Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins

Philip Chambers, Craig R. Pringle and Andrew J. Easton*

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, U.K.

Extensive regions of heptad repeat units consistent with an \( \alpha \)-helical coiled coil conformation are located adjacent to hydrophobic, potentially fusion-related regions in the amino acid sequences of paramyxovirus fusion and retrovirus envelope glycoproteins. Similar arrangements of hydrophobic peptides and heptad repeat units exist in coronavirus peplomer proteins and influenza virus haemagglutinins. This suggests that there may be similarities in the structures of these proteins and in the functions of the hydrophobic fusion-related regions during virus entry.

It has been proposed that many transmembrane proteins may include sections of \( \alpha \)-helical coiled coil, in which two or more \( \alpha \)-helices associate lengthwise by coiling around each other to maintain contact along hydrophobic faces (Cohen & Philips, 1981; Cohen & Parry, 1986). Coiled coils may form in proteins where extensive regions that are devoid of potential helix-breaking prolines occur, and consist of seven residue (heptad) repeats of amino acids in a sequence periodicity \((a \ b \ c \ d \ e \ f \ g)\) in which the side-chains of amino acids in positions a and d are predominantly bulky and hydrophobic or neutral. These features in amino acid sequences are capable of generating hydrophobic faces on each interacting \( \alpha \)-helix, within or between polypeptide chains (Cohen & Parry, 1986; McLachlan & Karn, 1983). The structures of two oligomeric transmembrane glycoproteins, the influenza virus haemagglutinin (HA) and the trypansomone variable surface glycoprotein, include elements of such a supersecondary structure (Wilson et al., 1981, Freyman et al., 1984). Coronavirus peplomer glycoproteins may also contain such a structure (de Groot et al., 1987).

We have determined the nucleotide and deduced amino acid sequences of cDNA clones from pneumonia virus of mice, a paramyxovirus classified in the genus Pneumovirus (Chambers et al., 1990). A region containing heptad repeat units was detected in the amino acid sequence of the fusion glycoprotein (Fig. 1). The heptad repeat pattern is conserved in the amino acid sequence of the fusion glycoprotein of another pneumovirus, human respiratory syncytial virus, and also in the amino acid sequences of the fusion glycoproteins of members of the more distantly related Paramyxovirus and Mollivirus genera of the family Paramyxoviridae (Fig. 1). The heptad repeat regions are relatively poor in glycines, contain no helix-breaking prolines and charged amino acid side-chains are scattered in all heptad positions except a and d. The heptad repeat pattern, shown in Fig. 1, is located in the F1 portion of the paramyxovirus fusion glycoproteins, adjacent to the hydrophobic amino terminus (Fig. 2). Furthermore, the heptad regions of the paramyxovirus fusion glycoproteins shown in Fig. 1 represent that part of the protein which is most similar between paramyxoviruses and pneumoviruses (Fig. 3 in Chambers et al., 1986). A second region of similarity, which is located near the transmembrane domain, also consists of heptad repeats and is discussed below (Buckland & Wild, 1989).

This comparison was extended to the cleavage/activation sequences of retroviruses because similarity between the amino terminus of F1 of respiratory syncytial virus and the envelope glycoprotein of human immunodeficiency virus has been suggested (Gallaher, 1987). A region of heptad repeat units has been located on the carboxy-terminal side of the retroviral cleavage site (Fig. 1 and 2). These heptad repeats commence approximately 20 to 40 amino acids from the cleavage sites in paramyxoviruses and the oncovirus and lentivirus subgroups of the retroviruses. For consistency between the oncovirus and lentivirus sequences, heptad positions a and d are indicated only in the region where most of the aligned residues have bulky side-chains (Cohen & Parry, 1986). Several other paramyxovirus fusion and retrovirus envelope glycoprotein sequences have also been examined; that of equine infectious anaemia virus shown in Fig. 1 is the only sequence identified thus far that contains a charged amino acid residue at any of the heptad positions a or d. Although this is probably un-
amino acid side-chains in heptad positions a or d favourable, there are many precedents for the presence of polar (N and Q) or charged (K, R or E; D is rare) amino acid side-chains in heptad positions a or d (McLachlan & Karn, 1983). Statistically, there is a highly significant association by the $\chi^2$ test of hydrophobic amino acids (I, L, V, M, F and Y) in heptad positions a and d for paramyxovirus and retrovirus sequences in the regions indicated in Fig. 1 (data not shown). It seems unlikely that the regions of heptad repeats described above are present by chance in all of these glycoproteins because extensive patterns of heptad repeats have not been detected in multiple alignments of other paramyxovirus proteins (Rima, 1989) or in alignments of sequences from proteins of known non-coiled coil structure (influenza virus HA head region, Hiti et al., 1981; influenza virus neuraminidase, Air et al., 1985; haemoglobin, Barton & Sternberg, 1987).

Potential long $\alpha$-helices formed by these heptad repeat regions may be terminated at the site of the first proline residue, which is conserved in most of the paramyxovirus sequences, or the paired glycines, the presence and approximate location of which are conserved in most of the retroviral sequences. These potential breaks are closely followed by conserved cysteines which may be involved in disulphide bond formation with the other peptide of the fusion protein monomer (Buckland et al., 1986).
Fig. 2. Location of hydrophobic and heptad repeat regions in virus fusion glycoproteins. Also shown are the predicted probabilities of $\alpha$-helix and $\beta$-sheet structure (Garnier et al., 1978) and the hydrophobicity profiles (Hopp & Woods, 1981), with a window of 20 residues, determined using the Microgenie sequence analysis program (Beckman Instruments). The key beneath the figure indicates the scale, cleavage/activation sites, heptad repeat regions and hydrophobic signal, transmembrane and potential fusion-related regions. A representative paramyxovirus (Newcastle disease virus, a), retrovirus (Rous sarcoma virus, b), coronavirus (infectious bronchitis virus, c) and influenza A virus (A/Aichi/2/68, d) is indicated. From the amino terminus at the left, the Newcastle disease virus protein is cleaved into F2 and F1 portions the Rous sarcoma virus protein is cleaved into gp85 and gp37 portions, the infectious bronchitis virus protein is cleaved into S1 and $S_2$ portions and the influenza A virus protein is cleaved into HA1 and HA2 portions. For display purposes, a break has been introduced into the infectious bronchitis virus $S_1$ portion, which is 537 amino acids in length. $\text{pentagon}$, Signal sequence; $\text{square}$, cleavage site; $\text{vertical bar}$, transmembrane region; $\text{closed oval}$, heptad region; $\text{open oval}$, fusion-related region.

1987; Shinnick et al., 1981). The spumavirus subgroup of the retroviruses differs somewhat in that a rather longer region of heptad repeats commences 49 residues from the proposed cleavage site (Fig. 1). Paramyxovirus fusion glycoproteins and retroviral envelope glycoproteins are thought to be oligomeric (Sechoy et al., 1987; Pinter & Fleissner, 1979; Pepinsky et al., 1980) and therefore regions of heptad repeats may be of great structural importance if they form $\alpha$-helical coiled coils between the subunits.

Heptad repeat regions have also been described in coronavirus peplomer glycoprotein sequences (de Groot et al., 1987). For simplicity, coronavirus sequences are displayed in Fig. 1 showing the phasing of heptads adjacent to the hydrophobic regions described below, extending only until it becomes necessary to position gaps to align similar sequences. A further parallel between the coronavirus peplomer and paramyxovirus fusion glycoprotein sequences is the presence of a second series of heptad repeats immediately adjacent to the presumed transmembrane domain of all of these glycoproteins (de Groot et al., 1987; Buckland & Wild, 1989).

The corresponding regions of the paramyxovirus fusion proteins contain heptad repeats which have been likened to a leucine zipper motif (Buckland & Wild, 1989; Landschultz et al., 1988), which may form part of a helical bundle, interacting with the longer heptad regions shown in Fig. 1 to stabilize the base of fusion protein oligomers. This has been proposed for coronaviruses and is demonstrated in the trypanosome variable surface glycoprotein (de Groot et al., 1987; Freymann et al., 1984). Retrovirus envelope glycoprotein sequences also contain short regions predicted to be helical adjacent to the transmembrane domain but no extended patterns of heptads are conserved.

Hydrophobicity and predicted $\alpha$-helical and $\beta$-sheet structure profiles of representative paramyxovirus, retro-, coronavirus and influenza viruses are shown in Fig. 2. The heptad repeat regions are predicted to be largely $\alpha$-helical in all these cases (and also for other representatives of these virus groups; not shown). Studies using circular dichroism also suggest that paramyxovirus fusion proteins contain large amounts of $\alpha$-helical secondary structure (Hsu et al., 1982). The heptad regions of
paramyxovirus fusion and coronavirus peplomer glycoproteins are reasonably hydrophilic despite the regular pattern of hydrophobic amino acids in positions a and d, consistent with the formation of an extended structure (Cohen & Parry, 1986). The heptad regions of some retroviruses, for example human immunodeficiency virus (not shown), are somewhat more hydrophobic, suggesting that they may be less exposed to solvent than those of the other examples shown.

Heptad repeat regions have been detected in several other virus glycoprotein sequences. Glycoprotein gB of herpes simplex virus, and its homologues in other herpesviruses, contains a short run of heptad repeats that continues into a non-heptad region predicted to be helical (herpes simplex virus residues 502 to 546; Bzik et al., 1984; Pellett et al., 1985); in arenaviruses the glycoprotein precursor contains a reasonably long run of heptad repeats with a high prediction of an α-helix (Lassa virus residues 309 to 346; Auperin et al., 1986). Again, residues adjacent to membrane-associated regions are hydrophilic and predicted to be in a helical conformation (not shown).

Hydrophobic regions thought to be involved in virus entry and cell fusion are located between the cleavage/activation site and the heptad repeat regions in both paramyxovirus and retroviruses (Fig. 2). Similarity can be detected amongst all paramyxoviruses if gaps are inserted into the pneumovirus sequences (Fig. 1). These hydrophobic regions also have sequences consistent with predictions of an α-helix (Server et al., 1985; Richardson et al., 1986) but, if the heptad pattern is simply extended back towards the cleavage site, positions a and d would be predominantly occupied by glycine or alanine rather than the bulkier hydrophobic residues preferred in these positions in α-helical coiled coils (Cohen & Parry, 1986). In contrast, the hydrophobic sequences between cleavage sites and the heptad repeat regions of the retroviral glycoproteins are quite diverse. Cleavage is not required for fusion activity of coronaviruses and there is no hydrophobic region adjacent to the cleavage site (Spaan et al., 1988). There is, however, a hydrophobic region on the amino-terminal side of the longer area of heptad repeats, as in paramyxovirus and retrovirus glycoproteins (Fig. 1 and 2). By analogy to paramyxovirus and retroviruses, these regions may function in virus entry and cell fusion. These hydrophobic regions in the retro- and coronavirus glycoproteins are usually rich in alanine and turn-inducing amino acid residues (glycines or prolines).

These predicted features have some similarities to the known structure of the influenza virus HA, the virus fusion protein (Huang et al., 1980). The HA is a trimer built around a bundle of α-helices. It is primed for function by proteolytic cleavage which generates a new, glycine-rich, hydrophobic amino terminus on the HA2 peptide (Fig. 2) but is not proficient in fusion until a conformational change is induced by incubation at acid pH (Wilson et al., 1981; Klenk et al., 1975; Waterfield et al., 1979; Skehel et al., 1982). The amino acid sequences of the HA2s continue with two series of heptad repeats (compressed to one region in Fig. 2); the first corresponds to the shorter helices of the stem, the second corresponds to the inner surface at the top of the longer helices where they make close (1 nm) contact. The conformational change necessary to activate the protein at pH 5 may involve upward movement by the shorter stem helices to form the knob which appears at the top of the stem and might therefore contain elements of anti-parallel α-helical coiled coil (Ruigrok et al., 1986).

In conclusion, heptad repeats may be associated with the presence of long, inter-subunit, α-helical coiled coils which could form the backbones of the projections or spikes on the viral envelopes, as predicted previously in general terms and demonstrated in influenza virus HA and (in a non-viral example) the trypanosome variable surface glycoprotein (Cohen & Philips, 1981; Wilson et al., 1981; Freymann et al., 1984). The heptad repeat regions shown in Fig. 1 may form a single long helix in paramyxoviruses and retroviruses. The hydrophobic regions that are located at the amino-terminal ends of the heptad repeat regions (Fig. 2) are thought to function in virus entry and cell fusion (Gething et al., 1978; Richardson et al., 1980; Shinnick et al., 1981). Assuming that the proposed long helices are oriented more or less perpendicular to the viral membrane, the hydrophobic regions would be in position to function during virus entry. If the amino terminus of the long helix pointed away from the virus core, the hydrophobic region could interact with, and perhaps destabilize, the cell membrane; if it pointed towards the core, it could interact with the viral membrane. If the subunit interface of the virus fusion proteins can form a hydrophobic channel, lipid movements between membranes to be fused may be allowed. Thus, diverse groups of enveloped viruses may have developed similar structures to enable the fusion of virus and cellular membranes.

References


adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proceedings of the National Academy of Sciences, U.S.A.* 80, 3618–3622.


(Received 26 June 1990; Accepted 28 August 1990)