Spectroscopic study of an HIV-1 capsid protein major homology region peptide analog

Clary B. Clish[a], David H. Peyton[a,*], Eric Barklis[b]

[a]*Department of Chemistry, Portland State University, PO Box 751, Portland, OR 97207–0751, USA*
[b]*Vollum Institute for Advanced Biomedical Research and Department of Microbiology and Immunology, Oregon Health Sciences University, Portland, OR 97201, USA*

Received 17 November 1995

Abstract The capsid (CA) domain of retroviral Gag proteins posseses one subdomain, the major homology region (MHR), which is conserved among nearly all avian and mammalian retroviruses. While it is known that the mutagenesis of residues in the MHR will impair virus infectivity, the precise structure and function of the MHR is not known. In order to obtain further information on the MHR, we have examined the structure of a synthetic peptide encompassing the MHR of human immunodeficiency virus type I (HIV-1) CA protein. Multiple sequence alignment and secondary structure prediction indicate that the peptide could form 50% α-helix and 10% β-sheet. In addition, circular dichroism studies indicate that, in the presence of 50% trifluoroethanol (TFE), the peptide adopts an α-helical structure over half of its length. Further analysis by proton nuclear magnetic resonance spectroscopy suggests that the C-terminal portion of the MHR forms a helix in aqueous solution. Upon the addition of TFE, the position of the helix remains nearly constant, but the magnitude of the changes in H_α chemical shifts of the residues indicate a more stable helix. These results suggest that a helical C-terminus of retroviral MHRs may be integral to the function of this region.

Key words: CD; NMR; Human immunodeficiency virus type 1; Secondary structure; MHR peptide

1. Introduction

The structural proteins of human immunodeficiency virus type 1 (HIV-1) are translated as a polyprotein precursor ($Pr55^{gag}$) [1,2] from the *gag* gene mRNA, which also is the full-length HIV-1 viral RNA. During or after virus assembly, virus protease-catalyzed processing of $Pr55^{gag}$ yields the viral proteins matrix (MA), capsid (CA) and nucleocapsid (NC), and p6 [3]. Of these proteins, structures have been provided for MA [4,5], which interacts with the viral membrane [6] and the viral envelope protein [7,8]; and for the NC protein [9,10], which helps encapsulate the viral RNA during virus assembly [11]. To date, no structure determination has been reported for the HIV-1 p6 protein, which facilitates particle budding [12]; or for CA, which is required for the formation of the capsid in immature and mature virus particles.

Interestingly, retrovirus CA primary sequences are poorly conserved, with the exception of a sequence termed the major homology region (MHR) (Fig. 1) [1]. The MHR exists, invariant in its location, in the capsid proteins of all avian and mammalian retroviruses for which the *gag* sequence is known, with the exception of spumaretroviruses [13]. Given the sequence conservation of the MHR, it seems likely that this CA subdomain plays a conserved role in virus replication, either via an interaction with cellular factors or another viral component. However, molecular genetic studies have only hinted at the functional significance of the MHR. Some mutations, for instance, alteration of MHR residues glutamine-155 and glutamate-159, blocked virus particle assembly [13]. Other mutations, including a deletion of HIV-1 MHR residues, blocked virus replication and reduced virus assembly [13,14]. However, mutations, such as an arginine to alanine substation at residue-167, while blocking virus replication, had no apparent effect on the efficiency of virus particle assembly [13]. Similarly, cysteine substitutions for C-terminal MHR residues permit efficient virus particle assembly, RNA encapsulation and $Pr55^{gag}$ protein processing, but block virus infectivity (McDermott, Farrell, Zhang and Barklis, unpubl. obs.). Thus, the MHR may effect an early stage of virus particle assembly, a later post-assembly stage of the virus life cycle, or multiple stages.

In order to help elucidate the role of retroviral MHRs and capsid proteins, we wished to determine the secondary structure of the HIV-1 MHR. To date, the only information regarding the MHR structure has been based on a sequence comparison of retroviral CA and picornaviral VP2 proteins, suggesting β-strand stretching from tyrosine-164 to threonine-171 [15]. In contrast, we present here spectroscopic evidence for the existence of an α-helix that extends from phenylalanine-161 to arginine-176 in an HIV-1 CA protein MHR peptide analog. Our results suggest that the C-terminal portion of the MHR may form an amphipathic helix which may interact with cellular and/or viral components as an essential step in the HIV life cycle.

2. Materials and methods

2.1. Peptide synthesis and purification

The 31-residue MHR containing peptide (Fig. 1) was obtained from a commercial supplier (Peptide Express, Colorado State University, Fort Collins, CO). The peptide was synthesized with a cysteine residue located at the N-terminus for use in complimentary studies. The crude peptide was purified by preparative reverse-phase HPLC and its identity confirmed by mass spectrometry.

2.2. Secondary structure prediction

Secondary structure predictions according to Rost and Sander [16–18] were obtained interactively from the PredictProtein server at EMBL (http://www.embl-heidlberg.de/predictprotein/phd_pred.html, on the World Wide Web).

2.3. Circular dichroism spectroscopy

Aliquots of an aqueous 2 mg/ml stock peptide solution were diluted with appropriate amounts of water and TFE to prepare samples con-

*Corresponding author. Fax: (1) (503) 725 3888.

SSDI 0014-5793(95)01419-5

taining 1.0 mg/ml peptide and 0% or 50% TFE (wt/wt). The CD spectrum for each sample was measured over a wavelength range of 180–260 nm on a JASCO J-500 CD Spectrometer (Easton, MD) using 4 scans per sample with a scan rate of 10 nm/min at 25°C. The resulting spectra were analysed for secondary structure using the program VARSLC1 [19,20] in which the basis set included 33 proteins.

2.4. Nuclear magnetic resonance spectroscopy

NMR sample concentrations were 5–7 mM peptide and 0% or 50% TFE-d_2 (ISOTEC, Miamisburg, OH) (wt/wt) in aqueous solutions containing 10% D_2O. The pH was adjusted to 3.1 to minimize H_N peak broadening due to exchange with the solvent [21]. All spectra were referenced to the solvent (H_2O) peak at 4.76 ppm. A crystal of dithionite was added to each sample to prohibit disulfide bond formation between peptide molecules.

NMR spectra were acquired at 298 K on a Bruker AMX-400 spectrometer, using a dedicated 5-mm proton probe, locked to the deuterium resonance of the solvent and without spinning. In order to assign the proton chemical shifts, nuclear Overhauser spectroscopy (NOESY) [22], double-quantum-filtered J-correlated spectroscopy (DQF-COSY) [23] and total correlation J-spectroscopy (TOCSY) [24] spectra were recorded in the phase-sensitive mode by the time proportional phase incrementation method (TPPI) [25]. TOCSY spectra were recorded at mixing times of 53 and 90 ms to aid the assignment of the longer side-chain resonances. NOESY spectra were acquired at mixing times of 100, 150 and 300 ms, with 512 t_1 values, and 2048 data points in t_2. Data were processed on an SGI Indy (Silicon Graphics Computer Systems, Mountain View, CA) workstation using FELIX (Biosym Technologies, San Diego, CA). Solvent supression was achieved by presaturation during the relaxation delay. 2D spectra were zero-filled in both dimensions to give 1024 × 1024 TOCSY and COSY real matrices and 2048 × 2048 NOESY real matrices and baseline corrections were applied as needed. Spectra were assigned much as in Wüthrich [21].

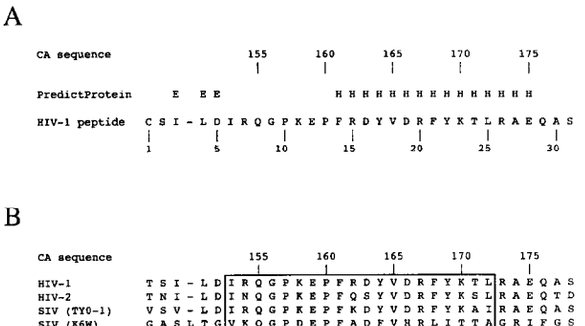

Fig. 1. Primary structure of the MHR. (A) Sequence of the studied peptide analog of the CA MHR (HIV-1 peptide) and results of PredictProtein secondary structure prediction of the HIV-1 peptide. The peptide numbering scheme is indicated below the sequence, while the corresponding CA sequence numbering is shown above. (H) α-helix; (E) β-strand. (B) Alignment of CA sequences of different retroviruses in the major homology region (MHR). The absolutely conserved residues occur at positions 155, 159 and 167, while an aromatic residue consistently occurs at position 164. HIV-2, human immunodeficiency virus type 1; SIV (TYO-1) and SIV (K6W), simian immunodeficiency virus isolates TYO-1 and K6W; MPMV, Mason-Pfizer monkey virus; EIAV, equine infectious anemia virus; HTLV-I and HTLV-II, human T-cell leukemia virus types I and II; Mo-MLV, Molony murine leukemia virus; FLV and BLV, feline and bovine leukemia viruses; MMTV, mouse mammary tumor virus; RSV, Rous sarcoma virus.

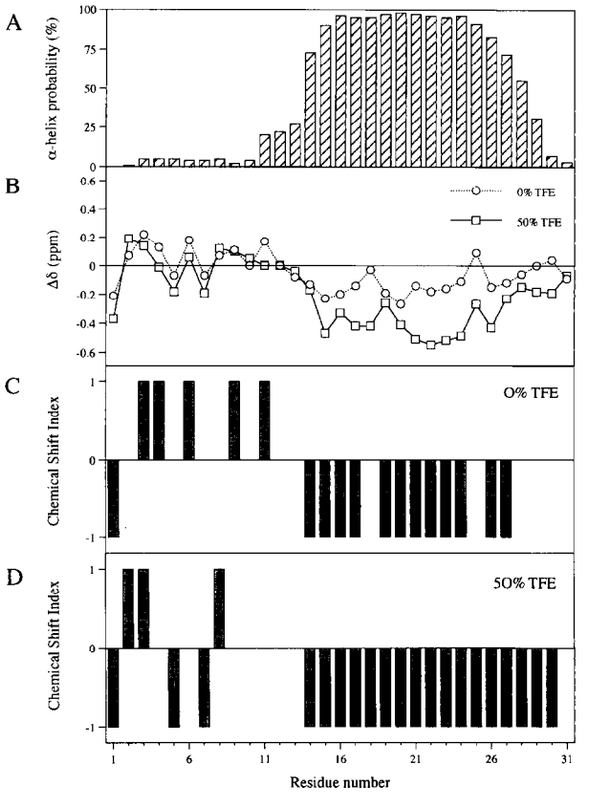

Fig. 2. Secondary structure of the MHR by various methods. (A) Results from PredictProtein secondary structure prediction indicating the probability of each residue being part of an α-helix. (B) Difference between random coil [29] and observed H_α chemical shifts of the HIV-1 MHR-containing peptide in 0% TFE (circles) and 50% TFE (boxes). (C) Chemical shift index of α-protons of the peptide in 0% TFE. (D) Chemical shift index of α-protons of the peptide in 50% TFE.

3. Results and discussion

The sequence of the peptide as shown in Fig. 1 was submitted to the PredictProtein server for multiple sequence alignment and secondary structure prediction. According to the prediction, residues isoleucine-3 through aspartate-5 might adopt a β-strand structure while residues phenylalanine-14 through glutamate-28 would form an α-helix (Fig. 1). The percentage probability of each residue being in an α-helix, as determined by the prediction, is shown in Fig. 2A. This result is in disagreement with prior predictions as to the tertiary structure of HIV-1 CA via sequence alignment with picornaviral proteins [15], in which it was suggested that residues tyrosine-17 through threonine-24 of the peptide analog form a β-strand in the CA protein.

We also examined the CD spectrum of the MHR peptide in aqueous and TFE containing solutions. As indicated by the CD spectrum in Fig. 3, the peptide adopts very little secondary structure in aqueous solution. However, in aqueous solution containing 50% TFE an intense positive band at 193 nm and strong negative bands at 208 and 218–222 nm were observed, indicating the presence of α-helix. TFE allows for the stabilization of intramolecular hydrogen bonds, such as those between amide protons and carbonyl oxygens in α-helices. Hydrogen bonding to the solvent is reduced because TFE is a weaker base

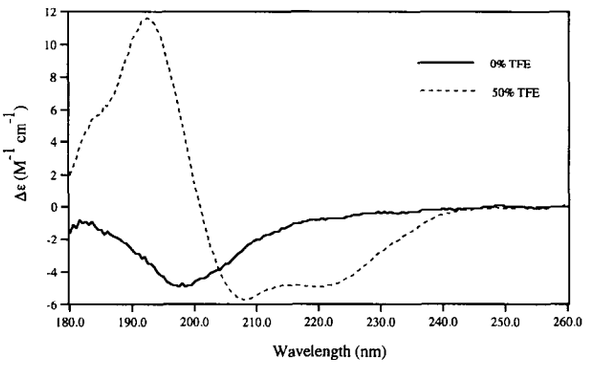

Fig. 3. CD spectra of the HIV-1 MHR containing peptide. The solid line is for aqueous solution only and the dotted line is for aqueous solution containing 50% TFE.

than water and as a consequence intramolecular hydrogen bonds are favored [26]. The relative amounts of secondary structure were estimated using the program VARSLC1 [19,20]. By this calculation, the helicity increased to 49% in 50% TFE from 5% in the absence of TFE. The absence of substantial helicity at 0% TFE could be due to a distortion of the ideal α-helical geometry caused by motion at the ends of the peptide, leading to a diminished CD signal [27]. The calculated amounts of β-sheet were 19% and 10% for the 0% and 50% TFE solutions, respectively. However, β-sheet formation is unlikely in such a short peptide and so CD signals corresponding to β-sheet-like conformations may have been due to chance orientation of amide bonds.

To test our interpretation of the HIV-1 MHR peptide CD spectra, the peptide was subjected to analysis by NMR. ^{1}H NMR assignments were carried out using the general strategy of Wüthrich [21]. Essentially complete amino acid spin systems were identified from the 2D TOCSY and DQF-COSY spectra. A representative plot of the fingerprint region of the DQF-COSY spectrum is shown in Fig. 4. The cross-peaks corresponding to $^3J_{HN\alpha}$ couplings were identified for all residues except glutamate-12, where we believe the major complications to finding the resonances from this residue arise from spectral overlap with signals from other protons. We note, however, that this residue is not in the region identified as helical below. Sequence specific assignments were then determined from short-range NOEs between H_N, H_α and H_β of residue i and H_N of residue $i + 1$. The assignments of the peptide in 50% TFE are summarized in Table 1, with the H_α resonances of the peptide in 0% TFE also indicated.

H_α chemical shifts have been shown to be dependent on the type of secondary structure of which the residue is a part [28]. Protons with upfield shifts relative to random coil values tend to occur in helices while those with downfield shifts occur in β-structures. An empirical chemical shift index (CSI) technique for fast approximations of secondary structure has been devel-

Table 1
Proton chemical shifts of HIV-1 peptide in 50% TFE[a]

No.	Residue	NH	Hα	Hβ	Hγ	Other
1	Cys[b]		4.28	3.03, 3.11		
2	Ser	8.82	4.69 (4.57)[c]	4.02, 3.85		
3	Ile	8.33	4.09 (4.17)	1.86	1.25, 1.46	γCH_3 0.91 δCH_3 0.86
4	Leu	7.71	4.16 (4.30)	1.59, 1.59	1.57	δCH_3 0.81, 0.88
5	Asp	7.85	4.58 (4.69)	2.85, 2.85		
6	Ile	7.60	4.01 (4.13)	1.97	1.21, 1.50	γCH_3 0.90 δCH_3 0.83
7	Arg	8.05	4.19 (4.31)	1.83, 1.88	1.65, 1.72	δCH 3.15, 3.15 εNH 7.28
8	Gln	7.96	4.41 (4.36)	2.02, 2.02	2.18, 2.38	
9	Gly	7.90	4.07, 4.07 (4.08, 4.08)			
10	Pro		4.49 (4.44)	1.93, 2.30	2.01, 2.01	δCH 3.60, 3.60
11	Lys	8.41	4.36 (4.53)	1.79, 1.79	1.48, 1.48	δCH 1.69, 1.69 εCH 3.01, 3.01
12	Glu[d]					
13	Pro		4.40 (4.36)	1.80, 2.26	1.97, 1.97	δCH 3.56, 3.69
14	Phe	7.50	4.49 (4.53)	3.15, 3.15		δCH 7.15 εCH 7.25
15	Arg	8.05	3.91 (4.15)	1.87, 1.87	1.64, 1.71	δCH 3.20, 3.20 εNH 7.21
16	Asp	8.29	4.43 (4.56)	2.77, 2.77		
17	Tyr	7.77	4.18 (4.46)	3.12, 3.12		δCH 7.01 εCH 6.73
18	Val	8.09	3.53 (3.92)	2.00	0.83, 0.94	
19	Asp	8.25	4.50 (4.57)	2.81, 2.96		
20	Arg	7.83	3.97 (4.11)	1.85, 1.85	1.51, 1.65	δCH 3.10, 3.10 εNH 7.83
21	Phe	8.41	4.15 (4.52)	2.89, 3.14		δCH 6.94 εCH 7.17
22	Tyr	8.54	4.05 (4.42)	3.10, 3.10		δCH 7.13 εCH 6.79
23	Lys	8.24	3.84 (4.20)	1.86, 1.92	1.41, 1.41	δCH 1.65, 1.65 εCH 2.95, 2.95
24	Thr	7.87	3.86 (4.24)	4.25	1.15	
25	Leu	7.98	3.90 (4.26)	1.43, 1.43	1.43	δCH 0.70, 0.70
26	Arg	8.05	3.95 (4.23)	1.74, 1.74	1.44, 1.53	δCH 3.05, 3.05 εCH 7.12
27	Ala	7.82	4.12 (4.23)	1.48		
28	Glu	8.05	4.14 (4.23)	2.15, 2.15	2.46, 2.58	
29	Gln	7.94	4.19 (4.37)	2.06, 2.14	2.39, 2.47	
30	Ala	7.80	4.16 (4.39)	1.41		
31	Ser	7.78	4.43 (4.41)	3.94, 3.94		

[a]The chemical shifts are given in ppm and are referenced to H_2O (4.76 ppm at 298 K).
[b]NH resonance not found due to exchange with solvent.
[c]H_α values in parentheses are for 0% TFE as the solvent.
[d]Glu 12 resonances not identified due to extensive overlap.

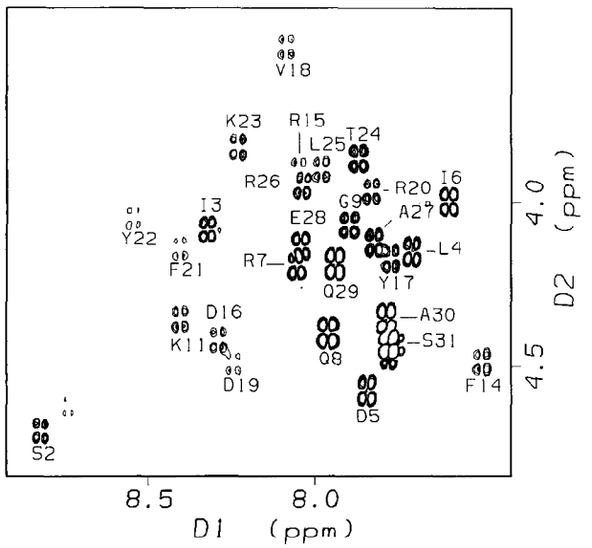

Fig. 4. Part of a DQF-COSY spectrum of the MHR peptide. The peptide was 5 mM, in aqueous 50% TFE (by weight) solution, at pH 3.1 and 298K. Cross-peaks for NH to H_α correlations are labeled for all residues except those of cysteine-1, glutamate-12, proline-10 and proline-13.

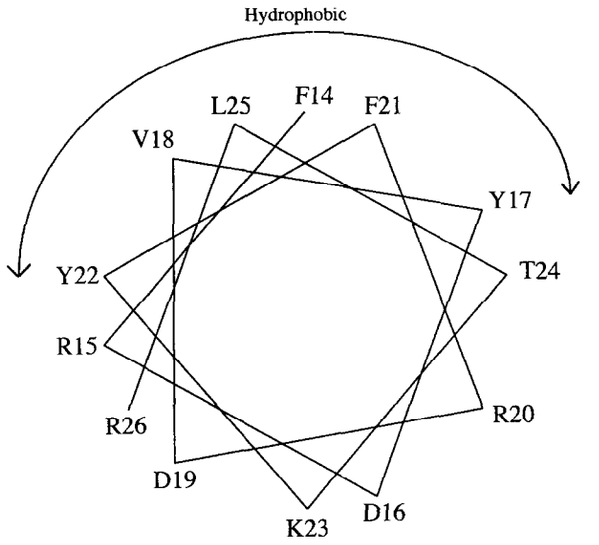

Fig. 6. Helix-wheel model of the amphipathic helix formed by residues 14 through 26 of the HIV-1 MHR peptide. Residues F14, Y17, V18, F21, Y22 and L25 form a hydrophobic face while residues R15, D16, D19, R20, K23 and R26 occupy a hydrophilic side of the helix.

oped in which each H_α is assigned a value of 1 if it has been downfield shifted by >0.1 ppm, a value of 0 if the shift is ±0.1 ppm or a value of −1 if it has been upfield shifted by >0.1 ppm [29]. Four consecutive H_α values with values of −1 are in α-helices while three consecutive values of 1 constitute β-structures.

The chemical shift differences between the H_αs of the peptide in both 0% and 50% TFE and random coil values are shown in Fig. 2B. As shown, the H_α chemical shift differences are small on the N-terminal part of the peptide and are similar for both concentrations of TFE, suggesting that this portion of the peptide is disordered. In contrast, upfield chemical shift differences are evident in the C-terminal portion of the peptide, from phenylalanine-14 to the C-terminus in the absense of TFE. Furthermore, the magnitude of the chemical shift differences increased significatly when in 50% TFE solution. Chemical shift index plots of the peptide in both TFE concentrations are shown in Fig. 2C,D. This analysis is indicative of an α-helix extending between residues 14–27 in 0% TFE and between residues 14–30 in 50% TFE. The magnitude of the chemical shift difference increased when TFE is present in solution in the α-helical portion of the peptide, suggesting that a more stable helix forms.

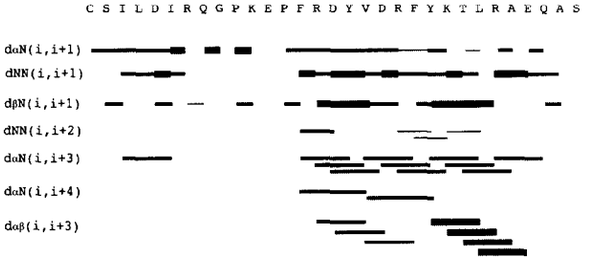

Fig. 5. Summary of the observed short- and medium-range NOEs for the MHR containing peptide in 50% TFE. The thick, medium and thin lines correspond to strong, medium and weak NOEs, respectively.

The chemical index data suggests that the peptide is 52% α-helix in 50% TFE, but gives no evidence for the presence of β-strand, all of which is in agreement with our CD and PredictProtein analyses.

The presence of medium-range NOEs, such as $d_{NN}(i,i+2)$, $d_{\alpha N}(i,i+3)$, $d_{\alpha N}(i,i+4)$ and $d_{\alpha\beta}(i,i+3)$, are further evidence of helical regions [12]. In addition to sequential NOEs of the types $d_{\alpha N}(i,i+1)$, $d_{NN}(i,i+1)$ and $d_{\beta N}(i,i+1)$, observed medium-range NOEs are summarized in Fig. 5. Spectral overlap prohibited the observation of several NOEs, nevertheless the pattern of NOEs shown in Fig. 5 serves as strong evidence for the presence of an α-helix that stretches from about residue 14 to residue 29.

Our NMR data is, therefore, in agreement with the CD results, suggesting that TFE stabilizes MHR helix formation. Furthermore, our assignment of the C-terminal portion of the MHR peptide as helical based on NMR chemical shift differences is in remarkable agreement with PredictProtein secondary structure predictions. While we do not yet have a complete structure determination for the HIV-1 MHR, it is noteworthy that peptide residues 14–26 can be modeled to form an amphipathic α-helix. By such a model (Fig. 6), peptide residues phenylalanine-14, tyrosine-17, valine-18, phenylalanine-21, tyrosine-22 and leucine-25 would form a hydrophobic helical face, while residues arginine-15, aspartate-16, aspartate-19, arginine-20, lysine-23 and arginine-26 would occupy the polar side of the helix. It is intriguing to speculate that such an amphipathic helix on one Gag protein monomer may interact with another helix on an adjacent monomer. Alternatively, an amphipathic MHR helix may bind to other viral components or cellular factors to facilitate virus assembly and/or infection. Such possibilities currently are under investigation.

Acknowledgements: We wish to thank H.P. Bachinger for his help in acquiring the CD spectra and W.C. Johnson for providing us with a

copy of his VARSLC1 software. Essential initial support for this work was provided by an American Foundation for Aids Research (AmFAR) grant to E.B.

References

[1] Wills, J.W. and Craven, R.C. (1991) AIDS 5, 639–654.
[2] Hunter, E. (1994) Semin. Virol. 5, 71–83.
[3] Mervis, R.J. Ahmand, N., Lillehoj, E.P., Raum, M.G., Salazar, F.H.R., Chan, H.W. and Venkatesan (1988) J. Virol. 62, 3993–4002.
[4] Massiah, M.A., Starich, M.R., Paschall, C., Summers, M.F., Christensen, A.M. and Sundquist, W.I. (1994) J. Mol. Biol. 244, 198–223.
[5] Mathews, S., Barlow, P., Boyd, J., Barton, G., Russel, R., Mills, H., Cunningham, M., Meyers, N., Burns, N., Clark, N., Kingsman, S., Kingsman, A. and Campbell, I. (1994) Nature (London) 370, 666–668.
[6] Bryant, M. and Ratner, L. (1990) Proc. Nat. Acad. Sci. USA 87, 523–527.
[7] Fäcke, M., Janetzko, A., Shoeman, R.L. and Kräusslich, H.-G. (1993) J. Virol. 67, 4972–4980.
[8] Wang, C.-T., Zhang, Y., McDermott, J. and Barklis, E. (1993) J. Virol. 67, 7067–7076.
[9] Summers, M.F., Henderson, L.E., Chance, M.R., Bess, J.W. Jr., South, T.L., Blake, P.R., Sagi, I., Perez-Alvarado, G., Sowder, R.C., Hare, D.R. and Arthur, L.O. (1992) Protein Sci. 1, 563–74.
[10] Delahunty, M.D., South, T.L., Summers, M.F. and Karpel, R.L. (1992) Biochemistry 31, 6461–6469.
[11] Aldovini, A. and Young, R.A. (1990) J. Virol. 64, 1920–1926.
[12] Göttlinger, H.G., Sodroski, J.G. and Haseltine, W.A. (1989) Proc. Nat. Acad. Sci. USA 86, 5781–5785.
[13] Mammano, F., Öhagen., A., Höglund, S. and Göttlinger, H.G. (1994) J. Virol. 68, 4927–4936.
[14] Dorfman, T., Bukovsky, A., Öhagen., A., Höglund, S. and Göttlinger, H.G. (1994) J. Virol. 68, 8180–8187.
[15] Argos, P. (1989) EMBO J. 8, 779–785.
[16] Rost, B. and Sander, C. (1993) J. Mol. Biol. 232, 584–599.
[17] Rost, B. and Sander, C. (1994) Proteins 19, 55–77.
[18] Sander, C. and Schneider, R. (1991) Proteins 9, 56–68.
[19] Compton, L.A. and Johnson, W.C. Jr. (1986) Anal. Biochem. 155, 155–167.
[20] Manavalan, P. and Johnson, W.C. Jr. (1987) Anal. Biochem. 167, 76–85.
[21] Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids. Wiley, New York, NY.
[22] Kumar, A., Ernst, R.R. and Wüthrich, K (1980) Biochem. Biophys. Res. Commun. 95, 1–6.
[23] Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. (1983) Biochem. Biophys. Res. Commun. 117, 479–485.
[24] Braunschweiler, L. and Ernst, R.R. (1983) J. Magn. Reson. 53, 521–528.
[25] Marion, D. and Wüthrich, K. (1983) Biochem. Biophys. Res. Commun. 113, 976–974.
[26] Sönnichsen, F.D., Van Eyk, J.E., Hodges, R.S. and Sykes, B.D. (1992) Biochemistry 31, 8790–8798.
[27] Dyson, H.J. and Wright, P.E. (1991) Annu. Rev. Biophys. Chem. 20, 519–538.
[28] Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) J. Mol. Biol. 222, 311–333.
[29] Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) Biochemistry 31, 1647–1651.