From Department of Biosciences and Nutrition
Karolinska Institute, Stockholm, Sweden

ACTIVATION OF THE SPIKE PROTEINS OF ALPHA- AND RETROVIRUSES

Shang-Rung Wu

Stockholm 2009

Cover illustration - Artists view of the metastable, but still inactive spike complex of chaperon p62 and the fusion subunit E1 in Semliki Forest virus.

Viruses build up their infection machinery with great care to prevent premature release of their cell entry capacity. In its acid-sensitive fusogenic design, the fusion active E1 of Semliki Forest virus is protected from action during transport by chaperon p62. When installed in the virus, the precursor is cleaved and matures to E2. On this, the E3 sequence that locks the machinery is removed. The E2 is shaped like a supportive cap on a sliding rod, anchored, as E1, in the virus membrane. Instead of oil, this machine works best on acid, and a low pH will lubricate the parts. **Shang-Rung Wu** reveals by electron cryomicroscopy that the E3 domain of the precursor forms a cotter-like connection between the precursor and E1. This locks the acid responsiveness of the whole spike complex, as indicated in this drawing.

Published by Karolinska Institutet. Printed by Larserics AB

ISBN 978-91-7409-736-8

ABSTRACT

Enveloped viruses like human immunodeficiency virus type 1 (HIV-1) enter cells by fusing their membrane with that of the cell plasma membrane, while others, like influenzavirus and alphavirus enter via uptake into endosomes and fusion with the endosomal membrane. For this they carry spike proteins. The spikes are composed of three copies of a subunit pair. One subunit has membrane fusion activity, whereas the other one binds to a structure (receptor) present on the cell surface. It also controls (chaperones) the activation of the fusion active subunit so that fusion does not occur prematurely. The fusion subunit is anchored in the virus membrane at one end and carries a fusion peptide (fp) at the other one. While hidden in the native spike the fp becomes exposed after spike activation (trigger). The fusion subunit interacts via the fp with the target membrane in an extended conformation and then forces the viral and the cell membranes together for fusion by a backfolding reaction. This model is mostly based on biochemical and structural studies using isolated subunits. However, a full understanding of the spike activation mechanism can only be obtained by studying the complete spike, i.e the trimer of the subunit pair. This was the aim of my thesis work.

I used low dose electron cryomicroscopy (cryo-EM) to capture images of my objects that had been frozen in liquid ethane in their native and partly activated states. This procedure facilitates the analysis of the virus and the spikes in their intact form, free from e.g. staining artifacts. By computer aided processing of the particle images, the three-dimensional structures were obtained. For my studies I used an alphavirus, Semliki Forest virus (SFV), and two retroviruses, Moloney mouse leukemia virus (Mo-MLV) and HIV-1. The SFV spike is triggered by low pH in the cell endosome and the retrovirus spikes by receptor binding.

One dilemma for SFV is that its spike during biosynthesis has to pass acidic compartments. How can it avoid activation? The chaperone and the fusion subunits are called E2 and E1. E2 is made as a precursor p62, which is cleaved by cellular furin into E2 and E3. The p62-E1 complex is acid resistant. I analyzed an SFV mutant with uncleaved p62-E1 spikes by cryo-EM and found that the E3 portion of p62 forms an extra contact with E1, which can explain the acid resistance of the precursor spike. A subsequent question was how low pH alleviated the E2 chaperoning of E1. In the native spike antibody mapping showed that the E1 fp is protected by E2 in the tip region. Partial triggering by low pH resulted in subunit dissociation in the membrane proximal but not in the distal fp containing part of the spike. This suggests a tight regulation of the steps in the fusion process.

While the SFV spike was studied in the context of the whole virus the retrovirus spikes were solubilized and isolated as trimers. Cryo-EM showed that the retrovirus spikes were hollow cage-like structures where the three subunit pairs formed a common roof and floor and separated lobes on the sides. Analyses of a partially activated form of the Mo-MLV spike showed an outward rotation of the top domain of the chaperone subunit. This opened up the cage from above, most likely to allow the fusion subunit to reach out to the cell membrane. I conclude that cryo-EM offers a powerful approach to study virus activation.

Keywords: Alphavirus, Retrovirus, Fusion protein, electron cryomicroscopy, intermediate structure

To my beloved parents

LIST OF PUBLICATIONS

This thesis is based on the following papers.

I. **Wu SR**, Haag L, Hammar L, Wu B, Garoff H, Xing L, Murata K, Cheng RH. *J Biol Chem.* 2007 Mar 2;282(9):6752-62. The dynamic envelope of a fusion class II virus. Prefusion stages of Semliki Forest virus revealed by electron cryomicroscopy.

II. **Wu SR**, Haag L, Sjöberg M, Garoff H, Hammar L. The dynamic envelope of a fusion class II virus. E3 domain of glycoprotein E2 precursor in Semliki Forest virus provides a unique contact with the fusion protein E1. *J Biol Chem.* 2008 Sep 26;283(39):26452-60.

III. **Wu SR**, Sjöberg M, Wallin M, Lindqvist B, Ekström M, Hebert H, Koeck PJ, Garoff H. *EMBO J.* 2008 Oct 22;27(20):2799-808. Turning of the receptor-binding domains opens up the murine leukaemia virus Env for membrane fusion.

IV. **Wu SR***, Löving R*, Lindqvist B, Hebert H, Koeck PJ, Sjöberg M, Garoff H. The solubilized HIV-1 spike is a hollow cage-like structure with separated legs. (*manuscript*)
*the authors contributed equally to the study

Unpublished results will also be presented.

CONTENTS

LIST OF ABBREVIATIONS

3D	three-dimensional
Å	Ångström
ASLV	avian sarcoma and leukosis virus
ALV	avian leukosis virus
BHK	baby hamster kidney
BLV	bovine leukemia virus
BN-PAGE	blue native polyacrylamide gel electrophoresis
cryo-EM	electron cryomicroscopy
Env	envelope protein
ER	endoplasmic reticulum
Fr-MLV	Friend murine leukemia virus
GALV	gibbon ape leukemia virus
gp	glycoprotein
HA	hemagglutinin
HIV-1	human immunodeficiency virus type 1
HTLV-1	human T-cell leukemia virus type 1
IAS	isomerization arrested state
Mo-MLV	Moloney murine leukemia virus
MPMV	Mason-Pfizer monkey retrovirus
NC	nucleocapsid
PFT	polar Fourier transform
Rev-A	avian reticuloendotheliosis virus
SARS	severe acute respiratory syndrome coronavirus
SFV	Semliki Forest virus
SIV	simian immunodeficiency virus
SRV	squirrel leukemia virus
SU and TM	surface and transmembrane fusion subunits
T	triangulation number
TBE	tick-borne encephalitis virus
TM	transmembrane
VLP	virus like particle
VSV	vesicular stomatitis virus
wt	wild type

INTRODUCTION

Viral membrane fusion

Enveloped viruses are wrapped in a membrane that they obtained by budding from one of the host cell membranes. This envelope protects the viral capsid with its genetic cargo during the extracellular transport. At the target cell, the membrane barriers have to be passed to deliver the genome into the cell and start a new infectious cycle. The enveloped viruses do that by unwrapping through fusion between the viral and the target membrane. In this process the viral and cellular membranes merge so that the capsid can enter into the cell. During this process virus-essential and unique structures may be available for antiviral intervention or vaccine approaches, explaining the interest to reveal the intermediate structures of this activity.

Biological membranes are composed of polar lipids in a double layered sheet arrangement with their hydrophobic tails within, shielded from the aqueous environment by their polar head groups at the outer surfaces. This arrangement is thermodynamically stabile and fusion or merging of two such membranes is, from the energetic point of view, an unlikely event. Membranes constitute the principle of biological compartmentalization, why catalytic tools have evolved to manipulate the membranes in different ways. In enveloped viruses the fusion of the virus membrane with the target membrane is catalyzed by the viral fusion proteins.

For initiation of membrane fusion a contact need to be established that overcomes the repulsion between the meeting polar surfaces and creates a local distortion to allow the hydrophobic inner domains to physically come in contact and rearrange. This would be the trick performed by the viral fusion proteins (Chernomordik &Kozlov, 2003; Kozlovsky &Kozlov, 2002). In artificial systems it is observed that two meeting membranes may be joined in a “hemifusion stalk” where the contacting outer lipid leaflets of the two bilayer membranes merge, while the inner leaflets remain intact. This stage may be followed by another hemifusion intermediate, where the inner leaflets meet in a

diaphragm. Consecutively, the membranes merge into one continuous membrane bilayer and form a fusion pore, which then expands to complete the membrane fusion process (**Fig. 1**) (Chanturiya *et al.*, 1997; Chernomordik *et al.*, 1987; Kozlovsky &Kozlov, 2002).

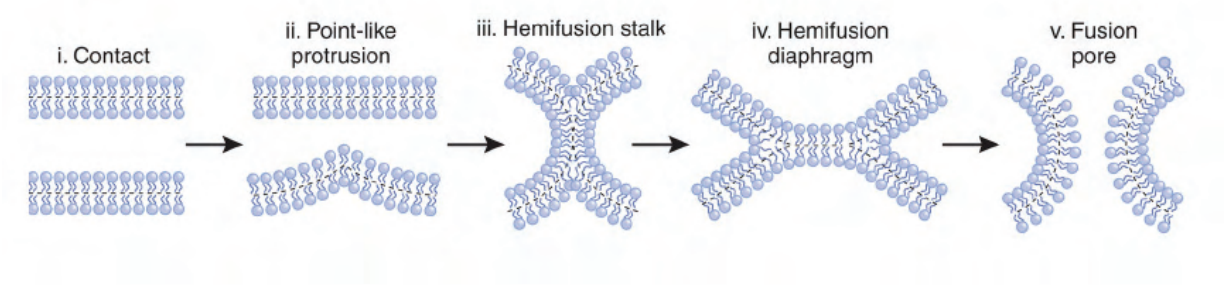

Fig. 1. *Model of membrane fusion through hemifusion pathway: (i) Pre-fusion encounter. (ii) A point-like membrane protrusion allows the membranes to come in close proximity. (iii) A hemifusion stalk with meeting membrane leaflets merged, but inner leaflets remaining continuous. (iv) Stalk expansion may yield a hemifusion diaphragm intermediate, with meeting inner leaflets. (v)Eventually, a fusion pore forms. (From (Chernomordik &Kozlov, 2008) with copyright permission from Nature Structural & Molecular Biology)*

The fusion process of enveloped viruses is catalyzed by their fusion proteins - The viral fusion proteins, also called spikes because of their morphology, are parts of the viral envelope. Typically, they are composed of dimers of two different subunits and form spike like protrusions around the virus particle. The fusion subunit has a transmembrane domain, anchored in the viral membrane, and carries a hydrophobic sequence (the fusion peptide, fp) to establish contact with the target membrane. Initially, viral fusion proteins are in an inactive state on the viral surface and their fusion peptides are hidden (**Fig. 2**). When activated, by interaction with a receptor on the cell surface, or by the acidic environment in the endosome, the viral fusion protein undergoes a sequence of conformational changes to expose the fusion peptide, catch the target membrane and establish fusion – The activated intermediate may be an extended trimer of the fusion peptide carrying subunit. In the vicinity of a target membrane the fusion peptide would penetrate the polar membrane surface and establish a “pre-hairpin structure”, linking the

viral and target membranes (**Fig. 2**). This releases further refolding - the extended fusion oligomer folds back on itself and brings the viral transmembrane anchors into close proximity to the fusion peptides that are inserted in the target membrane. The fusion protein ends up in a “trimer of hairpins” configuration in the merged membranes. The folding activity emanates from the metastable structure created during biosynthesis of the fusion protein. On trigger, this provides the energy for approaching, contacting and forcing the viral and cellular membranes into close proximity and establishes membrane fusion.

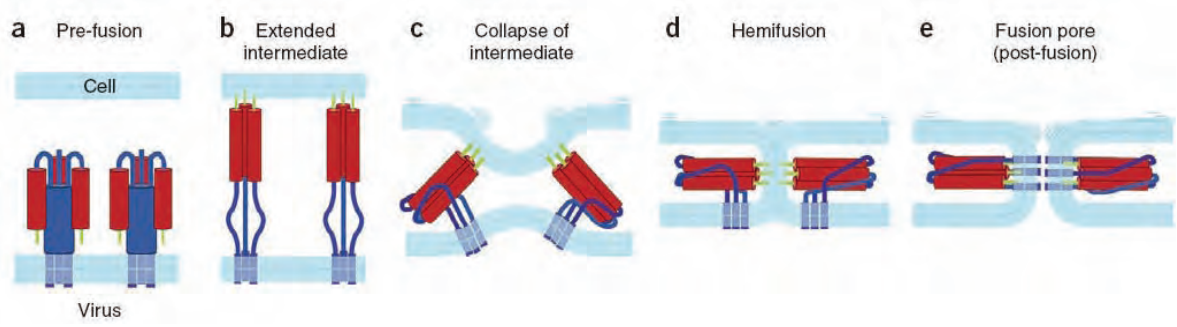

Fig. 2. *Sequence of events in membrane fusion promoted by viral fusion proteins. (a) The fusion protein oligomer is in the inactive pre-fusion state, with the fusion peptides (green) hidden. (b) When activated the oligomeric fusion protein opens up and extends the fusion peptide sequence to interact with the target membrane. This extended intermediate is also called the pre-hairpin structure. (c) Collapse of the extended intermediate proceeds to enforce (d) the two bilayers into contact. A hemifusion stalk forms. (e) The fusion proteins end up in a backfolded “trimer of hairpins” configuration. (From (Harrison, 2008) with copyright permission from Nature Structural & Molecular Biology)*

The viral fusion proteins

Atomic structures of fusion subunits of some fusion proteins and/or portions of them, in pre- and post-fusion forms have been determined. Although the fusion proteins are quite diverse in structure, they all mediate membrane fusion by forming a common membrane-embedded principally backfolded configuration - a “trimer of hairpins”. In this post-fusion structure the fusion peptide and the transmembrane domains of the fusion subunit are inserted in the merged membrane in close proximity. This indicates that they

may mediate fusion through a similar mechanical transformation. To date, three major classes of the fusion proteins are identified based on the distinct structure present at the post-fusion state. The class I fusion proteins are characterized by a central α-helical coiled-coil structure; the class II fusion proteins are dominated by β-sheet structures; and the class III fusion proteins have mixed secondary structures **(Fig. 3)** (Eckert &Kim, 2001; Skehel &Wiley, 1998; Weissenhorn *et al.*, 2007; White *et al.*, 2008)**.**

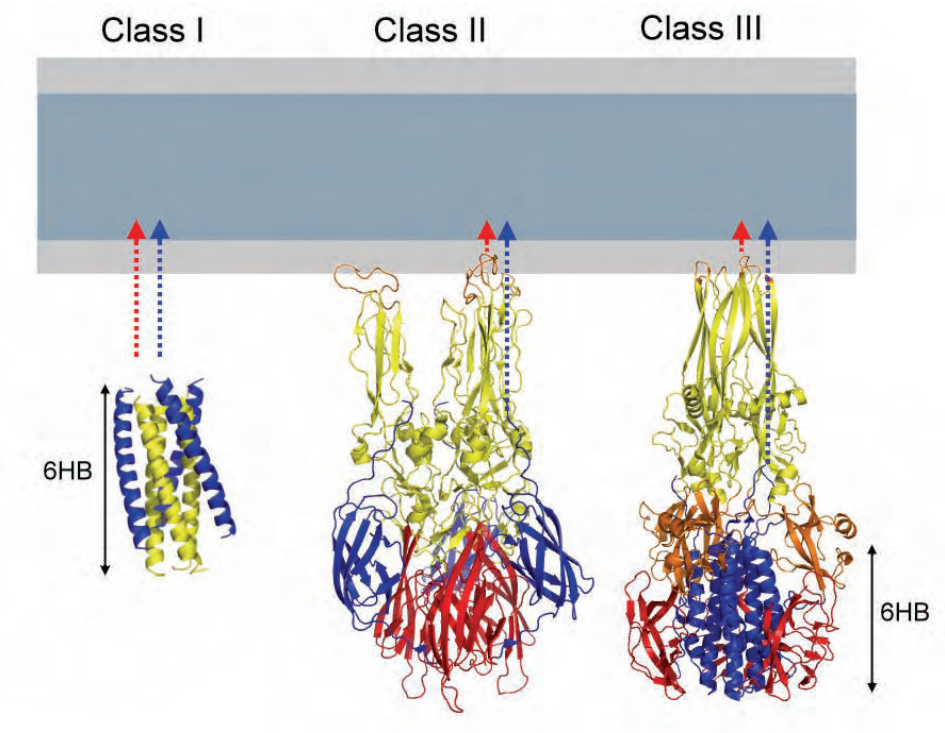

Fig. 3. *"Trimer of hairpins" structures of viral fusion proteins at their post-fusion state. The terminology "hairpin" is because the backfolded structure remotely resembles a hairpin. Ribbon diagrams of HIV-1 gp41 core structure (left), SFV fusion protein E1 (middle) and VSV protein G (right) represent the class I, II and III fusion proteins, respectively. The "six helix bundle" (6HB) is the defining feature in class I fusion protein, where the trimeric N-terminal helical coiled-coil forms a central core packed by outer C-terminal helices. A 6HB is also found in the membrane distal domain of the VSV G post-fusion structure. Both the membrane anchored fusion peptide (red arrow) and the transmembrane domains (blue arrow) are positioned close to each other as indicated in the figure. Structural elements are shown in different colors.*

Class I fusion protein

Fusion proteins of retroviruses, coronaviruses, orthomyxoviruses, paramyxoviruses and filoviruses belong to this group (Colman &Lawrence, 2003; Eckert &Kim, 2001; Skehel &Wiley, 1998; White *et al.*, 2008). The fusion proteins mature by proteolytic cleavage of a precursor, yielding a transmembrane fusion subunit with an N-terminally located hydrophobic fusion peptide sequence and a peripheral subunit. Their fusion subunits are trimers at both their pre- and post-fusion states and are featured by dominatingly α-helical configurations. The pre-fusion forms are metastable trimers, oriented perpendicularly to the viral surface, while the post-fusion conformations are highly stable trimeric hairpin structures, with a central 6HB (**Fig. 3**).

The hemagglutinin (HA) from influenza virus is the prototype of class I fusion proteins (Skehel &Wiley, 2000). HA is synthesized as a fusion incompetent precursor, HA0 which is then proteolytically cleaved into two subunits, the peripheral receptor binding subunit (HA1) and the transmembrane fusion subunit (HA2) with the fusion peptide at its N-terminus. In the pre-fusion state (at neutral pH), HA2 is masked by HA1 and the N-terminal fusion peptide buried within the molecule. The C-terminus extends through the viral membrane. HA is triggered for activation after virus uptake into the acidic endosome. Upon acidification, HA1 dissociates from HA2. This allows HA2 to undergo dramatic structural changes in its secondary and tertiary structure, which are shown in **Fig. 4** (Bullough *et al.*, 1994) and reviewed in (Skehel &Wiley, 2000; White *et al.*, 2008). In its pre-fusion state the HA2 is characterized by a long helix (yellow-green-blue in **Fig. 4**) that forms a coiled-coil in the spike trimer. At its N-terminal end the helix is linked to a shorter helix (red in **Fig. 4**) by a loop (orange in **Fig. 4**). This shorter helix connects to the fusion peptide, which in the trimer is buried between the subunits. Acidification changes the N-terminal loop into a helix resulting in an extension of the original helix with not only this region but also with the distal shorter helix. This relocates the fusion peptide about 100 Å, from its buried location to the tip of the helix. Furthermore, acidification changes a short region in the C-terminal part of the original HA2 helix into a loop (green in **Fig. 4**). This allows the distal part of the helix (blue in **Fig. 4**) and the flanking

C-terminal polypeptide to backfold and pack against the N-terminally extended coiled-coil. According to the fusion model of HA the relocation of the fusion peptide will allow it to interact with the cell membrane and the backfolding will approach the virus membrane to the cell membrane for fusion. Thus, in this model there is first the formation of a trimer of pre-hairpins where the HA2 fusion protein grabs the cell membrane with the fusion peptides and then it forces the virus membrane towards the cell membrane by bending its C-terminal part against the N-terminal part forming a trimer of hairpins. Through this conformational change the HA2 also changes from a metastable structure into a very stable one, which is characterized by a bundle of six helices as shown in **Fig. 3**, left panel.

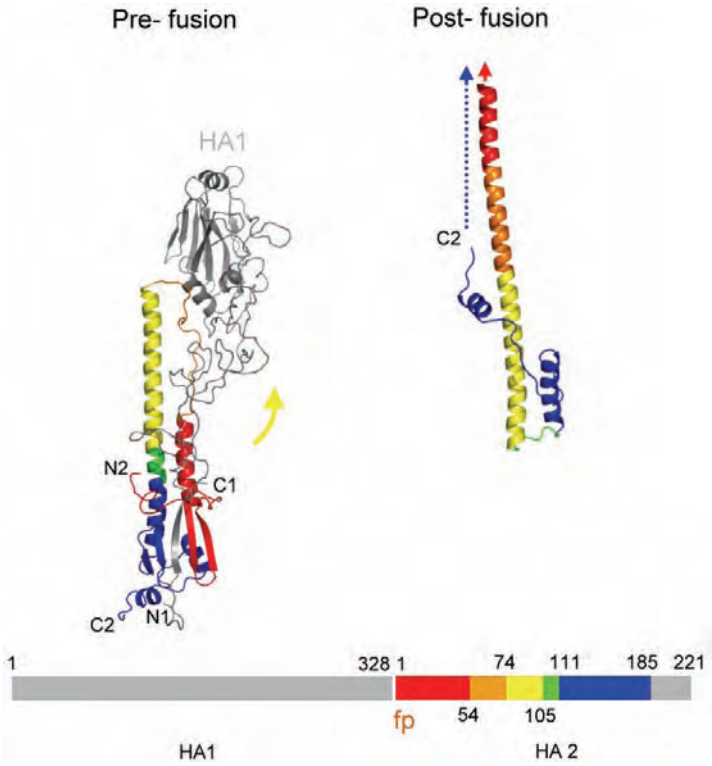

Fig. 4. *Acid induced structural rearrangements in the influenza virus subunitHA2. The structure of the ecto-domain of the native non-triggered HA1-HA2 subunit pair is shown to the left and the fully activated form of the HA2 subunit ecto-domain without its fusion peptide to the right as ribbon representations. The HA1 is gray and the HA2 multicolored. For the sake of clarity only one subunit pair of the trimeric complex is shown. A linear map of the HA polypeptide is shown below with corresponding colored regions and the position of the fusion peptide (fp) indicated. The N- and C-terminal ends of the subunits are indicated N1, N2, C1 and C2. The prominent feature of HA2 in the native spike trimer is a central coiled-coil. The acid induced change involves a loop-to-helix transition in the orange region N-terminal to the long helix and a helix-to-loop transition in the green C-terminal part of the long helix. As a result there is large relocation of the red helix, flanking the fusion peptide, onto the top of the long helix and a backfolding of the most C-terminal part of the long helix and parts distal to it. The blue and red arrows in the post-fusion structure model indicate the missing transmembrane anchor and fusion peptide. The PDB accession numbers for pre- and post-fusion forms are 2HMG and 1QU1.*

Class II fusion protein

Alphaviruses and flaviviruses carry class II fusion proteins. They are made as trimers of a pair of subunits, both of which are transmembrane proteins. One is called the transmembrane fusion subunit and the other the chaperone subunit. The former carries the membrane fusion activity, whereas the latter controls this activity via its proteolytic

processing in the virus producing cell (Salminen *et al.*, 1992; Yu *et al.*, 2008). The chaperone subunit also binds to the cell receptor. Structural determinations of the fusion subunits of class II fusion proteins, such as the E protein of tick-borne encephalitis virus (TBE, a flavirirus) and E1 protein of Semliki Forest virus (SFV, an alphavirus) show that they are elongated three-domain molecules, in which the N-terminal domain I is in the middle, the finger-like domain II on one end and the immunoglobulin (Ig)-like domain III at the other end (**Fig. 5**, left panel) (Lescar *et al.*, 2001; Rey *et al.*, 1995). The internal fusion loop is located at the free tip of the finger-like domain. Most of the polypeptide is folded as β-sheet. So far it has not been possible to determine the atomic structure of the chaperone subunit or the heterodimeric subunit complex. However, cryo-EM reconstructions of alphavirus particles show that the chaperone subunits are positioned centrally and the fusion subunits peripherally in a trimeric spike (Zhang *et al.*, 2002). The chaperone subunits stand upright and protect from above the fusion subunits, which are tangentially oriented relative to the viral membrane. Interestingly, in mature flavivirus, i.e. virus where the chaperone subunit has been proteolytically processed, the fusion subunits are lying flat along the membrane plane as homodimers with the fusion peptides buried at the interdimeric interphase (Kuhn *et al.*, 2002; Rey *et al.*, 1995).

The class II fusion proteins are triggered by low pH in cell endosomes (Marsh *et al.*, 1983; White &Helenius, 1980). This dissociates the dimeric subunit interactions and allows the fusion subunits to rise up from the virus membrane to grab the target membrane, trimerize and fuse the virus membrane with that of the cell through a backfolding reaction (Allison *et al.*, 1995; Klimjack *et al.*, 1994; Wahlberg *et al.*, 1989; Wahlberg *et al.*, 1992). The atomic structure of the low pH forms of the alpha and the flavivirus fusion subunit ecto-domains have been determined (Bressanelli *et al.*, 2004; Gibbons *et al.*, 2004b; Modis *et al.*, 2004). These show that the subunits trimerize via domains I and II and that the domain III and the stump of the stem connecting to the viral membrane anchor have folded back on the groove formed between the domain II of adjacent subunits (**Fig. 5**, right panel). Thus typical for class II fusion subunit refolding is the backfolding reaction

as in class I fusion subunits, but different from these there is very little refolding on the level of secondary structure.

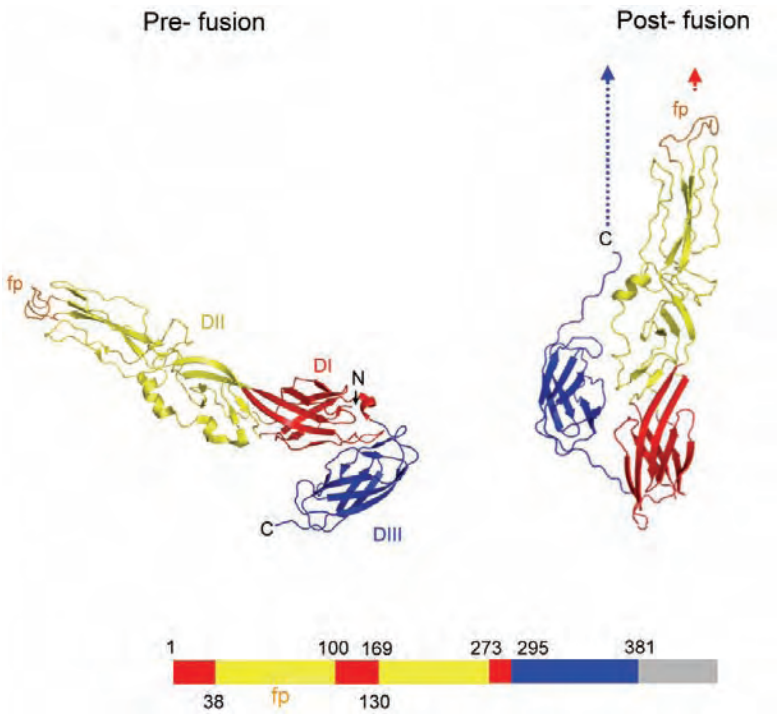

Fig. 5. *The acid induced conformational change of the fusion subunit E1 of SFV. The fusion subunit is shown in its pre- and post-fusion states. Only one monomer is shown for both states. The fusion protein is divided into three domains, a β-barrel domain I (red) at the N-terminus, a finger-like projecting domain II (yellow) containing the fusion peptide (fp, orange), and an Ig-like domain III (blue) at the C-terminus. N- and C-termini are indicated by N and C. The fusion peptide and the transmembrane domain, which both are missing in the post-fusion structure model, are indicated by red and blue arrows, respectively. The bottom panel shows the schematic domain architecture of fusion subunit E1 as a linear map. The position of the fusion peptide is indicated. The PDB accession numbers for the pre- and post-fusion forms are 2ALA and 1RER.*

Class III fusion protein

This newly identified class includes viral fusion proteins from rhabdoviruses and herpesviruses. They are classified as a new class because they mix features of class I and II. The G protein of vesicular stomatitis virus (VSV) is an example of a class III fusion protein. The atomic structure of this trimeric protein has been determined at the pre- and post-fusion states (Roche *et al.*, 2006; Roche *et al.*, 2007). Class III fusion proteins are four- or five-domain molecules which have a molecular architecture distinct from class I and class II fusion proteins (**Fig. 6**). However, similar to class I fusion proteins, the class III fusion proteins orient vertically to the viral surface. Upon low pH exposure, which is the trigger for the VSV G protein, a series of conformational changes occur, including the reposition of the fusion peptide sequences towards the target membrane and backfolding of the protein so that the C-terminal domain comes close to the N-terminal domain thereby bringing the viral and target membranes together for fusion. This post-fusion structure displays a central trimeric N-helical coiled-coil core onto which three C-helices are packed forming a six helical bundle, a hallmark of class I proteins. However, most of the G protein domains are made of β-sheets and the fusion peptide sequences are located internally revealing the similarity with class II fusion proteins. Unlike class I and II fusion peptides which are well conserved, class III fusion peptides appear to be more complex in their function (Backovic &Jardetzky, 2009). Notably is also that the VSV G protein is not maturing by cleavage into a metastable protein, the folding of which is triggered by low pH into a stable state. Instead it changes its conformation reversibly between the pre- and the post-fusion states depending on the surrounding pH.

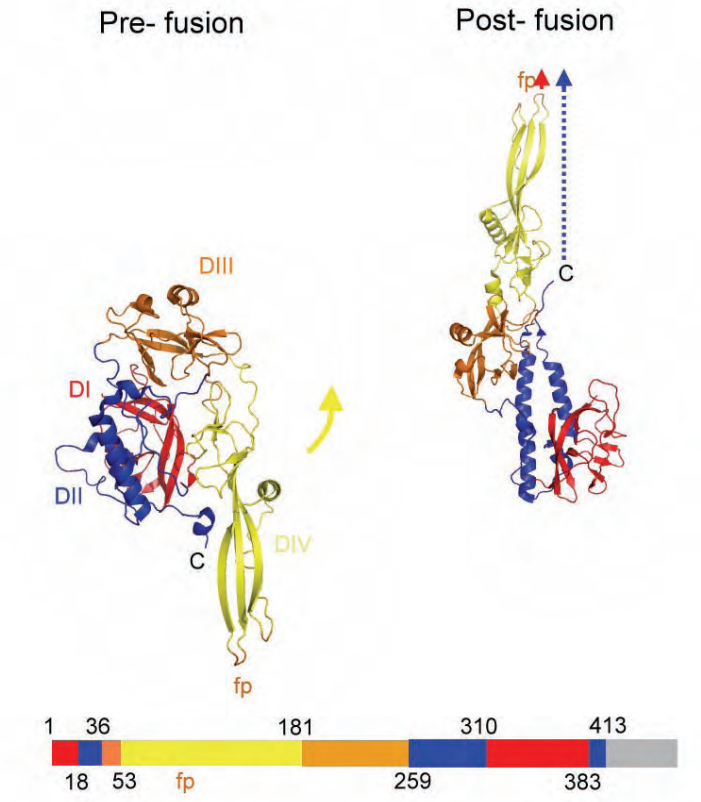

Fig. 6. *Acid induced conformational change in the VSV G protein. Left panel shows the pre-fusion and the right panel the post-fusion structure with protein domains colored. Bottom panel shows protein domains on a linear map of G. Domain I is colored red, domain II, which links to the membrane anchor is blue, domain III is orange and domain IV, which contains the fusion peptide loop at its tip, is yellow. The position of the fusion peptide is indicated. Activation of this spike involves a change in the domain II-IV hinge region that results in the relocation of the domain IV as indicated by the yellow arrow. Through this change the G protein might grab the cell membrane. Further, the domain II is partly refolded so that a C-terminal helix is formed. This packs itself anti-parallel into the grooves formed by the trimerizing central helices of this domain. The reorganization orients the stump of the stem leading to the membrane anchor into the same direction as the elongated domain IV with the fusion loop (indicated by arrows in the right panel). Most likely this backfolding reaction leads to membrane fusion. The PDB accession numbers for pre- and post-fusion forms are 2J6J and 2CMZ.*

Triggering of spike protein activation

The metastable spike

The membrane fusion proteins, i.e. the spike proteins of the enveloped viruses are made in the endoplasmic reticulum (ER) of the cell, where they become glycosylated and also trimerized. The complexes are then routed to the cell surface via the Golgi complex. Although the spike proteins are made in a stable form they usually mature during transport into a metastable form, which is incorporated into the virus particle. This form has obtained the potential to refold further into a stabile form along a kinetic folding pathway (Carr *et al.*, 1997). The refolding requires that the energy thresholds that are preventing it from folding are overcome. This is possible by spike protein interactions during receptor binding, low pH induced conformational changes of the spike in the cell endosome or by proteolytic processing of the spike by the target cell.

Spike maturation

The most notable maturation event that all spikes undergo, but those of the rhabdo viruses, is a proteolytic cleavage event. This is usually done by cellular furin in the trans Golgi or in compartments that are subsequently passed by the spike on its way to the cell surface. The cleavage releases the transmembrane fusion subunit of class I spike proteins from the peripheral one. In the case of the class II spike proteins furin cleaves the chaperon subunit that associates with the transmembrane fusion subunit in the heterodimer. As many of the class I and II spike proteins are triggered for activation by low pH, the fact that the furin compartment(s) also are acidic increases the risk of premature spike activation. How the virus can overcome this problem has been studied in detail for the dengue virus, a flavivirus with a typical class II spike protein (Li *et al.*, 2008; Yu *et al.*, 2008). In this case the whole virus is assembled in the ER and the spikes mature on the viral surface during transport of the particle out of the cell. The acidic milieu of the furin compartment causes a structural change of the spike heterodimer, where the chaperone protein prM protects the fusion peptide-carrying part of the membrane fusion protein E. The change facilitates prM cleavage by furin, but the acidic milieu prevents the release of the peripheral prM fragment until the virus with the spikes is removed from the acidic compartments. As the

fragment inhibits low pH triggering of the spike it is possible for the virus with the spikes to safely undergo furin cleavage in the acidic compartment.

Receptor induced triggering

In the mature form of the class I spikes it is the interactions between the peripheral and the transmembrane fusion subunits that suppress the activation, i.e. the refolding of the latter subunits. In the case of many retroviruses the receptor binding of the peripheral subunit induces sufficient changes in the subunit interaction to activate the transmembrane fusion subunit. As a consequence these retroviruses fuse at the cell surface. For instance HIV-1 interacts subsequently with the CD4 and a chemokine receptors and then fuses at the cell surface. Mo-MLV interacts with a basic amino acid transporter at the cell surface and then fuses here. Typically, cells expressing these spikes are also able to fuse with target cells carrying the corresponding receptors.

Low pH mediated triggering

In the case of the mature class II spike proteins it is the interaction between the transmembrane fusion subunit and the chaperone subunit that control activation (Lobigs *et al.*, 1990a; Wahlberg *et al.*, 1989). The effect of receptor binding on the activation has not been studied because the receptors are still not well characterized. However, the major trigger is low pH, which the virus encounters when it is taken up by endocytosis into the cell (Marsh *et al.*, 1983; White &Helenius, 1980). Consequently, these viruses will fuse in the endosome. The low pH will protonate in particular His residues and thus introduce a positive charge. Such residues at critical sites of subunit interactions, like those between the chaperone and the fusion subunits, could affect triggering. One possible advantage with the endosomal pathway is that it will release the viral genome deeper into the cell cytoplasm, by passing e.g. the cortical actin layer.

Triggering by receptor binding and low pH

In some cases virus spike triggering by receptor binding or low pH alone is not sufficient, but a combination is required. This situation was found for the spike protein of the

alpharetrovirus, avian leukosis virus A, ALV-A, a class I membrane fusion protein (Mothes *et al.*, 2000). Receptor binding to the disulfide linked peripheral subunit apparently caused the formation of a trimeric pre-hairpin structure of the transmembrane fusion subunit (Netter *et al.*, 2004). This was able to interact with a target membrane via its fusion peptide, but could not refold further into a trimeric hairpin structure if not acidified. Another combination was found to be required for triggering the activation of the class I spike protein S of the severe acute respiratory syndrome coronavirus (SARS) (Simmons *et al.*, 2005). This spike is not cleaved by furin in the producer cell. Binding to the angiotensin converting enzyme 2 receptor probably only triggered an intermediate form of the spike, because fusion required uptake into the endosome. Here it was not the low pH *per se* that caused the final activation but cleavage of the spike by the acid dependent cathepsin L enzyme.

Artificial spike triggering

The elucidation of the triggering conditions for various virus spikes has offered the possibility to mimic these conditions and trigger the spikes of a virus preparation *in vitro*. This possibility has been very useful when studying the conformational transition that the spikes undergo during the activation process. For instance in cases when two receptors are used like in HIV-1 or when receptor binding and low pH are used like in ALV-A the first triggering event should yield an intermediate structure of the spike that is not yet fully activated. However, in practice spike triggering by this approach requires that the receptor fragment can be isolated in a soluble and active form, which was the case for both examples. Also when low pH is used as the sole triggering event like in alphavirus, treatment of the virus with buffers of suboptimal pH for fusion might result in corresponding intermediate forms of the alphavirus spikes. Finally it should be noted that several class I spike proteins do not necessarily need the natural receptor mediated triggering at all, but unspecific protein perturbation treatment will do as well. Thus, for instance heat or urea treatment will trigger the class I spikes of Mo-MLV, ALV and influenza virus (Carr *et al.*, 1997; Smith *et al.*, 2004; Wallin *et al.*, 2005). The structural

transformation pathway of the spike appears at least in the case of Mo-MLV to be correct as it was shown to mediate fusion.

Alphavirus

To introduce my study on spike activation in Semliki Forest virus I will here give a general introduction on the alphavirus biology and structure. A comprehensive recent review of the major events in alphavirus infection, entry, replication, assembly and budding is given by Jose, Snyder and Kuhn (Jose *et al.*, 2009). See also the classical review by Strauss & Strauss (Strauss &Strauss, 1994).

Pathology

Alphaviruses with the ability to cause human disease are present worldwide. Contrary to the other genus of the *Togaviridae* - the rubivirus with humans as the only known host - the alphaviruses are replicating in arthropod insects as well as in vertebrate animals, including humans, in rodents, birds, and even fish. Vector transmission by mosquitoes is common. The viruses gain entry into the bloodstream and further into the central nervous system where it is able to grow and multiply within the neurons. Virus infection causes fever, rash and/or arthritis or even encephalitis.

Genome and proteins

The alphavirus is a genus of small enveloped viruses (~70 nm in diameter) of the *Togaviridae* family[1]. They enter the cell by receptor mediated endocytosis in a clathrin-dependent manner (Helenius *et al.*, 1980; Marsh *et al.*, 1983). The genome is a linear single-stranded RNA (ssRNA) molecule of positive polarity. The 5'-end has a methylated nucleotide cap and the 3'-end a poly-A tract, i.e. the viral genome will function as an mRNA in the cytoplasm of a cell. It contains two open reading frames (ORFs); the first encodes the non-structural proteins (the viral enzymes) and the second the structural proteins that build up the virion. The non-structural proteins form a

[1] http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/

replication complex, which transcribe the genome into a copy of negative polarity. With this as a template new full length genomes and a subgenomic mRNA species, the 26S RNA, are synthesized. The structural proteins are translated as a polyprotein precursor from the 26S RNA in the order C-pE2-6k-E1. (**Fig. 7**), where pE2 (called p62 in SFV) is the precursor of E2 and E3.

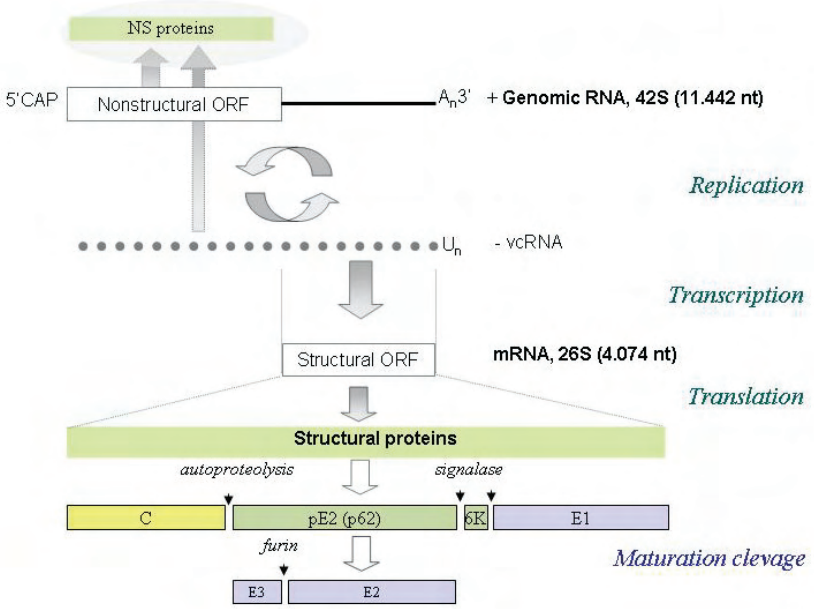

Fig. 7. *Alphavirus genome and the structural proteins. In the infected cell the genomic (42S) RNA is directly acting as a messenger for a set of non-structural (ns) proteins encoded by the non-structural open reading frame (non-structural ORF). The ns proteins enable replication of the genomic RNA and transcription of sub-genomic mRNA. The structural ORF, which occupies the 3' terminal third of the genome, is separately transcribed into a 26S mRNA. From this the structural proteins are translated in the order indicated. The mature virion consists of the structural proteins, C, E1, E2, and E3, of which the external E3 may be released at an early stage after budding. A small portion of the membrane spanning 6K may remain in the virion.*

General structure

SFV and other alphaviruses have been studied by cryo-EM, and virion structures with resolution levels down to 9 Å have been reported (Mukhopadhyay *et al.*, 2006). A cryo-EM derived structure of SFV at 15 Å is shown in **Fig. 8**. The center of the virion is the nucleocapsid (NC) where a single copy of the RNA genome is complexed by 240 copies of a capsid protein (C) (Zeng *et al.*, 2002). The NC is spherical with a diameter of ~40 nm and the C proteins are arranged in pentagonal and hexagonal capsomers in an icosahedral lattice with T=4 symmetry (**Fig. 8**, right panel). The NC is surrounded by the viral envelope; a lipid bilayer containing 80 glycoprotein spike complexes that are also arranged in T=4 icosahedral symmetry (**Fig. 8**, left and middle panels). The spikes are trimeric assemblies of the viral membrane glycoprotein; a heterodimer of two subunits called E1 and E2. The E1 carries the membrane fusion activity of the complex and is chaperoned by the receptor binding subunit E2. The spike forms a shell that connects the spikes on the membrane surface, a spike stalk and a head. Below the protruding spikes and the shell, the E1 and E2 form a common limb structure that approaches the virus membrane (**Fig. 8**, middle panel) (Mukhopadhyay *et al.*, 2006). Both subunits traverse the lipid bilayer, and under the membrane the C-terminal domain (tail) of E2 form contacts with the C protein in the NC (Skoging &Liljestrom, 1998; Skoging *et al.*, 1996). The legs, i.e. the limb and tail regions, of a single spike will connect to C proteins in three different capsomers creating a very stable structure (**Fig. 8**, compare the left and right panels).

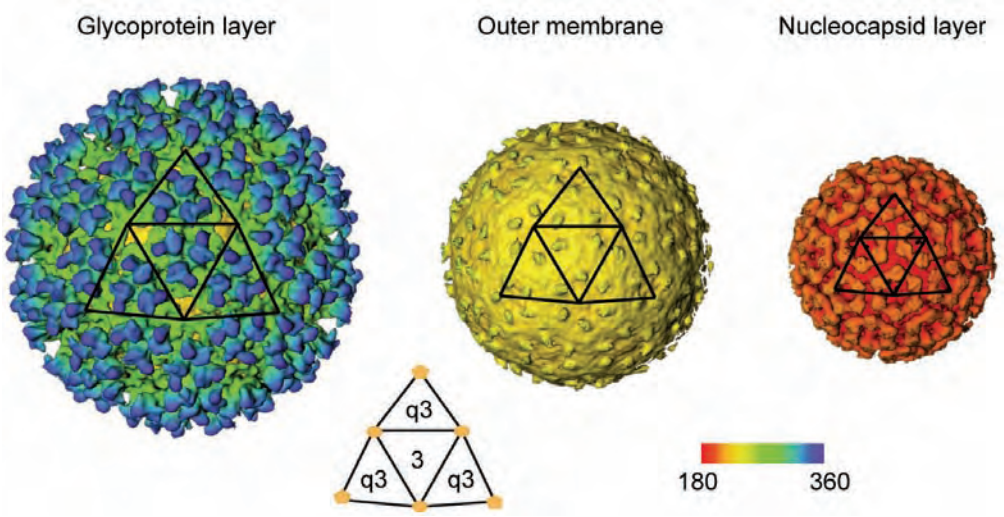

Fig. 8. *Cryo-EM reconstruction of Semliki Forest virus at 15Å resolution. The virion has icosahedral symmetry and is viewed along one of the three fold symmetry axes. The left panel shows the glycoprotein layer with its trimeric spike protrusions (blue) rising above the spike shell layer (green) that cover most of the membrane (yellow). The middle panel shows the outer surface of the membrane and the membrane proximal so called limb region of the spikes as small protrusions. The right panel shows the outer surface of the nucleocapsid with the C proteins arranged in pentagonal or hexagonal rings. The color code gives radial distance (Å).*

Spike assembly involves an inactive form which is activated by proteolytic cleavage

Upon cell entry the virus is exposed to low pH in the endosomes. This will cause the fusion protein to refold and catalyze a membrane fusion reaction that will merge the viral and the endosomal membranes and release the NC in the cytoplasm. The genome is uncoated from C protein and then replicated and transcribed. Further, the polyprotein precursor, C-pE2 (or p62)-6K-E1, is translated from the 26S RNA (**Fig. 7**). As soon as the translation of the C protein is completed it will cleave itself off the growing polypeptide chain in an autocatalytic reaction. This uncovers a signal peptide at the N-terminus of pE2 (or p62) and the rest of the polyprotein is co-translationally translocated into the ER (Garoff *et al.*, 1990). The signal sequence is glycosylated and retained as part of pE2 (or p62) (Lobigs *et al.*, 1990b). The proteins that follow; the 6k viroporin, which is found in minor amounts in the mature virus (Gonzalez &Carrasco, 2003; Melton *et al.*, 2002) and the E1 are released sequentially by signalase cleavage (**Fig.**

7). The membrane bound glycoproteins pE2 (or p62) and E1 form a heterodimer already in the ER, with pE2 (or p62) acting as a chaperone for E1, protecting both its metastable fold and hydrophobic fusion peptide (residues 83 to 100) (Andersson *et al.*, 1997; Barth *et al.*, 1995; Ziemiecki *et al.*, 1980). The heterodimer is transported via Golgi to the cell surface. During the passage the carbohydrates become modified and at a late stage, in the trans-Golgi or trans Golgi network, the pE2 (or p62) is cleaved by cellular furin into E2 and E3 (de Curtis &Simons, 1988; Garoff *et al.*, 1994; Jain *et al.*, 1991; Salminen *et al.*, 1992). Mature E1-E2 heterodimers, in the form of trimeric spike complexes, are incorporated into virus particles that bud off from the plasma membrane of the infected cell. The peripheral E3 may be released from the virus during or shortly after budding (Lobigs &Garoff, 1990; Salminen *et al.*, 1992; Tubulekas &Liljestrom, 1998). Both the chaperon arrangement and the furin cleavage are essential for correct assembly into virions and function of the alphavirus spike (Andersson *et al.*, 1997; Bron *et al.*, 1993; Kielian *et al.*, 1990).

Spike structure

The structures of both pre- and post-fusion states of the E1 subunit have been solved at atomic resolution and improved our understanding of the fusion process (Martin *et al.*, 2009). Except for the helical transmembrane domains the alphavirus spike is dominated by β-sheet secondary structures. This holds true both for the E1 subunit, for which the atomic structure of the external domain, E1*, has been solved (Lescar *et al.*, 2001; Roussel *et al.*, 2006), and as judged form the sequence, also for the E2 subunit. The crystal structure of the E1 ecto-domain was fitted into a cryo-EM derived density map of the virus and this showed that E1 is the main component of the shell region of the viral envelope (Lescar *et al.*, 2001; Mancini *et al.*, 2000; Wu *et al.*, 2007). From here it raises diagonally forming the sides of the spike. Subtraction of the E1 parts from such a fitting demonstrated the contribution of the E2 subunit (Mukhopadhyay *et al.*, 2006; Pletnev *et al.*, 2001; Zhang *et al.*, 2002). The E2 appears as a structure that rises on the inside of E1 in the spike, and like a bow covers the E1 fusion peptide at the top of the complex. The

pre- and post-fusion structures are discussed in the section on fusion proteins above and the SFV E1 protein in pre- and post-fusion configuration is shown in **Fig. 5**.

Fusion activation

The fusion function of an alphavirus resides in the E1 subunit. As discussed in the section on membrane fusion, it is required that the E1 fusion peptide is inserted into the target membrane. The viral and the target membranes are then approached by the formation of the homotrimeric post-fusion hairpin structure of E1. By this the organization of the two membranes will be distorted enough to first mix their outer leaflets and then form a pore (**Fig. 1**). The process is controlled by the E1 interaction with the E2 protein. The trigger to release the E1 from E2 and activate its fusion function is the low pH encountered in the endosomes upon virus entry. Experiments, both *in vivo* and *in vitro*, have shown that low pH triggers a series of conformational changes that convert the trimeric complex of E2-E1 heterodimers into a homo-trimeric E1 (Fuller *et al.*, 1995; Haag *et al.*, 2002; Wahlberg *et al.*, 1992; Wahlberg &Garoff, 1992). However, the details of this transformation and how the E2 may control the E1 is still largely unknown. The identification of the structural interplay between the two subunits in the spike upon fusion activation is the focus in my work (See the result section and papers I and II).

Considerations on the metastable construct of the virion structure

The folding of the polypeptide chain is the result of the local environment during the process, as is the stability of any possible polypeptide conformation. It is therefore possible for one polypeptide to guide the folding of another. Such polypeptides are referred to as chaperons and they may protect an otherwise unfavorable configuration from release into a more stable structure. In the envelope of the alphaviruses, the pE2 (or p62), precursor of the subunit E2, seems to take on such a chaperon function. In this way a metastable construction is created, which upon the correct trigger may be released to provide the mechanical energy for drastic reconfigurations during virus entry. The demands of this construction form the architecture of the whole virion, and in that sense the virus is one giant molecule. It is conceivable that a disturbance in one part of such a

construction would affect the structure as a whole and that the energy stored may be released in a cooperative action, when required. In other words, the strategy by which the virus assembles itself for transport to a new host cell is crucial for its capacity to enter and infect that cell.

Therefore, the study of the whole virion is important to understand the dynamic aspects of the virus architecture and its impact on the infection process. Although the major driving force of the fusion process was revealed when the pre- and post-fusion structural forms of the fusion subunit E1 were solved, the contacts that guide this transformation and how they change are still obscure. Furthermore, the nature of the protection against acid induced triggering that is provided by the E3 part of pE2 (or p62) has not yet been demonstrated. Therefore, my aim has been to start filling this gap in knowledge.

Retrovirus

Retroviruses like Mo-MLV and HIV-1 belong to the viral family *Retroviridae.* They are ssRNA viruses that replicate in a host cell using the viral reverse transcriptase to produce dsDNA from its RNA genome. The dsDNA is inserted by the viral integrase into the chromosomal DNA of the cell and then used for transcription of new viral RNA genomes and mRNAs. Retroviruses have infected animals during thousands of years and when inserted into germ cell DNA they have stayed in the linage. This has given rise to the large amounts of endogenous retroviral DNA found in the cell chromosomes. Retroviruses can pick up cell growth regulating genes and mutate these so that the expressed protein is constitutively active and then transmit cancer between animals. Also the mere fact that they integrate into the chromosomal DNA can cause malignancy through insertional mutagenes.

Like alphaviruses the retroviruses are enveloped by lipid bilayers that carry protein spikes. These are often referred to as the envelope protein, Env (e.g. Mo-MLV), or sometimes the glycoprotein, gp (e.g. HIV-1). However the retrovirus spikes are apparently not symmetrically organized in the membrane as is the case with the spikes of the alphavirus.

Furthermore, the internal protein, the group specific antigen or simply Gag, is cleaved in the virus into several pieces, including a membrane binding matrix protein, MA, a genome encapsidating capsid protein, C and the RNA binding nucleocapsid protein, NC.

The integrated retrovirus genome, the so called provirus, is transcribed by host cell polymerase into fullsized and spliced mRNA species, which are transported from the nucleus to the cell cytoplasm for translation by ribosomes. The unspliced mRNA is also used as genomic RNA for new virus particles. Translation of the latter yields the Gag protein and by stop codon suppression or frame shifting (about 5%) a longer polyprotein, Gagpol, which includes units for the three retroviral enzymes, the protease (PR), the polymerase (Pol, i.e. the reverse transcriptase) and the integrase (Int). Translation of the spliced mRNA yields the Env.

Virus assembly is facilitated by the multivesicular-body sorting machinery and occurs at the cell surface (Morita &Sundquist, 2004). Here the Gag and the Gagpol proteins bind to portions of the internal side of the plasma membrane forming a lattice. This combines with a dimeric form of the viral RNA genome and with the trans-membrane Env. The assembly reaction drives also an outward curvature of the membrane so that a bud is formed. This matures into a free retrovirus particle by first forming a stalk-like connection to the cell and then undergoing a membrane fission event. The Gag and the Gagpol proteins are targeted and attached to the plasma membrane by myristylation of their N-terminal Gly residue and a patch of positively charged residues in the MA unit. The Env protein is made as a precursor in the ER of the cell and associates into trimers (Hunter, 1997). These are transported via the Golgi complex to the cell surface. At a late stage during the transport the Env precursor is cleaved by cellular furin into the peripheral, N-terminal SU and the C-terminal TM subunit (Hallenberger *et al.,* 1992). This cleavage potentiates Env for further structural changes that are needed to activate the fusion function. It is now in a typical metastable state of polypeptide folding.

In the free retrovirus particle the viral protease cleaves the Gag and Gagpol precursors to the mature proteins MA, CA, NC, PR, Pol and Int. In some retroviruses the protease additionally cleaves a short peptide from the internal endo-domain of the TM subunits (e.g. Mo-MLV and Mason-Pfizer monkey virus, MPMV). The latter cleavage further potentiates Env for fusion activation (Brody *et al.*, 1994; Rein *et al.*, 1994). The SU subunits target the virus to receptors on the surface of uninfected cells. Sometimes the virus binds first to one receptor, which has its binding domain far from the cell membrane and then to another one, which has its binding domain close to the cell membrane (e.g. HIV-1) (Feng *et al.*, 1996; Maddon *et al.*, 1986). The latter binding domains are often external loops of multispanning transport proteins of the plasma membrane (Wang *et al.*, 1991). The receptor binding usually activates the TM subunit to fuse the viral and the cell membrane through refolding as described for class I membrane fusion proteins.

The membrane fusion reaction introduces the internal components of the virus to the cytoplasm of the cell. In here the internal viral proteins serve important functions in transporting the viral genome to the cell nucleus, while the Pol reverse transcribes it into dsDNA (Gallay *et al.*, 1997). The transport must pass the actin cortex below the plasma membrane and also for some retroviruses the nuclear pore (e.g. HIV-1). In other cases the retrovirus waits in the cytoplasm until the cell dissolves its nuclear membrane during cell division (e.g. Mo-MLV). Recently several cellular factors have been described that restrict the replication of incoming retrovirus. These retrovirus restriction factors can hypermutate the viral genome like the APOBEC3G or bind to the viral capsid and distort its productive intracellular transport and disassembly like TRIM5alpha (Goff, 2004) .

The retroviruses have been classified into several groups as shown in **Table 1**. Although sharing general features they differ by their structure, replication and target cells. For instance the capsid of the gammaretroviruses is spherical, whereas that of lentiviruses is conical in shape. Further, the replication of lentiviruses is complex and involves several accessory proteins, like vpu, vif, vpr and nef, which are not used by the other retroviruses.

A list of retrovirus receptors is also shown in **Table 1**. As examples of class I membrane fusion proteins I have chosen to study the spike proteins of Mo-MLV and HIV-1.

Table 1. Classification of retroviruses

Genus	Example	Virion morphology	Genome	Receptor
Alpharetrovirus	Avian leukosis virus (ALV)	Central, spherical core	Simple	Low density lipoprotein (LDL) related protein[1]
Betaretrovirus	Mouse mammary tumor virus	Eccentric, spherical core	Simple	Transferrin Receptor 1[2]
Gammaretrovirus	Moloney murine leukemia virus (Mo-MLV)	Central, spherical core	Simple	Basic amino acid transporter (mCAT-1)[3]
Deltaretrovirus	Bovine leukemia virus (BLV)	Central, spherical core	Complex	Not known[4]
Lentivirus	Human immunodeficiency virus type 1 (HIV-1)	Cone-shaped core	Complex	CD4/chemokine receptors[5]

[1] (Barnard *et al.*, 2006), [2] (Ross *et al.*, 2002), [3] (Albritton *et al.*, 1989), [4] (Lavanya *et al.*, 2008), [5] (Doranz *et al.*, 1997)

The spike protein of Mo-MLV

The Env precursor of Mo-MLV is 665 amino acid residues long and can be divided into a major ecto-domain (571 residues), a transmembrane peptide (27 residues) and a minor endo-domain (33 residues) (Rothenberg *et al.*, 2001). The ecto-domain is heavily glycosylated. It contains 7 Asn-linked sugar units and several O-linked ones (Felkner &Roth, 1992; Geyer *et al.*, 1990). After trimerization in the ER, it is cleaved in the trans-Golgi by furin into a 70 kD SU (436 residues) and 15 kD TM subunit (195 residues) (Sjoberg *et al.*, 2006). The cleavage releases the fusion peptide at the N-terminal end of the TM subunit. However, it is not until the R-peptide, constituting the 16 C-terminal residues of the TM endo-domain has been removed by the viral protease in the newly assembled particle that the mature Env trimer or spike has been formed (Green *et al.*, 1981). This metastable protein complex is prevented from fusion activation, i.e. refolding

into its stable conformation, by the subunit interactions (Wallin *et al.*, 2004). Ultimately it is the binding of the spike by its SU subunit(s) to the receptor, a multispanning amino acid transporter (**Table 1**), that triggers spike activation and subsequent virus-cell membrane fusion (Wang *et al.*, 1991). The important question is how this occurs. In the SU subunit one can discriminate a receptor binding N-terminal domain (RBD, about 230 residues) that is linked to a C-terminal domain (about 160 residues) by a proline rich region (PRR, about 40 residues) (Kayman *et al.*, 1999; Lavillette *et al.*, 1998). The atomic structure of the RBD from the closely related Friend (F)-MLV has been solved and it revealed a bent finger-shaped molecule with receptor binding residues in the upper parts of the molecule **(Fig. 9)** (Fass *et al.*, 1996). Receptor bound RBD transmits an activation signal to the C-terminal domain (Barnett &Cunningham, 2001; Lavillette *et al.*, 2001). The conserved SPH_8Q-motif found in the N-terminal region of RBD appears to be important in activation signaling, since in particular the His_8 mutation abrogates activation but not receptor binding (Bae *et al.*, 1997). Surprisingly, the mutants can be complemented by receptor bound wt RBD that has been added *in trans* (Lavillette *et al.*, 2000). This suggests that activation signal transmission requires the establishment of a new RBD-C-terminal domain interaction using the conserved His containing region in RBD either directly or indirectly. Further studies using PRR swapping between different MLV strains have suggested that the PRR modulates the activation signaling from the receptor triggered RBD to the C-terminal domain (Lavillette *et al.*, 1998). The way by which this SU activation is further transmitted to the TM has remained elusive until recently, when an interesting clue was found. The SU and TM are disulfide linked and it was shown that upon receptor binding the linkage was rearranged into an intrasubunit disulfide isomer in SU (Wallin *et al.*, 2004). This released SU from TM so that it became possible for TM to refold and activate membrane fusion. The rearrangement was facilitated by the fact that the intersubunit disulfide Cys of SU was part of an isomerization active $C_{336}XXC$ motif residing in the C-terminal domain. In this motif the other Cys residue carried a free thiol, which became activated (deprotonated) by receptor binding to attack the intersubunit disulfide and cause the rearrangement (**Fig. 10**).

Apparently the RBD induced activation of the C-terminal domain, reorganized the locale of the free thiol in the CXXC-motif in a way that decreased its pK_a value.

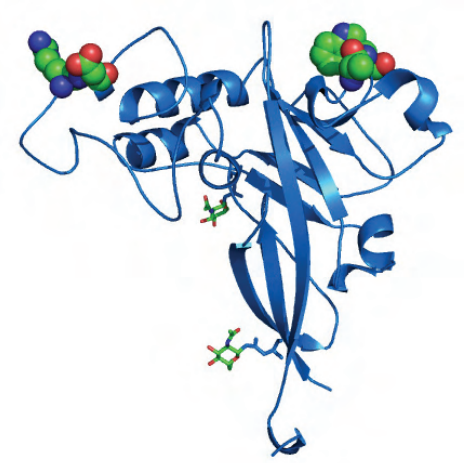

Fig. 9. *Structure of Fr-RBD (PDB: 1AOL). Residues in the two receptor-binding sites (front site: Arg85 and Asp86; back site: Trp144) are shown as spheres and the first residue of the two sugar units as sticks.*

The TM refolding of Mo-MLV is thought to follow the pre-hairpin model of Influenza HA. The TM ecto-domain contains the predicted helical regions in the N- and C-terminal parts and these are separated by a conserved motif, $C(X)_6CC$, where the last Cys residue is disulfide linked to the CXXC motif in SU **(Fig. 10)** (Kobe *et al.*, 1999). When this disulfide is rearranged by the receptor induced isomerization reaction, the N-helices of the three TM subunits are thought to reorganize into a coiled-coil that presents the fusion peptides at the TM tips, to the target membrane. The N-helix contains a typical heptad repeat trimerization signal with hydrophobic residues in positions *a* and *d* in a helical wheel depiction. This pre-hairpin structure is then, according to the model, converted into a hairpin structure by the association of the C-helices into the grooves of the central coiled-coil in an anti-parallel orientation. As the TM subunits are bound to the cell membrane by their fusion peptides and to the viral membrane by their transmembrane segments the latter event will approach the two membranes for fusion.

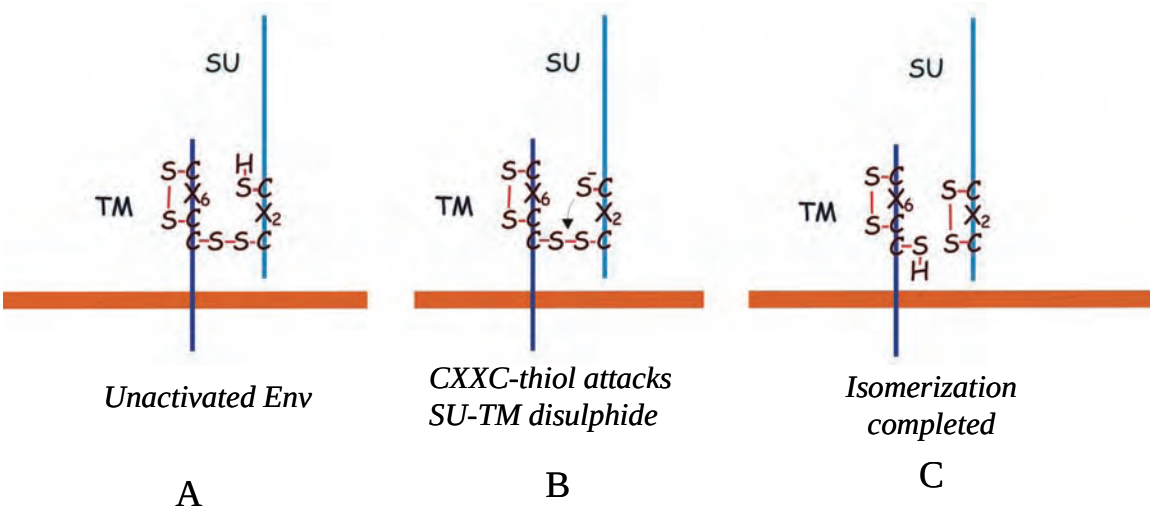

Fig. 10. *SU-TM disulphide-bond isomerization reaction in Env of Mo-MLV. (A) In the unactivated Env, the SU and TM are linked by an intersubunit disulfide-bond which relates to the CXXC motif in SU and CX_6CC motif in TM. (B) Upon receptor binding, the free thiol in the CXXC motif attacks the intersubunit disulfide bond and causes its rearrangement into another disulfide isomer, which is contained within the CXXC motif (C). This rearrangement is called isomerization .*

The model is supported by the atomic structure of a crystallized TM ecto-domain fragment (Fass *et al.*, 1996). This associated into very stable trimers, where the N-helices formed a central coiled-coil upon which the C-terminal parts were bound (**Fig. 11**). Apparently, the free fragments adopted the stable state of TM spontaneously. Furthermore the general activation model of TM is corroborated by the fact that N-helix peptides can inhibit virus fusion with cells (Wallin, 2006). In this case it is thought that the peptides either interfered with pre-hairpin formation in their monomeric form or that they form trimers in solution and that the trimers then interfered with the backfolding reaction of the TM C-terminal parts.

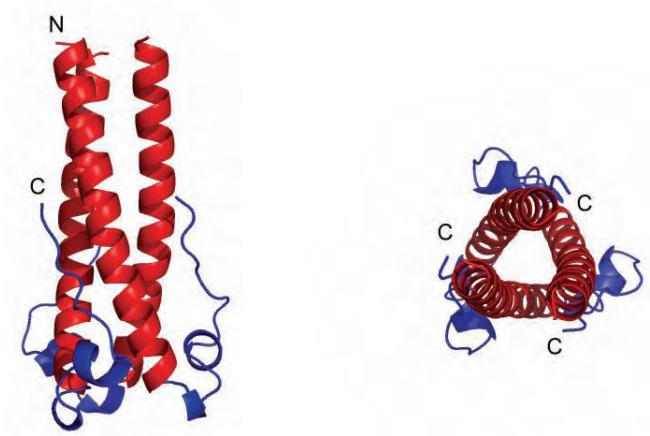

Fig. 11. *The "trimer of hairpins" structure of Mo-MLV TM (PDB: 1MOF) from side (left) and top (right) views. The N-helices (red color) formed a central coiled-coil core onto which the C-terminal parts were packed (blue color). The top view is looking down from the amino terminus of the coiled-coil core.*

The ultimate challenge for researchers in the field is to understand how the Env subunits are organized in the native unactivated spike and how they refold step by step in the context of the trimer into the final stable form. This will require unique ways of blocking the structural transformation path at intermediate steps as well as gentle procedures for isolation of the non-covalently associated trimers of the disulfide linked SU-TM complexes for structure determination by e.g. EM. In this respect it is important to note that one early intermediate form of the Mo-MLV Env has been generated. If a thiol alkylator is present during the activation of the spikes in virus particles then the isomerization active CXXC-thiol becomes modified by alkylation before it attacks the intersubunit disulfide and the Env refolding is arrested at an intermediate stage (Wallin *et al.*, 2004). Further, as the receptor mediated activation of the spike involves removal of a stabilizing Ca^{2+} from the spike, it is possible to mimic the activation trigger by depleting a virus preparation of Ca^{2+}. Therefore it is possible to convert all the spikes of a virus preparation into the isomerization arrested state (IAS) by Ca^{2+} depletion in the presence of an alkylator.

An elegant way to avoid the difficult isolation of the trimeric forms of the spikes for structure determination was introduced by cryo-electron tomography. In this approach the spikes are studied *in situ* on the surface of virus particles. Tomograms are created of the particles and the spike images then used for average-based reconstruction of the spike structure. This approach was used to study the Mo-MLV spike. The result showed a spike head with multiple projections standing on the membrane with tripod legs (**Fig. 12**) (Forster *et al.*, 2005).

Fig. 12. *The Mo-MLV Env structure. Note the tripod legs and the spiky head. (from (Forster et al., 2005) with copyright (2009) from National Academy of Sciences, U.S.A.)*

The spike protein of HIV-1

The HIV-1 spike is larger (420kD) than that of Mo-MLV (225 kD). The peripheral protein is called gp120 and the transmembrane gp41. In contrast to Mo-MLV the two subunits are non-covalently associated in the HIV-1 spike. This is illustrated in **Fig. 13**, which summarizes the nature of subunit association in retroviruses. Whereas the gamma (e.g. Mo-MLV)- and delta-retroviruses (e.g. human T-cell leukemia virus, HTLV-1) share the typical $C(X)_6CC$ and CXXC motifs in TM and SU, respectively, and probably also the intersubunit disulfide control mechanism of spike activation, the lentiviruses only have a modified motif in gp41 ($C(X)_5C$). The alpharetroviruses on the other hand (e.g. ALV-A)

have a typical TM motif, which also participates in intersubunit bond formation, but no isomerization motif in SU. Interestingly, an intersubunit disulfide has been introduced into the HIV-1 spike by substituting Cys residues directly after the TM-like motif and in the C-terminal portion of gp120 (Binley *et al.*, 2000). This gp mutant (SOSgp) matures normally in the cell into furin cleaved gp120 and gp41 subunits and is incorporated into particles if expressed in the context of the other viral genes. The viral SOS spikes can interact with the CD4 primary receptor and also with the co-receptor, a chemokine receptor, but the gp41 subunit is not activated for membrane fusion. However, fusion can be rescued if the intersubunit disulfide is reduced by DTT treatment after receptor engagement (Abrahamyan *et al.*, 2003; Binley *et al.*, 2003).

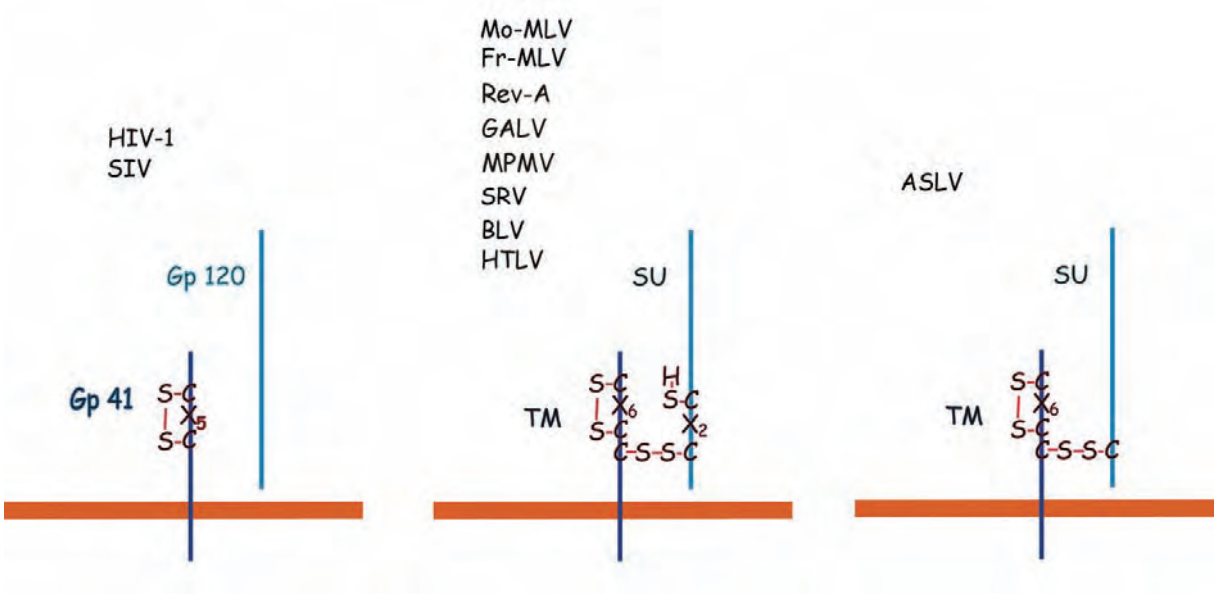

Fig. 13. *Intersubunit disulfides in retrovirus representing the lentivirus (HIV-1 and SIV), gamma- and deltaretrovirus (Mo-MLV, Fr-MLV, Rev-A, GALV, MPMV, SRV, BLV, HTLV-1) and alpharetrovirus (ASLV).*

A truncated form of the gp120 in complex with the CD4 ecto-domain and the Fab of the neutralizing monoclonal antibody 17b was possible to crystallize in a form that facilitated the determination of the atomic structure by X-ray analysis (Kwong *et al.*, 1998). The gp120 core was lacking most of the variable loops and parts of the N- and C-terminal regions. It should represent an activated intermediate form of gp120 because of its

reaction with the primary receptor; although it remained unclear to what extent the Fab binding modified the structure. The 17b antibody represents a CD4 induced antibody, which means that it reacts with an epitope generated by the CD4 induced structural change of gp120 (Thali *et al.*, 1993). The fact that this complex was the first atomic structure of gp120 demonstrates the extreme difficulties in obtaining a suitable preparation for structural studies of this protein. The major reason for this resides is its structural flexibility. Subsequently, the atomic structure of a V3 loop containing gp120 core in complex with the CD4 ecto-domain and the Fab of the neutralizing antibody X5 was determined (Huang *et al.*, 2005). The V3 loop is the major target for the secondary receptor. Finally, the atomic structure of a disulfide stabilized gp120 core in complex with the Fab of a CD4 binding site antibody, b12, was determined (Zhou *et al.*, 2007). These structural analyses showed that the gp120 core was a heart-shaped molecule with an external and inner domain linked together by a bridging sheet (**Fig. 14**). The latter structure was not complete in the b12 complexed structure of the gp120 core, suggesting that it represented an earlier form of gp120 in its structural transition. The V3 loop formed an extended finger-like projection (about 50 Å).

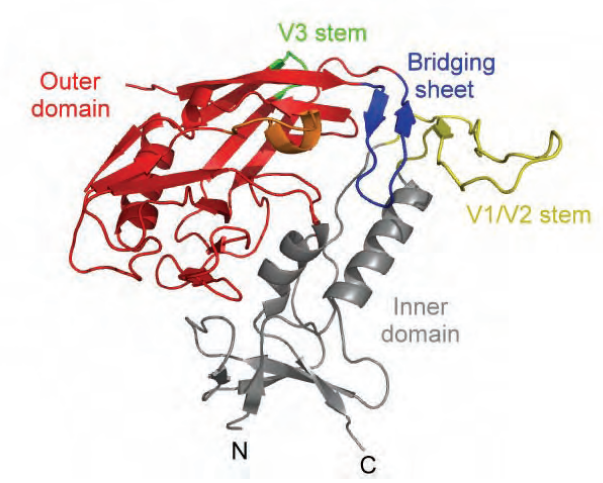

Fig. 14. *The atomic structure of gp120 core, derived from its complex with CD4 and the 17b Fab (PDB: 1GC1). The outer (red) and inner (gray) domains of the gp120 core are linked by a bridging sheet (blue). The stem of the V1/V2 loop is yellow whereas the stem of the V3 is green. The truncated N- and C-terminal ends are labeled and the CD4 binding loop is shown in orange color.*

Recently, the HIV-1 spike was analyzed *in situ* on vitrified virus particles by cryo-electron tomography (Liu *et al.*, 2008). The reconstruction of the spike structure showed a hollow head, which was connected to the viral membrane by a compact stem (**Fig. 15A**). In the head region the protomeric units formed lobes on the sides and interprotomeric connections at the three fold axis at top. The atomic structures of the gp120 core could be fitted into the lobe, such that the V3 loop and the stems of the truncated V1/V2 loops pointed towards the head top and the putative gp41 binding region of the inner domain of gp120 towards the lower part of the head. In this position the heavily glycosylated face of the gp120 was peripherally exposed, whereas the non-neutralizing face was internally oriented (Wyatt *et al.*, 1998). The CD4 binding site on the neutralizing face of the molecule was exposed on one of the flattened surfaces of the lobe. The co-receptor binding site was, according to this model, located to the V3 loop region close to the three fold axis of the head top. Interestingly, somewhat earlier a different reconstruction of the SIV spike was presented using the same *in situ* approach (**Fig. 15B**) (Zhu *et al.*, 2006). According to this the spike had a compact head standing on tripod legs. Thus at present there are two conflicting lentivirus spike structure models.

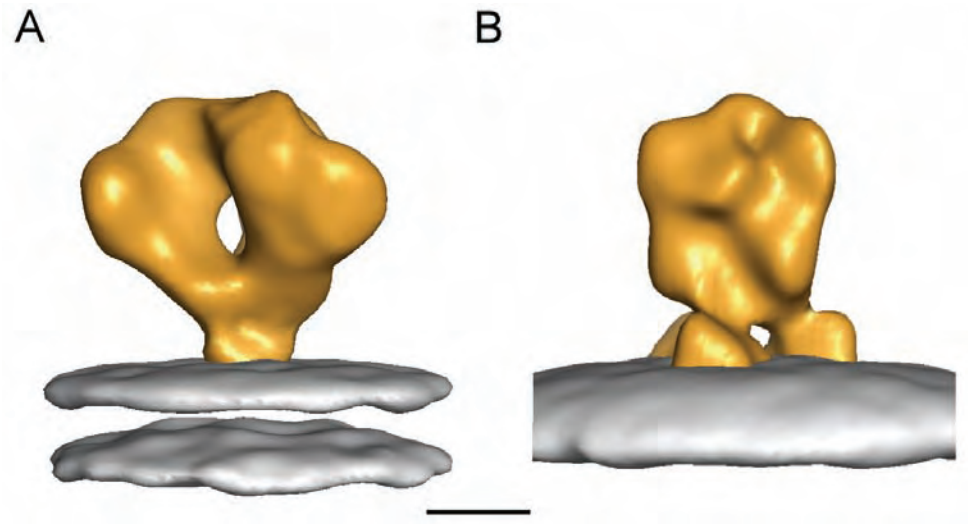

Fig. 15. *Cryo-electron tomographic reconstructions of the lentivirus spike. (A) The HIV-1 spike model with a cavity containing head connected to the membrane by a compact stem. (B) The SIV spike model with a compact head standing on tripod legs. Scale bar=5 nm.*

The activation of the HIV-1 spike for membrane fusion has been the subject of intensive studies. In particular thermodynamic measurements of the gp120 binding reaction with soluble CD4 have indicated that the complex formation is mediated by major structural changes in gp120 (Kwong *et al.*, 2002). This has been corroborated by the change in antibody binding pattern. Specifically, CD4 binding induced the epitopes for antibodies like 17b, which overlap the co-receptor binding site (Thali *et al.*, 1993). However, as long as the atomic structure of the native, non-complexed form of the gp120 remains elusive the complete nature of the CD4 induced changes also remain unclear. The atomic structure of the gp41 ecto-domain, lacking the fusion peptide region and the membrane proximal external region (MPER), but including both predicted helical regions flanking the $C(X)_5C$ motif, has been determined (Weissenhorn *et al.*, 1997). This showed the typical structure of an activated transmembrane subunit ecto-domain, i.e. a central coiled-coil composed of the N-helical region of three subunits, onto which the C-terminal helical regions were attached in an anti-parallel orientation. CD4 binding alone cannot activate the gp41 for membrane fusion, for this the CD4 induced changes must be complemented by those of the co-receptor binding. Nevertheless, CD4 has been shown to reveal binding sites for peptides representing the C-terminal helix region of the gp41 (Haim *et al.*, 2009). This suggests that at least a portion of the N-terminal coiled-coil is present and exposed. Whether this structure has been formed as a result of CD4 binding or whether it already is present in the native form of the spike and has only been revealed by CD4 induced changes in gp120 is unknown. The cryo-electron tomographic studies showed that there were major changes in the viral spike structure by the binding of CD4 (Liu *et al.*, 2008). The gp120 lobe was rotated clockwise along an axis parallel to the three fold axis 45° and out of the plane 15°. Also there appeared to be alterations in the parts assigned to the gp41. However, as the spike complex was studied in the presence of bound 17b Fab, for reasons of stability, it remains unclear to what extent the changes observed were the result of CD4 alone rather than a combined effect of both ligands. Therefore, the structure of the HIV-1 spike and its transition into the final activated form remains a formidable challenge for future studies.

AIMS

The goal in my thesis is to reveal the dynamic transformation of viral structures preceding membrane fusion. This is done by mimicking the fusion initiation. The whole virion or the spikes are trapped at the intermediate stages and their structures are revealed by electron cryomicroscopy and three-dimensional image reconstruction.

The work has been focused on the following aspects:

- The structure of SFV
  - The acid triggered fusogenic structure of SFV (*paper I*)
  - Mapping functional domains in the SFV envelope glycoprotein (*unpublished results*)
  - Control of premature acid activation in SFV (*paper II*)
- The activation of the Mo-MLV envelope glycoprotein (*paper III*)
- The structure of the HIV-1 envelope glycoprotein (*paper IV*)

METHODOLOGY

The methods are described in detail in the individual articles. The virus purification and the retrovirus spike isolation as well as the EM technique and 3D reconstruction are summerized below.

Virus preparation

Purification of SFV and the low pH treatment

For production of SFVwt, monolayers of Baby hamster kidney (BHK) cells were infected with SFVwt (clone pSFV4 (Liljestrom &Garoff, 1991)); while for the p62 cleavage deficiant mutant SFV, SFVmSQL, the cells were infected with chymotrypsin-activated mutant, for one hour at 37ºC. The cells were washed, further incubated at 37°C and virus harvested at 18h post infection. The cell supernatant was cleared from cell debris and the virus collected as a pellet by centrifugation. The virus pellet was soaked in TNM (50 mM Tris, 50 mM NaCl and 10 mM $MgCl_2$) buffer, pH 7.4, and further purified by centrifugation to equilibrium in a linear potassium tartrate gradient. The virus was eluted from the gradient, diluted with TNM buffer, and pelleted by centrifugation. The finally obtained virus pellet was soaked in TNM buffer and kept at 4°C to avoid structural damage. For acid treatment, the virus was mixed with a buffer containing 50 mM MES, 50 mM NaCl, 10 mM $MgCl_2$, pH 5.6, to reach the desired pH.

Purification of the Mo-MLV spike

Mo-MLV was grown in MOV3 cells, an NIH 3T3 clone transformed with wt Mo-MLV. The virus in the cell supernatant was purified and concentrated by sedimentation in a 20/50% (w/w) sucrose step gradient in HN buffer (20 mM Hepes, 135 mM NaCl, pH 7.45) containing 1.8 mM $CaCl_2$. The virus was collected at the sucrose interphase and dialyzed against HN buffer with Ca^{2+} and then further purified by sedimentation in a small (650 µl) 20/50% sucrose step gradient. Native spikes were prepared by solubilizing the purified virus in HN buffer containing 1% Triton X-100 and 1.8 mM Ca^{2+} for 15 min on ice and, after removing sucrose by gel filtration, subjecting the sample to sedimentation in a

continuous 5-20% sucrose gradient in HN buffer containing 0.05% Triton X-100. The Env trimers, i.e. the spikes, were identified by blue native gel electrophoresis (BN-PAGE) and silver staining. These were shown to be separated from Env dimers and monomers as well as viral internal proteins. After sucrose removal the material in the fractions containing pure Env trimers were used for EM analyses. The intermediate form of the spike, the IAS form, was prepared by solubilizing the virus in HN buffer containing 1% Triton X-100, 10 mM EDTA instead of Ca^{2+} and 10 mM NEM for 20 min at 37°C. The Ca^{2+} depletion activated the spike and the alkylator NEM arrested it at the intermediate, IAS, stage. The spike intermediate was then isolated as described for the native form of the spike.

Purification of the HIV-1 spike

Thirty 150 cm^2 flasks with 293T cells were transfected, each flask with 20 µg HIV-1 VLP DNA and 20 µg SOSgp DNA. VLPs were collected from 8-48 h after transfection and purified by sedimentation in a 20/60% (w/w) sucrose step gradient in HNC buffer (50 mM Hepes, 100 mM NaCl, 1.8 mM $CaCl_2$, pH 7.4). The VLPs at the sucrose interphase were recovered, diluted and concentrated by ultra-filtration before further purification on a small, 650 µl, 20/60% sucrose step gradient. For isolation of spikes, the purified VLPs were solubilized in HNC buffer containing 2% Triton X-100 for 10 min at 37°C and then sedimented in a continuous 20-50% sucrose gradient, superimposed with a 0-0.2% glutaraldehyde gradient, in HNC buffer containing 0.05% Triton X-100. The trimeric form of the HIV-1 gp was identified in gradient fractions by BN-PAGE. The pure and correctly oligomerized form of the spike was then used for EM analysis.

Structural analyses

The models of viral structures presented in this thesis were obtained by transmission electron microscopy (TEM) and computer aided three-dimensional (3D) image reconstruction. In short, projection images of protein complexes or viruses were collected from recording media in which all possible viewing angles were covered. The collected

data sets were then combined to reconstruct a 3D model of the protein or virus. The EM related techniques and the reconstruction processing are described below.

Specimen preparation

Biological molecules are susceptible to damage both from mechanical stress during sample preparation and from EM radiation. Furthermore, they have poor ability to scatter electrons, leading to low contrast images. Thus, an ideal specimen preparation should preserve and stabilize the structure under study and also increase the contrast. However, there is no single method that can serve all these needs. Instead, independent structural analyses based on different preparation techniques have to be compared. I chose to work with samples embedded in either negative stain or vitrified ice without stain for my reconstructions.

Negatively stained specimen

Negative staining is a technique where the specimen is first adsorbed to a carbon coated grid, then embedded in a heavy metal solution, rinsed with water, and air dried. This makes protein to be surrounded by a thick layer of metal on the grid. While imaging by EM, electrons are scattered differently by the protein and the metal, thus creating various electron densities in the image. The method was introduced as a quick and easy technique that significantly increases the specimen contrast and makes it possible to visualize much smaller molecules than before. The drawback is that the specimen is usually flattened against the carbon support film and is often distorted during the air drying process. An additional problem is that many proteins attached to the carbon support film with a preferred orientation, so that certain viewing angles may be completely omitted. Furthermore, microcrystals formed by the heavy metals upon drying also limit the possible resolution of the reconstruction.

A variety of heavy metal solutions are available for negative staining. The most commonly used are uranyl acetate (UA), uranyl formate, tungstate stains and ammonium molybdate. The uranyl stains often generate the highest contrast of the specimen. The

latter two stains have the advantage that the pH of the staining solutions can be adjusted close to 7. In my study, I chose UA because it provides the highest contrast and has a fixative effect. The high contrast helps to visualize small molecules, such as solubilized spikes of Mo-MLV (225 kDa) and HIV-1 (420 kDa). UA solution is acidic (pH ~4.2) and neutralization is not possible as it causes precipitation of the stain. A concern was that the viral fusion proteins under study would undergo conformational changes due to the low pH of the stain. However, it has been shown that UA solution fixes the protein on a millisecond timescale (Zhao &Craig, 2003), thus eliminating this problem.

In the negative staining protocol, 3.5 μl of the sample was applied on a glow-discharged carbon-coated grid, washed once with a drop of distilled water, and stained with a drop of freshly prepared 2% UA. Excessive stain solution was removed by a piece of filter paper to create an adequate stain embedding and the grid was air dried. The negatively stained specimen was transferred to a room temperature holder and then inserted into the TEM for inspection.

Vitrified specimen

The goal of vitrified sample preparation is to obtain a frozen-hydrated specimen free from possible deformation or damage due to the dehydration effect from the negative staining procedure (Adrian *et al.*, 1984; Dubochet *et al.*, 1988). Besides, proteins embedded in vitrified ice can normally adopt random orientations which results in better angular coverage in the projections of the molecule. Therefore, it provides a means to achieve higher resolution in the 3D reconstruction than the negatively stained specimen. However, the vitrification procedures are substantially more complicated and time-consuming than the negative staining procedures; in addition, without the heavy metals, images from vitrified specimens have very low contrast.

In the vitrification procedure, 3.5 μl of the sample was placed on a glow-discharged holey carbon grid, blotted with a piece of filter paper to produce a thin layer of sample solution across the holes of the carbon film, and subsequently plunged into liquid ethane cooled in

a liquid nitrogen bath, so that the specimen was preserved within vitrified ice. This rapid freezing causes vitrification of water molecules rather than polycrystalline ice formation. The liquid ethane cooled by liquid nitrogen is a good combination to prepare the cryogen because the liquid ethane does not boil as it cools the sample. In an ideal preparation, the specimen is embedded in a layer of ice which is slightly thicker than the specimen. If the ice is too thick, the particles may overlap each other and become inadequate for the following image processing. Therefore, in order to obtain the proper ice thickness, the humidity control of the environment and the removal of the excess liquid by blotting are the critical factors in the freezing procedure. In this thesis work, the freezing process was performed by FEI Vitrobot^{TM}, an automated vitrification device, which features a chamber with controllable humidity and temperature and a robotic blotting device where time and force can both be adjusted.

The vitrified specimen was transferred to a cryo-holder which contains a cooling dewar to keep the specimen below -170^{o}C throughout all subsequent steps. The mounting step was performed in a cryo-transfer station which is used to maintain the low temperature to prevent the ice contamination caused by atmospheric water during the transfer steps. The holder with the specimen was then inserted into the TEM.

Data imaging

Acquisition of high quality images of the protein complexes or viruses is the critical step to obtain a high resolution 3D reconstruction. While taking images of a negatively stained specimen is straightforward, imaging of a vitrified specimen is more complex.

Cryo-EM is a technique in which the specimen to be investigated is kept at low temperature in order to protect it during imaging. As the specimen is so sensitive to the electron radiation, it is essential to image under low electron dose conditions (below 10 electrons per Å^{2}). Because low-dose conditions are used, the signal-to-noise ratio becomes very low. In order to increase the signal-to-noise ratio, a large number of particles are required for credible statistical results.

In the SFV study, data were recorded on Kodak SO163 photographic films using a JEM3200F TEM equipped with field emission gun (FEG) and liquid helium cold stage at a magnification of 40,000x and an accelerating voltage of 300 kV. In the study of Mo-MLV and HIV-1, all cryo-images were recorded on a TVIPS TemCam-F415 CCD camera (Gauting, Germany) with 4k × 4k pixels at the nominal magnifications of 43,000x or 86,000x in a JEM2100F field emission TEM. Digitized micrographs and images were sub-sampled to 3.5 Å per pixel on the specimen scale for the following image processing. All data were taken at multiple defocus levels.

Image processing

The general purpose of single particle reconstruction is to create a 3D reconstruction from many two-dimensional (2D) projections of proteins (particle images) generated in the TEM. Before reconstruction processing, visual inspection is required to discard the images with too much contamination or obvious drift. This is done in a graphical program, such as *boxer* in the EMAN package (Ludtke *et al.*, 1999) or the program *RobEM*[2]. The power spectra of the images are then inspected to discard those with astigmatism or drift and those without Thon rings. The major part of the image processing procedure is to determine the origin and orientation of each 2D particle image, so that a 3D reconstruction can be computed. In order to achieve a high-resolution reconstruction, it is necessary to have many different views of the particle. In brief, the particles are firstly selected from the scanned micrographs or CCD frames. Secondly, the selected particles are used to generate an initial 3D model. The initial model is required for generating a set of model projections with known orientations. Each individual particle image is compared to the model projections and then assigned to the orientation of the projection having the best matches to generate a new 3D model. Finally, the initial model is updated with the newly computed 3D reconstruction for the next iteration of the refinement process until no improvement can be observed.

[2] http://cryoem.ucsd.edu/programDocs/runRobem.txt

I used several different strategies throughout this thesis study. For the whole SFV virion, a model-based polar Fourier transform (PFT) software package (Baker &Cheng, 1996) was used to compute the icosahedral 3D reconstructions. For the trimeric spikes of Mo-MLV and HIV-1, the EMAN package was used because it supports other symmetries (C3 in these cases).

Icosahedral reconstruction

Icosahedral reconstruction is a process of combining particle images to yield a structure which is aided by the intrinsic icosahedral symmetry of the particle. The icosahedral reconstruction takes advantage of the high symmetry to increase the averaging power in the final 3D reconstruction, i.e., an icosahedral virus contains 60 equivalent asymmetric units, so that a single particle image can be treated as 60 images by symmetry operation to reconstruct the 3D model. The initial origins and orientations of the particle images were determined using *PFTsearch* (Baker &Cheng, 1996), in which each particle image was compared against a set of projections, covering one-half of the asymmetric unit, from an existing 3D model in the polar coordinates. The particle image was then assigned to the orientation of the projection having the highest resemblance. The origin and orientation parameters determined in *PFTsearch* were used as the input in *OOR*[3] for further refinement, which compares the Fourier transform of the particle image against the Fourier transforms of the model projections within 7 x 7 search grids and finds the best matches close to the current estimate. The refined origins and orientations were then used to compute a new 3D model using *EM3DR* (Baker &Cheng, 1996) for the next iteration. Cycles of refinement against a successively improved 3D model was continuous until no further improvement in the resolution was observed.

[3] http://cryoem.ucsd.edu/programDocs/oor.pdf

Trimeric spike reconstruction

Reconstructions of the trimeric spikes of Mo-MLV and HIV-1 were processed by EMAN package, which also used the model-based projection-matching algorithm for the classification in the refinement procedure. EMAN has its own algorithm to build an initial 3D model. In the case of C3 symmetry using the program *startcsym*, EMAN first searches the images which have the best C3 rotational symmetry to use as top views, and those with good mirror but poor C3 rotational symmetry as side views. The selected particles within these two groups were aligned and averaged respectively to generate the class-averages. A preliminary 3D model was then constructed from these two orthogonal views. With the obtained initial model, the refinement was carried out as an iterative procedure of classification and model building. Generally, each particle was translationally aligned to a rotational averaged image, the average from all of the particles, to create a centered particle set. The centered particles were subjected to classification against a set of projections from the initial 3D model and then assigned to the projection group having the best matching value. When the orientations of all particles have been determined, the particles with the same orientations, that is, the particles within the same class, were further aligned to generate a class-average. These 2D class-averages were then used to build a new 3D model. The refinement procedure was continued iteratively to gradually higher resolutions until the model converged. To assess whether the classification procedure was successful, each class-average was visually compared to the corresponding projection of the 3D model.

Validation and visualization

The resolution of the 3D reconstruction was assessed using Fourier shell correlation (FSC) to measure the correlation coefficient between two reconstructions generated by randomly splitting the original dataset into two parts over corresponding shells in Fourier space (Saxton &Baumeister, 1982). The resolution value is defined by the point where the FSC curve first drops below a specified threshold (normally at 0.5). If the resolution fails to improve with further refinements, the target resolution has been reached. The stability of the model is also estimated by an FSC that compared the latest two 3D models in the

refinement cycles. When a stable 3D reconstruction with a sufficient resolution has been obtained, a quasi atomic resolution may be reached if the atomic structure of the protein (or a protein subunit) is known and can be fitted into the density map. All 3D visualizations were rendered using the graphical software *IRIX Explorer* (Numerical Algorithms Group, Oxford UK), *RobEM*, *Chimera* (Pettersen *et al.*, 2004) or *PyMol* (DeLano Scientific, Palo Alto, USA). The docking of atomic structures into density maps was done in the program *O* (Jones *et al.*, 1991).

RESULTS

The structures of SFV (papers I, II, and unpublished results[4])

SFV infects its host cells by endocytosis (Helenius *et al.*, 1980) and low pH-mediated membrane fusion (White *et al.*, 1980). The membrane fusion of SFV is strictly dependent on the exposure of the virus to a low pH during the endosomal pathway (Marsh *et al.*, 1983; White *et al.*, 1980). The acidification triggers conformational changes in the virus spike to allow exposure of the fusion loop as seen by antibody mapping (Hammar *et al.*, 2003). As discussed, the cell entry is catalyzed by a class II fusion protein or spike which is triggered for fusion activity by acidification. The virions are constructed such that a maturation cleavage is needed for the spike to become susceptible for acid trigger.

To explore the structural transformation of SFV into its acid activated fusogenic configuration, my approach has been to determine structures of the virus at intermediate stages to reveal relocations of the spike subunits E2 and E1 that are related to the fusion process. Therefore, purified virus was treated at a series of pH values, from pH 7.4 down to pH 5.8, for 60s and plunge-frozen to be inspected by low dose cryo-EM. Structural data were recorded and the structures were obtained by image analysis and 3D reconstruction. By comparing the solved structures of the viruses treated at different pH, relocation of subunit domains in the SFV spike is demonstrated (paper I). To strengthen the functional implications of the structures revealed, observations on SFV in complex with monoclonal antibodies, are discussed (unpublished results). Furthermore, I explore the structural nature of the acid resistance imposed by the chaperon subunit precursor p62, before its maturation cleavage into E3 and E2, using the p62 cleavage-impaired SFVmSQL (Berglund *et al.*, 1993) (paper II).

[4] Haag, Wu et al., Functional spike domains and their interplay in Semliki Forest virus, *in reparation*.

Morphological features of the spike structure related to pH variations

The structures of the SFV virion in the different low pH conditions (5.8 and 5.9) were obtained and compared to the structure of the virion at pH 7.4 (control). The control structure was improved from an earlier published reconstruction (Haag *et al.*, 2002) and received at the resolution of 15 Å. The higher resolution is mainly due to improved data collection using a 300 kV field emission electron microscopy equipped with a helium cooling system leading to improved high-frequency information. A comparison of the reconstructions of control and acid treated virions revealed that the overall particle morphology with its three-lobed spike organization was essentially retained in all the structures, including the pH 5.8 structure (**Fig. 16A**, Acid). The spikes display a head, stalk and a shell domain as well as a limb part that connects to the outer bilayer leaflet of the virus membrane. However, as the pH was lowered, the diameter of the acid treated virus increased from 707 Å at pH 7.4 to 727 Å at pH 5.8 (**Fig. 16**). In addition, the openings in the shell layer at the two and five fold axes were enlarged in the acid treated structures. The size expansion is mainly due to a lift in the glycoprotein i.e. the spike layer. This was evident by comparing the SFV structures in cross-section views and in one-dimensional radial average-density plots (1D plots) (**Fig. 16B**).

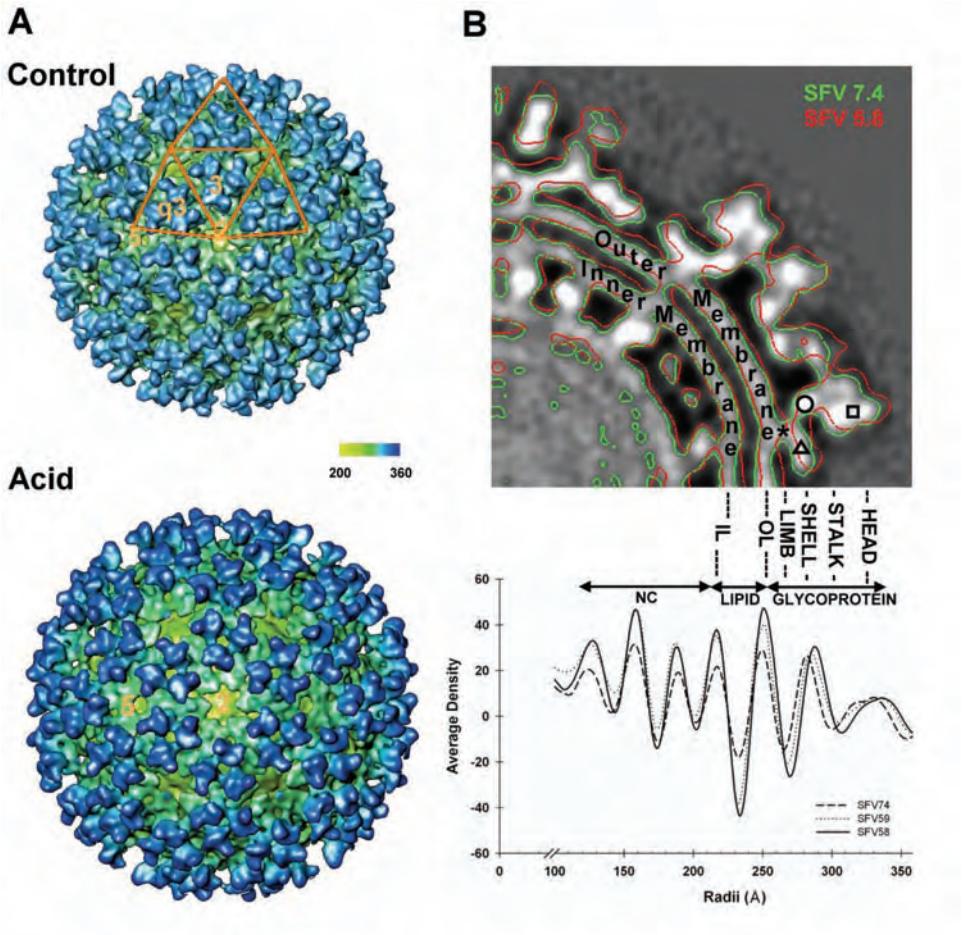

Fig. 16. *Cryo-EM reconstructions of SFV at control (pH 7.4) and acid triggered (pH 5.8) stages. (A) Surface rendered SFV structure is viewed along the two fold axes. The particles swelling in the acid structure is seen by the darker blue in the rendering of the pH 5.8 structure (see the radial color code between the renderings), and in the overlaid cross-section views (B, top). 1D plots of virus treated at pH 7.4 (hatched), pH 5.9 (dotted) and pH 5.8 (solid) show the gradual increase of the virion diameter in response to decrease in pH (B, bottom). Note that the radii of the nucleocapsid (NC) and membrane bilayer (inner, IL; outer, OL) remain the same, while the glycoprotein, i.e. the spike, layers are moved outwards in the acid structures (B).*

To analyze the structure variations further and find clues for how the interactions between the subunits E1 and E2 varied with pH, I had to identify the regions occupied by the two subunits. Therefore, the solved atomic structure of the ecto-domain of the pre-fusion E1, the E1* (Lescar *et al.*, 2001; Roussel *et al.*, 2006), was modeled into the cryo-EM derived structures. Furthermore, the limited knowledge of the folding of the E2 polypeptide chain

and the nature of its contacts with E1 demanded a way to better understand the structural relocations seen.

Assigning subunit domains to nodes in the cryo-EM derived structure

When modeling the E1* structure (PDB: 2ALA) (Roussel *et al.*, 2006) into the high resolution cryo-EM density map of SFV at the control pH of 7.4, we observed that high density centers, or nodes, derived at high stringency in the structural rendering, fitted well with beta sheet domains of the E1*. The **Fig. 17** shows the assignment of the individual nodes, denoted *a* to *i*, from the top of the glycoprotein layer and inwards, and the good agreement of the nodes *c*, *e*, g, *h*, and *i*, to secondary structural elements in the E1* atomic structure. The nodes *h* and *i* in the shell domain of the virus, represent the domains I and III, respectively. From this we conclude that the nodes reflect rigid structural domains. Therefore, nodes identified in the non-E1 part of the external spike, i.e. *a*, *b*, *d*, and *f*, would represent rigid secondary structures, or otherwise non-flexible domains, in the E2 subunit. The nodes, which could be followed in the series of structures from different pH conditions, allowed the pH dependent rearrangements in, and between, the E1 and E2 subunits to be followed (**Fig. 17**, bottom panel and **Fig. 18**). Based on the node assignments, one can start to discuss the dynamics of the two glycoproteins in response to the pH changes.

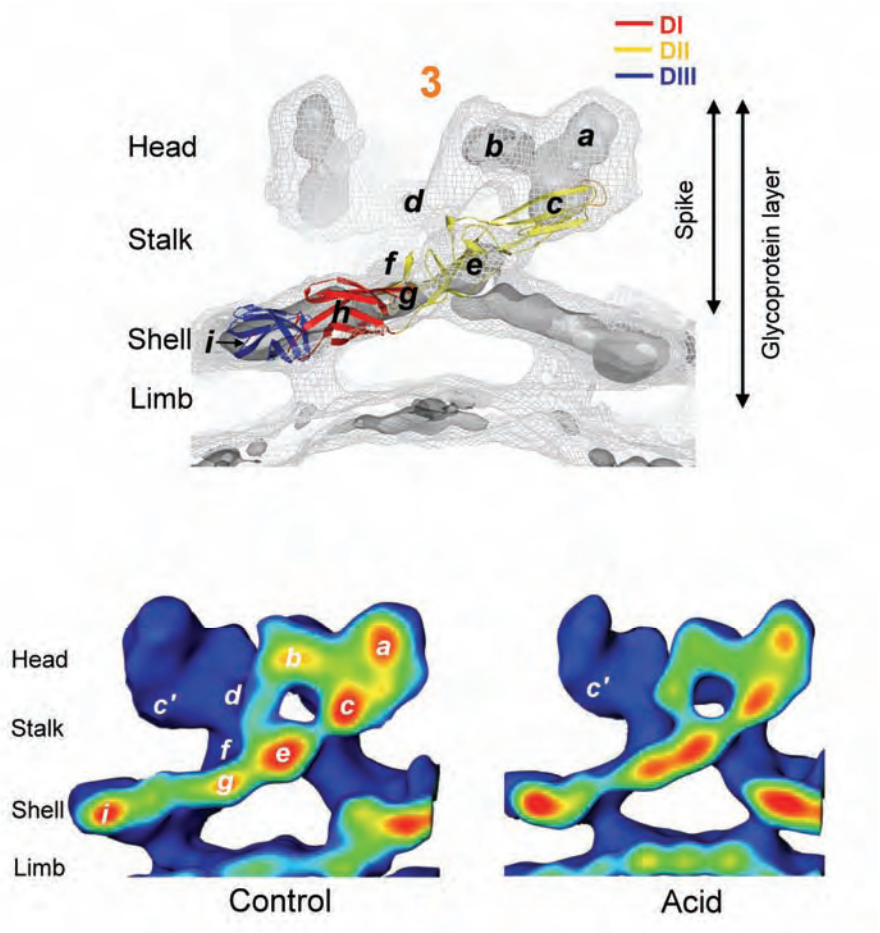

Fig. 17. *Assignment of density nodes in the cryo-EM reconstruction of SFV (top panel) and the demonstration of pH dependent movement within the spike (bottom panel). High density centers in the reconstruction, referred to as nodes, are envisioned as sigma level 3 surface rendering, within a sigma level 1 surface net. The nodes are numbered alphabetically from the top of the spike towards the shell region. Best fit of the atomic structure of E1* shows that rigid secondary structures match the nodes c, e, g, h and i in the molecule (top panel; domain I of E1 is red; domain II, yellow; and domain III, blue). The remaining nodes a, b, d and f are assigned to E2. The bottom panel shows virtual vertical cuts of the three fold spike in the control (pH 7.4) and acid (pH 5.8) structures with a gradient color code that represents the density level in the reconstructions. The high density nodes are in red. The node h is out of plane. The E2 nodes a and b, and E1 node c are seen as slightly less dense as their centers have partly moved out of plane. The c' refers to the position of node c in another E1 molecule.*

Demonstration of pH dependent node relocation in the spike

All the nodes identified in the control structure can also be traced in the acid structure and their relative positions followed. On acidification of the virus, the nodes *c*, *e*, *g*, *h*, and *i*, representing the E1 protein, have all moved outwards in the virus and adopted a more vertical position relative to the membrane (**Fig. 17**, bottom panel. The *h* node is out of plane). This upward movment extends the spike layer. Concomitantly the spike shell is split up with widening of the holes around the two and five fold axes (**Fig. 16**). As the shell is constituted of E1 domains I and III, this reflects a release of the E1-E1 interactions, both between the spikes and between the domain I and III in the same molecule, as observed from the distortions between nodes *i* and *h* (paper I).

Release of subunit contacts in the stalk domain of the spike

In order to demonstrate details of E1 and E2 subunit rearrangements on acid trigger, a series of virtual sections along the three fold spike were done and further analyzed. By comparing corresponding regions of the spike of control and acid structures, a reciprocal relocation of E1 and E2 densities in the bottom of the head and the stalk regions were found. The nodes *c*, *e*, *g* and *h* which represent E1, follow an anticlockwise rotation (**Fig. 18**); the E2 node *d* that represent the spike wing connecting point, sharpens into two separating structures (the black arrows in **Fig. 18**, Acid, top panel); whereas the node *f*, an E2 density that becomes more intense, is turning slightly clockwise (**Fig. 18**, bottom panel). This low pH induced reciprocal turning of E1 and E2 subunits separates the heterodimers of the spike.

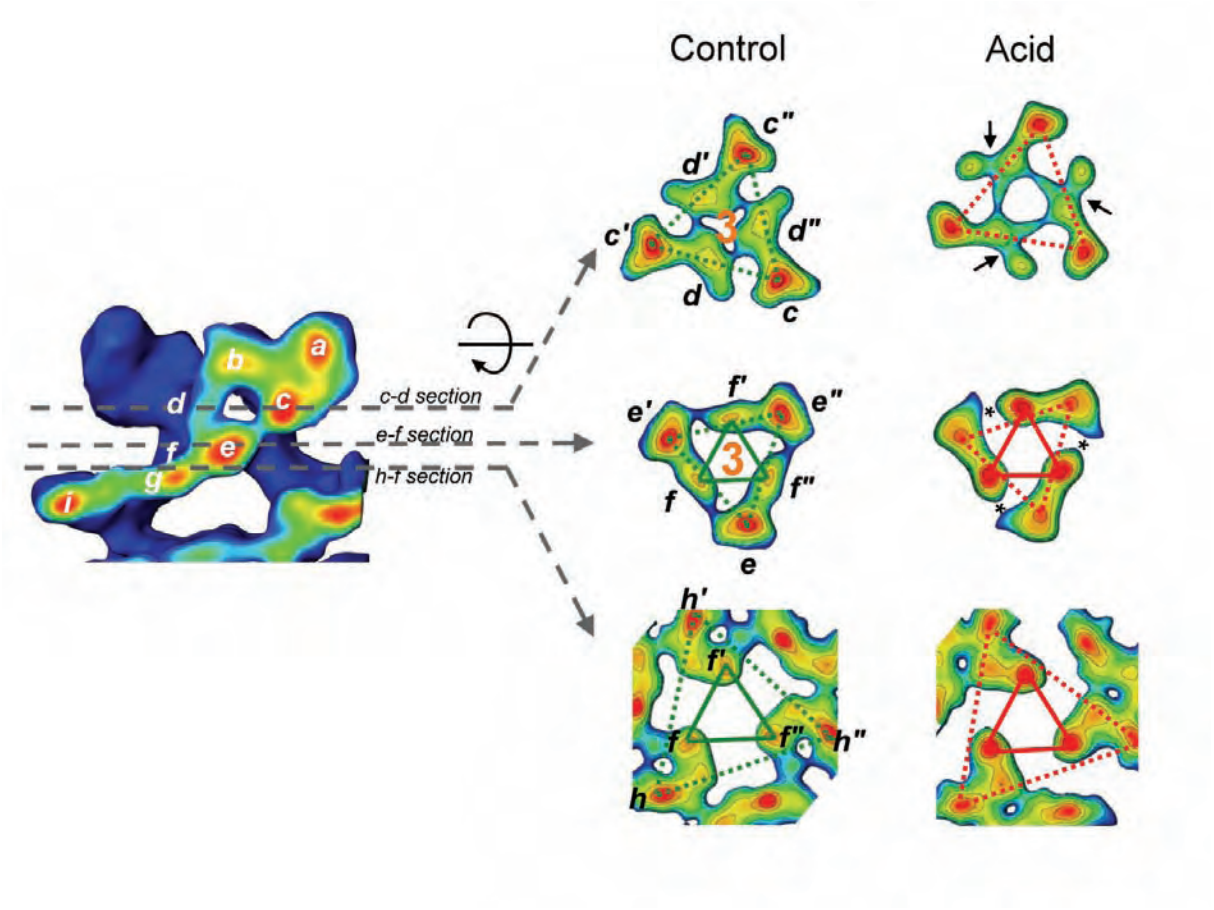

Fig. 18. *A series of radial sections show the node rearrangement in the spike in response to variations in pH. Major nodes are labeled as assigned in the* ***Fig. 17****. Here the sections through the c-d, e-f and h-f plains are shown. The dashed triangles connecting nodes c (top), e (middle), and h (bottom), which represent E1, follow the same anticlockwise turning. The solid triangles that connect node f, representing E2, turn slightly clockwise. Consequently, this dissociates the subunits in the stalk region, here seen as a separation between nodes e and f (black asterisk). The c' and c" refer to the neighbors of node c within the spike, etc.*

Rearrangements in the head domain of the spike

In the acid treated virus the E2 nodes *a*, *b*, and E1 node *c* are raised with the head of the spike (**Fig. 17**, bottom panel). The *a* and *b* in the E2 assigned part of the spike head show a slight relocation relative to each other and the *c* is moved sideways under node *a*. As judged from the fitted E1* atomic structure, the fusion peptide loop, residing in the distal part of the domain II of E1, would be located at the outer edge of node *c*. The detailed analyses of the acid structure indicate that this would allow the fusion peptide to be

exposed at the side of the spike head, in accord with its availability for detection by a monoclonal antibody E1f (Hammar *et al.*, 2003). To verify that this is a correct interpretation, a complex of the virus and the fusion peptide binding antibody was made and applied to structural analysis.

Location of a putative receptor binding site and the fusion peptide loop in the spike at low pH

In order to accurately reveal the location of a putative receptor binding site in E2 and the fusion peptide in E1, two monoclonal Fab-virion complexes were studied. Earlier it was observed that a set of monoclonal antibody (mAb) epitopes on SFV were exposed for antibody interaction in a pH dependent manner (Hammar *et al.*, 2003). Of the antibodies studied, only the neutralizing E1n bound to the virus at neutral pH. A complex between the antibody Fab fragment and the virus showed that E1n binds at the rim of the central cavity in the spike and with occupancy of at the most one per spike (Haag, 2006). The epitope was not linear and was lost on acidification. Here I introduced two other monoclonal Fab-virion complexes to demonstrate the location of a putative receptor binding site and location of the fusion peptide loop in the spike structure at pH 5.9, representing optimal binding condition for both antibodies. The specificity and epitopes had been identified by western blot and PepScan analysis (Hammar *et al.*, 2003). The mAbE2r$_{211\text{-}218}$ epitope is overlapping in the E2 sequence with a protective epitope reported in Ross River (Vrati *et al.*, 1988) and Sindbis viruses (Mendoza *et al.*, 1988). The E2r epitope is only 12 residues away from N^{199} where the complex type carbohydrate moiety in SFV is attached. It binds the virus at close to neutral pH, but optimally at around pH 5.9. The mAbE1f$_{85\text{-}95}$ epitope represents part of the fusion peptide sequence (Garoff *et al.*, 1980), a conserved region in E1 that is hidden in the neutral structure. Optimal binding is at about pH 6.0. The Fab complexes were prepared by mixing gradient purified virus with an about four-fold excess of the Fab and equilibrate the mixture by dialysis at pH 5.9. The excess Fab was removed by gel filtration and the complex applied to cryo-EM analysis.

The mAbE2r$_{211-218}$ epitope location

The cryo-EM reconstruction of SFV in complex with the Fab fragment of mAbE2r was computed to a resolution of 30 Å (**Fig. 19**). The reconstruction shows well preserved virus morphology. There are 240 external protrusions sitting on the outermost tips of the SFV trimeric spikes (green structures, **Fig. 19**, left), which indicate full occupancy of the ligand. The location of the antibody grip on the top of the spike is shown in **Fig. 20**. It coincides with that reported by Smith *et al.* as a putative receptor binding site (Smith *et al.*, 1995).

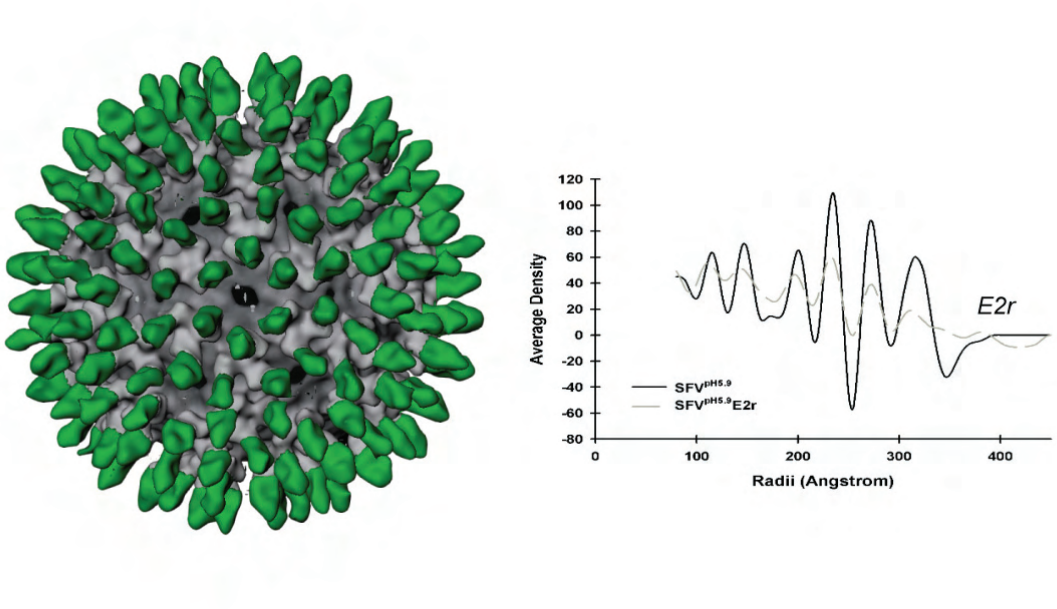

Fig. 19. *Cryo-EM reconstruction of the SFV in complex with the Fab fragment of mAbE2r (SFV/E2r). The mAb epitope represents a putative receptor binding site in E2. The complex has a well-preserved overall virus morphology and, except for the Fab protrusions, a comparable size with SFV at pH 5.9, as seen from the 1D plot (right).*

The mAbE1f$_{85-95}$ epitope location

The cryo-EM reconstruction of SFV in complex with Fab fragment of mAbE1f$_{85-95}$ (SFV/E1f) also displayed well preserved virus morphology and a diameter, disregarding the ligands, as that of a pH 5.9 unliganded virus (not shown). The structure of one wing of

the spike head of the Fab E1f complex is shown in the right panel of **Fig. 20**. The Fab E1f density in the complex reconstruction is much slimmer than that expected. The height and thickness of the density corresponding to the Fab E2r and Fab E1f in the electron density maps are about 90% and 40% of the atomic structure Fab density. The poor density of the Fab in the case of E1f, would partly be the result of structural averaging at poor occupancy - in Biacore experiments the maximal binding obtained with the E1f is one ligand per spike (Hammar *et al.*, 2003). It is obvious that the Fab gripping site is poorly available. However, the location outside node *c* is clear and confirms the location in the spike of the fusion peptide, as modeled by the atomic structure of E1 (**Fig. 20**, orange loop in the blue ribbon representation of E1).

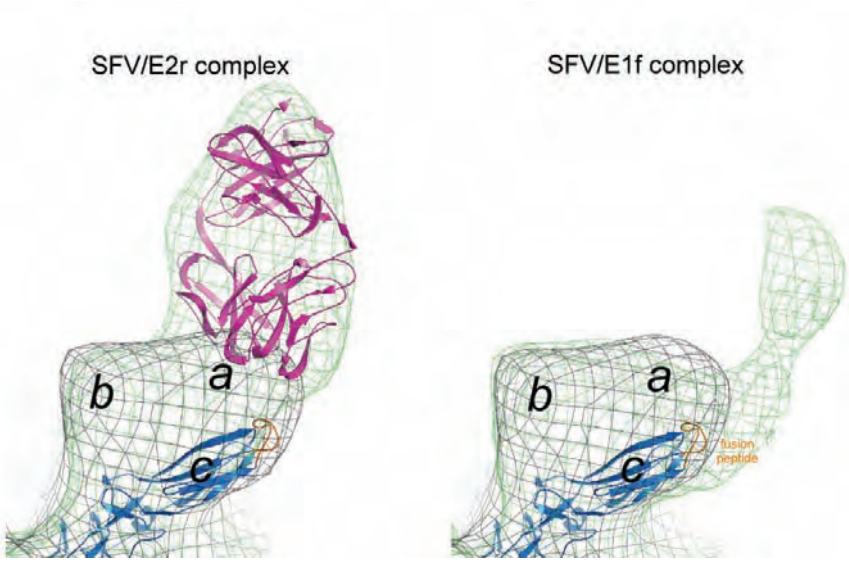

Fig. 20. *The locations of the mAbE2r and mAbE1f epitopes in the head domain of the SFV spike. Only one of the three wings is shown. Letters a, b and, c refer to position of high stringency nodes in the structure. The nodes a and b represent rigid domains in E2, and the c the twisted beta sheet in distal part of E1 domain II. The left structure represents the SFV/E2r complex – the protruding Fab density is footing above the node a, and holds the shape and volume of a Fab fragment (mauve ribbon structure). The right structure shows the SFV/E1f complex. Here the attached Fab grips at a position down at the side of the spike wing and confirms the modeled location of the fusion loop. The E1 atomic structure is modeled into the spike structure (blue ribbon, fusion loop in orange).*

Taken together, the SFV/E2r complex demonstrates a reasonable location of a receptor binding domain at the outermost tip of the spike, above the node *a* - the most external structures of the virus that would be the first part to come in contact with the cell surface on infection. On the other hand, the fusion peptide in the SFV seems to be much less available, in spite of the fact that the virus is fully fusion competent (pH 5.9) for membrane fusion if meeting a susceptible target membrane at pH 5.9. It is obvious that in the pH 5.9 structure of the virus, a pH at which the shell domain is zipping up and interactions in the stem region are released, the E2 top domain of the spike is still holding a dominating control over the E1 subunit.

Identification of the E3 part of p62 in the activation impaired virus spike and the suppression of the premature acid activation in SFV

In the previous studies I pointed to regions of the SFV envelope and in particular in the spike, where relocations were seen in response to variations in pH. The structures were solved at pH conditions known to promote virus fusion when in contact with a susceptible target membrane. The threshold for fusion is reported to be ~6.2 (Barth &Garoff, 1997; Bron *et al.*, 1993; Justman *et al.*, 1993). In the light that the relocations occurring during the fusion process are drastic and require that the spike subunit pair of E1 and E2 is completely releasing the hold on each other, the structural changes observed were astonishingly moderate, leaving the two proteins attached at the head of the spike. Nevertheless, it has to be assumed that they have relived critical interactions that, at higher pH prevent the fusion process.

During the biosynthetic pathway the two spikes has to pass through cell compartments of low pH. To escape premature activation for fusion, the E1 folds cotranslationally with the E2 precursor, termed p62. The p62 contains the sequence of the small E3 protein N-terminally to that of E2. In the linker region there is a furin cleave site that has to be cleaved to create the acid sensitivity needed for fusion activation during infection. A virus in which the maturation cleavage site is modified to prevent furin cleavage, may still be able to fuse, but at much lower pH (Salminen *et al.*, 1992); therefore, it is assumed that

p62 would protect the pre-spike structure against acid induced reconfiguration by establishing a tighter connection with E1 than the E2. However, it has never been shown that such a contact exists. In this study, I would like to approach the contact between p62 and E1 from the structural point of view, so as to explain how the precursor plays such an essential role for the maturation of the virion. To do so, the cryo-EM structure of a p62 cleavage-impaired mutant virus (SFVmSQL) where the furin cleavage site of p62 was impaired by amino acid substitutions (from R_{63}HRR to S_{63}HQL) (Berglund *et al.*, 1993) was determined and compared with earlier publish structure of SFVwt.

The pH profile of fusion peptide exposure

The exposure of the fusion peptide on acidification was demonstrated by virus binding to the mAbE1f (**Fig. 21**). This mAb recognizes the fusion peptide sequences, as discussed above (Hammar *et al.*, 2003). The pH profile of the mAb interaction shows an optimum for the SFVwt at pH 6.2, while the peak for the mutant appears first at below pH 5 (**Fig. 21A**). In addition, the comparison of SFVwt and SFVmSQL acid structures shows that the E3 domain of p62 prevents the swelling of the particle until a much lower pH is reached at pH ~4.3 (**Fig. 21B**).

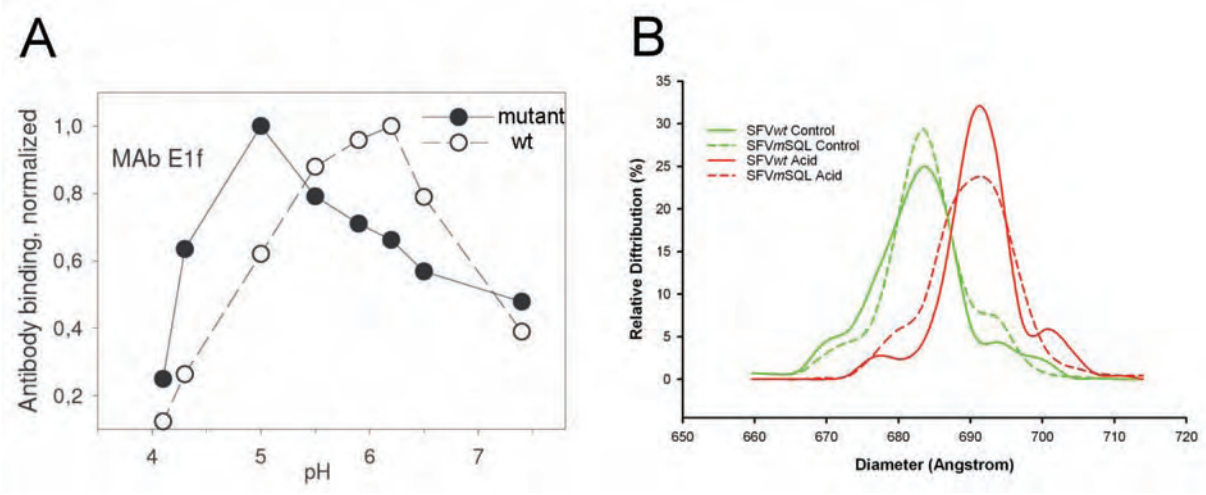

Fig. 21. *(A) Fusion peptide exposure relative to pH in SFVwt and the SFVmSQL particles, demonstrated by ELISA. (B) Cryo-EM reconstructions of SFVwt and SFVmSQL at pH 7.4 show a similar particle size distribution (green curves). The pH sensitive structure of the SFVwt virus responds to acidification by swelling and can be titrated down to pH 5.8 before particle aggregation. The acid resistant SFVmSQL is essentially unaffected and does not swell until close to pH 4.3, at which pH it show a similar size distribution (red hatched curve) as the SFVwt at pH 5.8 (solid red curve).*

Location of the E3 protein

The structure of SFVmSQL was obtained at a resolution calculated to 17 Å. For an accurate comparison between the two particles, the SFVwt structure was low-pass filtered back to 17 Å. By superimposing these two structures, it is seen that there is a good agreement in large parts. However, there is an extra density in the SFVmSQL structure, seen in the spike renderings in **Fig. 22A** as a white mesh. The main part of the difference structure protrudes out into the space between the spikes (open arrow). A careful analysis at different stringency reveals its details; **Fig. 22B** shows a radial section with density contours for sigma level 1 (σ=1) to σ=5, to demonstrate an additional structure assigned to the SFVmSQL. It extends from the sidewise bulky protrusion (open arrows) at radius 322 Å and continues downwards. At radius of 310 Å the solid arrow points to this

SFVmSQL specific node that finally connects to the shell region. We refer to this structure as the cotter-like node[5].

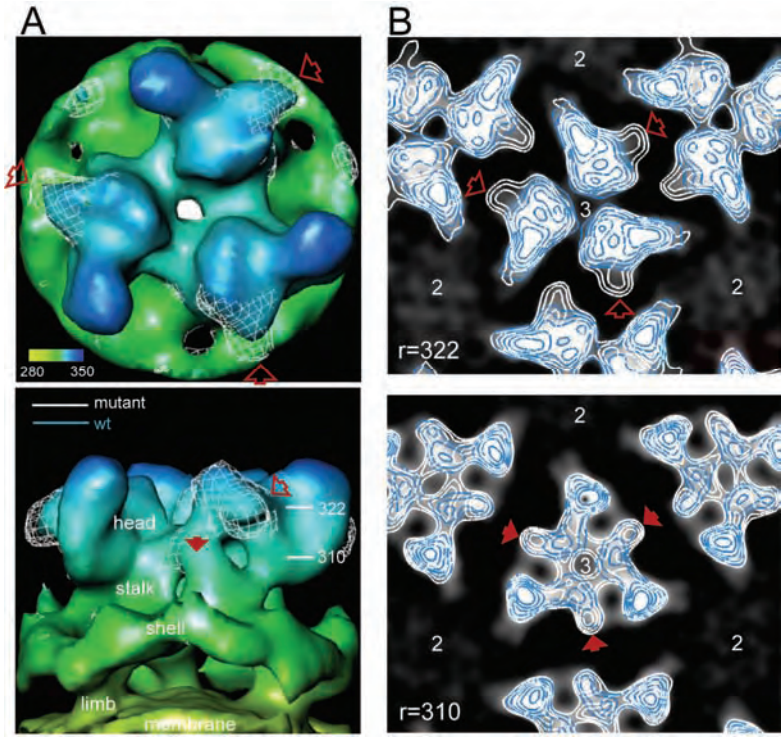

Fig. 22. *(A) Top (top panel) and side (bottom panel) views of the superimposed spikes of SFVwt (solid surface rendering) and SFVmSQL (white net). The protruding white net is the extra volume representing E3 density, as indicated by red arrows. (B) Superimposed radial section with density contours for σ=1 to σ=5 through the spike (same orientation as in A, top view) of SFVwt (blue contour) and SFVmSQL (white contour) at radii 322 Å (top panel) and 310 Å (bottom panel), to reveal mutant specific configurations. The solid and open arrows refer to the positions marked in A.*

[5] In an earlier paper I introduced "nodes" in the SFV wt spike structure Wu, S.R., L. Haag, L. Hammar, B. Wu, H. Garoff, L. Xing, K. Murata, and R.H. Cheng (2007) The dynamic envelope of a fusion class II virus. Prefusion stages of Semliki Forest virus revealed by electron cryomicroscopy. *J Biol Chem* **282**(9)**:** 6752-6762**.** These represent the more rigid domains in the structure.

E3 protein and its contact with E1

The spike structures of the SFVwt and SFVmSQL were superimposed at high stringency (σ=3) so that the details in the more stable domains can be compared. I found that the high stringency structure of the SFVmSQL is coinciding with that of the SFVwt, except for the specific sidewise bulky node appearing as a gripper (open red arrows in **Figs. 22 and 23**), a more tight association in between the spike wings (white stringency rings around the three fold spike center at r=310, **Fig. 22B**, bottom panel), and the cotter-like node (solid red arrows). Since the difference between SFVwt and SFVmSQL is the E3 domain, the E3 appears to protrude out from node *b* in the mutant spike (**Fig. 23**, bottom, left panel). The cotter-like node is connecting the gripper vertically down to the fitted E1 β-strand-a, β-strand-e and β-strand-f (not to be mixed with the node nomenclature), a hydrophobic cluster in E1 domain II (**Fig. 23**, bottom panel). These precursor-specific structures, i.e. the gripper and the cotter, represent high stringency structures of the E3 domain.

According to the fitted E1, the amino acid residues that would be very close or directly involved in the contact with the cotter-like structure are Leu44 and Ile47 in β-strand-a, and Tyr122 and Ser120 in β-strand-e, Val 178 in β-strand-f, Val208 and Glu209 in the h-h'-loop (**Fig. 23**, bottom, right panel). Especially, the Leu44 and Val178 of this domain have been observed in mutation studies to be related to lipid dependence (Chatterjee *et al.*, 2002). Finally, this hydrophobic surface domain finds, through a slight torsion, an internal location in the post-fusion trimeric form of the E1 (Mukhopadhyay *et al.*, 2006). Together, I therefore conclude that this cotter-like E3 structure constitutes a functional center in the spike.

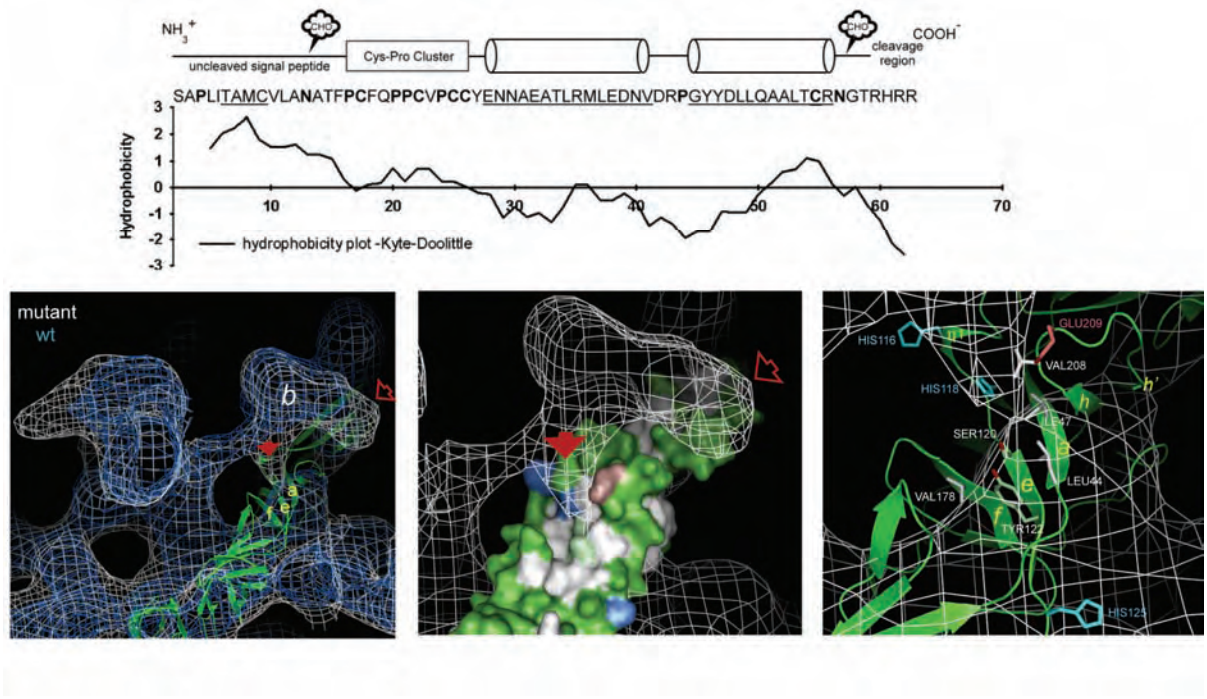

Fig. 23. *Structural elements in the E3 sequence and its Kyte-Doolittle hydropathy plot (top panel). The SFVmSQL specific high stringency cotter-like structure and its contact with E1 are shown in bottom panel. The SFVwt structure at high stringency is represented as a blue net, and the superimposed SFVmSQL structure as a white. Node b, in the E2 part of p62, is indicated by letter b. The mutant has a specific sidewise bulky protrusion and a cotter-like structure which goes vertically under the spike head, connecting down to the domain occupied by the fitted E1 structure, β-strands a, e, and f (bottom, left) with a patch of hydrophobic side chains (white surface, middle panel). The enlargement shows the cotter bolt structure of the E3 domain and its interaction with E1. The histidyls are in blue, Ser120 and Tyr122 in pale green, and Glu209 in light red.; His116, 118 and 125 are in blue.*

Cotter bolt theory and implications

The E3 (66 residues long) contributes the N-terminal part of p62. Its sequence and predicted secondary structure elements are given in the top panel of **Fig. 23**. The hydrophobic N-terminal region of the E3 sequence is functional as a signal sequence during p62 synthesis (Garoff *et al.*, 1990). I propose that the Pro-Cys cluster in the N-terminal half of the E3 sequence forms the bottom of the gripper structure and the hydrophobic signal peptide the cotter, connecting to the hydrophobic cluster in E1 (**Fig. 23**, bottom, middle panel). Two amphipatic helices, a predicted folding of the C-terminal half of the E3 sequence, folded back in an antiparallel fashion, would then form the bulky

gripper domain. Such a configuration is also supported by the structures connecting to the cotter. The hydrophobic signal sequence of E3 would be stabilized by hydrophobic effects created by the surrounding residues, and the N-terminal amino group of it be stabilized in a salt bridge with the γ-carboxyl of E1 Glu209; the hydroxyl of the N-terminal serine of the signal peptide sequence would fit in position to interfere with E1 Tyr122. By such a configuration the E3 domain would lock the release of E1 down to a pH providing carboxyl protonation, which is lower than that required for wt fusion. Thus, it is reasonable to interpret the cotter-like connection to be formed by the hydrophobic signal sequence constituting the N-terminal region of the E3 protein. This postulation explains the pH-resistance of the cleavage-impaired mutant virion - a structural finding and mechanistic view that has not been explored earlier.

A structural comparison of the SFVwt and SFVmSQL structures and detailed density analyses revealed that the E3 domain in the p62 provides a gripper-like structure, protruding out from the spike head into the space between the spikes and a cotter-like contact with a hydrophobic cluster in the E1 molecule. This configuration would prevent E1 from a premature collapse into homotrimers during the passage through the low pH cell compartments of the biosynthetic pathway. The cotter-like structure of the precursor locks the E1 and therefore gives raise to a pH resistant precursor spike. At the mature state, the furin cleavage would allow the release of E3 from the locking position, like the removal of a cotter bolt from its holder. In the mature spike the E1 still remains held by E2 at several points, to await further trigger, like acidification and membrane contact, to release the metastable spike complex in a way promoting virus and cell membrane fusion.

The activation of the Mo-MLV spike (paper III)

To study the activation of the Mo-MLV spike I analyzed the spike structure by EM in its non-activated form after triggering activation by Ca^{2+} depletion and arresting the process by alkylation at IAS, i.e. the isomerization arrested stage. The Ca^{2+} depletion mimics the receptor induced spike activation and the alkylator blocks the structural transition of the spike at an early stage by modifying the exposed CXXC-thiol before it attacks the intersubunit disulfide (Wallin *et al.*, 2004).

Isolation of native and activated intermediate forms of spikes in their trimeric forms

Mo-MLV was produced in MOV3 cells and purified by centrifugation in a sucrose step gradient. The virus was then solubilized in Triton X-100 buffer and the spikes isolated by centrifugation in a sucrose gradient. I used blue native polyacrylamide gel electrophoresis (BN-PAGE) to detect the trimeric forms of the disulfide linked SU-TM complexes, i.e. the complete spikes. For isolation of the native spikes the virus was solubilized on ice and in the presence of the spike stabilizing Ca^{2+}. Solubilization at 37°C could not be used as this resulted in dissociation of the trimer, $(SU\text{-}TM)_3$, into monomers, SU-TM, and dimers $(SU\text{-}TM)_2$. For isolation of the activated intermediate form of the spike the virus was solubilized at Ca^{2+} depleting conditions (EDTA) in the presence of the alkylator, NEM, and at 37°C. This isomerization arrested stage (IAS) is stabile at 37°C in contrast to the native form of the spike. In the sucrose gradient, which was used for spike isolation, it was possible to separate the trimeric forms of the SU-TM complexes from incomplete complexes and also from higher oligomeric forms. The trimers were then used for structural analyses by EM and image processing using EMAN package. As we isolated trimers for the analyses we imposed three fold symmetry to the model building.

Structures of native and IAS Env trimers

The spikes were analyzed by cryo- and negative stain-EM. The cryo-EM structure of the native spike, which was resolved at 19 Å, showed a hollow, fenestrated molecule, where the protomeric units united into a roof and floor, but separated on the sides (**Fig. 24**). One could distinguish a top, middle and bottom protrusion of the protomer. The height and

width of the trimeric molecule was around 90 Å. The IAS spike, also resolved at 19 Å, looked similar, but with a distinct change at the top. Here the top protrusions formed a ring-like structure instead of the roof seen in the native spike. This change opened up the internal cavity from above. Negative stain-EM, at the resolution of 22 Å, confirmed the major features observed by the cryo-EM.

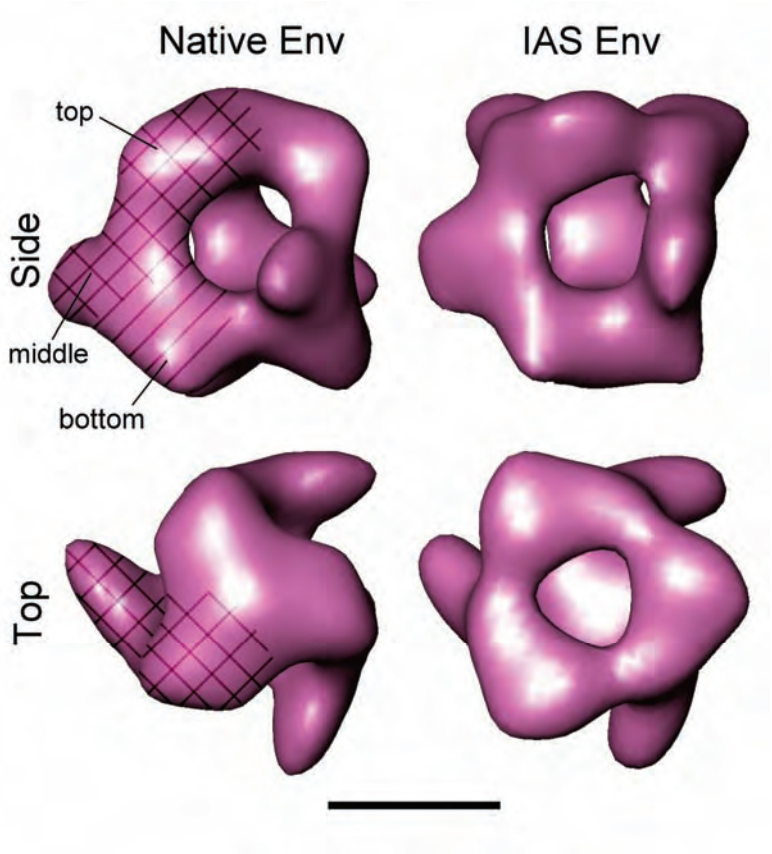

Fig. 24. *Side and top views of the surface rendered cryo-EM reconstructions of native (left) and the intermediate (IAS, right) spike trimers. The hatched region indicates one protomeric unit with the SU subunit cross-hatched and the TM subunit single-hatched. The three protrusions are indicated. Scale bar represents 5 nm. Notice the opening of the cavity from above after activation of the native spike into the IAS form.*

Identification of subunit domains in the spike reconstructions

Fortunately the atomic structure of the RBD of Friend (F)-MLV has been determined (Fass *et al.*, 1997). Its amino acid sequence is very similar to that of Mo-MLV and therefore I used it to make a molecular model of the Mo-MLV RBD. This was then used

for fitting into the density map of the native Mo-MLV spike. I found a unique fit of the RBD structure in the top protrusion of the protomer (**Fig. 25**). In here the tips of the bent finger-like RBD were slightly intercalated at the three fold axis in the roof region, exposing the receptor binding regions of the RBD at the top of the spike. This positioning of the RBD suggested that the middle protrusion most likely represented the SU C-terminal domain, with the disulfide isomerase activity, and the lower protrusion, including the floor were probably formed by the ecto-domain of the TM subunit. The transmembrane domain with bound detergent and the small endo-domain of TM I suppose are not visible in the reconstruction because of its postulated flexibility. The RBD structure model also found a unique fit into the top protrusion of the protomer in the IAS spike. However, in this case the tip of the bent finger-like RBD was facing the bend region of the RBD of a preceding protomer in a ring like organization. Thus, during activation the RBDs had rotated clockwise about 80° around and axis parallel to the three fold axis, opening up the spike cavity from above. At the bottom of this well-like structure the postulated TM seemed to reorganize into a thicker layer.

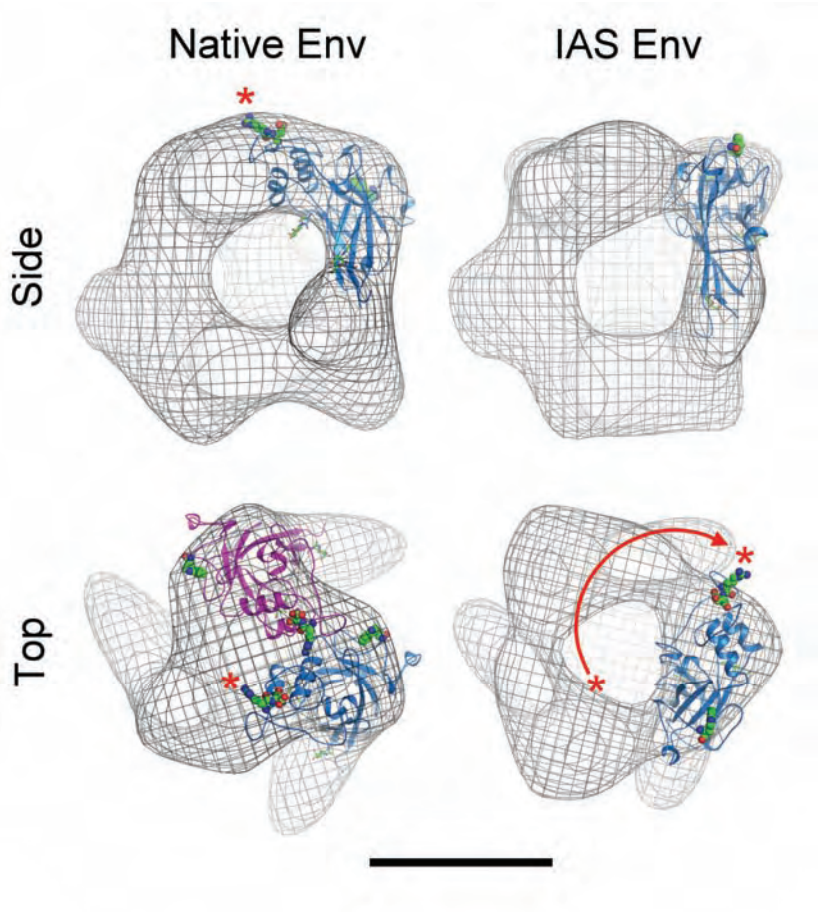

Fig. 25. *Fitting of the modeled atomic structure of the Mo-MLV RBD into the density maps of the native and IAS spikes. Note the outward rotation of the RBD that occurs upon switching from the closed native to open intermediate form of the spike (red arrow). The spheres in the ribbon structured RBD indicate receptor binding amino acid residues. For clarity only one or two RBD molecules are fitted. Scale bar represents 5 nm.*

The structure of the HIV-1 spike (paper IV)

To study the activation of the HIV-1 spike I so far only studied the structure of the spikes solubilized from virus by EM. My intention is to compare it with the spike, which has been complexed with a soluble form of the primary receptor CD4. The CD4 binding should trigger the spike into an intermediate form.

Isolation of HIV-1 spikes

The analysis of the HIV-1 spike created several problems not encountered with the Mo-MLV spike. Firstly, the HIV-1 spike was an extremely labile structure lacking a covalent association between its peripheral gp120 and transmembrane gp41 fusion subunits (McKeating *et al.*, 1991). Secondly, much less spikes were incorporated into the particle than into Mo-MLV (Zhu *et al.*, 2003). Thirdly, there were no transformed cell lines efficiently producing virus like Mo-MLV in MOV3. Fourthly, the risk of HIV-1 infection required containment of the laboratory work. Therefore, I used HIV-1 virus like particles, VLPs, with expression-blocking mutations in the gp gene and in the accessory genes vif, vpr and nef (Binley *et al.*, 2000). The gp gene was expressed from a separate unit and this was engineered such that the endo-domain was deleted to increase spike incorporation into particles and it carried the intersubunit disulfide of the SOSgp mutant to increase spike stability (Binley *et al.*, 2000). The VLPs carrying gps were produced in 293T cells transfected with corresponding DNAs. The particles were incompetent for spreading, although they could infect cells if DTT treated after binding to the primary and the secondary cell receptors. The VLPs were purified essentially as Mo-MLV, solubilized in Triton X-100 buffer and the spikes isolated by centrifugation in a density gradient. The trimeric form of the gp was detected by BN-PAGE and used for EM analyses.

The structure of the HIV-1 spike

The cryo-EM reconstruction of the HIV-1 spike structure showed a hollow cage-like molecule (**Fig. 26**). The resolution was estimated to 18 Å. When following one protomeric unit from the top of the molecule to the bottom, one could discriminate a common roof, a separated flat lobe and a leg on the side and a common bottom part. The

lobe was angled so that it formed a peripheral (outer) and central (inner) edge. The atomic structure of the gp120 core from its complex with the b12 Fab found a unique fit into the lobe (Zhou *et al.*, 2007). The stems of the deleted V3 loop and the V1-V2 loops were directed towards the roof and the gp41 binding region of the inner domain β-sandwich towards the leg below the lobe (**Fig. 27**, top panel). The binding site for the primary receptor, CD4, on the neutralizing face of the gp120 core was on the outer surface of the lobe, the extensively glycosylated silent face of gp 120 on the peripheral edge and the non-neutralizing phase on the central edge of the lobe (Wyatt *et al.*, 1998). The positioning suggested that the roof was composed of the variable loops 1-3 and showed that the leg and the bottom were composed of gp41. As with the Mo-MLV spike reconstruction I believe that the transmembrane segments with bound detergents were not resolved because of their postulated flexible nature. The gp41 endo-domain cannot be part of the structure as this has been deleted in the construct I used.

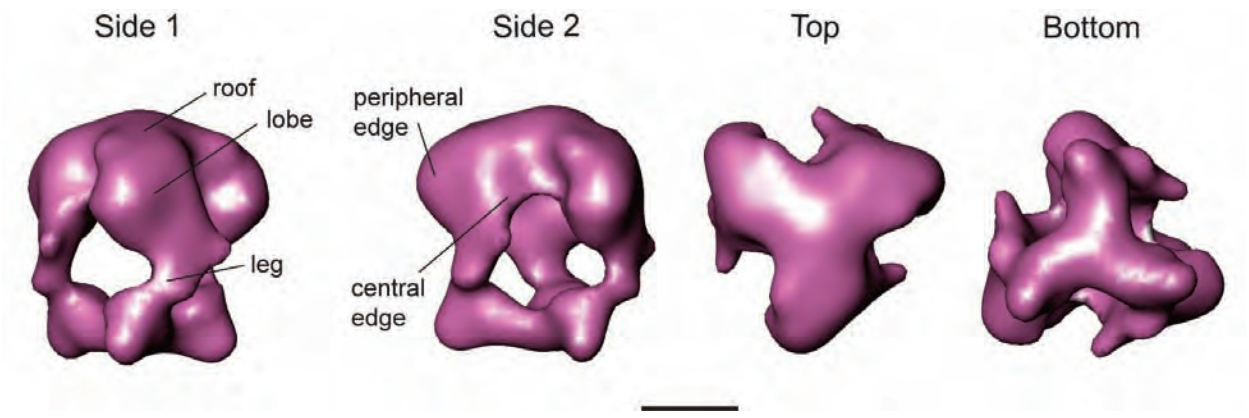

Fig. 26. *Surface rendered cryo-EM reconstructions of the HIV-1 spike from side, top and bottom views (left to right). The cavity roof, the lobe with its peripheral and central edges, the slanting leg and the bottom floor are indicated. Scale bar represents 5 nm.*

As I did not have a comparable reconstruction of the CD4 activated intermediate structure I was uncertain to what extent the solubilization of the HIV-1 spike preserved its native features. For this purpose I cross-linked the VLP particles with glutaraldehyde before

solubilization and spike isolation. The rationale was that if the spikes were cross-linked before solubilization then possible protein domain movements caused by the treatment should be prevented. I choose conditions for the cross-linking that coupled the gp120-gp41 units into trimers, as shown by SDS resistance in BN-PAGE, but still allowed trimer extraction from the particles by Triton X-100. The cryo-EM reconstruction of the spike structure revealed essentially the same features as those found in spikes from non-cross-linked virus. I also tested the fitting of the atomic structure of the CD4 activated gp120 core from its complex with the Fab of the neutralizing antibody X5 (Huang *et al.*, 2005). This gp120 core structure included the entire V3 loop. It fitted into the lobe almost exactly in the same position as the gp120 core from the Fab b12 complex (**Fig. 27**, bottom panel). However, the V3 loop and the stem of the V1-V2 loops pointed out of the spike density map. This suggested that my spike reconstruction did not correspond to the CD4 activated form of the spike. Finally, I made a biochemical experiment to follow whether the solubilized spike could be activated by CD4. The binding of the primary receptor has been shown to induce epitopes for certain neutralizing antibodies like 17b. Therefore, I analyzed the binding of the monoclonal 17b antibody to solubilized HIV-1 spikes before and after CD4 binding in a spike (trimer) band-shift assay using BN-PAGE. I found that the trimer band shifted quantitatively to a slower moving band when the spikes had been reacted with CD4. There was no reactivity with the 17b antibody without CD4. However, all of the CD4 bound spikes were able to bind 17b as shown by a further retardation of band migration. Thus, this showed that the solubilized spike I analyzed by EM was able to change into a 17b antibody binding conformation, consistent with it representing a native form of the spike.

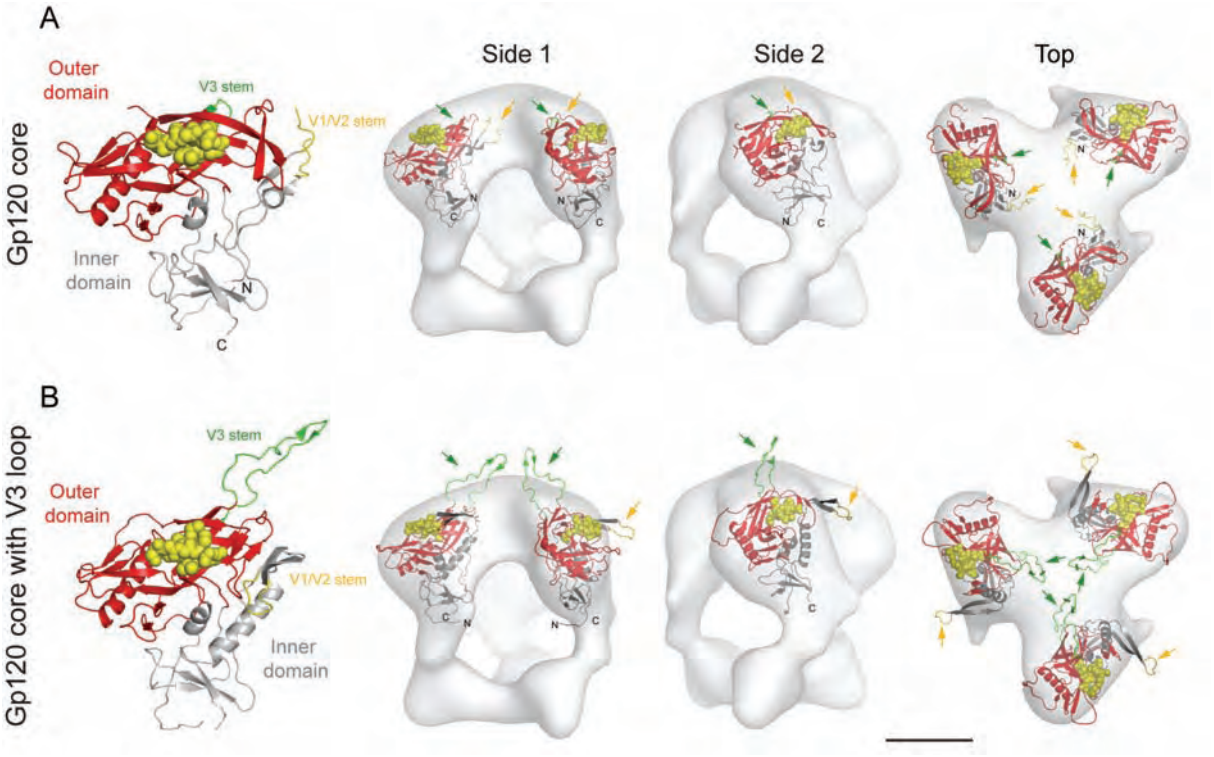

Fig. 27. *The atomic structure of the gp120 core and the gp120 core with V3 loop derived from its complex with the Fabs of the b12 and X5 antibodies were fitted into the HIV-1 spike density map in panels (A) and (B), respectively. The gp120 cores are shown in cartoon representation. The outer and inner domains are colored in red and gray, respectively. The stem of the V1-V2 loops is yellow and indicated by a yellow arrow while that of the V3 loop is green and indicated by a green arrow. The C- and N-termini of the truncated gp120 are also indicated. The CD4 binding loop is shown as yellow spheres. Scale bar=5 nm.*

GENERAL DISCUSSION AND FUTURE PERSPECTIVES

In this thesis my aim has been to reveal some of the structural interplay that occurs within a virus fusion protein as it prepares for facilitating fusion between virus and target cell membranes. The studies focus on the fusion protein in its functional assembly as a viral spike. The retroviruses Mo-MLV and HIV-1, and the alphavirus SFV, here exemplifies fusion proteins of class I and II, respectively. The structural data have been collected by cryo-EM and 3D reconstruction procedures of the whole virion or isolated spikes.

Structure determination

As my main approach, I used cryo-EM combined with 3D reconstruction processes to obtain the structures. Atomic structure information of my target viruses was available on protein fragments or artificial substructures. This has proven useful to classify the fusion protein and envision portions of the rearrangements that occur during the fusion process, especially at the final stages. However, the cryo-EM and single particle reconstruction approach make it possible to study protein dynamics in the context of the functional assembly of the whole spike, even in the context of the whole virus. Although this approach will not yield atomic resolution, I still can expect resolution in the 15-20 Å range, which can generate a wealth of information on the activation mechanisms at the level of individual protein domains. Moreover, where atomic structures of the subcomponents are available these can fit into our density maps to give a pseudo atomic structure.

The alphavirus spike activation

Implications from the post-fusion E1 structure

From a mechanistic point of view one can note that the post-fusion E1 homotrimers (Gibbons *et al.*, 2000; Gibbons *et al.*, 2003) are very stable molecular arrangements (Wahlberg *et al.*, 1992) , the formation of which would represent an essentially irreversible state of the process. However, for the final refolding to be possible, the E1 subunit has to be separated from the framework of the native spike, i.e. from its

interactions with the E2 subunit in the heterodimer and from its interactions between spikes. As recently pointed out (Roussel *et al.*, 2006; Wu *et al.*, 2007), and nicely demonstrated (Mukhopadhyay *et al.*, 2006) , the E1 has to turn around its long axis during the activation process, since in the post-fusion hairpin configuration the regions that in the spike are in contact with E2 towards the spike center are on the external surface of the post-fusion trimer, and structures that makes intermonomer contacts in the post-fusion state, are located away from the spike center in the non-triggered structure.

In the strictly organized SFV, the structural reorganization into the post-fusion state awakes many questions. The icosahedral symmetry in both the nucleocapsid and the envelope structure, their proven contact (Skoging *et al.*, 1996), and the large portion of the nucleocapsid protein that is embedded in the RNA genome, imply a relatively stable organization among its components. This well organized construct – how is it released in a fusion promoting way by the acid environment encountered in the endosome on cell entry?

The shell crack and rise on acidification

The most prominent feature of acid treatment of the SFV is the size expansion and the concomitant raise of the external parts of the envelope, above the bilayer membrane, demonstrated in the cryo-EM reconstructions of the virus structures (**Fig. 16**) (Haag *et al.*, 2002; Wu *et al.*, 2007). Although both effects may be related, and the raise of the spike be causing the cracking of the shell domain, it is plausible that all the effects are a result from histidyl protonation in the spike. This could result in repulsion between the shell forming parts, and thereby forcing an enlargement of the shell surface, which, to gain space, lifts the structure outwards. As discussed in the introduction and result sections, the shell is mainly constituted by the E1 domains I and III. More or less preserved histidyl side chains (Roussel *et al.*, 2006) are located at the rim of the separate spike shell regions (Lescar *et al.*, 2001). Thereby the intermolecular E1 contacts, together with the intramolecular contact between domains I and III would act as shell zippers. During passage through the low pH compartments of the ER-Golgi, on its way to the plasma

membrane, the pE2 (or p62) chaperoned E1 in the heterodimer or possibly in the trimeric spike, would not establish these intermolecular shell contacts, while at the neutral milieu outside the cell, the rims zip together forming the spike shell layer. Again, during cell entry through the endosomal pathway, here mimicked by low pH treatment, they unzip and the shell cracks. The histidyl side chain protonation, with its pK_a around pH 6, coincides with the pH threshold for susceptibility for fusion, observed for SFV. Accordingly, a reasonable contributor to the structural shell widening would be related to histidyl protonation.

The E3 part of p62 locks the spike stalk and shell dissociation

The maturation cleavage impaired SFVmSQL is not triggered by acidification in the way that the cleaved, wt virus is. The protective E3 part of p62, attached to the spike in the way here demonstrated, would not directly affect the histidyl protonation in the shell structure. However, it does prevent the swelling of the particle until a lower pH is reached (pH 4.3, **Fig. 21B**). Therefore, additional control points exist that affect the dissociation of the spike subunits and allows the shell to ascend. From the structural organization discussed in the result section, it would be assumed that this critical structure is located in the E2.

The E3 domain in the SFVmSQL configuration has a broad contact with the node *b* in the spike head (**Fig. 23**), which connects at an angle down and sideways to node *d*. This represents the part of the spike stalk region that keeps the E2 subunits together. In a top view of the spike, the stalk contact is released in the SFVwt at pH 5.8 (**Fig. 18**), but not in the SFVmSQL (data not shown). At pH 7.4 the SFVmSQL spike shows a somewhat tighter association in this region than present in the SFVwt (**Fig. 22**, the p62 specific white contours show connection around the three fold spike center). In consequence, it might be that the shell expansion is initiated by the dissociation of this stalk contact and that the E3 controls this process.

The p62 dimerization with E1 occurs in ER (Barth *et al.*, 1995) before transfer to the Golgi system, where p62 chaperones against acid induced self aggregation of E1. However, it is also here that trimerization might occur. Therefore, the p62 chaperon may not promote trimerization in ER, but allow trimers in an acidic milieu. One does not yet understand the nature of the modifications in the molecule that creates the trimerization, and it could well be that trimerization includes the signal for transfer out of ER. Trimers remain in the maturation-cleaved virus structure, but are here susceptible to dissociation on acidification around pH 6. I suggest that the E3 contacts, both the cotter-like and the gripper-like connections, indicate control points that become released by the maturation cleavage. In the blocked state, both are permitting activation trigger and fusion in a more acidic environment. Therefore, when that happens, the N-terminal domain of the E3 sequence should move out from the main structure, not to shield the direct interaction with the target membrane.

Internal spike contacts are retained, while the fusion peptide is exposed

In addition to the shell crack and rise, there is another quiz in the acid triggered, activated state of the virus, resulting from my observations. It is obvious that the contact between the two subunits in the head of the spike is retained at pH 5.8 (**Figs. 16 and 17**), a fusion permissive pH. The E2 location at the top of the spike and that of the fusion peptide at the rim of the spike under the E2 site are confirmed by Fab probing (**Figs. 19 and 20**). The epitopes for these antibodies have been mapped to E2 sequence 211-218 and E1 sequence 85-95, respectively (Hammar *et al.*, 2003). The fusion peptide represents the E1 residues 83-100 (Garoff *et al.*, 1980). Functional assays with the E1 ecto-domain, E1*, have shown that mAbE1f binding at neutral pH prevented subsequent low-pH-triggered E1* interaction with target membranes and trimerization (Gibbons *et al.*, 2004a). In the present study I noted that bound Fab E1f allowed a further acidification of the virus, before particle aggregation occurred, than what was otherwise possible (not shown). Therefore, whilst the fusion peptide was exposed and both the shell and the internal spike assembly were loosened, the top of the spike remained as one morphological unit. This implies that additional activation steps, after fusion loop exposure, are needed before E1 relocates into

the assumed extended homotrimer and folds into a hairpin conformation. It could be that the membrane insertion of the fusion loop of E1 is necessary for the complete release from E2, homotrimerization and backfolding. Here, cholesterol in the target membrane, which has been reported to be required (Ahn *et al.*, 2002; Chatterjee *et al.*, 2002; Kielian *et al.*, 2000; Lu &Kielian, 2000; Samsonov *et al.*, 2002; Umashankar *et al.*, 2008; Waarts *et al.*, 2002), could be the final trigger.

The retroviral spike

The retroviral spikes are hollow cage-like structures with separated legs

My studies about the spike structure of two retroviruses, the Mo-MLV and the HIV-1, show that these are remarkably hollow structures, where the three protomeric units form a cage-like surface (**Fig. 28**). In both reconstructions the top parts of the protomeric units assemble into a common cage-roof. In the case of Mo-MLV the top part is represented by the bent finger-like protrusion, RBD (i.e. the N-terminal part of SU) and in the case of HIV-1 probably by the V3 and the V1-V2 loops. The protomers continue down the sides of the cage-like molecule forming a lobe (HIV-1) or a middle protrusion (Mo-MLV). The HIV-1 lobe is represented by the gp120 core and the Mo-MLV protrusion most likely by the C-terminal isomerase domain of SU. Further down the spike protomers form leg-like structures that lead to the lower part of the cage. Here the legs transform into protrusions pointing downwards and into a cage-bottom with three fold interactions. The legs and the bottom part are formed by the HIV-1 gp41 subunit and probably by the Mo-MLV p15E subunit, respectively. Thus, the retrovirus spikes are kept together by three fold interactions only at the top and bottom, whereas most of the protomers are separated from each other on the sides.

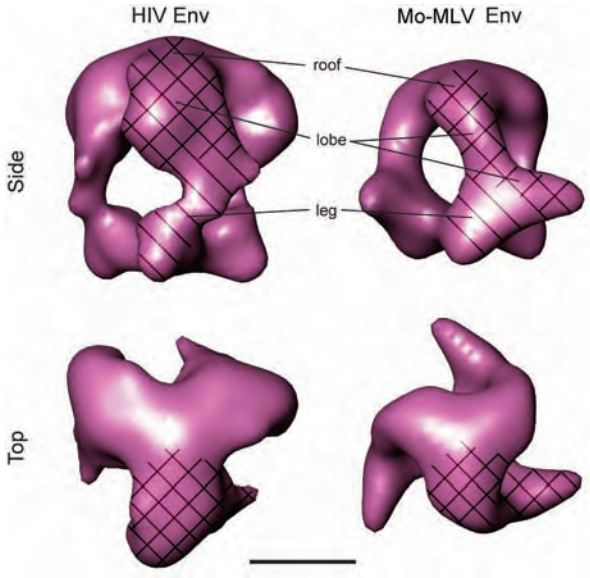

Fig. 28. *Comparison of cryo-EM reconstructions of HIV-1 and Mo-MLV spikes. The gp41 and the TM parts of the protomeric unit of each spike are hatched and the gp120 and SU parts cross-hatched. The spike roof, lobe and leg are indicated. Scale bar=5 nm.*

Validity of the structures

The facts that the two retrovirus spikes showed a similar structure and, that this was obtained using both cryo- and negative stain-EM, makes me confident about the reconstructions. Furthermore, the unique fitting of the gp120 core into the lobe of the HIV-1 spike and the RBD into the top protrusion of the Mo-MLV spike validate at least these parts of the reconstructions. One might argue that my model building was biased, because I imposed three fold symmetry from the beginning of the image processing procedure. However, as I specifically isolated trimers from the gradient I find that this was correct. A further concern was whether solubilization of the spikes from the viral membrane with the mild detergent Triton X-100 caused structural changes in the structure. In particular in the case of the HIV-1 spike, where I could not compare the native spike structure with an activated intermediate structure like in the case of the Mo-MLV spike, this was problematic. Therefore, I also reconstructed the 3D structure of the HIV-1 spike using gps that were extracted from cross-linked VLPs, where structural changes should be restricted. This was found to be very similar to the gp reconstruction

using non-crosslinked virus. Furthermore, I showed biochemically that the non-cross-linked solubilized spike was able to react with a soluble form of the primary HIV-1 receptor CD4, in a way that induced the expected conformational changes in the gp120 subunit. However, it will be important to solve the cryo-EM structure of the solubilized CD4-spike complex of HIV-1 and document structural changes to settle this question.

The transmembrane segments

The Mo-MLV spike has a transmembrane segment with a 33 amino acid residues long cytoplasmic domain at its C-terminal end and the mutated HIV-1 spike a transmembrane segment without its normal endo-domain. In both cases I believe that these protein domains are not visualized in my reconstructions. This is also the case with the Triton X-100 attached to the transmembrane segment. I expect these structures to extend out from the bottom projections seen in both reconstructions, but because of their expected flexibility I believe they remain invisible after image averaging. The same might be true for most of the protein bound carbohydrate.

The cryo-electron tomographic spike models

The studies of retrovirus spikes *in situ* in virus particles, by cryo-electron tomography, have suggested two morphologically different types of spikes; one with a compact head with tripod legs standing on the membrane and another one with a cavity containing head connected to the membrane by a compact stalk **(Fig. 15)** (Forster *et al.*, 2005; Liu *et al.*, 2008; Zanetti *et al.*, 2006; Zhu *et al.*, 2006). At present this contradiction is not settled. One possible explanation is that the averaging process during image processing has obscured the presence of separated legs in studies where a compact stem was found (Zhu *et al.*, 2008). This could be the case if the spikes studied have had an incomplete oligomeric structure, as suggested by biochemical studies (Moore *et al.*, 2006). My single particle reconstruction clearly highlights the separated legs and the cavity containing head of the earlier models, but not their compact features. Indeed, among all of the viral

spikes known so far, my model with the cage-like hollow spike structure appears unique (Lee *et al.*, 2008; Roche *et al.*, 2007; Wilson *et al.*, 1981; Yin *et al.*, 2006).

Interprotomeric interactions

One important question about my spike structure is the nature of the interprotomeric interactions, which are restricted to the top and bottom parts of the cage-like molecule. In the Mo-MLV spike the top interactions seem to be mediated by the tips of the RBD. However, the interactions must be easy to dissociate to facilitate formation of the IAS form of the spike, where the RBDs have rotated outward. The RBD tips also contain plenty of variable loops, which are less suitable for molecular interactions (Fass *et al.*, 1997). As Ca^{2+} ion(s) have been shown to stabilize the native form of the spike, it is tempting to suggest that Ca^{2+} ions are mediating the three fold interactions (Wallin *et al.*, 2004). Further, according to this model receptor binding to the RBD tips disturbs the coordination of the ion(s) and causes the subsequent RBD relocation. In the case of the HIV-1 spike the V3 loop might have an analogous function to RBD in Mo-MLV, although Ca^{2+} might not be involved. The V3 is known to be an extended structure that could mediate three fold interactions via some conserved elements of its otherwise very variable structure (Huang *et al.*, 2005). The V3 loop was not included in the gp120 core from the b12 Fab complex used for fitting in my study, but the remaining stub was pointing into the unassigned roof density. The roof location of the V3 loop was also supported by the fitting of the V3 containing gp120 core from its activated CD4 complex, although the triggering seemed to have lifted the V3 tips up from the density formed by the roof of the native spike. The fitting using the b12 complexed gp120 core also pointed the V1-V2 stub into the roof, suggesting a corresponding location and the possibility for the loops to mediate interprotomeric interactions, either dimeric or trimeric ones.

The three fold interactions at the cage bottom of the Mo-MLV and HIV-1 spike structures are much more enigmatic than those at the top. However, they must be formed by the transmembrane subunits p15E and gp41. These are both disulfide bonded in the middle of their ecto-domains to the SU C-terminal isomerase domain and to the gp120 lobe,

respectively. Therefore it is reasonable to believe that both the N- and C-terminal parts of the p15E and gp41 ecto-domains descend the leg together to the bottom part of the cage. Here the C-terminal part might continue to the bottom protrusion, whereas the N-terminal part might make the three fold interactions.

Functional considerations

The fitting of the gp120 core into the HIV-1 spike reconstruction places the primary CD4 receptor binding site well exposed on the upper-outer surface of the lobe. The three sites are separated by about 70 Å, which means that three CD4 molecules could bind simultaneously. This binding is expected to induce structural changes in the spike. Among others the V3 and the V1-V2 loops should reorganize so that the former are exposed to the co-receptor, the chemokine receptor, binding. In my model this would occur close to the three fold axis. In here there would be space for this interaction even in the presence of bound CD4. The formation of corresponding complexes has been demonstrated biochemically (Mkrtchyan *et al.*, 2005). The combined receptor interactions with the gp120 subunit should then activate gp41 refolding. I believe that this is initiated by extension of the three fold interaction in the bottom region. This is postulated to be represented by a short coiled-coil formed by a portion of the N-terminal part of the gp41 ecto-domain. After activation further parts of it might be available for coiled-coil formation. Thus, it is possible that the pre-hairpin structure starts to grow inside the hollow cage structure. As I assume that the gp41 still remains attached to the gp120 by the $C(X)_5C$ region in the middle of the ecto-domain, and that the reclocation of the N-terminal part for coiled-coil formation might transfer the coil towards the roof of the cage. Here the receptor binding might have displaced the V1-V3 loops from the central roof region making space for the fusion peptides at the tip of the growing coiled-coil to reach and interact with the lipid bilayer of the cell membrane. First at this critical stage the gp41 and gp120 interaction is dissociated. This leads to the release of the gp120-receptor complex from the fusion site and the possibility for the gp41 to further refold from its pre-hairpin configuration in the extended coiled-coil into a hairpin conformation causing membrane fusion.

Although I have no data about the structural changes occurring in the HIV-1 spike during activation, my reconstructions of the native and the IAS form of the Mo-MLV spike lend support for this kind of a general retrovirus spike activation model. Firstly, in the case of Mo-MLV only a cell membrane-close receptor molecule, the basic amino acid transporter, has been characterized (Wang *et al.*, 1991). This binds to the RBD tips close to the three fold axis of the roof region in the spike, much like the co-receptor is thought to bind to the V3 loop in the HIV-1 spike (Fass *et al.*, 1997). Secondly, in the activated IAS form of the Mo-MLV spike, where the $C(X)_6CC$ region, in the middle of the TM ecto-domain, is still disulfide linked to the CXXC motif of the C-terminal SU domain, there was a noticeable thickening of the bottom region (**Fig. 24**, top panel). This might reflect N-terminal coiled-coil formation. Thirdly, during activation the RBDs rotated outward, opening up the roof for a potential protrusion of the fusion peptide at the tip of a TM coiled-coil. In the case of the Mo-MLV there is an elegant mechanism to control the SU-TM association, i.e. through intersubunit disulfide isomerization. However, the dissociation should not occur before TM refolding, i.e. coiled-coil formation, has proceeded to a stage of fusion peptide insertion into the target membrane. The isomerization reaction requires that the free CXXC thiol in the C-terminal SU domain is activated to attack the intersubunit disulfide. For this a structural change must take place in its surrounding that decreases its pK_a value. The change is probably caused by the RBD rotation that follows receptor binding. In the case of HIV-1, which does not have an isomerization prone intersubunit disulfide one can speculate that V3 loop relocation after co-receptor binding causes corresponding changes in the gp41 binding region of the gp120 inner domain, which affects subunit association.

The structures of the retrovirus spikes I have determined also have profound implications on antibody binding. In the case of HIV-1 the gp120 fitting in the lobe was such that the heavily glycosylated and immunologically silent face was exposed on its peripheral and mostly accessible surface (Wyatt *et al.*, 1998). The neutralizing face of the gp120, with the CD4 binding site was exposed on the external flat surface of the lobe. However, here so called conformational masking makes it possible only for a restricted set of antibodies

to bind (Kwong *et al.*, 2002). The large amount of non-neutralizing anti HIV-1 antibodies is directed to the inner domain of the molecule. This was localized to the inner and lower edge of the lobe where it interacted with gp41. Thus the structure of the spike can explain the general way antibodies react. Furthermore, its hollow nature with the possible generation of an initial gp41 pre-hairpin inside offers an interesting possibility for this early intermediate to escape antibody attacks.

Future perspectives

One of the most important challenges for further understanding of the structure and function of the retrovirus spike appears to be the identification of protein regions, in particular those of the transmembrane subunit (gp41 and p15E), in my structure models. What protein region is forming the three fold interaction at the bottom of the cage-like spike structures? Is it the N-terminal region of the TM/gp41 ecto-domain, whereas the C-terminal region is making up the peripheral, downward directed protrusions at the bottom as proposed? Fab mapping represents one approach to solve the question. However, as already mentioned, sterical hindrance inside the cage might create problems. Another important challenge is to use my EM approach to solve the structure of the CD4 activated form of the HIV-1 spike. This might show the opening of the cage roof by the postulated V1-V3 loop reorganization as well as bottom thickening by the proposed gp41 coiled-coil formation. In general I think that the isolation of the trimeric forms of the spike proteins in their native or activated intermediate forms and structure analyses by single particle cryo-EM might be superior to the *in situ* analysis of the spike structure by cryo-electron tomography as an approach to study the retrovirus spike structure. The advantages of the former approach appear to be more homogenous material and better final resolution of the reconstruction. As a third challenge I like to mention the elucidation of those structural changes in the peripheral subunit that leads to the release of the interaction with the transmembrane subunit. However, this question will require atomic resolution, which is difficult to reach by EM. Furthermore, it will probably also require more advanced intermediate forms of the spikes than the IAS spike of Mo-MLV and the CD4 activated spike of HIV-1.

ACKNOWLEDGEMENTS

The work presented here has been carried out in the Department of Biosciences and Nutrition in Karolinska Institutet.

I would like to express my deepest gratitude to my supervisors. I am very fortunate to have two wonderful mentors guiding me every step in my study.

Professor Henrik Garoff, thank you very much for taking me as part of your group. Without your support and guidance, I would not be able to achieve my PhD. I admire your clear mind and wide-ranging knowledge. There are lots of things I have learned from you, the greatest scientist in the virology field.

Associate Professor Lena Hammar, I cannot thank you enough. You, like a mother, have enormous patience on me and always listen to me. Those nights we spent together on paper writing are always in my mind. I deeply appreciate your constant encouragement. Without you, I would not know where I was.

I have the pleasure to work with the fantastic former and current colleagues:

Mathilda Sjöberg, your well-organized experiments and high productivity truly impress me. Thank you for your great support through all projects. Without your input, none of these works would be possible. I also thank you a lot for revising for my thesis. **Michael Wallin,** the works we performed together really lift me up to a higher level. Thanks for the enlightening discussions and all the valuable hints during the work. I believe that we can accomplish the following missions together. **Birgitta Lindqvist,** you are such an energetic person. I really want to get some from you! Thanks for creating such a happy atmosphere in the lab. I also look forward to learning the excellent experimental skills from you. **Robin Löving,** thanks for your patience answering my stupid bio- questions when I started the wet bench. Running the reconstruction process is like taking drugs, it definitely is. **Rickard Nordström,** I am impressed by your brilliant intelligence. I wish you the best in your future. **Kejun Li** and **Shujing Zhang,** thanks for all your supports during these years and all the thoughtful words. Those really warm me. I

love the flowers you planted, and I will continue taking care of them. **Malin Kronquist,** thank you for creating such a warm atmosphere in the lab and for the great lab supports. **Maria Ekström,** thanks for all high-quality sample preparation. **Helena Andersson** and **Maria Hammarstedt,** thanks for all the encouragements and advices. I really wish you could stay a little longer. I wish you all the best in your future careers. **Lars Haag** and **Josefina Nilsson,** I am fortunate to have such fantastic friends and colleagues like you. Thanks for all the inspirations and encouragements that make me believe I can achieve the goal. I have crossed the hill that I would never dream about. I believe that our friendship will continue forever no matter where we are. **Leif Bergman,** for helping me set up software for the computations and for sharing me the beauty of programming scripts. I also thank **Bomu Wu,** for the instructions on data processing and **Li Xing,** for the excellent data collection and helpful advices.

I would like to express my gratitude to the group members in Hebert's group:
Professor Hans Hebert, I enjoy a lot from our collaboration. Thanks for the enlightening discussions in our joint meetings and for providing excellent EM core facilities. **Philip Koeck,** thanks for patiently instructing me how to use the data processing programs and the EM technique, and for always answering my queries. **Kimberley Cheng,** thanks for sharing me all the important skills for EM. I really hope we can have the dancing course again. You are such a fantastic dancing teacher and I will never be absent from your course. **Pasi Purhonen,** thanks for always being there when I need help in the microscopy room and all the technical supports. **Caroline Jegerschöld,** thanks for being my half-time committee and the valuable criticisms. **Qie Kuang,** thanks for all the encouragements. I wish you the best in your future study.

I also would like to thank all personnel in the Department of Biosciences and Nutrition in Karolinska Institutet for the great assistance and for establishing such a warm working environment.

I would like to express my gratitude to **Professor R. Holland Cheng,** for introducing me to the structural virology field. I continue it and never give up. I am also grateful to the group members in UC Davis for all the helpful discussions and advices.

I am deeply grateful to: **Antony, Carol, Fei, Chieh-Hsin, Te-Yu, Helena, Yi-Ta, Jack** and others from the Taiwanese student association, I really enjoy the trips, BBQ, lunches, and dinners we had together. My life would have been boring without you. **Vivian** and **Stacy** in the Cultural Division of Taipei Mission, I appreciate all the valuable information and the supports. **Julie,** thanks for the fantastic Christmas dinners; those warm my heart. **Alka,** thanks for the funs we had together. I am always thinking of the great milk tea and Indian food you prepared. I wish you all the best in your future, and we will keep in touch forever. **Torbjörn,** thanks for many wonderful things you have brought to me. I cherish them all. I miss a lot the great families you have; they will be always in my mind. Tack så mycket! Many thanks to **RuRu, Entz, Cheng-Yu,** for having been such closest friends since we are teenagers. Thanks for giving me all the supports whenever I needed it. Because of you, I realize the true friendship. Remember our promise that our friendship will continue no matter what happen in the future. I would like to express my special thank to **Che-Yen,** all the words are not enough to express my gratitude. Thanks for being with me throughout my ups and downs. You really light up my life.

The biggest thanks go to my beloved families, without your support, I have nothing.
My sisters, **Huei-Ru** and **Ai-Jen,** I do not think there will be any sisters greater than you in the world. It is my fortunate to have you as my sisters.
My parents, **Shuei-Shiang Wu** and **Bi-Yun Weng,** for your endless support and love throughout my entire life. Thanks for always being with me. I love you!

The Swedish Knowledge Foundation through the Industrial PhD programme in Medical Bioinformatics at Corporate Alliances, Karolinska Institutet for the financial support is gratefully acknowledged.

Shang-Rung Wu in Stockholm, 2009

REFERENCES

Abrahamyan, L.G., R.M. Markosyan, J.P. Moore, F.S. Cohen, and G.B. Melikyan (2003) Human immunodeficiency virus type 1 Env with an intersubunit disulfide bond engages coreceptors but requires bond reduction after engagement to induce fusion. *J Virol* **77**(10)**:** 5829-5836

Adrian, M., J. Dubochet, J. Lepault, and A.W. McDowall (1984) Cryo-electron microscopy of viruses. *Nature* **308**(5954)**:** 32-36

Ahn, A., D.L. Gibbons, and M. Kielian (2002) The fusion peptide of Semliki Forest virus associates with sterol-rich membrane domains. *J Virol* **76**(7)**:** 3267-3275

Albritton, L.M., L. Tseng, D. Scadden, and J.M. Cunningham (1989) A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. *Cell* **57**(4)**:** 659-666

Allison, S.L., J. Schalich, K. Stiasny, C.W. Mandl, C. Kunz, and F.X. Heinz (1995) Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. *J Virol* **69**(2)**:** 695-700

Andersson, H., B.U. Barth, M. Ekstrom, and H. Garoff (1997) Oligomerization-dependent folding of the membrane fusion protein of Semliki Forest virus. *J Virol* **71**(12)**:** 9654-9663

Backovic, M., and T.S. Jardetzky (2009) Class III viral membrane fusion proteins. *Curr Opin Struct Biol* **19**(2)**:** 189-196

Bae, Y., S.M. Kingsman, and A.J. Kingsman (1997) Functional dissection of the Moloney murine leukemia virus envelope protein gp70.PG - 2092-9. *J Virol* **71**(3)**:** 2092-2099.

Baker, T.S., and R.H. Cheng (1996) A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *J Struct Biol* **116**(1)**:** 120-130

Barnard, R.J., D. Elleder, and J.A. Young (2006) Avian sarcoma and leukosis virus-receptor interactions: from classical genetics to novel insights into virus-cell membrane fusion. *Virology* **344**(1)**:** 25-29

Barnett, A.L., and J.M. Cunningham (2001) Receptor binding transforms the surface subunit of the mammalian C-type retrovirus envelope protein from an inhibitor to an activator of fusion.PG. *J Virol* **75**(19)**:** 9096-9105.

Barth, B.U., and H. Garoff (1997) The nucleocapsid-binding spike subunit E2 of Semliki Forest virus requires complex formation with the E1 subunit for activity. *J Virol* **71**(10)**:** 7857-7865

Barth, B.U., J.M. Wahlberg, and H. Garoff (1995) The oligomerization reaction of the Semliki Forest virus membrane protein subunits. *J Cell Biol* **128**(3)**:** 283-291

Berglund, P., M. Sjoberg, H. Garoff, G.J. Atkins, B.J. Sheahan, and P. Liljestrom (1993) Semliki Forest virus expression system: production of conditionally infectious recombinant particles. *Biotechnology (N Y)* **11**(8)**:** 916-920

Binley, J.M., C.S. Cayanan, C. Wiley, N. Schulke, W.C. Olson, and D.R. Burton (2003) Redox-triggered infection by disulfide-shackled human immunodeficiency virus type 1 pseudovirions. *J Virol* **77**(10)**:** 5678-5684

Binley, J.M., R.W. Sanders, B. Clas, N. Schuelke, A. Master, Y. Guo, F. Kajumo, D.J. Anselma, P.J. Maddon, W.C. Olson, and J.P. Moore (2000) A recombinant human immunodeficiency virus type 1 envelope glycoprotein complex stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits is an antigenic mimic of the trimeric virion-associated structure. *J Virol* **74**(2)**:** 627-643

Bressanelli, S., K. Stiasny, S.L. Allison, E.A. Stura, S. Duquerroy, J. Lescar, F.X. Heinz, and F.A. Rey (2004) Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *EMBO J* **23**(4)**:** 728-738

Brody, B.A., S.S. Rhee, and E. Hunter (1994) Postassembly cleavage of a retroviral glycoprotein cytoplasmic domain removes a necessary incorporation signal and activates fusion activity. *J Virol* **68**(7)**:** 4620-4627

Bron, R., J.M. Wahlberg, H. Garoff, and J. Wilschut (1993) Membrane fusion of Semliki Forest virus in a model system: correlation between fusion kinetics and structural changes in the envelope glycoprotein. *EMBO J* **12**(2)**:** 693-701

Bullough, P.A., F.M. Hughson, J.J. Skehel, and D.C. Wiley (1994) Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**(6492)**:** 37-43

Carr, C.M., C. Chaudhry, and P.S. Kim (1997) Influenza hemagglutinin is spring-loaded by a metastable native conformation. *Proc Natl Acad Sci U S A* **94**(26)**:** 14306-14313.

Chanturiya, A., L.V. Chernomordik, and J. Zimmerberg (1997) Flickering fusion pores comparable with initial exocytotic pores occur in protein-free phospholipid bilayers. *Proc Natl Acad Sci U S A* **94**(26)**:** 14423-14428

Chatterjee, P.K., C.H. Eng, and M. Kielian (2002) Novel mutations that control the sphingolipid and cholesterol dependence of the Semliki Forest virus fusion protein. *J Virol* **76**(24)**:** 12712-12722

Chernomordik, L.V., and M.M. Kozlov (2003) Protein-lipid interplay in fusion and fission of biological membranes. *Annu Rev Biochem* **72:** 175-207

Chernomordik, L.V., and M.M. Kozlov (2008) Mechanics of membrane fusion. *Nat Struct Mol Biol* **15**(7)**:** 675-683

Chernomordik, L.V., G.B. Melikyan, and Y.A. Chizmadzhev (1987) Biomembrane fusion: a new concept derived from model studies using two interacting planar lipid bilayers. *Biochim Biophys Acta* **906**(3)**:** 309-352

Colman, P.M., and M.C. Lawrence (2003) The structural biology of type I viral membrane fusion. *Nat Rev Mol Cell Biol* **4**(4)**:** 309-319

de Curtis, I., and K. Simons (1988) Dissection of Semliki Forest virus glycoprotein delivery from the trans-Golgi network to the cell surface in permeabilized BHK cells. *Proc Natl Acad Sci U S A* **85**(21)**:** 8052-8056

Doranz, B.J., J.F. Berson, J. Rucker, and R.W. Doms (1997) Chemokine receptors as fusion cofactors for human immunodeficiency virus type 1 (HIV-1). *Immunol Res* **16**(1)**:** 15-28

Dubochet, J., M. Adrian, J.J. Chang, J.C. Homo, J. Lepault, A.W. McDowall, and P. Schultz (1988) Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* **21**(2)**:** 129-228

Eckert, D.M., and P.S. Kim (2001) Mechanisms of viral membrane fusion and its inhibition. *Annu Rev Biochem* **70:** 777-810

Fass, D., R.A. Davey, C.A. Hamson, P.S. Kim, J.M. Cunningham, and J.M. Berger (1997) Structure of a murine leukemia virus receptor-binding glycoprotein at 2.0 angstrom resolution.PG - 1662-6. *Science* **277**(5332)**:** 1662-1666.

Fass, D., S.C. Harrison, and P.S. Kim (1996) Retrovirus envelope domain at 1.7 angstrom resolution. *Nat Struct Biol* **3**(5)**:** 465-469.

Felkner, R.H., and M.J. Roth (1992) Mutational analysis of the N-linked glycosylation sites of the SU envelope protein of Moloney murine leukemia virus. *J Virol* **66**(7)**:** 4258-4264

Feng, Y., C.C. Broder, P.E. Kennedy, and E.A. Berger (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**(5263)**:** 872-877

Forster, F., O. Medalia, N. Zauberman, W. Baumeister, and D. Fass (2005) Retrovirus envelope protein complex structure in situ studied by cryo-electron tomography. *Proc Natl Acad Sci U S A* **102**(13)**:** 4729-4734 Epub 2005 Mar 4717

Fuller, S.D., J.A. Berriman, S.J. Butcher, and B.E. Gowen (1995) Low pH induces swiveling of the glycoprotein heterodimers in the Semliki Forest virus spike complex. *Cell* **81**(5)**:** 715-725

Gallay, P., T. Hope, D. Chin, and D. Trono (1997) HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc Natl Acad Sci U S A* **94**(18)**:** 9825-9830

Garoff, H., A.M. Frischauf, K. Simons, H. Lehrach, and H. Delius (1980) Nucleotide sequence of cdna coding for Semliki Forest virus membrane glycoproteins. *Nature* **288**(5788)**:** 236-241

Garoff, H., D. Huylebroeck, A. Robinson, U. Tillman, and P. Liljestrom (1990) The signal sequence of the p62 protein of Semliki Forest virus is involved in initiation but not in completing chain translocation. *J Cell Biol* **111**(3)**:** 867-876

Garoff, H., J. Wilschut, P. Liljestrom, J.M. Wahlberg, R. Bron, M. Suomalainen, J. Smyth, A. Salminen, B.U. Barth, H. Zhao, and et al. (1994) Assembly and entry mechanisms of Semliki Forest virus. *Arch Virol Suppl* **9:** 329-338

Geyer, R., J. Dabrowski, U. Dabrowski, D. Linder, M. Schluter, H.H. Schott, and S. Stirm (1990) Oligosaccharides at individual glycosylation sites in glycoprotein 71 of Friend murine leukemia virus. *Eur J Biochem* **187**(1)**:** 95-110

Gibbons, D.L., A. Ahn, P.K. Chatterjee, and M. Kielian (2000) Formation and characterization of the trimeric form of the fusion protein of Semliki Forest Virus. *J Virol* **74**(17)**:** 7772-7780

Gibbons, D.L., A. Ahn, M. Liao, L. Hammar, R.H. Cheng, and M. Kielian (2004a) Multistep regulation of membrane insertion of the fusion peptide of Semliki Forest virus. *J Virol* **78**(7)**:** 3312-3318

Gibbons, D.L., I. Erk, B. Reilly, J. Navaza, M. Kielian, F.A. Rey, and J. Lepault (2003) Visualization of the target-membrane-inserted fusion protein of Semliki Forest virus by combined electron microscopy and crystallography. *Cell* **114**(5)**:** 573-583

Gibbons, D.L., M.C. Vaney, A. Roussel, A. Vigouroux, B. Reilly, J. Lepault, M. Kielian, and F.A. Rey (2004b) Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* **427**(6972)**:** 320-325

Goff, S.P. (2004) Retrovirus restriction factors. *Mol Cell* **16**(6)**:** 849-859

Gonzalez, M.E., and L. Carrasco (2003) Viroporins. *FEBS Lett* **552**(1)**:** 28-34

Green, N., T.M. Shinnick, O. Witte, A. Ponticelli, J.G. Sutcliffe, and R.A. Lerner (1981) Sequence-specific antibodies show that maturation of Moloney leukemia virus envelope polyprotein involves removal of a COOH-terminal peptide. *Proc Natl Acad Sci U S A* **78**(10)**:** 6023-6027

Haag, L. (2006) The dynamic envelope of a fusion class II virus; Molecular reorganizations during pre-fusion stages of Semliki Forest virus. Ph.D Thesis, Karolinska Institute, Stockholm, SWEDEN

Haag, L., H. Garoff, L. Xing, L. Hammar, S.T. Kan, and R.H. Cheng (2002) Acid-induced movements in the glycoprotein shell of an alphavirus turn the spikes into membrane fusion mode. *EMBO J* **21**(17)**:** 4402-4410

Haim, H., Z. Si, N. Madani, L. Wang, J.R. Courter, A. Princiotto, A. Kassa, M. DeGrace, K. McGee-Estrada, M. Mefford, D. Gabuzda, A.B. Smith, 3rd, and J. Sodroski (2009) Soluble CD4 and CD4-mimetic compounds inhibit HIV-1 infection by induction of a short-lived activated state. *PLoS Pathog* **5**(4)**:** e1000360

Hallenberger, S., V. Bosch, H. Angliker, E. Shaw, H.D. Klenk, and W. Garten (1992) Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature* **360**(6402)**:** 358-361

Hammar, L., S. Markarian, L. Haag, H. Lankinen, A. Salmi, and R.H. Cheng (2003) Prefusion rearrangements resulting in fusion Peptide exposure in Semliki forest virus. *J Biol Chem* **278**(9)**:** 7189-7198

Harrison, S.C. (2008) Viral membrane fusion. *Nat Struct Mol Biol* **15**(7)**:** 690-698

Helenius, A., J. Kartenbeck, K. Simons, and E. Fries (1980) On the entry of Semliki forest virus into BHK-21 cells. *J Cell Biol* **84**(2)**:** 404-420

Huang, C.C., M. Tang, M.Y. Zhang, S. Majeed, E. Montabana, R.L. Stanfield, D.S. Dimitrov, B. Korber, J. Sodroski, I.A. Wilson, R. Wyatt, and P.D. Kwong (2005) Structure of a V3-containing HIV-1 gp120 core. *Science* **310**(5750)**:** 1025-1028

Hunter, E. (1997) Viral entry and receptors. In *Retroviruses*, J.M. Coffin, S.H. Hughes, and H.E. Varmus (eds), pp 71-119. Cold Spring Harbour Laboratory Press

Jain, S.K., S. DeCandido, and M. Kielian (1991) Processing of the p62 envelope precursor protein of Semliki Forest virus. *J Biol Chem* **266**(9)**:** 5756-5761

Jones, T.A., J.Y. Zou, S.W. Cowan, and M. Kjeldgaard (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47 (Pt 2):** 110-119

Jose, J., J.E. Snyder, and R.J. Kuhn (2009) A structural and functional perspective of alphavirus replication and assembly. *Future Microbiol* **4:** 837-856

Justman, J., M.R. Klimjack, and M. Kielian (1993) Role of spike protein conformational changes in fusion of Semliki Forest virus. *J Virol* **67**(12)**:** 7597-7607

Kayman, S.C., H. Park, M. Saxon, and A. Pinter (1999) The hypervariable domain of the murine leukemia virus surface protein tolerates large insertions and deletions, enabling development of a retroviral particle display system. *J Virol* **73**(3)**:** 1802-1808

Kielian, M., P.K. Chatterjee, D.L. Gibbons, and Y.E. Lu (2000) Specific roles for lipids in virus fusion and exit. Examples from the alphaviruses. *Subcell Biochem* **34:** 409-455

Kielian, M., S. Jungerwirth, K.U. Sayad, and S. DeCandido (1990) Biosynthesis, maturation, and acid activation of the Semliki Forest virus fusion protein. *J Virol* **64**(10)**:** 4614-4624

Klimjack, M.R., S. Jeffrey, and M. Kielian (1994) Membrane and protein interactions of a soluble form of the Semliki Forest virus fusion protein. *J Virol* **68**(11)**:** 6940-6946

Kobe, B., R.J. Center, B.E. Kemp, and P. Poumbourios (1999) Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding

protein chimera reveals structural evolution of retroviral transmembrane proteins.PG - 4319-24. *Proc Natl Acad Sci U S A* **96**(8)**:** 4319-4324.

Kozlovsky, Y., and M.M. Kozlov (2002) Stalk model of membrane fusion: solution of energy crisis. *Biophys J* **82**(2)**:** 882-895

Kuhn, R.J., W. Zhang, M.G. Rossmann, S.V. Pletnev, J. Corver, E. Lenches, C.T. Jones, S. Mukhopadhyay, P.R. Chipman, E.G. Strauss, T.S. Baker, and J.H. Strauss (2002) Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* **108**(5)**:** 717-725

Kwong, P.D., M.L. Doyle, D.J. Casper, C. Cicala, S.A. Leavitt, S. Majeed, T.D. Steenbeke, M. Venturi, I. Chaiken, M. Fung, H. Katinger, P.W. Parren, J. Robinson, D. Van Ryk, L. Wang, D.R. Burton, E. Freire, R. Wyatt, J. Sodroski, W.A. Hendrickson, and J. Arthos (2002) HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420**(6916)**:** 678-682

Kwong, P.D., R. Wyatt, J. Robinson, R.W. Sweet, J. Sodroski, and W.A. Hendrickson (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**(6686)**:** 648-659

Lavanya, M., S. Kinet, A. Montel-Hagen, C. Mongellaz, J.L. Battini, M. Sitbon, and N. Taylor (2008) Cell surface expression of the bovine leukemia virus-binding receptor on B and T lymphocytes is induced by receptor engagement. *J Immunol* **181**(2)**:** 891-898

Lavillette, D., B. Boson, S.J. Russell, and F.L. Cosset (2001) Activation of membrane fusion by murine leukemia viruses is controlled in cis or in trans by interactions between the receptor-binding domain and a conserved disulfide loop of the carboxy terminus of the surface glycoprotein.PG - 3685-95. *J Virol* **75**(8)**:** 3685-3695.

Lavillette, D., M. Maurice, C. Roche, S.J. Russell, M. Sitbon, and F.L. Cosset (1998) A proline-rich motif downstream of the receptor binding domain modulates conformation and fusogenicity of murine retroviral envelopes. *J Virol* **72**(12)**:** 9955-9965

Lavillette, D., A. Ruggieri, S.J. Russell, and F.L. Cosset (2000) Activation of a cell entry pathway common to type C mammalian retroviruses by soluble envelope fragments.PG - 295-304. *J Virol* **74**(1)**:** 295-304.

Lee, J.E., M.L. Fusco, A.J. Hessell, W.B. Oswald, D.R. Burton, and E.O. Saphire (2008) Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* **454**(7201)**:** 177-182

Lescar, J., A. Roussel, M.W. Wien, J. Navaza, S.D. Fuller, G. Wengler, and F.A. Rey (2001) The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* **105**(1)**:** 137-148

Li, L., S.M. Lok, I.M. Yu, Y. Zhang, R.J. Kuhn, J. Chen, and M.G. Rossmann (2008) The flavivirus precursor membrane-envelope protein complex: structure and maturation. *Science* **319**(5871)**:** 1830-1834

Liljestrom, P., and H. Garoff (1991) Internally located cleavable signal sequences direct the formation of Semliki Forest virus membrane proteins from a polyprotein precursor. *J Virol* **65**(1)**:** 147-154

Liu, J., A. Bartesaghi, M.J. Borgnia, G. Sapiro, and S. Subramaniam (2008) Molecular architecture of native HIV-1 gp120 trimers. *Nature* **455**(7209)**:** 109-113

Lobigs, M., and H. Garoff (1990) Fusion function of the Semliki Forest virus spike is activated by proteolytic cleavage of the envelope glycoprotein precursor p62. *J Virol* **64**(3)**:** 1233-1240

Lobigs, M., J.M. Wahlberg, and H. Garoff (1990a) Spike protein oligomerization control of Semliki Forest virus fusion. *J Virol* **64**(10)**:** 5214-5218

Lobigs, M., H.X. Zhao, and H. Garoff (1990b) Function of Semliki Forest virus E3 peptide in virus assembly: replacement of E3 with an artificial signal peptide abolishes spike heterodimerization and surface expression of E1. *J Virol* **64**(9)**:** 4346-4355

Lu, Y.E., and M. Kielian (2000) Semliki forest virus budding: assay, mechanisms, and cholesterol requirement. *J Virol* **74**(17)**:** 7708-7719

Ludtke, S.J., P.R. Baldwin, and W. Chiu (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* **128**(1)**:** 82-97

Maddon, P.J., A.G. Dalgleish, J.S. McDougal, P.R. Clapham, R.A. Weiss, and R. Axel (1986) The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* **47**(3)**:** 333-348

Mancini, E.J., M. Clarke, B.E. Gowen, T. Rutten, and S.D. Fuller (2000) Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. *Mol Cell* **5**(2)**:** 255-266

Marsh, M., E. Bolzau, and A. Helenius (1983) Penetration of Semliki Forest virus from acidic prelysosomal vacuoles. *Cell* **32**(3)**:** 931-940

Martin, C.S., C.Y. Liu, and M. Kielian (2009) Dealing with low pH: entry and exit of alphaviruses and flaviviruses. *Trends Microbiol*

McKeating, J.A., A. McKnight, and J.P. Moore (1991) Differential loss of envelope glycoprotein gp120 from virions of human immunodeficiency virus type 1 isolates: effects on infectivity and neutralization. *J Virol* **65**(2)**:** 852-860

Melton, J.V., G.D. Ewart, R.C. Weir, P.G. Board, E. Lee, and P.W. Gage (2002) Alphavirus 6K proteins form ion channels. *J Biol Chem* **277**(49)**:** 46923-46931

Mendoza, Q.P., J. Stanley, and D.E. Griffin (1988) Monoclonal antibodies to the E1 and E2 glycoproteins of Sindbis virus: definition of epitopes and efficiency of protection from fatal encephalitis. *J Gen Virol* **69 (Pt 12):** 3015-3022

Mkrtchyan, S.R., R.M. Markosyan, M.T. Eadon, J.P. Moore, G.B. Melikyan, and F.S. Cohen (2005) Ternary complex formation of human immunodeficiency virus type 1 Env, CD4, and chemokine receptor captured as an intermediate of membrane fusion. *J Virol* **79**(17)**:** 11161-11169

Modis, Y., S. Ogata, D. Clements, and S.C. Harrison (2004) Structure of the dengue virus envelope protein after membrane fusion. *Nature* **427**(6972)**:** 313-319

Moore, P.L., E.T. Crooks, L. Porter, P. Zhu, C.S. Cayanan, H. Grise, P. Corcoran, M.B. Zwick, M. Franti, L. Morris, K.H. Roux, D.R. Burton, and J.M. Binley (2006) Nature of nonfunctional envelope proteins on the surface of human immunodeficiency virus type 1. *J Virol* **80**(5)**:** 2515-2528

Morita, E., and W.I. Sundquist (2004) Retrovirus budding. *Annu Rev Cell Dev Biol* **20:** 395-425

Mothes, W., A.L. Boerger, S. Narayan, J.M. Cunningham, and J.A. Young (2000) Retroviral entry mediated by receptor priming and low pH triggering of an envelope glycoprotein. *Cell* **103**(4)**:** 679-689.

Mukhopadhyay, S., W. Zhang, S. Gabler, P.R. Chipman, E.G. Strauss, J.H. Strauss, T.S. Baker, R.J. Kuhn, and M.G. Rossmann (2006) Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. *Structure* **14**(1)**:** 63-73

Netter, R.C., S.M. Amberg, J.W. Balliet, M.J. Biscone, A. Vermeulen, L.J. Earp, J.M. White, and P. Bates (2004) Heptad repeat 2-based peptides inhibit avian sarcoma and leukosis virus subgroup a infection and identify a fusion intermediate. *J Virol* **78**(24)**:** 13430-13439

Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**(13)**:** 1605-1612

Pletnev, S.V., W. Zhang, S. Mukhopadhyay, B.R. Fisher, R. Hernandez, D.T. Brown, T.S. Baker, M.G. Rossmann, and R.J. Kuhn (2001) Locations of carbohydrate sites on alphavirus glycoproteins show that E1 forms an icosahedral scaffold. *Cell* **105**(1)**:** 127-136

Rein, A., J. Mirro, J.G. Haynes, S.M. Ernst, and K. Nagashima (1994) Function of the cytoplasmic domain of a retroviral transmembrane protein: p15E-p2E cleavage activates the membrane fusion capability of the murine leukemia virus Env protein. *J Virol* **68**(3)**:** 1773-1781

Rey, F.A., F.X. Heinz, C. Mandl, C. Kunz, and S.C. Harrison (1995) The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. *Nature* **375**(6529)**:** 291-298

Roche, S., S. Bressanelli, F.A. Rey, and Y. Gaudin (2006) Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science* **313**(5784)**:** 187-191

Roche, S., F.A. Rey, Y. Gaudin, and S. Bressanelli (2007) Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. *Science* **315**(5813)**:** 843-848

Ross, S.R., J.J. Schofield, C.J. Farr, and M. Bucan (2002) Mouse transferrin receptor 1 is the cell entry receptor for mouse mammary tumor virus. *Proc Natl Acad Sci U S A* **99**(19)**:** 12386-12390

Rothenberg, S.M., M.N. Olsen, L.C. Laurent, R.A. Crowley, and P.O. Brown (2001) Comprehensive mutational analysis of the Moloney murine leukemia virus envelope protein. *J Virol* **75**(23)**:** 11851-11862

Roussel, A., J. Lescar, M.C. Vaney, G. Wengler, and F.A. Rey (2006) Structure and interactions at the viral surface of the envelope protein E1 of Semliki Forest virus. *Structure* **14**(1)**:** 75-86

Salminen, A., J.M. Wahlberg, M. Lobigs, P. Liljestrom, and H. Garoff (1992) Membrane fusion process of Semliki Forest virus. II: Cleavage-dependent reorganization of the spike protein complex controls virus entry. *J Cell Biol* **116**(2)**:** 349-357

Samsonov, A.V., P.K. Chatterjee, V.I. Razinkov, C.H. Eng, M. Kielian, and F.S. Cohen (2002) Effects of membrane potential and sphingolipid structures on fusion of Semliki Forest virus. *J Virol* **76**(24)**:** 12691-12702

Saxton, W.O., and W. Baumeister (1982) The correlation averaging of a regularly arranged bacterial cell envelope protein. *J Microsc* **127**(Pt 2)**:** 127-138

Simmons, G., D.N. Gosalia, A.J. Rennekamp, J.D. Reeves, S.L. Diamond, and P. Bates (2005) Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci U S A* **102**(33)**:** 11876-11881

Sjoberg, M., M. Wallin, B. Lindqvist, and H. Garoff (2006) Furin cleavage potentiates the membrane fusion-controlling intersubunit disulfide bond isomerization activity of leukemia virus Env. *J Virol* **80**(11)**:** 5540-5551

Skehel, J.J., and D.C. Wiley (1998) Coiled coils in both intracellular vesicle and viral membrane fusion. *Cell* **95**(7)**:** 871-874

Skehel, J.J., and D.C. Wiley (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* **69:** 531-569

Skoging, U., and P. Liljestrom (1998) Role of the C-terminal tryptophan residue for the structure-function of the alphavirus capsid protein. *J Mol Biol* **279**(4)**:** 865-872

Skoging, U., M. Vihinen, L. Nilsson, and P. Liljestrom (1996) Aromatic interactions define the binding of the alphavirus spike to its nucleocapsid. *Structure* **4**(5)**:** 519-529

Smith, J.G., W. Mothes, S.C. Blacklow, and J.M. Cunningham (2004) The mature avian leukosis virus subgroup A envelope glycoprotein is metastable, and refolding induced by the synergistic effects of receptor binding and low pH is coupled to infection. *J Virol* **78**(3)**:** 1403-1410

Smith, T.J., R.H. Cheng, N.H. Olson, P. Peterson, E. Chase, R.J. Kuhn, and T.S. Baker (1995) Putative receptor binding sites on alphaviruses as visualized by cryoelectron microscopy. *Proc Natl Acad Sci U S A* **92**(23)**:** 10648-10652

Strauss, J.H., and E.G. Strauss (1994) The alphaviruses: gene expression, replication, and evolution. *Microbiol Rev* **58**(3)**:** 491-562

Thali, M., J.P. Moore, C. Furman, M. Charles, D.D. Ho, J. Robinson, and J. Sodroski (1993) Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J Virol* **67**(7)**:** 3978-3988

Tubulekas, I., and P. Liljestrom (1998) Suppressors of cleavage-site mutations in the p62 envelope protein of Semliki Forest virus reveal dynamics in spike structure and function. *J Virol* **72**(4)**:** 2825-2831

Umashankar, M., C. Sanchez-San Martin, M. Liao, B. Reilly, A. Guo, G. Taylor, and M. Kielian (2008) Differential cholesterol binding by class II fusion proteins determines membrane fusion properties. *J Virol* **82**(18)**:** 9245-9253

Vrati, S., C.A. Fernon, L. Dalgarno, and R.C. Weir (1988) Location of a major antigenic site involved in Ross River virus neutralization. *Virology* **162**(2)**:** 346-353

Waarts, B.L., R. Bittman, and J. Wilschut (2002) Sphingolipid and cholesterol dependence of alphavirus membrane fusion. Lack of correlation with lipid raft formation in target liposomes. *J Biol Chem* **277**(41)**:** 38141-38147

Wahlberg, J.M., W.A. Boere, and H. Garoff (1989) The heterodimeric association between the membrane proteins of Semliki Forest virus changes its sensitivity to low pH during virus maturation. *J Virol* **63**(12)**:** 4991-4997

Wahlberg, J.M., R. Bron, J. Wilschut, and H. Garoff (1992) Membrane fusion of Semliki Forest virus involves homotrimers of the fusion protein. *J Virol* **66**(12)**:** 7309-7318

Wahlberg, J.M., and H. Garoff (1992) Membrane fusion process of Semliki Forest virus. I: Low pH-induced rearrangement in spike protein quaternary structure precedes virus penetration into cells. *J Cell Biol* **116**(2)**:** 339-348

Wallin, M., Ekström, M. and H. Garoff (2006) Receptor triggered but alkylation-arrested Env of murine leukaemia virus reveals the transmembrane subunit in a prehairpin conformation. *J Virol* **80:** 9921-9925

Wallin, M., M. Ekstrom, and H. Garoff (2004) Isomerization of the intersubunit disulphide-bond in Env controls retrovirus fusion. *Embo J* **23**(1)**:** 54-65 Epub 2003 Dec 2011

Wallin, M., M. Ekstrom, and H. Garoff (2005) The fusion-controlling disulfide bond isomerase in retrovirus env is triggered by protein destabilization. *J Virol* **79**(3)**:** 1678-1685

Wang, H., M.P. Kavanaugh, R.A. North, and D. Kabat (1991) Cell-surface receptor for ecotropic murine retroviruses is a basic amino-acid transporter. *Nature* **352**(6337)**:** 729-731

Weissenhorn, W., A. Dessen, S.C. Harrison, J.J. Skehel, and D.C. Wiley (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**(6631)**:** 426-430.

Weissenhorn, W., A. Hinz, and Y. Gaudin (2007) Virus membrane fusion. *FEBS Lett* **581**(11)**:** 2150-2155

White, J., and A. Helenius (1980) pH-dependent fusion between the Semliki Forest virus membrane and liposomes. *Proc Natl Acad Sci U S A* **77**(6)**:** 3273-3277

White, J., J. Kartenbeck, and A. Helenius (1980) Fusion of Semliki forest virus with the plasma membrane can be induced by low pH. *J Cell Biol* **87**(1)**:** 264-272

White, J.M., S.E. Delos, M. Brecher, and K. Schornberg (2008) Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol* **43**(3)**:** 189-219

Wilson, I.A., J.J. Skehel, and D.C. Wiley (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution.PG - 366-73. *Nature* **289**(5796)**:** 366-373.

Wu, S.R., L. Haag, L. Hammar, B. Wu, H. Garoff, L. Xing, K. Murata, and R.H. Cheng (2007) The dynamic envelope of a fusion class II virus. Prefusion stages of semliki forest virus revealed by electron cryomicroscopy. *J Biol Chem* **282**(9)**:** 6752-6762

Wyatt, R., P.D. Kwong, E. Desjardins, R.W. Sweet, J. Robinson, W.A. Hendrickson, and J.G. Sodroski (1998) The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**(6686)**:** 705-711

Yin, H.S., X. Wen, R.G. Paterson, R.A. Lamb, and T.S. Jardetzky (2006) Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. *Nature* **439**(7072)**:** 38-44

Yu, I.M., W. Zhang, H.A. Holdaway, L. Li, V.A. Kostyuchenko, P.R. Chipman, R.J. Kuhn, M.G. Rossmann, and J. Chen (2008) Structure of the immature dengue virus at low pH primes proteolytic maturation. *Science* **319**(5871)**:** 1834-1837

Zanetti, G., J.A. Briggs, K. Grunewald, Q.J. Sattentau, and S.D. Fuller (2006) Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. *PLoS Pathog* **2**(8)**:** e83

Zeng, J., P. Fournier, and V. Schirrmacher (2002) Induction of interferon-alpha and tumor necrosis factor-related apoptosis-inducing ligand in human blood mononuclear cells by hemagglutinin-neuraminidase but not F protein of Newcastle disease virus. *Virology* **297**(1)**:** 19-30

Zhang, W., S. Mukhopadhyay, S.V. Pletnev, T.S. Baker, R.J. Kuhn, and M.G. Rossmann (2002) Placement of the structural proteins in Sindbis virus. *J Virol* **76**(22)**:** 11645-11658

Zhao, F.Q., and R. Craig (2003) Capturing time-resolved changes in molecular structure by negative staining. *J Struct Biol* **141**(1)**:** 43-52

Zhou, T., L. Xu, B. Dey, A.J. Hessell, D. Van Ryk, S.H. Xiang, X. Yang, M.Y. Zhang, M.B. Zwick, J. Arthos, D.R. Burton, D.S. Dimitrov, J. Sodroski, R. Wyatt, G.J. Nabel, and P.D. Kwong (2007) Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445**(7129)**:** 732-737

Zhu, P., E. Chertova, J. Bess, Jr., J.D. Lifson, L.O. Arthur, J. Liu, K.A. Taylor, and K.H. Roux (2003) Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. *Proc Natl Acad Sci U S A* **100**(26)**:** 15812-15817 Epub 12003 Dec 15810

Zhu, P., J. Liu, J. Bess, Jr., E. Chertova, J.D. Lifson, H. Grise, G.A. Ofek, K.A. Taylor, and K.H. Roux (2006) Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441**(7095)**:** 847-852

Zhu, P., H. Winkler, E. Chertova, K.A. Taylor, and K.H. Roux (2008) Cryoelectron tomography of HIV-1 envelope spikes: further evidence for tripod-like legs. *PLoS Pathog* **4**(11)**:** e1000203

Ziemiecki, A., H. Garoff, and K. Simons (1980) Formation of the Semliki Forest virus membrane glycoprotein complexes in the infected cell. *J Gen Virol* **50**(1)**:** 111-123