Quasilattice-Based Models for Structural Constraints on Virus Architecture

David Georges Salthouse

PhD

University of York

Department of Mathematics

September 2013

Abstract

Crick and Watson were the first to recognise the importance of symmetry in the structures of viral capsids. This observation was the departure point for Caspar-Klug's theory in which the possible positions and orientations of the protein building blocks are predicted and classified in terms of T-numbers. Whilst this theory predicts the layouts of the protein containers, it provides no information on the thickness of the capsid or its surface features. The creation of icosahedrally invariant point arrays via affine extension of the icosahedral symmetry group and their mapping to viral capsids in [37] has shown that they provide geometrical constraints on viral structure that not only correlate positioning of proteins on the capsid, but also relate structural features on different radial levels including genome organisation.

In this study we have extended this approach using the quasilattices embedding these point arrays. To derive further geometric constraints on virus architecture we firstly show how classifying the possible transitions between the quasilattices modelling the structure of the virus before and after the transition allows us to derive information on the most likely transition paths taken by the protein shell during the structural transformation. Next, a new algorithm matching tile sets to viral capsids has been implemented to investigate further the geometrical constraints quasilattices place on these structures over and above the point arrays in [37].

Contents

List of Figures

List of Tables

Acknowledgements

There are many people I would like to thank for making my research such a great experience. First my supervisor, Prof. Reidun Twarock, without whom this work would not have been possible. For your persistant help, support and enthusiasm which has made the thesis so exciting, thank you. I would also like to thank Julie Wilson and Gustav Delius, members of the thesis committee, for their help and their pertinent questioning. My gratitude also goes to Giuliana Indelicato and Paolo Cermelli without whom quasilattice transition would have been hardly as much fun.

I would also like to take this opportunity to acknowledge my friends who made York feel like home. Tom Keef and Jessica Wardman, my mentors and friends, sharing offices with you has been a privilege and an experience I will never forget. I am indebted to Nick Grayson, a dear friend, who never gave up on trying to taught me English. Also thanks to Adele, Emma, Carl, Tom, Mark, Frances, Myrddin, Jenny and Roger for all those fun evenings. I could have done little without my family whom was never short on encouragement in moments of doubt. Last but not least, I should thank my sweetheart Marie. Your love was the highlight of our months apart.

Author's Declaration

The Introductory chapter *Viruses and Symmetry* summarises previous work on the determination of virus structure using symmetry groups and surface tesselations. The *Cut-and-Project method* chapter compiles the mathematical background necessary for the construction of tilings with non-crystallographic symmetry based on [3, 42]. *Structural transitions in quasicrystals* condenses the work published in the two papers [26] and [27] and includes the figures from these papers, to which I contributed. *The Matching Algorithm* and *Application to viruses structure* constitute entirely my own work.

Chapter 1

Viruses and Symmetry

Viruses are well known for their devastating impact on health and economy, but the development of technology has enabled us to deepen our understanding of these entities and the possible applications of these results in the medical field and nanotechnology are promising. For example, the construction of protein cages from viral proteins (lacking viral genome), referred to as virus-like particles, can be used as gene vectors, i.e. to transport genetic material into cells for therapeutic purposes [49]. These (non-infectious) particles provide containers preventing premature degradation of drugs which, combined with the high host-cell specifity of viruses, can deliver these drugs to the specific targeted tissues. In cancer therapy, viruses carrying gold particles have been used to target cells for photothermal cancer treatment [21].

Understanding how to alter viruses via protein engineering also finds applications in nanotechnology. The detection of biological markers often involves proteins as "catcher" molecules that fix them onto a surface. However, using viruses instead of proteins offers advantages. For instance, once the viral genetic material has been genetically engineered so that its protein shell binds to specific proteins, it can easily be reproduced through infection of new cells and harvested by ultracentrifugation [22]. Microarrays made of viruses have been used for the detection in sera of autoimmune antibodies specific to prostate cancer [81].

During its life cycle a virus needs to infect a cell in order to reproduce. Once the cell is infected, the virus replicates within it using its host's machinery. The newly formed viruses then exit the cell (sometimes killing the cell in the process) and are ready to infect other cells completing the cycle. To prevent degradation of its genetic material (either RNA or DNA) from the extracellular medium and to penetrate cells for infection, viruses have adopted different techniques. These are distinguishable by the following morphologies [18]: helical, icosahedral, enveloped and more complex viruses such as HIV that contains a nucleocapsid with a cone-shaped morphology (see Figure 1.1).

Helical and icosahedral viruses use proteins to protect their genomic material and to interact with the host-cell during the infection process. Other viruses, such as the enveloped viruses, add an extra layer surrounding their protein shell with an outer lipid membrane acquired during the release from the host-cell (also known as "budding process"). Viruses with more complex morphologies exist; one example is the Human immunodeficiency virus (HIV). Although HIV's nucleocapsid is surrounded by an envelope, its nucleocapsid, made up of about 2000 proteins, has a cone-shaped morphology [54].

The work presented here focuses solely on the study of icosahedral protein shells, also called icosahedral capsids or nucelocapsids. In these capsids, proteins are arranged in such a way that the overall structure exhibits the same symmetry as the icosahedron, see Figure 1.2. In structural biology it is common to distinguish different levels of structure when considering proteins. At the lowest level, the primary structure of a protein is made up of a sequence of amino acids. Each amino acid is made up of a backbone which constitutes the "skeleton" (also called main-chain) of the protein, and a side-chain defining the type of the amino acid. About 500 different amino acids were known at the time of publication [80].

Folding of the primary structure, through hydrogen bonding between the protein's main chain, into regular structures constitutes elements of the pro-

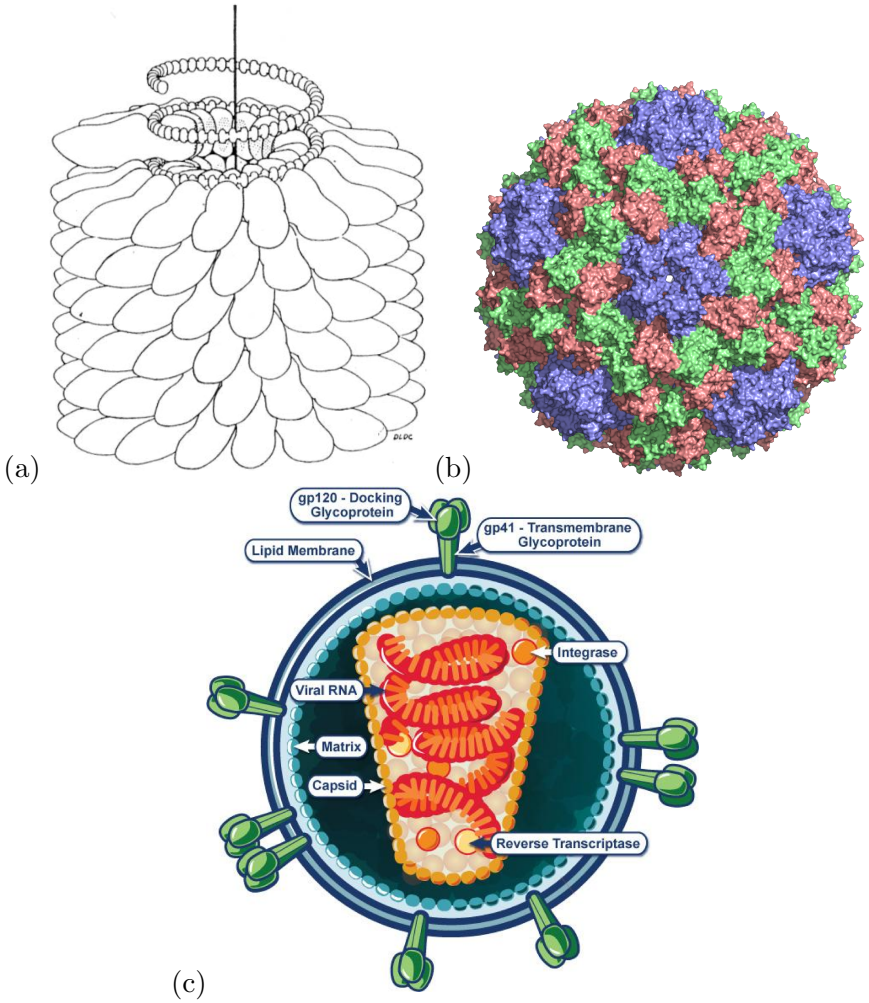


Figure 1.1: Different morphologies of viruses protective shells: (a) representation of the Tobacco Mosaic virus whose RNA is protected by a helical shell made of identical protein subunits [13]; (b) the icosahedral capsid of Cowpea Chlorotic Mottle virus (CCMV) shown along a five-fold axis and rendered using the PyMol software [65]; (c) the cone-shaped nucleocapsid of the Human Immunodeficiency Virus (HIV) type1 surrounded by a lipid membrane [54].

tein's secondary structure (mainly α-helices and β-sheets). The globular form of a single protein unit results from the subsequent folding of the secondary structure's elements, resulting in the so called tertiary structure. The next organisational level is the quaternary structure, a 3D structure of an assembly

of proteins. In some icosahedral capsids, such as in Cowpea Chlorotic Mottle virus, the quaternary structure is formed via assembly of pentamers (cluster of 5 proteins) and hexamers (cluster of 6 proteins).

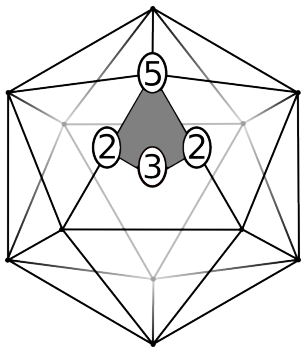


Figure 1.2: Illustration of the icosahedral group. The 5-, 3-, 2-fold symmetry axes bordering the fundamental domain (grey kite) are indicated with reference to an icosahedron.

To study these viral capsids, whose radius may vary from 88 Å to 925 Å, libraries (Protein database [9] and Viperdb [12]) of Protein database files (pdb files) containing the positions of atoms have been created based on X-ray and Cryo-EM experiments [72]. In the rest of this chapter, earlier mathematical models developed to investigate the structures of these capsids through symmetry will be introduced, starting with Caspar-Klug's theory of quasi-equivalence.

1.1 Previous Mathematical models

Crick and Watson [18] observed that many virus capsids have icosahedral or helical shape if these are made of structurally equivalent subunits. The underlying reason for this was believed to be related to genetic economy. Using symmetry, the virus can code the maximal number of identical protein subunits using the same genetic sequence. A consequence is that fewer genetic material is needed to code the entire capsid making it easier to package. This was the first step towards quasi-equivalence theory.

1.1.1 Quasi-Equivalence theory

A protein shell where all proteins occupy equivalent positions is composed of precisely 60 proteins, one for each element of the icosahedral group. In this case the same type of bonds between each protein are used all over the protein shell. To account for viruses with more than 60 proteins, Caspar and Klug introduced the notion of *quasi-equivalence* [13]. Here, the capsid is formed of $60T$ (with $T \geq 1$) protein subunits but the bonds between proteins not related by symmetry can be different, and the proteins may have different conformations, i.e. they have different tertiary structures. To represent the possible arrangements of proteins in the capsid, one could classify all ways of paving a sphere by the folding of a planar net. It has been shown [57], that these planar nets could only have 4-fold (the square lattice with square units) or 6-fold symmetry (hexagonal lattice with triangular units). Because a square net can only be folded orthogonaly, the triangular net is preferred as its folding requires smaller bendings between triangles, i.e. allowing for the maximal degree of quasi-equivalence.

In quasi-equivalence theory, different protein organisations are given by triangulations with icosahedral symmetry, which are called icosideltahedra. These are represented as onto planar nets as demonstrated for the icosahedron in Figure 1.3. The positions of the proteins in the corresponding capsid are then encoded by the corners of the triangular faces. At each vertex of the icosahedron, five triangular facets meet forming clusters of 5 proteins (i.e. pentamers).

Enumerating all icosadeltahedra is equivalent to finding all possible sub-triangulations of the icosahedron. First, we choose one vertex as the origin O and two vectors $\boldsymbol{h}$ and $\boldsymbol{k}$, at an angle of $\frac{\pi}{3}$, forming a basis of the hexagonal lattice, see Figure 1.4. Second, we represent the edge of one icosahedral face by the segment OS defined by the coordinates (H, K) of vertex S in the basis $(\boldsymbol{h}, \boldsymbol{k})$, where H and K indicate the numbers of steps between midpoints of hexagons along the directions $\boldsymbol{h}$ and $\boldsymbol{k}$, respectively. The icosahedral face is

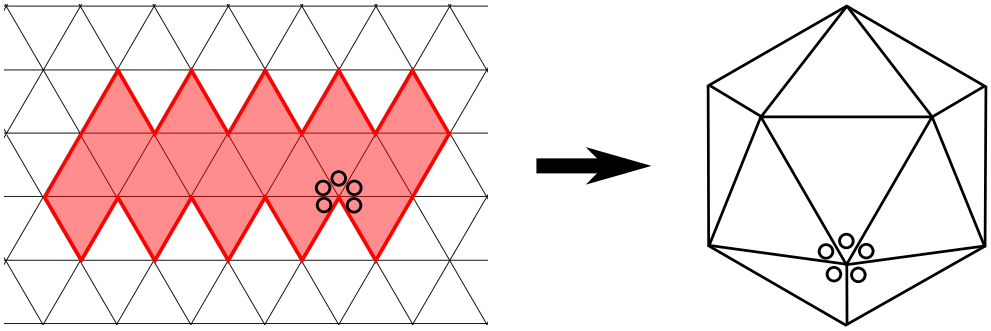


Figure 1.3: Folding of the icosahedron from the triangular net.

then subdivided by the smaller triangles congruent to the (O, h, k) triangle as shown in Figure 1.4.

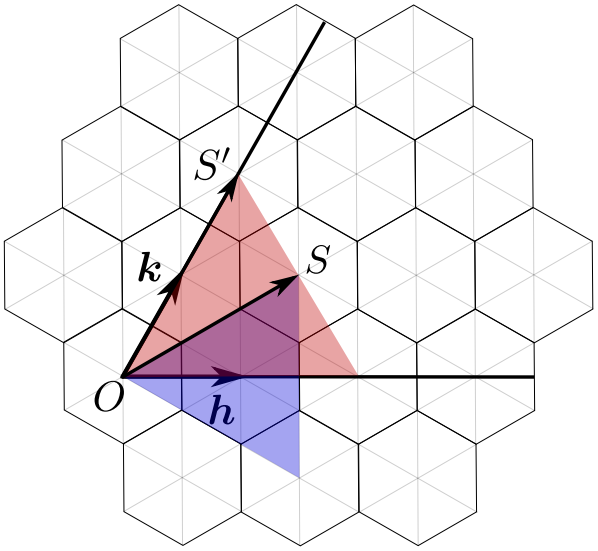


Figure 1.4: Subtriangulations of an icosahedral face are shown for two edges OS and OS' corresponding to icosadeltahedra $T = 3$ and $T = 4$ in Caspar-Klug's classification. A flattened and a folded version of these deltahedra are shown in Figure 1.5.

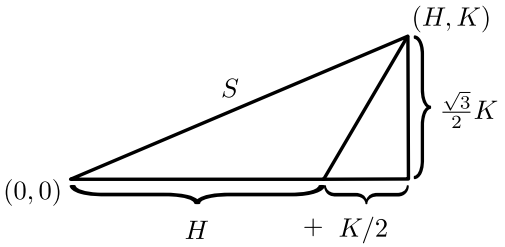
Finding all possible subtriangulations is then equivalent to determining all possible lengths, in terms of the $(\boldsymbol{h}, \boldsymbol{k})$ basis, of segment OS. Note that, because of symmetry, we can limit the analysis to positive values for the coordinates of H and K. The triangulation number T, which quantifies the

subtriangulation, is defined as the ratio of the area of an icosahedral face over that of a triangle:

$$T = \frac{\frac{\sqrt{3}}{4}S^2}{\frac{\sqrt{3}}{4}} = S^2$$

Using Pythagoras' theorem, we obtain:

$$\begin{aligned} S^2 &= (H + \frac{K}{2})^2 + (\frac{\sqrt{3}}{2}K)^2 \\ &= H^2 + HK + K^2 \end{aligned}$$



Hence:

$$T = H^2 + HK + K^2 \text{ with } (H, K) \in \mathbb{Z}^2.$$

The lowest non-tivial value of T is 1 for $(H, K) = (0, 1)$ or $(H, K) = (1, 0)$ which corresponds to the standard icosahedron. Higher values of T correspond to further subdivisions of the icosahedral faces. Examples of folded nets corresponding to $T = 3$ and $T = 4$ icosadeltahedra are shown in Figure 1.5. Notice that there exist no values for H and K such that $T = 2$ and that the lowest non trivial value of T is 3. These icosadeltahedra encode the positions of the $60T$ proteins with precisely 12 pentamers (one around each of the 5-fold axes of the icosahedron) and $10(T\text{-}1)$ hexamers (i.e. clusters of six proteins).

We have seen that, due to the symmetry of the hexagonal lattice, only positive values of H and K need be considered. Note that the line $H = K$ is also a symmetry axis. This mirror symmetry leads to enantiomorphous structures for values of $T = 7$, 13 , 19... as can be seen in Figure 1.6. To distinguish between two enantiomers for a given T-number, we call the enantiomer *laevo* if $H > K > 0$ and *dextro* if $K > H > 0$.

The structures of many viruses can be represented by quasi-equivalence, provided that the same type of bonding environment occurs for every protein

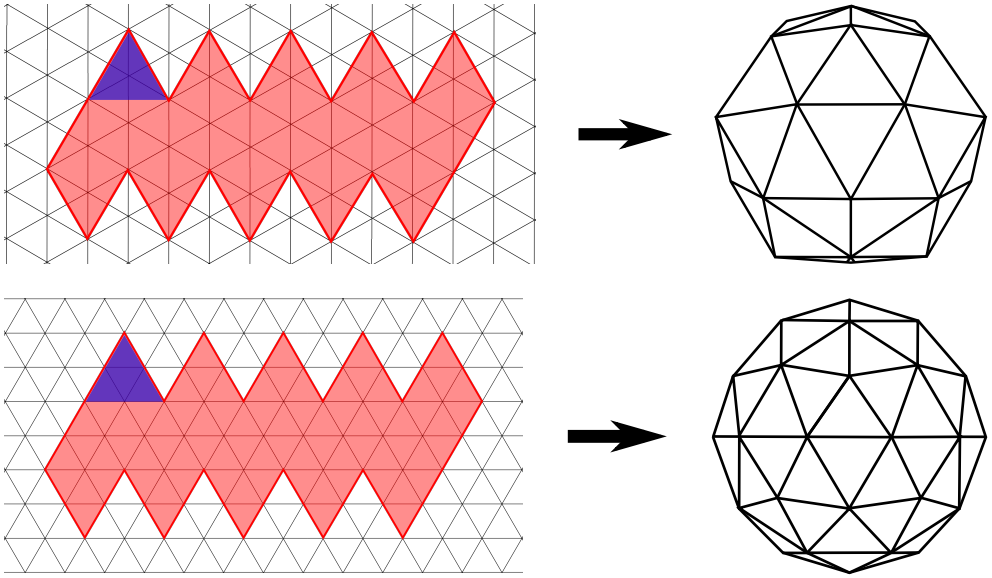


Figure 1.5: Folding of the triangular net of the icodeltahedra corresponding to the T=3 (top row) and T=4 (bottom row) in Caspar and Klug's classification. In each case one of the faceted icosahedral faces has been highlighted in blue.

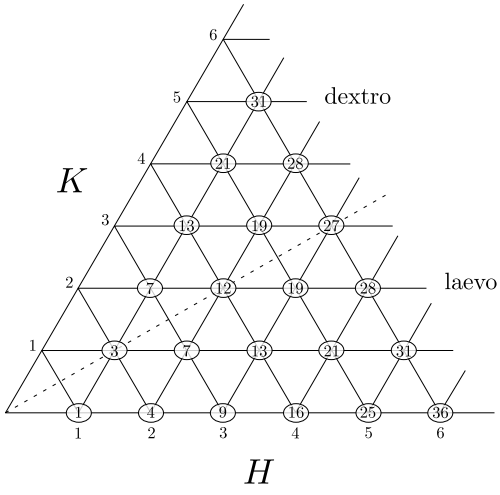


Figure 1.6: Possible values of T depending on H and K. For $T = 7$ mirror images exist and $(H > K)$ or $(K > H)$ distinguishes laevo from dextro configurations.

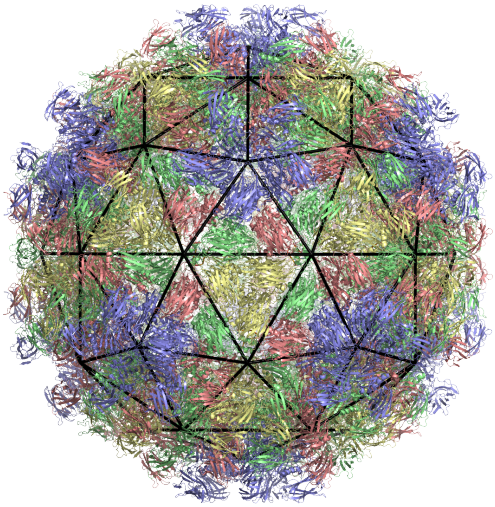


Figure 1.7: A $T = 4$ triangulation superimposed on the capsid of Providence virus pinpoints the position of proteins at the corners of the triangular faces.

in the capsid. In addition, it has been shown in [86] that symmetry occurs as a consequence of energy minimisation. To demonstrate this, equilibrium states are explored using Monte Carlo techniques applied to a coarse-grain model. To describe the capsomere-capsomere interaction potential, a van der Waals type potential energy $V(r)$ (with r the distance between the centers of the interacting capsomeres) includes a short range repulsion (representing rigidity of the capsomere) and a long-range attraction (i.e. the driving force of aggregation). This coarse-grained model includes two different morphological units, hexamers and pentamers represented by the internal state H and P, respectively. The minimum of $V(r)$ acknowledges the geometrical size difference between pentamers and hexamers of the same edge length, and an energy difference ΔE between a P and an H capsomere (taking into account different individual contact interactions and folding conformations of the proteins) is implemented. The N interacting capsomeres are allowed to freely move on a spherical surface, each able to switch between the P and H states to explore all possible configurations. The energy ε is plotted as a function of N in Figure 1.8, using the capsomere-capsomere binding energy ε_0 as energy unit.

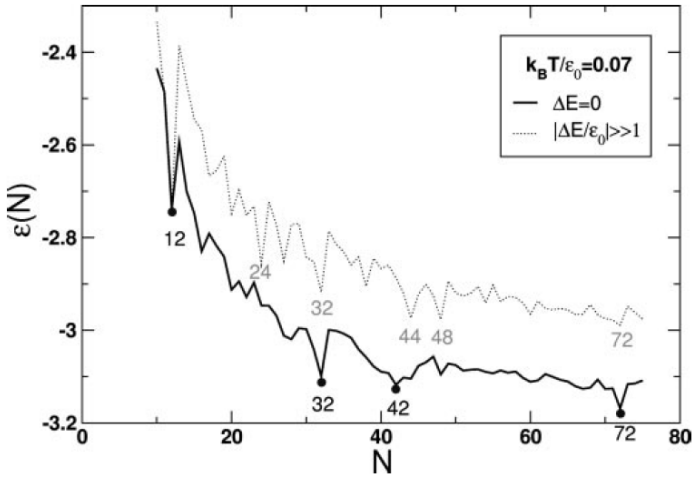


Figure 1.8: Searching for capsids composed of N capsomeres which correspond to a minimisation of the overall structure's energy. Energy per capsomere for $\Delta E = 0$ (black curve) and for $\dfrac{\Delta E}{\varepsilon_0} >> 1$ (dotted curve) are plotted as functions of the number of capsomeres N in the capsid, figure adapted from [86].

In the case $\Delta E = 0$, one state (P or H) is not preferred over the other and capsids are composed of a combination of the two. The minima of the structure's energy are located at $N = 12, 32, 42$ and 72 which correspond to Caspar-Klug's $T = 1, 3, 4$ and 7 structures, respectively. This supports the hypothesis that icosahedral symmetry and T number structure are consequences of energy minimisation. For $\|\Delta E\| >> \varepsilon_0$ capsids are uniquely made up of hexamers or pentamers depending on the sign of ΔE. Forcing all capsomeres to be in the same state (either P or H) has a dramatic effect on icosahedral symmetry and only the Caspar-Klug structure $T = 1$ ($N = 12$) and $T = 3$ ($N = 32$) are still observed. Other minima of $\varepsilon(N)$ are observed at $N = 24$ and $N = 48$, both exhibiting octahedral symmetry. Under these circumstances (i.e. $\|\Delta E\| >> \varepsilon_0$), it has been shown that external pressure on the capsid is necessary to facilitate icosahedral symmetry.

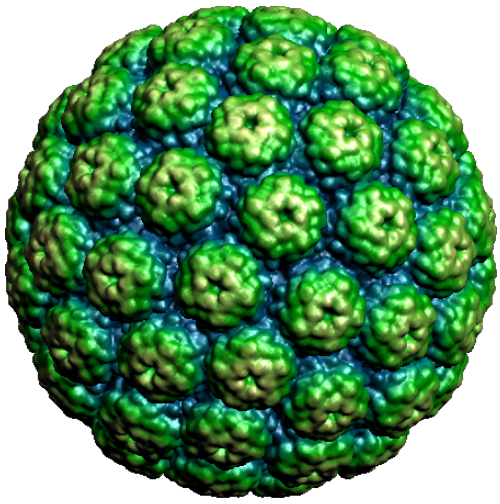


Figure 1.9: The Simian virus 40 (*pdb-id*:1sva from [12]) falls outside quasi-equivalence's scope as its capsid is entirely composed of pentamers which is not possible with a uniform type of inter-subunit interaction.

1.1.2 Viral Tiling Theory

Although many capsid structures conform to Caspar-Klug theory, the improvement of experimental techniques has led to the discovery of an increasing number of capsids which fall outside quasi-equivalence's scope. This is the case for Polyoma-like viruses (a family of cancer-causing viruses) whose protective shells are entirely composed of pentamers. The example of the Simian virus 40 is shown in Figure 1.9 where the 360 proteins form 72 pentamers, for which the positions of the capsomeres correspond to those of the $T = 7d$ (i.e. 7 dextro) capsid, albeit with different bonding environment.

To include these capsids into a broader theory required the inclusion of non-quasi-equivalent bonding environments. A generalisation, called Viral Tiling Theory was developed [79], and extends spherical tesselations with overall icosahedral symmetry to non-triangular building-blocks. In the following, we refer to these building blocks as tiles using tiling theory terminology. This work is closely related to Penrose tilings [66] as the tiles have similar shape, but here these tesselate spherical objects instead of a planar surface. Like

the generalisation of tesselations to non-triangular tiles and the mixture of different tiles, the concept of quasi-equivalence needs to be broadened, too to include different types of bonding environments (such as a mix of dimer and trimer interactions). The generalised quasi-equivalence guarantees that identical types of proteins occupy similar positions on the tile by locating protein subunits at the corners of the tile subtending the same angle. This generalised concept encompasses quasi-equivalence, for which proteins are located at the corners of equilateral triangular tiles.

By allowing more than one bonding environment, Viral Tiling Theory better describes inter-capsomere and intra-capsomere interactions between proteins. The example of the Simian Virus 40 is shown in Figure 1.10, where representation of the 72 pentamers requires two different tiles, the rhomb and the kite. Note that here colours represent positions that are equivalent under icosahedral symmetry. Viral Tiling Theory allows for different inter-capsomere bonds. These are either dimer interactions, i.e. interactions between two protein subunits (represented here by the rhomb-shaped tiles), or trimer interactions, in which three protein subunits interact with each other (represented here by a kite tile). Interactions take place via the C-terminal arms of the protein subunits. In a dimer interaction, this occurs via an exchange of C-terminal arms by the two protein subunits represented on the tile (see double line on the rhomb tiles); in a trimer interaction, each subunit receives a C-terminal arm from a protein subunit different from the one it extends its C-terminal arm to (see single lines on the kite tile).

Even though quasi-equivalence theory and its generalisation to Viral Tiling Theory predict the possible arrangements of proteins on a 2 dimensional layout, they provide no information on the 3D structure of the virus, e.g. in terms of radial extent and surface features. We introduce in the following section previous mathematical models which investigate the full 3D organisation of virus capsids.

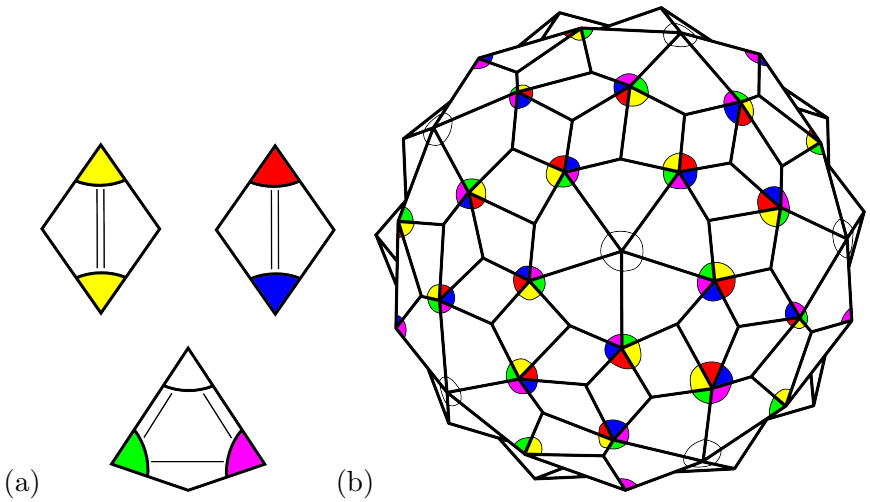


Figure 1.10: Application of Viral Tiling theory to Simian Virus 40 whose capsid is formed of 72 pentamers. Here the Viral Tiling theory encodes three different bonding environments: the red and blue and the yellow rhombs (i.e. two protein subunits interacting) and a kite (i.e. three protein subunits interacting).

1.1.3 Janner's approach

Janner was the first to use lattices to study the structure of clusters of proteins and icosahedral viruses. We will briefly explain here his method [29, 30] applied to different variants (or serotypes) of the human rhinovirus capsids. This T=3 virus counts approximatively 100 different serotypes of which serotypes 1a, 2, 3, 14, and 16 have been succesfully crystallised and used as a tutorial example in [29] and [31]. Note that even though two different serotypes have the same overall shape and the same architecture (that is same number and same overall arrangement of coat proteins) there are differences in the protein structures and in particular the antigens located on the surface of the capsid proteins are different. Since the crystallisation of an ensemble of particles depends on these surface features, two serotypes may crystallise with different symmetries. In this series of papers, Janner investigated the relationship between serotype and crystal symmetry and was able to define *fingerprints* using 3D lattices

which encode the esssential properties (packing and symmetry) of the crystal structure of the viruses.

First the crystal structure of viruses is represented by the crystal lattice Λ, where each of its vertices coincides with the centers of the capsids in the crystal, see Figure 1.11.

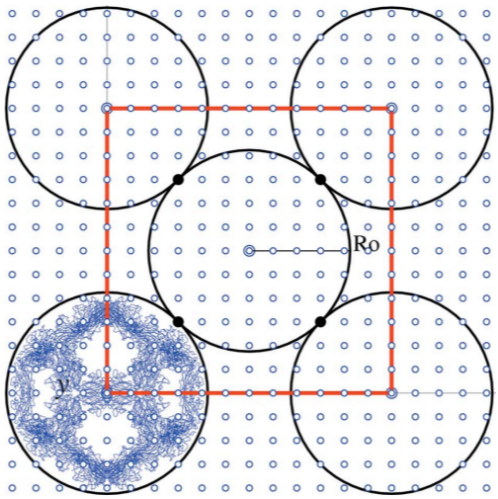


Figure 1.11: Example of crystal lattice Λ and packing Λ_P corresponding to the serotype 14 of Human rhinovirus adapted from [31]. The vertices of Λ, located at the centre of each capsid in the crystal, are shown as double circles. Spheres of radius R_0 circumscribe capsids in the crystal and kissing points (shown as filled circles) are located wherever two spheres touch. To index these kissing points, the finer lattice Λ_P (shown as open circles) is used.

Using spheres circumscribing each protein shell in the crystal, the proximity between two neighboring capsids is encoded by the *kissing points* defined as the points where two spheres "touch". The residues closest to these kissing points, called the kissing point residues (or KPR), are believed to play an important role in the clusterisation process due to their proximity with neighboring capsids.

To account for the symmetry of the capsid, an additional enclosing form is chosen for the rhinovirus capsids (independently of their serotype): an ico-

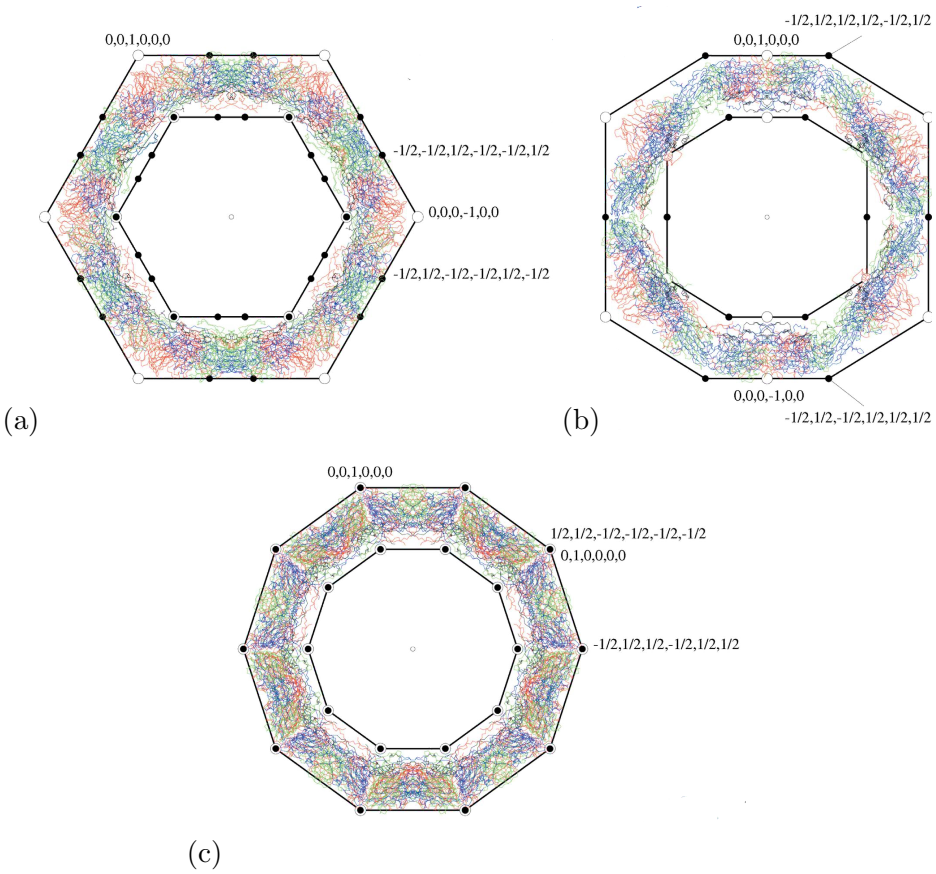


Figure 1.12: Capsid of the human rhinovirus with the ico-dodecahedra used as an enclosing form in [29], seen along a 3-fold (a), 2-fold (b) and 5-fold axis (c). Vertices of the ico-dodecahedra can be indexed using either the 6-dimensional simple cubic lattice (represented as open circles) or the body-centered cubic lattice (shown as filled circles). Projection of vertices from 6D lattices (explained in detail in Chapter 2) are indexed by their coordinates in the 6D standard Euclidean basis.

dodecahedron (i.e. the projection of the unit cube in the 6D SC lattice) scaled to the outer surface and a $\frac{1}{\tau}$-rescaled ico-dodecahedron matching the inner surface of the protective shell, see Figure 1.12. However, it is known that no 3D lattice can embed the vertices of the ico-dodecahedron because of the crystallographic restriction [66]. To circumvent this problem and better

approximate the capsid without involving higher dimensional lattices, a finer *packing* lattice Λ_P is used. Since the vertices of the ico-dodecahedron used to approximate the virus structure do not form a lattice in 3D, Janner uses the vertices of the cubes circumscribing the outer ico-dodecahedron and inscribed into the inner ico-dodecahedron. The lattice Λ_P is constructed to contain all vertices of these cubes. By construction Λ is a sub-lattice of Λ_P, see Figure 1.11. Λ_P can be considered as a representation of both the crystal lattice and the orientation of each capsid. Indexing the kissing points in Λ_P provides a unique fingerprint for each of the studied serotypes. The example of the fingerprints for serotypes 3 and 14 is shown in Figure 1.13, superimposed with the coat protein VP2 and the corresponding lattice Λ_P.

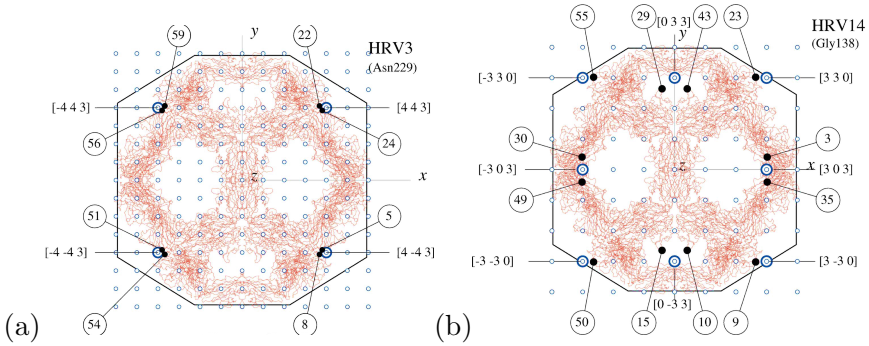


Figure 1.13: The VP2 coat protein of human rhinovirus superimposed on the Λ_P-lattice for serotypes 3 (a) and 14 (b). The positions of residues closest to these kissing points, believed to play a role in clusterisation, are marked by filled circles and indexed by the number of its associated chain as given in pdb-file. Together with the indexed kissing points (represented by double circles) in Λ_P, a fingerprint of a capsids crystal structure is constructed enabling the reconstruction of its essential features.

1.2 Extended Symmetry Group

The work presented in this section extends that by Janner as it considers enclosing forms of the viral capsids using icosahedrally invariant points sets, which are related to the vertex sets of lattice-like structures (quasilattices) with non-crystallographic symmetry. We will discuss in more depth the construction of these point sets as they will be used in the study of the dynamics in the transition undergone by the Cowpea Chlorotic Mottle virus in Chapter 3.

As mentioned above icosahedral symmetry is non-crystallographic in 3D space therefore no lattice can be constructed in 3D with the same symmetry as the viral capsids. To gain knowledge of the radial distribution of material in a virus, such as the thickness of its capsid and the features of its genome organisation (which are not accessible via Caspar-Klug theory or Viral Tiling theory) an affine extension of the icosahedral group has been used [35].

In general, the construction is as follows. Using group theory terminology, a group element is a letter and combination of these letters forms a word, point sets correspond to words in the extended symmetry group. They are iteratively constructed using a base shape that is invariant under a desired symmetry group and represents the action of the symmetry group geometrically. After translation along one of its symmetry axes corresponding to the affinisation of the group, i.e. the introduction of a translation, and subsequent application of the symmetry group, a point set with the same symmetries as the base shape is constructed. For arbitrary translations, the infinite group corresponding to the resulting point array corresponds to the free group, i.e. has only trivial mathematical structure, because there are no relations between words in the group generators (elements in the infinite group) other than those already present in the finite group. Therefore we introduce the concept of admissible translation. Following [35], we call *admissible* those translations leading to degeneracy of the points (given by the vertices of the translated and rotated copies of the base shape) after application of the whole symmetry group. We

denote the translation by $\lambda \boldsymbol{t}$, where $\boldsymbol{t}$ is the unit vector along a symmetry axis and λ a scaling factor. Each vertex of these point sets can be obtained by application of a product of the affine extended group elements to a vertex of the base shape. The construction of these point sets is analogous to the construction of lattices from affine extensions of symmetry groups. A 2D example of such a construction is shown in Figure 1.14 demonstrating the construction of an hexagonal lattice and of a point set exhibiting the 2D non-crystallographic 5-fold symmetry.

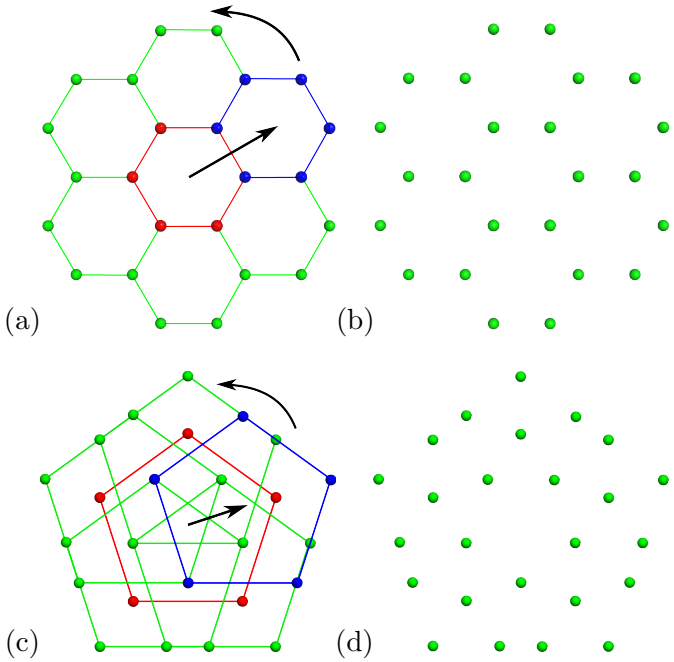


Figure 1.14: Comparison between the 2D construction of an hexagonal lattice and a point set with the non-crystallographic 5-fold via affine extension of symmetry groups. *first row*: creation of an hexagonal lattice through translation of a red hexagon to a blue hexagon and subsequent rotation (green hexagons) (a). Vertices are part of the hexagonal lattice (b). *second row*: Similarly, translating the red pentagon through a symmetry axis followed by 5-fold rotation (green pentagons) in (c), a point set with 5-fold symmetry (non-crystallographic) is constructed (d).

For a set of vertices with icosahedral symmetry, the base shape can be

chosen as one of the three polygons invariant under the icosahedral group and with minimal cardinality, i.e. the icosahedron, dodecahedron or the icosidodecahedron represented in Figure 1.15. As will be shown in Chapter 2, these base shapes appear naturally within the construction of icosahedral quasilattices (i.e. set of ordered points that lacks lattice translational symmetry).

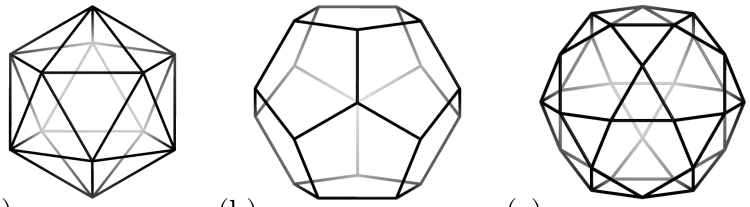


Figure 1.15: The icosahedron (a), the dodecahedron (b) and the icosidodecahedron (c) are the three base shapes used in the construction of the 3D point sets with icosahedral symmetry in [35]. Note that they correspond to projections of the three icosahedral lattice types in 6D (see Chapter 2).

Exhaustive computation of admissible translations corresponding to the three icosahedrally invariant base shapes resulted in a library of 55 point sets [35] which correspond to all words containing at least one occurence of the translation λt. Allowing for multiple uses of the non-compact operator $\lambda \boldsymbol{t}$ leads to spatially extended and denser point sets. Keeping in mind that these must model structural constraints on viral capsids, application of $\lambda \boldsymbol{t}$ should be limited. An analysis of a range of viruses [83] suggests that, although the thickness of the viral capsid may vary, allowing a single translation $\lambda \boldsymbol{t}$ in the construction of the point array, is enough to represent the structure of small viruses (up to $T = 4$ and in some cases $T = 7$). A preferred way to create denser point sets is by combination of existing ones. Two point sets A_1 and A_2 are defined *compatible* if their associated translations t_1 and t_2 are along the same symmetry axis and A_1 is λ-rescaled such that $\lambda\lambda_1 = \lambda_2$. Note that in this case, $\lambda\lambda_1 t_1 = \lambda_2 t_2$ is an admissible translation of both base shapes

λA_1 (i.e. A_1 rescaled by λ) and A_2. Combining the 55 point sets in this way produced a library of 569 vertex sets which we call *viral configurations* or *point arrays*.

To compare point arrays with a given viral capsid the only degree of freedom is the relative size of the point array to the protein shell, i.e. an overall scaling factor, since the orientation of the capsid as well as the point set are fixed by the axes of symmetry. Fixing the size of the protein container, an automated algorithm has been developed in [37] to scale point arrays such that the outer vertices match the outermost features of the viral capsid. Each of the rescaled point arrays is then scored according to its proximity to atomic positions in the viral capsid. As only structural information regarding the capsid is used, three radii have been defined. The *inner radius* is given as the minimal distance of any atoms in the capsid from its centre, the *virus radius* (maximal distance of any capsid's atoms from the centre of the capsid plus a van der Waals radius of 1.9Å) and the *mean radius* as mean of inner and virus radii. Vertices positioned below a 4Å threshold below the inner radius are not included in the score as they may correspond to geometrical constraints on the genomic material. Vertices below the mean radius contribute to the scoring if they represent a good match to the capsid, that is if they are within 4Å of the capsid's material. By contrast, vertices above the mean radius cannot possibly match to the genomic material and they are thus required to match features of the capsid. Note that, by construction, two point arrays may have the same "outer" points, but different "inner" points. It is therefore possible to have more than one good match for a given virus, possibly representing pleomorphic interiors. The example presented in Figure 1.16 shows for the $T = 3$ Pariacoto virus how these point arrays can be thought of as geometrical constraints and how these are related at different radial levels.

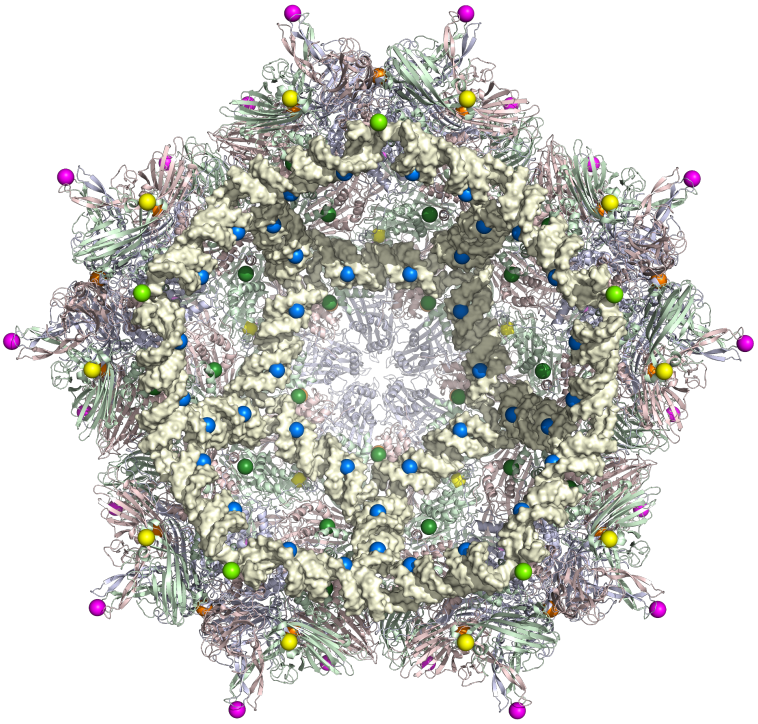


Figure 1.16: Application of the matching algorithm to the Pariacoto Virus [37]: Here the scaling was defined by the positions of the magneta points fitting the 15Å-high protrusions located at the quasi-three fold axes. Points in magenta, yellow, orange and light green are scored to the protein shell. The RNA dodecahedral cage (not used in the fitting algorithm) is predicted by the dark green and blue vertices providing information on the major and minor grooves of the RNA, respectively.

1.3 Motivation for this study

The point arrays are by construction a subset of the vertex set of a quasilattice as demonstrated later in Chapter 3. In this study we have constructed the full 3D tilings, of which these point arrays are part, to investigate if these structures correspond to additional geometric constraints and to better understand the role of symmetry in virus structure. Other models based on quasilattice theory have been developed [43, 71] to understand the underlying

principles of new ordered non-crystallographic structures (such as i-AlPdMn quasicrystals), see Figure 1.17.

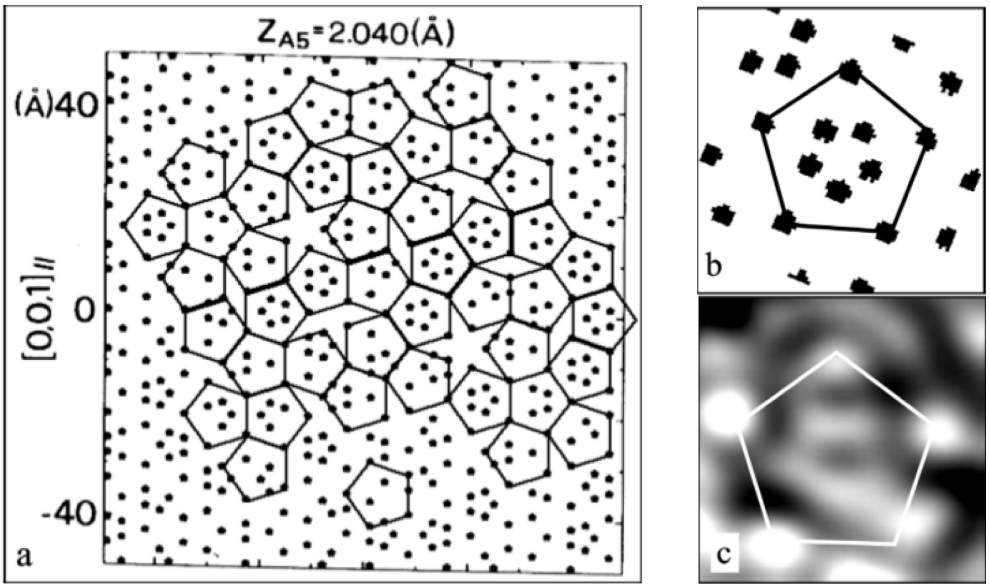


Figure 1.17: Match of a tiling to the i-AlPdMn quasicrystal determined via Scanning Tunneling Microscopy (STM) experiment. A tiling is constructed using the acute rhombus, crown and pentagon as tiles whose vertices match atomic positions in the quasicrystal (a). A close up view shows that tiles have an internal structure (b), i.e. these also contain atoms, shown as high contrast region from the STM output (c). Figures adapted from [43].

The present study investigates how a completed quasilattice-based model can help us understand the underlying principles of very different structures such as viruses. If we can indeed use tilings to model the structure of viral capsids, we can assume that it supports the energy minimisation structure hypothesis as in the case of tilings mapping quasicrystal alloys. If so, these geometrical constraints may have a direct influence on the evolution of viruses. Indeed inferring evolutionary relationship between viruses using sequence comparison of genomic material is difficult for long-range viral relationships because of high mutation rates. Instead morphological similarities between capsids are used as indicators [5, 6, 23]. Thanks to improved experi-

mental techniques, a growing number of structures have been determined and it has become clear that some protein shells are strikingly similar even if the corresponding viruses are from different families and infect different domains of life (i.e. archæa, bacteria or eukaryota). Whether protective shells have similar morphologies due to minimisation of their structure's energy or as a result of being closely related viruses, i.e. due to descendance from a common ancestor, is still an open question. However, this study supports the hypothesis that these similarities may be the consequences of structural constraints on viral evolution towards a more stable capsid design.

Chapter 2

Cut-and-Project method

2.1 Introduction

The cut-and-project method is the most commonly used technique to construct an aperiodic tiling with a given symmetry group [42, 56]. This technique projects elements (such as vertices, edges, etc) from a higher dimensional lattice $\mathcal{L}$ onto two subspaces that are invariant under the symmetry group, see Figure 2.1. We call these two subspaces $E_{\parallel}$ and $E_{\perp}$, often referred as the *parallel* or *physical* space and the *orthogonal* or *control* space, respectively. As will be seen in subsequent sections, the *physical* space $E_{\parallel}$ corresponds to the space in which the tiling is constructed. The selection rule, defined in the orthogonal space $E_{\perp}$, specifies which elements from $\mathcal{L}$ project onto $E_{\parallel}$ such that the constructed tiling contains no holes, no overlapping tiles and is invariant under a specific point group G. The two requirements for a projected quasilattice and its associated tiling to have G as a symmetry group are that: (i) G must be a subgroup of the holohedry (i.e. the maximal point group symmetry) of the lattice $\mathcal{L}$, and (ii) both subspaces $E_{\parallel}$ and $E_{\perp}$ must be invariant under G. As we will see for 5-fold symmetric and icosahedral tilings, the desired symmetry group G fixes the minimal dimensionality of the lattice $\mathcal{L}$ from which the tiling is obtained via projection. Other techniques could have been used to construct quasilattices such as the grid method [66]. Even though this

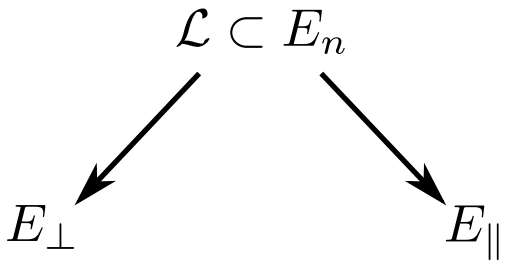


Figure 2.1: Illustration of the cut-and-project method. Let $\mathcal{L}$ be a lattice, embedded into an Euclidean space E_n, and let $E_\parallel$ and $E_\perp$ be two subspaces of E_n that are invariant under a given symmetry group. The tiling, constructed in $E_\parallel$, is made of projected elements from $\mathcal{L}$. To prevent tile overlaps or the presence of holes in the tiling, the cut-and-project method defines the selection criteria on lattice elements in the control space $E_\perp$.

technique has been shown to be equivalent to the cut-and-project method [24], we prefer the latter as it provides a link with the higher-dimensional lattices used by Janner to index his models. Moreover, the point arrays discussed in Chapter 1 are uniquely embedded into a 6D lattice. Projecting the other elements of this lattice is a natural extension of the point arrays to a quasilattice and introduces additional geometric constraints. Next, the embedding into higher dimensional lattices allows the use of 3D crystal transition techniques to study structural transitions of viral capsids with minimal symmetry loss.
In this chapter, we will focus on the creation of tilings with 5-fold symmetry (in Section 2.3) and icosahedral symmetry (in Section 2.4). The icosahedral tilings are used, in Chapters 4 and 5, to construct models for viral capsids that encase the protein structure akin to Janner's encasing forms. Structural transitions of tilings with 5-fold or icosahedral symmetry are studied in Chapter 3 and applied to the study of capsid transitions.

2.2 The cut-and-project method

We review here the cut-and-project method for the construction of aperiodic tilings from higher-dimensional lattices. In the remainder of this chapter we assume that $\mathcal{L}$ is an n-dimensional lattice embedded into an Euclidean space E_n of dimension n and that the two subspaces $E_\parallel$ and $E_\perp$, with $E_n = E_\parallel \otimes E_\perp$ and $E_\parallel \perp E_\perp$, are invariant under a symmetry group G, which is a subgroup of the holohedry of $\mathcal{L}$. Before we define the selection rules specifying which elements of $\mathcal{L}$ will be part of the projected tiling in $E_\parallel$, let us recall a few definitions from lattice theory (see [66] and [17]):

Definition 2.2.1. *Given a lattice $\mathcal{L}$ in E_n and $q \in \mathcal{L}$, the set:*

$$V(q) = \{u \in E_n; \mid q - u \mid \leq \mid y - u \mid, \forall y \in \mathcal{L} - \{q\}\}. \tag{2.1}$$

is called the Voronoi cell of the centre vertex q.

Note that because all vertices of a lattice are equivalent modulo translations so are the associated Voronoi cells, that is $V(q) = V(q')+q-q'$ for any $q, q' \in \mathcal{L}$. An example of the construction of Voronoi cells within a 2D lattice is presented in Figure 2.2. Having defined the Voronoi cell of a lattice $\mathcal{L}$, we call $X(q)$ an m-boundary (i.e. a boundary of dimension m such that $m<n$) of $V(q)$, and define the dual $X^*(q)$ of X as follows.

Definition 2.2.2. *Let $S(X(q))$ be the set of vertices $k \in \mathcal{L}$ whose Voronoi cells $V(k)$ share the same $X(q)$ as a boundary, i.e:*

$$S(X(q)) = \{k \in \mathcal{L} \mid X(q) \subseteq V(k)\} \tag{2.2}$$

Then the convex hull of $S(X(q))$ is called the dual of $X(q)$ and is denoted as $X^*(q)$.

By definition the dual of $q \in \mathcal{L}$ is its associated Voronoi cell $V(q)$. The duals of vertices of $V(q)$ (also known as "holes" of $\mathcal{L}$) are the so-called Delone cells [66].

This definition of dual leads to the following properties:

1. The dual of X^* is $X^{**} = X$;

2. if X is an m-boundary then X^* is an (n-m)-boundary;

3. if $X_1 \subset X_2$ then $X_1^* \supset X_2^*$.

Note that, by construction, Voronoi cells, their boundaries and the associated duals are convex polytopes. Calculations of these have been performed using the Qhull package [7].

2.2.1 Selection rule

This section summarises the selection rule used to prevent the overlapping of tiles in $E_\parallel$ using the so-called klotz tiles. To illustrate this concept, we assume that $E_\perp$ is an m-dimensional subspace (with $m<n$) and $E_\parallel$ is (n-m)-dimensional (since $E_n = E_\parallel \otimes E_\perp$). We define $X(q)$, an m-dimensional boundary of the Voronoi cell $V(q)$, and $X^*(q)$ its (n-m)-dimensional dual.

Definition 2.2.3. *The n-dimensional klotz tiles constructed from $V(q)$ are given by:*

$$X^*_{\parallel j}(q) \otimes X_{\perp j}(q) \quad j = 1, ..., N(m) \tag{2.3}$$

where $N(m)$ is the number of m-boundaries in $V(q)$, $X^*_{\parallel j}(q)$ is the projection of $X^*_j(q)$ into $E_\parallel$ and $X_{\perp j}(q)$ the projection of $X_j(q)$ into $E_\perp$.

It has been shown in [42] that since the lattice $\mathcal{L}$ spans E_n, the klotz tiling paves E_n, i.e. there are no two overlapping tiles and every point of E_n belongs to a klotz tile. A 2D example of this construction is shown in Figure 2.2.

We can now state the *selection rule* defined by the cut-and-project method which prevents the overlapping of any two tiles in $E_\parallel$ as follows:
$X^*_\parallel(q)$ is included into the tiling in $E_\parallel$, which we call $\mathcal{T}^*$, if and only if $E_\parallel$ *cuts* the corresponding klotz tile. If we assume moreover that $E_\parallel \cap \mathcal{L} = \{O\}$, with O the origin of $\mathcal{L}$, then $E_\parallel$ cuts a klotz tile if $O_\perp \in X_\perp(q)$. By construction of the klotz tiling in E_n, the cut-and-project method guarantees that two tiles $X_{\parallel i}$ and $X_{\parallel j}$ do not overlap in $E_\parallel$ if and only if their projections $X_{\perp i}$ and $X_{\perp j}$

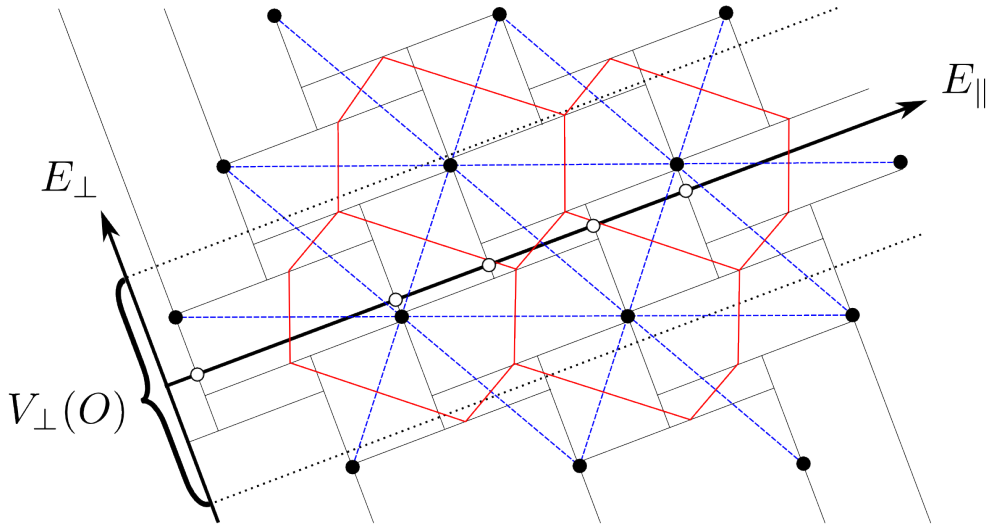


Figure 2.2: Example of the klotz construction in 2D based on [42]. The 1-boundaries of the Voronoi cells are represented as red lines and their 1-dimensional duals as blue lines. The edges of the klotz tiles $X^*_{\parallel} \otimes X_{\perp}$ are represented as black lines.

into $E_{\perp}$ share an interior point. The situation where two tiles may overlap in $E_{\parallel}$ is explained in detail in Section 2.2.2.

Due to translational invariance of $\mathcal{L}$, we can also consider the m-boundaries X of a single Voronoi cell. Recalling that $V(q) = V(O) + q$ for $q \in \mathcal{L}$, it follows that for any $X(q) \in V(q)$ there is a single $X(O) \in V(O)$ such that $X(q) = X(O) + q$. The condition $O_{\perp} \in X_{\perp}(q)$ is then equivalent to $-q_{\perp} \in X_{\perp}(O)$ and the tile $X^*_{\parallel}(q) = X^*_{\parallel}(O) + q_{\parallel}$ is part of $\mathcal{T}^*$. Similarly, a vertex $q_{\parallel}$ belongs to the tiling $\mathcal{T}^*$ if there is at least one $X_{\perp}(O) \in V_{\perp}(O)$ such that $-q_{\perp} \in X_{\perp}(O)$. This is equivalent to $-q_{\perp} \in V_{\perp}(O)$.

The tiling $\mathcal{T}^*$ can then be constructed iteratively by projecting each vertex $q \in \mathcal{L}$ such that $-q_{\perp} \in V_{\perp}(O)$ and completed by projection of all tiles obeying the cut-and-project selection rules and sharing $q_{\parallel}$ as a vertex. These tiles form the so-called *vertex configuration* of $q_{\parallel}$ which is constituted by the set of tiles $X^*_{\parallel}(O) + q_{\parallel}$ such that $-q_{\perp} \in X_{\perp}(O)$. Examples of vertex configurations with 5-fold symmetry and icosahedral symmetry are shown later in Figures 2.5 and

2.7. An example of a 1-dimensional tiling is represented in Figure 2.2.

It is worth noting that the tiling $\mathcal{T}^*$ is composed of a finite number of non-congruent tiles (as opposed to meshes, i.e. structures with unlimited number of different shapes). Indeed $V(O)$ is a finite polytope and the number of boundaries of $V(O)$ and their duals is therefore finite. Moreover, these can be related by point group symmetry or by translation operations in $\mathcal{L}$, and hence the tiling is composed of a finite number of non-congruent tiles.
Following the terminology in [66], we introduce the following definition.

Definition 2.2.1. *A set of tiles S of a tiling $\mathcal{T}^*$ is called a protoset, if S contains no two congruent tiles and if all tiles in $\mathcal{T}^*$ are congruent to a tile in S. Tiles within a protoset are called prototiles.*

Notice that one could choose a different section of the klotz tiling by translating the physical subspace $E_{\parallel}^{'} = E_{\parallel} + \boldsymbol{\gamma}$. This will be useful for the creation of 5-fold symmetric tilings in Section 2.3 where shifting $E_{\parallel}$ allows us to maintain rotational symmetry but not inversion symmetry. If $\boldsymbol{\gamma} \in E_{\parallel}$ then the tilings in $E_{\parallel}$ and $E_{\parallel}^{'}$ are simply related by translation. If $\boldsymbol{\gamma} \in E_{\perp}$ this may result in a change of tiles in $E_{\parallel}^{'}$ since this subspace may *cut* different klotz tiles in E_n as shown in Figure 2.3.

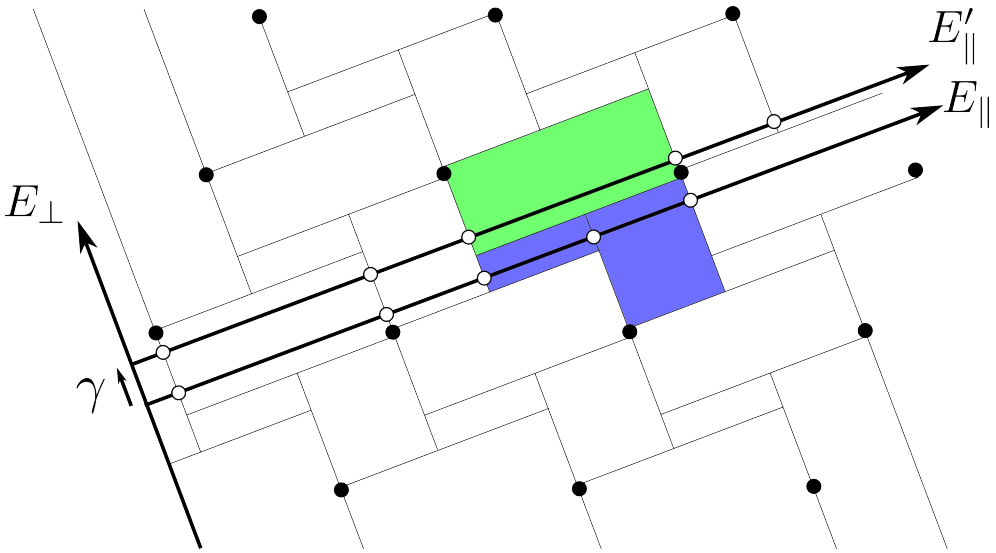


Figure 2.3: $E_{\parallel}$ and the translated subspace $E'_{\parallel} = E_{\parallel} + \boldsymbol{\gamma}$ with $\boldsymbol{\gamma} \in E_{\perp}$ cut different klotz tiles resulting in different projected tilings. Examples of the different intersected tiles are highlighted in pale green and blue.

2.2.2 Undefined vertex configurations

In this section, we show that the current construction does not allow us to determine all vertex configurations, and we propose a corrective approach based on [40]. We have developed a new algorithm to deal with these undefined vertex configurations. A different approach will be discussed in Section 2.3.1. First notice that, if G is non-crystallographic then the subspace $E_{\parallel}$ must be totally irrational, that is there is a single vertex $q \in \mathcal{L}$ such that $E_{\parallel} \cap \mathcal{L} = \{q\}$ (recalling that we have chosen the case were q corresponds to the origin O). If $E_{\parallel}$ were not irrational, then there would be at least two vertices q_i, q_j such that $q_i, q_j \in E_{\parallel} \cap \mathcal{L}$ and the projected tiling would have to be periodic along the vector $(q_i - q_j)_{\parallel}$. This would contradict the fact that G is aperiodic.

In turn, the irrationality of $E_{\parallel}$ implies a one-to-one correspondance between $q \in \mathcal{L}$ and its orthogonal projection in $E_{\perp}$, $q_{\perp}$. It can then be shown that the set of vertices whose projection in $E_{\perp}$ lies in $V_{\perp}(O)$ is dense and uniformly distributed [66]. When constructing the vertex configuration of $q_{\parallel}$

(i.e. the set of tiles sharing $q_\parallel$ as a vertex), if $-q_\perp$ intersects the boundaries of some $X_\perp$ then overlapping between the corresponding $X^*_\parallel$ may occur. Recall that the cut-and-project method implies that two tiles, $X_{\parallel i}$ and $X_{\parallel j}$ do not overlap in $E_\parallel$ if and only if their projections $X_{\perp i}$ and $X_{\perp j}$ in $E_\perp$ share an interior point. Because the projection of vertices into $E_\perp$ is dense, cases for which $-q_\perp$ lie on boundaries of non-intersecting polytopes $X_\perp(O) \in V_\perp(O)$ will occur. These cases can also occur via a shift of the subspace $E_\parallel$ as shown in Figure 2.4.

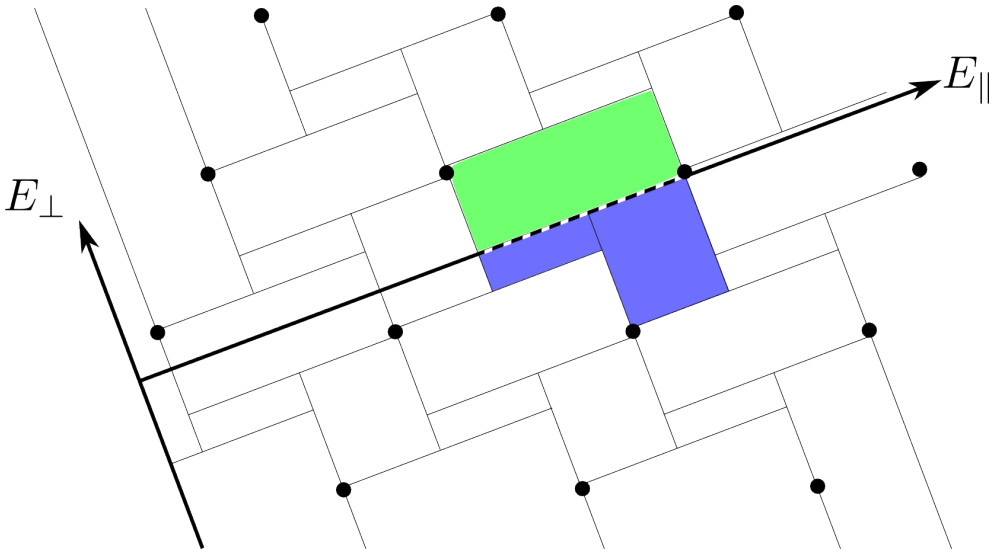


Figure 2.4: Demonstration of the occurence of glue-tiles. Here $E_\parallel$ is chosen such that it cuts E_n at the junction of two klotz tiles, hence leading to overlapping tiles in the resulting tiling. In this case the selection rule does not specify which of the blue or green tiles should project onto $E_\parallel$. The one-dimensional glue-tile, whose construction is explained in the text, is represented as a dashed line.

Let us assume that $-q_\perp$ belongs to the boundaries $\partial X_{\perp j}(O)$ of N polytopes $X_{\perp j}(O)$ $j = 1, ..., N$, that is $-q_\perp \in \bigcap_{j=1}^{N} \partial X_{\perp j}(O)$, and any two polytopes $X_\perp(O)$ do not share an interior point, i.e. $X_{\perp i}(O) \cap X_{\perp j}(O) = \varnothing$ if $i \neq j$. We introduce $\mathcal{D} = \left(\bigcap_{j=1}^{N} \partial X_j(O)\right) + q = \left(\bigcap_{j=1}^{N} \partial X_j(q)\right)$ with $\partial X_j(q) :=$

$\partial X_j(O) + q$ and call $\mathcal{D}^* = \left(\bigcap_{j=1}^{N} \partial X_j(O)\right)^* + q = \left(\bigcap_{j=1}^{N} \partial X_j(q)\right)^*$ its dual. We then choose the convex hull of $\mathcal{D}^*_{\parallel}$ as a new tile. To differentiate these new tiles from *regular* tiles, we use the same terminology as in [41] and call these *glue*-tiles. This corresponds to the cut-and-project construction, which can be summarised as follows: if $E_{\parallel}$ cuts $X(q)$, project its dual $X^*(q)$. According to the properties of duals, $\bigcap_{j=1}^{N} \partial X_j(q) \subset X_j(q)$ implies that $X_j^* \subset \left(\bigcap_{j=1}^{N} \partial X_j(q)\right)^*$. Thus all tiles $X_j^*(q)$ which may overlap are contained within the glue-tile. We verify that this construction technique preserves the desired properties of the projected tiling. Here the symmetry G is, by construction, preserved upon projection and the absence of holes in $\mathcal{T}^*$ is guaranteed by the klotz tile paving of E_n. In addition, the projection onto $E_{\parallel}$ of the (n-m)-boundaries $X^*_{\mathcal{D}}$ of $\mathcal{D}^*$ covers $\mathcal{D}^*_{\parallel}$. The set of polytopes $X^*_{\mathcal{D}}$ of dimension n-m is given by the duals of all m-dimensional polytopes X sharing $\mathcal{D}$ as a boundary. By definition of $\mathcal{D}$, m-dimensional polytopes X (whose projections into $E_{\perp}$ contain $-q_{\perp}$) either share $\mathcal{D}$ as a boundary or overlap in $E_{\perp}$. Thus $\mathcal{D}^*_{\parallel}$ does not overlap with other tiles in $E_{\parallel}$. Finally, glue-tiles contain all tiles susceptible to overlap and no two glue-tiles can overlap. An example of a one-dimensional glue-tile is shown in Figure 2.4.

In the following sections, we specify the spaces E_n and the subspaces $E_{\parallel}$ and $E_{\perp}$ for the construction of 5-fold symmetrical tilings (Section 2.3) and for icosahedral tilings (Section 2.4). In both cases, we choose the subspaces $E_{\parallel}$ and $E_{\perp}$ such that $E_{\parallel} \cap E_{\perp} = \{O\}$.

2.3 Quasilattices with 5-fold symmetry

In this section we apply the cut-and-project method to the construction of tilings with 5-fold symmetry. Although these tilings have not been applied directly to the study of structural constraints on proteins shells, they have been used to classify the possible conformational changes that appear in 5-fold symmetric tilings when transitions with minimal symmetry loss occur in

their associated higher dimensional lattice as explained in Chapter 3.
From [44], SC, BCC and FCC lattices are known for any dimension n and are given by:

$$\begin{aligned}
\mathcal{L}_{SC} &= \{\boldsymbol{x} = (x_1, \ldots, x_n) : \ x_i \in \mathbb{Z}, i = 1, \ldots, n\}, \\
\mathcal{L}_{BCC} &= \{\boldsymbol{x} = \tfrac{1}{2}(x_1, \ldots, x_n) : \ x_i \in \mathbb{Z}, \ x_i = x_j \ (\text{mod } 2), i, j = 1, \ldots, n\}, \\
\mathcal{L}_{FCC} &= \left\{ \boldsymbol{x} = \tfrac{1}{2}(x_1, \ldots, x_n) : x_i \in \mathbb{Z}, \ \sum_{j=1}^{n} x_j = 0 \ (\text{mod } 2) \right\}.
\end{aligned} \tag{2.4}$$

Note that since the three different lattices (SC, BCC and FCC) have Voronoi cells of different shapes, the tilings produced in $E_{\|}$ will be different.

To generate the group C_5, we use the generator G_5:

$$G_5 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{2.5}$$

This generator can be block-diagonalised using the following matrix M:

$$M = \begin{pmatrix} 1 & \cos(2\pi/5) & \cos(4\pi/5) & \cos(6\pi/5) & \cos(8\pi/5) \\ 0 & \sin(2\pi/5) & \sin(4\pi/5) & \sin(6\pi/5) & \sin(8\pi/5) \\ 1 & \cos(4\pi/5) & \cos(8\pi/5) & \cos(2\pi/5) & \cos(6\pi/5) \\ 0 & \sin(4\pi/5) & \sin(8\pi/5) & \sin(2\pi/5) & \sin(6\pi/5) \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{2.6}$$

giving the block-diagonalised form of G_5:

$$MG_5M^{-1} = \begin{pmatrix} \cos(2\pi/5) & -\sin(2\pi/5) & 0 & 0 & 0 \\ \sin(2\pi/5) & \cos(2\pi/5) & 0 & 0 & 0 \\ 0 & 0 & \cos(4\pi/5) & -\sin(4\pi/5) & 0 \\ 0 & 0 & \cos(4\pi/5) & \sin(4\pi/5) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Without loss of generality, we choose the first two (respectively last three) rows of M to be the projection matrix corresponding to $E_{\parallel}$ (respectively $E_{\perp}$) and call $\pi_{\parallel}$ (respectively $\pi_{\perp}$) the associated projection matrices.
Note that the vector $n = (1, 1, 1, 1, 1)$ is invariant under 5-fold rotation according to 2.5 and belongs to either of the 5-dimensional SC, BCC or FCC lattice. This additional constraint allows to reduce to 4 the minimal dimension of the embedding space necessary to generate a lattice with this symmetry. However, we consider in the present study 5D lattices to simplify the descriptions of the lattice transitions in Chapter 3.

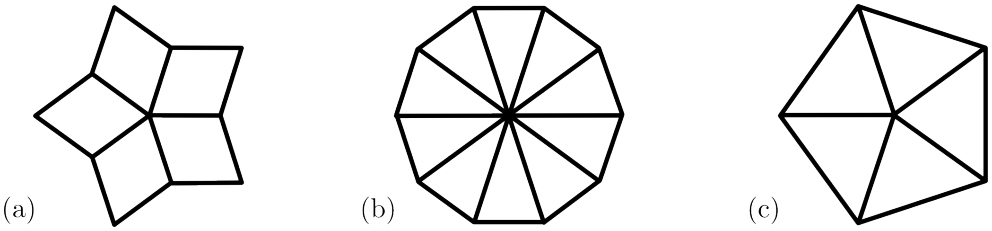


Figure 2.5: Vertex configurations of the origin for the 5-fold symmetrical tilings SC (a), the BCC (b) and FCC (c) tilings. A way of achieving rotational symmetry, but not inversion symmetry, is by a shifting $E_{\parallel}$ by $\gamma = \frac{1}{2}(1, 1, 1, 1, 1)$.

2.3.1 A 5-fold tiling from the A_4 lattice

Here we present an example, based on [3], where a different approach to dealing with the undefined boundaries of $X_{\perp}(O) \in V_{\perp}(O)$ is implemented. In this case, 5-fold symmetry has been constructed from the 4-dimensional FCC lattice (also known as A_4 lattice) and both subspaces $E_{\parallel}$ and $E_{\perp}$ are 2D, making the visualisation of $V_{\perp}(O)$ and its decomposition in terms of $X_{\perp}(O) \in V_{\perp}(O)$ easier. In [3], the A_4 lattice, embedded into the 5-dimensional Euclidean space

E_5, is generated using the following basis:

$$B_{A_4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Notice that all vertices generated by this basis are in a hyperplane orthogonal the to vector $n = (1, 1, 1, 1, 1)$. The block diagonalised form is obtained using the same matrix M as in 2.6 (up to a factor) and the projection of A_4 lattice points into $E_\parallel$ is done using the same projection matrix (up to a factor) $\pi_\parallel$ (given as the first two lines of 2.6). However projection into $E_\perp$ is done using the third and fourth lines of 2.6 (again up to a factor). A careful analysis of the projection of $V(O)$ onto this subspace $E_\perp$ (see for instance Figure 2.6) and its 2D boundaries reveals that the $X_\perp(O) \in V_\perp(O)$ are congruent to one of two polytopes. Each tile in the tiling $\mathcal{T}^*$ is therefore congruent to one of two prototiles.

To deal with undefined cases, i.e. in the situation when $-q_\perp$ lies on boundaries of non-intersecting polytopes $X_\perp(O)$, boundary conditions on $X_\perp(O)$ have been defined [3]. In the case discussed here, boundaries between two non-overlaping polytopes $X_\perp(O)$ are 1D edges in $E_\perp$. To define in which of the two polytopes $-q_\perp$ belongs, one side of this edge is hatched, see Figure 2.6. Should $-q_\perp$ belong to this boundary, it will be considered within the window whose side is hatched. Note that after choosing one side of an edge to hatch, translational invariance in A_4 requires that parallel boundaries must be hatched on the same side (as represented in Figure 2.6) and all constraints on the others edges are obtained by subsequent application of 5-fold symmetry. Although an arbitrary choice has been made (choosing which side of the boundary is hatched), this technique permits the creation of a complete tiling with 5-fold symmetry, no-overlaping tiles and without creating new tiles, i.e. glue-tiles.

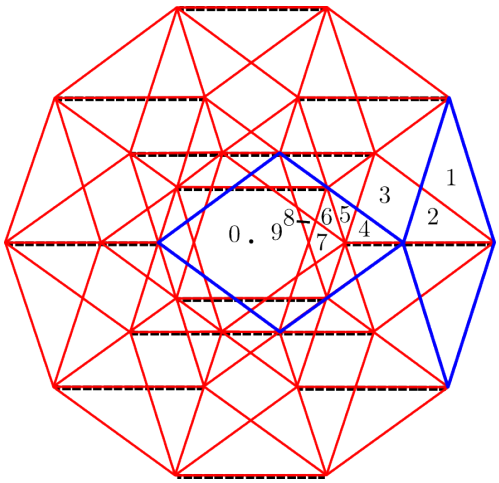


Figure 2.6: Illustration of $V_\perp(O)$ and its partition into its 2-boundaries $X_\perp(O)$ adapted from [3]. We highlight in blue the two polytopes to either of which all $X_\perp(O) \in V_\perp(O)$ are congruent. For clarity only one set of parallel edges is hatched to define boundary conditions, the constraints on the other edges are obtained through 5-fold symmetry. The numbers 1 to 9 count the different ways the $X_\perp(O)$ may overlap, each corresponding to different vertex configurations in $E_\parallel$.

However this technique is not applicable for symmetry allowing 2-fold rotations, as is the case for icosahedral tilings. Indeed, conservation of translational symmetry in the higher dimensional lattice and of the 2-fold rotation requires that both sides of the boundaries be hatched. We thus use instead the creation of glue-tiles as explained in Section 2.2.2.

2.4 Icosahedral tilings

It has been shown [44] that the lowest dimension for which lattices have icosahedral symmetry $\mathcal{I}$ as a subgroup of their holohedry is 6 and that these lattices can only be SC, BCC or FCC. These are given by (2.4). The 6D matrix representation of $\mathcal{I}$ is generated using the two-fold G_2 and the three-fold G_3 rotation matrices:

$$G_2 = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad G_3 = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{2.7}$$

Following [33], we choose the projection matrix M as follow:

$$M = \begin{pmatrix} \tau & 0 & -1 & 0 & \tau & 1 \\ 1 & \tau & 0 & -\tau & -1 & 0 \\ 0 & 1 & \tau & 1 & 0 & \tau \\ \tau & -1 & \tau & 0 & 0 & -1 \\ 1 & 0 & -1 & \tau & -\tau & 0 \\ 0 & \tau & 0 & 1 & 1 & -\tau \end{pmatrix}$$

with $\tau = \frac{1}{2}(1 + \sqrt{5})$.

M block-diagonalises G_2 and G_3, giving:

$$MG_2M^{-1} = \frac{1}{2} \begin{pmatrix} -\tau & \tau' & 1 & 0 & 0 & 0 \\ \tau' & -1 & -\tau & 0 & 0 & 0 \\ 1 & -\tau & -\tau' & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -\tau & -\tau' \\ 0 & 0 & 0 & -\tau & -\tau' & -1 \\ 0 & 0 & 0 & -\tau' & -1 & -\tau \end{pmatrix}$$

and

$$MG_3M^{-1} = \frac{1}{2} \begin{pmatrix} 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tau & -\tau' & -1 \\ 0 & 0 & 0 & \tau' & -1 & -\tau \\ 0 & 0 & 0 & -1 & \tau & \tau' \end{pmatrix}$$

with $\tau' = 1 - \tau$.

The two subspaces $E_\parallel$ and $E_\perp$ that are both invariant under $\mathcal{I}$ are therefore 3-dimensional subspaces. We choose the first three (respectively last three) rows of M to be the projection matrix corresponding to $E_\parallel$ (respectively $E_\perp$) and call $\pi_\parallel$ (respectively $\pi_\perp$) the associated projection matrices.

We take the opportunity to introduce here the dilatation matrix $\mathcal{D}$ which will be needed to embed point arrays into the SC, BCC or FCC 6D lattices in Chapter 3:

$$\mathcal{D} = \frac{1}{2}\begin{pmatrix} 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{2.8}$$

The FCC and BCC lattices are invariant under $\mathcal{D}$ (i.e. $\forall v \in \mathcal{L}$, $Dv \in \mathcal{L}$ for $\mathcal{L}$ either the FCC or BCC lattice) while the SC lattice $\mathcal{L}_{SC}$ is invariant under the $\mathcal{D}^3$ matrix (i.e. $\forall v \in \mathcal{L}_{SC}$, $D^3 v \in \mathcal{L}_{SC}$) [20].

$\mathcal{D}$ has the following property related to the projections $\pi_\parallel$ and $\pi_\perp$ into the parallel and orthogonal subspaces:

$$\pi_\parallel \mathcal{D} v = \tau \pi_\parallel v \text{ and } \pi_\perp \mathcal{D} v = -\frac{1}{\tau}\pi_\perp v. \tag{2.9}$$

In particular, $\mathcal{D}$ acts as a dilation in $E_\parallel$ and as a contraction in $E_\perp$.

Using the cut-and-project method we constructed finite sized tilings associated with the SC, the BCC and the FCC lattice. As an example, the vertex configurations of the origin O corresponding to the three different tilings are shown in Figure 2.7. The protoset of the SC tiling is composed of two regular prototiles and two glue-tiles. In accordance with [56], the BCC tiling is composed of 8 prototiles (four tetrahedra and four pyramids) and no glue-tiles were required. The FCC tiling is composed of 6 tetrahedral prototiles (also

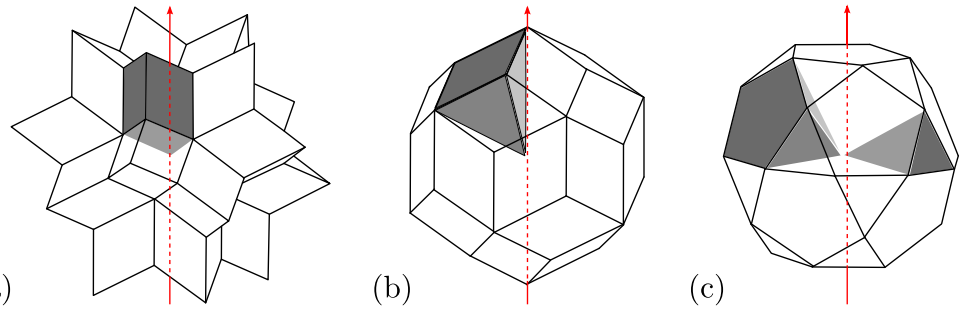


Figure 2.7: Vertex configurations of the origin for the three icosahedral tiling types: (a) SC, (b) BCC and (c) FCC. For each case the different tiles are represented in shades of grey.

found in [55]) and the addition of 10 glue-tiles was necessary to complete the tiling.

2.5 Remarks on Tiling Construction

Using the cut-and-project method we have constructed 2-dimensional 5-fold symmetric and 3-dimensional icosahedrally symmetric tilings corresponding to SC, BCC and FCC lattices in 6D and have shown that these contain neither holes nor overlapping tiles. Because it is computationally impossible to generate infinite size tilings we grow a tiling around the origin O until its size is "large enough" to model virus architecture. We will explain in Chapter 4 what large enough means in the particular case of viral capsids embedded into icosahedral tilings. To compute bigger icosahedral tiling patches, we have retained only tiles inside the fundamental domain (shown in Figure 2.8). This reduces by a factor of approximately 60 the number of tiles needed to be computed using the cut-and-project method. When needed the entire tiling can then be obtained by application of the icosahedral symmetry. Note that the smaller fundamental domain Coxeter group H_3 [25], i.e. the group composed of all reflections of the icosahedron, could have been used instead. However experimentally determined structures of viral capsids (on which is based the

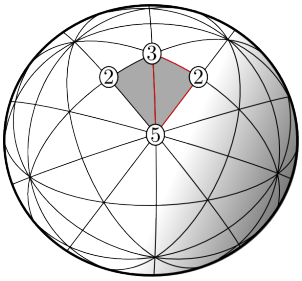


Figure 2.8: The half kite (indicated by the red boundaries) is the fundamental domain of the Coxeter group H_3 and the full kite (shaded in grey) is the fundamental domain of $\mathcal{I}$ which corresponds to the rotational symmetry only. Only tiles overlaping with the fundamental domain of $\mathcal{I}$ are retained in the analysis as subsequent application of $\mathcal{I}$ allows us to obtain all other tiles via symmetry.

matching algorithm discussed in Chapter 4) do not generally have this additional symmetry.

It is important to note that not all icosahedral lattices have been constructed in this study. Other examples of tilings would include the dual tiling $\mathcal{T}$ [66], where $X_{\parallel}(q)$ is included into the tiling in $E_{\parallel}$ if and only if $E_{\parallel}$ *cuts* $X^*_{\perp}(q)$. Here, the choice of $\mathcal{T}^*$ tilings over $\mathcal{T}$ tilings is motivated by the fact that all point arrays used to fit viral capsids in [37], described in the introductory Chapter 1, can be embedded into the constructed tilings $\mathcal{T}^*$, see Chapter 3 for proof.

2.5.1 From Point Arrays to Tilings

Now that the mathematical background necessary for the construction of tilings has been presented, we give here an insight into how point arrays are embbeded into tilings. Figure 2.9 is a 2D graphical example showing the construction of a 10-fold symmetric point array. The superimposed tiling was created using the five-dimensional SC lattice and the projection matrix explained in Section 2.3. In order for the tiling to have inversion symmetry, and hence 10-fold rather than 5-fold symmetry, note that no shift of the physical

space $E_{\parallel}$ in the lattice has been applied in contrast to Figure 2.5(a).

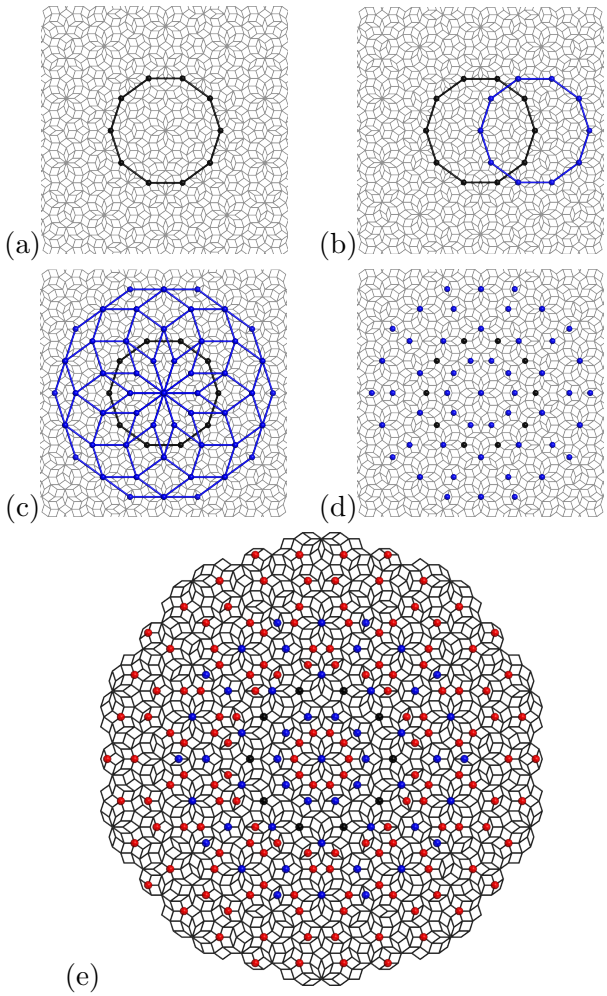


Figure 2.9: Example of the embedding of a 10-fold invariant point array into a tiling adapted from [37]. At each step of the construction, the embedded point array is shown superimposed on the tiling. The construction starts with the decagon shown in black in (a). Application of an *admissible* translation, i.e. a non-compact operator of the affine extension group of C_{10}, is applied giving the blue decagon (b). The nine other copies of the translated decagon are obtained by subsequent application of 10-fold rotation as shown in (c). Vertices of the point array correspond to all words of the affine extended group which contain at least one copy of the translation operator (black and blue vertices)(d). How the point array "grows" with multiple use of the non-compact operator is shown in (e), where red points correspond to the words of the affine extended group containing exactly two non-compact operators.

Chapter 3

Structural transitions in quasicrystals

3.1 Introduction

During a viral life-cycle, the viral capsid may undergo conformational changes necessary for the particle to become infective (such as maturation events [16]). Understanding the mechanisms governing these structural transformations is therefore important for the design of new anti-viral strategies. Experimental techniques allow observation of pre- and post-transitional states [16, 64, 67], and theoretical models are developed to determine how these structural transitions occur. For example, normal mode analysis can be used to determine how protein subunits rearrange during the transition (see for instance [74, 75]). However, this technique requires intensive computation due to the many degrees of freedom to be considered.

In this chapter we are presenting the work developed in [26] and [27] in which the point array descriptors from [36] and concepts from crystallography are combined to infer information on possible transition paths, characterised by minimal symmetry loss, between pre- and post-transitional states. We first present in Section 3.2, the mathematical tools derived from reconstructive structural phase transitions of crystalline solid studies [1, 4, 11, 34, 60, 61, 69,

77]. Transitions between the 3D quasilattices can be infered from the possible transitions between higher-dimensional lattices using the cut-and-project method. The classification of the possible changes undergone by the quasilattices during a transition is presented in Section 3.3, and applications to the study of viral transitions, using the Cowpea Chlorotic Mottle virus as an example, are explained in Section 3.4.

3.2 Mathematical Background

Let $GL(n,\mathbb{Z})$ be the group of $n\times n$ unimodular integer matrices, $GL(n,\mathbb{R})$ the group of $n\times n$ invertible real matrices, $O(n)$ and $SO(n)$ the orthogonal and special orthogonal group of $\mathbb{R}^n$, and $Sym^+(n,\mathbb{R})$ the set of $n\times n$ symmetric positive definite matrices with real coefficients. For a basis $\{\boldsymbol{b}_\alpha\}_{\alpha=1,\dots,n}$ of $\mathbb{R}^n$, we write $B\in GL(n,\mathbb{R})$ for the matrix with column vectors $\boldsymbol{b}_\alpha$. We denote by

$$\mathcal{L}(B)=\{\boldsymbol{x}=\sum_{\alpha=1}^{n} m^\alpha \boldsymbol{b}_\alpha \,:\, m^\alpha\in\mathbb{Z}\}$$

the lattice with basis $\{\boldsymbol{b}_\alpha\}$. All other lattice bases have the form $\{\sum_{\beta=1}^n M_\alpha^\beta \boldsymbol{b}_\beta\}$, with $M_\alpha^\beta\in GL(n,\mathbb{Z})$. Moreover we write

$$\Lambda(B)=\left\{ M_\alpha^\beta\in GL(n,\mathbb{Z}) \,:\, \exists Q\in O(n) \text{ such that } Q\boldsymbol{b}_\alpha=\sum_{\beta=1}^{n} M_\alpha^\beta\boldsymbol{b}_\beta \right\}, \tag{3.1}$$

for the *lattice group* of $\mathcal{L}(B)$, and

$$\mathcal{P}(B)=\left\{ Q\in O(n) \,:\, \exists M_\alpha^\beta\in GL(n,\mathbb{Z}) \text{ such that } Q\boldsymbol{b}_\alpha=\sum_{\beta=1}^{n} M_\alpha^\beta\boldsymbol{b}_\beta \right\}, \tag{3.2}$$

for its *point group*. The following notations are equivalent:

$$Q\boldsymbol{b}_\alpha=\sum_{\beta=1}^{n} M_\alpha^\beta\boldsymbol{b}_\beta \qquad\Leftrightarrow\qquad QB=BM.$$

The point and lattice groups are related via the identity

$$\Lambda(B)=B^{-1}\mathcal{P}(B)B, \tag{3.3}$$

and moreover

$$\mathcal{P}(RB) = R\mathcal{P}(B)R^\top, \qquad \mathcal{P}(BM) = \mathcal{P}(B),$$
$$\Lambda(RB) = \Lambda(B), \qquad \Lambda(BM) = M^{-1}\Lambda(B)M,$$

for $R \in O(n)$ and $M \in GL(n,\mathbb{Z})$. We therefore will also write $\mathcal{P}(\mathcal{L})$ for the point group of the lattice $\mathcal{L}$.

A lattice basis is characterised (modulo rotations) by its *lattice metric*

$$C = B^\top B \in Sym^+(n,\mathbb{R}),$$

and the lattice group is the subgroup of $GL(n,\mathbb{Z})$ that fixes the metric [61]:

$$M \in GL(n,\mathbb{Z}), \quad M^\top C M = C \qquad \Leftrightarrow \qquad M \in \Lambda(B). \tag{3.4}$$

For a matrix group $\mathcal{G} \subset GL(n,\mathbb{Z})$ we have the following standard definition:

Definition 3.2.1. *The centraliser $\mathcal{Z}(\mathcal{G},\mathbb{R})$ of $\mathcal{G}$ in $GL(n,\mathbb{R})$ is the group*

$$\mathcal{Z}(\mathcal{G},\mathbb{R}) = \left\{N \in GL(n,\mathbb{R}) \,:\, N^{-1}GN = G, \quad \forall G \in \mathcal{G}\right\}.$$

The centraliser of $\mathcal{G}$ can be obtained by solving the linear equations $G_i C = C G_i$ in the unknown C, where G_i are generators of $\mathcal{G}$. Hence, the centralisers of a finitely generated group in general depend linearly on a finite list of real parameters.

A lattice transition is defined as a continuous transformation between two lattices $\mathcal{L}_0$ and $\mathcal{L}_1$ along which some symmetry is preserved, described by a common subgroup $\tilde{\mathcal{G}} \subset GL(n,\mathbb{Z})$ of the lattice groups of the intermediate lattices.

Definition 3.2.2. *Let $\mathcal{L}_0$ and $\mathcal{L}_1$ be two lattices, and $\mathcal{G} \subset \mathcal{P}(\mathcal{L}_0)$. We say that there exists a transition between $\mathcal{L}_0$ and $\mathcal{L}_1$ with intermediate symmetry $\mathcal{G}$ if there exist bases B_0 and B_1 of $\mathcal{L}_0$ and $\mathcal{L}_1$, and a continuous path $B : [0,1] \to GL(n,\mathbb{R})$, with $B(0) = B_0$ and $B(1) = B_1$, such that, for*

$$\tilde{\mathcal{G}} = B_0^{-1}\mathcal{G}B_0 \subset \Lambda(B_0),$$

one has

$$\tilde{\mathcal{G}} \subset \Lambda(B(t)), \qquad t \in [0,1]. \tag{3.5}$$

We call the linear mapping

$$T := B_1 B_0^{-1} : \mathcal{L}_0 \to \mathcal{L}_1, \tag{3.6}$$

the transition*, while the curve* $T(t) = B(t)B_0^{-1}$ *is the* transition path*.*

Notice that, by continuity, $\det B_0$ and $\det B_1$ have the same sign, so that $\det T > 0$.

The following equivalent statements characterise lattice transitions.

Proposition 3.2.3. *Let* $\mathcal{L}_0$ *and* $\mathcal{L}_1$ *be two lattices, and* $\mathcal{G} \subset \mathcal{P}(\mathcal{L}_0)$*. The following statements are equivalent:*

i) There exists a transition between $\mathcal{L}_0$ *and* $\mathcal{L}_1$ *with intermediate symmetry* $\mathcal{G}$*.*
ii) There exist bases B_0 *and* B_1 *of* $\mathcal{L}_0$ *and* $\mathcal{L}_1$*, such that for* $\tilde{\mathcal{G}} = B_0^{-1}\mathcal{G}B_0$

$$\tilde{\mathcal{G}} \subset \Lambda(B_0) \cap \Lambda(B_1). \tag{3.7}$$

iii) There exist bases B_0 *and* B_1 *of* $\mathcal{L}_0$ *and* $\mathcal{L}_1$ *such that*

$$B_1 = RUB_0, \quad \textit{with} \quad R \in SO(n) \quad \textit{and} \quad U \in \mathcal{Z}(\mathcal{G}, \mathbb{R}) \cap Sym^+(n, \mathbb{R}). \tag{3.8}$$

iv) There exists a basis B_0 *of* $\mathcal{L}_0$*, and continuous paths*

$$R : [0,1] \to SO(n), \quad \textit{and} \quad U : [0,1] \to \mathcal{Z}(\mathcal{G}, \mathbb{R}) \cap Sym^+(n, \mathbb{R}),$$

such that $R(0) = U(0) = I$ *and* $R(1)U(1)B_0 = B_1$ *is a basis of* $\mathcal{L}_1$*.*
v) There exist bases B_0 *and* B_1 *of* $\mathcal{L}_0$ *and* $\mathcal{L}_1$ *and a continuous path* $C : [0,1] \to Sym^+(n, \mathbb{R})$ *such that, letting* $C_0 = B_0^\top B_0$*,* $C_1 = B_1^\top B_1$ *and* $\tilde{\mathcal{G}} = B_0^{-1}\mathcal{G}B_0$*, then* $C(0) = C_0$*,* $C(1) = C_1$*, and*

$$M^\top C(t) M = C(t) \qquad \textit{for all} \quad M \in \tilde{\mathcal{G}} \quad \textit{and} \quad t \in [0,1]. \tag{3.9}$$

Proof. The implications i) $\Rightarrow$ ii) and i) $\Leftrightarrow$ v) are immediate. To prove that ii) $\Rightarrow$ iii), we notice that, by ii), letting $\mathcal{G}_1 = B_1\tilde{\mathcal{G}}B_1^{-1}$, then $B_1^{-1}\mathcal{G}_1 B_1 =$

$B_0^{-1}\mathcal{G}B_0$. Hence $\mathcal{G}$ and $\mathcal{G}_1$ are conjugate in $GL(n,\mathbb{R})$: $\mathcal{G} = T^{-1}\mathcal{G}_1 T$, with $T = B_1 B_0^{-1}$. By the polar decomposition theorem, writing $T = RU$ with $R \in SO(n)$ and $U \in Sym^+(n,\mathbb{R})$, it follows that, for all $G \in \mathcal{G}$ there exists $H \in \mathcal{G}_1$ such that $G = U^{-1}R^{-1}HRU$, i.e., $UG = R^{-1}HRU$. By the uniqueness of the polar decomposition, since G and $R^{-1}HR \in O(n)$, one has

$$G = R^{-1}HR \qquad \text{and} \qquad G = U^{-1}GU,$$

Hence U belongs to the centraliser of $\mathcal{G}$ in $GL(n,\mathbb{R})$. This proves iii). Notice that, as a by-product of the above argument, $\mathcal{G}$ and $\mathcal{G}_1$ are conjugate in $SO(n)$, indeed

$$\mathcal{G} = R^{-1}\mathcal{G}_1 R.$$

To prove that iii) $\Rightarrow$ iv), notice first that $SO(n)$ is arcwise connected, so that there exist paths connecting R to the identity (one could choose $R(t) = exp(Wt)$ where W is not unique).

Further, since U is positive definite, we can write $U = Q^\top DQ$, with $Q \in SO(n)$, $D = \mathrm{diag}(\lambda_1,\ldots,\lambda_n)$, and $\lambda_i > 0$. Then a path connecting U to the identity is for instance $U(t) = Q^\top D(t)Q$, with $D(t) = \mathrm{diag}((\lambda_1 - 1)t + 1,\ldots,(\lambda_n - 1)t + 1)$, which is still in the centraliser. In fact, if $U \in \mathcal{Z}(\mathcal{G},\mathbb{R})$, then $D \in \mathcal{Z}(Q\mathcal{G}Q^\top,\mathbb{R})$, which means that $DH = HD$ for every $H \in Q\mathcal{G}Q^\top$. This is equivalent to $\lambda_i h_{ij} = \lambda_j h_{ij}$ for all $i,j = 1,\ldots,n$, where h_{ij} are the entries of H. On the other hand, this identity is true if and only if $[(\lambda_i - 1)t + 1)]h_{ij} = [(\lambda_j - 1)t + 1]h_{ij}$ for all $t \in [0,1]$, which implies that $D(t) \in \mathcal{Z}(Q\mathcal{G}Q^\top,\mathbb{R})$, so that $U(t) \in \mathcal{Z}(\mathcal{G},\mathbb{R})$ for all $t \in [0,1]$.

Finally, iv) $\Rightarrow$ i): in fact, letting $B(t) = R(t)U(t)B_0$, one has

$$\begin{aligned} B(t)\tilde{\mathcal{G}}B(t)^{-1} =& R(t)U(t)B_0\tilde{\mathcal{G}}B_0^{-1}U^{-1}(t)R^{-1}(t) \\ =& R(t)U(t)\mathcal{G}U^{-1}(t)R^{-1}(t) = R(t)\mathcal{G}R^{-1}(t) \subset O(n), \quad t \in [0,1]. \end{aligned}$$

Hence, $\tilde{\mathcal{G}} \subset \Lambda(B(t))$ for every $t \in [0,1]$ which proves the assertion. □

If $T(t) = R(t)U(t)$ is a transition path with symmetry $\mathcal{G}$ between the lattices $\mathcal{L}_0$ and $\mathcal{L}_1$, the symmetry of the intermediate phase is also described

by the group of orthogonal transformations

$$\mathcal{G}_t = T(t)\mathcal{G}T^{-1}(t) = R(t)\mathcal{G}R^{-1}(t), \tag{3.10}$$

which is a subgroup of the point group $\mathcal{P}(B(t))$ of the intermediate phase. The following result is a consequence of the above characterisation of lattice transitions, and shows that any centraliser of $\mathcal{G}$, not necessarily symmetric, defines a transition with that symmetry.

Proposition 3.2.4. *Any continuous path*

$$T : [0,1] \to \mathcal{Z}(\mathcal{G}, \mathbb{R}), \qquad T(0) = I, T(1) = B_1 B_0^{-1},$$

with B_0 and B_1 lattice bases for $\mathcal{L}_0$ and $\mathcal{L}_1$, defines a transition between $\mathcal{L}_0$ and $\mathcal{L}_1$ with intermediate symmetry $\mathcal{G}$.

Proof. From 3.1, $\tilde{\mathcal{G}} \subset \Lambda(B(t))$ if $\mathcal{B}(t)\tilde{\mathcal{G}}\mathcal{B}(t)^{-1} \subset O(n)$.
However $\mathcal{B}(t)\tilde{\mathcal{G}}\mathcal{B}(t)^{-1} = \mathcal{B}(t)\mathcal{B}_0^{-1}\mathcal{G}\mathcal{B}_0\mathcal{B}(t)^{-1} = T(t)\mathcal{G}T(t)^{-1} = \mathcal{G} \subset O(n)$ proving the claim. □

3.2.1 Structural transformations of cut-and-project quasicrystals

We now show how structural transformations of cut-and-project quasicrystals can be induced by transitions between higher-dimensional lattices. Consider two n-dimensional lattices $\mathcal{L}_0$ and $\mathcal{L}_1$, with point groups $\mathcal{P}(\mathcal{L}_0)$ and $\mathcal{P}(\mathcal{L}_1)$, and two subgroups $\mathcal{H}_0 \subset \mathcal{P}(\mathcal{L}_0)$ and $\mathcal{H}_1 \subset \mathcal{P}(\mathcal{L}_1)$. Assume that $\mathcal{H}_0$ and $\mathcal{H}_1$ have invariant subspaces E_0 and E_1, respectively, with $\dim E_0 = \dim E_1 = k$. Without loss of generality we can rotate the lattice $\mathcal{L}_0$ such that E_0 maps to E_1, i.e.

$$E_0 = E_1 =: E_\parallel,$$

and consider the cut-and-project quasicrystals $(\mathcal{L}_0, E_\parallel)$ and $(\mathcal{L}_1, E_\parallel)$.

Definition 3.2.5. *We say that there exists a transition between the cut-and-project quasicrystals $(\mathcal{L}_0, E_\parallel)$ and $(\mathcal{L}_1, E_\parallel)$ with intermediate symmetry $\mathcal{G} \subset$*

$\mathcal{P}(\mathcal{L}_0)$ if there exists a transition with intermediate symmetry $\mathcal{G}$ between $\mathcal{L}_0$ and $\mathcal{L}_1$ according to Definition 3.2.2, and if

$$\mathcal{P}(B(t)) \supseteq \mathcal{G}_t \equiv \mathcal{G}, \quad t \in [0,1]. \tag{3.11}$$

Any such transition defines a family of cut-and-project quasicrystals $(\mathcal{L}_t, E_\parallel)$ in the same projection space $E_\parallel$, all of which have symmetry $\mathcal{G}$.

From this definition, the following proposition can be derived.

Proposition 3.2.6. *For a transition between the cut-and-project quasicrystals, the following statements are equivalent:*

i) There exists a transition between $\mathcal{L}_0$ and $\mathcal{L}_1$ with intermediate symmetry $\mathcal{G}$.
ii) There exist a basis B_0 of $\mathcal{L}_0$, and continuous paths

$$T : [0,1] \to \mathcal{Z}(\mathcal{G}, \mathbb{R}), \tag{3.12}$$

such that $T(0) = I$ and where $T(1)B_0 = B_1$ is a basis of $\mathcal{L}_1$.

Proof. Demonstration that ii) implies i) has been shown in Proposition 3.2.4 and we only need to prove that i) implies ii).
First recall that from 3.5 and if 3.11 holds, we then have

$$\mathcal{G}_t = \mathcal{G} = B(t)\tilde{\mathcal{G}}B^{-1}(t), \qquad t \in [0,1],$$

yielding

$$\mathcal{G} = B(t)B_0^{-1}\mathcal{G}B_0B^{-1}(t), \qquad t \in [0,1]\,.$$

Hence, $T(t) = B(t)B_0^{-1}$ belongs to the normaliser of $\mathcal{G}$ in $GL(n,\mathbb{R})$. Recalling that the centraliser is a normal subgroup of the normaliser, we can then write $T(t) = N(t)C(t)$ for $t \in [0,1]$, with N a representative of the left cosets of $\mathcal{Z}(\mathcal{G},\mathbb{R})$ in $\mathcal{N}(\mathcal{G},\mathbb{R})$ and $C(t) \in \mathcal{Z}(\mathcal{G},\mathbb{R})$.
Hence for any $G \in \mathcal{G}$ there exist $H \in \mathcal{G}$ such that

$$T(t)GT(t)^{-1} = H \Rightarrow N(t)GN^{-1}(t) = H, \qquad t \in [0,1]$$

In particular, for $t = 0$, $T(0) = I$ (the identity matrix) leaving $G = H$. By construction $T(t)$ is continuous, the connected component of the identity in

the normaliser of $\mathcal{G}$ in $GL(n,\mathbb{R})$ is then contained in the centraliser, which proves the claim. □

It is useful to discuss the relation of our method with other higher dimensional approaches used in the literature to study experimentally observed transitions between icosahedral quasicrystals and cubic phases in 3D [2, 15, 39, 45, 46, 47, 53, 73, 78], and for deformations of Penrose tilings of the plane (see for instance, [28, 68, 84]). The procedure used to obtain a transformation of the projected 3D point set in these works is either (i) by rotating the projection plane with respect to the 6D lattice or (ii) by deforming the 6D lattice through certain linear transformations called phason strains. These two approaches can be shown to be related [15].

The technique presented here generalises (i) and (ii), since it allows for both deformations and changes of symmetry of the lattice, controlling the intermediate symmetry. In detail our approach relates to the rotation-plane method [39] as follows. For a fixed 6D cubic lattice $\mathcal{L}$, it is known that suitable choices of 3D subspaces E_0 and E_1 in $\mathbb{R}^6$ yield sets $(\mathcal{L}, E_0)$ and $(\mathcal{L}, E_1)$ that correspond to a FCC crystal or to an icosahedral quasicrystal in $\mathbb{R}^3$. Suppose that $\mathcal{H}_0$ and $\mathcal{H}_1$ are 6D representations of the 3D cubic group and of the icosahedral group, respectively, so that E_0 is invariant under $\mathcal{H}_0$ and E_1 is invariant under $\mathcal{H}_1$. Then there exist orthonormal bases $\tilde{B}_0$ and $\tilde{B}_1$ of $\mathbb{R}^6$ (not necessarily lattice bases) that block-diagonalise $\mathcal{H}_0$ and $\mathcal{H}_1$ as the direct sum of two 3D blocks. These bases are adapted to the subspaces E_0 and E_1.

Consider now the tetrahedral group $\mathcal{T}$, which is a subgroup of both $\mathcal{H}_0$ and $\mathcal{H}_1$: both $\tilde{B}_0$ and $\tilde{B}_1$ block-diagonalise $\mathcal{T}$, but since $\mathcal{T}$ has just a single 3D irreducible representation, we may choose $\tilde{B}_0$ and $\tilde{B}_1$ such that $\tilde{B}_1^{-1}\mathcal{T}\tilde{B}_1 = \tilde{B}_0^{-1}\mathcal{T}\tilde{B}_0 = \mathcal{T}'$. Hence, there exists $C \in SO(6)$ in the centraliser of $\mathcal{T}'$ such that $\tilde{B}_1 = \tilde{B}_0 C$. Actually, there exists a continuous path $C(t)$ in the centraliser of $\mathcal{T}'$ that connects C to the identity (referred to as the "Schur rotation" in [39]). This in turn means that the path $R(t) = \tilde{B}_0 C(t) \tilde{B}_0^{-1}$ belongs to the

centraliser of $\mathcal{T}$ in $SO(6)$, and is such that $\tilde{B}_1 = R(1)\tilde{B}_0$. To summarise, the Schur rotation defines a path in the centraliser of $\mathcal{T}$, connecting E_0 to E_1, i.e. such that $E_1 = R(1)E_0$, and $R(t)E_0$ is invariant under $\mathcal{T}$ for every $t \in [0,1]$.

The Schur rotation can be used to define a transition with intermediate symmetry $\mathcal{T}$ in the sense of the present paper because rotating the projection plane with respect to the lattice is equivalent to rotating the lattice with respect to a fixed projection plane: in fact, E_0 is invariant under both $R(1)^\top \mathcal{H}_1 R(1)$ (as $R(1)^\top \mathcal{H}_1 R(1)B_0 = R(1)^\top \mathcal{H}_1 B_1 = R(1)^\top B_1$ belongs to E_0) and $R(t)^\top \mathcal{T} R(t)$, and since $\mathcal{P}(R(1)^\top \mathcal{L}_0) = R(1)^\top \mathcal{P}(\mathcal{L}_0)R(1)$, this means that the Schur rotation defines a family of rotated cubic 6D lattices

$$\mathcal{L}_t := R^\top(t)\mathcal{L}_0, \quad \mathcal{L}_1 := R^\top(1)\mathcal{L}_0,$$

with the property that $R(t)^\top \mathcal{T} R(t)$-orbits in $\mathcal{L}_t$ project on tetrahedral orbits in E_0. Consider now any path

$$U : [0,1] \to Sym^+(n,\mathbb{R}) \cap \mathcal{Z}(\mathcal{T},\mathbb{R}), \qquad U(0) = U(1) = I.$$

Then, choosing any lattice basis B_0 for $\mathcal{L}_0$, the path

$$B(t) = R(t)U(t)B_0, \qquad t \in [0,1],$$

defines a lattice transition with tetrahedral symmetry between $\mathcal{L}_0$ and $\mathcal{L}_1$. Furthermore, since $R(t)^\top \mathcal{T} R(t)$ and $R(1)^\top \mathcal{H}_1 R(1)$ have E_0 as invariant subspace, this transition defines in turn a transition between the quasicrystals $(\mathcal{L}_0, E_0)$ and $(\mathcal{L}_1, E_0)$ with tetrahedral symmetry as defined in Definition 3.2.5.

3.3 Transformation between aperiodic tilings

3.3.1 Transformations between planar aperiodic tilings preserving the five-fold symmetry

In this section we present three examples of transformations that preserve the global five-fold symmetry between planar tilings having that same symmetry, in particular the Penrose tiling. For the latter, we adopt here a 5-dimensional

approach instead of the usual one based on a 4D minimal embedding [3], because in this way it is simpler to describe the transitions in terms of deformations of the unit cubic cell in $\mathbb{R}^5$.

Consider the SC, BCC, and FCC lattices, and the standard basis $(\boldsymbol{e}_\alpha)$, $\alpha = 1, \ldots, 5$, in $\mathbb{R}^5$, together with the group $\mathcal{G} = C_5 \subset SO(5)$ of five-fold rotations about the body diagonal $\boldsymbol{n}$ of the unit cube:

$$\boldsymbol{n} = \sum_{\alpha=1}^{5} \boldsymbol{e}_\alpha.$$

Recall from Chapter 2 that the group C_5 leaves all the above three 5D cubic lattices invariant and has two mutually orthogonal invariant subspaces: the two-dimensional subspace $E_\parallel$ and the three-dimensional subspace $E_\perp$ with projections

$$\pi_\parallel = \begin{pmatrix} 1 & \cos(2\pi/5) & \cos(4\pi/5) & \cos(6\pi/5) & \cos(8\pi/5) \\ 0 & \sin(2\pi/5) & \sin(4\pi/5) & \sin(6\pi/5) & \sin(8\pi/5) \end{pmatrix}$$

and

$$\pi_\perp = \begin{pmatrix} 1 & \cos(4\pi/5) & \cos(8\pi/5) & \cos(2\pi/5) & \cos(6\pi/5) \\ 0 & \sin(4\pi/5) & \sin(8\pi/5) & \sin(2\pi/5) & \sin(6\pi/5) \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

With reference to Chapter 2, we choose to apply a shift $\boldsymbol{\gamma} = \frac{1}{2}\boldsymbol{n}$ to the physical space $E_\parallel$.

Projection of the SC lattice and the related Delone tiling on $E_\parallel$ produces the well-known Penrose tiling of the plane, while projecting the FCC and BCC lattices on $E_\parallel$ gives more complex aperiodic planar tilings. Examples can be seen in Figure 3.4. All these aperiodic structures have a global five-fold symmetry about the origin, and we concentrate on their structural transformations which preserve this symmetry.

Proposition 3.2.3 guarantees that the C_5-preserving transition paths for the associated 5D lattices are parametrised by the centralisers of C_5 in $GL(5, \mathbb{R})$. We therefore compute the set of centralisers in $GL(5, \mathbb{R})$ of the group C_5

solving the group of matrices M such that $G_{C_5}M = MG_{C_5}$ with

$$G_{C_5} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix},$$

Using the Maple software [50] we obtain

$$M = \begin{pmatrix} x & y & z & s & v \\ v & x & y & z & s \\ s & v & x & y & z \\ z & s & v & x & y \\ y & z & s & v & x \end{pmatrix}.$$

and consider three specific transition paths: the first two are paths between the SC and the BCC lattices in 5D:

$$T_1(t) = \begin{pmatrix} 1-t/2 & t/2 & t/2 & -t/2 & -t/2 \\ -t/2 & 1-t/2 & t/2 & t/2 & -t/2 \\ -t/2 & -t/2 & 1-t/2 & t/2 & t/2 \\ t/2 & -t/2 & -t/2 & 1-t/2 & t/2 \\ t/2 & t/2 & -t/2 & -t/2 & 1-t/2 \end{pmatrix}, \tag{3.13}$$

and

$$T_2(t) = \begin{pmatrix} 1-t/2 & t/2 & -t/2 & t/2 & -t/2 \\ -t/2 & 1-t/2 & t/2 & -t/2 & t/2 \\ t/2 & -t/2 & 1-t/2 & t/2 & -t/2 \\ -t/2 & t/2 & -t/2 & 1-t/2 & t/2 \\ t/2 & -t/2 & t/2 & -t/2 & 1-t/2 \end{pmatrix}, \tag{3.14}$$

while the third one joins the SC to the FCC lattice:

$$T_3(t) = \begin{pmatrix} 1-t/2 & 0 & 0 & 0 & t/2 \\ t/2 & 1-t/2 & 0 & 0 & 0 \\ 0 & t/2 & 1-t/2 & 0 & 0 \\ 0 & 0 & t/2 & 1-t/2 & 0 \\ 0 & 0 & 0 & t/2 & 1-t/2 \end{pmatrix}. \qquad (3.15)$$

Each intermediate lattice along these paths has global five-fold symmetry by construction. Since

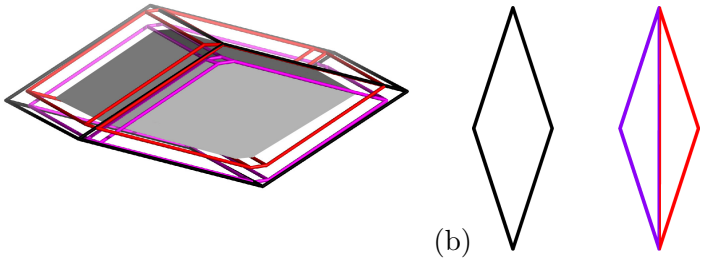
$$T_1(t)\boldsymbol{n} = (1-t/2)\boldsymbol{n} \qquad \text{and} \qquad T_2(t)\boldsymbol{n} = (1-t/2)\boldsymbol{n}$$

the first two transition paths above involve a compression of the unit cube along a body diagonal $\boldsymbol{n}$. These paths provide, through the cut-and-project method applied at each step, transformations between the Penrose and the BCC and FCC tilings.

To explain transformations in tilings in $E_\parallel$, it is useful to look at the changes, in $E_\perp$, of the projected Voronoi boundaries into which $\boldsymbol{\gamma}-\boldsymbol{q}$ projects. As a general observation, we see that the transformations of the aperiodic structures proceed through a combination of three basic mechanisms:

(1) *Splitting of a tile into two:* This occurs when a facet of the Voronoi cell splits into two. The projection on $E_\perp$ of the facet before the split, and the two new facets, are shown in Figure 3.1(a). A lattice point $\boldsymbol{q}$ such that $\pi_\perp(\boldsymbol{\gamma}-\boldsymbol{q})$ belongs to the region where the perpendicular projections overlap will be a vertex of a single tile in the first step and two tiles in the subsequent step (Figure 3.1(b)). Equivalently, two tiles join to become a single tile if the projection of $\boldsymbol{\gamma}-\boldsymbol{q}$ falls out of the intersection of two regions into a region covered by only one.

(2) *Tile flips:* Rearrangements of tiles within limited areas, which we call tile flips, occur due to points $\pi_\perp(\boldsymbol{\gamma}-\boldsymbol{q})$ in $E_\perp$ moving from one projected facet in $E_\perp$ to another, such as going through the shaded area in Figure



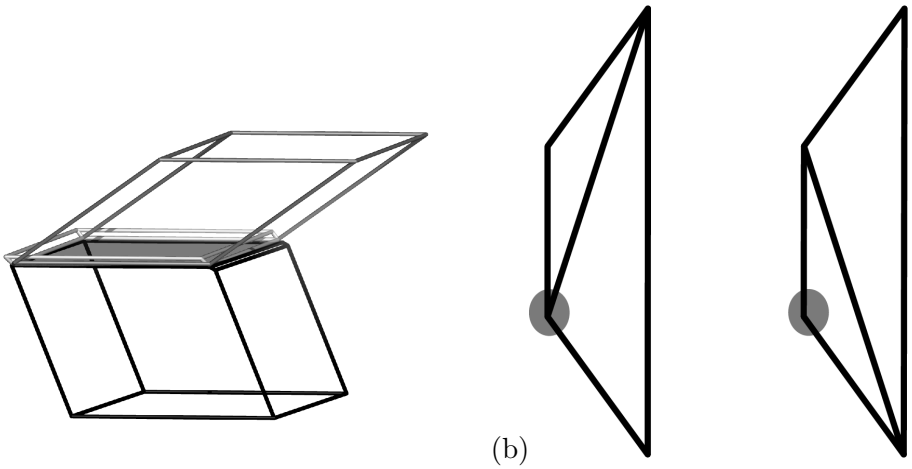
(a) (b)

Figure 3.1: (a) The splitting of a Voronoi facet into two: the larger rhombohedron is the projection on $E_\perp$ of the facet before splitting, the two smaller overlapping regions with shaded intersection being the projection of the split facet. (b) The rhomb on the left corresponds to the projection on $E_\parallel$ of the dual of the facet before splitting, while the two triangular tiles correspond to the duals of the two split facets.

3.2(a). In the upper half of the figure, the point $\boldsymbol{\gamma} - \boldsymbol{q}$ is projected into the projection of two Voronoi regions in $E_\perp$, and therefore $\pi_\parallel(\boldsymbol{q})$ is a vertex of two tiles as shown on the left of Figure 3.2(b). When the projected point goes through the shaded face, it is now only within the projection of a single Voronoi facet and $\pi_\parallel(\boldsymbol{q})$ is thus a vertex of a single tile. Since the transitions are continuous, there is a time when a projected lattice point lies in the intersection of these regions producing overlapping tiles, and at this time we insert in the tiling a glue-tile, in this case a quadrilateral tile.

(3) *Tile mergers:* During the lattice transition the Voronoi cells change, so that the projection on $E_\perp$ may involve different sets of points, see for instance Figure 3.3(a). Points $\boldsymbol{q}$ such that $\pi_\perp(\boldsymbol{\gamma} - \boldsymbol{q})$ fall outside the resulting acceptance window are deleted from the tiling, along with any tiles with these points as vertices. This produces tile mergers as seen for instance in Figure 3.3(b).

All the changes in the transforming quasicrystals and tilings can be de-



(a) (b)

Figure 3.2: (a) The projection on $E_\perp$ of three 3D facets of $\mathcal{V}(\mathbf{0})$ for some lattice $\mathcal{L}$, two above the shaded face and one below. (b) Along a transition path a projected lattice point, corresponding to the shaded vertex in the tiles in $E_\parallel$, passes from the two regions at the top of the shaded face to the lower region when projected on $E_\perp$. This results in a tile flip.

scribed through the three mechanisms above. Figure 3.4 shows four steps in the transformations from SC to BCC along the paths defined by (3.13) and (3.14), as well as the transformation from SC to FCC defined by the path in (3.15). Figure 3.4(a) shows the tiling for $t = 0$, obtained via projection from a simple cubic lattice, which then branches into three pathways. For the path (3.13) shown in Figure 3.4(b),(c) and (f), the first step in the transition is a splitting of all tiles along their long diagonal, followed by flips, deletions of vertices and further splits/recombinations. The second path (3.14) shown in Figure 3.4(d),(e) and (f) also displays at first a splitting of all the tiles, in this case along the shorter diagonal, and then proceeds as above to further splits/recombinations. For the transition from SC to FCC (3.15), Figure 3.4(g),(h) and (i), only the thin rhombs split at first, followed by further changes in the tiles, finally resulting in a much coarser tiling than the one produced by the BCC lattice.

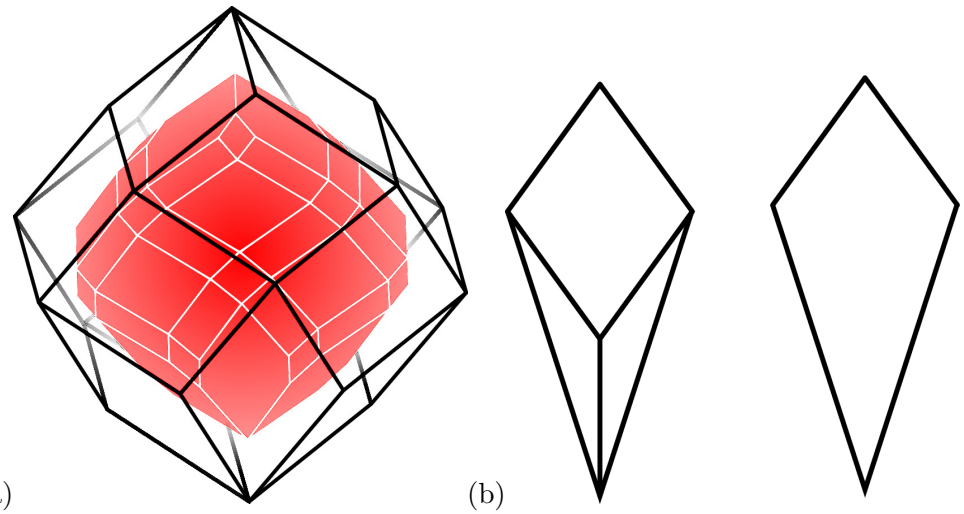


Figure 3.3: (a) The projections on $E_\perp$ of the Voronoi cells on the path (3.15), at $t = 0$ (lines) and $t = 0.3$ (shaded area). (b) The central vertex is the image of a lattice point projected near the boundary of the shrinking Voronoi cell. As the Voronoi cell shrinks, the point is no longer in the acceptance window, and is hence removed from the tiling.

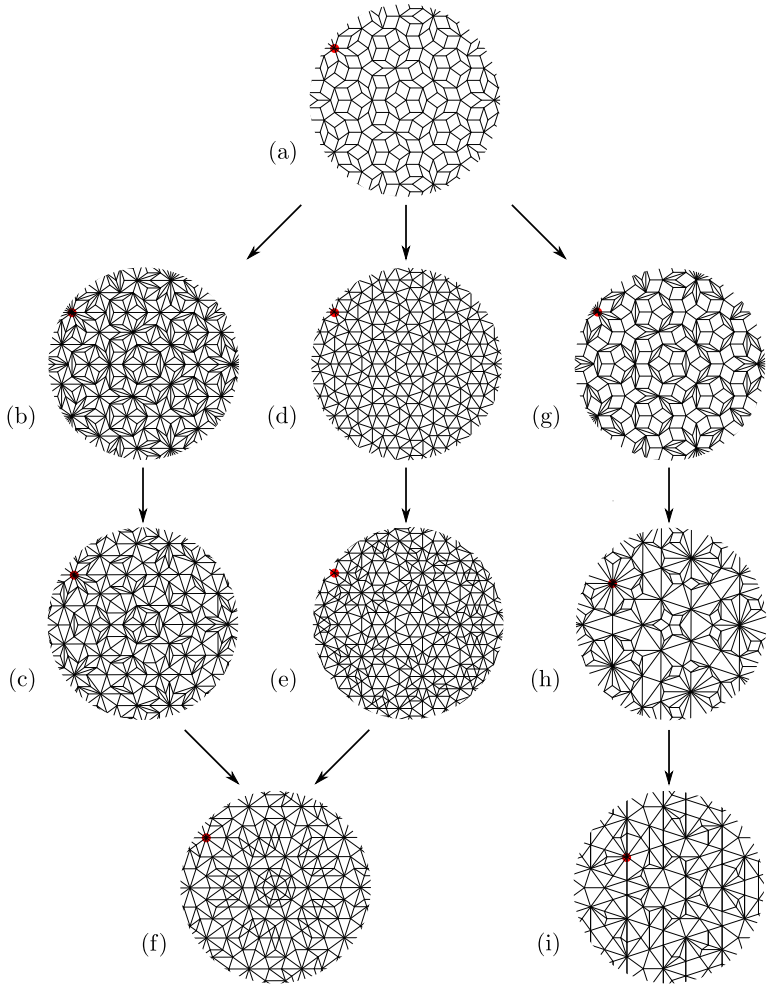


Figure 3.4: (a), (b), (c), (d), (e), (f): transition from the Penrose tiling (a) to the BCC tiling (f) via two pathways; (a), (b), (c), (f) is the path defined by (3.13), while (a), (d), (e), (f) is the path defined by (3.14). (a), (g), (h), (i): the transition from the Penrose tiling (a) to the FCC tiling (i) via the pathway defined by (3.15). Notice that, at the first time step, some vertices have disappeared and tiles have flipped, but some of the original Penrose tiles are still visible. The highlighted vertex corresponds to the 'same' lattice point and shows the relative scaling of the corresponding tilings.

3.3.2 Example of a transformation between icosahedral 3D quasicrystals

We briefly discuss here a transformation between icosahedral quasicrystals in $\mathbb{R}^3$ and their associated tilings.

We recall from Chapter 2, the definition of the two- and three-fold rotations

$$G_2 = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad G_3 = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \tag{3.16}$$

which generate a 6D integral representation of the icosahedral group $\mathcal{I}$.

The representation of $\mathcal{I}$ on $\mathbb{R}^6$ is the sum of two non-equivalent irreducible representations of degree 3, and $\mathbb{R}^6$ splits into the direct sum of two 3D subspaces, $E_\parallel$ and $E_\perp$, invariant under the icosahedral group. We recall from Chapter 2 the matrix representation of the projection on $E_\parallel$

$$\pi_\parallel = \begin{pmatrix} \tau & 0 & -1 & 0 & \tau & 1 \\ 1 & \tau & 0 & -\tau & -1 & 0 \\ 0 & 1 & \tau & 1 & 0 & \tau \end{pmatrix}$$

with $\tau = \frac{1}{2}(1+\sqrt{5})$.

The icosahedral group $\mathcal{I}$ has three maximal subgroups: the tetrahedral group $\mathcal{T}$ and the dihedral groups D_5 and D_3. Since we are interested in transitions between the SC and the FCC, BCC lattices with maximal intermediate symmetry, we therefore compute the centraliser in $GL(6,\mathbb{R})$ for the three maximal subgroups. Using the Maple software [50], for each subgroup we solve the set of equations $MG_i = G_iM$ for each generator G_i of the subgroup and with unknown M.

The tetrahedral group can be generated using the two- and three-fold gen-

erators $G_{\mathcal{T},2}$ and $G_{\mathcal{T},3}$

$$G_{\mathcal{T},2} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad G_{\mathcal{T},3} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Solving $\begin{cases} MG_{\mathcal{T},2} = G_{\mathcal{T},2}M \\ MG_{\mathcal{T},3} = G_{\mathcal{T},3}M \end{cases}$ for $M \in GL(6,\mathbb{R})$ leads to the centraliser

$$C_{\mathcal{T}} = \begin{pmatrix} z & -x & -y & -r & r & -x \\ r & z & r & x & x & y \\ -y & -x & z & r & -r & -x \\ x & -r & -x & z & y & r \\ -x & -r & x & y & z & r \\ r & y & r & -x & -x & z \end{pmatrix}, \tag{3.17}$$

with the four parameters $x, y, z, r \in \mathbb{R}$.

In a similar way, we compute the centraliser of D_5, i.e. C_{D_5}, choosing the generators $G_{D_5,2}$ and $G_{D_5,5}$ corresponding to two- and five-fold rotations as:

$$G_{D_5,2} = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}, \quad G_{D_5,5} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

leading to

$$C_{D_5} = \begin{pmatrix} z & x & y & y & x & r \\ x & z & x & y & y & r \\ y & x & z & x & y & r \\ y & y & x & z & x & r \\ x & y & y & x & z & r \\ v & v & v & v & v & w \end{pmatrix}, \tag{3.18}$$

with six parameters $x, y, z, r, u, w \in \mathbb{R}$.

The subgroup D_3 is generated using the two- and three-fold generators:

$$G_{D_3,2} = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}, \quad G_{D_3,3} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

and the centralisers of D_3 have the form

$$C_{D_3} = \begin{pmatrix} u & w & -w & x & s & s \\ -r & y & v & -v & z & -r \\ r & v & y & v & r & -z \\ z & -v & v & y & -r & -r \\ s & x & -w & w & u & s \\ s & w & -x & w & s & u \end{pmatrix}, \tag{3.19}$$

with parameters $x, y, z, r, u, w, v, s \in \mathbb{R}$.

We show that no tetrahedral transition exist between the 6D cubic lattices using transitions between the SC and FCC lattices as example. A similar proof being possible also for transitions to a BCC lattice.

First, assume that a tetrahedral transition exists. Then, according to Proposition 3.2.3-(iii) the transition operators T can be decomposed as $T = RU$, with $R \in SO(n)$ and $U \in \mathcal{Z}(\mathcal{T}, \mathbb{R}) \cap Sym^+(n, \mathbb{R})$ such that

$$B_1 = RUB_0,$$

with B_0 a basis of a SC lattice and B_1 a basis of an FCC lattice.
Any $U \in \mathcal{Z}(\mathcal{T}, \mathbb{R}) \cap Sym^+(n, \mathbb{R})$ has the form:

$$U = \begin{pmatrix} z & -x & -y & x & -x & -x \\ -x & z & -x & x & x & y \\ -y & -x & z & -x & x & -x \\ x & x & -x & z & y & -x \\ -x & x & x & y & z & -x \\ -x & y & -x & -x & -x & z \end{pmatrix},$$

with $x, y, z \in \mathbb{R}$.

Recalling the definition of an FCC lattice as:

$$\mathcal{L}_{FCC} = \left\{ \boldsymbol{x} = \tfrac{1}{2}(x_1, \ldots, x_n) : x_i \in \mathbb{Z}, \sum_{j=1}^{n} x_j = 0 \bmod 2 \right\}$$

the metric of an FCC lattice $C = B_1^\top B_1$ has entries in $\mathbb{Z}/4$, and since

$$C = B_1^\top B_1 = B_0^\top U^\top R^\top R U B_0 = B_0^\top U^2 B_0$$

and $B_0 \in GL(n, \mathbb{Z})$, then $U^2 = B_0^{-\top} C B_0^{-1}$ also has entries in $\mathbb{Z}/4$:

$$U^2 = \begin{pmatrix} a & -b & -c & b & -b & -b \\ -b & a & -b & b & b & c \\ -c & -b & a & -b & b & -b \\ b & b & -b & a & c & -b \\ -b & b & b & c & a & -b \\ -b & c & -b & -b & -b, & a \end{pmatrix},$$

with $a = z^2 + 4x^2 + y^2$, $b = 2xz$, $c = 2yz$, and $a, b, c \in \mathbb{Z}/4$.

Then $det(U^2) = det(B_0^{-\top} C B_0^{-1})$ leads to the following equation

$$[a^2 - 4b^2 - c^2]^3 = \left(\frac{1}{2}\right)^{10}$$

that cannot be fulfilled for any choice of $a, b, c \in \mathbb{Z}/4$. In fact, for $a = \alpha/4$,$b = \beta/4$,$c = \gamma/4$ with $\alpha, \beta, \gamma \in \mathbb{Z}$ the equation above reduces to

$$\left(\alpha^2 - 4\beta^2 + \gamma^2\right)^3 = 4,$$

which cannot be solved by any integer α, β, γ since there is no $n = \alpha^2 - 4\beta^2 + \gamma^2 \in \mathbb{Z}$ such that $n^3 = 4$. Similarly there cannot be any tetrahedral transitions between the SC and any rescaled FCC and BCC lattices. The same holds also for the FCC-BCC lattice transitions in 6D.

Next, we consider the possibility of transitions with D_5 symmetry of the form 3.18. As an example, we consider a path with intermediate symmetry D_5 that deforms an SC lattice (at $t = 0$) into an FCC lattice (at $t = 1$):

$$T(t) = \begin{pmatrix} 1-\frac{1}{2}t & 0 & 0 & 0 & 0 & 0 \\ 0 & 1-\frac{1}{2}t & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-\frac{1}{2}t & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-\frac{1}{2}t & 0 & 0 \\ 0 & 0 & 0 & 0 & 1-\frac{1}{2}t & 0 \\ \frac{1}{2}t & \frac{1}{2}t & \frac{1}{2}t & \frac{1}{2}t & \frac{1}{2}t & 1 \end{pmatrix}. \tag{3.20}$$

Figure 3.5 shows three snapshots of a patch of the corresponding 3D tiling around a fixed vertex, i.e. projections on $E_{\|}$ of a suitable portion of the Delone tiling of the lattices $\mathcal{L}(B(t))$, for $t = 0; 0.233$; and 1. Throughout the transition the tile arrangements have D_5 symmetry, and we observe that they evolve through mechanisms similar to those discussed in detail for the planar case in Section 3.3.1.

The main purpose of this work is to explore the effect these structural transformations have on the plane or space tilings associated with each quasicrystal. The local tile rearrangements can be understood in terms of the change of geometry of the Voronoi cell of the higher dimensional lattice during the transition, and we analyse the effect this has on the projected tiling. Our results suggest that the possible ways in which an aperiodic tiling can change, while still conserving some intermediate symmetry, reduces to the three basic mechanisms of tile splitting, tile flipping, and tile merger. The general patterns identified here may provide a basis for further analysis of structural

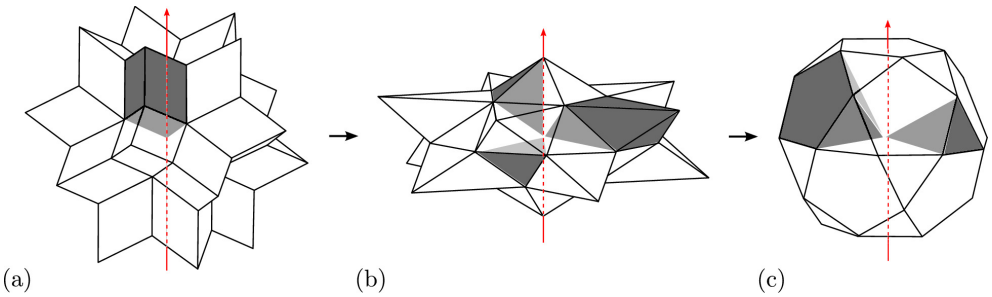


Figure 3.5: Vertex stars at the origin of the tilings obtained by projection of (a) the 6D SC lattice, (b) an intermediate D_5 lattice and (c) the FCC lattice. The vertex star is composed by repetitions of a number of suitable tiles, highlighted in the figure. The pentagonal pyramid in (c) is a glue tile.

transitions in quasilattices and for a possible classification of transitions in aperiodic structures. This is of interest both from a theoretical viewpoint and for the applications. In the following section, we show how these structural transitions in quasilattices may help us understand the structural rearrangements of viral capsids necessary for infection.

3.4 Viral transitions via point arrays

The majority of viral capsids exhibit icosahedral symmetry before as well as after their structural transition. As explained in Chapter 1, the pre- and post-transition configurations can then be mapped using icosahedrally invariant point arrays from [37]. Note that the approximating point array is not necessarily unique as different members of an ensemble of point arrays may provide equally good approximations of the capsid geometry according to [37]. As we will explain shortly, these point arrays can be obtained as projections into 3D space of suitable subsets of the icosahedral Bravais lattices in 6D (i.e. the SC, BCC and FCC lattices). We can therefore, in principle, study the transitions between two viral configurations (represented by the two corresponding icosahedral point arrays) in terms of transitions between the two 6D lattices that generate them via the cut-and-project method (see Figure 3.6).

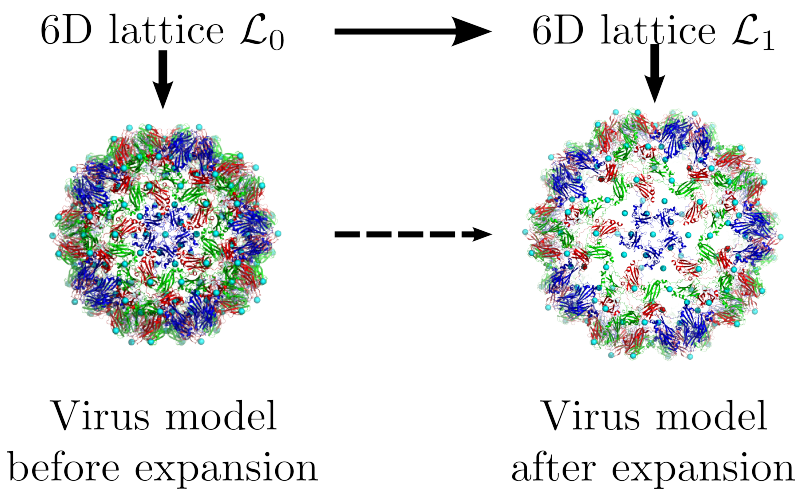


Figure 3.6: Sketch of the procedure: transitions in 6D can be used to study conformational changes of models of 3D viral capsids.

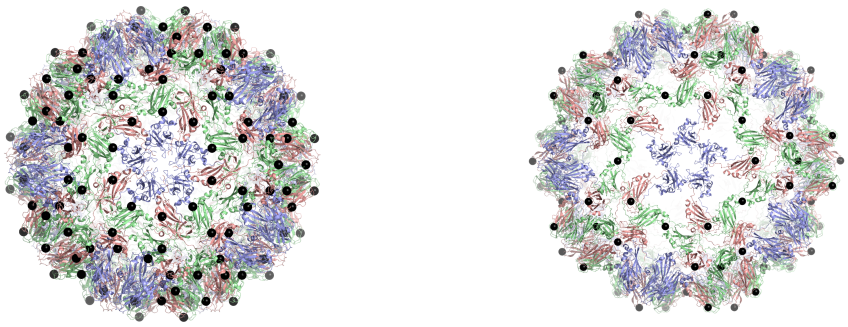
Recall from Chapter 1 that these point arrays are obtained using affine extension of the 3D icosahedral group $\mathcal{I}_3$ onto two nested polyhedra (either an icosahedron, a dodecahedron or an icosadodecahedron). We will later refer to these nested polyhedra as the *skeletal structure*. Therefore,

Definition 3.4.1. *For every element S in the classification of icosahedral point arrays listed in [35], there exists $\boldsymbol{u}$, $\boldsymbol{r}$ and $\boldsymbol{s}$ such that*

$$S \equiv S(\boldsymbol{u}, \boldsymbol{r}, \boldsymbol{s}) = \mathcal{I}_3\boldsymbol{u} \cup \mathcal{I}_3\boldsymbol{r} \cup (\mathcal{I}_3\boldsymbol{u} + \mathcal{I}_3\boldsymbol{s}) \cup (\mathcal{I}_3\boldsymbol{r} + \mathcal{I}_3\boldsymbol{s}). \tag{3.21}$$

where $\boldsymbol{u}$, $\boldsymbol{r}$ and $\boldsymbol{s}$ are vectors that point along a two, a three or a fivefold symmetry axis and $\mathcal{I}_3\boldsymbol{u}$, $\mathcal{I}_3\boldsymbol{r}$ and $\mathcal{I}_3\boldsymbol{s}$ their respective icosahedral orbits.

We will first show that point arrays are finite subsets of the vertex sets of suitable 3D icosahedral quasicrystals and that they can be uniquely associated with icosahedral Bravais lattices in 6D. In Section 3.4.2 the procedure to compute all transitions of the form (3.12) such that point arrays corresponding to the native state map to those of the final state is presented. To illustrate our approach we study the deformations for the 6D lattices related to the Cowpea Chlorotic Mottle Virus (CCMV). The point arrays corresponding to the



(a) (b)

Figure 3.7: (a) The point array $10-44$ superimposed on the pre-transition state of CCMV (pdb-id: *1cwp*). (b) The point array $27-51$ superimposed on the post-transition state of CCMV (pdb-id: *ccmv_swln_1*).

start and end configuration have been determined based on the pdb-files with pdb-id *1cwp* for the pre-, and *ccmv_swln_1* for the post-transition configuration [82]. As an example we display the point arrays corresponding to these configurations in Figure 3.7.

3.4.1 Embedding of the capsid geometry into a 6D icosahedral lattice

In order to relate a point array of the form (3.4.1) with a 6D lattice, we use the following facts:
(i) The projection $\pi_{\parallel}$ is one-to-one onto its image when restricted to $\mathcal{L}$, since $E_{\parallel}$ is totally irrational (i.e., $E_{\parallel} \cap \mathcal{L} = \{0\}$) with respect to $\mathcal{L}_{SC}, \mathcal{L}_{FCC}$ and $\mathcal{L}_{BCC}$.
(ii) The icosahedral group commutes with the projection, so that the 6D preimages of the standard icosahedral polyhedra are, in turn, icosahedral orbits.
(iii) A dilatation by a factor of τ in 3D corresponds to a symmetry operation of the cubic lattices in 6D.

We use the following three-step procedure:

- **Step 1:** We embed the standard polyhedra (i.e. icosahedron, dodec-

ahedron and icosadodecahedron) into 6D according to Table 3.1: The standard icosahedron can be obtained via projection of the icosahedral orbit of the canonical basis vector $\boldsymbol{e}_1$ of the SC lattice, the standard icosidodecahedron via projection of the icosahedral orbits of the FCC lattice vector $\frac{1}{2}(\boldsymbol{e}_1 + \boldsymbol{e}_2)$, and the standard dodecahedron via projection of the icosahedral orbit of the BCC lattice vector $\frac{1}{2}(\boldsymbol{e}_1 + \boldsymbol{e}_2 - \boldsymbol{e}_3 + \boldsymbol{e}_4 - \boldsymbol{e}_5 + \boldsymbol{e}_6)$.

- **Step 2:** Since the skeletal structure of the point arrays constructed in [37] is a combination of two standard polyhedra at different scalings of τ^k with $k \in \mathbb{Z}$ (see Tables 3.2, 3.3), we create the 6D counterparts of standard polyhedra rescaled by τ via the action of the quasidilatation D (see 2.8 of Chapter 2). For example, the 6D embedding of a icosahedron of length $\tau^k\sqrt{2+\tau}$, $k \in \mathbb{Z}$ is given by the icosahedral orbit of the rescaled vector $D^k\boldsymbol{e}_1$.

- **Step 3:** The translation vectors $\boldsymbol{s} \in \mathbb{R}^3$, along which the skeletal structure (i.e. the two nested standard polyhedra) is translated to generate the point arrays, also belongs to a rescaled standard polyhedron. Therefore $\boldsymbol{t} = (\pi_{\parallel})^{-1}(\boldsymbol{s})$ and all its orbits belong to one of the icosahedral lattices.

Hence, we associate with each point array S in (3.4.1) a unique set Σ of points in either $\mathcal{L}_{SC}$, $\mathcal{L}_{FCC}$ or $\mathcal{L}_{BCC}$. It is called the *lifted viral configuration* or *lifted point array*, and fulfills $\pi_{\parallel}(\Sigma) = S$. It follows from (3.4.1) that Σ is the union of icosahedral orbits in $\mathbb{R}^6$ and their suitable translates:

$$\Sigma = \Sigma(\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{t}) = \mathcal{I}\boldsymbol{v} \cup \mathcal{I}\boldsymbol{w} \cup (\mathcal{I}\boldsymbol{v} + \mathcal{I}\boldsymbol{t}) \cup (\mathcal{I}\boldsymbol{w} + \mathcal{I}\boldsymbol{t}), \tag{3.22}$$

where $\pi_{\parallel}\boldsymbol{v} = \boldsymbol{u}$, $\pi_{\parallel}\boldsymbol{w} = \boldsymbol{r}$, $\pi_{\parallel}\boldsymbol{t} = \boldsymbol{s}$.

By construction, all points of a given lifted viral configuration Σ are points of some icosahedral lattice, and there exists a unique minimal such lattice that contains a given lifted viral configuration. We say that the lifted point array is *embedded* into such a lattice, and the following characterisation holds:

Table 3.1: 6D vectors and their 3D projections corresponding to the vertices of the standard polyhedra. Vectors are listed modulo inversion.

	Orbits in 6D modulo inversion	3D representation
Icosahedron	$(1,0,0,0,0,0)$	$(\tau,1,0)$
	$(0,1,0,0,0,0)$	$(0,\tau,1)$
	$(0,0,1,0,0,0)$	$(-1,0,\tau)$
	$(0,0,0,1,0,0)$	$(0,-\tau,1)$
	$(0,0,0,0,1,0)$	$(\tau,-1,0)$
	$(0,0,0,0,0,1)$	$(1,0,\tau)$
Dodecahedron	$\frac{1}{2}(1,-1,1,1,-1,1)$	$(0,1-\tau,\tau)$
	$\frac{1}{2}(-1,1,-1,1,1,1)$	$(1,-1,1)$
	$\frac{1}{2}(1,-1,-1,-1,-1,1)$	$(1,1,-1)$
	$\frac{1}{2}(1,1,-1,1,-1,1)$	$(1,1,1)$
	$\frac{1}{2}(1,1,1,1,-1,-1)$	$(-1,1,1)$
	$\frac{1}{2}(1,1,1,-1,1,-1)$	$(\tau-1,\tau,0)$
	$\frac{1}{2}(-1,1,-1,1,-1,-1)$	$(-\tau,0,1-\tau)$
	$\frac{1}{2}(-1,-1,1,-1,-1,1)$	$(-\tau,0,-1+\tau)$
	$\frac{1}{2}(1,-1,1,1,1,-1)$	$(-1+\tau,-\tau,0)$
	$\frac{1}{2}(1,-1,-1,1,-1,-1)$	$(0,1-\tau,-\tau)$
Icosidodecahedron	$\frac{1}{2}(1,0,0,-1,0,0)$	$\frac{1}{2}(\tau,\tau^2,-1)$
	$\frac{1}{2}(0,1,0,0,0,1)$	$\frac{1}{2}(1,\tau,\tau^2)$
	$\frac{1}{2}(0,1,1,0,0,0)$	$\frac{1}{2}(-1,\tau,\tau^2)$
	$\frac{1}{2}(0,0,0,1,0,1)$	$\frac{1}{2}(1,-\tau,\tau^2)$
	$\frac{1}{2}(1,0,0,0,0,1)$	$\frac{1}{2}(\tau^2,1,\tau)$
	$\frac{1}{2}(1,0,0,0,1,0)$	$(\tau,0,0)$
	$\frac{1}{2}(0,0,0,1,1,0)$	$\frac{1}{2}(\tau,-\tau^2,1)$
	$\frac{1}{2}(0,0,0,0,1,1)$	$\frac{1}{2}(\tau^2,-1,\tau)$
	$\frac{1}{2}(0,0,1,1,0,0)$	$\frac{1}{2}(-1,-\tau,\tau^2)$
	$\frac{1}{2}(0,0,1,0,-1,0)$	$\frac{1}{2}(-\tau^2,1,\tau)$
	$\frac{1}{2}(1,0,-1,0,0,0)$	$\frac{1}{2}(\tau^2,1,-\tau)$
	$\frac{1}{2}(1,1,0,0,0,0)$	$\frac{1}{2}(\tau,\tau^2,1)$
	$\frac{1}{2}(0,0,1,0,0,1)$	$(0,0,\tau)$
	$\frac{1}{2}(0,1,0,-1,0,0)$	$(0,\tau,0)$
	$\frac{1}{2}(0,1,0,0,-1,0)$	$\frac{1}{2}(-\tau,\tau^2,1)$

Proposition 3.4.2. *The minimal lattice containing a given lifted viral configuration Σ has icosahedral symmetry.*

Proof. Let $\Sigma = \{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_N\}\cup\{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_M\}\cup\{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_K\}$, where $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_N\}$ and $\{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_M\}$ are distinct $\mathcal{I}$-orbits, and let $\{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_K\}$ be the icosahedral orbit of the translation vector. Consider the $\mathbb{Z}$-module

$$\mathcal{L}' = \left\{ \boldsymbol{x}\in\mathbb{R}^6 \,:\, \boldsymbol{x} = \sum_{i=1}^{N} n^i\boldsymbol{v}_i + \sum_{j=1}^{M} m^j\boldsymbol{w}_j + \sum_{k=1}^{K} p^k\boldsymbol{t}_k \,,\, n^i,m^j,p^k\in\mathbb{Z} \right\}.$$

Since $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_N\}$, $\{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_M\}$, and $\{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_K\}$ are icosahedral orbits, $\mathcal{L}'$

Table 3.2: The list of shells (i.e. standard polyhedra and their corresponding translations) relevant to the description of CCMV uses the same numbering as in [37]. For the construction of a shell, the standard polyhedron can either be an icosahedron (ICOS), a dodecahedron (DOD) or an icosidodecahedron (IDD). The corresponding translation vector points to the vertices of either an ICOS, a DOD or an IDD.

shell #	st. polyhedron	translation	scaling factor
10	ICO	DOD	τ^2
11	ICO	ICO	$-\tau'$
12	ICO	ICO	1
13	ICO	ICO	τ
19	DOD	ICO	$-2\tau'$
26	DOD	DOD	τ^2
27	DOD	ICO	τ'^2
29	DOD	ICO	1
30	DOD	ICO	τ
44	IDD	DOD	$\frac{1}{2}\tau$
51	IDD	ICO	$\frac{1}{2}$
52	IDD	ICO	$\frac{1}{2}\tau$
53	IDD	ICO	1
54	IDD	ICO	$\frac{1}{2}\tau^2$
55	IDD	ICO	τ

Table 3.3: List of generating vectors and minimal lattices for each viral configuration relevant to CCMV. The corresponding viral configurations are denoted by the two shells (listed in Table 3.2) from which they are composed. For example viral configuration 10-44 combines shell 10 and 44 as explained in Chapter 1.

viral conf.	lattice	$\boldsymbol{v}$	$\boldsymbol{w}$	$\boldsymbol{t}$
10-44	SC	$(1,0,0,0,0,0)$	$(1,1,0,0,1,1)$	$(1,1,0,0,0,1)$
26-44	BCC	$(1,0,0,1,0,-1)$	$(1,0,0,0,1,0)$	$\frac{1}{2}(1,1,1,1,-1,-1)$
11-27	BCC	$\frac{1}{2}(1,-1,1,1,-1,-1)$	$\frac{1}{2}(1,1,-1,1,1,-1)$	$\frac{1}{2}(3,-1,1,1,-1,-1)$
12-27	SC	$(1,0,0,0,0,0)$	$(1,1,0,0,0,1)$	$(1,0,0,0,0,0)$
13-27	BCC	$(1,1,1,1,1,-2)$	$\frac{1}{2}(1,1,1,1,-1,-1)$	$\frac{1}{2}(3,-1,1,1-1,-1)$
27-51	SC	$(0,1,1,0,-1,0)$	$(0,1,0,0,-1,0)$	$(1,0,0,0,0,0)$
27-52	SC	$(0,1,0,-1,-1,0)$	$(0,0,1,0,1,0)$	$(1,0,0,0,0,0)$
27-53	FCC	$\frac{1}{2}(1,-1,-1,-1,-1,1)$	$\frac{1}{2}(1,0,1,1,0,-1)$	$\frac{1}{2}(3,-1,1,1,-1,-1)$
27-54	SC	$(0,1,0,-1,-1,0)$	$(1,0,1,0,1,-1)$	$(1,0,0,0,0,0)$
27-55	FCC	$\frac{1}{2}(1,-1,1,1,1,-1)$	$\frac{1}{2}(1,1,0,2,0,-2)$	$\frac{1}{2}(3,-1,1,1,-1,-1)$
27-29	BCC	$\frac{1}{2}(1,1,-1,1,-1,1)$	$\frac{1}{2}(1,3,-1,3,1,-3)$	$(1,0,0,0,0,0)$
27-30	SC	$(1,1,0,0,0,1)$	$(1,-1,1,0,0,0)$	$(1,0,0,0,0,0)$

is invariant under $\mathcal{I}$. Furthermore, $\mathcal{L}'$ is a lattice, because $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$ and $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_M\}$, and $\{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_K\}$ are all lattice vectors of the maximal lattice FCC. It is always possible to extract a basis of $\mathbb{R}^6$ from Σ, since otherwise the subspace generated by Σ, being invariant under $\mathcal{I}$, would coincide either with $E_\parallel$ or $E_\perp$, both irrational with respect to the cubic lattices. But the elements of Σ are lattice vectors, which leads to a contradiction. Further, $\mathcal{L}'$ contains Σ, and it follows that $\mathcal{L}'$ is a 6D icosahedral lattice. Finally, $\mathcal{L}'$ is also minimal, since every lattice that contains Σ must also contain all integer linear combinations of vectors of Σ, that is, $\mathcal{L}'$. □

A given viral configuration can trivially be embedded in infinitely many other lattices but the minimal lattice constructed above is unique. It follows that, since the point arrays can be obtained by projection of a 6D lattice onto a completely irrational icosahedrally-invariant subspace in 3D, they are subsets of an icosahedral 3D quasicrystal [66].
The following definition is central in our approach to the study of structural transformations of viral configurations.

Definition 3.4.3. *Given a lifted viral configuration $\Sigma = \Sigma(\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{t})$, any basis $\{\boldsymbol{b}_\alpha\}_{\alpha=1,\ldots,6}$ of $\mathbb{R}^6$ such that*

i) $\{\boldsymbol{b}_\alpha\}$ is a basis for the minimal icosahedral lattice containing Σ;

ii) every basis vector $\boldsymbol{b}_\alpha$ belongs to either $\mathcal{I}\boldsymbol{v}$, $\mathcal{I}\boldsymbol{w}$, or $\mathcal{I}\boldsymbol{t}$;

iii) each orbit contains at least one basis vector;

is called an admissible basis *for the lifted viral configuration.*

The second condition requires that the vectors of an admissible basis are representatives of icosahedral orbits of the skeletal vectors and of the translation. Since the icosahedral orbits of these three vectors are used to construct Σ, there exists a unique Σ for each basis B. We write this unique viral configuration as

$$\Sigma = \Sigma(\boldsymbol{b}_\alpha) = \Sigma(B). \tag{3.23}$$

Note that all admissible bases for a given lifted viral configuration are bases of the same 6D lattice. Furthermore, applying a rotation $R \in \mathcal{I}$ to the vectors of an admissible basis yields another admissible basis for the same lifted viral configuration, and the same is true for permutations and change of signs of the basis vectors.

3.4.2 Viral transitions

Consider two viral configurations S_0 and S_1, with corresponding lifted viral configurations Σ_0 and Σ_1 in 6D. Let Σ_0 be embedded into $\mathcal{L}_0$, and Σ_1 into $\mathcal{L}_1$.

Definition 3.4.4. *A viral transition between two viral configurations S_0 and S_1 in 3D, with intermediate symmetry $\mathcal{G}$ is a transition obtained via the cut-and-project method from a transition between the lattices $\mathcal{L}_0$ and $\mathcal{L}_1$ in 6D, such that*
i) B_0 and B_1 (the bases associated with the transition) are admissible for Σ_0 and Σ_1, i.e.,

$$\Sigma_0 = \Sigma(B_0), \qquad \Sigma_1 = \Sigma(B_1);$$

ii) the transition is compatible with the cut-and-project method as stated in Definition 3.2.5.

By Proposition 3.2.6, given a transition T, the possible viral transition paths are curves in $\mathcal{Z}(\mathcal{G}, \mathbb{R})$ connecting the identity with T. To derive from these paths the information about the actual intermediate structure of a viral capsid, let

$$T(t) \in \mathcal{Z}(\mathcal{G}, \mathbb{R}), \quad \text{for} \quad t \in [0, 1],$$

be a viral transition path with intermediate symmetry $\mathcal{G}$. Let $\boldsymbol{v}_0$, $\boldsymbol{w}_0$ and $\boldsymbol{t}_0$ be three vectors of the basis B_0 such that

$$\Sigma_0 = \Sigma(\boldsymbol{v}_0, \boldsymbol{w}_0, \boldsymbol{t}_0),$$

i.e., $\boldsymbol{v}_0$ and $\boldsymbol{w}_0$ belong to the skeletal shells, and $\boldsymbol{t}_0$ is in the orbit of the translation vector. For $t \in [0, 1]$, define

$$\boldsymbol{v}(t) = T(t)\boldsymbol{v}_0, \quad \boldsymbol{w}(t) = T(t)\boldsymbol{w}_0, \quad \boldsymbol{t}(t) = T(t)\boldsymbol{t}_0.$$

By definition, $\boldsymbol{v}(t)$, $\boldsymbol{w}(t)$ and $\boldsymbol{t}(t)$ are vectors of the basis $B(t) = T(t)B_0$ of the intermediate lattice. We associate with the transition path $T(t)$, for any $t \in [0,1]$, a lifted viral configuration $\Sigma(t)$ defined as

$$\Sigma(t) = \mathcal{G}\boldsymbol{v}(t) \cup \mathcal{G}\boldsymbol{w}(t) \cup (\mathcal{G}\boldsymbol{v}(t) + \mathcal{G}\boldsymbol{t}(t)) \cup (\mathcal{G}\boldsymbol{w}(t) + \mathcal{G}\boldsymbol{t}(t)) \subset \mathcal{L}(B(t)). \quad (3.24)$$

The resulting point array, when projected to $\mathbb{R}^3$ via $\pi_{\|}$, yields a family of non-icosahedral point sets $S(t)$ parametrised by t, with constant $\mathcal{G}$-symmetry, that represents the intermediate viral configurations, i.e.

$$S(t) = \mathcal{G}_3\boldsymbol{u}(t) \cup \mathcal{G}_3\boldsymbol{r}(t) \cup (\mathcal{G}_3\boldsymbol{u}(t) + \mathcal{G}_3\boldsymbol{s}(t)) \cup (\mathcal{G}_3\boldsymbol{r}(t) + \mathcal{G}_3\boldsymbol{s}(t)), \quad (3.25)$$

where $\mathcal{G}_3$ denotes a representation of $\mathcal{G}$ on $E_{\|}$, and $\boldsymbol{u}(t) = \pi_{\|}\boldsymbol{v}(t)$, $\boldsymbol{r}(t) = \pi_{\|}\boldsymbol{w}(t)$ and $\boldsymbol{s}(t) = \pi_{\|}\boldsymbol{t}(t)$.

3.4.3 A procedure to determine transitions between viral configurations

As shown in Section 3.3.2, the centralisers of the three maximal subgroups of the icosahedral group in $GL(6,\mathbb{R})$ are 6×6 matrices whose coefficients depend linearly on a finite number n of parameters, with $n = 4, 6, 8$ for tetrahedral, D_5 and D_3 transitions respectively (see equations 3.17, 3.18 and 3.19). In order to find transitions between two given lifted viral configurations, we use the following strategy to determine their parameter values: we solve a linear system of equations that formulates the requirement that the vectors acting as descriptors of the pre-transitional configuration, be mapped into descriptors of the final configuration. For this, we write $T = T(\boldsymbol{p})$, with $\boldsymbol{p} = (p_1, \ldots, p_n) \in \mathbb{R}^n$,and $\Sigma_0 = \Sigma(\boldsymbol{v}_0, \boldsymbol{w}_0, \boldsymbol{t}_0)$, $\Sigma_1 = \Sigma(\boldsymbol{v}_1, \boldsymbol{w}_1, \boldsymbol{t}_1)$. Choosing $\overline{\boldsymbol{v}}_0 \in \mathcal{I}\boldsymbol{v}_0$, $\overline{\boldsymbol{w}}_0 \in \mathcal{I}\boldsymbol{w}_0$, $\bar{\boldsymbol{t}}_0 \in \mathcal{I}\boldsymbol{t}_0$, $\overline{\boldsymbol{v}}_1 \in \mathcal{I}\boldsymbol{v}_1$, $\overline{\boldsymbol{w}}_1 \in \mathcal{I}\boldsymbol{w}_1$, and $\bar{\boldsymbol{t}}_1 \in \mathcal{I}\boldsymbol{t}_1$, we require that

$$\overline{\boldsymbol{v}}_1 = T(\boldsymbol{p})\overline{\boldsymbol{v}}_0, \quad \overline{\boldsymbol{w}}_1 = T(\boldsymbol{p})\overline{\boldsymbol{w}}_0, \quad \bar{\boldsymbol{t}}_1 = T(\boldsymbol{p})\bar{\boldsymbol{t}}_0\,. \quad (3.26)$$

This yields a system of equations for the unknown parameter $\boldsymbol{p}$, whose solution (if it exists) we denote by $\overline{\boldsymbol{p}}$. Notice that, by construction, $T(\overline{\boldsymbol{p}})$ is in the

centraliser of $\mathcal{G}$, so that $T(\overline{\boldsymbol{p}})G\boldsymbol{v} = GT(\overline{\boldsymbol{p}})\boldsymbol{v}$ for every $G \in \mathcal{G}$ and $\boldsymbol{v} \in \mathbb{R}^6$. Hence, $\mathcal{G}$-orbits in Σ_0 are mapped into $\mathcal{G}$-orbits in Σ_1. If a basis $\overline{B}_0$ for the minimal lattice containing Σ_0 can be extracted from the $\mathcal{G}_0$-orbits of $\overline{\boldsymbol{v}}_0$, $\overline{\boldsymbol{w}}_0$, and $\bar{\boldsymbol{t}}_0$, then this basis is automatically admissible for Σ_0. Moreover, its image $\overline{B}_1 = T(\overline{\boldsymbol{p}})\overline{B}_0$ is also admissible for Σ_1, and $T(\overline{\boldsymbol{p}})$ is a viral transition between Σ_0 and Σ_1.

By repeated application of the above procedure for all possible representatives of the orbits $\mathcal{I}\boldsymbol{v}_0$, $\mathcal{I}\boldsymbol{w}_0$, $\mathcal{I}\boldsymbol{t}_0$ and $\mathcal{I}\boldsymbol{v}_1$, $\mathcal{I}\boldsymbol{w}_1$, $\mathcal{I}\boldsymbol{t}_1$, we obtain all possible viral transitions between Σ_0 and Σ_1.

3.4.4 Application to CCMV

We have applied the procedure developed in the previous sections to CCMV capsid transitions during maturation. An analysis of the CCMV capsid structure via the algorithm in [37] shows that the pre-transition geometry of the CCMV capsid is given by one of the two following viral configurations (see Tables 3.2, 3.3 for the notations used to label the point-arrays):

$$S_0 \in \{\text{10-44}, \text{26-44}\}\,, \tag{3.27}$$

while the swollen form of CCMV is best approximated by one of the ten following viral configurations:

$$S_1 \in \{\text{11-27}, \text{12-27}, \text{13-27}, \text{27-29}, \text{27-30}, \text{27-51}, \text{27-52}, \text{27-53}, \text{27-54}, \text{27-55}\}. \tag{3.28}$$

We have determined the possible transitions with capsid configurations given by a point array S_0 in (3.27), and post-transition configuration given by an array S_1 in (3.28). The results are as follows:

i) In Section 3.3.2 we demonstrated that there is no transition with tetrahedral symmetry between SC, FCC and BCC lattices. Therefore no transition between initial and final configurations with this symmetry can exist.

ii) There exist no transitions with D_5 symmetry between any of the initial and final configurations.

iii) There exist four transitions with D_3 symmetry, mapping the initial configuration 10-44 into one of the final configurations 27-52, 11-27 and 12-27.

Here we list these transitions together with the bases of the initial configuration involved in each transition.

- D_3, 10-44→27-52

$$T_{1044-2752} = \begin{pmatrix} 0 & -1 & 1 & -3 & 1 & 1 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 1 & 0 & 0 & 0 \\ 1 & -3 & 1 & -1 & 0 & 1 \\ 1 & -1 & 3 & -1 & 1 & 0 \end{pmatrix}, \quad B_{1044} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 & 0 & 1 \\ -1 & 0 & 0 & 1 & -1 & 1 \end{pmatrix},$$

- D_3, 10-44→27-52

$$T'_{1044-2752} = \begin{pmatrix} -1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & -1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 \\ -1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & -1 & 0 \\ 0 & 1 & -1 & 1 & 0 & -1 \end{pmatrix}, \quad B_{1044} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & -1 & -1 & 1 & 0 \\ 1 & 0 & -1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix},$$

- D_3, 10-44→11-27

$$T_{1044-1127} = \begin{pmatrix} 2 & -1/2 & 1/2 & 1/2 & -1 & -1 \\ 0 & -1/2 & 1/2 & -1/2 & 1 & 0 \\ 0 & 1/2 & -1/2 & 1/2 & 0 & -1 \\ 1 & -1/2 & 1/2 & -1/2 & 0 & 0 \\ -1 & 1/2 & 1/2 & -1/2 & 2 & -1 \\ -1 & -1/2 & -1/2 & -1/2 & -1 & 2 \end{pmatrix}, \quad B_{1044} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & -1 \end{pmatrix},$$

- D_3, 10-44→12-27

$$T_{1044-1227} = \begin{pmatrix} 1 & -1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 \\ -1 & 0 & -1 & 0 & -1 & 1 \\ -1 & 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 \end{pmatrix}, \quad B_{1044} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

3.5 Conclusion

In this chapter we have determined the possible transitions between SC, BCC and FCC lattices such that the projected quasilattices conserve maximal symmetry. Our procedure describes structural transformations in cut-and-project quasilattices, and all structural transitions with all possible intermediate symmetries have been considered. We have characterised all local changes in these quasilattice transitions via three mechanisms: splitting of a tile, tile flips and tile mergers.

We have also demonstrated that quasilattice transitions can be used to study structural transformations of viral capsids when combined with point arrays descriptors from [36]. Note that our approach cannot determine the exact transition path (as the energy landscape of transitions is unknown), however the maximal symmetry of the intermediates has been determined. In particular, by embedding the point descriptors corresponding to the pre- and post-transitional states of an examplar virus (the Cowpea Chlorotic Mottle Virus- CCMV), into higher dimensional lattices, the possible transitions between these configurations have been computed. Provided that maximal symmetry is conserved along the transition path, we have shown that icosahedral symmetry is lost at intermediate states and that the maximal symmetry preserved is D_3, i.e. a single three fold axis is preserved in the case of this virus. This result is consistent with the coarse-grained study in [14] which suggests that different local environments favour loss of icosahedral symmetry and that the structural transition is triggered at a given symmetry axis of the viral capsid. Here, conservation of a single three-fold axis indicates that the transformation propagates along this axis like a circular wave.

Chapter 4

The Matching Algorithm

We have seen in Chapter 1 that the icosahedrally invariant point arrays in [37] can be used as descriptors of viral capsids. This implies that the structures of these capsids are constrained by icosahedral symmetry at different radial levels. In Chapter 3 we have shown that these point arrays are in fact a subset of the vertex set of an icosahedral tiling. Here we investigate further the icosahedral symmetry constraints by embedding these point arrays into complete tilings and associating tilings, rather than points arrays, with virus architecture. In this chapter we explain the algorithm used to match capsids to subsets of tiles from either of the three icosahedral tilings (i.e. SC, BCC or FCC), that have been obtained by the cut-and-project method in Chapter 2. Application to a selection of $T = 1$ and $T = 3$ viral capsids will be discussed in Chapter 5.

4.1 The Matching Algorithm

The experimental data of viral structure used by the algorithm is provided by the Viperdb website [12], a database for capsid structures derived from X-ray and Cryo-EM experiments. Each pdb-file contains the cartesian coordinates of the atoms of the protein shell and, if available, of the genomic material present in the viral capsid. First we describe how these data are processed

before applying the model. The algorithm which compares the structure of the viral capsid to the icosahedral tiling (be it SC, BCC or FCC) is explained in detail in Section 4.1.2 and can be summarised as:

1. Data preparation and alignment to tilings:

   Information about the atomic positions of the capsid proteins are extracted, removing atoms related to genomic material or water molecules. A surface representation of the data is generated via PyMol, and a discrete subset of the vertices in the corresponding mesh is retained for further analysis. Viral capsids are embedded into the tilings via alignment of the icosahedral symmetry axes, such that the only degree of freedom remaining is the scaling of tiling size relative to the data. Different scalings of the tilings for fixed data size will be assessed in increments of 0.1 Å in the following.

2. Identification of a maximal scaling:

   An upper bound s_{max} on the scaling of tilings relative to the data is determined by requiring that the vertex star of the origin does not contain any atoms. Otherwise the tiling would be too large to account for the interior surface of the capsid.

3. Computation of tile occupancy:

   For every scaling, starting from s_{max} and decreasing in increments of 0.1 Å, the occupancy of tiles by atoms in the data set is computed for every tile contained in the fundamental domain of the symmetry group.

4. Creation of a library of tile configurations as candidates for representing the data:

   For a given tiling and scaling, all tiles with an occupancy of at least 50% are retained. Denoting the occupancy labelled in increasing order of the N tiles in the fundamental domain with at least 50% occupancy as ρ_i (i=1 . . . N), we call $\mathcal{S}(s, \rho_j)$ the s-scaled tiling subset consisting of all tiles with at least occupancy ρ_j. The tile sets $\mathcal{S}$ are computed for every

tiling and scaling. If such a tile set, contains at least 90% of the C_α atoms of the viral capsid (modulo experimental precision) then this tile set is added to the library of tile configurations that contain all possible candidates to represent the data.

5. Computation of integral square error for each tile configuration in the library:

   The integral square error σ_{sd}, a concept from polynomial approximation, is computed for each tile configuration in the library and plotted against the scaling s corresponding to the tile configuration. A cut-off $\Delta\sigma_{sd}$ is determined depending on the experimental precision of the data such that approximations with a value of σ_{sd} below that would be too precise given the uncertainty of the data. The minimal scaling s_{min} is defined as the scaling below which tile sets $\mathcal{S}$ with lowest scores σ_{sd} are below $\Delta\sigma_{sd}$.

6. Computation of a renormalised integral square error for each tile configuration in the library:

   A renormalisation procedure is introduced that gives different components in the integral square error comparable weight. It takes into account all scalings s between s_{min} and s_{max} and all tile configurations in the library.

7. Interpretation of renormalised integral square error plots:

   Tile configurations in the library are grouped according to the number of facets, and the renormalised score σ is then plotted against the number of facets. Minima in this plot correspond to the tile configurations in the library that best approximate the data. The smaller the number of facets, the coarser the approximation, and the smaller the score σ the better the approximation.

4.1.1 Surface representation of the Viral Capsid

Although viral capsids from [12] are invariant under icosahedral symmetry, the axes of symmetry are in general different from those we choose in our model. Therefore, all atoms in the corresponding pdb-files are first rotated such that the new axes of symmetry of the viral capsid align with those of the tilings.

Since our algorithm focuses only on the viral capsid, pdb-files which contain information on the position of the genetic material of the virus must have this information removed prior to analysis with our algorithm. From the modified pdb-file, the shape of the viral capsid is calculated as the Solvent Excluded Surface (or SES) of the proteins, the computation of which is performed using the PyMol software [65]. The output from PyMol is a mesh of points on the SES (depending on the radius of each of the atoms) that will in the next section be compared with the tilings.

For simplicity, we assume that all atoms are represented by spheres centered on the atomic positions (given by the pdb-file) whose radius is equal to the mean van der Waals radius computed over all types of atoms present in the protein shell. The type (for example: Carbon, Nitrogen, etc...) of each atom contained in the protein shell is also available in the pdb-file and the values of the van der Waals radii used to compute the mean radius have been taken from [10]. To reduce computation time the number of vertices representing the mesh is limited to one every 2Å. This subset of vertices, which we call $\mathcal{M}$, enables faster simulation while keeping the essential features of the capsid (see Figure 4.1). It is also consistent with the fact that the finest resolution of experimental data is usually higher than 3Å.

4.1.2 Description of the algorithm

Using the cut-and-project method we have constructed the 3 different tilings corresponding to the 3 lattices invariant under icosahedral symmetry in 6D (i.e. the SC, BCC and FCC lattices). We now need to embed the virus into the

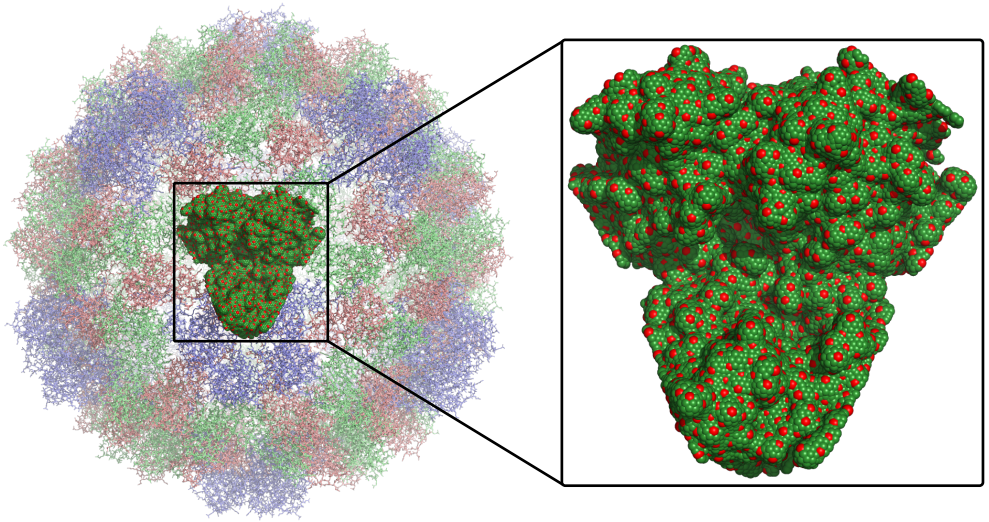


Figure 4.1: Illustration of the surface representation procedure described in 4.1.1 based on the example of the native form of Cowpea Chlorotic Mottle Virus (pdb-id: 1cwp [12]). The surface intersecting the fundamental domain, rendered by PyMol [65], is represented in green. On the right hand side a close up view representing the vertices of the un-reduced mesh (green spheres), is shown together with the reduced mesh (red spheres).

tilings allowing for different scalings. This is equivalent to finding the scaling for which the tiling (derived from SC, BCC or FCC lattices) best describes the capsid of a given virus. The orientation of the virus with regards to the tiling is fixed by the fact that both must share the axes of symmetry of the icosahedral group, so the only degree of freedom we have is an overall scaling factor.

The algorithm, which is described here, scores a selected set of tiles to the viral capsid to find the most appropriate scaling. In this section we will define the scaling of a tile and the selection rules used to pick the set of tiles which approximate the surface of the capsid's layout at a given scale. The ranking of these sets of tiles is made using the scoring system explained in subsequent subsections.

Using the same notation as in Chapter 2, a tiling will be called $\mathcal{T}^*$ regard-

less of it being projected from a SC, BCC or FCC lattice.

Selection of tiles

Here, we choose to leave the viral capsid unchanged and rescale only the tiling $\mathcal{T}^*$. Also, we choose the scaling $s = 1$ if the tiling corresponds to the one projected from the lattices given by the bases:

$$B_{SC} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad B_{BCC} = \frac{1}{2}\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix},$$

$$B_{FCC} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \tag{4.1}$$

We refer to these lattices as the "gauge" lattices to which all others are scaled. As the scaling increases or decreases, the size and the volume of each tile within $\mathcal{T}^*$ increases or decreases, respectively. The set of tiles overlapping with the protein shell may vary with the scaling. We next assume a fixed value of s and describe the rules which select the subset of tiles within a tiling $\mathcal{T}^*$ that will be matched to the viral capsid. The process is repeated for different values of s in a given range (whose boundaries are explained in detail in Section 4.1.4). Intuitively, for a tile to be selected it must contain at least a minimum number of atoms of the viral capsid. It is more convenient to work in terms of occupancy in terms of % of overlap rather than in terms of the number of atoms

within a tile, since tiles may have different sizes. For simplicity, we say that an atom is inside a tile if the center of the atom is inside the tile, bearing in mind that in our model each atom is represented as a sphere of fixed radius. This can lead to cases where the volume occupied by atoms within a tile is bigger than the volume of the tile (i.e. $\rho > 1$), but these cases occur primarily for small values of s. Overall, this simplification remains good enough while reducing the computation time to reasonable levels.

Definition 4.1.1. *Let V be the fixed volume of an atom and $t_i \in \mathcal{T}^*$ a tile, and define the occupancy of the tile t_i as:*

$$\rho_i = \frac{N_i V}{|t_i|} \tag{4.2}$$

where N_i is the number of atoms of the viral capsid whose center lies in the tile t_i, and $|t_i|$ the volume of t_i.

We fix a minimal threshold occupancy ρ as follows: a tile is taken into consideration by the algorithm if its occupancy satisfies $\rho_i \geq \rho$. The Quick Hull algorithm [7] is used for the computation of convex hulls, volumes and of the surface areas of the tiles.

Definition 4.1.2. *We call $\mathcal{S}_{\mathcal{T}^*}(s, \rho)$, the subset of tiles in $\mathcal{T}^*(s)$ with occupancy level above the minimal threshold ρ and ordered in increasing order of tile occupancy ρ_i, as:*

$$\mathcal{S}_{\mathcal{T}^*}(s, \rho) = \{t_i \in \mathcal{T}^*(s) : \rho_i \geq \rho \text{ and } \rho_i \leq \rho_j \text{ for } i \leq j\} \tag{4.3}$$

where ρ_i is the fraction of the volume occupied by the atoms in tile t_i, and $\mathcal{T}^*(s)$ is $\mathcal{T}^*$ rescaled by s.

Because the set of tiles in $\mathcal{S}_{\mathcal{T}^*}(s, \rho)$ can be the same for different values of ρ, we relabel $\mathcal{S}_{\mathcal{T}^*}(s, \rho)$ as $\mathcal{S}_{\mathcal{T}^*}(s, [\rho_{min}, \rho_{max}])$ where ρ_{min} and ρ_{max} are, respectively, the minimal and maximal ratio value for which $\mathcal{S}_{\mathcal{T}^*}(s, \rho_{min}) = \mathcal{S}_{\mathcal{T}^*}(s, \rho_{max}) = \mathcal{S}_{\mathcal{T}^*}(s, \rho)$ for every $\rho \in [\rho_{min}, \rho_{max}]$. Note that ρ_{max} can also

be thought of as the lowest value of ρ_i for every tile t_i in $\mathcal{S}_{\mathcal{T}^*}(s,\rho)$.
In addition, we define $\mathcal{A}_{\mathcal{T}^*}(s,\rho)$ the boundary surface of this set of tiles, defined as the surface given by the facets of the tiles in $\mathcal{S}_{\mathcal{T}^*}(s,\rho)$ which are not shared by any two tiles in this set. The set of facets, edges (intersection of two facets) and vertices (intersection of more than two facets) corresponding to $\mathcal{A}_{\mathcal{T}^*}(s,\rho)$ is called $\mathcal{B}_{\mathcal{T}^*}(s,\rho)$ and $|\mathcal{B}_{\mathcal{T}^*}(s,\rho)|$ denotes its cardinality. By the same reasoning as for $\mathcal{S}_{\mathcal{T}^*}(s,\rho)$, we can define $\mathcal{B}_{\mathcal{T}^*}(s,[\rho_{min},\rho_{max}]) = \mathcal{B}_{\mathcal{T}^*}(s,\rho)$ for every $\rho \in [\rho_{min},\rho_{max}]$.

To restrict choices for $\mathcal{S}_{\mathcal{T}^*}(s,\rho)$, we add two conditions:

R1 $\rho \geq 0.5$, i.e. we consider no set $\mathcal{S}_{\mathcal{T}^*}(s,\rho)$ in which any tile is less than 50% occupied;

R2 90% of the total main chain atoms of the capsid (i.e. the backbone given by C, C_α and N atoms to which the side chains are attached) must be within a distance of smaller than the experimental precision from the tiles in $\mathcal{S}_{\mathcal{T}^*}$, i.e. if data are given at xÅ resolution, the distance must be less than or equal to xÅ.

These additional limitations reduce the number of sets to be considered (and hence reduce computation time) to evaluate only good candidates for a fit. On the one hand, restriction R1 ensures that we consider only sets with tiles that are mostly fully occupied. An intuitive notion of a good match suggests that the surfaces of the tiles would be as close as possible to the vertices of the mesh $\mathcal{M}$. This is the case if all tiles selected have a high occupancy and very few atoms of the protein shell are in none of the selected tiles $\mathcal{S}_{\mathcal{T}^*}(s,\rho)$. Tiles with high occupancy (i.e. $\rho \sim 1$), should not be affected by the limitation on the minimal value of ρ. Instead $\rho \geq 0.5$ limits the possible choice of tiles selected nearer the outside of the protein shell. Accepting a tile if it is mostly full allows us to consider sets whose external surface is more likely to be nearer the mesh $\mathcal{M}$. The scoring system (described in subsequent sections) then

selects the set which best matches the capsid. On the other hand, restriction R2 forbids sets for which the protein shell is not adequately covered by the tiles. Without this restriction, sets covering a small portion of the viral capsid would be selected and scored. A good scoring system should penalise such configurations but allowing these sets and computing their scores would involve unnecessary computations.

The scoring system aims at finding sets $\mathcal{S}_{\mathcal{T}^*}(s, \rho)$ which provide good quality fits. To do this, sampling of the two variables s and ρ is necessary. Because the sampling of s is a more complex process, it will be explained separately in Section 4.1.6. If we assume a fixed scaling s, scores should be different for different values of ρ if and only if the corresponding sets $\mathcal{S}(s, \rho)$ are different. Notice that $\mathcal{T}^*(s) \supset \mathcal{S}(s, \rho) \supseteq \mathcal{S}(s, \rho')$ for $\rho' \geq \rho$. The values of ρ corresponding to different sets of tiles can easily be computed as follows:
Recall that for every tile t_i in $\mathcal{S}(s, \rho = 0.5)$, $\rho_i \leq \rho_j$ for $i < j$. Further notice that $\mathcal{S}(s, \rho = \rho_i) \supset \mathcal{S}(s, \rho = \rho_j)$ if $i < j$, and $\mathcal{S}(s, \rho) = \mathcal{S}(s, \rho = \rho_i)$ if $\rho \in (\rho_{i-1}, \rho_i]$. Choosing $\rho = \rho_i$ for $i \in [1, |\mathcal{S}(s, 0.5)|]$ generates all possible subsets of tiles with regards to ρ. If there exists i such that $\rho_i = \rho_{i+1}$ then $\mathcal{S}(s, \rho_i) = \mathcal{S}(s, \rho_{i+1})$ and the previous method would create duplicated sets. This is not a problem as the elements of $\{\mathcal{S}(s, \rho_i) : i \in [1, |\mathcal{S}(s, 0.5)|]\}$ can easily be made unique.

For example, suppose $\mathcal{S}(s, 0.5)$ is composed of three tiles t_1, t_2 and t_3 with respective ratios ρ_1, ρ_2 and ρ_3 such that $\rho_1 < \rho_2 < \rho_3$. Then three possible sets $\mathcal{S}(s, \rho)$ can be generated:
$\mathcal{S}(s, 0.5) = \mathcal{S}(s, \rho_1) = \{t_1, t_2, t_3\}$, $\mathcal{S}(s, \rho_2) = \{t_2, t_3\}$ and $\mathcal{S}(s, \rho_3) = \{t_3\}$. Note that in the above example, the set $\{t_1, t_3\}$ is not considered.

4.1.3 Scoring system

At this point finding the best set of tiles that approximates the surface of the protein shell can easily be related to a polygonal approximation problem [38, 51, 52, 58, 59, 62, 63, 85] where one polygon is approximated by another.

In our case, the polygon to be approximated is represented by the vertices $\mathcal{M}$ on the SES of the viral capsid, and the approximating polygon is defined by the boundary surface $\mathcal{A}(s,\rho)$ of $\mathcal{S}(s,\rho)$. In this section we describe the scoring system we use to quantify the approximation error.

We introduce some notation which will be needed. Let $v_i \in \mathcal{M}$ and denote by $P_{\mathcal{B}}(v_i)$ the projection of v_i onto $\mathcal{B}(s,\rho)$, i.e. the point of $\mathcal{B}(s,\rho)$ having minimal distance from v_i (assuming this is unique). Moreover, denote by d_i the minimal distance between v_i and $\mathcal{B}(s,\rho)$.

Definition 4.1.3. *The integral square error σ_{sd} is defined as:*

$$\sigma_{sd} = \sum_{\{i:v_i \in \mathcal{M}\}} d_i^2. \tag{4.4}$$

σ_{sd} is widely used in polygonal approximation problems [38, 51, 52, 58]. This score penalises each vertex in $\mathcal{M}$ with regards to its distance from $\mathcal{A}(s,\rho)$. The bigger the value of σ_{sd}, the poorer the quality of the approximation.

As the scaling s diminishes, the size of the tiles diminishes and more of them are needed to cover the viral capsid. As a result of this and conditions R1 and R2, σ_{sd} tends to decrease with s. An example is shown in Figure 4.5 and we observed the presence of local minima of σ_{sd}. These local minima reflect the fact that for a given set of tiles $\mathcal{S}$, there exists a scaling s for which σ_{sd} is minimal. Varying the scaling around s by a small amount will result in an increase of σ_{sd}.

In our case, however, for a simple example, such as the one presented in Figure 4.2, σ_{sd} would favour sets which do not match our intuitive notion of a good approximation of the shell's surface by the tiles. Therefore, we instead decompose σ_{sd} into two scores σ_1 and σ_2. Calling $\mathcal{M}_j$ the set of vertices whose projection belongs to the j^{th} element of $\mathcal{B}(s,\rho)$ and $D_j = \dfrac{1}{|\mathcal{M}_j|} \displaystyle\sum_{\{i:v_i \in \mathcal{M}_j\}} d_i$, the mean distance of the j^{th} element of $\mathcal{B}(s,\rho)$ to vertices of $\mathcal{M}_j$, we obtain:

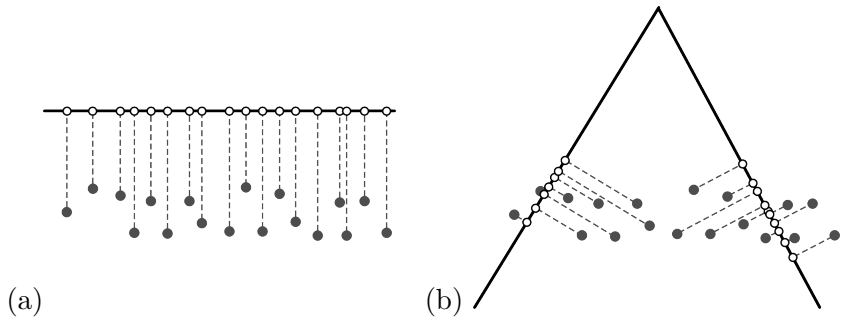


Figure 4.2: The vertices in $\mathcal{M}$ (grey dots) and their projections (empty circles) onto the facets of $\mathcal{A}(s,\rho)$ (black lines) are shown for two different tile selections in (a) and (b). The score σ_{sd} is higher in case (a) than in case (b), contradicting our intuitive notion of a good fit.

$$
\begin{aligned}
\sigma_{sd} &= \sum_{j=1}^{|\mathcal{B}(s,\rho)|} \sum_{\{i:v_i\in\mathcal{M}_j\}} d_i^2 \\
&= \sum_{j=1}^{|\mathcal{B}(s,\rho)|} \sum_{\{i:v_i\in\mathcal{M}_j\}} (d_i - D_j + D_j)^2 \\
&= \sum_{j=1}^{|\mathcal{B}(s,\rho)|} 2D_j \underbrace{\sum_{\{i:v_i\in\mathcal{M}_j\}} (d_i - D_j)}_{0} + \underbrace{\sum_{j=1}^{|\mathcal{B}(s,\rho)|} \sum_{\{i:v_i\in\mathcal{M}_j\}} D_j^2}_{\sigma_1} + \underbrace{\sum_{j=1}^{|\mathcal{B}(s,\rho)|} \sum_{\{i:v_i\in\mathcal{M}_j\}} (d_i - D_j)^2}_{\sigma_2}
\end{aligned}
$$

The definition of σ_1 and σ_2 is such that the lower the score, the better the approximation of the set of tiles to the viral capsid. σ_1 penalises sets for which the distance between the mean position of the approximated points in $\mathcal{M}_j$, and the corresponding element in $\mathcal{B}(s,\rho)$, is bigger. σ_2 mesures the difference between the distance of the vertices in $\mathcal{M}_j$ to the j^{th} element in $\mathcal{B}(s,\rho)$ and the mean distance D_j. It can be thought of as a score which quantifies the alignment of this boundary with the surface of the protein shell modelled by $\mathcal{M}$. In the case where all vertices in $\mathcal{M}_j$ are equidistant to the j^{th} element in $\mathcal{B}(s,\rho)$, the contribution of these vertices to σ_2 is null, see Figure 4.3.

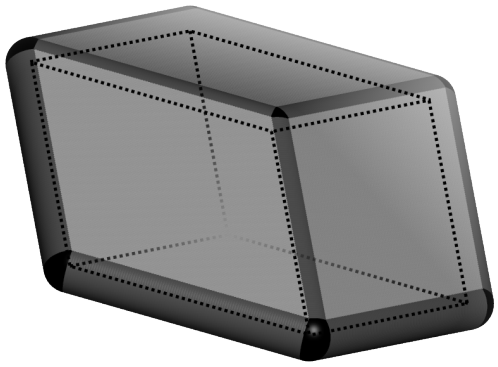


Figure 4.3: Here we assume an approximating polygon made up of a single tile whose edges are represented by dashed lines and where vertices of $\mathcal{M}$ are located on the greyed surface. σ_2 is minimised if the vertices projecting on a plane of $\mathcal{B}(s, \rho)$ are located on a plane parallel to this facet (light grey in the Figure). Similarly, vertices projecting on an edge of $\mathcal{B}(s, \rho)$ minimise σ_2 if they are on a cylinder with axis this edge (dark grey in the Figure) and vertices projecting on a vertex of $\mathcal{B}(s, \rho)$ are be located on a sphere centered on this vertex (black in the Figure).

The reason why σ_{sd} is not a suitable score is due to the fact that in general σ_1 is bigger than σ_2 and a renormalisation of the two scores must be applied in order to make them comparable. The calculation of the renormalisation factor (which will be discussed in Section 4.1.5) should only consider sets which are *plausible* approximations in order to save computation time. Restriction R1 already limits the possible values for the unknown ρ, while restriction R2 excludes sets of tiles which do not represent enough of the viral capsid. But the size of tiles also needs to be bounded as very small tiles can approximate any structure indiscriminately. To do this we also restrict the values the scalings can take.

4.1.4 Defining boundaries for the scaling value s

If the size of the tiles is too large compared to the size of the protein shell (for large values of s), then no information can be obtained about the structure of the viral capsid. In this case the approximation is considered too rough. In a similar way, tiles could be rescaled to a very small size (corresponding to small values of s), but we would gain no meaningful information by doing so. Hence, we need to restrict the values that s can take by defining a lower bound s_{min} and an upper bound, s_{max}, outside of which we consider $\mathcal{S}_{\mathcal{T}^*}(s, \rho)$ irrelevant to the study of the structure of the viral capsid. We discuss here the choices made to define these boundaries for s.

For this model to provide additional information with regards to the Caspar-Klug model, we require that the algorithm finds sets of tiles which give information on the thickness of the capsid. First let us consider the capsid as a container whose interior is empty. We require that *some* tiles within the capsid remain empty for any sampled value of s. For large scaling values, this is the case if at least the tiles in the vertex configuration of the origin O (see Figure 4.4) are within the protein shell. We then choose s_{max} as the maximal scaling such that no atoms of the capsid are within tiles with vertex O. In practice, however, the calculations for large values of s are fast and the scalings for which the tiles are too big to match the viral structure are penalised by the scores (especially by σ_1).

To estimate a lower bound for s, we use the score σ_{sd} and the precision on the experimental data.

Definition 4.1.4. *We call $\Delta\sigma_{sd}$, which is defined in terms of experimental precision ε, as:*

$$\Delta\sigma_{sd} = \varepsilon^2|\mathcal{M}| \tag{4.5}$$

the error on σ_{sd}.

If $\sigma_{sd} \leq \Delta\sigma_{sd}$ then the uncertainty on the experimental data does not allow us to differentiate the value of σ_{sd} from 0. In the case where $\Delta\sigma_{sd} = \sigma_{sd}$, each

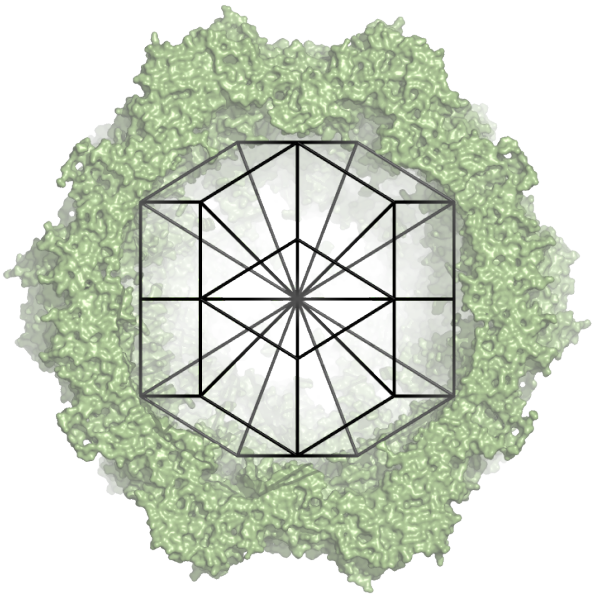


Figure 4.4: The maximal scaling s_{max} is defined as the maximal scaling for which the tiles of the vertex configuration of the origin O are empty. As an example we show the surface representation of the Minute Mice virus (pdb-id: *1mvm*) computed using PyMol software [65] with the edges of the tiles from the BCC tiling sharing the vertex O and scaled to s_{max} shown in black. For $s > s_{max}$ no sets $\mathcal{S}$ could provide information on the capsid's thickness.

vertex in $\mathcal{M}$ could be translated by ε onto the boundaries of the approximating polygon $\mathcal{B}(s, \rho)$. Remembering that globally σ_{sd} decreases when s decreases. We can therefore use $\Delta\sigma_{sd}$ to estimate the minimal value s_{min}.

Because we can assume that high quality fits may have a very low σ_{sd} (possibly lower than $\Delta\sigma_{sd}$), we cannot choose s_{min} as the highest value of s for which $\sigma_{sd} \leq \Delta\sigma_{sd}$. In particular, we need to check that the condition $\sigma_{sd} \leq \Delta\sigma_{sd}$ is not due to a local decrease of σ_{sd} near a possible good fit. If $\sigma_{sd}(s, \rho) \leq \Delta\sigma_{sd}$ for any value $\rho \geq 0.5$ and for $s \in [s_1, s_2]$ and if $[s_1, s_2]$ is wide enough to exclude a local effect, we choose $s_{min} = s_2$. An example is shown in Figure 4.5. To estimate the value of s_{min}, σ_{sd} has been computed for s varying by a fixed interval step Δs of 0.1.

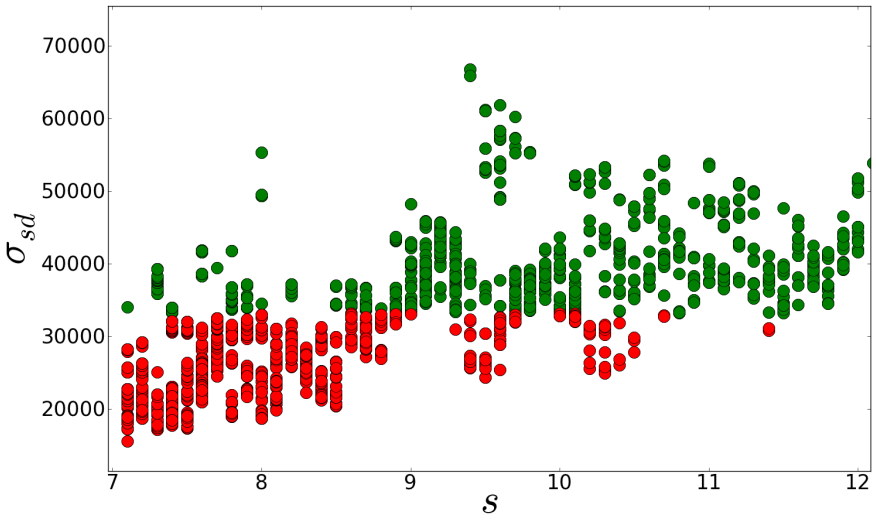


Figure 4.5: Determination of s_{min} for the matching of the BCC-tiling to the viral capsid of the Minute Mice Virus (pdb-id: *1mvm* [12]). The values of σ_{sd} are represented (y-axis) as a function of the scaling s (x-axis) for configurations $\mathcal{S}(s, \rho \geq 0.5)$. The different configurations due to variations of ρ are represented for each scaling and sets for which $\sigma_{sd} \leq \Delta\sigma_{sd}$ are colored in red. σ_{sd} displays local minima for scalings s=9.5, 10.3 and 11.4 which are candidates for good fits. For $s \leq 9.0$, all sets of tiles with lower σ_{sd} values fulfill $\sigma_{sd} \leq \Delta\sigma_{sd}$. The measure of σ_{sd} is then limited by the uncertainty of the experimental data. Therefore we choose $s_{min} = 9.0$.

4.1.5 Renormalisation

Now that suitable boundaries for the occupancy ρ and the scalings s have been defined, we can calculate the renormalisation factor r, i.e. $\sigma = r\sigma_1 + \sigma_2$, such that σ_1 and σ_2 have comparable weights. We will call σ the total score. To do this, after determining s_{min} and s_{max} (which are tiling-dependent), we select every possible configuration $\mathcal{S}(s_{min} \leq s \leq s_{max}, \rho)$ such that conditions R1 and R2 of Section 4.1.2 are fulfilled.

Definition 4.1.5. *We define* $\mathcal{R}(\gamma)$ *given by:*

$$\mathcal{R}(\gamma) := \sum_i (\gamma\sigma_{1,i} - \sigma_{2,i})^2 \tag{4.6}$$

where the index i *runs over all possible configurations* $\mathcal{S}(s_{min} \leq s \leq s_{max}, \rho \geq 0.5)$ *for any tiling.* $\sigma_{1,i}$ *and* $\sigma_{2,i}$ *correspond to* σ_1 *and* σ_2 *for the* i^{th} *configuration, respectively.*

$\mathcal{R}(\gamma)$ can be thought as the sum of squared distances between $\gamma\sigma_{1,i}$ and $\sigma_{2,i}$ over all configurations i. To minimise the difference between the contributions of $\gamma\sigma_1$ and σ_2 to σ, when considering the whole sets of configurations i, we minimise $\mathcal{R}(\gamma)$ with respect to γ.

Definition 4.1.6. *Keeping the same notations as in Definition 4.1.5, we call* r *such that* $\mathcal{R}(r)$ *is minimal, the renormalisation factor. Using* $\frac{\partial \mathcal{R}}{\partial \gamma}(r) = 0$*, we obtain:*

$$r := \frac{\sum_i \sigma_{1,i}\sigma_{2,i}}{\sum_i \sigma_{1,i}^2}. \tag{4.7}$$

$\mathcal{R}(r)$ *is minimal as* $\frac{\partial^2 \mathcal{R}}{(\partial\gamma)^2} \geq 0$.

Note that in the calculation of the renormalisation factor, the acceptable configurations considered come from either of the SC, BCC and FCC tilings such that there is a single factor r for each virus regardless of the tiling. This is to ensure that the contribution of σ_1 in the total score σ is the same regardless of the tiling. Once the renormalisation factor is computed, σ is determined for each acceptable (according to conditions R1 and R2 in Section 4.1.2) configuration of each tiling.

4.1.6 Exhaustive Sampling of s

Up to now we have sampled the scaling with a fixed interval of $\Delta s = 0.1$ to estimate, through the score σ, the approximation error of $\mathcal{S}$ as s varies.

However, $\Delta s = 0.1$ may not be a fine enough sampling to determine all possible sets $\mathcal{S}$ that can approximate the viral capsid. For large values of s, $\Delta s = 0.1$ is small enough for all possible sets $\mathcal{S}$ to be selected at least once, i.e. there is no scaling $s' \in [s, s + \Delta s]$ such that $\mathcal{S}(s', \rho') \neq \mathcal{S}(s, \rho)$ and $\mathcal{S}(s'\rho') \neq \mathcal{S}(s + \Delta s, \rho)$ for at least a pair of occupancy values ρ and ρ'. However, this is not true for lower values of s where the sizes of the tiles are small compared to the capsid. To prevent this, the scaling interval needs to be sampled such that any set $\mathcal{S}(s, \rho)$ that could obey conditions R1 and R2 for any value s within the sampled interval is scored at least once.

To discretise the scaling interval, notice that a change in the sets of tiles can only occur when an atom enters or exits a tile as s varies. All possible changes of $\mathcal{S}$ can then be listed by computing the sets of tiles satisfying R1 and R2 each time an atom "hits" a tile surface as s changes. Also note that the same tile set may be selected for different scalings, i.e $\mathcal{S}(s_i, \rho) = \mathcal{S}(s_j, \rho)$ for $s_i \neq s_j$. In this situation the scores σ associated with the two configurations $\mathcal{S}(s_i, \rho)$ and $\mathcal{S}(s_j, \rho)$ may be different. Computing the scores for each of these scalings is too time consuming so we choose to score the set with highest covering. We define covering in Section 4.1.2 as the ratio of main chain atoms whose distance to the tile set is less than the experimental precision).

Unfortunately the smaller the scaling, the smaller the size of tiles and the quicker the changes in $\mathcal{S}$ arise as s varies. For this reason we cannot sample s fine enough for the determination of s_{min} as the computation increases rapidly as s diminishes. Similarly the renormalisation factor r can only be estimated (using a sampling interval of $\Delta s = 0.1$) as an exhaustive sampling of the scaling over the range $[s_{min}, s_{max}]$ would be computationally too intensive. We assume that $\Delta s = 0.1$ is fine enough to provide us with a reasonable sampling of the sets $\mathcal{S}$. An exhaustive sampling of the variable s is then performed over smaller intervals in the vicinity of minima of σ to guarantee that no sets $\mathcal{S}$ which may correspond to good approximation of the capsid surface have been omitted.

4.1.7 Interpretations

In 2D polygonal approximation studies, two approximating polygons with the same number of edges are compared. In our 3D case, we compare different approximations if they have the same number of facets approximating $\mathcal{M}$ (i.e. $|\mathcal{A}|$), and represent σ as a function of $|\mathcal{A}|$. Since more than one configuration may exist for a single value of $|\mathcal{A}|$, we plot the one with lowest σ for clarity. For example, see Figure 5.1, 5.4, 5.6 and 5.10.
In these plots, we distinguish between two regions of $|\mathcal{A}|$:

1. For low values of $|\mathcal{A}|$ (i.e. when the sizes of the tiles in $\mathcal{S}$ are large) σ varies between high and low values as the change of tiles in $\mathcal{S}$ corresponds to large changes in the shape of the approximating polygon. In some cases, tiles overlapping with the protein shell represent such a poor approximation that condition R1 or R2 are not fulfilled, leaving values of $|\mathcal{A}|$ to which no sets $\mathcal{S}$ can be attributed.

2. For large values of $|\mathcal{A}|$ the size of the tiles compared with that of the capsid is small. In this case, conditions R1 and R2 are no longer too restrictive and a score σ is associated to each value of $|\mathcal{A}|$. As $|\mathcal{A}|$ increases, and when sufficiently many facets are used in the fit, smaller features of the capsid (such as protrusions) can be approximated. When these features can be fitted, σ decreases to a local minimum or a plateau.

Care must be taken when interpreting results for high $|\mathcal{A}|$-values. Indeed, given sufficiently many facets, the algorithm attempts to minimise the score by fitting smaller features including those related to the tertiary structures of the proteins. For this reason, coupled with the restricted resolution of the input data and the fact that some features may be subject to thermal fluctuation, the reliability of the results is reduced for high $|\mathcal{A}|$-values. We therefore only consider $|\mathcal{A}|$-values which result in a steep minima of σ as these indicate a region of interest despite these factors.

To find good quality fits, we refer to the two main problems addressed in polygonal approximation studies:

P1 Which configuration is the best fit given a fixed number of facets (referred to as the min-ε problem in [38])?

P2 What is the minimal number of facets needed to have a good match (referred to as the min-# problem in [38])?

Note that the case P1 is already dealt with since we consider only sets with minimal σ for a fixed $|\mathcal{A}|$-value. To address P2, we would need to arbitrarily choose a maximum tolerance for σ. However, σ depends on the number of vertices in $\mathcal{M}$, making it difficult to define a maximum tolerance.

Ideally, we would prefer a configuration whose corresponding σ and $|\mathcal{A}|$-values are both low. Because we cannot define the maximum tolerance for σ, we analyse minima of σ with increasing values of $|\mathcal{A}|$. If the approximation matches the main surface features, we consider no finer fits. In the case of high values for $|\mathcal{A}|$ corresponding to a minimum of σ, more than one set $\mathcal{S}$ may represent a good match to the capsid shell. As the difference between these configurations is difficult to judge, we will refer to them as a *class* of approximations.

As by definition the renormalisation factor r is independent of the type of tiling (i.e. σ_1 is equally weighted in the three cases), we represent the results corresponding to the three different tilings on the same plot for comparison. See Figures 5.1, 5.4, 5.6 and 5.10 for examples. A tiling is preferred over another if its approximation of the capsid surface has lower σ for equal or lower $|\mathcal{A}|$.

To clarify these concepts, we explain the results obtained for a selection of $T = 1$ and $T = 3$ viral capsids in the following chapter.

Chapter 5

Application to viruses structure

In this chapter, we apply the matching algorithm described in the previous chapter to a range of $T = 1$ and $T = 3$ viral capsids. The present approach will be compared with the icosahedrally invariant point array approach [37] described in the introductory chapter. For all plots, we keep the same color code to distinguish between sets from different tilings and choose to represent scores corresponding to SC, BCC and FCC tilings as blue, red and green dots, respectively, on the figures. An exhaustive search of tile sets $\mathcal{S}$ corresponding to local minima of σ has been carried out and the results are included in the plots. When representing the fits, edges of the tiles are pictured as black lines and, for $T = 3$ viruses, the different chain types are represented in different colors.

5.1 Pariacoto Virus (PaV)

The PaV is a $T = 3$ virus that infects insects. The 180 coat proteins cluster into 60 trimers with protrusions, which are about 15 Å high along the quasi-three-fold axis [76]. These protrusions, see Figure 5.1 (a), are believed to be involved in host-cell recognition and are therefore important features which

need to be included into our approximation. The tile matching algorithm was run on the pdb-file (pdb-id: 1f8v) from [12]. Because the matching of tile is performed only on the viral capsid, information on the genetic material available in the pdb-file has been removed prior analysis. Results are plotted in Figure 5.1 (b).

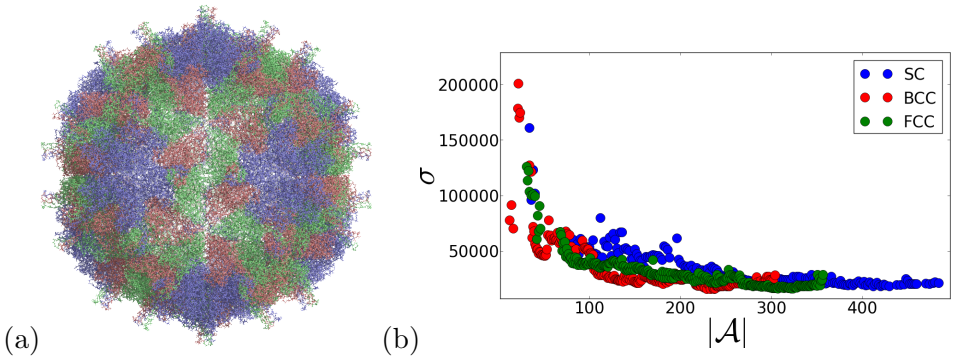
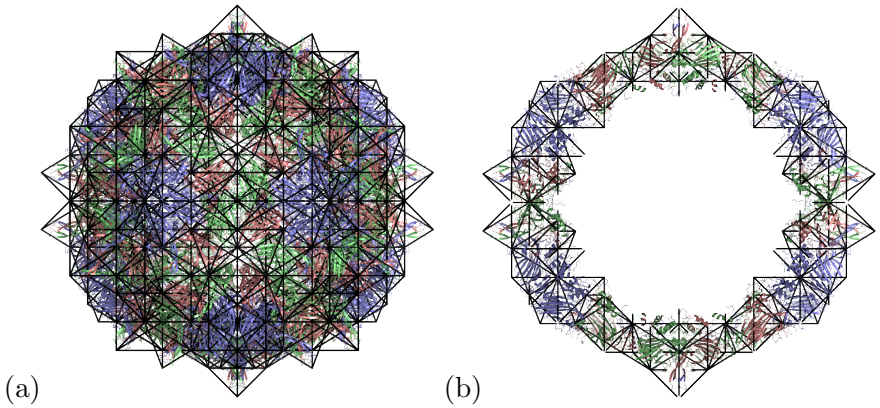


Figure 5.1: The tile matching algorithm is applied to the PaV's capsid shown in (a) with proteins labelled as "A", "B" and "C" chains in pdb-file colored in blue, red and green, respectively. The output of the algorithm for each of the three tilings is shown in (b) and a renormalisation factor r_{1f8v} of about 0.369 between the two scores σ_1 and σ_2 as explained in the Chapter 4 Definition 4.1.6.

First note that the BCC tiling provides a better match for equal or lower values of $|\mathcal{A}|$ up to $|\mathcal{A}| = 74$. The first local minimum of the BCC tiling is located at $|\mathcal{A}| = 16$, but the corresponding approximation by the tiles is too coarse, i.e. the outer and inner capsid surfaces are not well represented by the tile set. To match the protrusions, whose size is small when compared to the size of the virus (with an outer radius $\sim$ 175 Å), a finer fit is needed. The second minimum of σ in the BCC case belongs to the interval $|\mathcal{A}| \in [40, 52]$. Matches corresponding to $|\mathcal{A}| \in [41, 52]$, with exception of $|\mathcal{A}| = 40$ or 44 (which do not fit the protrusions), are visually very similar and because the difference between these configurations is difficult to appreciate, we refer to

them as a *class of approximations*. We choose to represent $|\mathcal{A}| = 41$ in Figure 5.2 since it is the simplest representation of this class of BCC matches. For

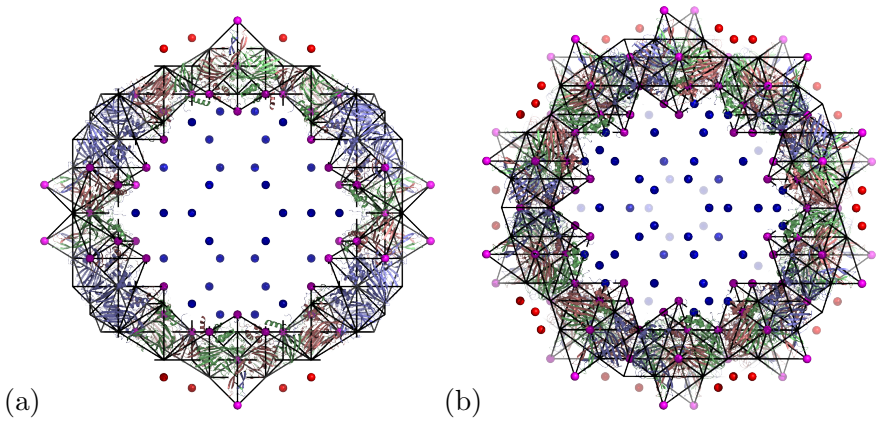


(a) (b)

Figure 5.2: Best match for the Pariacoto virus with a front view along the two-fold axis (a) and a slice of 50 Å along the same axis (b).

this viral capsid we consider no finer fits as much higher values of $|\mathcal{A}|$ are required for FCC or BCC tilings to display lower σ values.

We can compare our approach with the point array approach developed in [37] by analysing their respective results when fitting PaV's viral capsid. For clarity, we will call $\mathcal{P}$ the point array obtained in [37] that best matches PaV and $\mathcal{S}_{PaV}$ the tile subset corresponding to $|\mathcal{A}| = 41$ obtained by fitting tiles to PaV. Recall from Chapter 1 that a library of 569 point arrays has been created using affine extension of the icosahedral group. After rescaling of these vertex sets such that the outermost vertices match the protrusions of PaV, an RMSD-based score measuring the distance between vertices and the surface capsid has been used to rank these point arrays according to the approximation error (which is scored according to its proximity to atomic position in the viral capsid, see [37] for more details). First $\mathcal{P}$ and the tile subset $\mathcal{S}_{PaV}$ are both subsets of a BCC-tiling. Once scaled to the PaV capsid, $\mathcal{P}$ can be embedded into the gauge BCC tiling (defined in Chapter 4 as the BCC tiling with scaling factor $s = 1$) rescaled by a factor $s = 15.51$ whereas the scaling factor obtained by fitting $\mathcal{S}_{PaV}$ is $s = 16.05$. This small difference



(a) (b)

Figure 5.3: We compare the point array $\mathcal{P}$ obtained in [37] with our selected tile set $\mathcal{S}_{PaV}$ using two cross-sectional views: (a) a 50Å thick slice along a two-fold axis is shown and (b) a 100Å thick slice along a five-fold axis. For clarity vertices of $\mathcal{P}$ are displayed with a 4Å radius and are rescaled to match the vertices of the tile set $\mathcal{S}_{PaV}$. We colored in red, vertices contributing to a poorer match. All vertices in $\mathcal{P}$ matching the capsid features (in magenta) are included within the tile set selected by our algorithm. Whereas vertices which contribute to a lower approximation error (in red) and those (in blue) providing geometrical constraints on the genomic material (not shown), are not included within $\mathcal{S}_{PaV}$.

in the scalings of the BCC tilings corresponding to $\mathcal{P}$ and $\mathcal{S}_{PaV}$ can easily be associated with the difference in the matching algorithms, i.e. minimising the capsid to vertices distance (in the point array approach) as opposed to minimising the capsid to tile subset distance (in the current study). As can be shown in Figure 5.3 the two approaches converge toward similar embedding of the PaV's capsid.

5.2 Carnation Mottle Virus (CarV)

Carnation Mottle virus is a $T = 3$ virus responsible for mild mottling and chlorosis (i.e. whitening of green plant tissues due to choropyllic deficiency) in

carnation crops. The pdb file (pdb-id: 1opo) which displays protrusions near the two-fold axes, see Figure 5.4 (a), was matched against the three types of tilings and the results are shown in Figure 5.4(b).

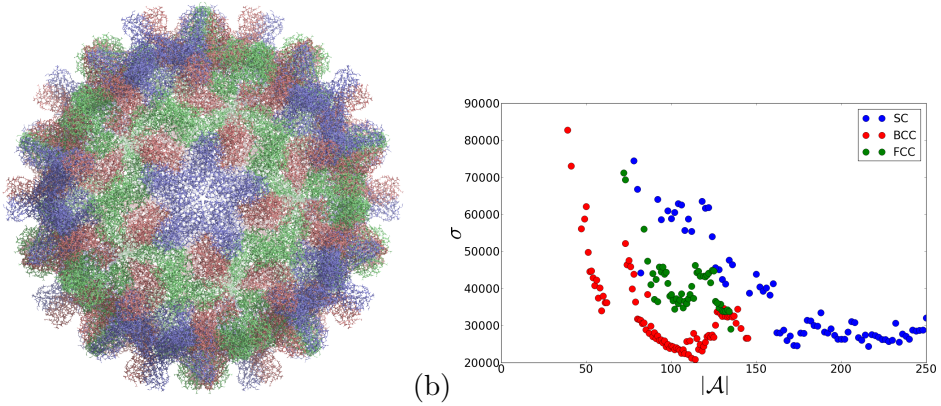


Figure 5.4: The tile matching algorithm is applied to the CarV capsid shown in (a) using the same color coding as in Figure 5.1(a). The output for each of the three tilings is shown in (b) with a renormalisation factor r_{1opo} of about 0.362.

As in the case of Pariacoto virus, the BCC-tiling offers better fits than the SC or FCC-tilings for equal or lower values of $\mathcal{A}$. Also, the first minimum at $|\mathcal{A}| = 59$ provides only a poor description of the inner capsid surface and of the protrusion contour. We therefore require a finer approximation to better match the capsid contour. A steep decrease in σ for $|\mathcal{A}| \in [73, 80]$ suggests that smaller features of the viral capsid are matched. After this σ decreases almost linearly over $|\mathcal{A}| \in [80, 114]$. A closer analysis reveals that for $|\mathcal{A}| \in [80, 101]$ the set of approximating tiles is indeed very similar and the difference between them is mostly due to fine-tuning of the variables s and ρ. From $|\mathcal{A}| = 102$ to 114 fits of the protrusions are less good, and we suspect that the lower score is solely due to the high number of facets scoring the protein tertiary structures. Since $|\mathcal{A}| = 80$ corresponds to the configuration for which the decrease of σ is the most significant, we choose this as the simplest representation of the essential features of the viral capsid (see Figure 5.5).

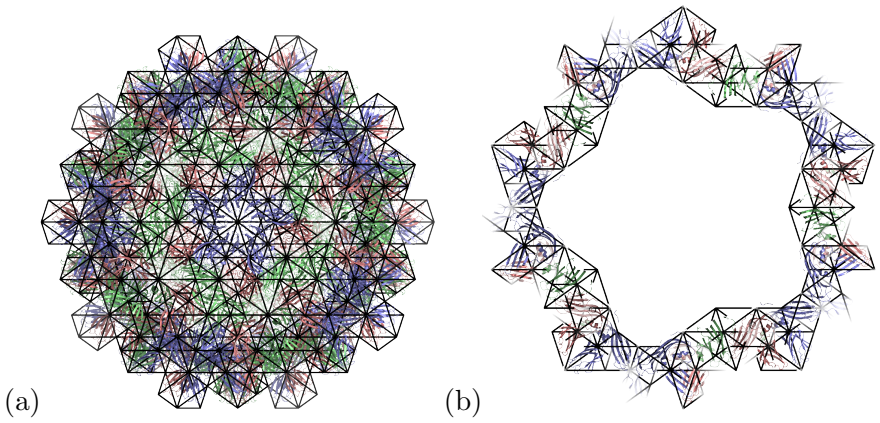


Figure 5.5: Best match for the Carnation virus with a front view along the five-fold axis (a) and a 30Å thick slice along the same axis (b).

5.3 Cowpea Chlorotic Mottle Virus (CCMV)

This $T = 3$ virus can undergo a reversible pH-dependent conformational change. For $pH > 6.5$ the expansion (or "swelling") of the viral capsid is accompagnied by the formation of openings at the pseudo three-fold axis and allows molecular exchange between the capsid cavity and the external medium. When the pH is reduced below 6.5, these openings close and no further exchange has been observed. Using this property, crystals of paratungstate ions have succesfully been packaged in CCMV capsids [19]. Subsequent modification of the outer surface proteins could provide specific biological targeting and make the CCMV capsid a good vector for drug targeting and delivery.

As in CarV and PaV, the BCC-tiling presents lower scores than FCC and SC tilings as shown in Figure 5.6(b). Note that no fits have been scored for $|\mathcal{A}| \in [45, 53]$ and that the total score σ fluctuates greatly for $|\mathcal{A}| < 45$, indicating that the size of tiles is large compared to the size of the capsid. However, the bulky structure of the CCMV capsid, see Figure 5.6 (a), is well approximated by fits corresponding to $|\mathcal{A}| = 35, 37$. These two configurations differ very slightly in both variables s and ρ and the exchange of a single tile relates the two sets of tiles. The configuration for $|\mathcal{A}| = 37$ is represented in

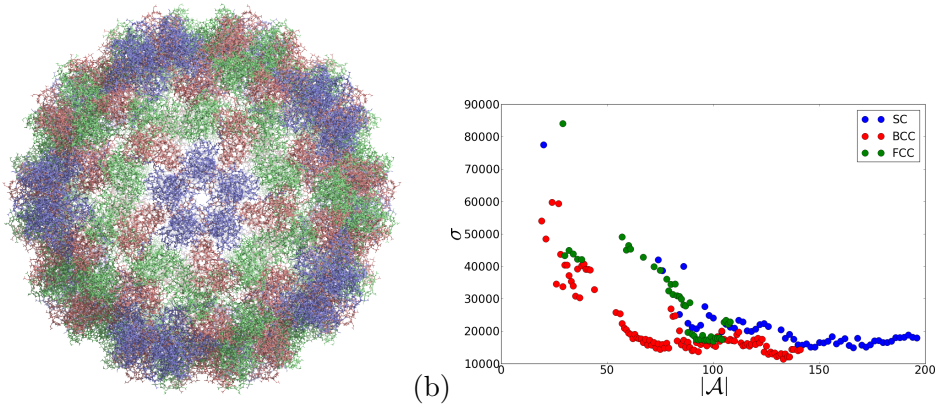


Figure 5.6: The matching algorithm is applied to the CCMV capsid shown in (a) using the same color coding as in Figure 5.1(a). The output for each of the three tilings is shown in (b) with a renormalisation factor r_{1cwp} of about 0.392.

Figure 5.7. As the thickness of the capsid shell, as well as the structure of

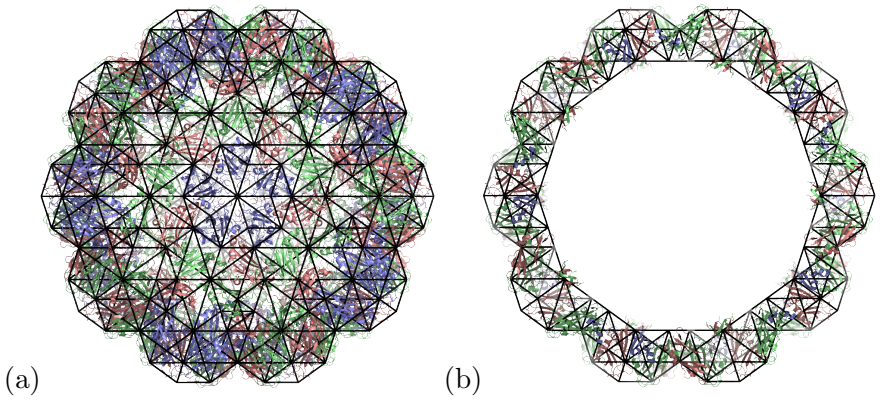


Figure 5.7: Best match for CCMV capsid with a front view along a five-fold axis (a) and a slice of 60Å along the same axis (b).

hexamers and pentamers, are well approximated within the BCC tile set no finer configuration is considered.

5.4 Physalis Mottle virus (PhyMV)

PhyMV is another T=3 plant infecting virus. The pdb-file (pdb-id: 1e57) used in the tile matching alogrithm was created from X-ray crystallography with capsids already devoid of genomic material [70]. The tile matching algorithm output is plotted in Figure 5.8(b).

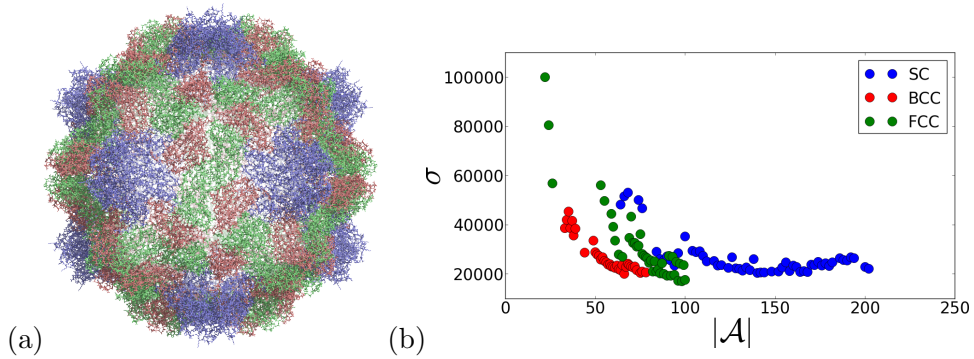
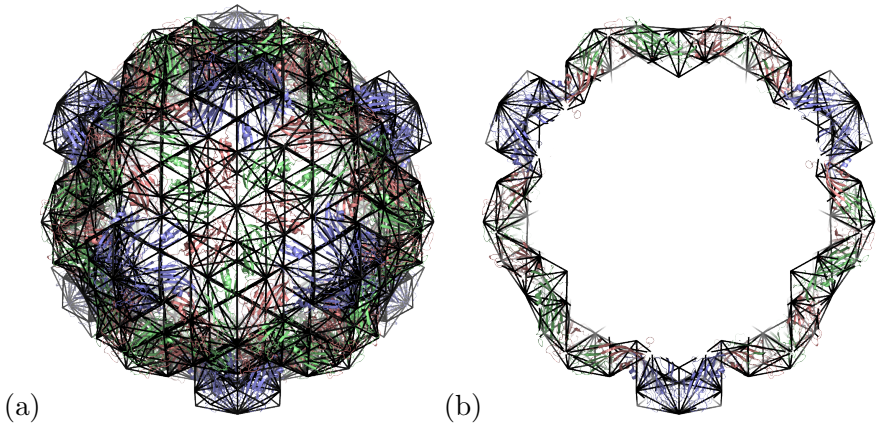


Figure 5.8: Application of the tile matching algorithm to the PhyMV capsid shown in (a) in a frontal view along the two-fold axis. The output of the algorithm for each of the three tilings is shown in (b) with a renormalisation factor r_{1e57} of about 0.415.

For this capsid shell the BCC tiling provides better fits for coarser approximations. For $|\mathcal{A}| < 50$, configurations provide poor matchings of the hexamers outer surface, and we therefore look for finer approximations. $|\mathcal{A}| \in [50, 66]$ correspond to fits which share a common subset of tiles fitting the bulk of the hexamers. The main differences between these fits correspond to two different fittings of the pentamer's outer surface. Our procedure cannot distinghuish between these two approximations of the outer surfaces. This may be due to the fact that the spikes on top of the pentamer's surface could be subject to fluctuations. Note that no information was available from the pdb-file, i.e. the temperature factor available in the pdb-file has been set to a constant value of 15.00. However, as $|\mathcal{A}|$ increases, i.e. as the approximation is made finer, one of the two approximations appears consistently. We therefore dis-

play the simplest representation (corresponding to $|\mathcal{A}| = 53$) of this class of approximations in Figure 5.9.



(a) (b)

Figure 5.9: Best match for PhyMV with a front view along a three-fold axis (a) and a 65Å thick slice along the same axis (b).

5.5 Minute Mice Virus (MVM)

This $T = 1$ virus replicates in cells which are undergoing division and is responsible for behaviour modifications of biological systems where cell multiplication is important, such as in cancer research studies. The pdb-file from [12] (pdb-id:1mvm) includes information on the RNA of the virus with chain identifiers "R" and "S", that is removed prior to analysis.

As can be seen from the results in Figure 5.10(b), the minimal score σ corresponds to configurations from BCC tilings (at $|\mathcal{A}| = 14$) then SC tilings (at $|\mathcal{A}| = 26$) and finally for FCC tilings (for $|\mathcal{A}| \in [30, 38]$). Over $|\mathcal{A}| \in [30, 38]$, the prefered subset of FCC tiling oscillates in turn between two classes of approximations. As such $|\mathcal{A}| = 33, 35$ and 37 correspond to one class and $|\mathcal{A}| = 30, 32, 34, 36, 38$ (no configurations with $|\mathcal{A}| = 31$ have been found) to another. Since each representative of the first class has higher values of σ over the same interval $|\mathcal{A}|$, we only consider the second. As an example we show, in Figure 5.11, a representative of tile sets corresponding to each of the

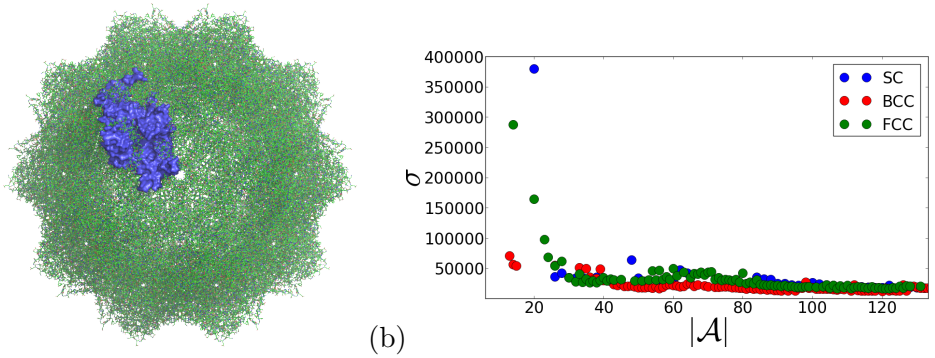


Figure 5.10: Application of the matching algorithm to the MVM capsid. This $T = 1$ capsid is shown in (a) with one protein subunit represented in blue. The output of the algorithm for each of the three tilings is shown in (b) with a renormalisation factor r_{1mvm} of about 0.324.

SC, BCC and FCC minima. From Figure 5.11, it is clear that the FCC tiling represents a better approximation to the viral shell. This is also reflected in the scores. Since the difference between the configurations of the FCC class for $|\mathcal{A}| = 30, 32, 34, 36$ and 38 mainly consists of a fine-tuning of the scaling of the tiles and that the size of these tiles is large compared to the capsid, we are confident that the decrease of σ is not dominated by the fit of the capsids tertiary features. Therefore we chose to represent the configuration with lowest score, i.e. $|\mathcal{A}| = 36$ in Figure 5.12. As the contours of the inner and outer surfaces of the viral capsid are well approximated by the FCC tile set (including the β-barrel involved in host-cell recognition [48]), there is no need to consider finer configurations.

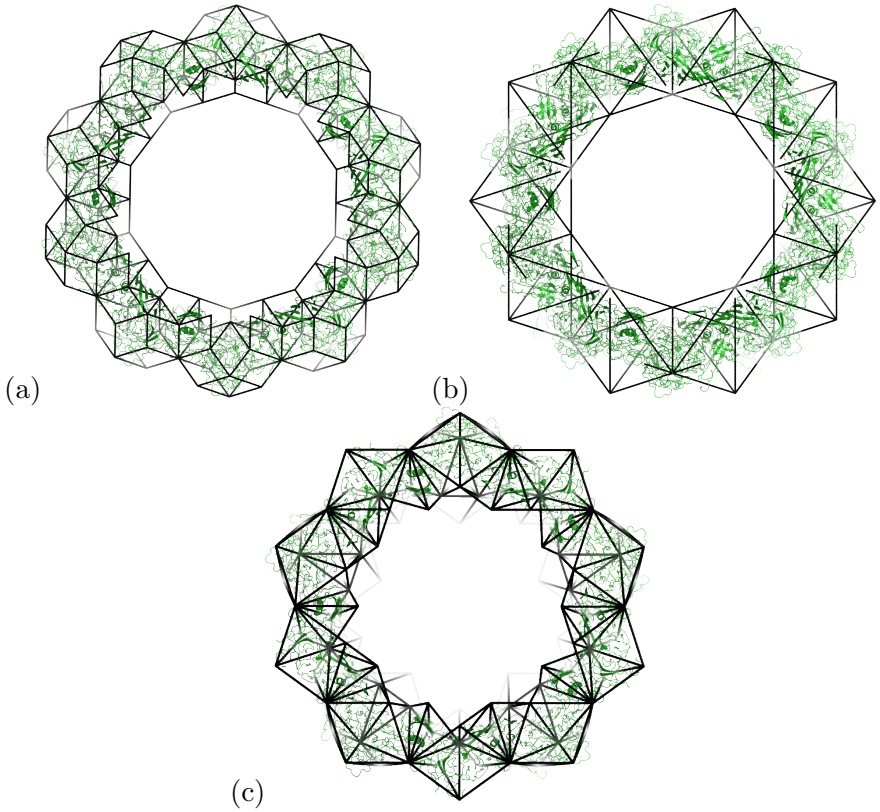
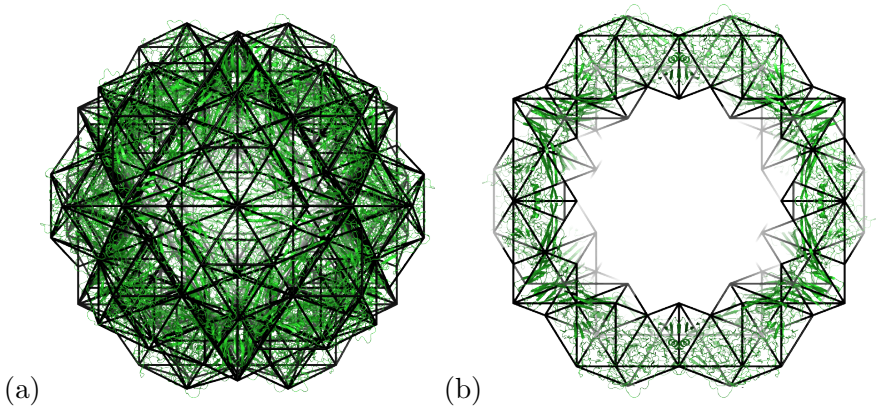


Figure 5.11: Comparison of the configurations corresponding to the first minima of σ for the three tilings when matching the MVM capsid using cross-sectional views along the five-fold axis. For the SC tiling (50Å cross-sectional view shown in (a)), the minimum is located at $|\mathcal{A}| = 26$ with $\sigma \sim 35697$, BCC has a minimum at $|\mathcal{A}| = 14$ with $\sigma \sim 56391$ (70Å cross-sectional view shown in (b)), and FCC has a minimum at $|\mathcal{A}| = 36$ with $\sigma \sim 26275$ (40Å cross-sectional view shown in (c)). Despite corresponding to minima of the approximation error σ, the tile sets obtained by matching the SC or the BCC tilings do not offer good approximations of the inner and outer surfaces of the MVM capsid whereas the FCC tiling provides a better approximation of the capsid layout.



(a) (b)

Figure 5.12: Best match for the Minute Mice virus with a front view along the two-fold axis (a) and a slice of 45Å along the same axis (b).

5.6 Bacteriophage $\alpha 3$

Fitting of the $T = 1$ Bacteriophage $\alpha 3$ capsid to icosahedral tilings was performed using the pdb-file with pdb-id:1m06 created from X-ray crystallography in [8]. The structure is mostly spherical save the G-proteins positioned at each of the five-fold axes as shown in Figure 5.13(a).

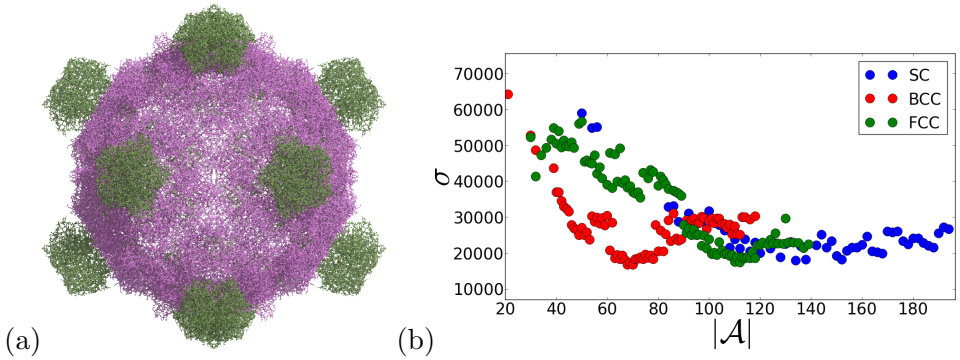


Figure 5.13: The $T = 1$ bacteriophage $\alpha 3$ capsid is shown along a two-fold axis in (a). The output of the algorithm for each of the three tilings is shown in (b) with a renormalisation factor r_{1m06} of about 0.360.

The fit of Bacteriophage $\alpha 3$ is more subtle. The configurations within the first minima, i.e. $|\mathcal{A}| \in [48, 53]$ correspond to very similar approximations (i.e. each tile set contains about 29 tiles and change by only one or two tiles as $|\mathcal{A}|$ varies) enclosing forms for the viral capsid and therefore constitute a single class of approximations. The configuration corresponding to $|\mathcal{A}| = 53$ is shown in Figure 5.14(a) and (b). For $|\mathcal{A}| = 53$, the protrusions (labelled as G chains in the pdb-file) are not very well approximated and we can assume that the minisation of σ is due to a rather good approximation of the capsid core (i.e. no G chains). A finer fit to the G-proteins may be found with higher values of $|\mathcal{A}|$, but then the fit of the capsid's core is less good. A good compromise has been found for $|\mathcal{A}| = 62$ (see Figure 5.14(c) and (d)). However the σ score associated with configuration $|\mathcal{A}| = 62$ is higher than the one with $|\mathcal{A}| = 53$ despite fits on G-proteins being finer and having similar tiles fitting the capsid's

core. The tile sets corresponding to $|\mathcal{A}| = 53$ and $|\mathcal{A}| = 62$ are both shown in Figure 5.14 for comparison. The difference in scores can be explained by the different scalings of the two tile sets corresponding to $|\mathcal{A}| = 53$ and $|\mathcal{A}| = 62$, respectively $s \sim 14.09$ and $s' \sim 13.32$. For $|\mathcal{A}| = 53$, σ is minimised due to a better fit of the capsid core. As the scaling decreases towards configuration $|\mathcal{A}| = 62$, the σ_1 component (i.e. the measure of the capsid-to-tiles mean distance) of σ increases. In this case a better fitting of the protrusions does not compensate for a poorer fit of the overall capsid.

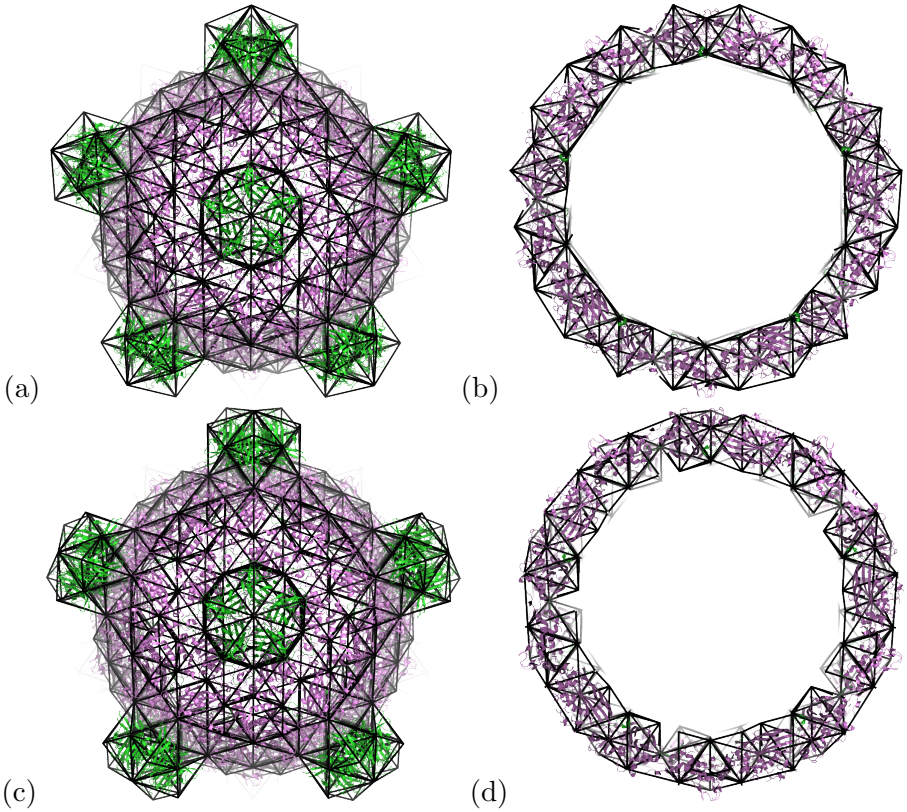


Figure 5.14: Comparison of tile sets matching the Bacteriophage $\alpha 3$ capsid corresponding to $|\mathcal{A}| = 53$ (first row) and $|\mathcal{A}| = 62$ (second row) along a five-fold axis. A 50Å thick cross-sectional view along the same axis in (b) and (d) demonstrates that a subset of tiles common to the two configurations are used to fit the capsid core despite displaying different outer surfaces approximating the G-chains (shown in green in the front views along five-fold axes (a) and (c)).

5.7 Discussion

The work presented here extends previous mathematical models for the determination of viral structures. Starting with the prediction and classification of the possible two dimensional layouts for protein capsids [13], it was shown in [37] that the use of icosahedrally invariant point arrays created from affine extensions of the icosahedral symmetry group provides geometrical contraints for virus structure at different radial levels. In this study we have extended this approach further, and have constructed tilings of which these point arrays are part, to better understand the role of symmetry in virus structure.

Because these icosahedral tilings fill the 3D space and provide more geometrical restrictions (such as the edges and faces of the tiles), a new algorithm as been devised here to match them to protein shells. Since the orientations of the tilings and the capsids are fixed by the axes of the icosahedral group, the set of tiles approximating the capsid shell is determined by two unknowns: the scaling s of the tiling and the minimal occupancy ρ for each tile in the set. Upper and lower bounds for these two variables have been defined, taking experimental precision into account, and the sampling of these have been carried out such that all sets of tiles which may provide a suitable fit are considered.

A scoring system, derived from polygonal approximation studies, has been implemented to locate good fits. Although the optimal solution can be found in 2D polygonal approximation it is computationally intensive and near-optimal algorithms are being developed instead to reduce computations. Here the 3-dimensional character of the problem adds further complexity and the search of an optimal is beyond the scope of our IT resources. Another problem, inherited from polygonal approximation studies, is the definition of a "low enough" approximation error, i.e.: "How fine an approximation needs to be to represent the main capsid features?". Unfortunately this question remains unanswered as structural features involved in host-cell recognition may require, due to their size, a very fine approximation. Finally capsids are dynamical systems with structural features more or less subject to thermal fluctuations.

Depending on the experimental measurement, these fluctuations may not be readily accessible in pdb-files.

Despite these difficulties we have shown that conditions on the tile set selection and minimisation of the approximation error σ suggested the intervals of $|\mathcal{A}|$ which may provide good matches using a selection of $T = 1$ and 3 capsids as tutorial examples. Moreover we have shown that our approach and the one developed in [37] converge towards similar results, using PaV's capsid as an example the two matching algorithms embed the viral capsid into BCC tilings of size very alike.

The objection has been raised that, given sufficient resolution on the experimental data (allowing a small value of s_{min}), all features of the capsid could be matched well enough with sufficiently small tiles. We therefore need to clarify under which circumstances embedding viral capsids into tilings is relevant. As can be seen for viruses with sufficiently high experimental resolution (for which a smaller value s_{min} is allowed), high values of $|\mathcal{A}|$ correspond to a plateau (or a shallow slope) of σ. This is particularly noticeable, for example, on Figure 5.1(b). Under these conditions the error approximation σ is practically the same independently of the $|\mathcal{A}|$-values and it is clear that the approximation of the viral capsid is too fine, i.e. any tile set is as good as any other. However note that in all cases studied explained above, the preferred fit always corresponds to a coarser approximation (lower values of $|\mathcal{A}|$) and in most cases corresponds to local minima of σ such that the plateau of σ is not reached. In the case of PhMV no match, corresponding to local minima of σ, provided a reasonable approximation of the pentamers and hexamers. The preferred match (corresponding to $|\mathcal{A}| = 53$) was therefore chosen as the coarser approximation fitting sufficiently well the capsid before the plateau of σ (observed for $|\mathcal{A}|$ approaching 60 and higher).

Conclusions

As the structure of more and more viral capsids are determined experimentally, it has become clear that symmetry plays a major role in capsid structure. The aim of this work is to develop mathematical methods to describe the full 3D structure of these protein shells and show that their structural features (such as capsid size and thickness, and the location of protrusions etc) are in fact related to each other.

Janner was the first to use 3D lattices to study the crystal structures of viruses. Although the essential features of these crystals (packing and symmetry) could be modeled, the icosahedral nature of viral capsids requires the use of non-crystallographic symmetry to understand the geometrical constraints imposed on the shell structure. As opposed to Caspar-Klug theory, the creation of icosahedrally invariant point arrays via affine extension of the underlying symmetry group and their mapping to viral capsids in [37] has shown that the geometrical constraints not only correlate positioning of proteins in the capsid, but also relate structural features at different radial levels including genome organisation.

In this study we have extended the point array approach using quasilattice theory to demonstrate that quasilattices provide additional structure constraints on virus architecture. We have used this to derive the most likely transition paths taken by the protein shell during the structural transformation. Our model does not involve any chemical or physical interactions between the atoms of the capsid, but it is not suggested that these are not important for understanding the shell's structure. However, we show that these inter-

actions are under strong geometric constraints restricting the possible ways these can affect the capsid structure.

In Chapter 3 we have used the point arrays as descriptors of virus structure to represent pre- and post-transitional state of viral capsids. We have lifted these non-crystallographic 3D constraints in 6D icosahedral lattices, and have used crystallographic techniques to compute the transitions with minimal symmetry loss between these 6D lattices. Via projection into 3D we have then induced transitions between the corresponding quasilattices. By inference this method allows us to determine the maximal symmetry that can be conserved by the intermediate configurations. For example, application of this method to CCMV suggests that the structure of the capsid has at most D_3 symmetry throughout the transition. Note that the symmetry analysis used here cannot predict the exact transition path, as this would require a complete analysis of the energy landscape. Instead, using symmetry considerations it is possible to consider only the most favourable subspaces and derive statements on the most likely symmetry properties of the intermediate states.

Each point array from [35] is in fact a subset of the vertex set of one of three quasilattices. We have constructed these quasilattices by projecting from higher dimensional lattices. These provide in addition the elements, such as the complete vertex set, edges and tile surfaces, that the point array descriptors lack. To investigate further the geometrical constraints on viral capsids imposed by these new elements, an algorithm has been developed in Chapter 4 to match the viral capsid to patches of quasilattices. The algorithm depends on two parameters that capture the relative size of the tiling to the viral capsid, and the minimal occupancy of tiles by atomic positions in the capsid proteins, respectively. An exhaustive search of the parameter space has been performed and a scoring system (based on polygonal approximation studies) used to measure the approximation error of the capsid surface by the

tile sets.

In Chapter 5, we have applied this algorithm to the quaternary structures of a range of test viruses. Scoring each of the possible tile sets necessitates intensive computation and the quality of the fitting is more easily observable for viral capsids whose structural features (such as protrusions) have a size comparable to the capsid itself. Bigger T-number capsids already involve more computation due to their size and often display relatively smaller protrusions requiring finer approximations. For both these reasons we have restricted our study cases to $T = 1$ and $T = 3$ viruses, although the same algorithm is in principle also applicable to viruses with higher T numbers. Note that, we have neglected changes in structure due to thermal fluctuation, but such information is not always readily available in pdb-files (via b-factor). However, we expect that experimental precision, which is included in our model, would be poorer for highly mobile features, which should be borne in mind when interpreting the results. Despite this our results provide valuable new insights into structural constraints on virus architecture can serve as a coarse-grained approximations of virus architecture.

Outcome of the model

Our technique permits a finer classification of viral capsids than that provided by Caspar-Klug theory (i.e. in terms of T-numbers) as the tile matching algorithm may associate different tile sets to two viruses with same the T-number. Our model also provides finer icosahedrally invariant enclosing forms than those used by Janner. Indeed, only two triacontrahedra (one scaled to the inner capsid surface and a τ-rescaled one for the outer surface) are used in [32] as enclosing forms for the $T = 3$ Pariacoto and Cowpea Chlorotic Mottle viruses. We have shown in Chapter 5 that two different tile sets match these two viruses better, hence distinguishing between them.

In our approach, for two viruses to be mapped to the same tile set would

require that, they have similar outer and inner capsid surfaces modulo some scaling factor, that is: the position of protrusions should be the same with regards to the icosahedral symmetry axes; the sizes of the protrusions relative to the size of the capsid should be the same; they should have comparable thickness to capsid size ratio. The tile matching algorithm applied to Carnation and Pariacoto viruses in Chapter 5 shows how the two $T = 3$ viruses (with similar capsid and protrusion sizes but with different locations of these protrusions) are matched to different tiles sets. On the contrary, some viral capsids cannot be differentiated using the tile matching algorithm. An example of this are Canine Parvovirus (pdb-id: *1c8h*) and the Minute Mice Virus (pdb-id: *1mvm*), see Figure 5.10(a). Both are small animal viruses infecting different hosts, that exhibit strickingly similar structural features despite their capsid proteins sharing only 52% amino acids [48].

A parallel can be drawn with the study in [86], which shows using a coarse-grain model that the occurence of quasi-equivalence in viral capsids is a consequence of overall energy minimisation. By analogy, quasilattice-based models can also be viewed as resulting from energy minimisation [43, 71]. The fact that viral capsids can be embedded into icosahedral tilings suggests that not only the quasi-equivalent positions of proteins, but also the whole 3D structure of viral capsids should be a consequence of energy minimisation, albeit in the context of a more complex energy landscape.

Being able to categorise protein shells using tile matching has implications for phylogeny. Indeed, criteria such as the genome type (eg. single- or double-stranded RNA or DNA) are only one option to classify viruses, and capsid structure may be an alternative. Indeed, it has been observed that despite infecting different hosts, viral capsids could display very similar morphologies. Phylogenetic trees of viruses have then been constructed using capsid similarities as indicators [5, 6, 23]. However the present study suggests that these unexpected similarities between capsids may be due to a convergent evolution. That is, two viral capsids of non-related viruses may still evolve towards sim-

ilar capsid designs due to the fact that they are subject to similar geometric constraints. Our study therefore provides a new view also on the constraints impacting on viral evolution.

Bibliography

[1] P. Alippi, P. M. Marcus, and M. Scheffler. Strained tetragonal states and Bain paths in metals. *Phys. Rev. Lett.*, 78(20):3892–3895, 1997.

[2] J. L. Aragón and M. Torres. A geometrical approach to icosahedral displacive phase transitions. *Europhys. Lett.*, 15:203–208, 1991.

[3] M. Baake, P. Kramer, M. Schlottmann, and D. Zeidler. Planar patterns with fivefold symmetry as sections of periodic structures in 4-space. *Int. J. Mod. Phys.*, B4:2217–2268, 1990.

[4] E. C. Bain. The nature of martensite. *Trans. AIME*, 70:23–35, 1924.

[5] D. H. Bamford, R. M. Burnett, and D. I. Stuart. Evolution of viral structure. *Theor. Pop. Biol.*, 61:461, 2002.

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

[7] C. B. Barber, D. P. Dobkin, and H. T. Huhdanpaa. The Quickhull algorithm for convex hulls. *ACM Trans. on Mathematical Software*, 22:469–483, Dec 1996.

[8] R. A. Bernal, S. Hafenstein, N. H. Olson, V. D. Bowman, P. R. Chipman, T. S. Baker, B. A. Fane, and M. G. Rossmann. Structural studies of bacteriophage $\alpha 3$ assembly. *J. Mol. Biol.*, 325:11–24, 2003.

[9] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. E. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The

protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.*, (112):535, 1977.

[10] A. Bondi. van der Waals volumes and radii. *Journal of Physical Chemistry*, 68(3):441451, 1964.

[11] C. Capillas, J. M. Perez-Mato, and M. I. Aroyo. Maximal symmetry transition paths for reconstructive phase transitions. *J. Phys.: Condens. Matter*, 19:275203, 2007.

[12] M. Carrillo-Tripp, C. M. Shepherd, I. A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J. E. Johnson, C. L. Brooks III, and V. S. Reddy. VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res.*, 37:D436–D442, 2009.

[13] D. L. D. Caspar and A. Klug. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.*, 27:1–24, 1962.

[14] P. Cermelli, G. Indelicato, and R. Twarock. Non-icosahedral pathways for capsid expansion. *Phys. Rev. E*, submitted.

[15] Y. F. Cheng, G. Z. Pan, and F. H. Li. Real space approach to the linear phason strain. *Phys. Stat. Sol. (b)*, 170:47–55, 1992.

[16] J. F. Conway, W. R. Wikoff, N. Cheng, R. L. Duda, R. W. Hendrix, J. E. Johnson, and A. C. Steven. Virus maturation involving large subunit rotations and local refolding. *Science*, (292):744–748, April 2001.

[17] J. H. Conway and N. J. A. Sloane. *Sphere packings, lattices and groups.* Springer-Verlag, NY, third edition, 1998.

[18] F. H. C. Crick and J. D. Watson. The structure of small viruses. *Nature*, 177:473–475, 1956.

[19] T. Douglas and M. Young. Host-guest encapsulation of materials by assembled virus cages. *Nature*, 393:152–155, 1998.

[20] Veit Elser. Indexing problems in quasicrystal diffraction. *Physical Review B*, 32, Oct 1985.

[21] M. Everts, V. Saini, J. L. Leddon, R. J. Kok, M. Stoff-Khalili, M. A. Preuss, C. L. Millican, G. Perkins, J. M. Brown, H. Bagaria, D. E. Nikles, D. T. Johnson, V. P. Zharov, and D. T. Curiel. Covalently linked Au nanoparticles to a viral vector: Potential for combined photothermal and gene cancer therapy. *Nano Lett.*, 6:587–369, 2006.

[22] M. Fischlechner and E. Donath. Viruses as building blocks for materials and devices. *Angewandte Chemie International Edition*, 46:3184–3193, 2007.

[23] A. Fokine, P. G. Leiman, M. M. Shneider, B. Ahvazi, K. M. Boeshans, A. C. Steven, L. W. Black, V. V. Mesyanzhinov, and M. G. Rossman. Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *PNAS*, 102(20):7163–7168, 2005.

[24] F. Gahler and J. Rhyner. Equivalence of the generalised grid and projection methods for the construction of quasiperiodic tilings. *J. Phys A: Math. Gen.*, 19:267–277, 1986.

[25] J. E. Humphreys. *Reflection Groups and Coxeter groups*, volume 29 of *Cambridge studies in advanced mathematics*. Cambridge University Press, 1992.

[26] G. Indelicato, P. Cermelli, D. G. Salthouse, S. Racca, G. Zanzotto, and R. Twarock. A crystallographic approach to structural transitions in icosahedral viruses. *J. Math. Biol.*, 64:745–773, 2012.

[27] G. Indelicato, T. Keef, P. Cermelli, D. G. Salthouse, R. Twarock, and G. Zanzotto. Structural transformations in quasicrystals induced by higher dimensional lattice transitions. *Proc. R. Soc. A*, 468:1452–1471, 2012.

[28] Y. Ishii. Structural transformation and phason dynamics in quasicrystals. *J. Non-Cryst. Solids*, 153-154:645–649, 1993.

[29] A. Janner. Crystallographic structural organization of Human Rhinovirus serotype 16, 14, 3, 2 and 1a. *Acta Crystallographica Section A*, 62:270–286, 2006.

[30] A. Janner. Form, symmetry and packing of biomacromolecules. I. Concepts and tutorial examples. *Acta Crystallographica Section A*, 2010.

[31] A. Janner. Form, symmetry and packing of biomacromolecules. II. Serotypes of human rhinovirus. *Acta Crystallographica Section A*, 66:312–326, 2010.

[32] A. Janner. Form, symmetry and packing of biomacromolecules. IV. Filled capsids of cowpea, tobacco, MS2 and pariacoto RNA viruses. *Acta Crystallographica Section A*, 67:517–520, 2011.

[33] A. Katz. *Introduction to the Mathematics of Quasicrystals*, chapter Some local properties of the 3-dimensional Penrose tilings, pages 147–182. Academic Press, 1989.

[34] E. Kaxiras and L. L. Boyer. Exploring the structure of solids through magic strains: Prediction of a new metallic phase of Si. *Modelling Simul. Mater. Sci. Eng.*, 1:91–100, 1992.

[35] T. Keef and R. Twarock. Affine extensions of the icosahedral group with applications to the three-dimensional organisation of simple viruses. *J. Math. Biol.*, 59(3):287–313, 2009.

[36] T. Keef and R. Twarock. Structural blueprints of GroEL-GroES via affine extensions of heptagonal symmetry. *In Preparation*, 2013.

[37] T. Keef, J. Wardman, N. A. Ranson, P. G. Stockley, and R. Twarock. Structural constraints on the three-dimensional geometry of simple viruses. *Acta Crystallographica Section A*, 69:140–150, 2013.

[38] A. Kolesnikov. ISE-bounded polygonal approximation of digital curves. *Pattern Recognition Letters*, 33:1329–1337, 2012.

[39] P. Kramer. Continuous rotation from cubic to icosahedral order. *Acta Crystallographica Section A*, A43:486–489, 1987.

[40] P. Kramer. Covering presentation and coloring of dual canonical tilings. *Structural Chem.*, 13:315–319, 2002.

[41] P. Kramer and Z. Papadopolos. *Covering of Discrete Quasiperiodic Sets*, volume 180, chapter Voronoi and Delone Clusters in Dual Quasiperiodic Tilings, pages 97–165. Springer-Verlag Berlin Heidelberg, 2003.

[42] P. Kramer and M. Schlottmann. Dualisation of Voronoi domains and klotz construction: a general method for the generation of proper space fillings. *Journal of Physics A: Mathematical and General*, 22:1097–1102, December 1989.

[43] J. Ledieu, R. McGrath, R. D. Diehl, T. A. Lograsso, D. W. Delaney, Z. Papadopolos, and G. Kasner. Tiling of the fivefold surface of AlPdMn. *Surface Science Letters*, (492):L729–734, 2001.

[44] L.S. Levitov and J. Rhyner. Crystallography of quasicrystals; application to icosahedral symmetry. *J. Phys. France*, 49(49):1835–1849, 1988.

[45] F. H. Li, G. Z. Pan, and Y. F. Cheng. Transformation from icosahedral phase to body center cubic phase in cut description. *Chinese Phys. Lett.*, 6:465–468, 1989.

[46] F. H. Li, G. Z. Pan, S. Z. Tao, M. J. Hui, Z. H. Mai, X. S. Chen, and L. Y. Cai. From quasicrystals to ordinary crystals. *Phil. Mag. B*, 59(5):535–542, 1989.

[47] D. B. Litvin, V. Kopsky, and J. L. Birman. Symmetry and phase transitions in decagonal quasicrystals. *J. Phys A: Math. Gen.*, 20:6169–6184, 1987.

[48] A. L. Llamas-Saiz, M. Agbandje-McKenna, W. R. Wikoff, J. Bratton, P. Tattersall, and M. G. Rossmann. Structure determination of minute virus of mice. *Acta Cryst. D*, pages 93–102, 1997.

[49] Y. Ma, R. J. M. Nolte, and J. J. L. M. Cornelissen. Virus-based nanocarriers for drug delivery. *Advanced Drug Delivery Reviews*, 64:811–825, 2012.

[50] Maple. Maplesoft, a division of Waterloo Maple inc. , Waterloo, Ontario.

[51] M. Marji and P. Siy. A new algorithm for dominant points detection and polygonization of digital curves. *Journal of the Pattern Recognition Society*, 36:2239–2251, 2003.

[52] A. Masood. Optimized polygonal approximation by dominant point deletion. *Journal of the Pattern Recognition Society*, 41:227–239, 2008.

[53] N. K. Mukhopadhyay, N. Ishihara, S. Ranganathan, and K. Chattopadhyay. Rational approximant structures and phason strain in icosahedral quasicrystalline phases. *Acta metall. mater*, 39(6):1151–1159, 1991.

[54] NIAID. HIV virion, accessed 18th January 2013. `http://www.niaid.nih.gov/topics/HIVAIDS/Understanding/Biology/Pages/hivVirionLargeImage.aspx`.

[55] Z. Papadopolos and G. Kasner. Delone covering of canonical tilings $\mathcal{T}^{*(d_6)}$. *Ferroelectrics*, 250:409–412, 2001.

[56] Z. Papadopolos, R. Klitzing, and P. Kramer. Quasiperiodic icosahedral tilings from the six-dimensional bcc lattice. *Journal of Physics A: Mathematical General*, 30(6):L143–L147, 1997.

[57] G. S. Pawley. Plane groups on polyhedra. *Acta Cryst.*, 15:49–53, 1962.

[58] J. C. Perez and E. Vidal. Optimum polygonal approximation of digitized curves. *Pattern Recognition Letters*, 15:743–750, 1994.

[59] A. Pikaz and I. Dinstein. Optimal polygonal approximation of digital curves. *Journal of the Pattern Recognition Society*, 28(3):373–379, 1995.

[60] M. Pitteri and G. Zanzotto. Beyond space groups: The arithmetic symmetry of deformable multilattices. *Acta Cryst. A.*, 54:359–373, 1998.

[61] M. Pitteri and G. Zanzotto. *Continuum models for phase transitions and twinning in crystals.* CRC/Chapman and Hall, London, 2002.

[62] U. Ramer. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1:244–256, 1972.

[63] B. K. Ray and K. S. Ray. Determination of optimal polygon from digital curve using $\mathcal{L}_1$ norm. *Journal of the Pattern Recognition Society*, 26(4):505–509, 1993.

[64] I. K. Robinson and S. C. Harrison. Structure of the expanded state of tomato bushy stunt virus. *Nature*, (297):563–568, 1982.

[65] Schrödinger. *The PyMOL Molecular Graphics System.* LLC.

[66] M. Senechal. *Quasicrystals and Geometry.* Cambridge University Press, 1996.

[67] M. B. Sherman, R. H. Guenther, F. Tama, T. L. Sit, C. L. Brooks III, A. M. Mikhailov, E. V. Orlova, T. S. Baker, and S. A. Lommel. Removal of divalent cations induces structural transitions in red clover necrotic mosaic virus, revealing a potential mechanism for RNA release. *J. Virol.*, 80(21):10395–10406, 2006.

[68] B. Sing and T. R. Welberry. Deformed model sets and distorted Penrose tilings. *Z. Krystallogr.*, 221:621–634, 2006.

[69] H. Sowa and E. Koch. Group-theoretical and geometrical considerations of the phase transition between the high-temperature polymorphs of quartz and tridymite. *Acta Crystallographica Section A*, 58:327–333, 2002.

[70] S. Sri Krishna, M. Sastri, H. S. Savithri, and M. R. N. Murthy. Structural studies on the empty capsids of physalis mottle virus. *J. Mol. Biol.*, 307:10353–1047, 2001.

[71] P. J. Steinhardt, H.-C. Jeong, K. Saito, M. Tanaka, E. Abe, and A. B. Tsai. Experimental verification of the quasi-unit-cell model of quasicrystal structure. *Nature*, 396(55):262–265, 1998.

[72] P. G. Stockley and R. Twarock, editors. *Emerging Topics in Physical Virology.* Imperial College Press, 2010.

[73] J. Sun. Structure transitions of the three-dimensional Penrose tiling under phason strain field. *Chinese Phys. Lett.*, 10:449–452, 1993.

[74] F. Tama and C. L. Brooks III. The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. *J. Mol. Biol.*, 318:733–747, 2002.

[75] F. Tama and C. L. Brooks III. Diversity and identity of mechanical properties of icosahedral viral capsids studies with elastic network normal mode analysis. *J. Mol. Biol.*, 345(2):299–314, 2005.

[76] L. Tang, K. N. Johnson, L. A. Ball, T. Lin, M. Yeager, and J. E. Johnson. The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA. *Nat. Struct. Biol.*, 1(8):77–83, Jan 2001.

[77] P. Toledano and V. Dmitriev. *Reconstructive Phase Transitions.* World Scientific Publishing Company, 1996.

[78] M. Torres, G. Pastor, I. Jimenez, and J. Fayos. Geometric models for continuous transitions from quasicrystals to crystals. *Phil. Mag. Letters*, 59:181–188, 1989.

[79] R. Twarock. A tiling approach to virus capsid assembly explaining a structural puzzle in virology. *J. Theor. Biol.*, 226(4):477–482, 2004.

[80] I. Wagner and H. Musso. New naturally occurring amino acids. *Angewandte Chemie International Ed.*, 22:816–828, 1983.

[81] X. Wang, J. Yu, A. Sreekumar, S. Varambally, R. Shen, D. Giacherio, R. Mehra, J. E. Montie, K. J. Pienta, M. G. Sanda, P. W. Kantoff, M. A. Rubin, J. T. Wei, D. Ghosh, and A. M. Chinnaiyan. Autoantibody signatures in prostate cancer. *N. Engl. J. Med.*, 353:1224–1235, 2005.

[82] J. Wardman, 2010. Private communication.

[83] J. Wardman. *A Symmetry Approach to Virus Architecture*. PhD thesis, University of York, 2012.

[84] T. R. Welberry and B. Sing. Deformed Penrose tilings. *Phil. Mag.*, 87(18–21):2877–2886, 2007.

[85] A. Winzen and H. Niemann. Matching and fusing 3D-polygonal approximations for model-generation. *IEEE International Conference*, 1:228–232, 1994.

[86] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci.*, 101(44):15556–15560, 2004.