# Gresham College Main logo

20 May 2015

**Geometry:**

**a new weapon in the fight against viruses**

Professor Reidun Twarock

Thank you very much for the kind introduction. It is a pleasure to be here in London. I am a mathematician and, actually, I trained as a mathematical physicist but I made gradually a transition into mathematical biology, and the way I am working is highly interdisciplinary, so my mathematics is always at the heart of what I am doing and plays a crucial role in answering open questions. I very much work on the interface with biophysics, bioinformatics, computational chemistry and biology, as you will see in this lecture.

So what I am doing is working on viruses. You all certainly are familiar with viruses. I had one a week ago, a rhinal virus, a common cold, so if you hear me a bit coughing still, that is from that cold. So all of you, surely, are familiar and you have heard in the press about all sorts of viruses – surely you have heard about HIV and Hepatitis C. There are cancer-causing viruses. There are viruses that actually are linked with diabetes. So, viruses are really important and what I am going to show you today is that, actually, by using mathematics and by using geometry and actually staring at little objects like that, you can answer open questions and ultimately contribute to the design of novel antivirus strategies. So, that is the gist of my talk.

I want to go through a little bit of background before I am diving into mathematical applications. Viruses are known since the antiquity, so you see, in old frescos from Egypt, people with polio infections, remnants of polio infections here, so we know that those viruses have been around for a long time and have been documented.

Also what is interesting is that they infect all kingdoms of life: they infect humans – here is a smallpox infection; they infect plants; they infect bacteria.

What is interesting for the mathematician is that those rules that govern the organisation of those viruses are actually common to these different viruses. So, while the biologist often focuses on one of those viruses, we are looking for commonalities, for mechanisms. We are looking for understanding principles that underpin the functioning of those viruses.

I really, really want to show you this movie here because it is absolutely fascinating to see how this works because it gives you a really nice glimpse of how those viruses actually work. So this is a bacteriophage. It is a virus that infects bacteria. You see here this bacteriophage coming to approach its host cell, and what you see on top here is the [viral] capsid and that is what we are going to look at from a mathematical point of view. Also, here, you see this little tubular structure, which is used, in this case, as a tail sheaf which is used to inject genomic material. All of those components have beautiful mathematical properties, as we are going to see later, so these are aspects we can try and understand with mathematics. Here, you see how this bacteriophage, in this case, is penetrating through the wall of its host and then injected the genomic material.

What we have seen here then is that viruses are like little machines on the nanoscale, so you have seen these bacteriophages here and if you were to zoom in a little bit more, into those proteins that make up those biological building blocks that make up this tip of this sheath, you see it is very much like a syringe. So, it is a little nano-machine, and what we are going to see later is that, with the mathematics we are using, we can actually understand the functioning of those machines. It is not just a matter of describing the structure of those objects, but actually understanding how they eject their genomic material. That is what we want to do, and to contribute to that.

Before we are getting there, we have to talk a little bit about viruses to understand really how small they are, and here I have got for you a show graph that shows you how far you can look down with your bare eyes – obviously, humans, fish, ants. You can see all of that, and eventually, you will need a light microscope to see those structures, perhaps to see a cell, but then, if you want to look at viruses or even DNA or RNA, the genetic code of those viruses, you will have to look with what we call an electron microscope. We are radiating electron beams at the structures and then analyse those pictures, and as you are going to see later, in order to better understand those structures, you need to use these mathematical properties of those objects in order to make clearer how those structures really look.

So if I put an average sized virus next to a flea, it is as if you are standing next to twice the size of Mount Everest. So, these structures are really, really small, and that makes it so fascinating for the mathematician because, through the mathematical microscope, we are actually able to see things that are very, very small, understand intricate details of objects that are extremely small.

Here, I show you a movie of how, actually, viruses, in this case, again, phages, are infected their hosts. This is to demonstrate that viruses actually cannot replicate on their own. They need to invade a host. They basically need to hijack the host machinery in order to produce more viruses. So, when we are thinking about viruses in this research programme, we always have to be aware of the host interactions, the host immune systems, because all of that is important, and we are later on talking about the implications for evolution of those structural aspects that we are discussing in this talk.

What I therefore have shown you here is that - we have the first mathematical equation here in my talk – is that virus is actually Trojan horses, and this container, which is formed from proteins here, is basically the Trojan horse that contains the genomic material and brings it to the host cell and helps with the infection mechanism. What we are going to see today is that, actually, it is very important to look at this from a mathematical point of view, not just on the surface, on the container level, but actually to understand how genomic but actually to understand how genomic material is positioned inside of those containers, and we will see that the mathematics we are developed in my group can actually predict how this correlation works, and that this understanding then helps you to better understand how these viruses form. So if you want to think about what my research is all about, think about this – it is a little jar with RNAs inside.

Right, so let us work towards the maths. This is a cryo-electron micrograph of viruses. So we said earlier, we need electron beams to be radiated against those viral samples because they are so small, and that is the kind of quality you would see at that level. But we really want to understand a bit better what these viruses look like, so we need to actually arrive at these very fine reconstructions, so this would be a viral capsid after some mathematical procedures, averaging procedures, have been applied, and we will see later what they are. But if you were to zoom in to the surface, what you would see are these little doughnut-like shapes that are arranged in these periodic arrangements, these lattices as we call them as mathematicians, tessellations, and that is where my mathematical interest comes in. These are surface tessellations and in fact, if you were to zoom in on any of those, you would see, these shapes, they are formed from what we call helices and sheets. These are proteins, biological building blocks of these containers, but what we are really interested in, as mathematicians, is to try and understand this overall arrangement, the thickness of these containers, and the correlation with genomic material.

Now, how do we do this? Well, I am getting out my mathematical microscope here, so instead of the electron microscope, looking at this, actually, I am looking at that. So this is an icosahedron. It is a polyhedral shape, as we say – it is one of the Platonic solids, looks like this, and the reason why this is interesting in this context is that it shares symmetries, as we say, with the virus.

Now, what do I mean by that? Let us look a little bit at viruses and geometry…

So, symmetry operations, you are all familiar with them. You see them all around you. The simplest is thinking about an axis going through your body and you are reflecting one side on the other side – this would be reflection symmetry. But, also what you can have is so-called rotational symmetries. Imagine you have an axis sticking out of the middle here and you turn by 180 degrees and you get this same shape again. So, you can have that symmetry in different areas in life. For instance, in this playing card here, if I put my axis in the middle here, you can see that the heart would map onto that heart, and that is rotation. So, it is certainly not a reflection, because if I were to use this as a reflection line, the heart would go here, but it does not. So it is really a proper rotation. You can have different types of rotations. This is called a three-fold rotation because you have a third of 360 degree rotations, 120 degrees, keeping your structure invariant, as mapping onto itself. Another example, imagine axis sticking out here, would again do the same job… And four-fold rotations, and so on and so on…

When we are in three-dimensions, it is exactly that same idea, but now these axes are at different angles with respect to each other, and if I am speaking about the symmetry of this object, in my mind, I am thinking about a collection of those axes of different type, in this case, five-fold, three-fold, and two-fold axes, that have specific orientations with respect to each other.

So, let us see what that means. Here is my viral capsid. Here is a rendering of its surface structure. So I have taken this object here and I have basically superimposed a surface lattice, which is given by little hexagonal shapes.

Now, here, where I have marked the 3, if you imagine there is an axis that goes through the centre of this object and the number 3, then that is a three-fold rotation of my axis, three-fold rotation of my icosahedron.

Similarly, if I take the midpoint on this edge, where I have indicated a 2, and stick an axis through the 2 and the centre of my structure and rotate by 180 degrees, again, that is an invariance of my structure, and likewise, the corners here, the five-fold axis, with the centre of my structure. So you can see there are 12 five-fold vertices but there are always two on the same axis, so there are six five-fold axes, and similarly, you can count the three-fold, so there are 20 triangles, but again, they are opposite to each other, so 10 three-fold axes, and we have 30 edges, but they are again opposite, so we have 15 two-fold axes.

For a mathematician, all of those objects here are exactly the same from a symmetric point of view because they share the same symmetry axis. So, whether we are talking about this viral capsid, the icosahedron, a soccer ball, it is all the same – also the virus here.

Now, we have established another equation: a virus is actually a soccer ball, for me. It is a good analogy.

Let us think a little bit about why a virus has symmetry. It seems alien. Why is there symmetry at the nanoscale? If you think about it from a biological point of view, it makes a lot of sense for the virus to try and have as small as possible genomic sequence to code for the container because then it can create a container with a relatively large volume, into which a relatively small genome has to be packaged. So, the virus wants to optimise container volume, while, at the same time, minimising genome lengths. So, what is the way forward? It is to generate a minimal amount of different building blocks, different types of building blocks, in the easiest case, a sino-one and then repeatedly synthesize it and use it, and that is what the virus does, and the multiplicity with which those building blocks then come together to form these containers are determined by the symmetry, in this case the icosahedral symmetry. So, unless a building block is positioned on a symmetry axis, it would come in multiples of 60, in these containers. So the simplest virus that indeed exists in nature would have 60 proteins.

Now, why is that interesting? And it is a mathematical lecture so we want not to lose sight of why icosahedral symmetry is actually special. If you think about the classification of finite groups of these symmetries you can have in three-dimensions, or from a geometric point of view, if you ask the question “What kind of objects can I form that have the same edge lengths and the same types of faces and sort of are invariant in the symmetry groups?” I am ending up with my Platonic solids here, and the icosahedron, the dodecahedron, are the ones that both have the icosahedral symmetry and is, as we say, the largest symmetric group in three-dimensions. If I think about rotational symmetries, there can be reflection symmetries – let us go there later, but from the rotational point of view, it is the largest symmetry group, so it is not surprising that viruses pick this because they get the largest multiplicity for coding just for a single building block or for a limited number of building blocks. So it makes all sense from a biological and mathematical point of view, and I should say that, for me, the common cold again, thank God I have not been coughing so far, and another virus that causes cervical cancer, they all fulfil the same symmetry rule, but they look different. So, they are not, when you look at them, the same object, and the reason is that this one has many more different building blocks, so obviously, symmetry on its own cannot be the only determinant of viral structure. This is where my research comes in: I want to understand what the principles are, what are the other principles beyond the icosahedral symmetry that actually account for what we are seeing in the virus sphere.

Now, I have got my little friend here, who is always with me, so he is someone we are still not quite clear about. He certainly breaks symmetry, with his eyes glaring at you, but he is like our conundrum – we are still puzzling with him. But otherwise, we are getting a little bit of a grip of what is going on.

The first step in generalising those rule sets and understanding what determines viral symmetry was actually done by Casper and Klug, in the so-called quasi-equivalence theory in the ‘60s. So, here, people were asking the question, so, if I have a larger virus, what are the additional rules that complement icosahedral symmetry that account for what I am seeing? And they were biologists, so they looked at the problem from a biological point of view, and said, well, let us classify structures where the local bonding environment, the way that proteins interact with each other in the capsid, is similar in all these positions. So, in other words, these little dots here, they are place-holders for the position of a protein. We have seen proteins are complicated – they have sheets and helices, but for me, a single protein is just a dot right now. And this requirement, from a mathematical point of view, would mean that I could tessellate this surface into triangles, a triangulation could be formed, such that I have the positions of proteins marked in the corners of these little triangles, because, then, locally, every protein sees itself sitting in a triangular environment. So, locally, they all look the same, but obviously not globally. It is what we call local symmetries.

So this was then used as an idea to start classification. It was actually quite an important classification. So, you take the icosahedron – this is just the icosahedron, one of its 20 faces shown in blue. You are marking the position of the proteins in the corners of the triangle, and that would be what it looks like when you render it, when you take all the atomic positions of your proteins and look what it actually looks like.

The next largest object you can get has three times that triangles, so this is called a T3 virus therefore, and the way to interpret this triangulation on that surface would be to put a little dot in every corner, so the red ones are always around the five-fold axis and the green ones are around the three-fold axis, in this case, but the form clusters of six, and when they are local, it is local threes.

Now, we can continue that game, look at what triangulations are compatible with icosahedral symmetry, enumerate everything, and generate larger and larger viruses, and the beauty of this theory is that all of those virus structures have eventually been discovered. So, it is a quite powerful tool to know, a catalogue of what you can potentially have.

Obviously, I am a mathematician and I am attracted by open problem that challenge a little bit the mathematics as well, so all of that theory was out there when I entered the field, so I got attracted because there are viruses that do not fulfil those rules.

So, this is a cancer-causing virus here. Actually, that paper where this picture comes from was given to me by a biologist, who said, “Well, we have problems here – you are mathematicians, can you not do something about it?” So what is special about this virus is that every cluster, everywhere in the capsid shell, has five-fold local symmetry, is composed of five proteins. Now, we know, on the five-fold axis of icosahedral symmetry, expect that, indeed, but everywhere other than that, we do not, because we said earlier on we needed these triangulations, and they lead to hexagonal lattices, so that is obviously not compatible with this problem.

So why did I get excited about it? Because, actually, it is, from a mathematical point of view, a fundamental problem that is related to non-crystallographic groups, to a lot of beautiful maths, different types of number fields over irrational numbers. So, what is the problem? Think you want to tile your bathroom. The bathroom is now, I am sure, tiled with little squares, but you want something a bit more fancy. So, you try pentagons, but if you try and do this, and try to glue them to each other, you will see that you are getting these gaps here that you cannot fill unless you paste over it, which you do not want to do. So, this is known as the crystallographic restrictions. We cannot have periodic arrangements that have this five-fold symmetry.

But, we can have long-range order with that symmetry. So, Dan Schechtman discovered in the ‘80s, via defraction patterns, that there are alloys that organise their atoms in such a way that they actually have long-range order with this non-crystallographic symmetry. Actually, it caused him a lot of grief because people initially were quite critical of this approach and thought it cannot be, but the problem was that people equated periodicity with long-range order, and these are different things, and that, actually, was known to Roger Penrose, long ago before this discovery was made. You have probably heard about those famous Penrose tilings. These are tessellations but you can continue ad infinitum if you do it right, locally. In this case, of two shapes – there is these two rhoms, a large rhom and small rhom, and you can actually do this so that you have this five-fold symmetry here. Look here, there is a five-fold axis here. If I stick an axis through here, out of the plane, you will see that it has the five-fold symmetry, so it is actually possible. The price to pay is that you no longer have just one building block, and you have to choose your building blocks mathematically. So, it is all in the formulation of the problem.

Let us come back to the virus problem. So, this is this polyomavirus again, or papillomavirus there, working the same way, and what we did then was to realise that in order to solve that problem that we have been made aware of, we had to look at this problem from the point of view of these Penrose tiles and from the point of view of this non-crystallographic tilings. So, what we have done then is to tessellate the surface of this virus in terms of these kites and rhoms, and whenever a vertex is positioned in the tiling, at a five-connected vertex – so, let us look at this one, there are five edges coming out – then we are marking the position of a protein in its corner. So, in other words, the proteins are either sitting opposite each other, across such a rhom, in which case they would interact with each other, so there would be an interaction between these two proteins, or, like up here, they would be sitting in a triangular arrangement. So it is a different type of interaction. In other words, the idea of quasi-equivalence earlier on was that all the interactions are the same, the local environments are the same, but what we are saying here is that, actually, in some cases, you have to give that up. You have to admit that there can be environments that are different, and the mathematical way to formulate this is very much related to these ideas of Penrose tilings and quasi-crystals.

Now, we have done that in 2D initially, but then we have done this also in three-dimensions because we were asking the questions: where do these two-dimensional tilings come from? And you can actually rationalise them as sections in these higher dimensional tilings. So, think about a tessellation of three dimensions, with more complicated shapes like those, and then take surfaces in these and then look – we have superimposed a virus here for you to see – so then we look at the outer surface of the virus, the inner surface, and you can see that all of that is accounted for by different radial levers in these structures. So, the beauty is that in this mathematics, there is so much more information than we had before. It is not just about the position of proteins. It is actually also about the thickness of the containers and certain details about the proteins and how they are arranged and how they are interacting.

We have talked about the virus sphere and the structure of those containers, and at that point, they were fairly well understood, but the next really crucial open problem faced us, and this was that there was little information at that point about the organisation of the genome inside of the capsids. So, we are only talking about protein containers until now, which is very important, but then we want to understand how viruses function, how they form, how they evolve, and that all is intimately linked with the genomic material.

Here is an example. It is Pariacoto virus. It is an insect virus. When I was cutting this virus open and I look inside, I see this dodecahedral cage structure here. This is viral RNA. So, viruses can either store their genetic information in what we call RNA or DNA. This virus would do this with RNA. It is an RNA virus. And the question is: is there any mathematical rule that tells me how the genomic material should be organised, given that the container has a certain shape.

So, this is our little artist’s impression. We have an arts society, and also, I painted when I was thinking about the Pariacoto virus, thinking nothing else but Pariacoto virus, so all I could do is paint a Pariacoto virus. Here it is.

This is the research we are doing in YYCSA, in the York Centre for Complex Systems Analysis. Now, what did we do? So, we actually – and I will tell you a bit more about the mathematics in a minute – we worked out a way of correlating the structure of the protein layer with that of the genomic material. Here, you see outside the protein layer, inside this cage of RNA, and the points you are seeing – we will see in a minute where they come from – are our mathematical models, and they are such that every point is related to every other point by a symmetry group, so there is symmetry behind that. These are our biological collaborators, Peter Stockley and Neil Ranson, and these were two people from my team, a postdoc, at that time, PhD student, at that time, so we worked together looking at that problem.

Now, what is happening? Well, I told you about symmetry before and I told you about these different axes and an operation which is a rotation about an axis – you might want to think of as a person, so they form a group, we said there are symmetry groups, so it is like a group of people. And we said this group has 60 people in, but we get somebody else who is slightly different. In this case, it would be an operation that has a different character to it. These rotations then move objects on a sphere around, if you want, an origin, you want to imagine, so they are moving on a sphere, but you need something that moves away from the sphere, like a translation, so instead of rotating around, we can move. We can either walk, we can translate, so we can do other…we do an operation that does not preserve the distance from the origin and that is the little guy playing football.

Now, I apologise to the mathematicians. If anybody wants to know more, this is one of the articles on that, but I am happy to stay around and answer all the detailed questions. All I want to say is that there is really exciting maths in this. You have to work over different number fields, for reasons – with these non-crystallographic symmetries, you are no longer working with integers but you have quadratic extensions of integers. You have very nice extensions, matrices that you are extending to get those extensions. You can get them induced from higher dimensional projections – there is a reason for that, and it is all in what we call representation theory. So, do talk to me about that or if you want to send me an email, I am happy to send more, but for the purpose of this, I just want to say today so we have this extra element and that makes my group bigger.

So, this is, unfortunately, again, the same problem we had before. So what I am showing you now is how these objects are working on a viral…on data for virus, a real virus. You will see here our mathematical model superimposed on a surface rendering of a virus. This is again Pariacoto virus. You see it rotating, spinning. You see its surface at the moment. And now what we are going to do is to slap into the structure and see inside. So, I am taking away the surface so you get a glimpse of what is inside. The points are those members – they have been formed by combinations of these generators of my symmetry group. So, when this little guy who plays football is playing together with the other elements in that group, then I am getting these different points, and every point can be met on every other point by elements of this extended group? So, it gives me one point, and from a mathematical point of view, every other point is completely determined, so therefore I can classify those structures as well. So, as soon as I know what is the number of these football-playing guys I can get in, that are mathematically sensible with the rest of the group, I can catalogue all of those geometric models and see what they do.

Here you see this part of my point array has been used to pick out of the library the best fed one. We have on purpose not used any data on the genomic material that is employed by the mathematics, but as you can see, it maps beautifully in what we call the minor grooves of this RNA, and also sees those green vertices on those junctions. So what I am telling you here is actually that the mathematics is constraining the organisation of the genomic material, given the organisation of the protein container. There is a correlation, as we say, between the two things.

Let us have a look once more. So, let us have a look once more at this cancer-causing virus we started off with, initially, where we just had the tiling, and now, if we zoom in around those clusters, the points will tell me something about their extent at different radial levels. So, they are really almost like a blueprint. They are mapping around material boundaries. So you get a lot more structural information out of the mathematics here. You can also see these interactions, they are completely determined, and that will prove very important for us.

If you are a mathematician and you are developing a mathematical model inspired by an application, you want to see that this is more generic as a mathematical structure and has applications in other areas because, otherwise, as we can say, fiddled a tool set for your application, but if you have something generic, it should be having applications in other areas, so we were pleased to see that actually our new mathematical tools could also be applied to fullerenes. These are carbon cage structures and, actually, you probably have heard about the famous bucky-ball that was discovered by Sir Howard Kroto and for which he got the Nobel Prize in 1996. What has been subsequently seen is that these carbon stage structures can also occur as nested structures, like a Russian doll arrangement, so you have different structures, say, a C60 with 40 carbon atoms, a C240 with 240 carbon atoms, and so on and so forth, and as you can see in this rendering, there can be quite a few of these shells, and the question is: is there a mathematical rule that, given the position of the atom in this carbon onion, as it is called, gives me all the other atomic positions? And actually, we have been able to show that our mathematical structures fit that bill and actually can deliver that. That is a paper that appeared last year and we also got picked up Nature Physics for it, so it had some ramifications in other areas as well. We are obviously very happy about that because it means like we are having a generic mathematical structure.

But now I want to go onto something that really excites me because we got a lot of press for this this year, it is how geometry helps you to do code-breaking in viruses. So, the big problem that biologists and virologists are facing is to understand how those building blocks get together to form the container. We know what it looks like but can we actually understand the pathways of assembly, the production line? So, as if you are going to IKEA, you are getting your shelf, and you want some kind of instruction on how to put it together. So, what is the instruction – can we understand this from a mathematical point of view?

This movie that I am going to show you is quite instructive because it thinks about this problem just from the point of view of the proteins. This is actually the belief that was around in the community for a long while. The movie is actually, using this object – this is a structure formed from 12 pentagonal clusters. They are basically 3D prints, with magnets mimicking the interactions between the different building blocks. So, I have got 12 building blocks. Watch what happens when I am shaking… It breaks apart and eventually, the structure comes together. If I am shaking and shaking more and more – this is thermal energy – as you can see, I am getting the structure back. So, people knew about this, and you can also do this in the test-tube with real viruses. You can actually, as we say, purify the cups of proteins, put them in a test-tube and see how it all assembles. That is why people thought for a long while that genomic material would not be important in that context, and what we have been able to show, in contrast that, it is actually extremely important when you are thinking about efficiency of assembly, because one thing is to assemble but the other thing is to do it fast, do it efficiently and be able to outsmart the immune system, and that is really what the virus has to do.

Now, how did we approach that problem? Well, we have spoken quite a bit about the outside view and we have also looked at the mathematics that correlates it with the inside view. So, this is the outside of this bacteriophage MS2, and if I am slabbing in, I see these two rings, which is data for the RNA, the genomic RNA, inside, and what we really want to do is bring everything together now, from a mathematical point of view.

If I take a lot of these viruses, many of those, and I superimpose them according to their symmetry axis, then I am scrambling, if you want, around the genomic material because I do not know if I have got a given symmetry axis – do I have this axis, do I have this axis, or this axis? – because we said before its invariant under these rotations. So, suppose there is a defined organisation in the capsid… If you take many particles and you align them with regard to their symmetry axis, you generate automatically a structure inside that has symmetry. So, if I take this virus then and I align many of those viral particles according to their symmetry axis, I slice it open and I look at the RNA in proximity to the capsid, I see an RNA organisation that looks like this polyhedral shape. But in any single one of those of course, it is like a path, like a path on a polyhedral shell, but I have just, through my averaging, I have created this polyhedral shell.

So, what we really need to understand then is something like a travelling salesman problem. Hamilton himself actually introduced a board game that is very much appropriate for what we are doing here. So, as you can see here, what is on the board is the dodecahedron. It is something that distorts the 12 pentagonal faces, but it is actually the connectivity of the edges of a dodecahedron. In Hamilton’s game, the player had to find a path that visits every vertex, ticks every corner, precisely once, and that is a Hamiltonian path – they had to work out the Hamiltonian path. That is precisely what the virus has to do: it has to work out the Hamiltonian path. It needs to visit every vertex because there is a biologically important function happening there – we can see later. So, mathematically, it needs to be a Hamiltonian path on this polyhedron.

So, in other words now, bring everything together: we have the group theory that tells me the correlation that correlates the tessellation of the outside and the organisation of the inside. What I am doing here is actually I am taking an icosahedron, its surface, and I am cutting some of the edges open and I am putting this surface flat on the plane. So, imagine you were to identify all of those five points, you would generate one of those corners of the icosahedron. Likewise, you would glue those five corners together, you would get the opposite corner, so you would get the surface of the icosahedron back. But it is easier for the mathematician to actually look at this, see at the same time the tessellation of the virus surface, and see the organisation of genomic material. So that is the objects we are looking at.

Let us come back to our IKEA instruction manual problem. So what does this actually tell me about how to put the virus together?

Well, these viruses use something that is called nucleated assembly. They start with a contact between a protein and the viral RNA, and then, around that nucleus, the capsid grows, through both interactions with the RNA and interactions between the proteins, and the order in which those protein building blocks are coming in to form the container is specified by the geometry of this Hamiltonian path. So, if I know which Hamiltonian path it is using – again, a movie that does not work, so I am afraid I will talk you through the movie. So, the movie would have walked along…this is basically my polyhedral shell that gives me the arrangement of the RNA underneath the capsid, and depending on where my RNA is positioned, I know which of those protein building blocks will be recruited in which sequential order to form my container. So, actually, my instruction manual is very much encoded in the RNA, in the geometry of the RNA organisation. So that brings us to the question: so, how many of those paths are there? Let us play Hamilton’s game, so how many Hamiltonian paths do I have on my polyhedron, and it turns out – you can obviously enumerate them mathematically – there are over 40,000 of those paths. So that would be a lot of freedom, a lot of complexity for the virus, but you can show – and I will not be able to go into detail of every little bit of it – with different aspects, call it kinetic modelling, it is the biophysics of how the container builds up, bioinformatics analysis – I brush on this a bit later – and the analysis of cryotomograms, which I am afraid I will not have time to cover, that it is highly restricted. Among those many possibilities, there is one very dominant one, and in most cases, take a random virus out of your bag, you are likely to see this specific organisation.

Now, why is that? We have to understand a little bit of the biology to understand this. So, we have two types of building blocks here, and in solution actually, all those building blocks look similar. These building blocks are what we call diamides, two proteins bound together, and they have dynamic behaviour. They have these so-called FG loops on the sides that have symmetric and asymmetric motions, and these motions, these are the four dominant modes you can have. But when an RNA stem loop – this is a little shape that is formed from the RNA genomic material – is binding against this, you see that the motion is displaced. One of those motions will become stronger at the expense of the other one becoming weaker, and the one that becomes stronger actually is therefore able to easier flip over and form this sort of flipped asymmetric arrangement, and that is needed. It is crucial because you would not be able to fit the symmetric ones around the five-fold axes – these have to be asymmetric, otherwise you would get steric clashes around the five-fold axes. So, it is a very important functional role that those little shapes are playing here, and these little shapes – here comes the mathematics back – are actually positioned at the corners of those Hamiltonian paths. So depending on where they are sitting, what their affinities are, it will impact dramatically on how efficiently that viral capsid is building up.

Now, this allowed us to solve a really important open problem in the community, which was to answer the question whether or not there is specificity in the interactions between the RNA and the protein. People were originally thinking you just need protein and do this game we have played before. Then people realised that, actually, the genomic material, if it is present, will enhance the efficiency of that, but it was not clear whether that was just electrostatic interactions or whether there were actually quite specific interactions, so they were specific patterns in the genome, and the problem with those patterns is that they are a combination of a structure and a sequence. The sequence identity, the sequence component of the pattern is really minor – that is the problem. If you are going, as a bio-mathematician, and you look for those sequences, you will not have much luck. You find either too many or you cannot just tame the zoo because you do not quite know what is happening. You have to understand the geometry of those, and this we were able to do by bringing in the graph theory and our understanding of those Hamiltonian paths. That was the key for us to actually be able to determine that, and, actually, we were able to see, for viruses in the same family, that there are different viruses, GA and MS2, that even though the positions in the sequence, in the letter sequence – as we called a primary sequence – are different, the geometric organisation in those is conserved. So, geometry is very much a guiding principle in all this, and if you understand the geometry, you will have a leg up on the viruses because you understand how they actually form and what is the concerned determinant in what they are doing.

And that helped us, so we got a little bit of press at the beginning of the year, with our experimental colleagues, because we did this for a couple of viruses and they actually called it the viral “enigma machine”. So, it is basically finding a hidden code in the code. We all know the genomic sequence codes for the proteins, but actually, we did not know that it also codes for this efficient build-up of the capsid, and using this kind of mathematical approach and using the graph theory, we have been able, in a lot of these viruses now, to identify those patterns. So that was basically the little press release we had, so they said we had cracked the code that governs infections by a major group of viruses, including the common code and polio, very much in collaboration with colleagues at the Astbury Centre in Leeds, Peter Stockley, Roman Tuma and a lot of postdocs, my colleague Eric Dykeman and so on. So, lots of people involved in that.

And that is how these little patterns look. You see there are some conserved elements to them, but if you stare at them at first, you would almost not see it, unless you see those subtle rules that determine them, and suddenly they are in front of your eyes and they all make sense.

So, in other words, we have been able to contribute to what we call a paradigm shift in our understanding of virus assembly, away from a protein centric view or a view that just thinks about non-specific interaction, to one that actually recognises how important genomic material is in this context, and if you want to understand, there is a really beautiful picture that summarises it within the Huffington Post. So, here is the genome of this virus. What happens, it is actually like a self-packing suitcase: all your little clothes pieces, they know how to fold themselves and jump into the suitcase – would it not be nice if we travelled and we had that, right? And the virus has it actually, has evolved it over the years. It is a bit like that.

Now, as mathematicians, obviously we want to drill a little bit deeper and we want to really understand what is going on and we want to really also get at the evolutionary consequences of this discovery, and in order to do so, we are using a spherical cow, as we say. So, sorry, I did not find a nice picture of a cow, but for me, a spherical cow is always a dodecahedron, so it is this...

So, this is my prototype of a virus. It is the simplest structure with icosahedral symmetry because it has just 12 pentagonal shapes that come together to form the surface. It is a dodecahedron. And can you understand, based on a dodecahedron, how all of this works? That is a paper we had last year in PNAS. So, what we did was to actually take a place-holder for these viral RNAs by just assuming you have these 12 contact sides, with affinity to the protein that is variable, so we are allowing an evolutionary context for these affinities to vary, but basically, from a modelling point of view, you accept that you have 12 binding sides, each with a given affinity and a certain affinity bend, and we will present them as little circles, and we actually only look at rough sort of areas, so low affinity, intermediate or high affinity, and basically, the colour of the circle would signify which of them it is.

And then we have our building blocks here, and we want to form the capsid. So how does this happen? There are reactions, so the building blocks can bind to these interaction sites, and whenever you have two protein building blocks bound to adjacent packaging signals, as we call these little shapes, then they can also get together and form a nucleus, and then you have proteins binding and falling off, with certain rates, all across the RNA, but when they are binding next to the growing nucleus, they have a chance to bind and contribute to the growing of the capsid.

So, here, again, comes the insight from the geometry that I want to flagship here. This is a sort of larger programme, but I am diving in at that end. So, we tried to understand the complexity of this, a very famous problem in biology and protein folding, which is called Levinthal’s paradox. This is, when you have a protein, what are the pathways of it to fold into what we call the native state, its structure? If the protein were to try all possibilities, then it would take longer than biological time that is available, so there must be some ways of shortcutting of really understanding how you get to this folded state. And it is the same here: there’s a vast complexity of ways in which these structures can build up, but how does it find the most efficient one? And we could show, in this paper, that there is a very, very delicate interplay between the build-up of the concentration of the protein-building block. The affinities, that are finely tuned across the RNA to optimally help with that, so we can evolve these affinities and see how they are converging towards the appropriate distribution, and if you think about this from a geometric point of view, it means you are biasing your assembly towards a smaller number of those Hamiltonian paths. So you can understand, in terms of these Hamiltonian paths, or deviations from Hamiltonian paths, how it actually overcomes this complexity problem, and the better Hamiltonian paths that are the solutions you will see in nature are also the ones that go with the more energetically stable intermediates. But we can get a handle on this, from a mathematical point of view, by thinking about these Hamiltonian paths, and again, if that is all tuned, like the knobs on a radio, in the right way you get a very efficient build-up, a very efficient machine here.

So, that has helped us to actually come up with novel anti-viral strategies. So, we used these models then to understand what happens when you introduce drug molecules that block those interactions, and we could show that two very, very exciting things are happening here. One is that, through this interaction, you delay the build-up of those capsids, but what is even more important is there are cellular competitors. So, these cellular RNAs can also be packaged but then you do not have a viable virus. You have basically a protein shell with the wrong stuff inside, and they act like a vaccination because the surface of the protein shell is exposed to the immune system but what is inside is not functional, so we call it mis-encapsidation, and what this also does, it actually triggers a large amount of mis-encapsidation, so we can shift the equilibrium towards mis-encapsidation and that’s quite important for us.

And then the question is: is it really possible to find drugs that do things like this? And yes, our collaborators have shown in Leeds that this is possible. You can actually find compounds, and actually, these were licensed compounds already in this case, that can do that job of binding to those packaging signals and actually blocking those interactions. So, it is not just a mathematical model. It is really something that you can see happening in nature, and what is exciting for us about this is, when you think about the usual antivirus strategies you have got, they have the problem that the viruses, especially RNA viruses, are mutating very swiftly, and so they change their outside features that are important for the recognition of the drugs. But here, we are actually going against multiple dispersed features that have some common recognition motifs, and that is very powerful because they would not all vary at the same time, and we have also a better understanding why it is very difficult for the virus to vary those. So, it is actually an alternative to the key-lock principle and kind of helps us to get a leg up on viruses here.

So we were fortunate enough then, in joint efforts with our experiential colleagues to patent this. It was quite exciting for a mathematician to come from often extended groups and suddenly seeing that these things have implications for drug development, and being involved in that and seeing that story first-hand was really wonderful. So, we filed a patent and we are very much in the process at the moment of commercialisation. It was really nice for us to see. And we have identified now these so-called packaging signals for really important classes of viruses. I have named a few here, but there is many more by now, and it seems to work for really large classes of RNA viruses, which is fortunate. And it is, in a sense, also a new opportunity for vaccine design because, again, it is this idea of mis-encapsidation, triggering mis-encapsidation, and therefore triggering those shells that are triggering immune responses without actually containing genomic material inside.

Right, so what I also wanted to mention is we are working also with designers because the viruses obviously have an aesthetic element to them, and it is a very beautiful way of showing people how viruses work and how the mathematics can actually tell you something new.

Bryony Thomas is from Leeds, from the School of Design, or she’s now in the Mechanical Engineering, and she started off looking at surface tessellations, and we got together because she saw our tessellation of this cancer-causing virus that I covered at the beginning of the talk, and she started developing new tessellations as artwork, and then we felt, can we not get together and create new artwork that actually shows the dynamic aspects. So, for instance, what we are also working on in my group - and I did not have time to cover that – is to look at how these lattices can undergo structural rearrangements. Some viruses need to rearrange proteins in the container in order to create channels so this genomic material can be ejected, and that is one potential example for this. You see here ERAV – it is a horse virus – in the contracted and expanded state. That would be a mathematical model. And we had this kind of projection suite here in York, it is a 360 degree suite, where we are creating these models and you see them on all four walls simultaneously, and then our artwork that we did with two summer students last summer, that shows, in this case here, the expansion of such a capsid.

So, there are lots of interesting mathematical questions I did not have time to cover, which concern the rearrangements of those lattices, which are also of interest for material scientists in the community, and if you want to know more, we have the Festival of Ideas in June. There will be these exhibits and also there are movies to be seen.

Right, so I hope I could show you that mathematics can help to better understand open problems in virology and actually provide a new perspective that helps us to solve open problems. So I have shown you, starting with group theory, then going to tiling theory, using graph theory, how we can not only understand the outsides of those containers but actually also the correlation with genomic material inside, and how this impacts on the dynamics of formation of these containers and also I have brushed a little bit on the evolution of those containers, with regards to evolution of these packaging signals.

So, I should say mathematics has played a key role in all of this, but it is very much an exciting enterprise for me to be part of an interdisciplinary community and really have the dialogue with biologists, biophysicists, computational chemists, bioinformaticians. It is a really, really fun area to work with all these people, and therefore I would like to thank the people involved. This is my group here in York, so we are a larger group because obviously we are interdisciplinary and we need all these different aspects to tackle those questions. Then our collaborators, especially the experimental collaborators in Leeds, we are working for many, many years together. It is very much an integrated activity there, and also collaborators all over the world.

With this, I would like to thank you.

© Professor Reidun Twarock, 2015