PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Towards a Structure of the HIV-1 Envelope Glycoprotein gp120: An Immunochemical Approach

John P. Moore, Bradford A. Jameson, Quentin J. Sattentau, Ron Willey and Joseph Sodroski

Phil. Trans. R. Soc. Lond. B 1993 **342**, 83-88
doi: 10.1098/rstb.1993.0139

Towards a structure of the HIV-1 envelope glycoprotein gp120: an immunochemical approach

JOHN P. MOORE[1], BRADFORD A. JAMESON[2], QUENTIN J. SATTENTAU[3], RON WILLEY[4] AND JOSEPH SODROSKI[5]

[1]*Aaron Diamond AIDS Research Center, New York University School of Medicine, New York, New York 10016, U.S.A.*
[2]*Jefferson Cancer Institute, Philadelphia, Pennsylvania 19107, U.S.A.*
[3]*Centre d'Immunologie de Marseille-Luminy, 13288 Marseille Cedex 9, France*
[4]*National Institute of Allergy and Infectious Disease, National Institutes of Health, Bethesda, Maryland 20892, U.S.A.*
[5]*Division of Human Retrovirology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, U.S.A.*

SUMMARY

The HIV-1 surface glycoprotein gp120 binds CD4 in the initial state of virus–cell fusion. The extensive glycosylation of gp120 has thus far precluded definition of its structure by crystallographic methods. As an initial approach to a gp120 structure, the surface topology was mapped using antibodies. First, the regions of gp120 that are accessible on the surface of the native molecule, and those that are internal but exposed after denaturation, are identified. Second, epitopes for antibodies that recognize complex surface strutures comprising segments of different domains are identified. Third, we define how mutations in one domain of gp120 influence the binding of antibodies to defined epitopes on other domains. These latter approaches enable us to start to understand the inter-domain interactions that contribute to the overall structure of the gp120 molecule. Information from these studies is being used to model the structures of individual gp120 domains, and the way in which these interact in the folded protein.

1. INTRODUCTION

The CD4 glycoprotein has been subverted by the human immunodeficiency virus (HIV) surface glycoprotein gp120 as its primary cell surface receptor (Dalgleish *et al.* 1984; Klatzmann *et al.* 1984; Maddon *et al.* 1986; McDougal *et al.* 1986; reviewed by Moore *et al.* 1993). As the first step in the life cycle of HIV, the gp120–CD4 interaction is an attractive target for rational drug design and for vaccine strategies aimed at preventing HIV infection (reviewed by Moore & Sweet 1993). So far, however, no approach to blocking the CD4–gp120 interaction under *in vivo* conditions has been successful. A major obstacle to the rational design of therapeutics is our lack of knowledge of the molecular details of the gp120–CD4 interaction. Although a crystal structure for the first two domains of CD4 has been available for 3 years, and the principal details of the gp120 binding site on CD4 are consequently fairly well defined (Ryu *et al.* 1990; Wang *et al.* 1990), no such precise structural information is available for the co-ligand, gp120. The complex domain structure, delineated by nine intramolecular disulphide bonds, was described by Leonard *et al.* (1990), and is the essential starting point for structural analyses. However, a major obstacle is the lack of a crystal structure for gp120. The flexibility of gp120 together with its extensive glycosylation – more than half of the gp120 molecular mass is carbohydrate – has so far precluded its crystallization. To fill this lacuna, we have probed recombinant gp120 and mutants thereof with monoclonal antibodies to elucidate as much structural information about the molecule as we are able. We have also compared the antibody reactivity of monomeric gp120 with that of gp120 oligomers on the surface of HIV-infected cells to gain insights into the quaternary structure of the envelope glycoprotein complex in its most native form. Finally, we are using data on the antigenic structure of gp120 to assist in the preparation of computer models of sections of the gp120 molecule, and eventually of the entire protein.

2. PROBING THE ACCESSIBILITY OF gp120 LINEAR EPITOPES

The first part of our strategy is to use monoclonal antibodies (MAbs) that bind to defined, contiguous sections of gp120, judged by their reactivity with overlapping 20-mer gp120 peptides. We assess whether these MAbs bind preferentially to native recombinant gp120 (BH10 clone) in the absence of detergent, or to the same gp120 molecule after it has been denatured by boiling in the presence of SDS and dithiothreitol. In this way we assess which regions of gp120 are likely to be well exposed on the surface of the molecule, and which are poorly exposed, either because they are buried within the body of the molecule or shielded by carbohydrates. The results of these studies to date are summarized in table 1.

Phil. Trans. R. Soc. Lond. B (1993) **342**, 83–88
Printed in Great Britain

Table 1. *Surface-exposed and buried regions of recombinant HIV-1 gp120*

domain	well exposed	poorly exposed	buried or shielded	no information
C1	31–54	85–95	110–120	55–61
	62–71	100–110		71–85
V1	140–150			120–140
				150–158
V2	169–182	158–171		182–195
C2			210–231	195–210
				231–250
		275–280	250–280	280–294
V3	304–322	294–301	320–330	
C3			360–393	330–360
V4		391–405		405–423
C4	423–440		438–451	450–461
V5		461–470		
C5	499–511		465–498	

Few regions of recombinant gp120 are readily accessible to antibodies. Of those that are accessible on the monomer, preliminary results using FACS analysis suggest that regions in C1 and C5 are not well exposed on the gp120/gp41 oligomer on the surface of HIV-infected cells. Probably these sections of gp120 are occluded by other components of the oligomeric complex, such as gp41 or other gp120 molecules. Thus none of the conserved regions provides a major, contiguous target for antibodies on the mature envelope glycoprotein complex. In contrast, several of the variable regions provide readily accessible antibody binding sites, including the apices of the V1, V2 and V3 loops. The variability of these loops is almost certainly a product of immune selection pressures *in vivo*, so it is no surprise that they are antibody binding sites on both the gp120 monomer and oligomer.

It should be appreciated that the resolution of our gp120 'map' is relatively low; most epitopes for peptide-reactive MAbs have only been determined to within approximately 10–15 amino acids. Thus regions that we assess to be poorly exposed on recombinant gp120 may have short stretches that are well exposed. For example, amino acids within the stretches 250–260, 368–370 and 455–475 have been identified as important components of the discontinuous CD4 binding-site on gp120 (Olshevsky *et al.* 1990). These amino acids mostly lie within poorly exposed regions of monomeric gp120, but key residues are probably available on or near the surface of gp120 to contact CD4.

3. INTER-DOMAIN INTERACTIONS IN gp120

To gain insights into how the different domains of gp120 interact with each other in the folded molecule, we determined how amino acid substitutions in different regions of gp120 affect the binding of MAbs to linear or discontinuous epitopes within single or multiple domains of gp120 (Thali *et al.* 1991). Of very many MAbs now analysed in this way, a few are worthy of particular note.

G3–299 and G3–42 are two of a panel of six MAbs raised by Sun *et al.* (1989) and identified as being reactive with peptides from the C4 region of gp120. We can confirm this, and show that each of the MAbs recognizes a core peptide sequence spanning amino acids 429–438. However, both G3–299 and G3–42 differ from the other MAbs in that their binding is acutely dependent on native gp120 conformation; they do not react with gp120 once it has been denatured. Analysis with a panel of gp120 mutants indicates that both G3–42 and G3–299 are very sensitive to substitutions within the C4 peptide epitope, but also to substitutions in the V3 loop. In contrast, the other C4 Mabs are sensitive to C4 but not V3 substitutions (figure 1). Consistent with the

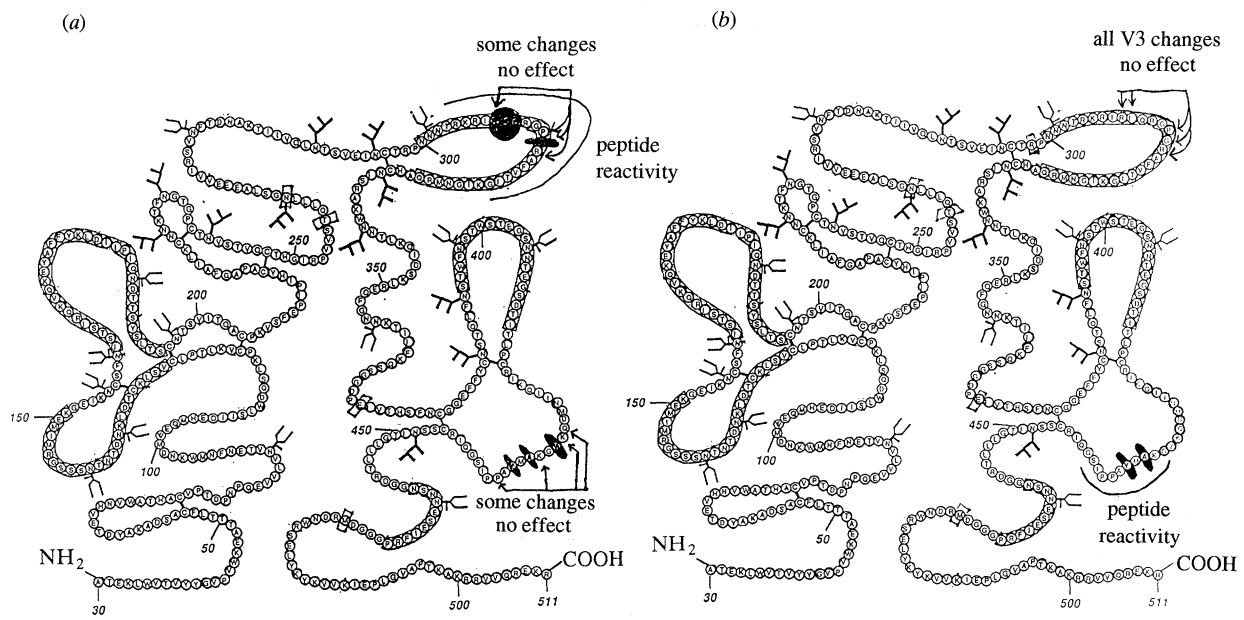

Figure 1. Effect of amino acid substitutions in V3 and C4 on binding of MAbs to gp120. Amino acid substitutions destroying or significantly impairing MAb binding are represented by solid and open circles respectively. Rectangles denote substitutions that enhance MAb binding. The MAbs tested were (*a*) G3–299; (*b*) G45–60. The depiction of gp120 in this and the following figures is derived from Leonard *et al.* (1990).

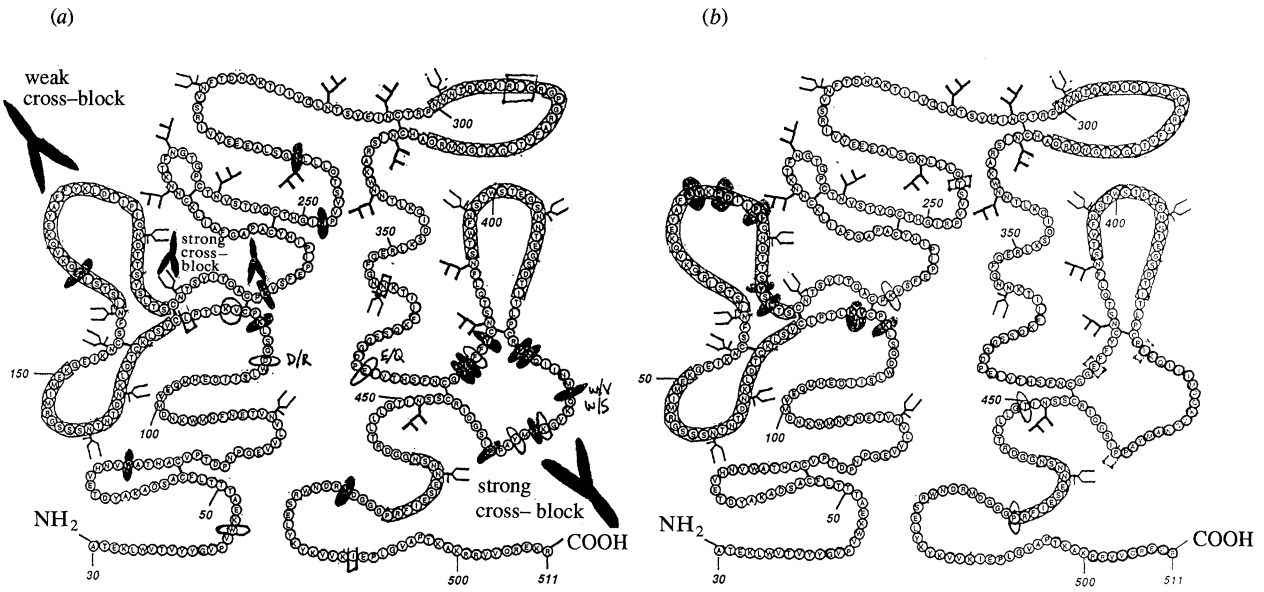

Figure 2. Effect of amino acid substitutions on binding of MAbs 48d and CRA-3 to gp120. Amino acid substitutions destroying or significantly impairing MAb binding are represented by solid and open circles respectively. Rectangles indicate substitutions that enhance MAb binding. The MAbs tested were (*a*) 48d; (*b*) CRA-3.

influence of V3 on the epitopes for G3–299 and G3–42, thrombin cleavage of the V3 loop reduces binding of these MAbs to gp120. Furthermore, G3–299 weakly binds to a V3 peptide as well as a C4 peptide (Moore *et al.* 1993*a*). Previous studies have ascribed an interaction between the V3 and C4 regions of HIV-1 gp120 (Wyatt *et al.* 1992); the present studies provide further evidence for such an association. Whether the epitopes for G3–299 and G3–42 are actually discontinuous ones spanning C4 and V3 is uncertain, but entirely possible.

Human MAbs 48d and 17b recognize discontinuous gp120 epitopes, and neutralize HIV-1, probably at a stage post-CD4 binding (Robinson *et al.* 1992; Thali *et al.* 1993). Consistent with this, their binding to gp120 is increased by prior CD4 binding. Mutational analysis indicates that multiple substitutions in gp120 abrogate 48d and 17b binding, but a major influence is provided by changes in C1 at the base of the V1/V2 loops and also by substitutions in C4 and in C3 at the base of the V4 loop (figure 2). Cross-competition studies indicate that 48d binding is strongly blocked by MAbs that bind to linear C4 epitopes, and also by MAb CRA – 3. This murine MAb is sensitive to amino acid changes in the V2 loop, and also to substitutions in C1 at the base of the V1/V2 loops (figure 2). Its epitope may span V2 and a small segment of C1/C2 (Moore *et al.* 1993*b*). Taken together, the cross-competition and mutational data suggest that the epitope for 48d spans C4 and the base of the V1/V2 loops, although other interpretations of the data are possible.

4. INTERACTIONS BETWEEN THE C1 AND C2 DOMAINS

During studies of gp120 antigenic structure using the gp120 point mutants, we noted that the binding of MAbs to linear epitopes in C1 (amino acids 80–120) was almost invariably increased by amino acid changes in C2 (amino acids 250–266). An interaction between the C1 and C2 regions of gp120 was described several years ago. It was demonstrated that an Asn-Gln (N/Q) mutation at position 262 of gp120 (267 in the original numbering system) that reduced HIV-1 infectivity was compensated for by a spontaneous Ser-Asn (S/N) reversion mutation at position 128 (Willey *et al.* 1989). We therefore explored the effect of these amino acid changes on the antigenic structure of gp120 solubilized from these viruses with non-ionic detergent. We first assessed the ability of the different gp120s to bind sCD4. All the gp120s containing the 262N/Q mutation show an approximately tenfold reduction in affinity for sCD4 compared to wild-type gp120, except for gp120 from the 262Q/128N revertant with wild-type replication phenotype: this bound sCD4 with wild-type affinity. We next determined whether the 262Q-containing gp120s had altered exposure of linear epitopes. Eight different MAbs mapping between amino acids 80–120 in C1 all bound with 5–10-fold increases in affinity to all the 262Q-gp120s compared to wild-type gp120, again except for the 262Q/128N revertant which behaved indistinguishably from wild-type gp120 (figure 3). No other regions of gp120 showed comparable increases in exposure, including epitopes around amino acid 262. Thus the 262Q gp120 mutant has an altered configuration, such that the geometry of the CD4-binding site is changed. Coincidentally or otherwise, this structural perturbation selectively increases the exposure of part of the C1 region. The second amino acid change at position 128 compensates for the deleterious effect of the 262Q mutation and restores the shape of gp120 to its natural one. We suppose that there must be a close structural proximity of amino acids 128 and 262 in the folded protein.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

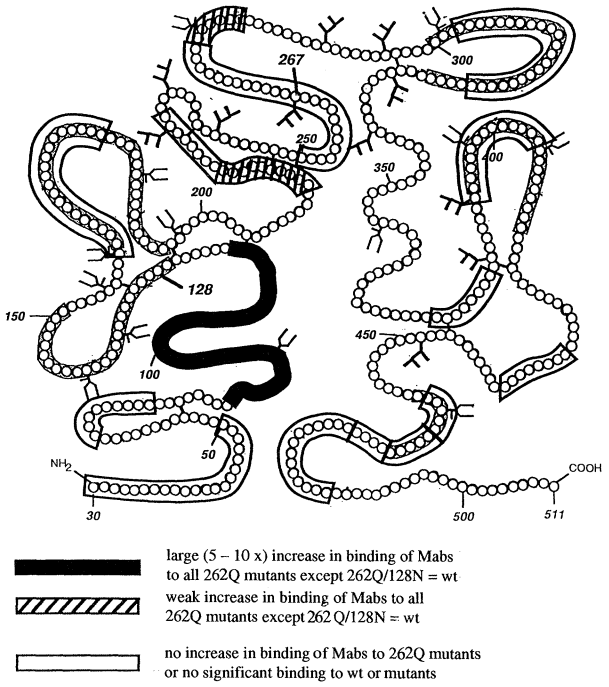

Figure 3. Effect of amino acid substitutions at position 267 on binding of MAbs to linear gp120 epitopes. Linear epitopes for MAbs binding with five- to tenfold higher affinity to the 262Q mutant than to wild-type gp120 or to the 262Q/128N revertant are marked with solid block. Linear epitopes with a more modest increase in exposure are hatched, and epitopes unaffected by the 262Q substitution are marked with an open block.

5. COMPUTER MODELLING OF gp120 DOMAINS

Information on the shape of gp120 derived from immunochemical approaches can be useful to assess the validity of computer-derived models of segments of gp120. As a first approach to combining antibody mapping data with computer modelling, we prepared a computer model of the C4/V4 domains and assessed how compatible were its predictions with epitope mapping analyses (figure 4). The molecular model was based on crude alignments of the gp120 C4/V4 domains (residues 363–457 of HxB2) with the Ig-superfamily and, in particular, with MHC Class I and Class II molecules. The crystal structure of the MHC Class I molecule (Bjorkman *et al.* 1987) was used as a template in the homology modelling of these gp120 domains. This was performed as described elsewhere (McConnell *et al.* 1992). The model was created with software (Biograf) provided by Molecular Simulations, Inc. on a Silicon Graphics Power Series 4D/48D computer. The model was based on a 'best guess' approach and, as such, is meant as a crude template to aid in experimental design, not as a precise picture of the tertiary structure of part of gp120.

It is clear that the model predicts the EVGKA-MYAPP sequence (429–438) to be exposed on the surface of C4, as implied by the binding of G3–519 and related MAbs to the native gp120 protein. This is consistent with antibody binding data for these MAbs. Regions flanking the above sequence are predicted to be largely buried, but neither the model nor the antibody binding data defines antibody epitopes more precisely than to about five amino acids. Trp-427 must be exposed for CD4 binding if it is a contact amino acid, although it may be partially shielded. The availability of additional MAbs to C4 would help to increase the precision of our model. The epitope for MAb B15 is present on the surface of V4 on the computer model, which is consistent with the ability of this MAb to bind, albeit weakly, to native gp120. The binding sites for MAbs B32 and C12 are also highlighted on figure 4. Both are present on the surface of the model, but neither MAb binds detectably to native gp120. However, these epitopes are comprised of extremely hydrophobic amino acids, and our view

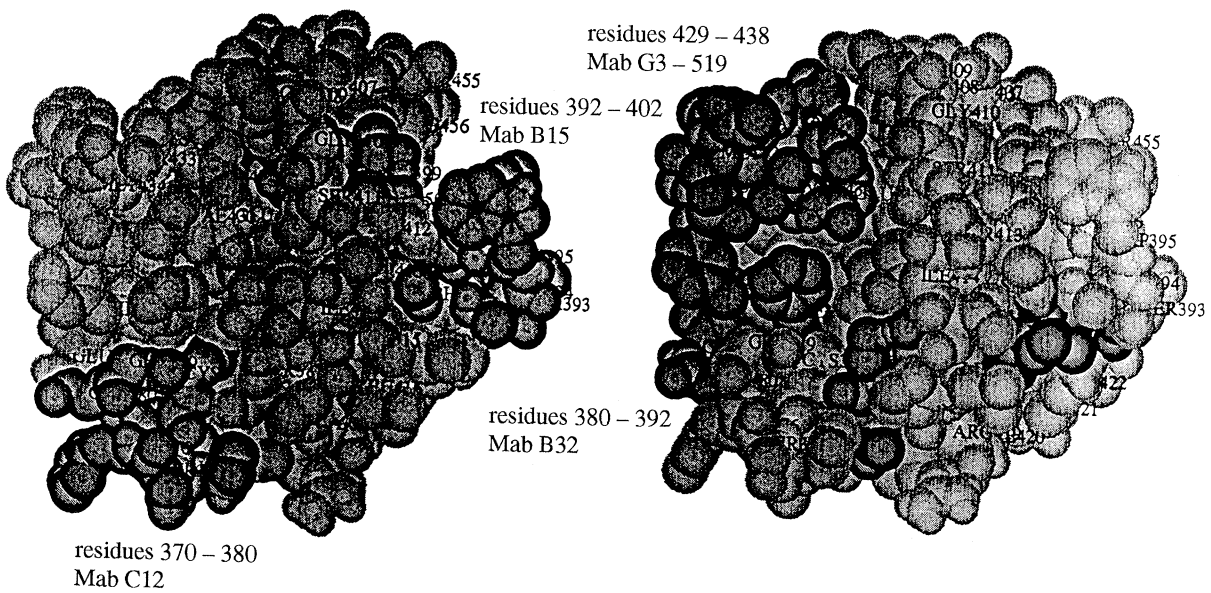

Figure 4. Computer models of the C4/V4 domains of gp120. The C4/V4 loops together with a small segment of C3 are depicted as a computer model. Two orientations of the same model are shown. Epitopes for MAbs G3–519, B15, B12 and C32 are highlighted.

is that it is probable that this region of C3 is buried by contact with other domains of gp120 in the native protein. Inter-domain interactions cannot be modelled until we have more information on gp120 structure–function relationships.

6. CONCLUSION

By a combination of mutational analysis, immunochemistry and computer modelling we hope to partially fill the structural gap left by the failure of gp120 to crystallize. It is premature to attempt to prepare even a crude model of naturally folded gp120, but we can draw some preliminary conclusions. Firstly, most of the conserved regions of gp120 are not well exposed on the oligmeric complex, either because they are buried by inter- or intra-molecular interactions with other segments of the oligomer, or because they are shielded by carbohydrates. The V3 loop is folded in such a way as to be in intimate association with the C4 region, and this may be, in turn, in reasonable proximity to the portion of C1/C2 located at the base of the V1/V2 structure. There is also a structural interaction between the C-terminal section of C1 and the C2 region around amino acid 260. Other studies (Helseth *et al.* 1991) have intimated that the extreme N- and C-terminal regions of gp120 in C1 and C5 are also likely to be in proximity to one another, and to form a binding site for gp41. The CD4 binding site comprises regions of several domains folded to be proximal in the mature molecule; these include portions of C2, C3, C4 and C5 (Olshevsky *et al.* 1991). The actual binding site itself may not be very accessible to antibodies, but regions in close proximity are. More information on the folded state of gp120 may assist in the design of better reagents for drug- and vaccine-related intervention against HIV infection.

These studies would have been impossible without the donation of monoclonal antibodies by many of our colleagues in the AIDS research community. We thank each of them.

REFERENCES

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. & Wiley, D.C. 1987 Structure of the human class I histocompatibility antigen, HLA-A2. *Nature, Lond.* **329**, 506–512.

Dalgleish, A.G., Beverley, P.C.L., Clapham, P.R., Crawford, D.H., Greaves, M.F. & Weiss, R.A. 1984 The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature, Lond.* **312**, 763–766.

Helseth, E., Olshevsky, U., Furman, C. & Sodroski, J. 1991 Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *J. Virol.* **65**, 2119–2123.

Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J.C. & Montagnier, L. 1984 T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature, Lond.* **312**, 767–768.

Leonard, C.K., Spellman, M.W., Riddle, L., Harris, R.J., Thomas, J.N. & Gregory, T.J. 1990 Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese Hamster ovary cells. *J. biol. Chem.* **265**, 10373–10382.

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

McDonnell, J.M., Blank, K.J., Rao, P.E. & Jameson, B.A. 1992 Direct involvement of the CDR3-like domain of CD4 in T helper cell activation. *J. Immunol.* **149**, 1626–1630.

McDougal, J.S., Kennedy, M.S., Sligh, J.M., Cort, S.P., Mawle, A. & Nicholson, J.K.A. 1986 Binding of HTLV-II/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule. *Science, Wash.* **231**, 382–385.

Moore, J.P., Jameson, B.A., Weiss, R.A. & Sattentau, Q.J. 1993 The HIV-cell fusion reaction. In *Viral fusion mechanisms* (ed. J. Bentz), pp. 233–289. Boca Raton, Florida: CRC Press.

Moore, J.P. & Sweet, R.W. 1993 The HIV gp120-CD4 interaction: a target for pharmacological or immunological intervention? *Persp. Drug Discov. Design* **1**, 235–250.

Moore, J.P., Thali, M., Jameson, B.A., Vignaux, F., Lewis, G.K., Poon, S.-W., Fung, M.S., Durda, P.J., Åkerblom, L., Wahren, B., Ho, D.D., Sattentau, Q.J. & Sodroski, J. 1993*a* Immunochemical analysis of the gp120 surface glycoprotein of human immunodeficiency virus type 1: probing the structure of the C4 and V4 domains and the interaction of the C4 domain with the V3 loop. *J. Virol.* **67**, 4785–4796.

Moore, J.P., Yoshiyama, H., Sattentau, Q.J., Thali, M., Poon, S.-W., Fung, M.S., Traincard, F., Robinson, J.E., Ho, D.D. & Sodroski, J. 1993*b* Probing the structure of the V1 and V2 domains of the human immunodeficiency virus type 1 surface glycoprotein gp120 with a panel of monoclonal antibodies. *J. Virol.* (In the press.)

Olshevsky, U., Helseth, E., Furman, C., Li, J., Haseltine, W. & Sodroski, J. 1990 Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 binding. *J. Virol.* **64**, 5701–5707.

Robinson, J.E., Yoshiyama, H., Holton, D., Elliott, S. & Ho, D.D. 1992 Distinct antigenic sites on HIV gp120 identified by a panel of human monoclonal antibodies. *J. cell. Biochem.* (Suppl. 16E), Abstract Q449, p. 71.

Ryu, S.-E., Kwong, P.D., Truneh, A., Porter, T.G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N.-h., Axel, R., Sweet, R.W. & Hendrickson W.A. 1990 Crystal structure of an HIV-binding recombinant fragment of human CD4. *Nature, Lond.* **348**, 419–426.

Sun, N.C., Ho, D.D., Sun, C.R.Y., Liou, R.-S., Gordon, W., Fung, M.S.C., Li, X.-L., Ting, R.C., Lee, T.-H., Chang, N.T. & Chang, T.W. 1989 Generation and characterization of monoclonal antibodies to the putative CD4-binding domain of HIV-1 gp120. *J. Virol.* **63**, 3579–3585.

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

Thali, M., Olshevsky, U., Furman, C., Gabuzda, D., Posner, M. & Sodroski, J. 1991 Characterization of a discontinuous HIV-1 gp120 epitope recognised by a broadly reactive neutralizing human monoclonal antibody. *J. Virol.* **65**, 6188–6193.

Wang, J., Yan, Y., Garret, T.P.J., Liu, J., Rodgers, D.W., Garlick, R.L., Tarr, G.E., Husain, Y., Reinherz, E.L. &

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Harrison, S.C. 1990 Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. *Nature, Lond.* **348**, 411–418.

Willey, R.L., Ross, E.K., Buckler-White, A.J., Theodore, T.S. & Martin, M.A. 1989 Functional interaction of constant and variable domains of human immunodeficiency virus type 1 gp120. *J. Virol.* **63**, 3595–3600.

Wyatt, R., Thali, M., Tilley, S., Pinter, A., Posner, M., Ho, D. Robinson, J. & Sodroski, J. 1992 Relationship of the human immunodeficiency virus type 1 gp120 third variable loop to elements of the CD4 binding site. *J. Virol.* **66**, 6997–7004.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES