Volume 271, number 1,2, 194–198 FEBS 08930 October 1990

Modelling of poliovirus

HIV-1 antigen chimaeras

M. James C. Crabbe, David J. Evans and Jeffrey W. Almond

Department of Microbiology, School of Animal and Microbial Sciences, University of Reading, London Road, Reading, RG1 5AQ, UK

Received 26 July 1990; revised version received 13 August 1990

We have used laboratory-based molecular modelling to identify structural features of antigen chimaeras of poliovirus expressing epitopes from human immunodeficiency virus (HIV-1) that may affect virus viability. Chimaeras were constructed by replacement of antigenic site 1 of VP1 by sequences corresponding to epitopes from HIV-1. Loop volume, estimated by approximating the loop to an ellipsoid was significantly ($P < 0.001$) lower in viable (2062.1 Å^3 ± 400.2) than in non-viable (3617 Å^3 ± 650.7) constructs. Our results suggest that viable virus will only be formed when antigen chimearas modified at antigenic site of VP1 have a loop occupying a similar volume in space to that occupied by the antigenic site 1 loop. In addition, the modified loop must fit with the peptide bond angles and distances at the top of the β-barrel of VP1.

Vaccine; Molecular modelling; β-Hairpin; Turn; AIDS virus; Glycoprotein; Antibody reaction

1. INTRODUCTION

Poliovirus shares a number of structural features with other picorna viruses, notably the eight-stranded anti-parallel β-barrel core of the capsid proteins, VPI-3. While the amino acids making up the core show some conservation between viruses, the loops which join the core elements vary considerably in sequence and size, and some of these constitute antigenic determinants. Fig. 1 shows the capsid protein VP1 of the type 1 (Mahoney) strain of poliovirus, with antigenic site 1 (amino acids 91–102) indicated by the arrow. This loop forms a distinct surface projection at the pentameric apex of the icosahedral particle [1], separating β strands B and C, and is known to elicit neutralising antibodies in the three serotypes of poliovirus [2]. Our studies to construct improved poliovirus vaccines have involved the use of genetic engineering techniques to replace antigenic site 1 of Sabin 1 (the live attenuated vaccine strain) with the analogous sequence from a type 3 virus [3]. The success of these studies led to the design of a cassette vector, pCAS1, [3] to facilitate the extensive modification of antigenic site 1, and to the subsequent exploitation of the Sabin 1 strain of poliovirus as an epitope expression vector. This system has been used to construct in excess of 50 antigen chimaeras, expressing known or predicted antigenic determinants from a wide range of pathogens. However, we have been unable to recover viable virus chimaeras from several cDNA constructs, presumably due to a block in a vital stage of the virus replication cycle, e.g. polyprotein folding, maturation, capsid assembly, receptor binding or virus uncoating. Studies are currently underway to determine whether chimaeras designed to express related sequences are blocked at the same point.

We have now initiated a programme of molecular modelling to characterise the modified antigenic site 1 of a number of viable and non-viable constructs to establish structural guidelines around which future viable chimaeras may be designed. Modelling topological elements is of interest in structural studies [4], and reasonably successful structure predictions have been made of immunoglobulin hypervariable regions [5,6]. In this paper we describe the modelling of chimaeras where antigenic site 1 has been replaced by residues corresponding to epitopes from human immunodeficiency virus type 1 (HIV-1). As the number of residues in the loop under consideration is 18 or less, we have used the convenient Desk Top Molecular Modeller (DTMM) and Alchemy II software packages, developed for use on IBM PC-compatible computers. These use molecular mechanics force-field calculations [7] based on the COSMIC system [8] for energy minimisation.

2. MATERIALS AND METHODS

2.1. *Hardware*

Database searches and protein co-ordinates from the Brookhaven Data Bank [9] were obtained from the SEQNET VAX computer at Daresbury, Cheshire, UK. Modelling was performed on an IBM PS/2

Correspondence address: M.J.C. Crabbe, Department of Microbiology, School of Animal and Microbial Sciences, University of Reading, London Road, Reading, RG1 5AQ, UK

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

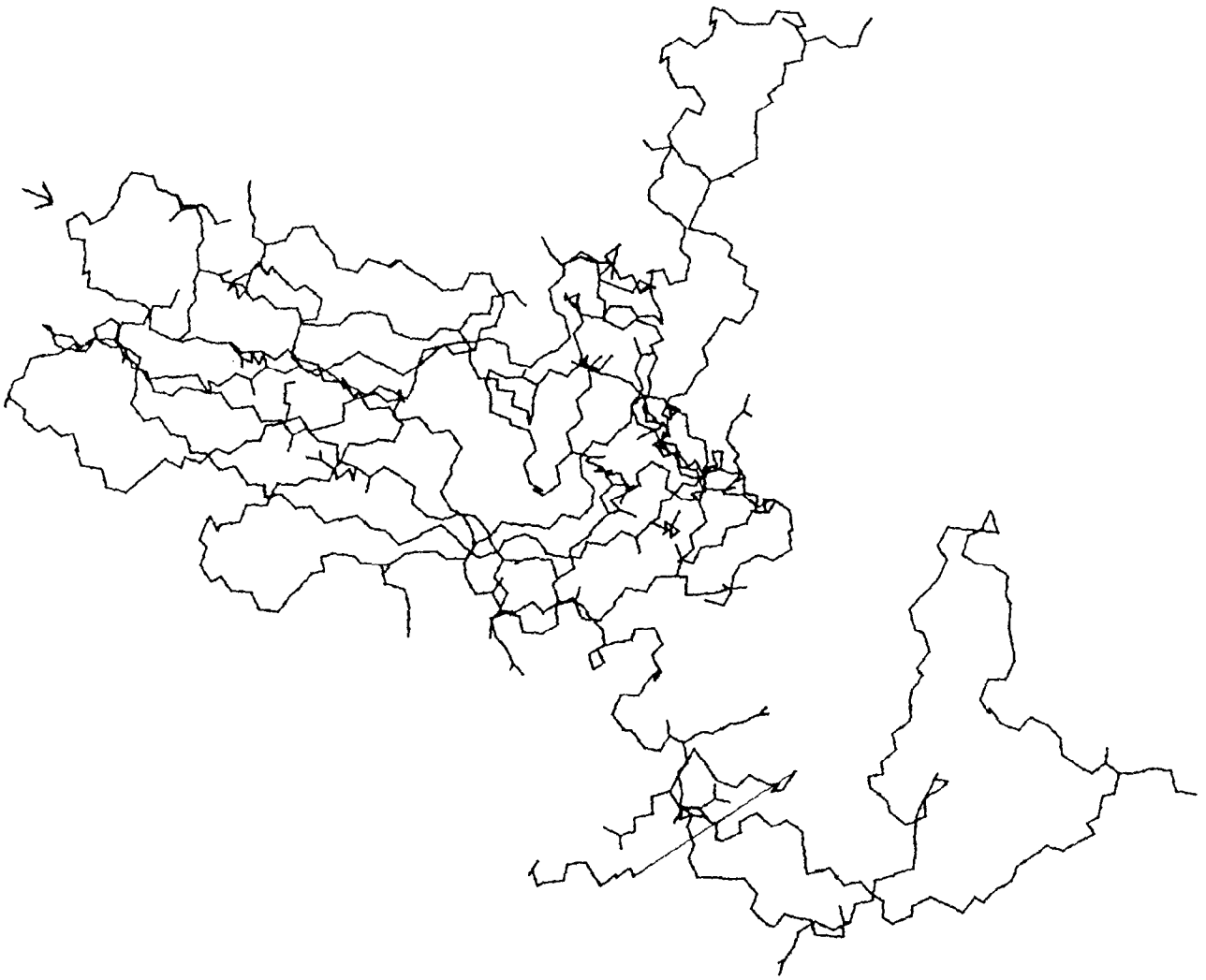

Fig. 1. The peptide chain, together with basic residue side chains, for VP1 of the Sabin type 1 (Mahoney) strain of polio virus. Atomic co-ordinates are from [1], downloaded from the Brookhaven Protein Data Bank. Antigenic loop 1 is indicated by an arrow.

Model 80 with 80387 co-processor, and a Victor V286C with 80287 co-processor.

2.2. *Modelling*

Desk Top Molecular Modeller (DTMM) version 1.2 (Oxford Electronic Publishing, Oxford University Press, Oxford OX2 6DP, UK) and Alchemy II (Tripos Associates Inc., St. Louis, MO 63144, USA) were used for molecular model building and quantitative calculations. We have used a knowledge-based approach to modelling rather than an ab initio conformational search algorithm [6]. The accuracy of the modelling method for loops was checked by building the antigenic site 1 loop of VP1, the β-hairpin loops from Immunoglobulin FAB New (light chain) (residues L64–L71 and 89–L98), and the β-hairpin loops from *S. griseus* protease A (residues 31–42 and 199–210) based on their amino acid sequences and including potential hydrogen bonds where appropriate [10], and checking their identity after minimisation against the X-ray structures, as described previously [11,12]. There were no significant differences ($< 10\%$) in hydrogen bonding distances or ϕ,ψ peptide angles between actual and predicted structures, while distance root mean square deviations between X-ray and minimised structures were ≤ 0.25 Å.

Fig. 2 shows the sequences of the loops studied. The Brookhaven database was searched for sequence similarities of loops in order to provide starting conformations where appropriate. Starting conformations of loops from viable gp41 (C), gp120 (D) and CD4 (N) chimaeras were based on similar sequences in turns capped by -Gly-Gly-,Pro-Gly- and the -Ile-Ser-Cys-Ser-Ser- loop region of pepsinogen respectively [10]. Otherwise, starting conformations were based on the antigenic site 1 loop of VP1. All loops exhibited minimum energy conformations of between 250–400 kJ/mol; there was no correlation between minimum energy and other parameters studied.

Accessible surface areas and hydrophilicities were estimated over each loop, summing values (based on solvent parameter values) for individual amino acids [13,14].

Loop lengths were estimated by measuring distances from the amide nitrogen of Asp (VP1 93) to the top of the loop.

Loop volumes were estimated by approximating the loop to an ellipsoid, measuring loop length and ellipsoid radii, and calculating volume according to:

$$^4/_3 \pi a b c$$

where a = loop length/2; b, c are the radii of the loop at 90° to each other and to the major axis of the loop. Ellipsoids have been used to assess protrusion indices in the protruding regions of proteins [15].

Virus viability was assessed as described previously [2,3].

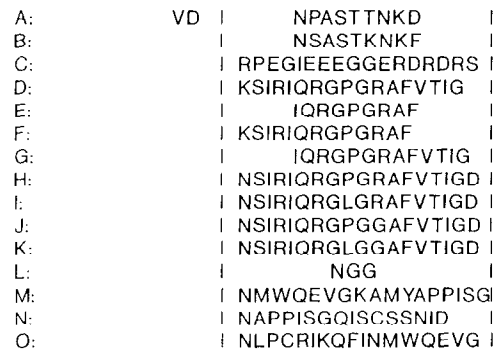

Fig. 2. Peptide sequences of the native and chimaeric polio virus constructs (A-O) studied. The residues VD (amino end) and KL (carboxy end), where the loop meets the beta barrel, were conserved throughout. (A) sequence of the loop in VP1 of the type 1 Mahoney strain of polio virus [1]; (B) sequence of the loop used in the cassette vector pCAS1 [3]; (C) sequence of the chimaera incorporating the epitope from transmembrane glycoprotein gp41 (residues 735-752) of HIV-1 (S1/env/3) [15]; (D-K) sequences of the chimaeras incorporating epitopes (GPGR) from the envelope glycoprotein of HIV-1; (L) sequence of the chimaera where the loop has been deleted; (M) sequence of the chimaera incorporating the epitope from the CD4 binding site of gp120; (N) sequence of the chimaera incorporating C-terminal residues from the epitope from the CD4 binding site of gp120; (O) sequence of the chimaera incorporating N-terminal residues from the epitope from the CD4 binding site of gp120.

3. RESULTS AND DISCUSSION

We have recently reported the construction of a Sabin 1-derived antigen chimaera expressing residues 735–752 of the transmembrane glycoprotein (gp41) of HIV-1 [16]. This sequence was chosen as it represents a relatively well-characterised group-specific neutralisation epitope of HIV-1, and therefore a potential candidate for inclusion in future subunit vaccines. The most extensively studied neutralisation epitope of HIV-1 occupies residues 307–321 of gp120 (all numbering from the BRU isolate [17]. This region of divergent virus isolates is characterised by the presence of a highly conserved NH_2-Gly-Pro-Gly-Arg-COOH sequence, which is predicted to form a β turn structure that caps a loop held at the base by a conserved disulphide bond.

We have used PC-based molecular modelling to compare energy-minimised loops from viable and non-viable antigen chimaeras to identify structural features that may affect viability. The sequences modelled (Fig. 2) include loops from poliovirus type 1 Mahoney (A) and the cassette vector pCAS1 (B), together with viable (C-D) and non-viable (E-K) poliovirus:HIV-1 chimaeras. Also shown are viable (M,N) and non-viable (O) constructs where antigenic site 1 was replaced by

Table I

Critical peptide distances, angles, hydrophilicity values, accessible surface areas, volumes and virus viability for a series of natural and chimaeric virus constructs based on antigenic site 1 of polio virus (VP1 amino acids 91–102). The sequences of the constructs are given in Fig. 2. Distances are given in Ångstroms (Å), angles in degrees, accessible surface areas in $Å^2$, volumes in $Å^3$

Cons.	Viable	N-C	C-C	C-N	N-N	Phi	Psi	Hphil	S.A.	Length	Volume
A	Yes	5.74	4.79	7.99	7.42	137.02	−84.36	5.4	1325	15.44	2438.6
B	Yes	5.75	4.83	7.13	7.41	140.11	−100.81	3.6	1415	15.33	2527.3
C	Yes	5.69	4.78	8.06	7.44	136.51	−96.38	31.5	2810	29.13	1499.2
D	No	6.01	5.87	8.37	7.89	−39.94	152.34	2.2	2685	24.04	4419.0
E	No	5.68	4.98	8.05	7.31	124.24	−100.91	1.4	1425	15.45	3198.3
F	No	16.50	15.10	16.94	17.45	−40.90	151.97	5.9	2140	23.24	4003.1
G	No	10.49	10.87	17.39	17.03	124.24	−100.91	−2.3	1970	16.46	2854.5
H	No	7.51	6.90	9.27	7.66	−39.94	152.34	2.4	2795	24.04	2758.3
I	No	7.43	6.90	9.29	7.63	−41.21	153.65	0.6	2820	23.77	3135.7
J	No	7.52	6.90	9.27	7.68	−41.44	153.34	−0.6	2645	24.04	3870.5
K	No	7.43	6.90	9.29	7.63	−41.21	153.65	−2.4	2670	23.77	3591.3
L	Yes	5.74	4.99	7.99	7.25	135.86	−83.86	0.2	310	7.56	1649.6
M	Yes	5.86	5.03	8.32	7.72	139.08	−98.92	−5.4	2700	28.50	2360.6
N	Yes	7.40	6.01	8.30	8.50	147.00	−115.00	−2.1	2250	28.74	1897.5
O	No	5.74	4.80	8.11	7.49	134.74	−118.65	−3.5	2975	27.60	4725.4

N-C-distance between the amide nitrogen of Asp (VP1 93) and the carboxyl carbon of Lys (VP1 103) where the loop meets the β-barrel. The mean distance for viable constructs was 6.03 Å ± 0.61 (non-viable 7.366 Å ± 2.75, $P>0.5$); C-C-distance between the side chain carboxyl carbon atom of Asp (VP1 93) and the carboxyl carbon of Lys (VP1 103). The mean distance for viable constructs was 5.07 Å ± 0.43 (non-viable 7.69 Å ± 3.10, $P>0.3$); C-N-distance between the side chain carboxyl carbon atom of Asp (VP1 93) and the epsilon amino nitrogen of Lys (VP1 103). The mean distance for viable constructs was 7.62 Å ± 0.41 (non-viable 10.65 Å ± 3.5, $P>0.3$); N-N = distance between the amide nitrogen of Asp (VP1 93) and the epsilon amino nitrogen of Lys (VP1 103). The mean distance for viable constructs was 7.62 Å ± 0.41 (non-viable 9.75 Å ± 4.00, $P>0.5$). Phi = the phi peptide angle between Asp (VP1 93) and the next residue in the loop. The mean phi angle for viable constructs was 139.26 ± 3.7 degrees; Psi, the psi peptide angle between Asp (VP1 93) and the next residue in the loop. The mean psi angle for viable constructs was −96.56 ± 10.5 degrees. Phi-psi angles were also measured at the other end of the loop where Lys (VP1 103) is the conserved residue; a similar pattern of results was obtained. HPhil = hydrophilicity values of each loop determined as described in the text. S.A. = accessible surface area for each loop determined as described in the text. Length = length of the peptide chain from the amide nitrogen of Asp (VP1 93) to the top of the loop. Volume = estimated volume for each loop, calculated as described in the text.

regions shown to be involved in binding the cellular receptor CD4 [18], and a loop derived from a viable construct in which antigenic site 1 was largely deleted and replaced with the sequence -N-G-G- (L). The X-ray structure of antigenic site 1 of VP1 [1] (Fig. 3A) was used as the basis for all comparisons. The distance root mean square deviation for the α-carbon atoms was 0.15 Å. The substitution of Asp-102 with Phe, due to the introduction of a restriction site during the construction of pCAS1, produced no significant quantitative differences in the peptide, and did not affect virus viability. Similarly, deletion of the loop and replacement with the sequence -Asn-Gly-Gly- (L in Fig. 2) did not affect viability or the positions of the β-barrel residues.

Table I shows critical peptide distances, phi,psi angles, hydrophilicity indices, accessible surface areas, loop lengths and estimated volumes for the sequences in Fig. 2, together with an assessment of virus viability. Statistical estimates for viable bond distances and angles are given in the table legend. Viable constructs all gave distances and angles to within about 10% of one another. While some non-viable constructs were outside this limit by over 100%, some non-viable constructs exhibited distances and angles not significantly different ($P<0.05$) from those in viable constructs. Thus while there are limits in key distances and angles where the loop joins the beta-barrel of VP1, at least one other parameter must also be important. Accessible surface area was lower in viable constructs (1801.6 $Å^2$ ± 878.11) than in non-viable constructs (2458.0 $Å^2$ ± 447.2), but this was not significant ($P<0.2$). Hydrophilicity was higher in viable constructs (11.53 ± 12.14) than in non-viable ones (6.41 ± 2.78), but again this was not significant ($P>0.5$). There was no signifi-

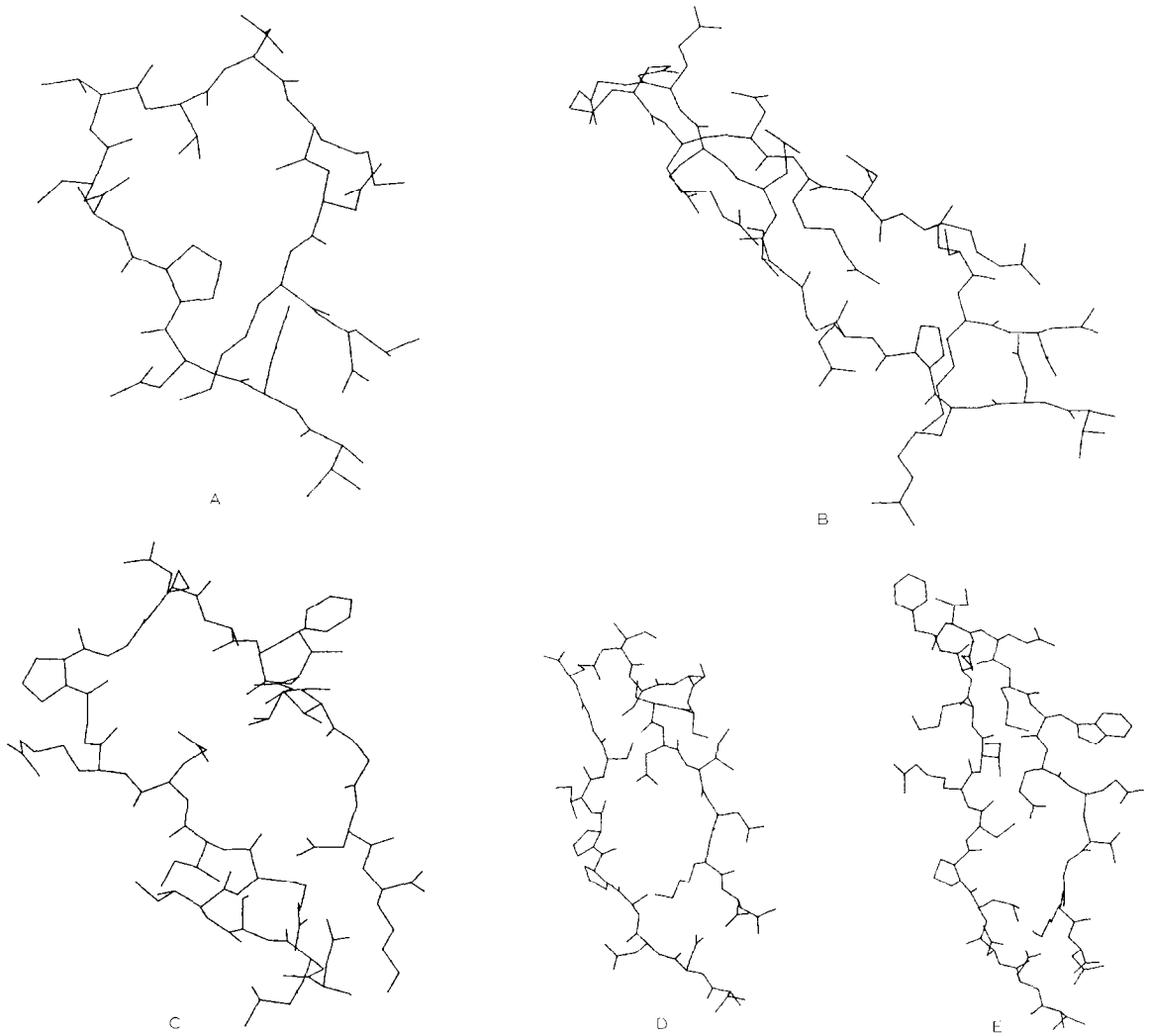

Fig. 3. Native and predicted structures of antigenic site 1 of poliovirus VP1 and chimaeras. (A) X-ray structure of antigenic site 1 on VP1 – atomic co-ordinates are from [1], downloaded from the Brookhaven Protein Data Bank. (B) Predicted structure of the viable construct incorporating the epitope from transmembrane glycoprotein gp41 of HIV-1 (S1/env/3) (C in Fig. 2). (C) Predicted structure of a non-viable construct incorporating an epitope (GPGR) from the envelope glycoprotein of HIV-1 (H in Fig. 2). (D) Predicted structure of a viable construct incorporating the C-terminal section of the CD4 binding site of gp120 (N in Fig. 2); (E) Predicted structure of a non-viable construct incorporating the N-terminal section of the CD4 binding site of gp120 (O in Fig. 2).

cant difference in mean length between viable and non-viable constructs (20.78 Å ± 8.42; 21.80 Å ± 6.0; $P > 0.6$).

Estimated loop volume, however, showed a very significant ($P < 0.001$) difference between viable (2062.1 $Å^3$ ± 400.2) and non-viable (3617 $Å^3$ ± 650.7) constructs. Of particular interest is the chimaera based on the transmembrane glycoprotein of gp41 (C, Fig. 3B), which was found to form a viable virus and to elicit broadly reactive HIV-1 neutralising antibodies [16]. Although the region analogous to antigenic site 1 of this virus was extended by 9 amino acids (Fig. 3B), the estimated volume of the loop was not significantly different from the native structure, and there were no significant changes in the parameters where the loop met the β-barrel residues. Similarly, there were no significant differences in estimated loop volume or parameters where the loop met the β-barrel for the viable constructs based on all (M in Fig. 2) or the C-terminal region section (N in Fig. 2; Fig. 3d) of the CD4 binding site of gp120.

In contrast, the gp120-derived sequences containing the -Gly-Pro-Gly-Arg- motif all showed significant changes in the peptide parameters, particularly in the distances across the peptide loop (Fig. 3C), as well as the estimated loop volumes, and all were non-viable. This was also true of the non-viable construct based on the N-terminal section of the CD4 binding site of gp120 (Fig. 3E), where exposed aromatic side chains may cause significant perturbations around the loop, as well as a significant increase in estimated loop volume.

These studies suggest that viable virus will only be formed when antigen chimaeras modified at antigenic site 1 of VP1 have a loop occupying a similar volume in space to that occupied by the antigenic site 1 loop. In addition, the modified loop must fit with the peptide bond angles and distances at the top of the beta-barrel of VP1. The B-C loop modified during these studies projects at the pentameric apex of the icosahedral virus particle, and is situated between the D-E loop (which is located closer to the pentameric apex) and the 'canyon' or cleft in the capsid face, which has been suggested as the receptor-binding region of the virus [19]. It is possible that the increased volume of the B-C loop in the nonviable chimaeras described here could interact adversely with the D-E loop, and so sterically hinder capsid formation and/or stability. Alternatively, the projecting loop may partially occlude the canyon, and so prevent the virus interacting with its cellular receptor.

Of the chimaeras containing the -G-P-G-R- sequence, E (-I-Q-R-G-P-G-R-A-F-) showed the most similarity to the native structure; only the predicted volume of the loop was significantly greater than that for the native structure. We would predict that replacing arginine residue 96 in construct E (Fig. 2) by an asparagine residue would lower the estimated volume to 2491.6 $Å^3$, almost to within 1 S.D. of the estimated mean volume for viable constructs, and might result in the formation of viable chimaeric virus. However, when this construct was produced, we could not recover viable virus constructs. This may be because replacement of one amino group on arginine by a carboxyl group on asparagine interferes with capsid formation and/or stability.

These results show that we have the foundation of a technique, based on estimates of loop volumes and key peptide distances and angles, to assess and predict the viability of related chimaeric virus constructs. Using PC-based modelling tools facilitates modelling of small loop constructs to be performed at the bench, thus expediting production of chimaeric viruses and potential vaccines. We have already initiated studies with poliovirus:SIV chimaeras and with chimaeras based on antigenic sites within VP3.

Acknowledgements: We thank the MRC, the Wellcome Trust, the Lord Dowding Fund, and the Humane Research Trust for support.

REFERENCES

[1] Hogle, J.M., Chow, M. and Filman, D.J. (1985) Science 229, 1358–1365.
[2] Minor, P.D., Ferguson, M., Evans, D.M.A., Almond, J.W. and Icenogle, J.P. (1986) J. Gen. Virol. 67, 1283–1291.
[3] Burke, K.L., Dunn, G., Ferguson, M., Minor, P.D. and Almond, J.W. (1988) Nature 332, 81–82.
[4] Sali, A. and Blundell, T.L. (1990) J. Mol. Biol. 212, 403–428.
[5] Chothia, C., Lesk, A.M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M. and Poljak, R.J. (1989) Nature 342, 877–883.
[6] Martin, A.C.R., Cheetham, J.C. and Rees, A.R. (1989) Proc. Natl. Acad. Sci. USA 86, 9268–9272.
[7] Allinger, N.L. and Burkert, U. (1982) ACS Monograph 177, Am. Chem. Soc., Washington, DC.
[8] Vinter, J.G., Davis, A. and Saunders, M.R. (1987) J. Computer-Aided Molecular Design 1, 31–51.
[9] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[10] Sibanda, B.L., Blundell, T.L. and Thornton, J.M. (1989) J. Mol. Biol. 206, 759–777.
[11] Crabbe, M.J.C. and Bicknell, G. (1989) Binary 1, 57–65.
[12] Davies, E.K. and Murrall, N.W. (1989) Computers Chem. 13, 149–156.
[13] Chothia, C. (1975) J. Mol. Biol. 105, 1–14.
[14] Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824–3828.
[15] Thornton, J.M., Edwards, M.S., Taylor, W.R. and Barlow, D.J. (1986) EMBO J. 5, 409–413.
[16] Evans, D.J., McKeating, J., Meredith, J.M., Burke, K.L., Katrak, K., John, A., Ferguson, M., Minor, P.D., Weiss, R.A. and Almond, J.W. (1989) Nature 339, 385–388.
[17] Myers, G., Rabson, A.B., Josephs, S.F., Smith, T.F., Berzofsky, J.A. and Wong-Staal, F. (1989) Human Retroviruses and AIDS 1989, Los Alamos National Laboratory, USA.
[18] Lasky, L.A., Nakamura, G.M., Smith, D.H., Fennie, C., Shimasaki, C., Patzer, E., Berman, P., Gregory, T. and Capon, D.J. (1987) Cell 50, 975–985.
[19] Rossman, M.G., Arnold, E., Erickson, J.W., Frankenburger, E.A., Griffith, J.P., Hecht, H.J., Johnson, J.E., Kamer, G., Luo, M., Mosser, A.G., Rueckert, R.R., Sherry, B. and Vriend, G. (1985) Nature 317, 145–153.