Reviews

The novel glycoproteins of coronaviruses

Lawrence Sturman and Kathryn Holmes

The two glycoproteins of coronavirus envelopes have different functions in intracellular budding, cytopathic effects and virus infectivity, and have markedly different biochemical properties.

Glycoproteins of viral envelopes mediate many of the biological activities of enveloped virions, such as attachment of virus peplomers (spikes) to receptors on the cell surface, entry of virions into cells by fusion of the viral envelope with cellular membranes, virus-induced cell fusion, and cellular and humoral responses to viral infection. In this brief review we will summarize recent studies on the envelope glycoproteins of coronaviruses, emphasizing the biochemical characteristics associated with their different roles in intracellular budding of virions, virus-induced cell fusion and immune responses to coronavirus infections.

A model for the structure of the coronavirus virion is shown in Fig. 1. Coronaviruses are large, enveloped viruses with helical nucleocapsids that contain RNA genomes of positive or message-sense polarity[1]. The viral envelope is a lipid bilayer with two viral-encoded glycoproteins which we have called E1 and E2. The glycoprotein which forms the viral peplomers or spikes, E2, is similar to the spike glycoproteins of other enveloped RNA viruses. In contrast, the membrane-associated glycoprotein, E1, which appears to function like the nonglycosylated M proteins or orthomyxo-, paramyxo-, and rhabdoviruses, differs markedly from other viral glycoproteins in its structure, processing, and intracellular transport. The major differences between the E1 and E2 glycoproteins of the A59 strain of murine hepatitis virus (MHV-A59) are summarized in Table I and will be discussed in detail after a summary of the virus replication strategy.

L. Sturman is at the Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany, New York 12201 and K. Holmes is at the Department of Pathology, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814, USA.

Upon entry into the cell, the positive-strand RNA genome (M_r 5.4 × 10^6) apparently acts as messenger RNA (mRNA) for the synthesis of RNA-dependent RNA polymerase[7]. This enzyme transcribes the genome, forming a genome-sized, negative-strand template[8]. From this template, new genomic RNA, and six subgenomic mRNAs are transcribed[7,9]. These capped and polyadenylated RNAs all share a common 3′ end and extend for different lengths in the 5′ direction, forming a set of seven overlapping transcripts[9]. Recent studies by Lai *et al.*[10,11] and Spaan *et al.*[12] show that a leader RNA at least 70 nucleotides in length, encoded by the 3′ end of the negative-strand template, is added to the 5′ end of each mRNA and genomic RNA. Although it is not yet certain how the leader RNA is added to the noncontiguous sequences of the mRNAs, the available evidence suggests that the leader is synthesized separately and utilized as a primer for transcription (Ref. 13 and M. Lai, personal communication). *In vitro* translation of the isolated viral mRNAs suggests that only the gene at the 5′ end of each mRNA species is translated, yielding a single polypeptide[14–16].

E2, the spike glycoprotein

Translation *in vitro* indicates that synthesis of E2 is directed by mRNA 3 (M_r 2.6 × 10^6). E2 is synthesized on ribosomes bound to the rough endoplasmic reticulum (RER) and is co-translationally glycosylated by transfer of N-linked oligosaccharides from dolichol phosphate carriers to asparagine residues[2–5]. The apoprotein appears to be about M_r 0.12 × 10^6 as shown by experiments in which attachment of the oligosaccharides was prevented by tunicamycin, and from translation *in vitro* in the absence of microsomal membranes[14,17,18]. After glycosylation, E2 has an apparent M_r of 0.18 × 10^6 (180 K). Palmitic acid is covalently bound to E2, and this acylation probably occurs as the glycoprotein is transported through the Golgi apparatus.

Virions form by budding from RER and Golgi membranes and are released from the cell apparently via the cellular secretory apparatus. Because viral components assemble at RER and Golgi membranes, subsequent processing of E2 may occur as intact virions are transported through intracellular compartments. Pulse-labeling shows that most of the E2 is incorporated into virions and released from the cell within 2 hours of synthesis[19]. A small proportion of E2 which is not incorporated into virions is transported to the plasma membrane where it may participate in cell fusion and render cells susceptible to cell-mediated cytotoxicity. This route for intracellular transport of E2 to the plasma membrane is the same as that of the envelope glycoproteins of orthomyxo-, paramyxo-, rhabdo- and alphaviruses.

Proteolytic cleavage of 180K E2 by host-cell proteases to form products of M_r 0.09 × 10^6 apparently occurs as a late step in transport of virions and E2 to the plasma membrane. Two such 90K species, which we have called 90A and 90B, can be separated by hydroxyapatite chromatography[20]. These cleavage products have different amino acid compositions, and only the 90A form is acylated. On virions released from different cell types, the ratio of 180K E2 to 90K E2 varies considerably, presumably reflecting differences in host-protease activity. Cell fusion occurs in coronavirus infections *in vivo* and *in vitro* and is mediated by E2[2]. Host-dependent cleavage of E2 may be an important determinant of coronavirus virulence.

Proteolytic cleavage of E2 to 90K forms is required for rapid cell fusion by exogenous virus[20]. Treatment of virions with trypsin, elastase, thermolysin or chymotrypsin activates cell-fusing activity so that cells can be fused in one hour in the absence of protein synthesis (F. Baker, C. Ricard and L. Sturman, unpublished results). As sequence data become available from molecular cloning and amino acid sequencing, it will be of considerable interest to compare the amino acid sequences near the protease sites on E2 with sequences of fusion glycoproteins of other viruses such as the F glycoprotein of paramyxoviruses and the HA glycoprotein of influenza

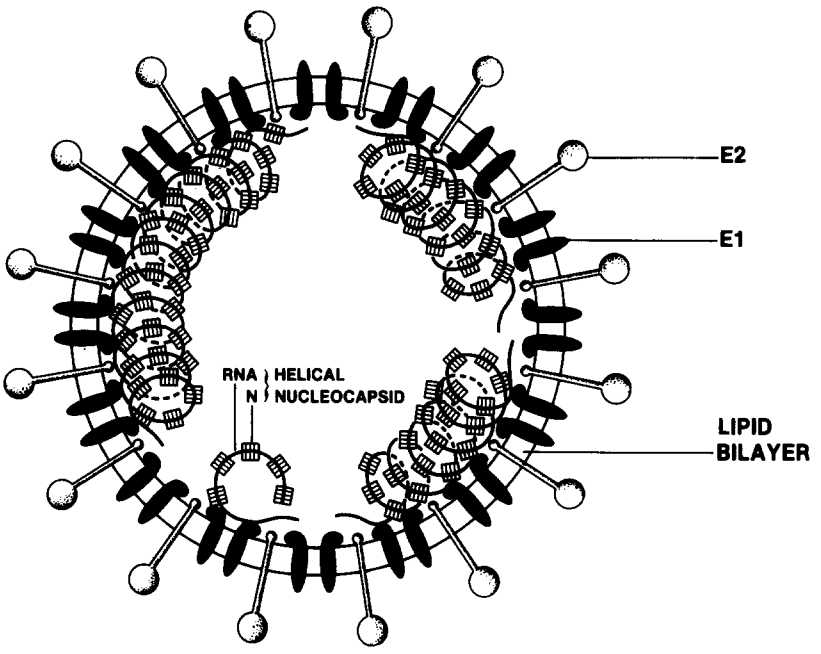


Fig. 1. Model of a coronavirus. The viral nucleocapsid is a long, flexible helix composed of the (+)-strand genomic RNA (M, 5–7 × 10⁶) and many molecules of the phosphorylated nucleocapsid protein, N (M, 0.05–0.06 × 10⁶). The viral envelope includes a lipid bilayer derived from intracellular membranes of the host cell and two viral glycoproteins, E1 (M, 0.02–0.03 × 10⁶) and E2 (M, 0.18–0.20 × 10⁶). The peplomers are composed of E2. The membrane glycoprotein E1 penetrates through the lipid bilayer and interacts with the nucleocapsid within the virion.

virus. In these cases, cleavage yields a new amino terminus with a markedly hydrophobic sequence of amino acids, which may participate in the cell-fusion reaction.

The E2 glycoprotein also appears to be responsible for the unusual pH-dependent thermolability of coronaviruses. There is a rapid loss of coronavirus infectivity above pH 6.5 associated with aggregation of the peplomeric glycoprotein. This suggests that there may be a pH-dependent conformational change in the E2 glycoprotein[21]. The amino acid sequence of E2 will soon be known from cloning and sequencing studies in progress in several laboratories. Comparison of this glycoprotein from a positive-stranded coronavirus with those of the large negative-stranded orthomyxo-, paramyxo- and rhabdoviruses may reveal important structural and functional determinants of envelope glycoproteins, and may clarify the evolutionary relationship of coronaviruses with these other virus groups.

E1, the membrane glycoprotein

The molecular characteristics of E1 differ from E2 and other viral envelope glycoproteins in many ways (Table I). Messenger RNA 6 (M_r 0.9 × 10⁶) which encodes E1 can be translated *in vitro* in the absence of membranes. When microsomal membranes were added after protein synthesis had begun, Rottier *et al.* found that about 65–70% of the E1 molecule was synthesized before the protein was inserted into the lipid bilayer[22], suggesting that an internal signal sequence was used. Although E1 lacks an amino terminal signal sequence, the cellular signal recognition particle (SRP) is required for integration into microsomal membranes, since addition of SRP to a cell-free translation system blocks elongation of E1 (P. Rottier, J. Armstrong, and D. Meyer, in press). The block is released by addition of microsomes, concomitant with assembly of the protein.

The primary structure of E1 deduced from analysis of cloned cDNA derived from viral mRNA reveals that there are two long sequences of uncharged amino acids which represent potential membrane-spanning regions of the protein[23]. Investigation of the disposition of E1 in microsomal membranes by digestion with protease indicates that only small portions at the N- and C-termini (plus a portion near the center of the molecule of M_r 500) are exposed or accessible to protease in the lumenal and cytoplasmic domains respectively[22].

Glycosylation of MHV E1 differs from that of most other viral glycoproteins in several ways[2]. First, it is glycosylated post-translationally rather than co-translationally like N-linked glycoproteins. Pulse-labeling studies of infected cells show that E1 is synthesized as a 20K apoprotein in association with the RER, and apparently migrates to the Golgi apparatus where glycosylation occurs after about 20–30 minutes. The oligosaccharides attached to E1 are different from those on E2 and most other viral glycoproteins. Glycosylation of E1 is not inhibited by tunicamycin, and the oligosaccharides are O-linked to serine and threonine residues at the amino terminus which projects from the external side of the viral envelope. Niemann and his colleagues have shown that the structures of these O-glycosidally linked oligosaccharides are identical to those of glycophorin A, a major sialoglycoprotein of the human erythrocyte membrane[24]. E1 and one type of glycophorin A possess the same amino terminal tetrapeptide sequence Ser–Ser–Thr–Thr. This region of the glycophorin molecule with its associated oligosaccharides constitutes the blood group M

Table I. Structure and function of coronavirus MHV-A59 glycoproteins[2-6]

	Spike glycoprotein (E2)	Membrane glycoprotein (E1)
Mol. wt	0.18 × 10⁶	0.023 × 10⁶
Cleavage fragments	90A, 90B	—
% in external domain	>90%	<15%
mRNA no. (mol. wt)	3 (2.6 × 10⁶)	6 (0.9 × 10⁶)
No. of epitopes	4 (MHV-JHM)	2 (MHV-JHM)
Processing	*N*-glycosylation, cleavage, acylation	O-glycosylation
Intracellular transport	RER to Golgi to plasma membrane	RER to Golgi
Functions	Cell attachment	Determines budding site
	Cell fusion	Forms viral envelope
	pH-dependent thermolability	Interacts with nucleocapsid and E2
	Neutralization (complement-independent)	Neutralization (complement dependent)
	Cell-mediated cytotoxicity	

determinant, and purified E1 also possesses blood group M activity.

Intracellular transport of E1 differs from that of E2. E1 is transported to the Golgi apparatus where it is glycosylated and accumulates in thiamine pyrophosphatase-positive cisternae (Ref. 25, and E. Dollar and K. Holmes, unpublished results). Unlike E2, E1 is not transported to the plasma membrane. The limited intracellular migration of E1 appears to determine the intracellular budding site of coronaviruses. Virions mature by budding in RER and Golgi membranes, presumably because this is the location of the E1 glycoprotein. Pulse–chase studies show that early in infection the synthesis of E1 is balanced with its release in virions. E1 is the only viral glycoprotein required for virus budding, since virions containing only nucleocapsid (N) and E1 proteins can be assembled and released in tunicamycin-treated cells[21].

As a consequence of intracellular budding, complete virions may be formed within the cell prior to insertion of large amounts of E2 glycoprotein into the plasma membrane. This would mean that enveloped virions are completed before the infected cell becomes susceptible to attack by humoral or cell-mediated immune responses directed against E2 on the cell surface. Clearly intracellular budding may provide an important mechanism for persistent infection. Indeed, coronaviruses cause persistent infections both *in vitro* and *in vivo*.

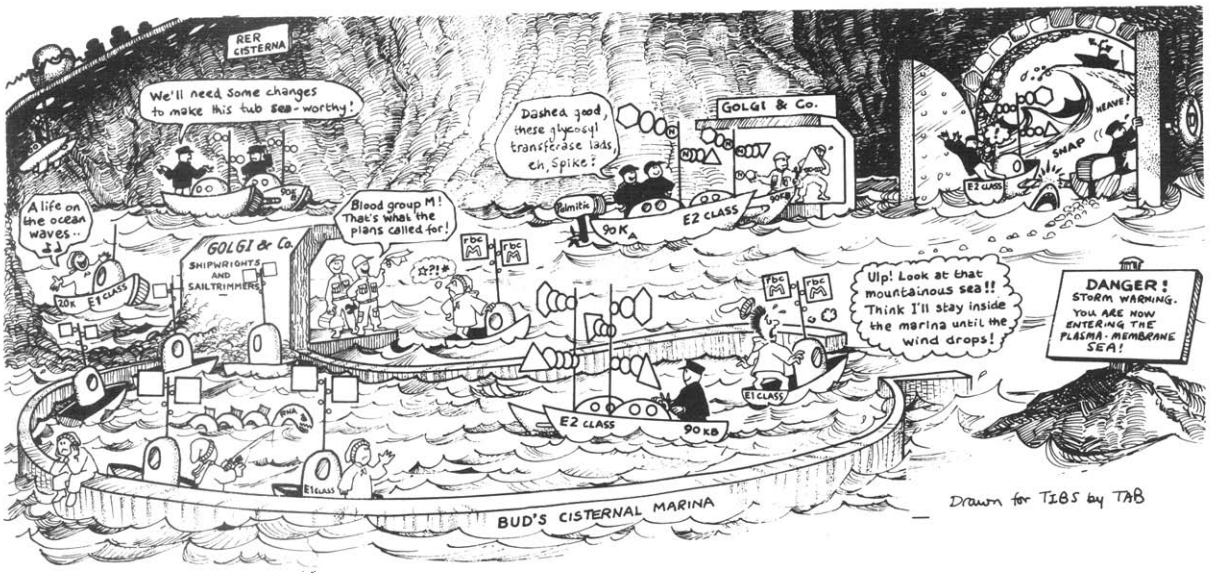
To analyse the signals responsible for the intracellular transport of E1 to the Golgi apparatus, cells have been microinjected with purified mRNA 6 or transfected with the cloned E1 gene and the distribution of E1 in these cells has been characterized by indirect immunofluorescence microscopy. Preliminary results indicate that at this level of resolution the distribution of E1 is identical to that of E1 in virus-infected cells. It is clear that the intracellular transport signal for E1 is not determined by the O-linked glycosylation of E1, since the E1 glycoprotein of the avian coronavirus IBV, which contains N-glycosidically linked oligomannosidic side-chains, also accumulates in the Golgi apparatus[26]. Comparison of the predicted amino acid sequences of MHV-A59 E1 and the IBV glycoprotein reveals that the sequences in the membrane-spanning region are highly conserved. Although their cytoplasmic domains are much less alike, both are rich in basic residues[27]. It is likely that studies with hybrid E1 proteins or site-directed mutagenesis of the cloned E1 gene will lead to identification of the signals on the E1 protein which determine its route of intracellular transport.

The chemical and physical properties of E1 are unusual, as might be expected from such a unique molecular structure. E1 is a highly hydrophobic glycoprotein which aggregates on heating at 100°C in SDS. E1 aggregates and binds to isolated nucleocapsids at 37°C. Aggregation of E1 molecules with each other in the plane of the lipid bilayer may be important in virion formation at the budding site. This lateral self-aggregation may be similar to that of the nonglycosylated M proteins of paramyxoviruses which has recently been reported[28]. E1 may form long tubules, 50 nm in diameter, which are found in the lumen of the RER in cells late in infection with MHV, or in cells infected in the presence of tunicamycin[2].

No details of the interaction between the E1 and E2 glycoproteins have yet been elucidated. We surmise, however, that at a domain in the membrane where E1 molecules have inserted and aggregated, the E2 glycoprotein on the lumenal side of the RER or Golgi membranes may assemble into spikes perpendicular to the lipid bilayer. The helical nucleocapsid of the coronavirus probably binds to the cytoplasmic domain of the E1 molecule, anchoring the flexible helix to the forming viral envelope. This interaction may stabilize the helix in a tight coil which inhibits transcription or translation of the plus-stranded genomic RNA. The nucleocapsids seen in the cytoplasm of infected cells appear to be much more flexible and possibly more loosely coiled than the nucleocapsids in the virion. Thus E1 might play a role in inhibition of transcription and in stabilization of the virion-associated nucleocapsid. E1 may also function in uncoating. Following fusion of the viral envelope with cell membranes, the binding between E1 and viral nucleocapsid may be released so that the nucleocapsid can be used in the cytoplasm for translation, and then for transcription.

Conclusion

The glycoproteins of coronaviruses are significant both for understanding of viral functions and as models for glycoproteins with different patterns of membrane insertion, glycosylation, and intracellular transport. The E2 glycoprotein most closely resembles the spike glycoproteins of enveloped viruses with negative-stranded RNA genomes in structure,



function, and requirement for protease activation. The E1 glycoprotein may represent a new class of glycoproteins resembling in some ways cellular glycoproteins associated with the plasma membrane, and in other ways, cellular glycoproteins associated with the Golgi apparatus.

Acknowledgements

We thank Drs J. Armstrong, M. Lai, H. Niemann and P. Rottier who shared their unpublished data with us. Studies in the authors' laboratories were supported by grants GM 31698 and A1 18997 from the NIH. The opinions expressed are the private views of the authors and should not be construed as official or necessarily reflecting the view of the Uniformed Services School of Medicine or the Department of Defense.

References

1 Siddell, S. G., Anderson, R., Cavanagh, D., Fujiwara, K., Klenk, H. D., Macnaughton, M. R., Pensaert, M., Stohlman, S. A., Sturman, L. and van der Zeijst, B. A. M. (1982) *Intervirology* 20, 181–189
2 Sturman, L. S. and Holmes, K. V. (1983) *Adv. Virus Res.* 28, 35–112
3 Siddell, S., Wege, H. and ter Meulen, V. (1983) *Curr. Top. Microbiol. Immunol.* 99, 131–163
4 Rottier, P. J. M., van der Zeijst, B. A. M., Spaan, W. J. M. and Horzinek, M. C. (eds) (1984) *Adv. Exp. Med. Biol.* 173, 1–409
5 ter Meulen, V., Siddell, S. and Wege, H. (eds) (1981) *Adv. Exp. Med. Biol.* 142, 1–438
6 Talbot, P. J., Salmi, A. A., Knobler, R. L. and Buchmeier, M. J. (1984) *Virology* 132, 250–260
7 Brayton, P. R., Lai, M. M. C., Patton, C. D. and Stohlman, S. A. (1982) *J. Virol.* 42, 847–853
8 Lai, M. M. C., Patton, C. D. and Stohlman, S. A. (1982) *J. Virol.* 44, 487–492
9 Lai, M. M. C., Brayton, P. R., Armen, R. C., Patton, C. C., Pugh, C. and Stohlman, S. A. (1981) *J. Virol.* 39, 823–834
10 Lai, M. M. C., Patton, C. C., Baric, R. S. and Stohlman, S. A. (1983) *J. Virol.* 46, 1027–1033
11 Lai, M. M. C., Baric, R. S., Brayton, P. R. and Stohlman, S. A. (1984) *Proc. Natl Acad. Sci. USA* 81, 3626–3630
12 Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B. A. M. and Siddell, S. G. (1983) *EMBO J.* 2, 1839–1844
13 Baric, R. S., Stohlman, S. A. and Lai, M. M. C. (1983) *J. Virol.* 48, 633–640
14 Siddell, S., Wege, H., Barthel, A. and ter Meulen, V. (1981) in *Biochemistry and Biology of Coronaviruses* (ter Meulen, V., Siddell, S. and Wege, H., eds), pp. 193–208, Plenum Press
15 Rottier, P. J. M., Spaan, W. J. M., Horzinek, M. and van der Zeijst, B. A. M. (1981) *J. Virol.* 38, 20–26
16 Leibowitz, J. L., Weiss, S. R., Paavola, E. and Bond, C. W. (1982) *J. Virol.* 43, 905–913
17 Rottier, P. J. M., Horzinek, M. C. and van der Zeijst, B. A. M. (1981) *J. Virol.* 40, 350–357
18 Niemann, H. and Klenk, H. D. (1981) *J. Mol. Biol.* 153, 993–1010
19 Holmes, K. V., Doller, E. W. and Behnke, J. N. (1981) in *Biochemistry and Biology of Coronaviruses* (ter Meulen, V., Siddell, S. and Wege, H., eds), pp. 133–142, Plenum Press
20 Sturman, L. S. and Holmes, K. V. (1984) in *Molecular Biology and Pathogenesis of Coronaviruses* (Rottier, P. J. M., van der Zeijst, B. A. M., Spaan, W. J. M. and Horzinek, M. C., eds), pp. 25–35, Plenum Press
21 Sturman, L. S. (1981) in *Biochemistry and Biology of Coronaviruses* (ter Meulen, V., Siddell, S. and Wege, H., eds), pp. 1–18, Plenum Press
22 Rottier, P., Brandenburg, D., Armstrong, J., van der Zeijst, B. and Warren, G. (1984) *Proc. Natl Acad. Sci. USA* 81, 1421–1425
23 Armstrong, J., Niemann, H., Smeekens, S., Rottier, P. and Warren, G. (1984) *Nature* 308, 751–752
24 Niemann, H., Geyer, R., Klenk, H. D., Linder, D., Sturm, S. and Wirth, M. (1984) *EMBO J.* 3, 665–670
25 Tooze, J., Tooze, S. and Warren, G. (1984) *Eur. J. Cell Biol.* 33, 281–293
26 Stern, D. F. and Sefton, B. M. (1982) *J. Virol.* 44, 804–812
27 Boursnell, M. E. G., Brown, T. K. D. and Binns, M. M. (1984) *Virus Research* 1, 303–313
28 Heggeness, M. H., Smith, P. R. and Choppin, P. W. (1982) *Proc. Natl Acad. Sci. USA* 79, 6232–6236

α_1-Antitrypsin and the serpins: variation and countervariation

Robin Carrell and James Travis

α_1-Antitrypsin is a plasma protein which protects elastic tissue from proteolytic attack. Consequently, genetic deficiency, or the oxidation of its reactive centre in cigarette smokers can result in the degenerative lung disease emphysema. Structural studies explain the mechanisms involved and have also drawn attention to a new family of serine proteinase inhibitors. The specificity of each of these inhibitors is primarily dependent on a single amino acid at its reactive centre. Site-directed mutagenesis is enabling the production of specifically designed inhibitors for therapeutic use, including an improved replacement for α_1-antitrypsin deficiency.

Human plasma contains several inhibitors of proteolytic enzymes that together form some 10% of its protein content[1]. A major example is the broad-spectrum inhibitor, α_2-macroglobulin, previously described in *TIBS* (James, 1980). However, the best studied of the inhibitors, and the one present in greatest concentration in plasma, is the more specifically targeted α_1-antitrypsin (also called α_1-proteinase inhibitor). Although this will inhibit most of the serine proteinases, it is really an anti-elastase and its prime physiological task is the inhibition of elastase released by neutrophil leucocytes.

The function of α_1-antitrypsin is known because its genetic deficiency[2] leads to a premature breakdown of connective tissue, to give a loss of elasticity in the lungs – a condition known as emphysema. Extensive studies on the molecular pathology of this genetic deficiency culminated recently in the completion of the structure[3,4] of the normal and variant molecules.

An unexpected bonus from these structural studies is that α_1-antitrypsin proved to be the archetype of a new superfamily[5] of homologous proteins. We have called this family, for convenience, the serpins since it is primarily a group of *S*erine *P*roteinase *I*nhibitors. It is becoming clear that the lessons learnt from α_1-antitrypsin are also applicable to the other members of the serpin family. One important conclusion, already apparent, is that the inhibitory specificity of the serpins is in each case primarily defined by a single amino acid at the reactive centre of the molecule. This is strongly supported by previous work[6] on the unrelated plant proteinase inhibitors (Table I). An exciting corollary is the ability to redesign the specificity of inhibition by a single substitution at the reactive centre. This ability to design inhibitors for a specific purpose is already being developed, for therapeutic purposes, using simple recombinant-DNA modifications of α_1-antitrypsin[7,8].

Robin Carrell is at the Pathology Department, Christchurch Clinical School of Medicine, Christchurch Hospital, New Zealand and James Travis was at this address on leave from the Biochemistry Department, The University of Georgia, Athens, GA, USA. In August 1984, R. Carrell was awarded the Pharmacia Prize by the New Zealand Biochemical Society. His presentation lecture covered material presented in this review.