Viral envelope glycoproteins swing into action

Analysis of tick-borne encephalitis virus E protein reveals considerable structural diversity in the glycoproteins that clothe enveloped viruses and hints at the conformational gyrations in this molecule that lead to viral fusion.

Structure 15 July 1995, **3**:645–648

Although the first envelope glycoprotein structure [1] — the haemagglutinin (HA) from influenza virus — predated that of an intact animal virion [2,3], intact virus work then stole the structural show for some years, yielding many, albeit generally static, biological insights. Now, studies on individual viral envelope proteins are again occupying centre stage. Less than a year ago, the Wiley group at Harvard published a low-pH conformation for a fragment of influenza virus HA, a first, compelling, insight into how massive conformational rearrangements of a metastable protein structure might drive key early events in the virus life cycle [4]. Now, also from Harvard, but this time from the Harrison stable, comes the first instalment of the corresponding story for the major envelope glycoprotein (E) of a flavivirus, tick-borne encephalitis (TBE) virus [5]. This structure is very different from that of the influenza virus HA, although the biological functions of the two proteins are similar. There is already an indication that the fuel for one of these functions, membrane fusion, may be the same. The molecules appear to act as one-stroke motors: proteolysis of a surface protein upon leaving the host cell primes the system to function later, as the pH drops during entry to another cell. A twist in the story is that in TBE virus this device seems to be in two parts, a relatively small transmembrane protein, prM, (which is cleaved to form M in the mature virion) and the larger E protein, which seems to be responsible for both receptor recognition and insertion into the cellular membrane.

Roles for surface glycoproteins

Enveloped viruses comprise half of the virus families that infect animals [6]. Some of these, such as influenza virus, are pleiotropic but even for these apparently variable viruses certain underlying principles dominate their structure and regulate their lifestyle. Of the various proteins encoded in the viral genome, the envelope glycoproteins (the E and M proteins in TBE virus) are major players in both of these aspects. During virus assembly they act both to exclude host proteins from the portion of the cell membrane hijacked by the virion, and to facilitate the budding of the virus from the infected cell. They are also central to the initiation of infection — they recognize the cellular receptor and hold the virus in place as it enters the cell. Finally, they drive perhaps the most dramatic phase of the viral life cycle, the conformational changes that lead to fusion of the host and viral membranes, the essential precursor to transfer of the viral genome into the host cell. In addition to having these votive roles, the envelope glycoproteins are the face of the virus presented to the host and therefore the focus of attack by the host immune system (for a review of these roles see [7]).

The 'flu paradigm

Viral envelope glycoproteins have been, until now, exemplified structurally by those from influenza virus types A and B. The two surface glycoproteins in these viruses, HA and neuraminidase (NA), are multimers, with a head group atop a stalk (the stalk being only inferred for NA) [1,8]. The head groups are architecturally different β-structures, each containing a binding site for the virus receptor, sialic acid. Attachment of the virus to a cell occurs by the binding of HA to sialic acid, which NA probably destroys during viral release [9]. Studies of the influenza virus, which has a mixture of surface glycoproteins and no real sign (at least in types A and B virus) of surface regularity [10], have given little insight into the role of the envelope proteins in viral assembly. They did, however, give a clear picture, consistent with electron micrographs, of the structure of the viral surface, in which the proteins stand up, bristling from the virus surface. A major result last year was the visualization of a conformational rearrangement in influenze virus HA, which occurs at low pH, to form a coiled-coil structure that had been predicted some years earlier [4,11,12]. This revealed the switch from a metastable conformation, in which a hydrophobic peptide (the fusion peptide) is tucked into the protein, well away from the cell membrane, to a more stable structure, with the fusion peptide moved 100 Å upwards, reaching towards the cellular membrane to which the head of the molecule previously attached. The fusion peptide then presumably embeds itself in the cell membrane, as a preliminary to fusion of the viral and cellular membranes.

TBE virus breaks the mould

TBE virus, at some 550 Å across, is one of the smallest enveloped viruses and is thought to possess icosahedral symmetry (S Fuller, personal communication), although no well-ordered crystals of the intact virus have been reported. Rey *et al.* [5] used a method that is now traditional for determining the structure of isolated envelope proteins. A soluble form of E was prepared by cleavage (using trypsin) from the viral membrane. This removed a little over 100 residues from the C terminus of the intact protein, including two C-terminal transmembrane

regions and almost 50 residues of the domain external to the virus membrane (the ectodomain) which leads into the membrane. The resulting protein fragment (395 amino acids in length) forms an elongated structure and individual molecules are formed by association of two of these fragments into a dimer. Although the stoichiometry of these proteins in the icosahedral surface lattice has not yet been determined, there is evidence [13] that, presumably similar, dimeric associations occur in the mature virion. If we assume that the observed molecular twofold axis is perpendicular to the viral envelope then it is immediately clear that the surface of the TBE virus formed by the E protein is radically different to that of influenza virus, formed by HA and NA. Whereas HA and NA sit up on stalks, the E protein extends laterally in the plane of the membrane, a little like two logs bound together and floating on the membrane, the bundle being 170 Å long, 60 Å wide but only 30 Å deep (Fig. 1) [5]. The question of which face of the molecule lies upwards on the viral surface is easily answered. The E protein molecule has a slight curvature, as if it were made to fit onto a spherical surface, and the carbohydrate and most of the antigenic sites map to the convex face of the molecule, whereas the C terminus projects out of the concave surface, towards the viral membrane (Fig. 1) [5].

Molecular architecture of TBE virus

Each of the E protein monomers is made up of three domains arranged along its length. The N-terminal portion of the polypeptide chain weaves between two domains, the central domain and a dimerization domain (Fig. 1). The central domain is all-β, forming a sandwich, and looks at first glance as if it might be a jelly-roll fold, a structure which is very common in viral structural proteins [14]. In fact it is different, belonging to the up-down family of β-sandwiches, the sandwich axis being roughly in the plane of the membrane (Fig. 1). Two large insertions in loops of the central domain form the second domain which is elongated and is principally responsible for dimerization. This dimerization domain is predominantly β in structure with two additional α-helices and, once again, shows only a superficial similarity to other viral structural protein folds. Intriguingly, the tip of this domain is largely formed from a three-stranded β-sheet with the topology of a kringle domain [15]. Although this structure has exactly the same pattern of disulphide bonds as a kringle domain, its three-dimensional structure is much more extended (this similarity has been anticipated; J Aaskov, personal communication). These first two domains lie end-to-end along the membrane and, to us, appear likely to form a rigid unit.

The third domain is a little separate from the others and has the topology of the immunoglobulin (Ig) superfamily, being a constant-type (C-type) fold [16]. This domain is oriented differently to the other two, with respect to the membrane: it essentially sits upright, so that the C terminus of the chain projects down towards the membrane (Fig. 1), reminiscent of the orientation assumed by Ig-like domains in cell-surface receptors [17]. This leads to the domain projecting slightly above and below the others. An extended, somewhat flexible, linker region allows a slight rocking of this domain with respect to the rest of the molecule. This is the first visualization of an Ig-like domain in a viral structural protein. Perhaps we should not be surprised to see it: Ig-like domains are ubiquitous in cellular proteins and viruses are tremendous molecular scavengers, appropriating molecules for their own ends (perhaps the clearest example being the trypsin-like serine protease fold seen in the core protein of Sindbis virus [18]). We are, after all, dealing with a multidomain membrane glycoprotein and some 40% of the proteins on the surface of the best characterized mammalian cell, the leukocyte, contain at least one such Ig-like domain [19]. But what about the biology?

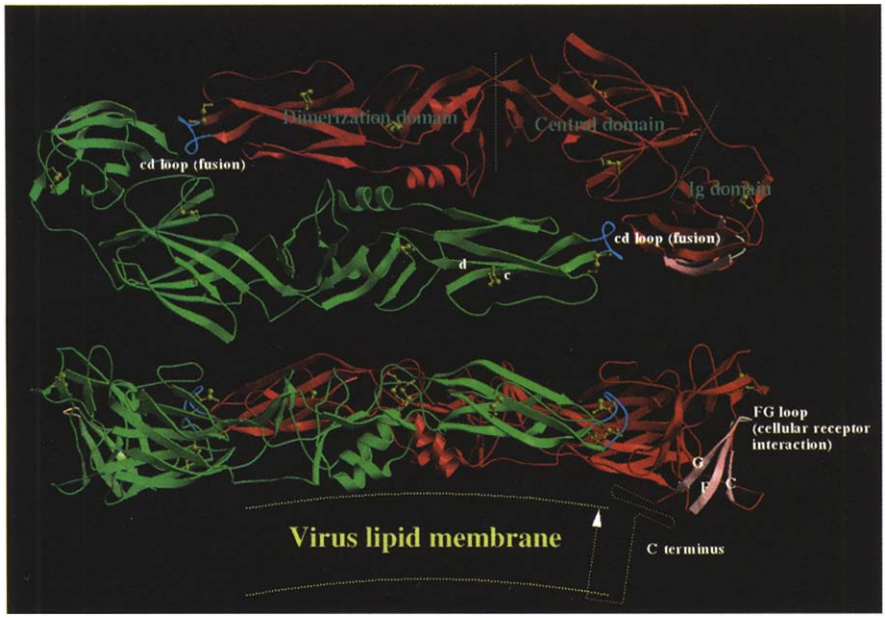

Fig. 1. Folding of the polypeptide chain of the envelope glycoprotein (E) of tick-born encephalitis (TBE) virus in the dimer. Two orthogonal views are shown. The division of E into domains is shown by labelling one subunit in the uppermost view. The central domain contains the N terminus. The dimerization domain contains the cd loop (blue), a hydrophobic loop between strands c and d that is supposed to be involved in the fusion process. The immunoglobulin (Ig)-like domain contains three strands at its surface, CFG, (highlighted in pink and seen most clearly in the lower view), thought to be involved in the binding to the cellular membrane during the virus infection. The FG loop (pale pink) accommodates an insertion that includes an RGD tripeptide in some flaviviruses. Drawn with MOLSCRIPT [26] as modified by R Esnouf and rendered with Raster3D [27]. Figure inspired by Figure 1 of [5]; coordinates supplied by F Rey and S Harrison.

Receptor interactions involve an Ig-like domain

The flaviviruses replicate in both mammalian hosts and insect vectors. Properties such as tissue tropism and host range are intimately connected with the properties of the E protein. As there is considerable conservation of the amino acid sequence of the E protein across the flaviviruses, biological information from other viruses can be directly mapped to the TBE virus structure. In this way, Rey *et al.* [5] have built up a convincing picture of the central role of the Ig-like domain in receptor interactions. There is a clustering of single mutations with altered virulence characteristics on the CFG face (Fig. 1) of the Ig-like domain, which points strongly to this portion of the molecular surface acting as the site of attachment to cellular receptors. Although the receptors for the flaviviruses have not been identified, there is an intriguing correlation between the structure of the FG loop (Fig. 1) and the particular sub-order of the Diptera (the true flies) that are used as vectors by these viruses. Thus, in tick-borne flaviviruses the FG loop forms a tight turn (Fig. 1) but this is enlarged in the mosquito-borne viruses by a four-residue insert, usually containing the tripeptide RGD, a characteristic motif of ligands for certain members of the integrin family of cell-surface receptors [20]. Other viruses are known to use integrins as receptors [21], and we now have structural information for Ig-like domains known to be used for integrin recognition at the cell membrane [22]. TBE, however, is the first example of a virus with an Ig-like domain appearing to conflate integrin recognition. The structure of a two-Ig-like domain integrin-binding fragment of vascular cell-adhesion molecule-1 (VCAM-1), taken together with mutational data for this protein, point to the use of the same face of the β-sandwich for the adhesion interaction as is employed in the Ig-like domain of the TBE virus E protein for receptor recognition [22]. The involvement of corresponding residues in cell-adhesive interactions in another member of the Ig superfamily, the cell-adhesion molecule, CD2 (reviewed in [23]), suggests a clear functional preference for the use of the CFG face of these structural scaffolds in molecular recognition.

The low-pH conformational transition

In influenza virus, the HA molecule alone is responsible for the conformational changes, triggered by low-pH, that are the prelude to membrane fusion [24]. The key event priming that molecule for the low-pH change is the proteolytic cleavage of HA into the two-chain form found in the mature virion. The situation is different in TBE virus, as the E protein is not cleaved but still undergoes a structural rearrangement at low pH — it is thought to switch from a dimer to a trimer. There must also be a maturation process as, during exocytosis, immature particles are exposed to low pH without this triggering a structural change. It is thought that prM (the precursor of the small viral transmembrane protein, M) interacts with E and modulates its properties. Thus, if we think of prM as part of E (cross-linking experiments suggest that they form heterodimers [25] and mutational evidence suggests interaction of prM and the dimerization domain of E), then we have a situation analogous to that in influenza virus. The initial prM–E complex is stable under acidic conditions. Then, during the virus maturation, prM is cleaved to M. This releases dimeric E, now primed for the low-pH conformational switch. Unfortunately, the fragment of E seen crystallographically cannot tell us the full story, as biochemical evidence shows that this particular fragment does not rearrange at low pH (FX Heinz, unpublished data), perhaps indicating that the C-terminal portion of the ectodomain, excised during preparation of a soluble form of E, is required to initiate the transition.

In spite of this, the X-ray structure, taken with other evidence, suggests a tentative model for at least part of the process. There is a very conserved hydrophobic peptide (the cd loop) at the end of the protein that is furthest from the membrane-anchoring C terminus (Fig. 1). Changes in this peptide are correlated with viral virulence, making it a likely candidate for a fusion peptide. If the dimerization domain were freed from its molecular partner then, because its long β-strands are pinned together by a ladder of three disulphide bonds, we might imagine it retaining its overall structure and swinging upwards. The pivot points could be either the junction between the dimerization domain and the central domain, or the extended linker to the Ig-like domain (where the crystal structure already provides some evidence of flexibility), or both. This movement would allow the putative fusion peptide to be projected some 100 Å upwards towards the cellular membrane, a result similar to that achieved by the low-pH conformational change in influenza HA. Mutational data support this theory [5]. Other, more complex arrangements may also be possible. An alternative scheme has been proposed by Allison *et al.* [13].

We find it tempting now to put these pieces together to produce a fuller, although speculative, model for the processes leading up to viral fusion. The mechanisms we propose are hypothetical, indeed an alternative scheme has been put forward [13] and our numerology is probably wrong; there are likely to be more molecules of E protein on the virus surface. The molecular dimensions of TBE virus are consistent with an icosahedral lattice of homodimeric E molecules, covering the 500 Å diameter viral membrane. It is likely that a portion of the molecule contained in the 100 amino acid fragment which is removed by trypsin cleavage, is partially disordered at neutral pH, but it is reasonable to imagine that this region of the molecule might initially be organized, by contacts between its membrane-spanning or cytoplasmic C-terminal residues at the threefold axes providing trimeric contacts close to the viral membrane. A low-pH conformational transition, perhaps further ordering this part of the molecule by increasing association about the isocahedral threefold axis, would place torque on the dimer contacts further along the molecule. In the absence of the stabilizing influence of prM these would be torn apart allowing the fusion peptide to swing upwards and also resulting in the formation of a trimeric E protein structure. This hypothetical model is shown in Figure 2.

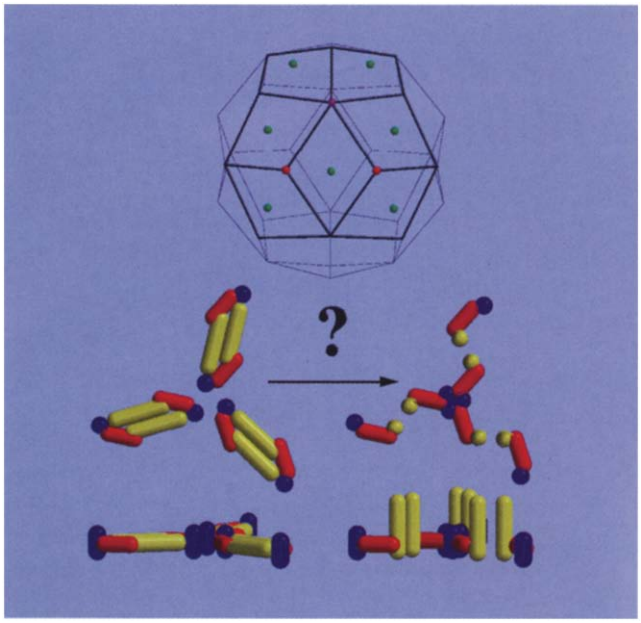

Fig. 2. Hypothesis for the conformational change in protein E during the fusion process in TBE virus. Monomers are represented schematically in the lower part of the figure with their central domains in red, their dimerization domains in yellow and their Ig-like domains in blue. Our proposal involves the disruption of the dimers at low pH, perhaps by increased association around the icosahedral threefold axes, followed by an elevation of the dimerization domain, which carries with it (on its extremity) the fusion peptide. The molecular dimensions of the E protein fit well with an organization of the trimers around the true icosahedral threefold axes (shown in red, top), the dimers being centred, in this model, on the points marked in green. (Drawn with MOLSCRIPT [26] as modified by R Esnouf and Raster3D [27].)

Viral fusion — a concept becomes concrete

With TBE virus E protein, structural biology once again has provided some rich biological pickings. Without the structure of Rey *et al.* [5] the basic principles discussed above could not have been guessed at. As structure determination becomes more routine, the trick for the structural biologist is to know what biological questions can be answered and then to sacrifice a little of their lives to them. Often the general significance of the results is teased out, not so much from the basic protein fold, as from the insights into mechanisms used to achieve function. Viral envelope glycoproteins provide an extreme example of this: the mechanism of conformational change shows a convergence in concept between influenza and TBE viruses — an example of molecular reification.

Acknowledgements: PG is supported by the European Human Capital and Mobility Programme (ERB4001GT941344).

References

1. Wilson, I.A., Skehel, J.J. & Wiley, D.C. (1981). Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* **289**, 366–373.
2. Rossmann, M.G., *et al.*, & Vriend, G. (1985). Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* **317**, 145–153.
3. Hogle, J.M., Chow, M. & Filman, D.J. (1985). Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* **229**, 1358–1365.
4. Bullough, P.A., Hughson, F.M., Skehel, J.J. & Wiley, D.C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371**, 37–43.
5. Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C. & Harrison, S.C. (1995). The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* **375**, 291–298.
6. Murphy, F.A. & Kingsbury, D.W. (1990). Virus taxonomy. In *Virology.* (2nd edn). (Fields, D.N. & Knipe, D.M., eds), pp. 9–35, Raven Press Ltd, New York.
7. Wiley, D.C. & Skehel, J.J. (1990). Viral membranes. In *Virology.* (2nd edn). (Fields, D.N. & Knipe, D.M., eds), pp. 63–85, Raven Press Ltd, New York.
8. Colman, P.M., Varghese, J.N. & Laver, W.G. (1983). Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature* **303**, 41–44.
9. Kingsbury, D.W. (1990). Orthomyxoviridae and their replication. In *Virology.* (2nd edn). (Fields, D.N. & Knipe, D.M., eds), pp. 1075–1089, Raven Press Ltd, New York.
10. Waterson, A.P., Huvell, J.M.W. & Jensen, K.E. (1963). The fine structure of influenza A, B and C viruses. *Arch. Ges. Virusforsch.* **12**, 487–495.
11. Ward, C.W & Dopheide, T.A. (1980). Influenza virus haemagglutinin. Structural predictions suggest that the fibrillar appearance is due to the presence of a coiled-coil. *Aust. J. Biol. Sci.* **33**, 441–447.
12. Carr, C.M. & Kim P.S. (1993). A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell*, **73**, 823–832.
13. Allison, S.L., Schalich, J., Stiasny K., Mandl, C.W., Kunz, C. & Heinz, F.X. (1995). Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. *J. Virol.* **69**, 695–700.
14. Stuart, D. (1993). Viruses. *Curr. Opin. Struct. Biol.* **3**, 167–174.
15. Magnusson, S., Petersen, T.E., Sottrup-Jensen, L. & Claeys, H. (1975). Complete primary structure of prothrombin: isolation, structure and reactivity of ten carboxylated glutamic acid residues and regulation of prothrombin activation by thrombin. In *Proteases and Biological Control.* (Reich, E., Rifkin, D.B. & Shaw, E., eds), pp. 123–149, Cold Spring Harbor conferences on cell proliferation volume 2, Cold Spring Harbor Laboratory.
16. Bork, P., Holm, L. & Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. *J. Mol. Biol.* **242**, 309–320.
17. Jones, E.Y. (1993). The immunoglobulin superfamily. *Curr. Opin. Struct. Biol.* **3**, 846–852.
18. Choi, H.K., *et al.*, & Wengler, G. (1991). Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature* **354**, 37–43.
19. Barclay, A.N., *et al.*, & Williams, A.F. (1993). *The Leucocyte Antigen.* Academic Press, London.
20. Ruoslahti, E. (1988). Fibronectin and its receptors. *Annu. Rev. Biochem.* **57**, 375–413.
21. Eisner, R. (1992). Finding out how a viral hitchhiker snags a ride. *Science* **255**, 1718–1720.
22. Jones, E.Y., *et al.*, & Stuart, D.I. (1995). Crystal structure of an integrin-binding fragment of vascular cell adhesion molecule-1 at 1.8 Å resolution. *Nature* **373**, 539–544.
23. Bodian, D.L., Jones, E.Y., Harlos, K., Stuart, D.I. & Davis, S.J. (1994). Crystal structure of the extracellular region of the human cell adhesion molecule CD2 at 2.5 Å resolution. *Structure* **2**, 755–766.
24. White, J., Helenius, A. & Gething, M.J. (1982). Haemagglutinin of influenza virus expressed from a cloned gene promotes membrane fusion. *Nature* **300**, 658–659.
25. Heinz, F.X., *et al.*, & Kunz, C. (1994). Structural changes and functional control of the tick-borne encephalitis virus glycoprotein E by the heterodimeric association with protein prM. *Virology* **198**, 109–117.
26. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.
27. Merrit, E.A. & Murphy, M.E.P. (1994). Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873.

David Stuart, Laboratory of Molecular Biophysics and Oxford Centre for Molecular Sciences and Patrice Gouet, Laboratory of Molecular Biophysics, Rex Richards Building, South Parks Road, Oxford OX1 3QU, UK.