Tetrahedron Letters, Vol.31, No.32, pp 4637-4640, 1990
Printed in Great Britain

0040-4039/90 $3.00 + .00
Pergamon Press plc

# PROTEIN ENGINEERING OF HIV VIRAL PROTEINS BY TOTAL CHEMICAL SYNTHESIS: THE C-TERMINAL 104 RESIDUE PEPTIDE FROM GAG p24.

Paolo Mascagni*[1], Dwo Y Sia[2], Anthony R M Coates[2] and William A Gibbons[1],

[1]Department of Pharmaceutical Chemistry, School of Pharmacy, University of London, 29/39 Brunswick Square, London WC1N 1AX.
[2]Department of Microbiology, The London Hospital, Turner Street, London E1.

Summary: The synthesis of the C-terminal 104 residue long peptide from HIV-1 gag p24 has been achieved by automated solid phase synthesis methods. An optimised chemical protocol based on the BOC-benzyl combination was used. After purification 15-20% (w/w) of homogeneous protein was obtained. This possessed the correct chemical structure and in vitro antigenic characteristics identical to those of the parent compound.

Recent developments of the original chemical protocol for solid phase peptide synthesis[1] have been used to made several proteins[2-6] possessing biological activity and correct three-dimensional structure[7]. The advantages of this approach to protein engineering include rapid and efficient chain synthesis, synthesis of protein analogues containing non-coded amino acids or other chemical modifications and the engineering of chemical immunogens containing multiple B and T lymphocyte epitopes. In an attempt to analyse the immunological properties of the HIV-1 (LAV) gag p24 protein we have synthesized the C-terminal half corresponding to the 274-377 sequence. Prediction studies [8] have shown that this region could contain 5 T-epitopes that are well conserved between 10 different HIV-1 isolates, and the only 100% conserved T- epitopes in the gag gene products (Figure 1).

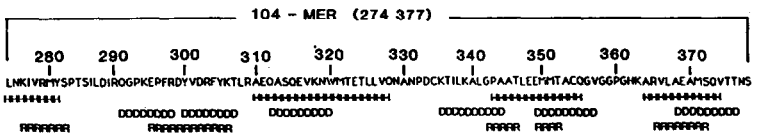


*Fig 1. Amino acid sequence of the C-terminal 104 residue peptide from HIV-I gag p24. Predictions: H = α-helical regions; D = residues within a De Lisi and Berzofsky T-epitope; R = residues within a Rothbard and Taylor T-site*[1]

The synthesis of the polypeptide was carried out in a stepwise fashion using a 430A ABI synthesiser and the chemical scheme based on the combination of N(α)-BOC, benzyl based side chain protected amino acids. Other side chain protecting groups included; tosyl for Arg, dinitrophenyl (DNP) for His, Cl- and Br-carbobenzoxy for Lys and Tyr respectively and formyl for Trp. The cysteine residues at 334 and 354 were replaced with two alanine residues. As the solid support we used the phenylacetamidomethyl resin[9] containing 0.5 mmoles of the first amino acid of the sequence (Ser, 0.72 mmol/g). All amino acids were double coupled using the chemical protocols developed by Kent[10] and the efficiency of each coupling step monitored by the quantitative ninhydrin test[11]. The average incorporation thus calculated was 99.5%[12] which, in absence of any other side reaction, predicted that about 60% of the molecules on the resin would have the correct amino acid sequence.
Several interruptions were made during the synthesis to allow for solvent and reagent replacement and to remove aliquots of peptide-resin. This last operation was both necessary and useful since, in the conditions used for the synthesis, inefficient swelling and reactivity would occur if more than about 3g of resin-peptide were used. Furthermore this provided a number of

[1] A = Ala; R = Arg; N = Asn; D = Asp; Q = Gln; E = Glu; G = Gry; H = His; I = Ile; L - Leu; K = Lys; M = Met; F = Phe; P = Pro; S = Sev; T = Thr; W = Trp; Y = Try; V =Val.

shorter sequences which together with the target peptide were used for the subsequent immunological studies. Taking into account these interruptions the overall synthesis time totalled seven days and about 3g of peptide-resin was obtained.

After the completion of chain synthesis, the DNP group from the His side chain was removed by treating an aliquot of peptide resin twice with a dimethylformamide (DMF) solution containing 2-mercaptoethanol (20% v/v) and di-isopropylethylamine (7.5% v/v). Prior to the removal of the t-BOC group from the last residue (50% TFA in dichloromethane, 1/2hr) tryptophane was deprotected with ethanolamine in aqueous DMF (pH 11).

To complete the side-chain deprotection and remove the peptide from the resin the low-high HF procedure was used[13]; scavengers used were dimethylsulphide, p-cresol and p-thiocresol (6:0.8:0.2 ml/0.5 g resin-peptide) for the low HF cleavage and p-cresol and p-thiocresol (0.8:0.2) during the high HF step.

After precipitation with diethylether the peptide was dissolved in 6M guanidine/HCl and dialysed for 48 hrs against 150 mM $NH_4OAc$ (pH 8.3). Any precipitate was then removed by centrifugation and the peptide desalted by semi-preparative reverse-phase (C4) HPLC. A large broad peak eluting between 70 and 80% acetonitrile was collected and shown by SDS-polyacrylamide gel electrophoresis to correspond to the expected molecular weight. Subsequent purification steps included size exclusion chromatography (TSK, G2000) and cation-exchange FPLC on a Mono-S column. In the latter case the peptide eluted as a single peak at 0.7 M KCl concentration (pH 5.2) (Figure 2).

Using the above procedure typical yields of homogeneous material ranged between 15 and 20% (W/W).

Acid hydrolysis (6N HCl, 24hrs, 110°C) of this polypeptide (1mg) indicated that, with the exception of Leu[14] (found 2.8; expected 4.0), the concentrations of all other residues were ±10% of the expected[14].

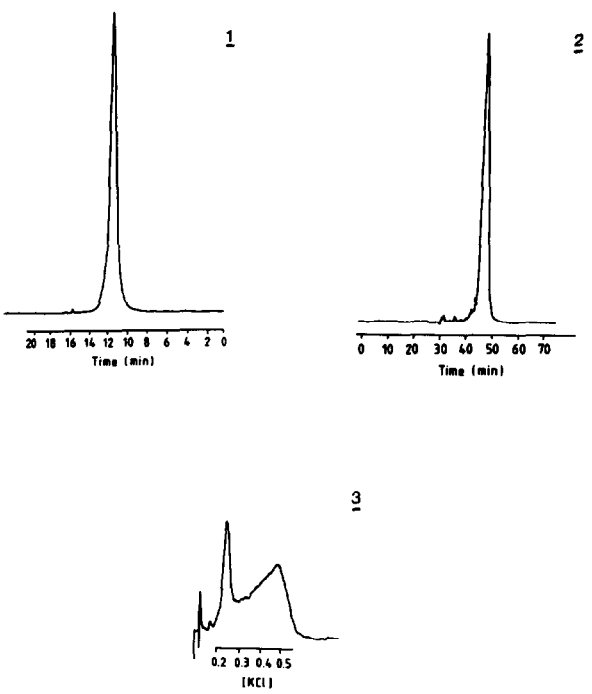


*Fig 2. Analytical HPLC profiles of purified synthetic 104 mer.*
*1 = size exclusion; 2 = C-4 reverse phase;*
*3 = Cationic exchange.*

SDS-gel electrophoresis (Figure 3A) and iso-electrofocussing[14] confirmed the essential homogeneity of the 104-mer. The chemical structure of the synthetic protein was then confirmed by amino acid sequencing and peptide mapping. The sequence from 39 successive Edman degradation cycles was consistent with that expected between residues Leu-1 and Ala-40[15]. The remainder of the sequence was determined by tryptic digestion (20 ug-enzyme, 1mg peptide in $NaHCO_3$ buffer, pH 8, 37°C, 2 hrs) followed by HPLC isolation and FAB-Mass Spectrometry characterisation of the tryptic fragments. The tryptic sequences thus determined were: 279-290, 277-298, 299-306, 319-355, 340-363 and 357-377.

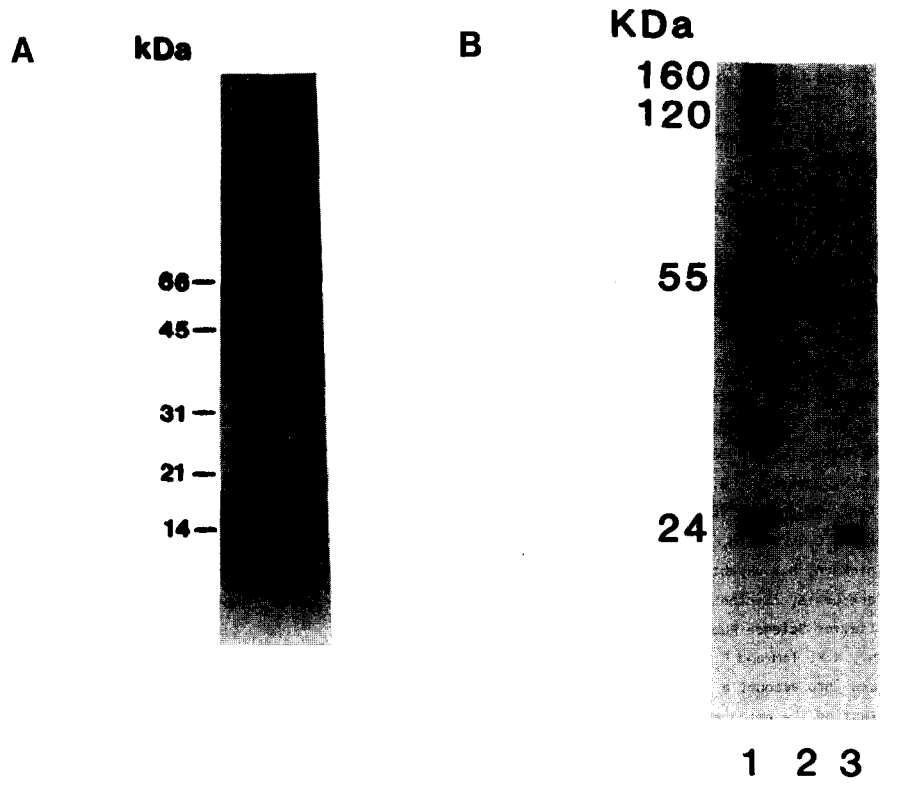


*Fig 3A. SDS-PAGE of purified 104-mer. A single band at about 12 KDa is consistent with the homogeneity of the synthetic protein and its calcuated molecular mass of 11,507.*

*Fig 3B. SDS-PAGE separated HIV-1 proteins blotted onto nitro-cellulose paper and incubated with: Lane 1 = human serum from an HIV-1 infected patient; Lane 2 = serum from an unimmunised mouse; Lane 3 = polyclonal serum from a mouse immunised with the 104-mer in Freund's incomplete adjuvant.*

Hence the combined results of peptide mapping, gas phase sequencing and FAB-MS were consistent with the expected sequence of the synthetic 104-mer. The only exception was sequence 315-318 which was not covered either by amino acid sequencing or tryptic fragmentation.

The synthetic protein possessed as many if not all the antigenic characteristics present in the parent compound which supports the fact that 3D structure is biologically intact. In particular human anti-HIV antibodies generated in the course of infection bound to the synthetic protein[15]. Furthermore, monoclonal antibodies generated against natural p24[16] also bound to two distinct B-epitopes whereas seven monoclonal antibodies made against the 104-mer bound to viral p24, the precursor p55 or both (unpublished results). In addition one monoclonal antibody bound to two different subcomponent peptides within the protein sequence.

The synthetic polypeptide was also immunogenic in mice. Figure 3b shows that polyclonal antibodies generated in mice immunised with 104-mer bound mainly to viral p24 and to a lesser extent to its precursor p55.
In conclusion, we report that protein engineering of large immunogenic proteins by total chemical synthesis has been established by rigorous quality control, the comparison of the immunoproperties of the synthetic and native molecule and finally the establishment of long-term T cell lines and monoclonal antibodies. The synthetic 104-mer molecule, furthermore, led to experimental verification of the T and B cell epitopes of the p24 gag.

Acknowledgement: We thank Applied Biosystems Inc. (Warrington, UK) for their generous support and Dr G K Toth for the tryptic digestion of the synthetic protein. We also thank Professor S B H Kent (Bond University, Brisbane, Australia) who provided the chemical protocols for the synthesis of the polypeptide.

REFERENCES AND NOTES

1. R.B. Merrifield, J.Am.Chem.Soc., 1963, 85, 2149.
2. I. Clark-Lewis, R. Aebersold, H. Ziltener, J.W. Schrader, L.E. Hood and S.B.H. Kent, Science, 1986, 231, 134.
3. R. Ramage, J. Green and O.M. Ogunjobi, Tetrahedron Letters, 1989, 30, 2149.
4. M.S.L. Lim, G.K. Toth, P. Mascagni and W.A. Gibbons, Proceedings of Innovation and perspectives in Solid Phase Synthesis symposium, Oxford 1989, in press.
5. P. Mascagni and S.B.H. Kent, J.Biol.Chem., 1989, submitted for publication.
6. J. Schneider and S.B.H. Kent, Cell, 1988, 54, 363.
7. A. Wlodawer, M. Miller, M. Jaskolski, B.K. Sathyanarayana, E. Baldwin, I.T. Weber, L.M. Self, L. Clawson, J. Schneider and S.B.H. Kent, Science, 1989, 245, 616.
8. A.R.M. Coates, J. Cookson, G.J. Barton, M.J.J. Zvelebil and M.J.E. Sternberg, Nature, 1987, 326, 549.
9. A.R. Mitchell, B.W. Erickson, M.N. Ryabtsev, R.S. Hodges and R.B. Merrifield, J.Am.Chem.Soc., 1976, 98, 7357.
10. S.B.H. Kent and I. Clark-Lewis, 'Synthetic peptides in biology and medicine' (K. Alitalo, D. Partanen and A. Vaheri Eds.), 1985, Elsevier Science Publishers, Amsterdam.
11. V.K. Sarin, S.B.H. Kent, J.P. Tam and R.B. Merrifield, Anal.Biochem., 1981, 117, 147.
12. This value does not take into account a background colour which accumulated during the synthesis. This was estimated to be about 0.25% by testing the amino acids which follow the Pro residues at positions 283, 293 and 296.
13. J.P. Tam, W.F. Heath and R.B. Merrifield, J.Am.Chem.Soc., 1983, 105, 2845.
14. Accurate determination of adjacent residues of sterically hindered amino acids, such as Leu-Leu-Val (325-327) is often inaccurate. See for example "Solid Phase Peptide Synthesis", J.M. Stewart and J.D. Young, Pierce, Rockford Ill., 1984.
15. P. Mascagni, D.Y. Sia, L.R. Haaheim, G.K. Toth, W.A. Gibbons, S.B.H. Kent and A.R.M. Coates. FEBS, 1989, submitted for publication.
16. R.P. Spence, M. Jarvill, R.B. Ferns, F.R. Tedder and D. Parker, J.Gen.Virol., 1989, in press.

(Received in UK 29 December 1989)