Journal of Peptide Science
J. Peptide Sci. **11**: 481–490 (2005)
Published online 10 January 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.641

JOURNAL OF PEPTIDE SCIENCE

Influenza A virus protein PB1-F2: synthesis and characterization of the biologically active full length protein and related peptides

PETER HENKLEIN,[a]* KARSTEN BRUNS,[b,c] MANFRED NIMTZ,[b] VICTOR WRAY,[b] UWE TESSMER[c] and ULRICH SCHUBERT[c,d,e]

[a] Humboldt University, Institute of Biochemistry, Berlin, Germany
[b] German Research Centre for Biotechnology, Department of Structural Biology, Braunschweig, Germany
[c] University of Hamburg, Heinrich-Pette-Institute, Germany
[d] Laboratory of Viral Diseases, NIAID, National Institutes of Health, Bethesda, MD 20892, USA
[e] University of Erlangen-Nürnberg, Institute of Clinical and Molecular Virology, Germany

Received 12 August 2004; *Revised 18 October 2004*; *Accepted 25 October 2004*

Abstract: Recently the discovery of a novel 87 amino acid influenza A virus (IAV) protein, named PB1-F2, has been reported that originates from an alternative reading frame in the PB1 polymerase gene and is encoded in most of the known human IAV isolates. Using optimized protocols, full length biologically active *s*PB1-F2 and a number of fragments have been synthesized by following either the standard elongation SPPS method or by native chemical ligation of unprotected *N*- and *C*-terminal peptide fragments at the histidine and cysteine residues located in position 41 and 42 of the native sequence, respectively. The ligation procedure afforded the most efficient synthesis of *s*PB1-F2 and facilitated the generation of various mutants of *s*PB1-F2 from pre-synthesized peptide fragments. During the synthesis of *s*PB1-F2, the formation of succinimide and subsequent conversion to the piperidine derivative at the aspartic acid residue in position 23 was observed. This reaction was forestalled by applying specific modifications to the SPPS protocol. The chain-elongation SPPS protocol is optimal for producing small peptides of *s*PB1-F2, their derivatives and precursors for a subsequent ligation protocol, while the full length protein, mutants and labelled derivatives are more conveniently and efficiently synthesized by SPPS protocols that include native chemical ligation. The molecular identity of *s*PB1-F2 was confirmed by peptide mapping, mass spectrometry, *N*-terminal sequencing, ^{1}H NMR spectroscopy and Western blot analysis. The latter analysis afforded direct evidence of the inherent tendency of *s*PB1-F2 to undergo oligomerization, a phenomenon observed both for full length *s*PB1-F2 and fragments thereof, as well as for its full length viral counterpart. Our synthesis protocols open the field for multiple biological and structural studies on *s*PB1-F2 that, similar to the molecule expressed in an IAV context, induces apoptosis and interacts with membranes *in vitro* and *in vivo*, as shown in previous studies.

Keywords: influenza A virus protein PB1-F2; protein synthesis; piperidide-formation; chemical ligation; MALDI-TOF mass spectrometry; TOCSY nuclear magnetic resonance spectroscopy

INTRODUCTION

Influenza A virus (IAV) represents one of the foremost pathogens threatening humans and animals. At least three major pandemics occurred during the last century, the most severe outbreak amongst the human population known as 'Spanish flu' was observed around 1918 with 20 to 40 million casualties (reviews [1,2]). As a zoonotic disease IAV naturally infects both avian and mammalian hosts, with aquatic birds being considered as the primary reservoir, and novel viruses sporadically enter humans from the animal reservoir in a process that usually involves genetic rearrangements between human and animal viruses [3,4]. According to IAV genetic data unravelled almost two decades ago, the virus genome consists of eight negative-strand RNA gene segments and was predicted to encode ten proteins [5]. Recently, an eleventh IAV gene product, named PB1-F2, was discovered and described as an 87 residue protein encoded by an alternative +1 reading frame within the coding sequence for the IAV viral RNA polymerase subunit PB1 [6]. While this newly described IAV protein was originally discovered for the IAV strain A/Puerto Rico/8/34/Mount Sinai(H1N1), also termed PR8, the PB1-F2 open reading frame was identified in 64 of 84 analysed IAV variants but not in influenza B viruses [2,6,7].

Compared with other IAV gene products, the PB1-F2 regulatory protein possesses several distinctive features. First, the PB1-F2 ORF is not present in all IAV isolates. While it is encoded by nearly all IAV strains isolated from humans, it is absent from a number of animal isolates, particularly those obtained from swine. Second, its levels of expression vary widely among infected cells in a manner unrelated to the levels of expression of other viral proteins. Third, its intracellular localization also varies considerably

* Correspondence to: Dr Peter Henklein, Humboldt University, Institute of Biochemistry, Monbijoustrasse 2, D-10117 Berlin, Germany; e-mail: peter.henklein@charite.de

between cells: in some cells it is almost exclusively mitochondrial, in others almost exclusively nuclear, with transitional arrangements recurrently observed. Fourth, it is degraded rapidly, at least in part by the ubiquitin–proteasome pathway (UPS) inasmuch as highly specific proteasome inhibitors increase its half life, although evidence for ubiquitinylation of PB1-F2 has not been reported yet. Fifth, the primary translation products migrate in three bands, all within a few kDa of the predicted M_r of 11 kDa, indicative of post-translational modification such as phosphorylation. Notably, when PB1-F2 is expressed by recombinant vaccinia virus (rVV) or CMV driven transgenes, the protein migrates with a single mobility. It therefore appears that other IAV proteins and/or cellular factors directly or indirectly influence post-translational modification, as well as potentially the transport and function of PB1-F2.

As is typical for a regulatory protein with an accessory function in the viral life cycle, data accumulated so far indicate that PB1-F2 is not required for IAV replication in tissue culture or the allantoic cavity of embryonated eggs. Further, it has not been formally demonstrated that the molecule is part of infectious IAV particles. However, there is mounting evidence that PB1-F2 represents another example of the growing list of pro-apoptotic proteins, such as HIV-1 Vpr or the HTLV-I p13 proteins [6]. The interest and importance of analysing PB1-F2 at the molecular level is enhanced by its similarity, in adopting mitochondrial and nuclear localization and exhibiting pro-apoptotic functionality, to two retroviral proteins, the HTLV type 1 p13 protein [8,9] and HIV-1 Vpr [10,11]. Although the molecular mechanism involved in the pro-apoptotic function of PB1-F2 is still unknown, a first insight into structure–function correlation of PB1-F2 domains stems from recent mutagenesis studies using GFP-PB1-F2 fusion proteins that mapped the inner mitochondrial membrane localization signal to a putative amphipathic helix located near the *C*-terminus of PB1-F2 [12].

As a small protein that can be completely synthesized as a functional entity, *s*PB1-F2 is an attractive subject for detailed structural–functional analysis. In previous investigations of *s*PB1-F2 it was shown to exhibit various biological phenomena that were also observed for its virally expressed counterpart [6]: following microinjection it localizes to mitochondria where it induces morphological alterations and causes cell death. It also induces transmembrane conductance of planar lipid bilayers and exhibits a behaviour similar to other pro-apoptotic proteins [13].

Here, as a prerequisite for further structural–functional studies, protocols have been established for the production of significant amounts of biologically active *s*PB1-F2 and related peptides on a routine basis that overcome the difficulties encountered in the production of recombinant material caused by the toxicity and the high tendency of PB1-F2 for self association and membrane interaction. The molecular identities of the synthetic products have been characterized in detail by peptide mapping, mass spectrometry, *N*-terminal sequencing, ^{1}H NMR spectroscopy and Western blot analysis. Our chemical ligation strategy of pre-synthesized peptide fragments is particularly attractive as it entails the use of a natural cysteine residue in the centre of the molecule that allows ready synthesis of both mutant molecules and various types of derivatives that provide material for further biochemical, structural and functional studies. Subsequently, the molecular identities of the synthetic products have been characterized in detail.

MATERIALS AND METHODS

Synthesis and Purification of Full Length sPB1-F2 and Related Fragment Molecules

Step by step synthesis. The sequence of *s*PB1-F2, described previously [6], is derived from influenza virus A/Puerto Rico/8/34/Mount Sinai(H1N1) strain (PR8). Initially a slightly modified solid-phase peptide synthesis (SPPS) protocol was used that had been optimized for the synthesis of the HIV-1 regulator protein *s*Vpr$^{1-96}$ [14]. The SPPS method was performed on an ABI 433A automated peptide synthesizer on a 0.1 mM scale with 250 mg TentaGel R-Trt-Ser(tBu) Fmoc-resin (capacity 0.19 mmol/g) using the Fmoc/tBu strategy. The following side-chain protecting groups were used: 2,2,4,6,7-pentamethyldihydrobenzofurane-5-sulfonyl (Arg), t-butoxycarbonyl (Trp, Lys), t-butyl ether (Thr, Ser, Tyr), t-butyl ester (Asp, Glu) and trityl (Asn, Cys, Gln and His). Couplings were performed with N-[1H-benzotriazol(1-yl)(dimethylamino)-methylene]-N-methylmethanaminium hexafluorophosphate-N-oxide (HBTU) in N-methylpyrrolidone as coupling agent for the first 50 amino acids with a cycle time of 45 min, while HATU (1-(bis(dimethylamino)methylene)-1H-1,2,3-triazolo-(4,5-b)pyridinium hexafluorophosphate-3-oxide) in N-methylpyrrolidone was used for the *N*-terminal 37 amino acids. In order to increase the efficiency of the synthesis double coupling steps were applied for the last 50 amino acids with cycle times of 75 min. Deprotection of the Fmoc group was performed during the complete synthesis with 20% piperidine in DMF, containing 0.1 M HCOOH to avoid aspartimide side reactions. After 55 coupling cycles a small aliquot of the resin was removed, the material was deprotected and the identity of the product was controlled by MS. As the crude peptide showed the expected molecular mass, the protocol was continued as indicated above, and the final deprotection from the resin was performed with 95% TFA in water containing 3% triisopropylsilane and 5% phenol. The crude protein (Figure 1Ca) was purified by reverse phase HPLC (rp-HPLC) on a VYDACC column (40 × 300 mm, 218TPB1520, 300 Å) with a linear gradient of 10% A to 100% B in 45 min (A: 1000 ml water, 2 ml TFA; B: 500 ml acetonitrile, 100 ml water, 1 ml TFA) at a flow rate of 70 ml min^{-1} with spectrophotometric monitoring at $\lambda = 220$ nm. The fractions were checked by RP-HPLC (Shimadzu LC10) on a VYDACC 218TPB54 column (4.6 × 250 mm, 300 Å) with a linear gradient of 10% to 100% B over 45 min to give the final pure product (Figure 1Cb).

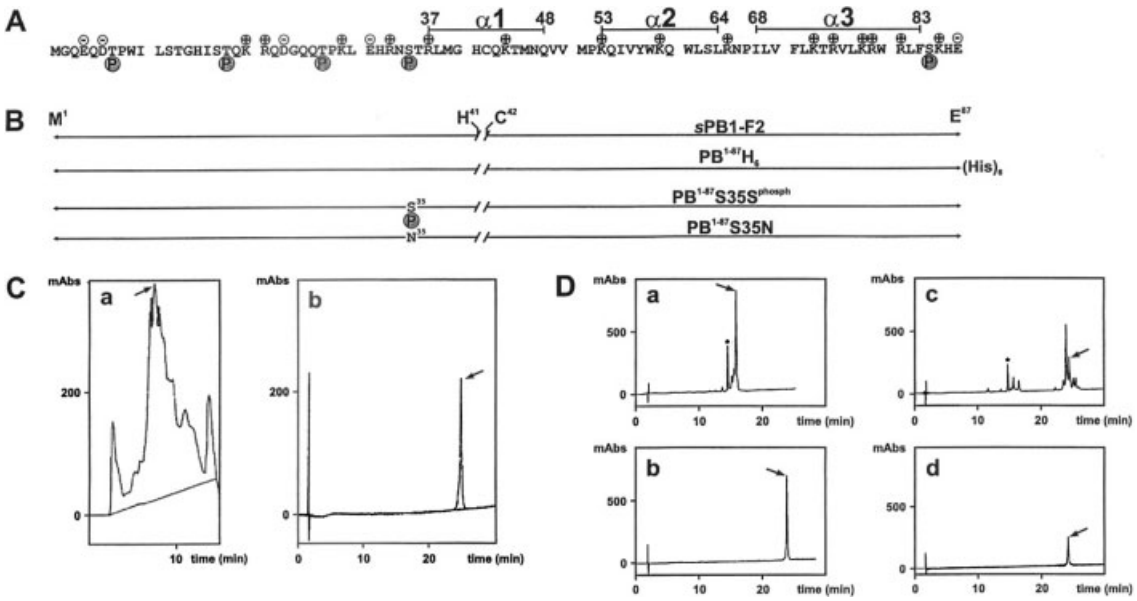


Figure 1 Synthesis of full length *s*PB1-F2, fragments and derivatives thereof. (**A**) Primary sequence of PB1-F2 derived from the IVA isolate PR8 is shown. Predicted secondary structures, phosphorylation sites, and positions of positively and negatively charged side chains are indicated (see Table 2). (**B**) Schematic depiction of the two ligation fragments and full length molecules synthesized to date by the chemical-ligation method: the wild type sequence as *s*PB1-F2 with either *C*-terminal His-tag ($PB^{1-87}H_6$), phosphoserine side chain in position 35 ($PB^{1-87}S35S^{phosph}$), or serine to asparagine exchange of the putative phosphorylation site ($PB^{1-87}S35N$). (**C**) Reverse-phase acetonitrile gradient HPLC chromatograms of the crude (a) and purified product (b) from the step by step method and (**D**) the chromatograms for the corresponding *N*-terminal thiolester and *C*-terminus (a and b, respectively), and the crude product after 12 h synthesis containing unreacted fragments and purified *s*PB1-F2 (c and d, respectively) for the chemical ligation method (**p*-acetaminothiophenol).

For various structural and functional analyses three overlapping fragments of *s*PB1-F2 were synthesized: PB^{1-40} comprising the 40 *N*-terminal residues of *s*PB1-F2, PB^{30-70} comprising the central residues 30–70 of *s*PB1-F2, and PB^{50-87} comprising the 38 *C*-terminal residues of *s*PB1-F2. The sequence selection was based on the secondary structure prediction that suggested the molecule could be conveniently divided into three similarly sized domains. The central PB^{30-70} and the *C*-terminal PB^{50-87} were produced following the same coupling and synthesis protocol used for full length *s*PB1-F2. In contrast to the standard SPPS protocols, several problems were encountered during the synthesis of the *N*-terminal fragment PB^{1-40}. For instance, beside the major full length synthesis product of PB^{1-40}, a sidechain artifact with a loss of 18 Da and a further product with a mass difference of +67 Da occurred that together accounted for approximately 60% of the total PB^{1-40} product. From the mass difference it was assumed that the artifact resulted from the formation of aspartimide at residue 23. Indeed, tryptic digestion and MS located the differences of −18 and +67 Da to the PB^{1-40} fragment Gln22-Lys29, confirming the occurrence of aspartimide as well as the β-piperidinyl ester through an initial cyclization to the succinimide and subsequent ring opening with piperidide under formation of α- and β- piperidide [15]. In order to avoid the formation of such artifacts, either the glycine in position 24 was replaced with Gly(Hmb, 2- hydroxy-4methoxybenzyl) or Gly(Dcpm, dicyclopropylmethyl), or the Asp-Gly sequence was replaced with the Asp-Gly(Hmb) dipeptide. As a consequence, synthesis of all sequences of *s*PB1-F2 that include Gly24 definitely requires the use of modified glycine residues for high yield synthesis.

In addition to the successful long chain synthesis, the overlapping peptide fragments of *s*PB1-F2 listed in Table 1 were synthesized. These fragments were also labelled with *N*-terminal biotin in a common manner on the resin before the final deprotection and purification.

Chemical-ligation synthesis of sPB1-F2 and derivatives. For structure–function correlation, it was necessary to create a mutant library of *s*PB1-F2 molecules. In order to improve the synthesis efficiency of the *s*PB1-F2 derivatives, chemical ligation of pre-synthesized peptide fragments of *s*PB1-F2 was established. The method involves the condensation of two unprotected peptide segments, one bearing a *C*-terminal α-thiolester and the other an *N*-terminal cysteine residue, to afford a full length polypeptide with a native amide linkage at the site of ligation [16]. Since the sequence of PB1-F2 used in this study contains only one cysteine residue in the middle of the molecule at position 42, the peptide appears as a *bona fide* candidate for this ligation approach. Initially, the *N*-terminal 41-residue fully protected fragment of *s*PB1-F2 with Boc-methionine on the *N*-terminus was synthesized by Fmoc/*t*-Butyl solid-phase strategy analogous to the method described previously [17] with an automated peptide synthesizer on 2-chlorotrityl (Cl-Trt) resin which was preloaded with the first *C*-terminal amino acid. As above, the Asp-Gly(Hmb) derivative in position 23/24 was employed in order to avoid aspartimide formation. The protected peptide was cleaved from the resin with acetic acid/trifluorethanol/DCM (1 : 1 : 8 v/v/v) for 1–2 h at room temperature. The thiolester of the 41-mer peptide was formed by treating the crude protected peptide (310 mg, 0.034 mmol)

Table 1 Overlapping Peptide Fragments of *s*PB1-F2 Synthesized by the Standard Elongation Method

Peptide	Sequence
PB^{1-40}	MGQEQDTPWILSTGHISTQKRQDGQQTPKLEHRNSTRLMG-OH
PB^{30-70}	LEHRNSTRLMGHCQKTMNQVVMPKQIVYWKQWLSLRNPILV-OH
PB^{50-87}	VMPKQIVYWKQWLSLRNPILVFLKTRVLKRWRLFSKHE-OH
PB^{1-20}	MGQEQDTPWILSTGHISTQK-NH_2
PB^{6-25}	DTPWILSTGHISTQKRQDGQ-NH_2
PB^{11-30}	LSTGHISTQKRQDGQQTPKL-NH_2
PB^{16-35}	ISTQKRQDGQQTPKLEHRNS-NH_2
PB^{21-40}	RQDGQQTPKLEHRNSTRLMG-NH_2
PB^{26-45}	QTPKLEHRNSTRLMGHCQKT-NH_2
PB^{31-50}	EHRNSTRLMGHCQKTMNQVV-NH_2
PB^{36-55}	TRLMGHCQKTMNQVVMPKQI-NH_2
PB^{41-60}	HCQKTMNQVVMPKQIVYWKQ-NH_2
PB^{46-65}	MNQVVMPKQIVYWKQWLSLR-NH_2
PB^{51-70}	MPKQIVYWKQWLSLRNPILV-NH_2
PB^{56-75}	VYWKQWLSLRNPILVFLKTR-NH_2
PB^{61-80}	WLSLRNPILVFLKTRVLKRW-NH_2
PB^{66-85}	NPILVFLKTRVLKRWRLFSK-NH_2
PB^{71-87}	FLKTRVLKRWRLFSKHE-NH_2

with DIPCDI (1.5 eq, 7.99 µl) and *p*-acetamidothiophenol (15 eq, 85.3 mg) in DCM (20 ml) overnight. The solvent was evaporated and the residue deprotected with TFA/water/TIPS (95 : 5 : 3 v/v/v). After precipitation with ice-cold diethyl ether the deprotected peptide thiolester was purified by preparative rp-HPLC on a VYDACC 218TPB1520 column (40 × 300 mm, 300 Å) with a linear gradient of 5% A to 55% B in 50 min in a H_2O-acetonitrile (0.1% TFA) system. The purified peptide thiolester ($[M + H]_{calc.}$: 4879, $[M + H]_{found}$: 4879) was obtained in a considerable yield of 45% (74 mg) based on the peptide content of the crude product (Figure 1Da).

Further a 46-residue *C*-terminal peptide of *s*PB1-F2, bearing the cysteine residue of *s*PB1-F2 at its *N*-terminus was synthesized using the standard method described above (Figure 1Db). For optimizing the synthesis, the amino acids Lys and Thr in positions 3 and 4 as well as 32 and 33 were replaced by the pseudoproline dipeptide Lys(Boc)-Thr($\Psi^{Me,\ Me}$ pro)-OH. The 41 aa segment thiolester (43.1 mg, 0.088 mmol) and the 46-residue *C*-terminal peptide of *s*PB1-F2 (51 mg, 0.088 mmol) were dissolved in a mixture of 0.05 M Tris/3 M urea, pH 8.5. The progress of the reaction was monitored by rp-HPLC (Figure 1Dc). The reaction was nearly complete (>80%) after 5 h. MALDI-MS confirmed the correct mass of the ligation product ($[M + H]_{calc.}$: 10510, $[M + H]_{found}$: 10510). After 15 h the reaction was complete. The ligation product was purified by following the same protocol as described above for *s*PB1-F2 (Figure 1Dd). Although the products from both syntheses are identical the chemical ligation technique is more efficient than chain-elongation for full length proteins and its derivatives as the intermediate peptides can be purified and yielded cleaner crude products with fewer side products after ligation. To date the optimized chemical-ligation protocol has been used successfully to synthesize efficiently not only wild type *s*PB1-F2, but also a mutant carrying a Ser to Asn exchange in position 35 (PB^{1-87}S35N), wild type *s*PB1-F2 carrying phosphoserine in position 35 ($PB^{1-87}S35S^{phosph}$) and wild type *s*PB1-F2 carrying a *C*-terminal tag of six histidine residues ($PB^{1-87}H_6$). Clearly this method is a very versatile procedure as numerous combinations of *N*- and *C*-terminal fragments can be fused together opening an easy path to synthetic mutagenesis of *s*PB1-F2.

Peptide Sequencing, Peptide Mapping and Mass Spectrometry

For *s*PB1-F2, six sequencing steps were completed on an Applied Biosystems 494A Procise HT sequencer according to a standard protocol. Electrospray ionization (ESI) mass spectra of all the peptides were recorded on a Micromass Q-TOF 2 mass spectrometer and matrix-assisted laser desorption/ionization time-of-flight (MALDI/TOF) mass spectra on a Bruker MALDI/TOF mass spectrometer. A peptide map of *s*PB1-F2 was determined using the above MALDI/TOF instrument after a standard tryptic digestion.

^{1}H NMR Spectroscopy

1D and 2D ^{1}H NMR spectra were recorded at various temperatures between 293 and 323 K on a Bruker Avance DMX 600 NMR spectrometer using a triple-resonance probe head with a gradient unit. *s*PB1-F2 was dissolved at a concentration of 1.0 mM (10 mg/ml) without pH adjustment in 50% and 90% aqueous TFE-d_2 to give final volumes of 600 µl. Spectra were referenced to the residual TFE signal at 3.95 ppm. 2D TOCSY spectra without spinning were accumulated with a mixing time of 110 ms.

Generation of *anti*-PB1-F2 Antibodies

Polyclonal antibodies directed against full length *s*PB1-F2 (*anti*-FL) were generated by immunization of rabbits with

*s*PB1-F2 in Titermac adjuvant. The titres of the antibodies were significantly increased when standard coupling of the peptide to keyhole limpet haemocyanin was omitted indicating that the peptide presents a *bona fide* antigen. The resultant antisera react with both *s*PB1-F2 as well as with its viral counterpart, PB1-F2, expressed in IAV_{PR8} infected cells or from CMV driven expression vectors by means of Western blot. Generation of antisera directed against the 15 *N*-terminal residues of *s*PB1-F2 (*anti*-N) was described previously [6].

Prediction of Secondary Structure and Phosphorylation Sites

A helical secondary structure was predicted using PSIPRED V2.4 [18], NNPREDICT [19], SSpro [20] and PROF V1.0 [21], and five possible phosphorylation sites were predicted using NetPhos 2.0 [22]. All programs were accessed through the ExPASy proteomics server of the Swiss Institute of Bioinformatics (www.expasy.org).

RESULTS AND DISCUSSION

Advantage of Chemical Total Synthesis of IAV Membrane Protein PB1-F2

A prerequisite for structural and functional studies is the provision of significant quantities of high quality PB1-F2 protein either by recombinant or synthetic methods. In general recombinant DNA-based expression systems have a number of shortcomings for the generation of the relatively large quantities of viral membrane proteins necessary for biochemical, functional, biophysical and structural studies. Cytotoxicity of the over-expressed recombinant material towards the producer systems generally results in low yields. In addition the amphipathic character of such molecules tends to favour strong intra- and intermolecular protein interactions and, in addition, they have the capacity to interact with membranes [13] that leads to self association and in many cases strong interactions with other hydrophobic proteins. As a consequence there is an inherent tendency in these virus proteins for aggregation and co-purification of contaminating cellular proteins. In particular for NMR investigations or membrane physiological studies these contaminations are critical and severely hinder such studies.

Hence, chemical peptide synthesis offers a viable alternative to recombinant strategies as synthetic full length viral membrane proteins usually have the ability to fold correctly and exhibit a functional status similar to their viral counterpart. In previous work we and others have successfully demonstrated efficient synthesis of viral proteins with similar characteristics and size to PB1-F2. For example, the synthetic HIV-1 regulatory protein Vpr [14,23] transduces cells, activates virus replication and induces apoptosis, while synthetic HIV-1 Vpu [24,25] and IAV M2 [26] both form ion channel systems in lipid bilayers. Based on our experience of long chain peptide synthesis of viral membrane proteins it was decided to develop optimized protocols for the total synthesis of high quality *s*PB1-F2 and its derivatives on a routine base.

Synthesis of sPB1-F2 by Chain Elongation and Ligation Reactions

Currently a considerable number of algorithms exist for the prediction of secondary structure and several of these have been used to predict the corresponding features and potential conserved phosphorylation sites in *s*PB1-F2 whose sequence corresponds to that derived from the isolate IAV_{PR8} [6]. The results of these are summarized in Figure 1A and Table 2. All algorithms predict a *C*-terminal helix terminating at residue 83 and comprising 9–20 residues in length. A shorter central helix is found in the region of residues 37–48 and has a maximum length of 12 residues. Further, the majority of the programs (Table 2) predict a third helix of approximately the same size as the latter and centred on residue 58. Clearly, all the programs predict the molecule to be divided essentially into two approximately equally sized domains. These correspond to an *N*-terminal domain for which little secondary structure was predicted and a *C*-terminal domain that exhibits a considerable propensity for α-helical structure. Furthermore, there is a relatively high number of positively charged side chains and Trp residues located within the *C*-terminal helix region: within the 44 residue *C*-terminal region a total of ten positive net charges are found, with a clustering of six in the region 73–85, and three Trp residues positioned at residues 58, 61 and 80, of which Trp61 is totally

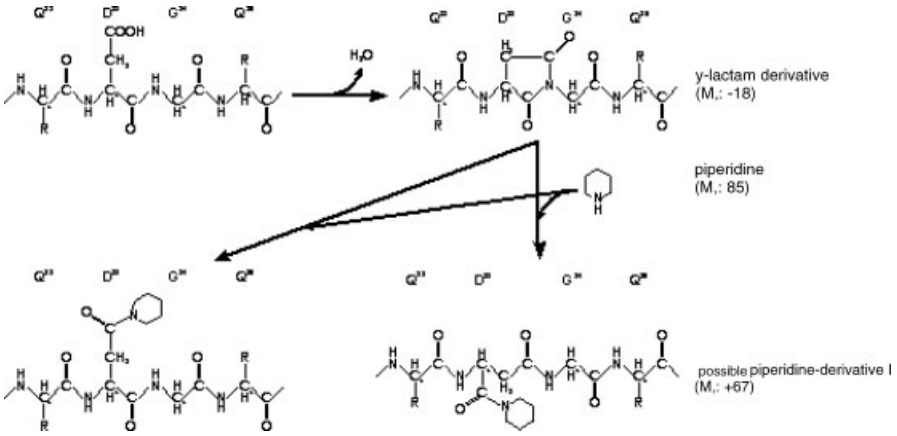
Table 2 Prediction of Helix Secondary Structure and Phosphorylation Sites in *s*PB1-F2

(a) Secondary structure			
Program (Ref.)	Residues in helix (longer than 2 residues)		
PSIPRED V2.4 (18)	37–48	53–64	68–83
Nnpredict (19)	38–41		64–83
sspro (20)	37–46*	53–63*	67–83
PROF (21)	38–48	54–63	75–83
(b) Phosphorylation sites are predicted (22) at Ser-35, Ser-84, Thr-7, Thr-18 and Thr-27.			

conserved [12] (Figure 1A). The high occurrence of cationic amino acids together with the high propensity for α-helix formation predict a model of the entire molecule that is amphipathic in character. Problems are usually experienced in the synthesis, processing and purification of such peptides through the build up of secondary structures on the resin during synthesis leading to reduced yields. As self association and aggregation cause problems during the purification steps or even irreversible denaturation in the non-native folded state, it was necessary to develop an optimized SPPS protocol that is specific for each peptide sequence. Our initial approach used a standard chain-elongation SPPS protocol that has been successfully employed for the synthesis of similar long chain linear viral peptides [24,14].

Although the desired product *s*PB1-F2 was detectable after purification, it was produced in relatively low yields. Mass spectrometric data of the crude peptide indicated the presence of a considerable number of smaller peptides and some side products of *s*PB1-F2, particularly those that originate from the *N*-terminal section of the peptide (data not shown here). These incomplete synthesis products could potentially arise from incomplete coupling and deprotection steps, possibly caused by either inter- and intrachain reaction with the resin matrix, or hydrogen bond mediated peptide aggregation, or side chain reactions. Most interestingly, there was a noticeable drop in coupling efficiency upon reaching the *N*-terminal half of the molecule that might be a consequence of a build up of secondary structure on the matrix, or even more likely, due to aggregation phenomena that were found in later investigations to be an inherent characteristic of PB1-F2 (paper in preparation). The coupling efficiency was improved by changing the coupling reagent after the first 50 amino acids and by increasing the coupling time for the last 50 coupling steps. Care was taken in the choice of deprotection conditions to avoid side reactions that were found to be prevalent in the *N*-terminal section of the molecule.

The origins of some of these problems became clearer during the synthesis of shorter fragments (Figure 1B) that were necessary for subsequent functional and structural studies. A summary of all the fragments and mutants of *s*PB1-F2 synthesized so far is shown in Figure 1B and Table 1. In an attempt to investigate the problematic areas within the molecule the sequence was divided into three similarly sized overlapping fragments, that approximate to the domains from the secondary structure predictions and correspond to the *C*-terminal, central and *N*-terminal sections of the molecule. In particular, side reactions were fully investigated during the synthesis of the *N*-terminal fragment PB^{1-40} and indicated that aspartimide formation was a particular problem associated with Asp23. Beside the intact fragment, a side artifact with a loss of 18 Da and a further product with a mass difference of +67 Da occurred that together accounted for approximately 60% of the PB^{1-40} product. From the mass difference it was assumed that the artifact resulted from the formation of aspartimide at residue 23. Indeed, tryptic digestion and MS located the differences of −18 and +67 Da to the PB^{1-40} fragment Gln22-Lys29, confirming the occurrence of aspartimide as well as the piperidine amide through an initial cyclization to the aspartimide and subsequent ring opening of this with piperidine (Scheme 1). This could be overcome by the use of 2-hydroxy-4-methoxybenzyl or dicyclopropylmethyl protected glycine for incorporation of residue 24 or the use of an Asp-Gly(Hmb) dipeptide. The optimized procedure gave crude material that afforded high quality protein after RP-HPLC. Thus we now have an optimized SPPS protocol that can be used efficiently to generate small fragments of *s*PB1-F2, their derivatives or mutants for use in the chemical ligation protocol below. Most importantly, the procedure definitely



Scheme 1 Generation of side products with mass differences of −18 and +67 during the synthesis of PB^{1-40} caused by cyclization to the γ-lactam through condensation of the Asp23 side chain with the backbone NH of Gly24 and subsequent ring opening by reaction with piperidine during the deprotection procedure to give two possible products.

requires the use of modified glycine residues (at position 24) if *N*-terminal fragments of PB1-F2 are to be synthesized. However, even with the considerably modified protocol the actual yield of full length protein was still relatively low.

As it was also necessary to create a mutant library of *s*PB1-F2 molecules for our ongoing structure–function correlation experiments, an alternative synthesis protocol was desirable in order to improve the synthesis efficiency and to allow easy access to such derivatives. Hence an alternative synthesis protocol was investigated that takes advantage of chemical ligation of two previously synthesized or expressed peptides and has previously been shown to be very efficient for long chain peptide synthesis (for recent review see [27]). This method makes use of the chemo selective trans thiolesterification of an unprotected peptide Cα-thiolester with an *N*-terminal Cys of a second peptide [16]. In particular native chemical ligation of unprotected peptide fragments has been demonstrated to be useful for water soluble proteins [28], and integral membrane viral proteins, such as the ion channel forming IAV M2 protein [26]. In our case this strategy appeared favourable as PB1-F2 has only one cysteine residue which is situated in a central position, 42, that divides the molecule into two almost equal size fragments (Figure 1A). Such an approach was particularly attractive as there was a specific interest in introducing point mutations within the two halves of the molecule. This stems from our approach to identify potential phosphorylation sites in PB1-F2 for which four highly conserved kinase acceptors sites in positions Thr7, Thr18, Thr27 and Ser35, and one in the *C*-terminal half at position Ser84 could be predicted (Figure 1A and Table 2). A further important argument in favour of the ligation method was the prediction of the secondary structure located within the *C*-terminal region, and thus would be the first part to build up on the resin with the corresponding loss in efficiency. Indeed, independent synthesis of *N*-terminal fragments was significantly more efficient than production of the same sequence in context of the full length molecule.

In our case, the chemical ligation strategy offers the advantage of increased efficiency and affords an extremely convenient tool for combining several wild type and mutant sequences in order to create a broad panel of mutant full length *s*PB1-F2 molecules in a relative short time with high efficiency. Two peptide segments were synthesized: the first consisted of the α-*C*-terminal thiolester of PB^{1-41}, comprising the *N*-terminal 41 residues of *s*PB1-F2, and a second fragment, PB^{42-87}, comprising the *C*-terminal 46 residues with the natural *N*-terminal cysteine residue (Figure 1Da and b). We thus made use of the one natural cysteine residue in the molecule at position 42 that separates the whole into two similar-sized fragments. The condensation reaction, carried out as further detailed in the Material and Methods section, afforded a native amide linkage at the site of ligation. The advantages of the chemical ligation method were indeed born out in practice as each fragment was synthesized with a significantly higher efficiency than the full length *s*PB1-F2 and these were then efficiently coupled.

In general, both synthesis procedures provided crude material that could be purified by HPLC to give single well defined peaks for *s*PB1-F2 (Figure 1Cb and Dd). The purification protocol was established first with the material produced following the standard elongation method. HPLC profiles of the crude and the purified protein products from both protocols are shown in Figure 1C and D, respectively. Both protocols produce the same material, although the purity of the crude product after chemical ligation was, as expected, considerably better than that from the elongation synthesis. For full length protein and its derivatives the ligation method has the advantage of using short peptides, 41 and 46 residues, intermediates that undergo more efficient coupling and can be purified more easily than the subsequent larger peptides and proteins generated in the elongation protocol.

To date the optimized chemical-ligation protocol has been used successfully to synthesize efficiently not only wild type *s*PB1-F2, but also a mutant carrying Ser to Asn exchange in position 35 (PB^{1-87}S35N), wild type *s*PB1-F2 carrying phosphoserine in position 35 ($PB^{1-87}S35S^{phosph}$) and wild type *s*PB1-F2 carrying a *C*-terminal tag of six histidine residues ($PB^{1-87}H_6$) (Figure 1B).

Molecular Characterization of sPB1-F2

There was no indication of any significant amounts of any by-product in the final material of the *s*PB1-F2 syntheses that was further characterized. MALDI mass spectrometry was used to determine the molecular weight of the product from both protocols and gave a molecular ion peak centred at 10484 Da that corresponds to the correct molecular mass of the peptide ($[M + H]_{calc.}$: 10484.3).

In order to verify the molecular integrity of the purified product, a peptide map of *s*PB1-F2 was obtained by MALDI/TOF-MS after a standard tryptic digestion (Figure 2). All predicted major peptide fragments with a molecular mass larger than 450 have been assigned. This almost complete coverage of the molecule, together with the detection of both carboxy terminal (T82–87) and in particular the amino terminal (T1–20) peptide, is strong evidence of the presence of a pure and original molecule comprising the complete sequence of *s*PB1-F2. There were no significant large signals that could be ascribed to either smaller incomplete molecules or those with a sequence different to that of *s*PB1-F2. In addition, the sequence of the *N*-terminal peptide (T1–20,

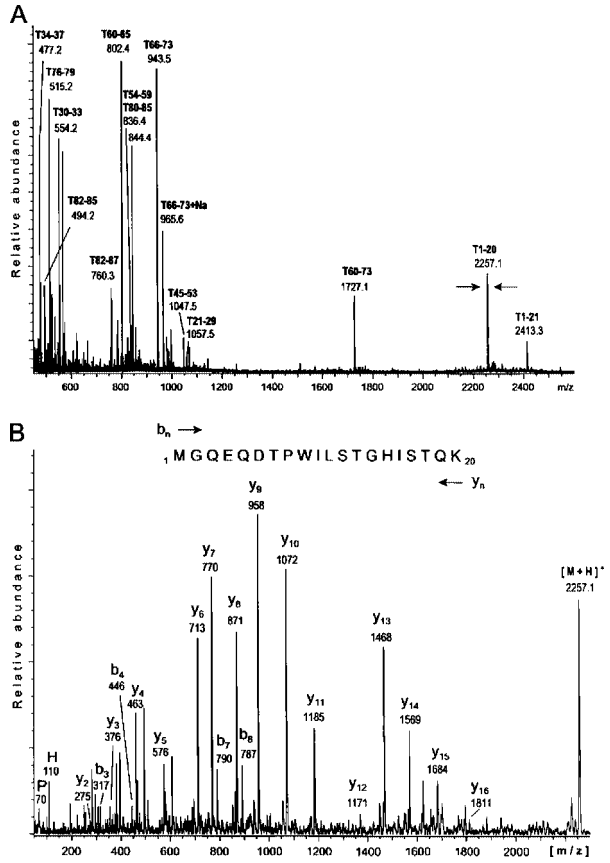


Figure 2 The identity of the purified *s*PB1-F2 peptide was confirmed by MALDI/TOF-MS after tryptic digestion (A) in which all major peptide fragments have been assigned, including a weak signal for T38–44 at 816.3 for the reactive cysteine-containing fragment. A daughter ion mass spectrum (B) of the amino terminal tryptic peptide from *s*PB1-F2, corresponding to the first 20 residues of mass 2257.1 Da indicated by arrows in A, was obtained by MALDI/TOF/TOF-MS (LIFT). Analysis of the fragmentation pattern of this ion affords the sequence shown.

marked with arrows in Figure 2A) was confirmed by a daughter ion spectrum recorded in the MS/MS (LIFT) mode as shown in Figure 2B.

The above data give indirect evidence that *s*PB1-F2 has been correctly synthesized. In addition to peptide mapping *N*-terminal sequencing was performed of purified *s*PB1-F2 which identified the correct six terminal residues as MGQEQD. From the sequencing data it was also concluded that the contamination from other peptides with free *N*-termini was estimated to be less than 10%. Thus the spectrometric data and combined sequencing data provided convincing evidence for an intact peptide with the correct sequence of *s*PB1-F2 with no detectable evidence of by-products in the purified material. Additional confirmation was afforded by inspection of the 2D ^{1}H NMR TOCSY spectrum for a 50% TFE solution of *s*PB1-F2 at 300 K (Figure 3). The through-bond vicinal correlation of ^{1}NH with ^{2}CH in the indole unit, enlarged in Figure 3C, confirms the presence of only four tryptophan residues. The same spectrum allows identification of four histidine residues from the observation of four-bond correlations between ^{2}CH and ^{4}CH (Figure 3B) corresponding to the histidine residues at positions 15, 32, 41 and 86 in the sequence. Although in each case it may be argued that additional peaks from impurities or shorter sequences could be overlapped with the signals shown here, data taken under different conditions (323 K and 90% TFE, data not shown), where the signal positions change, showed no evidence of such peaks. Thus the detection of these eight residues that are distributed throughout the length of *s*PB1-F2 is strong collaborative evidence for the presence of an intact molecule.

The molecular identity of *s*PB1-F2 was further defined by SDS-PAGE and Western blotting (Figure 4A). Similar to other small oligomeric membrane interacting viral proteins, such as the HIV-1 Vpu [29] or the IAV M2 protein [30] *s*PB1-F2 can only be separated with reasonable resolution in urea containing Tris/Tricine-gel according to Schagger and von Jagow [31]. In a standard SDS-PAGE according to Laemmli [32] *s*PB1-F2 could not be separated as a distinct species, as the peptide exhibits a high tendency for smearing and discontinuous migration indicating an unusually strong tendency for self association (data not shown). In Figure 4A serial dilutions of *s*PB1-F2 were separated and detected by Western blotting using antibodies directed either against a 15 residue *N*-terminal peptide of *s*PB1-F2 (*anti-N*) or the full length peptide (*anti-FL*). In Western blot analyses the majority of *s*PB1-F2 migrated with an apparent molecular mass of ~11 kDa, consistent with the predicted molecular weight of 10.5 kDa. Similar results were obtained by direct staining of *s*PB1-F2 in SDS PAGE with Coomassie blue (not shown). No additional smaller peptide species migrating below the major 10.5 kDa product were detected with antibodies directed against the full length *s*PB1-F2 peptide or the *N*-terminal fragment (Figure 4A) (or the *C*-terminal fragment, data not shown) attesting to the absence of degradation or chain termination products. However, a species was detected that migrated at the M_r position expected for a dimer, and at higher protein concentrations per lane, a species that migrated in the M_r range of trimers of *s*PB1-F2. It could be argued that the observation of a dimer may arise from disulphide bridge formation through the cysteine residue at position 42, although this is extremely unlikely under the conditions used. That this is not the case is evident from the similar behaviour of the *N*-terminal fragment PB$^{1-40}$ (Figure 4C) which does not possess such a cysteine residue yet again shows strong evidence of not only dimer formation but enhanced formation of higher oligomers. For comparison of *s*PB1-F2 with virally expressed PB1-F2, similar Western blot analyses were

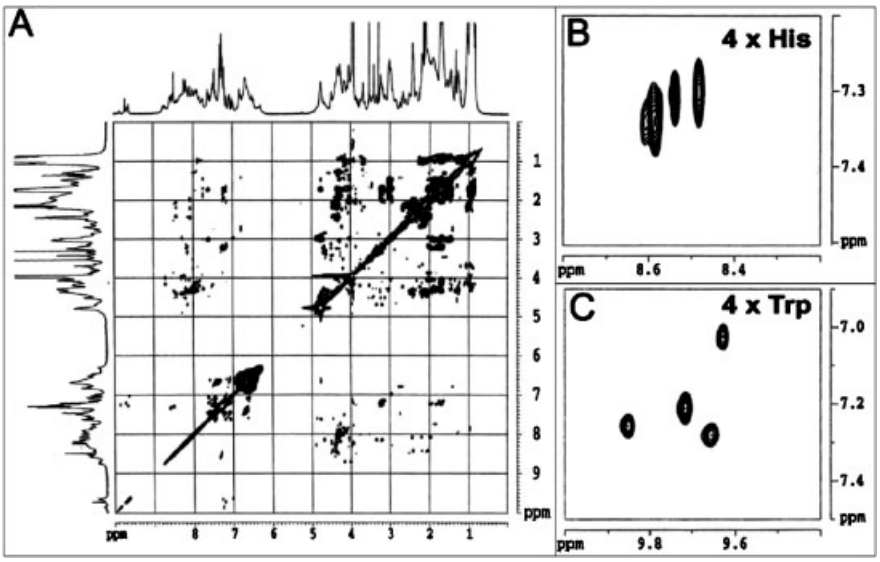


Figure 3 Two-dimensional ^{1}H TOCSY spectrum (mixing time 110 ms) of a 1 mM solution of *s*PB1-F2 in 1 : 1 (v/v) TFE-d_2/H_2O at 300 K (A). Expanded regions correspond to correlations between ^{2}CH and ^{4}CH of the histidines (B) and ^{1}NH with ^{2}CH (C) of the four tryptophans.

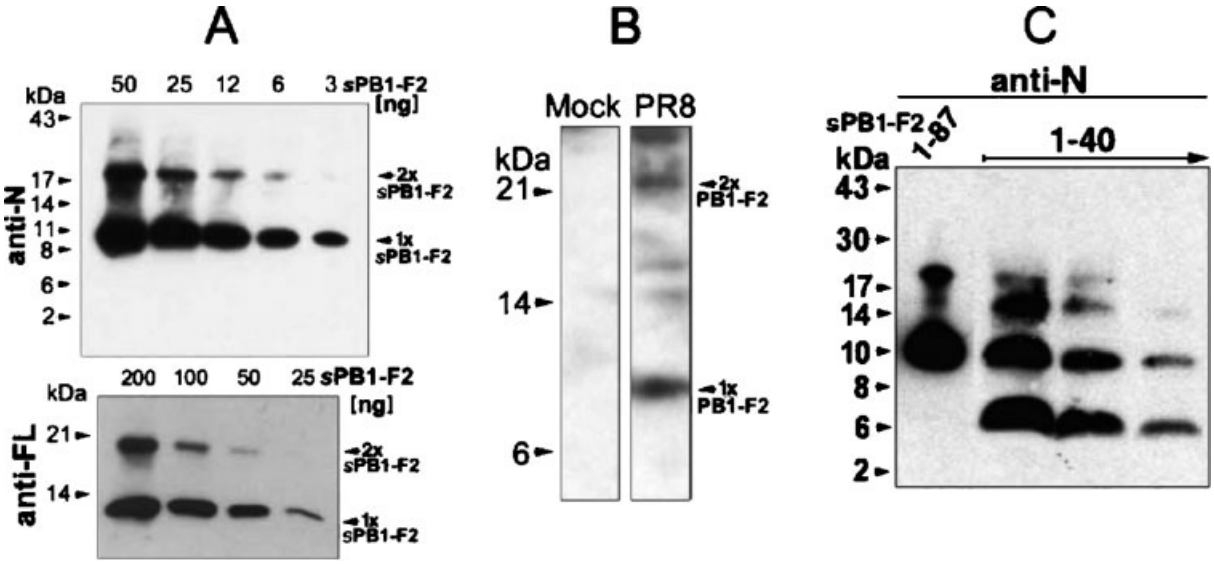


Figure 4 Analysis of *s*PB1-F2 by Western blot. (A) A serial dilution of 50 to 3 ng of *s*PB1-F2 per lane were separated in a 20% to 10% gradient SDS-PAGE, transferred to PVDF membrane, and stained with either *anti-N* or *anti-FL* antibodies. Positions of molecular weight standard marker proteins are indicated to the left, positions of monomers, dimers and trimers of *s*PB1-F2 are indicated to the right. (B) In a parallel Western blot analysis extract of L cells infected with IAV_{PR8} or mock infected cells were analysed with *anti-N* antibody. (C) Similar experiment to that of A with 100 ng *s*PB1-F2, and 5, 2.5 and 1.25 μg PB^{1-40}.

conducted with lysates of mouse L cells that were infected with IAV_{PR8} (Figure 4B). Despite the complex nature of extracts from acutely infected cells, when compared with the homogenous peptide solution of *s*PB1-F2, evidence for a similar oligomeric pattern of the native PB1-F2 was observed with a pronounced dimeric form of the molecule (Figure 4B). Clearly this phenomenon is related to the inherent tendency of PB1-F2 to self associate. This is currently under further investigation by us using chemical cross linking and dynamic light scattering and will be considered in detail in a future manuscript.

CONCLUSIONS

We have shown biologically active influenza A virus protein PB1-F2 and a number of full length derivatives are most easily synthesized using an optimized chemical ligation procedure afforded though the presence of

a conveniently, centrally positioned cysteine residue. Standard chain elongation SPPS methods are used to generate suitable fragments and precursors, and require precautions to avoid byproduct formation for *N*-terminal derivatives. The synthesized protein has been thoroughly characterized by physical methods.

Acknowledgements

We are indebted to J.W. Yewdell and J.R. Bennink (NIH, Bethesda, MD, USA) for their continuous support. This work was supported by NIH/NIDDK RO1 grant DK59537-01, by grant Schu11/2-1, by a Heisenberg grant from the Deutsche Forschungsgemeinschaft, and by grant IE-S08T06 from the German Human Genome Research Project to U.S. We thank Dr Beyermann for helpful discussion, and Prisca Kunert, Barbara Brecht and Christel Kakoschke for excellent technical assistance.

REFERENCES

1. Yewdell J, Garcia-Sastre A. Influenza virus still surprises. *Curr. Opin. Microbiol.* 2002; **5**: 414–418.
2. Lamb RA, Takeda M. Death by influenza virus protein. *Nat. Med.* 2001; **7**: 1286–1288.
3. Webster RG. Influenza virus: transmission between species and relevance to emergence of the next human pandemic. *Arch. Virol. Suppl.* 1997; **13**: 105–113.
4. Webby RJ, Webster RG. Emergence of influenza A viruses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 2001; **356**: 1817–1828.
5. Webster RG, Bean WJ Jr. Genetics of influenza virus. *Annu. Rev. Genet.* 1978; **12**: 415–431.
6. Chen W, Calvo PA, Malide D, Gibbs J, Schubert U, Bacik I, Basta S, O'Neill R, Schickli J, Palese P, Henklein P, Bennink JR, Yewdell JW. A novel influenza A virus mitochondrial protein that induces cell death. *Nat. Med.* 2001; **7**: 1306–1312.
7. Lowy RJ. Influenza virus induction of apoptosis by intrinsic and extrinsic mechanisms. *Int. Rev. Immunol.* 2003; **61**: 2001–2005.
8. D'Agostino DM, Ranzato L, Arrigoni G, Cavallari I, Belleudi F, Torrisi MR, Silic-Benussi M, Ferro T, Petronilli V, Marin O, Chieco-Bianchi L, Bernardi P, Ciminale V. Mitochondrial alterations induced by the p13II protein of human T-cell leukemia virus type 1. Critical role of arginine residues. *J. Biol. Chem.* 2002; **277**: 34 424–34 433.
9. Silic-Benussi M, Cavallari I, Zorzan T, Rossi E, Hiraragi H, Rosato A, Horie K, Saggioro D, Lairmore MD, Willems L, Chieco-Bianchi L, D'Agostino DM, Ciminale V. Suppression of tumor growth and cell proliferation by p13II, a mitochondrial protein of human T cell leukemia virus type 1. *Proc. Natl Acad. Sci. USA* 2004; **101**: 6629–6634.
10. Muthumani K, Hwang DS, Desai BM, Zhang D, Dayes N, Green DR, Weiner DB. HIV-1 Vpr induces apoptosis through caspase 9 in T cells and peripheral blood mononuclear cells. *J. Biol. Chem.* 2002; **277**: 37 820–37 831.
11. Sherman MP, Greene WC. Slipping through the door: HIV entry into the nucleus. *Microbes Infect.* 2002; **4**: 67–73.
12. Gibbs JS, Malide D, Hornung F, Bennink JR, Yewdell JW. The influenza A virus PB1-F2 protein targets the inner mitochondrial membrane via a predicted basic amphipathic helix that disrupts mitochondrial function. *J. Virol.* 2003; **77**: 7214–7224.
13. Chanturiya AN, Basanez G, Schubert U, Henklein P, Yewdell JW, Zimmerberg J. PB1-F2, an influenza A virus-encoded proapoptotic mitochondrial protein, creates variably sized pores in planar lipid membranes. *J. Virol.* 2004; **78**: 6304–6312.
14. Henklein P, Bruns K, Sherman MP, Tessmer U, Licha K, Kopp J, de Noronha CM, Greene WC, Wray V, Schubert U. Functional and structural characterization of synthetic HIV-1 Vpr that transduces cells, localizes to the nucleus, and induces G2 cell cycle arrest. *J. Biol. Chem.* 2000; **275**: 32 016–32 026.
15. Lauer J, Fields CG, Fields GB. Sequence dependence of aspartimide formation during 9-fluorylmethoxylcarbonyl solid-phase peptide synthesis. *Lett. Pept. Sci.* 1994; **1**: 197–205.
16. Dawson PE, Muir TW, Clark-Lewis I, Kent SB. Synthesis of proteins by native chemical ligation. *Science* 1994; **266**: 776–779.
17. von Eggelkraut-Gottanka R, Klose A, Beck-Sickinger AG, Beyermann M. Peptide α-thioester formation using standard Fmoc-chemistry. *Tetrahedron Lett.* 2003; **44**: 3551–3554.
18. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics* 2000; **16**: 404–405.
19. Kneller DG, Cohen FE, Langridge R. Improvements in protein secondary structure prediction by an enhanced neural network. *J. Mol. Biol.* 1990; **214**: 171–182.
20. Baldi P, Brunak S, Frasconi P, Soda G, Pollastri G. Exploiting the past and the future in protein secondary structure prediction. *Bioinformatics* 1999; **15**: 937–946.
21. Ouali M, King RD. Cascade multiple classifiers for secondary structure prediction. *Protein Sci.* 2000; **9**: 1162–1176.
22. Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 1999; **294**: 1351–1362.
23. Cornille F, Wecker K, Loffet A, Genet R, Roques B. Efficient solid-phase synthesis of Vpr from HIV-1 using low quantities of uniformly ^{13}C-, ^{15}N-labeled amino acids for NMR structural studies. *J. Pept. Res.* 1999; **54**: 427–435.
24. Henklein P, Schubert U, Kunert O, Klabunde S, Wray V, Kloppel KD, Kiess M, Portsmann T, Schomburg D. Synthesis and characterization of the hydrophilic C-terminal domain of the human immunodeficiency virus type 1-encoded virus protein U (Vpu). *Pept. Res.* 1993; **6**: 79–87.
25. Kochendoerfer GG, Jones DH, Lee S, Oblatt-Montal M, Opella SJ, Montal M. Functional characterization and NMR spectroscopy on full-length Vpu from HIV-1 prepared by total chemical synthesis. *J. Am. Chem. Soc.* 2004; **126**: 2439–2446.
26. Kochendoerfer GG, Salom D, Lear JD, Wilk-Orescan R, Kent SB, DeGrado WF. Total chemical synthesis of the integral membrane protein influenza A virus M2: role of its *C*-terminal domain in tetramer assembly. *Biochemistry* 1999; **38**: 11 905–11 913.
27. David R, Richter MP, Beck-Sickinger AG. Expressed protein ligation. Method and applications. *Eur. J. Biochem.* 2004; **271**: 663–677.
28. Wilken J, Kent SB. Chemical protein synthesis. *Curr. Opin. Biotechnol.* 1998; **9**: 412–426.
29. Schubert U, Strebel K. Differential activities of the human immunodeficiency virus type 1-encoded Vpu protein are regulated by phosphorylation and occur in different cellular compartments. *J. Virol.* 1994; **68**: 2260–2271.
30. Sugrue RJ, Hay AJ. Structural characteristics of the M2 protein of influenza A viruses: evidence that it forms a tetrameric channel. *Virology* 1991; **180**: 617–624.
31. Schagger H, von Jagow G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* 1987; **166**: 368–379.
32. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970; **227**: 680–685.

