Structural similarity between the p17 matrix protein of HIV-1 and interferon-γ

Stephen Matthews*†, Paul Barlow*, Jonathan Boyd†, Geoff Barton‡, Robert Russell‡, Helen Mills§, Mark Cunningham§, Nicola Meyers§, Nigel Burns§, Nigel Clark*, Susan Kingsman*, Alan Kingsman* & Iain Campbell*†‖

* Department of Biochemistry, † Oxford Centre for Molecular Sciences and ‡ Laboratory of Molecular Biophysics, University of Oxford, Oxford OX1 3QU, UK
§ British Biotechnology Ltd, Watlington Road, Oxford OX4 5LY, UK

THE human immunodeficiency virus (HIV) matrix protein, p17, forms the outer shell of the core of the virus, lining the inner surface of the viral membrane[1–4]. The protein has several key functions. It orchestrates viral assembly via targeting signals that direct the *gag* precursor polyprotein, p55, to the host cell membrane[1,5–7] and it interacts with the transmembrane protein, gp41, to retain the *env*-encoded proteins in the virus[8]. In addition, p17 contains a nuclear localization signal that directs the preintegration complex to the nucleus of infected cells[9]. This permits the virus to infect productively non-dividing cells, a distinguishing feature of HIV and other lentiviruses. We have determined the solution structure of p17 by nuclear magnetic resonance (NMR) with a root-mean square deviation for the backbone of the well-defined regions of 0.9 Å. It consists of four helices connected by short loops and an irregular, mixed β-sheet which provides a positively charged surface for interaction with the inner layer of the membrane. The helical topology is unusual; the Brookhaven protein database contains only one similar structure, that of the immune modulator interferon-γ.

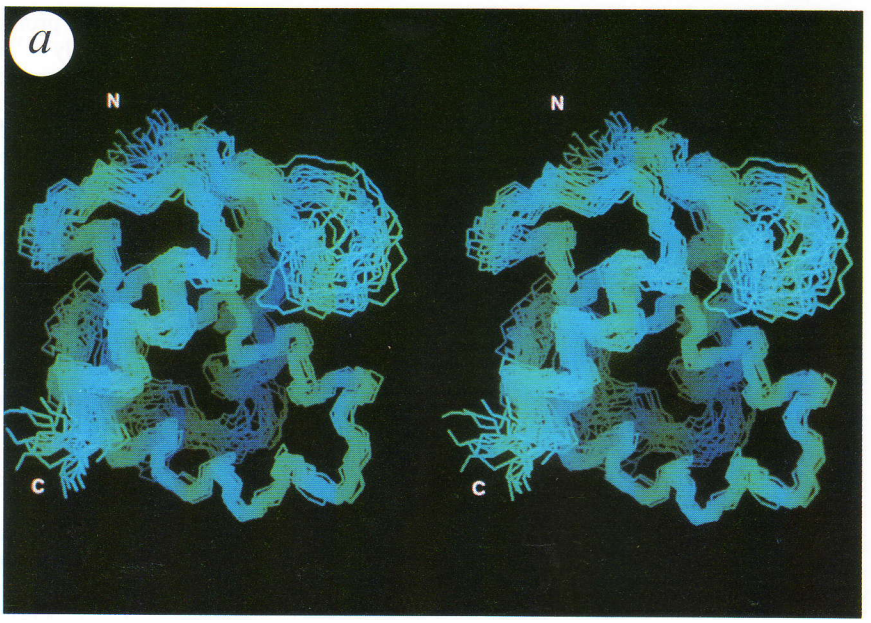
To produce sufficient protein for NMR analysis, p17 was expressed in *Escherichia coli* as a glutathione *S*-transferase fusion. The p17 coding sequence came from the HXB2 proviral clone. Following glutathione affinity purification, thrombin cleavage and ion-exchange chromatography, 20–30 mg of pure (>95%) protein was recovered from a 5-l fermentation. Uniformly ^{15}N-labelled and ^{15}N-leucine-labelled p17 gave sufficiently good spectra to allow a full backbone ^{1}H assignment using ^{15}N–^{1}H heteronuclear multiple quantum coherence nuclear Overhauser (HMQC-NOESY) and Hartmann–Hahn (HMQC-HOHAHA) spectroscopy[10,11]. The side-chain assignments were completed using ^{13}C–^{1}H HMQC-NOESY and ^{1}H–^{13}C–^{13}C–^{1}H-total correlated spectroscopy[12] (HCCH-TOCSY) on a uniformly labelled ^{13}C sample.

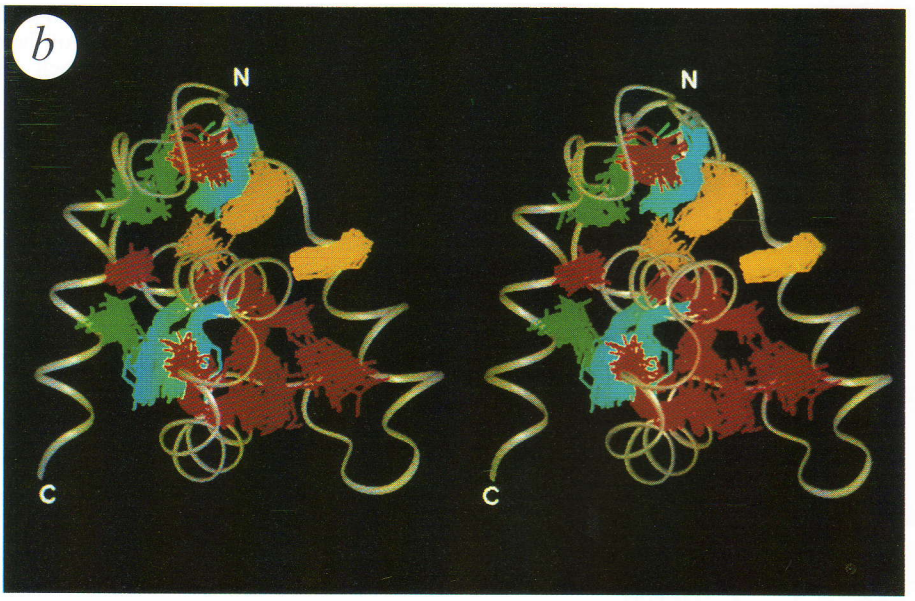
The structures were calculated on the basis of 1,067 NOE distance restraints, 75ϕ dihedral angle restraints and 26 hydrogen-bond distance restraints. The calculations were made within the programme XPLOR using a dynamical simulated annealing protocol[13,14]. Fifty structures were calculated, of which 26 contained no distance violations greater than 0.5 Å. The family of 26 structures is shown in Fig. 1*a* a superimposed on the mean coordinate position for the backbone atoms of residues 15–110. For these residues, the average root-mean-square deviation (r.m.s.d.) from the average structure is 1.34(±0.3) Å for the backbone atoms and 1.72(±0.4) for all atoms; for the regions with defined secondary structure these values are 0.92(±0.2) Å and 1.44(±0.3) Å, respectively.

The structure has a compact fold which contains four helices joined by short loops and a triple-stranded, irregular, mixed β-sheet. Helices B and C form an antiparallel coiled coil, which lies in the centre of the molecule. Helices A and D lie approximately parallel to each other on either side of the coiled coil. This arrangement gives rise to close packing of adjacent helices where the interhelical angles are 82°, 168° and 78° for AB, BC and CD, respectively. All the helices are accessible to the solvent and highly amphipathic, with the exception of helix C which runs through the centre of the hydrophobic core. Figure 1*b*, shows the radial distribution of the hydrophobic residues around this buried helix. Two regions (residues 19–23 and 26–29) towards the amino terminus, together with the residues between helices C and D (94–97), form three strands of an irregular, mixed β-sheet which constitutes a platform on 'top' of the structure. The lack of ^{1}H–^{1}H NOE data and a strong negative ^{15}N–^{1}H heteronuclear NOE[15] experienced by the N-amide resonances of the carboxy-terminal residues (data not shown) strongly suggest that the C-terminal 20 amino acids do not adopt a rigid conformation in solution. There is evidence for a turn centred about G10 and G11 (single-letter amino-acid code) within the N-terminal 14 amino acids, but it is not well defined with respect to the rest of the molecule.

An automated search of the Brookhaven protein structure database revealed similarity between p17 and the topology of the interferon-γ monomer[16–18] (Fig. 2). As Fig. 2*b* shows, there is essentially no sequence similarity between p17 and interferon-γ. Forty-one Cα atoms from the helical regions in p17 (A, B, C and D) superimpose on the equivalent regions (helices A, C, D and E) in interferon-γ with a r.m.s.d. of 2.6 Å. There are, however, two significant differences between the two structures. First, the β-sheet structure seen in p17 is not present in interferon-γ. Second, a short helical region exists within the long loop joining the major helices A and C of interferon-γ; in p17 this loop is significantly shorter and lacks any recognizable secondary structure. Interferon-γ is functional as a homodimer, in which there is unique interdigitizing of the helices across the subunit faces. The C termini of the monomers form helical arms which fit into a cleft between helices A and C of the adjacent monomer; they are not well structured[19,20]. It is conceivable that p17 dimerizes in an analogous way during viral assembly, as it also has a similar helical topology and contains a flexible C terminus. Extended arm- and helix-mediated interactions are common features in viral architecture[21]. The sequences of the simian immunodeficiency virus (SIV), HIV-1 and HIV-2 matrix proteins are highly homologous with each other. This strongly implies that any conclusions based on the structure of the HIV-1 matrix protein also hold true for all strains of HIV and SIV.

‖ To whom correspondence should be addressed.





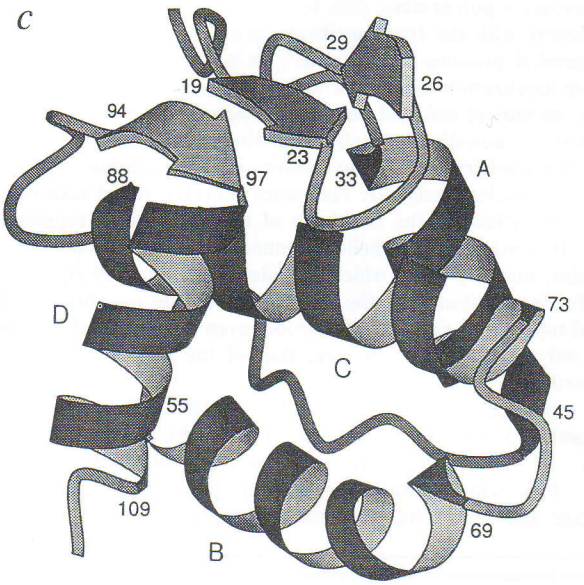


FIG. 1 *a*, Stereoview of backbone traces for the final 26 structures of HIV-1 p17 (residues 15–110) superimposed on the average coordinate position. *b*, Stereoview of the superposition of the principal hydrophobic side chains of the 26 structures. Leucine sidechains are shown in red, tyrosine in blue, isoleucine in green, and both valine and tryptophan are in yellow. The backbone of the average structure is represented by a white ribbon. *c*, Schematic representation of HIV-1 p17 showing the relative positions of the α-helices and β-sheet. The diagram was produced using the programme MOLSCRIPT[24].

METHODS. HIV-1 p17 was expressed using a pGEX-2T plasmid in *E. coli* JM103 cells. Labelled samples were obtained from growth of the *E. coli* on minimal media containing either $(^{15}NH_4)_2SO_4$ as the sole nitrogen source or U-^{13}C-glucose as the sole carbon source. The defined media for specifically labelling the leucine residues with ^{15}N have been described elsewhere[25]. All NMR spectra were recorded at 35 °C on 2 mM protein samples which contained 30 mM phosphate buffer at pH 6.0 and 15 mM NaCl. Three spectrometers were used, operating at proton frequencies of 500, 600 and 750 MHz. The structures were calculated on the basis of 1,067 NOE distance constraints, 75 ϕ dihedral angle restraints and 26 hydrogen-bond distance restraints. The NOE restraints were composed of 219 intra-residue, 304 sequential (residue i to residues $i+1$), 280 medium range (residue i to residues $1<i\leqslant 4$) and 264 long-range (residue i to residues $i>4$) connectivities. The distance restraints were categorized into three groups on the basis of estimated NOE cross-peak intensity: strong, 2.7 Å; medium, 3.8 Å; weak, 5.0 Å. Standard values for pseudo atom corrections[26] were added to the NOE distance constraints where appropriate. The ϕ restraints were derived from $^3J(NH_\alpha)$ by line-fitting traces from two-dimensional ^{15}N–^{1}H HMQC-*J* experiments[27]. Hydrogen-bonded NH groups were identified in helical regions by the presence of NH resonances 6 h after dissolving in D_2O. The final values for each term in the experimental terms in the force field used during simulated annealing were $F_{NOE}=31.5\pm9$ kcal mol^{-1}, $Fc_{dihed}=2.9\pm1$ kcal mol^{-1}, with force constants of 50 and 200 kcal mol^{-1} rad^{-2}, respectively. The average Lennard–Jones energy for the selected structures is -385.5 ± 20 kcal mol^{-1}, which is consistent with good non-bonded contacts. The r.m.s.d. from the experimental restraints and idealized geometry are as follows: NOE distances 0.023 ± 0.003 Å, dihedral restraints $0.62\pm0.1°$, bonds 0.0027 ± 0.0002 Å, angles $0.49\pm0.03°$ and improper angles $0.36\pm0.03°$. The coordinates will be deposited in the Brookhaven protein database.

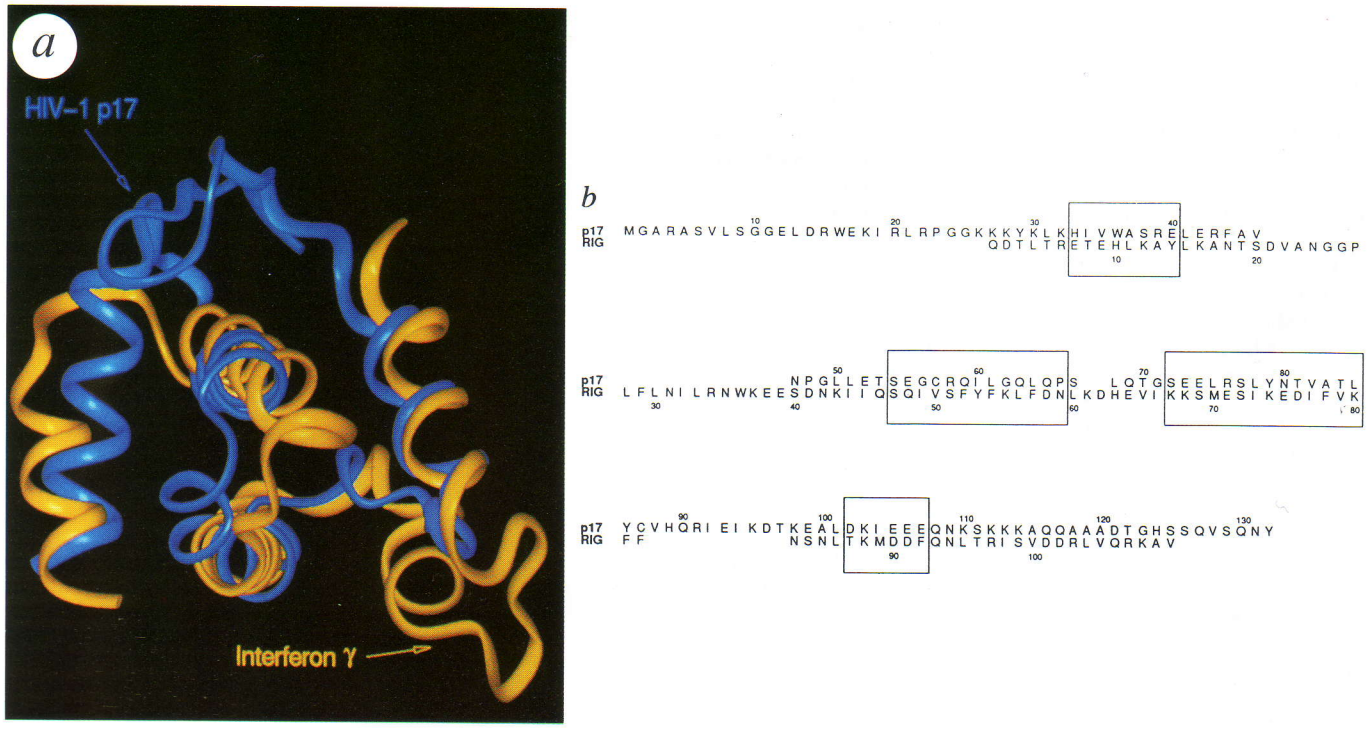


FIG. 2 *a*, A ribbon representation of the superimposition of the Cα atoms judged to be structurally equivalent within p17 and rabbit interferon-γ[16]. For clarity the C termini are not shown. *b*, The sequence alignment for p17 and rabbit interferon-γ. The 41 equivalent positions are indicated by boxes. These do not, of course, correspond exactly to the limits of the secondary structure regions shown in Fig. 1*c*.
METHODS. A database of 391 unique protein three-dimensional structural domains was searched by overlaying the sequences of p17 with every 5th residue in each database structure. Residues aligned by this procedure were used to derive an initial fit of p17 and the database structure. The initial fit was then refined using the protein-structure comparison algorithm described in ref. 28. The topology of p17 was judged using this method to be closest to that of the fold of interferon-γ. 41 equivalent Cα atoms were used to derive the superimposition shown, and gave a r.m.s.d. of 2.6 Å.

The irregular β-sheet is a prominent feature of p17. A glycine residue at the N terminus of p17 is myristoylated in the virus and many studies have established that this modification is essential for membrane targeting and virus assembly. It seems likely that this modified residue will be buried in the viral membrane. The solvent-exposed side of the β-sheet provides a surface which could associate with the inner face of the membrane, with several basic side chains (K18, R20, R22, K26–K28, K30, K32, K95) available for interaction with phospholipids headgroups. This is supported by mutagenesis studies in which mutations that alter the charge distribution have dramatic affects on virus assembly[6,7] and is consistent with the notion that the myristoylated glycine and the basic patch from residues 18 to 32 form a bipartite autonomous membrane-targeting signal[22]. This type of N-terminal basic region is common among the C-type retroviral matrix proteins. In HIV, the basic patch is also a major epitope for cytotoxic T lymphocytes[23] and a nuclear localization signal for the preintegration complex[9]. In addition, it is possible that the β-sheet interacts with gp41 to retain the viral envelope proteins within the particle[8].

We believe our study represents the first report of the structure of a retroviral matrix protein. The structural similarity of the HIV matrix to the immune modulator, interferon-γ, is intriguing and may provide a useful model for p17 dimerization as an initial step in virus assembly. It remains to be seen whether the p17/interferon-γ relationship is significant for the induction of immunodeficiency in HIV-infected individuals. □

Received 18 May; accepted 27 July 1994.

1. Spearman, P., Wang, J., Heyden, N. V. & Ratner, L. *J. Virol.* **68,** 3233–3242 (1994).
2. Gelderblom, H. R., Özel, M. & Pauli, G. *Archs Virol.* **106,** 1–13 (1989).
3. Nermut, M. V. *et al. Virology* **198,** 288–296 (1994).
4. Arnold, E. & Arnold, G. F. *Adv. Virus Res.* **39,** 1–58 (1991).
5. Bryant, M. & Ratner, L. *Proc. natn. Acad. Sci. U.S.A.* **87,** 523–527 (1989).
6. Yuan, X., Yu, X., Lee, T. & Essex, M. *J. Virol.* **67,** 6387–6394 (1993).
7. González, S. A., Affranchino, J. L., Gelderblom, H. R. & Burny, A. *Virology* **194,** 548–556 (1993).
8. Yu, X., Yuan, X., Matsuda, Z., Lee, T. & Essex, M. *J. Virol.* **66,** 4966–4971 (1992).
9. Bukrinsky, M. I. *et al. Nature* **365,** 666–669 (1993).
10. Marion, D. *et al. Biochemistry* **28,** 6150–6156 (1989).
11. Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T. & Gronenborn, A. M. *Biochemistry* **29,** 3542–3556 (1990).
12. Bax, A., Clore, G. M. & Gronenborn, A. M. *J. magn. Reson.* **88,** 425–431 (1990).
13. Nilges, M., Gronenborn, A. M., Brünger, A. T. & Clore, G. M. *Protein Engng* **2,** 27–38 (1988).
14. Brünger, A. T. *XPLOR Manual Ver 3.1* (Yale Univ. Press, New Haven, 1993).
15. Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. *Cell* **71,** 671–678 (1992).
16. Samudzi, C. T., Buston, L. E. & Rubin, J. R. *J. biol. Chem.* **266,** 21791–21797 (1991).
17. Ealick, S. E. *et al. Science* **252,** 698–702 (1991).
18. Farrer, M. A. & Schreiber, R. D. *A. Rev. Immun.* **11,** 571–611 (1993).
19. Lundell, D. *et al. Protein Engng* **4,** 335–341 (1991).
20. Grzesiek, S. *et al. Biochemistry* **31,** 8180–8190 (1992).
21. Stuart, D. *Curr. Opin. struct. Biol.* **3,** 167–174 (1993).
22. Zhou, W., Parent, L. J., Wills, J. W. & Resh, M. D. *J. Virol.* **68,** 2556–2569 (1994).
23. Johnson, R. P. *et al. J. Immun.* **147,** 1512–1521 (1991).
24. Kraulis, P. J. *J. appl. Crystallogr.* **24,** 946–950 (1991).
25. Matthews, S. J., Jandu, S. K. & Leatherbarrow, R. J. *Biochemistry* **32,** 657–662 (1993).
26. Wüthrich, K., Billeter, M. & Braun, W. *J. molec. Biol.* **169,** 949–961 (1983).
27. Kay, L. A. & Bax, A. *J. magn. Reson.* **86,** 110–126 (1990).
28. Russell, R. B. & Barton, G. J. *Proteins: Struct. Funct. Genet.* **14,** 309–323 (1992).

ACKNOWLEDGEMENTS. We thank N. Wood, G. Wingfield and S. Cole for help with *E. coli* fermentation, K. Jones and S. Champion for protein purification; and P. Cannon and I. Jones for discussion. This work is supported by the MRC AIDS Directed Programme. This is a contribution from the Oxford Centre for Molecular Sciences supported by the SERC and MRC.