$

density functions for the first five permitted values of l . The value of $f_l(\theta, \phi)$ is represented using false colors: variation of colors from red to violet corresponds to the function growth. Note that all $f_l(\theta, \phi)$ functions are anti-symmetric: they change their sign under the inversion operation or under the action of mirror planes of a regular icosahedron. Thus, for the sake of clarity, we present the positive part $f_l(\theta, \phi) > 0$ only. The protein centers are associated with the positions of maxima of $\Delta\rho_l$ function (2). The number of maxima of the density functions is equal to $60N$, where N is the number of different regular 60-fold orbits of the I group. In the viral capsid N corresponds to the number of different positions occupied by the proteins.

The density wave approach replaces nonuniversal geometric CK model and describes in a uniform way capsid structures which can be obtained using CK geometric model and those which cannot. It is also worth noting that our approach shows that there exist viruses with capsids constituted by the same number of proteins but with qualitatively different protein distributions. For example, Fig. 5 shows two different distributions with $N = 3$. One distribution, generated by the irreducible icosahedral density function with $l = 27$, satisfies CK model and shows “pentamer” and “hexamer” local order, while the distribution with $l = 25$ does not satisfy the CK model and shows typical “rhombus” pattern without any “hexamer”. The former distribution describes with a very good accuracy the experimental protein density distribution in CCMV virus (Fig. 4a), while the latter one fits the distribution in DENV virus (Fig. 4d), the capsids of both viruses being constituted of 180 proteins.

The proposed approach treats also the free energy of the transition from the isotropic aqueous solution of proteins to the assembled capsid shell (see the basic principles in [15, 18]). It gives an access to the thermodynamics of the virus self-assembly process and its relation with the symmetry of CSDW and irreducible icosahedral density functions. The invariant free energy functional Φ of the Landau crystallization theory adopted to describe the assembly process contains homogeneous part F dependent on local order parameter value and inhomogeneous part L expressing energy cost for deviation from spatial uniformity:

$$\Phi = \int_V [F((A_{l,m})) + L((\nabla A_{l,m}))] dV \quad (13)$$

where CSDW amplitudes $(A_{l,m}) = \langle A_{l,-l}, A_{l,-l+1}, \dots, A_{l,-l} \rangle$ are the order parameter components. The homogeneous free energy expansion can be taken in a standard for the crystallization theory form [15, 18] $F = F_0 + F_2 + F_3 + F_4 + \dots$ of successive invariant terms

$$\begin{aligned} F_2 &= A(T, c) \sum_{m=-l}^{m=l} A_{l,m} A_{l,-m}, \\ F_3 &= B(T, c) \sum_{m_1, m_2, m_3} a_{m_1, m_2, m_3} A_{l, m_1} A_{l, m_2} A_{l, m_3} \delta(m_1 + m_2 + m_3) \equiv 0, \\ F_4 &= \sum_k C_k(T, c) \sum_{m_1, m_2, m_3, m_4} a_{m_1, m_2, m_3, m_4}^k A_{l, m_1} A_{l, m_2} A_{l, m_3} A_{l, m_4} \delta(m_1 + m_2 + m_3 + m_4), \end{aligned} \quad (14)$$

where a_i are weight coefficients of the $SO(3)$ group (e.g. Clebsch-Gordan coefficients for the third order term F_3), $\delta(0) = 1$, $\delta(i \neq 0) = 0$, $A(T, c)$, $B(T, c)$, and $C_k(T, c)$ are temperature- and composition-dependent coefficients of the Landau theory. Because of the protein asymmetry and according to selection rule (3) free energy (14) can be written for an odd wave number l only. But for any odd wave number l the third-order term F_3 is identically zero. Then, the analysis of bifurcations of the invariant free energy functional shows [18] that in contrast with classical crystallization process the assembly of asymmetric proteins in an icosahedral shell can take place as a second-order phase transition without a free energy barrier between the isotropic and the icosahedral states.

3.4 Icosahedral Density Distribution Functions and Virus Infectivity

In the final part, we illustrate the relation between the parameters of the proposed theory and the problem of virus infectivity, extremely important for biology. We evidence the relation between the protein density distribution function and the probability of the host cell receptor molecule attachment in different sites of the viral capsid surface. This relation connects the proposed approach to the problem of virus entry into the infected cell. We take here an example of DENV virus interaction with dendritic host cells. Note that the DENV protein distribution obtained in the frame of our approach is the first DENV model based on physical principles and not on a simple empirical fit of the experimental data. Figure 6 shows the experimental cryoelectronic density of DENV virus (Fig. 6b) [19] with bound carbohydrate recognition domains (CRD) of a dendritic cell receptor. One can see a typical “rhombus” motif of the DENV protein arrangement identical to the pattern of maxima of the irreducible density function with $l = 25$ and $N = 3$ (Fig. 6a). Figure 6c [19] represents a schematic zoom of the “rhombus” motif with the fundamental domain of the I group (“asymmetric protein unit”) shown by a triangle. Three different capsid protein environments are given in different colors. It also shows the positions of glycosylation sites Asn 67 and Asn 153. In in vitro experiments with CRD molecules and individual DENV coat proteins in solution at correct temperature, CRD binds to Asn 67 sites of all DENV proteins with extremely high probability. However, in vivo CRD (given in cyan in Fig. 6) binds to two Asn 67 sites leaving the third Asn 67 residue vacant, though all three DENV proteins occupying three different positions in the distribution are identical. This result shows that infectivity promoted by the interaction of cell receptors with virus surface depends not only on bio-specific binding properties of individual molecules but also on the capsid proteins distribution. The interaction of CRD with the proteins in the capsid represents a collective phenomenon.

Comparison of the calculated distribution with the experimental data allows to establish the relation between the minima of $f_l(\theta, \phi)$ and binding sites on the capsid

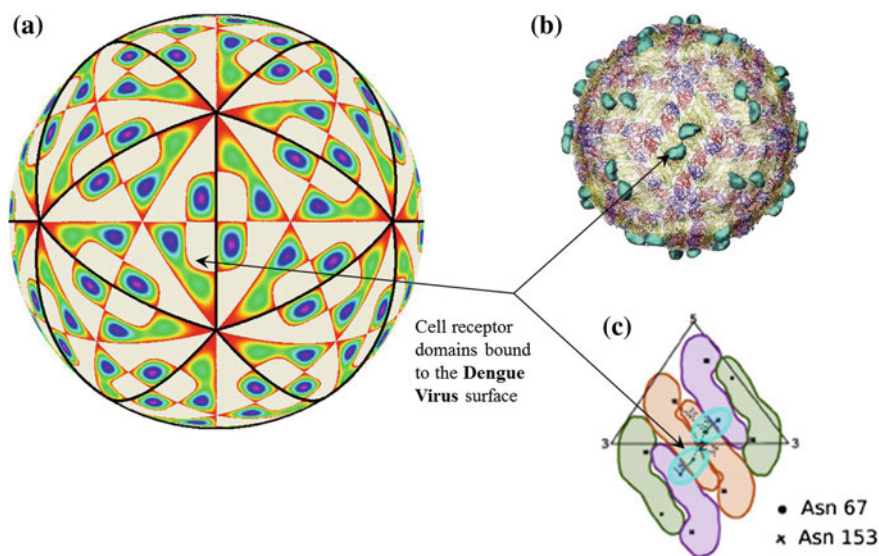


Fig. 6 Protein density distribution and Virus Infectivity. Interaction of the dendritic cell receptor molecules with the capsid of Dengue Virus: **a** irreducible density icosahedral density function with $l = 25$; **b** experimental cryoelectronique density of the Dengue Virus capsid with bound cell receptor domains; **c** schematic zoom of the “rhombus” motif in **(b)**

surface. In the present case it establishes a correspondence of the deepest minima of f_{25} and the binding sites for the carbohydrate recognition domains of the dendritic cell receptors on the DENV virus surface. Additional argument comes from the interpretation of the density function as a *probability* density distribution. From Fig. 6 it is evident that two of the three maxima of f_{25} are sharp picks with high amplitudes of occupation probability while the third one is a wide maximum with a low amplitude. Mean-square amplitude of protein position fluctuations is much high in the latter case than in two former ones. Thus, the attachment of CRD to the strongly fluctuating protein located in the position characterized by a wide maximum is much less probable than in two others.

Along the same line, the correspondence can be established between the positions of maxima for other irreducible density functions and peculiarities of protein distribution in other viruses. For example, sharp maxima of the function with $l = 33$ find themselves in positions situated very close one to another, and leaving between them a sequence of deep and narrow minima. This distribution is close to the experimental protein distribution in Human Rhino-Virus. Deep minima sequence at the surface of this virus is usual called a “canyon”. This canyon is known to be too narrow for human antibodies to enter and to neutralize viral binding sites, but large enough for narrow human cell receptors (more precisely for intracellular adhesion molecules-1, or ICAM-1, of the cell receptors) to enter into the canyon and to bind to the virus protein domains.

4 Perspectives

The approach developed in this work is only the first step of the group theory methods application in virology. Further development of these methods can clarify mechanical properties of viral capsids which have in some cases elastic constants comparable with the hardest materials. One of the most interesting extensions is the mathematical (and physical) analysis of the viral genome (DNA or RNA) dense packing in capsids. Another field where these methods can be applied is the shape transformation usually undergone by viruses during their maturation process which makes them infectious.

Acknowledgments The authors would like to acknowledge financial support by the Laboratory of Excellence NUMEV and RFFR grant N° 13-02-12085. All graphic presentations of the experimental cryoelectronic density images were made with the Multiscale extension [20] to the Chimera interactive molecular graphics package [21, 22].

References

1. Flint, S.J., Enquist, L.W., Racaniello, V.R., Skalka, A.M.: Principles of Virology: Molecular Biology Pathogenesis and Control. ASM, Washington (2000)
2. Baker, T.S., Olson, N.H., Fuller, S.D.: Microbiol. Mol. Biol. Rev. **63**, 862 (1999)
3. Crick, F.H.C., Watson, J.D.: Nature (London) **177**, 473 (1956)
4. Caspar, D.L.D., Klug, A.: Cold Spring Harbor Symp. Quant. Biol. **27**, 1 (1962)
5. Wikipedia: http://en.wikipedia.org/wiki/Icosahedral_symmetry. Cited 7 Jan 2015
6. Naitow, H., et al.: Nat. Struct. Biol. **9**, 725 (2002)
7. Kuhn, R.J., et al.: Cell **108**, 717 (2002)
8. Mukhopadhyay, S., et al.: Science **302**, 248 (2003)
9. Landau, L.D.: Phys. Zs. Sowjet. **11**, 26 (1937)
10. Landau, L.D.: Phys. Zs. Sowjet. **11**, 545 (1937)
11. Alexander, S., McTague, J.: Phys. Rev. Lett. **41**, 702 (1978)
12. Janot, C.: Quasicrystals. A Primer, 2nd edn. Oxford University Press, New York (1994)
13. Bak, P.: Phys. Rev. Lett. **54**, 1517 (1985)
14. Rochal, S.B., Kozinkina, Y.A.: Phys. Rev. B **72**, 024210 (2005)
15. Konevtsova, O.V., Rochal, S.B., Lorman, V.L.: Phys. Lett. A **377**, 1215 (2013)
16. Elliot, J.P., Dawber, P.G.: Symmetry in Physics. Macmillan Press, London (1979)
17. Springer, T.A.: Invariant Theory. Lecture Notes in Mathematics, vol. 585. Springer, Berlin (1997)
18. Lorman, V.L., Rochal, S.B.: Phys. Rev. Lett. **98**, 185502 (2007)
19. Pokidyshcheva, E., et al.: Cell **124**, 485 (2006)
20. Goddard, T.D., Huang, C.C., Ferrin, T.E.: Structure **13**, 473 (2005)
21. UCSF Chimera: <http://www.cgl.ucsf.edu/chimera>. Cited 7 Jan 2015
22. Pettersen, E.F., et al.: J. Comput. Chem. **25**, 1605 (2004)

Applications + Practical Conceptualization +
Mathematics = fruitful Innovation
Proceedings of the Forum of Mathematics for Industry
2014

Anderssen, R.S.; Broadbridge, P.; Fukumoto, Y.;
Kajiwar, K.; Takagi, T.; Verbitskiy, E.; Wakayama, M.
(Eds.)

2016, XIII, 278 p. 109 illus., 95 illus. in color., Hardcover
ISBN: 978-4-431-55341-0