Preventing the N-terminal processing of human interferon α-2b and its chimeric derivatives expressed in *Escherichia coli*

Fatima Ahsan[a], Qurratulann Afza Gardner[a,*], Naeem Rashid[a], Greg J. Towers[b], Muhammad Akhtar[a,c,*]

[a]School of Biological Sciences, University of the Punjab, New Campus, Lahore 54590, Pakistan, [b]Division of Infection and Immunity, University College London, London WC1E 6BT, UK, [c]Biological Sciences, University of Southampton, SO17 1BJ, UK

Running Title: *Interferon α-2b derivatives*

Key words: cytochrome b_5, interferon-cytochrome b_5 -chimera, interferon fusion, VSV-G (vesicular stomatitis virus G glycoprotein) pseudotyped HIV-1 derived vector particles, interferon bioassay, protein purification, mass spectrometry, protein expression, protein folding, protein chemistry.

*Corresponding Authors

*Professor M Akhtar FRS

E-mail: ma3@soton.ac.uk

*Dr Qurratulann Afza Gardner

Email: annegardner80@yahoo.co.uk

Tel:(+) 92-42-99230960 and (+) 44-23- 80767718

Fax: (+) 92-42-99230980

Abstract

We have previously shown that human interferon α-2b (IFN) produced in *Escherichia coli (E. coli)* is heterogeneous at the N-terminal, with three major species [1]. These are: (a) the direct translation product of the gene retaining the N-terminal methionine, (b) a species from which the methionyl residue has been removed by *E. coli* methionyl aminopeptidase to give the native interferon α-2b and (c) in which the N-terminal Cys residue of the latter contains an acetyl group. In this paper we overcome this heterogeneity, using engineered interferon derivatives with phenylalanine residue directly downstream of the N-terminal methionine (Met-Phe-IFN). This modification not only prevented the removal of the N-terminal methionine by *E. coli* methionyl aminopeptidase but also the subsequent N-acetylation. Critically, Met-Phe-IFN had enhanced activity in a biological assay. N-terminal stabilization was also achieved by fusing human cytochrome b_5 at the N-terminal of interferon (b_5-IFN-chimera). In this case also, the protein was more active than a reciprocal chimera with cytochrome b_5 at the C-terminal of interferon (Met-IFN-b_5-chimera). This latter protein also had a heterogeneous N-terminal but addition of phenylalanine following Met, (Met-Phe-IFN- b_5-chimera), resolved this problem and gave enhanced biological activity.

## *1. Introduction*

Following the original discovery in the late 1950s [2], in recent years, interferon (**IFN**, **1a** Figure 1) has gained recognition as a drug of choice for the treatment of many viral diseases and certain cancers [3]. Effective use of peptide and protein drugs in patients suffers from their instability *in vivo*, arising from rapid clearance rates, premature uptake by tissues (for instance the reticuloendothelial system), loss through the kidneys, and immunogenicity or antigenicity [4]. In order to prolong the plasma half-life of interferon a number of strategies have been employed, with the aim of increasing *in vivo* efficacy as well as reducing injection frequency hence helping to improve patients' life quality. Until now, conjugation of interferon with polyethylene glycol (PEG) has been the most successful approach. This, significantly reduces plasma interferon clearance and immunogenicity [4].

In general, pegylation occurs randomly on the ε-amino group of surface lysine residues [4], though, site-specific pegylation has also been reported [5, 6]. For more specific targeting, however, fusion technology to graft a protein specifically to the N- or C-terminal of a drug has been developed. Albumin fusion to interferon is one such example [6, 7]. Most commonly, *E. coli* is the host of choice for the heterologous expression of recombinant proteins. Expression of eukaryotic proteins

2

in a prokaryotic host such as *E. coli* leads, sometimes, to inclusion body formation requiring subsequent protein refolding. The presence of disulfide bridges in proteins makes their refolding a complex and delicate task, not tractable for multi-thiol-containing fusion partners such as albumin, which have been produced in yeast [7, 8].

Our previous work has shown that the expression of the **Met-IFN** encoding-DNA in *E. coli* leads to the production of a heterogeneous mixture of three products. These are: (a) the direct translation product of the DNA retaining the N-terminal methionine (**2** Figure 1), (b) a species from which the methionyl residue has been removed by *E. coli* methionyl aminopeptidase to give the native interferon α-2b (**1a**, Figure 1) and (c) a species in which the N-terminal Cys residue of the latter contains an acetyl group (**1b**, Figure 1). In the present study two related problems have been addressed. We firstly avoid the aforementioned processing of the N-terminal of interferon α-2b, by introducing a Phe residue following the initiator N-terminal Met of interferon α-2b (**3** Figure 1). Secondly, we constructed chimeras between interferon α-2b and human cytochrome b_5 (**4-6** Figure 1). Critically, cytochrome b_5 contains no cysteine residues and thus the two disulfide bridges in the final product will be formed by the four cysteines of interferon α 2b. We tested fusions of cytochrome b_5 at the carboxy- or amino- terminal of interferon α-2b. Construct **4**, (Figure 1) also had a counterpart containing a Phe residue penultimate to the N-terminal methionine. The biological activities of these constructs were determined functionally using a novel method, involving sensitivity of lentiviral vector infection to the recombinant interferon protein. *[Figure 1 near here]*
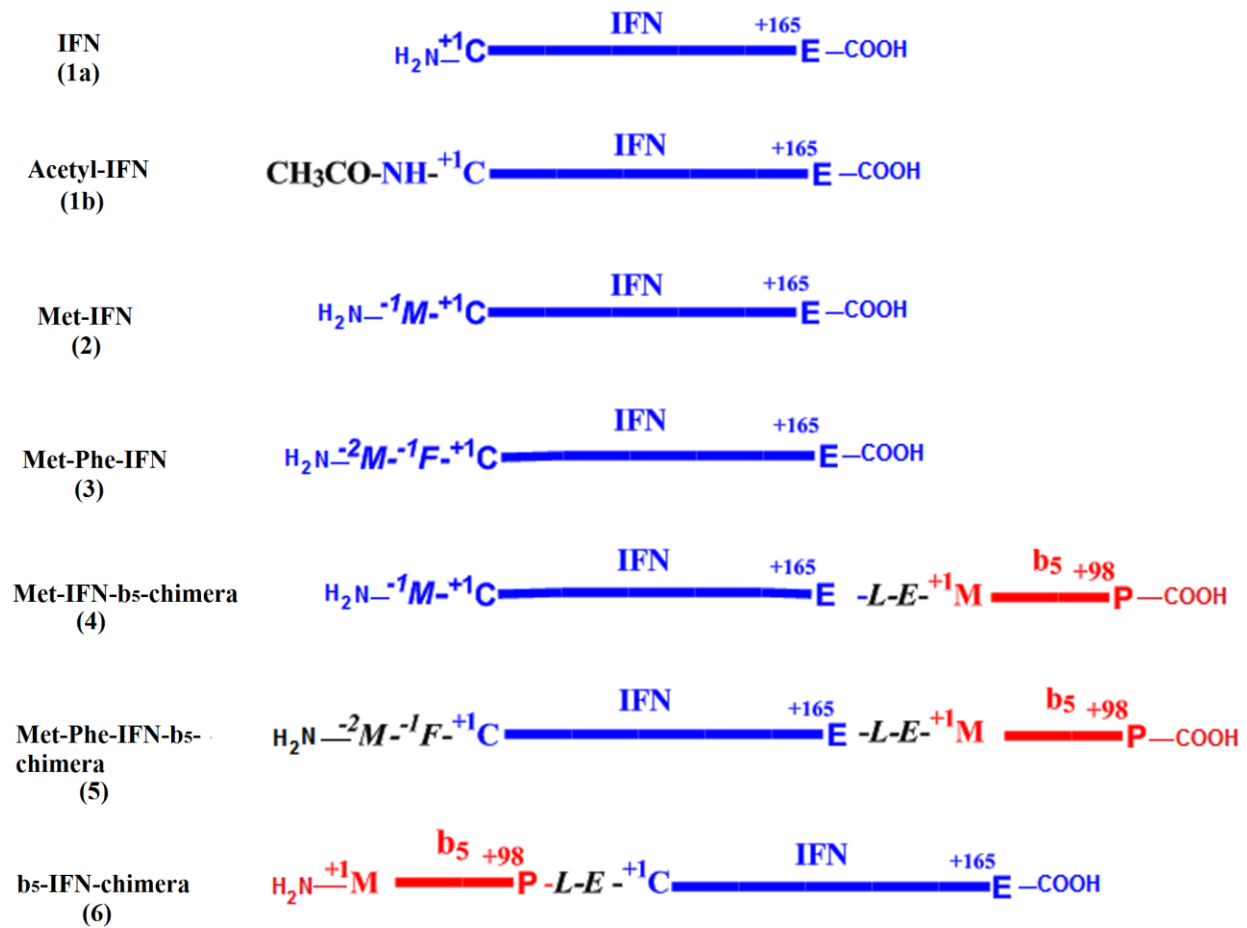
Figure 1. Interferon derivatives. Interferon core, referring to interferon α-2b, is shown in blue, its N- and C-termini are emboldened and additions to N-terminal are shown by (-) numbers. Cytochrome b_5 skeleton in the chimera is in red with the connecting residues (L-E) in black.

2. Results and Discussion

2.1 The N-terminal heterogeneity

Our previous study considering human **Met-IFN** engineering (structure **2** Figure 1) using *E. coli*, as a host, showed that the expressed protein consisted of three species: (a), IFN with an intact methionine (**Met-IFN** (**2**)), encoded by the DNA sequence, (b), **IFN** without its N terminal methionine (**1a**) formed by the action of host methionine aminopeptidase on the primary translation product of the gene (**2**) and (c), **N-acetyl-IFN** (**1b**) formed by the action of *E. coli* acetyl transferase on **IFN** (1a) itself.

We set out to prevent N-terminal heterogeneity taking several approaches. Unsuccessful approaches included engineering expression vectors that encoded **Met-IFN,** and methionine

4

aminopeptidase aiming to convert the primary translation product into native **IFN** [9]. However, soon it became evident that the main cause of heterogeneity was the action of an acetyl transferase which, as we will see later, efficiently modified the **IFN** N-terminal amino group of Cys^1. Notwithstanding, this heterogeneity, through careful HPLC separation we were able to obtain homogeneous native **IFN**. Commercial preparations of recombinant **IFN**, obtained from expression in *E. coli* are deemed to be homogenous, however, it is not known whether this is also the result of purification by HPLC or use of another genetic tool where terminal N-acetylation is avoided. In a novel PEGylation of interferon α-2b at one of the disulphide bonds, the MALDI spectrum in the paper, of the precursor interferon α-2b, was shown to have a mass of 19,306 Da [5] which corresponds to N-acetyl-interferon α-2b (**2,** Figure 1) and not interferon α-2b itself that would have a mass of 19,265 Da. The interferon used in the latter work was reported to be obtained from a yeast expression system. The specificities of yeast N-acetyl transferases have been studied, using a panel of proteins but in no case Cys modification was observed [10]. It would thus appear that a specific, but unknown, signal in interferon α-2b is recognized for acetylation by an acetyl transferase in yeast as well as *E. coli.*

2.2 Expression of the gene for **Met-Phe-IFN (3)***: analysis of the protein and of its tryptic peptides*

In order to avoid the heterogeneity, our subsequent efforts were aimed to use of an N-terminal sequence that was not susceptible to the action of methionine aminopeptidase. Here, we took advantage of our studies on the preparation of a variety of proinsulin derivatives. In general, these derivatives have a phenylalanine following the N-terminal methionine, that is, a Met-Phe sequence [11, 12]. In these derivatives we did not observe the removal of methionine or any other modification involving this residue.

In the light of these observations, interferon derivative containing Met-Phe was engineered, expressed in *E. coli* and the inclusion bodies solubilized at pH 11.00. We serendipitously discovered that use of pH 11 greatly improved the yield of the correctly folded protein compared to pH 8.0, used in our previous study [1]. Unlike the HPLC profile of the refolded **Met-IFN** (Figure 2A. Inset), **Met-Phe-IFN (3)** gave a single peak (Figure 2A) showing the desired M_r of 19543.7 (19544.4, theoretical) on MALDI analysis (Figure 2B). The purity of **Met-Phe-IFN** (**3**) was such that after the Q-sepharose step the mass spectral profile of the protein was indistinguishable from that of the HPLC peak (Figure 2B &C). *[Figure 2 near here]*
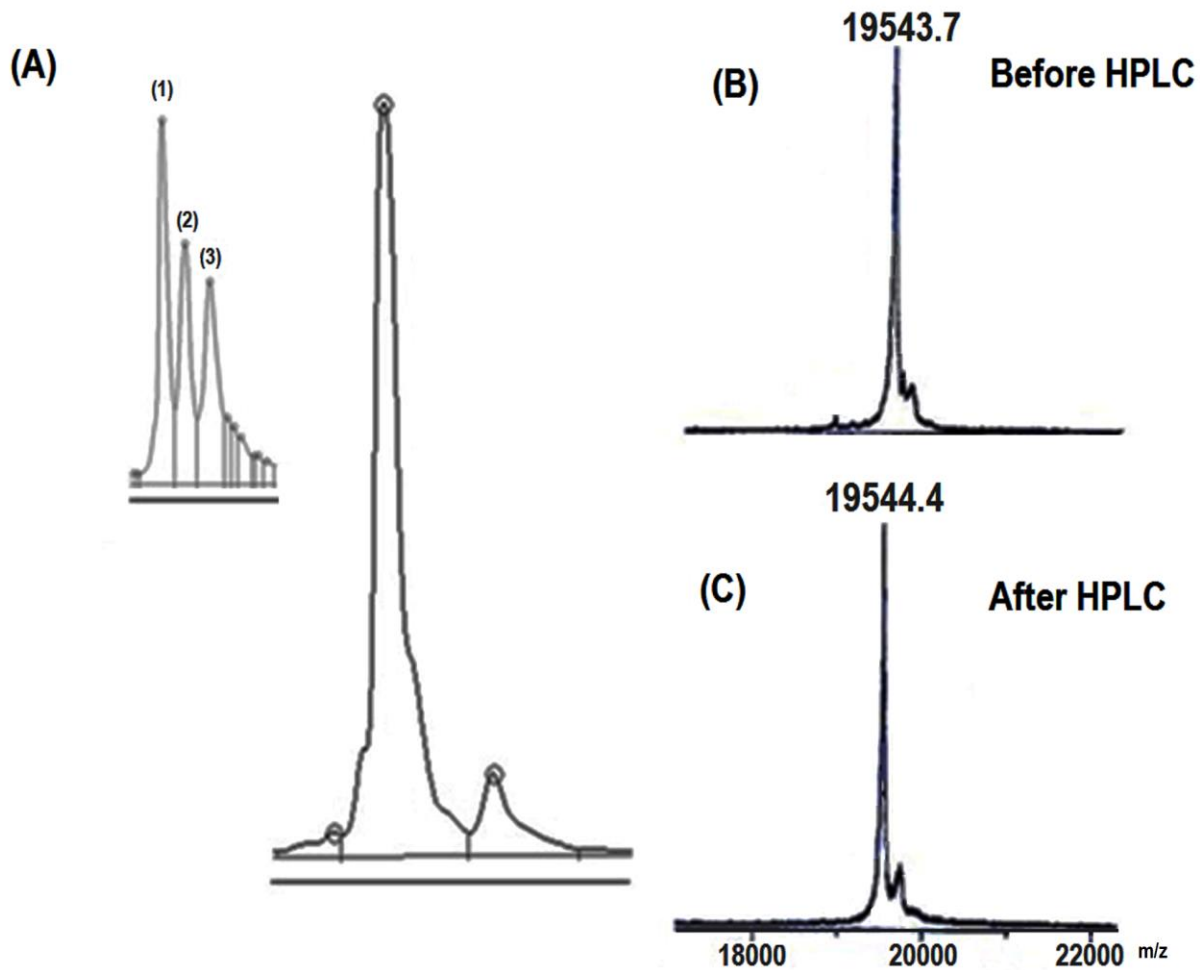
Figure 2. (A) HPLC chromatogram of Met-Phe-IFN (3 Figure 1) eluting with 59-61% acetonitrile. . Inset shows the equivalent profile from the expression of the **Met-IFN** gene containing three species, IFN, N-acetyl-IFN and Met-IFN (peaks 1.2 and 3 respectively)**. (B and C) MALDI-TOF spectrum of Met-Phe-IFN (3 Figure 1).** (**B**) Before and (**C**) after HPLC, showing the [M+H]$^+$ ion (Cal. 19544.4).

Tryptic digestion of **Met-Phe-IFN**, without prior cleavage of the disulphide bonds, followed by ESI-MS analysis, showed the presence of 92% of the theoretically possible peptides including those containing the correct disulphide bonds (Table S4 and structures **I & II** Figure S3a), in particular **I** which had the expected N-terminal sequence. The folding process was thus unaffected by the N-terminal modification in **Met-Phe-IFN.**

*2.3 Expression of the genes for the chimeras **4, 5 and 6** (**Figure 1**): analysis of the proteins and of their tryptic peptides*

Encouraged by the results on **Met-Phe-IFN**, we extended the use of this approach to fusion proteins involving interferon α-2b and cytochrome b_5. The most well used partner for bioactive peptide drugs is human serum albumin. The choice of this protein has been justified for a variety of reasons [13, 14] and its use is exemplified by human serum albumin-interferon fusion proteins [6, 7, 15-17], which are typically expressed in yeast [15]. The expression of such proteins in *E. coli* is problematic since serum albumin contains 35 cysteine residues engaged in 17 disulphide bonds. Since the putative conjugate of the protein with interferon will add four further cysteine residues, the possibility of the formation of native disulphide bonds within each partner will be greatly reduced. In order to avoid this complication we selected human cytochrome b_5 as a fusion partner for interferon since it does not possess cysteine residues.

Three fusion proteins were designed. Two of these contained the native interferon α-2b sequence ligated either to the N-or the C-terminal of cytochrome b_5 (**4** and **6** respectively, Figure 1). In the third derivative, the sequence Met-Phe-IFN was placed at the cytochrome b_5 N-terminal (**5** Figure 1).

These fusion proteins were expressed in *E. coli* as inclusion bodies, which were solubilized and refolded as described for **Met-Phe-IFN** (**3** Figure 1). HPLC analysis of **Met-IFN-b₅-chimera** (**5** Figure 1) showed the presence of three peaks. From their masses these were characterized as: (a) the translation product of the encoding-DNA retaining the N-terminal methionine (peak 3, Figure 3), (b) where methionine had been removed (peak 1, Figure 3) and (c) in which the amino group of Cys residue had been acetylated (peak 2, Figure 3). The profile with **Met-IFN-b₅-chimera** (**5** Figure 1) is, thus, identical to that obtained from the expression of **Met-IFN**. *[Figure 3 near here]*
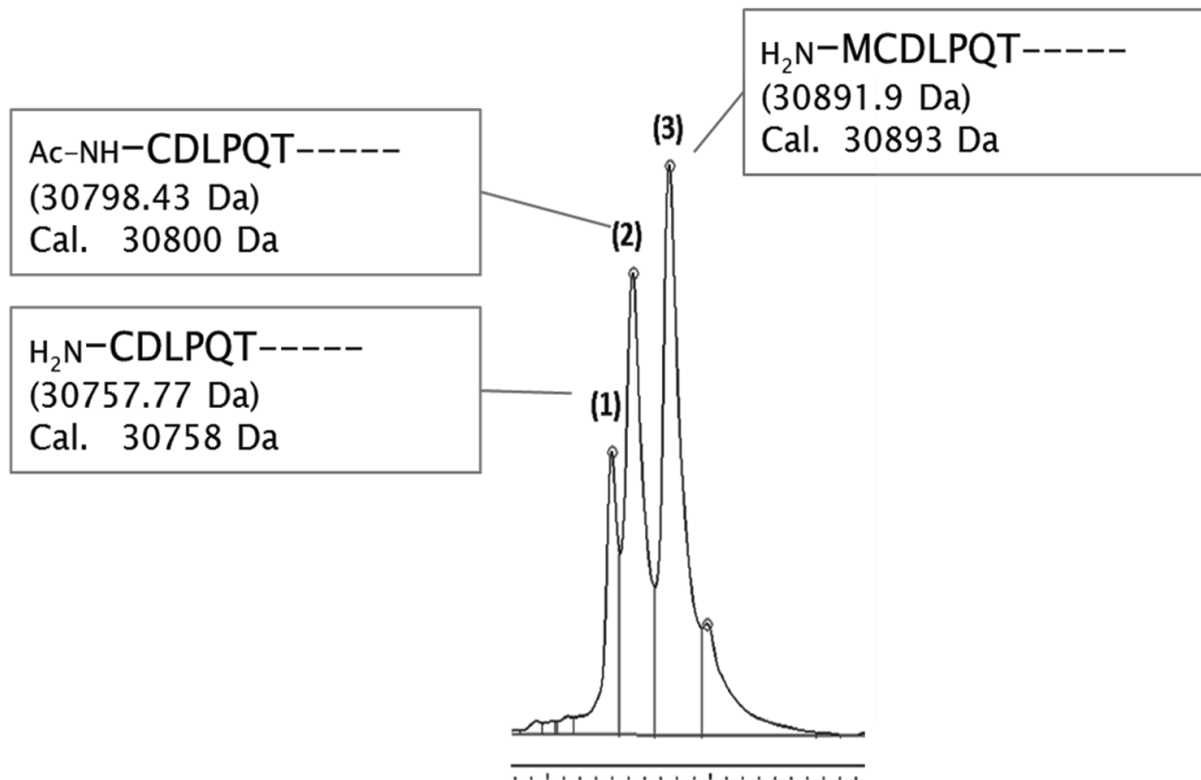
Figure 3. HPLC chromatogram of the products from the expression of the *Met-IFN-b₅-chimera*. Chromatogram containing three products is shown above. Masses as determined by ESI are above the peaks

The HPLC analysis of **b₅-IFN-chimera** (**4** Figure 1) again showed heterogeneity, giving three peaks (Figure 4A). These corresponded to the primary translation product, loss of methionine, by the action of methionine aminopeptidase and further removal of the newly generated Ala by an unknown *E. coli* peptidase (Figure 4B-D). We have previously observed similar processing of the N-terminal residues of proteins, when the N-terminal methionine is followed by a small residue.

Unlike, the preceding two chimeras (**3** and **6** Figure 1) the fusion protein, **Met-Phe-IFN-b₅-chimera** (**5** Figure 1) showed a single peak on HPLC, giving the expected mass Figure 5). *[Figure 4 & 5 near here]*
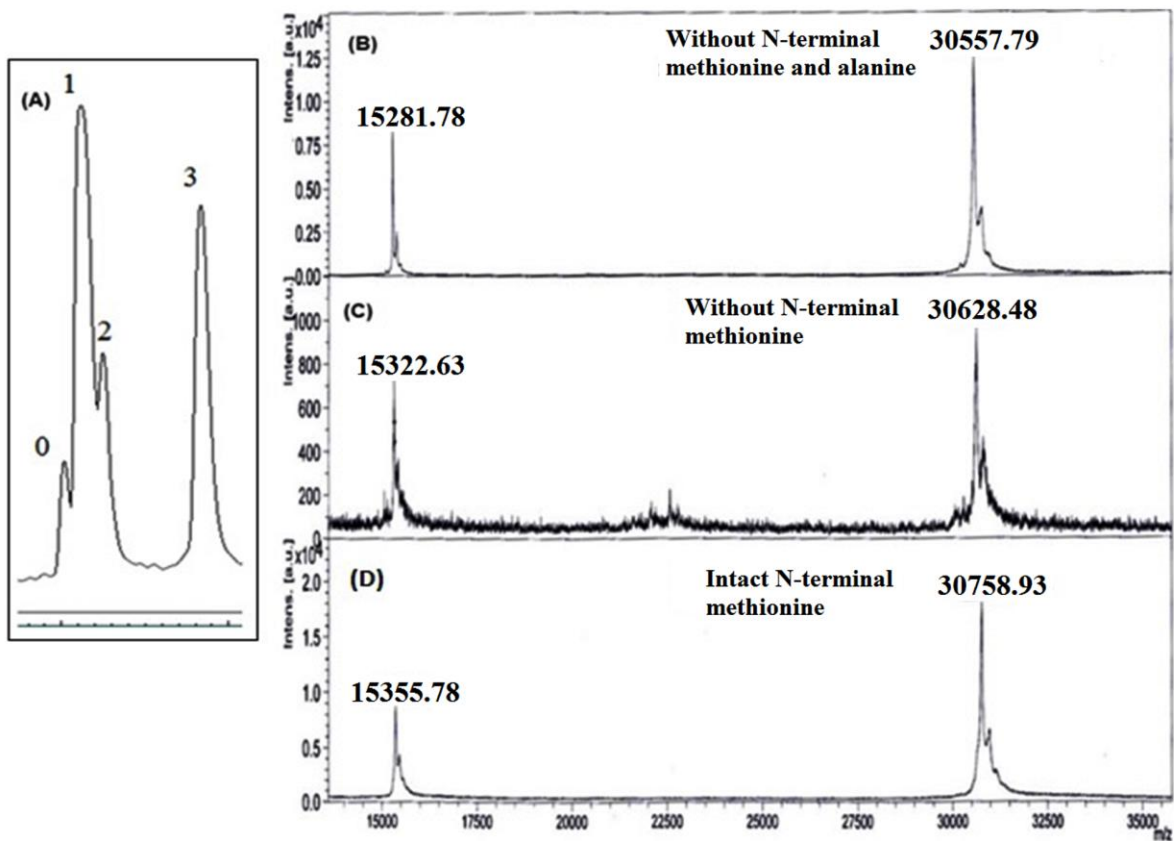
Figure 4. RP-HPLC chromatogram and mass spectra of b₅-IFN-chimera (6, Figure 1). **A**, Protein was eluted in four chromatographic peaks (peak 0, 1, 2 & 3). **B, C and D**, MALDI-TOF spectra of the three HPLC peaks showing [M+H]$^{+1}$ ions. (**B**), peak 1(Cal. 30556.8), (**C**) peak 2 (Cal. 30627.8) and (**D**) peak 3(Cal. 30758.8).
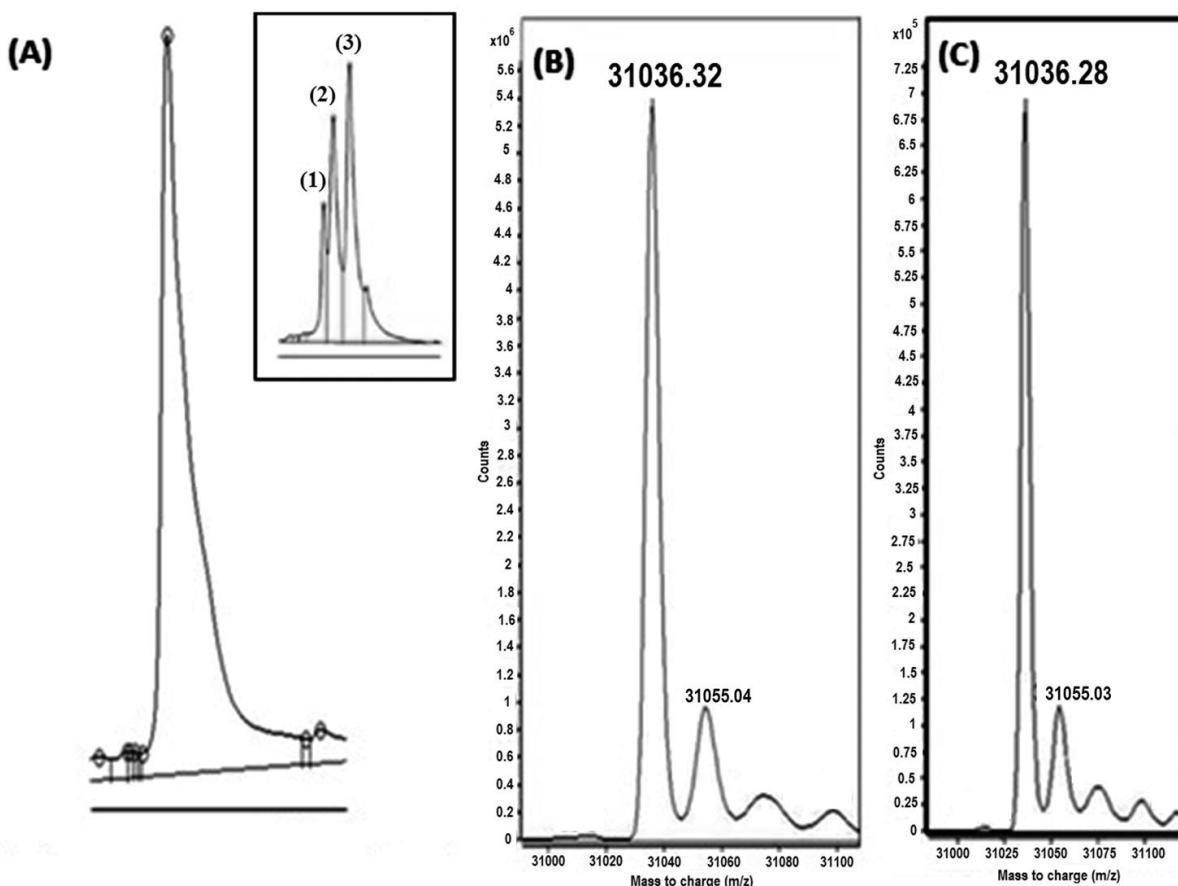
Figure 5. (A), RP-HPLC profile of Met-Phe-IFN-b₅-chimera (5, Figure 1) eluted at 50-61% acetonitrile. Inset shows the equivalent profile from the expression of the **Met-IFN-b₅-chimera** containing three species, **(B) ESI-MS of the protein before HPLC** showing its M_r (cal. 31036.1*)*; **(C) ESI-MS of the protein after HPLC** showing its M_r (cal. 31036.1).

The correct disulphide bond formation by the three fusion proteins, **4, 5 and 6** (Figure 1) was established by mass spectrometric analysis of tryptic digests. In the cases of the derivatives **4 and 5** (Figure 1), the expected disulphide-bonded fragments (IIIa/ IIIb) plus (II, Figure 3Sb, Supplementary section 6) and (I) plus (II, Figure S3c, section 6) respectively were obtained showing that in each case the two disulphide bonds were between Cys^1-Cys^{98} and Cys^{29}-Cys^{138}. With **b₅-IFN-chimera** (**6** Figure 1), the tryptic digest contained besides the Cys^{29}-Cys^{138} fragment (II), the second disulphide-bond-fragment (IV, Figure S3d, section 6), which as expected, had sequences from the C-terminal of b₅ as well as N-terminal of IFN. Thus, using the oxidative-refolding conditions used in this work, the interferon motif underwent correct refolding, tolerating the extensions on either end of the molecule.
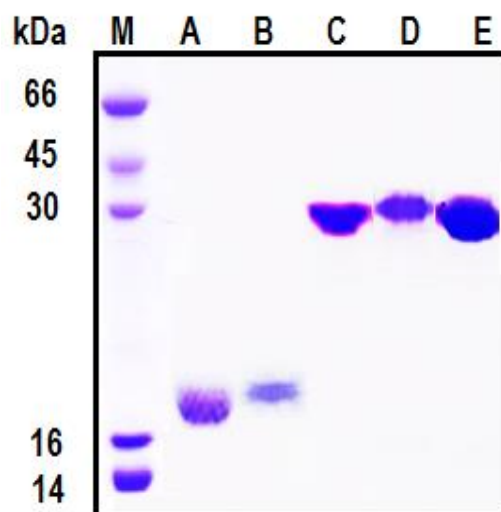
Figure 6: Reductive SDS-PAGE analysis of various interferon constructs after Q-sepharose column purification. Lanes: **M**; home-made ladder showing 66, 45, 30, 16 and 14 kDa species, **A**; Met-IFN, **2** (19.0 kDa), **B**; Met-Phe-IFN, **3** (19.5 kDa), **C**; Met-IFN-b5-chimera, **4** (30.8 kDa), **D**; Met-Phe-IFN-b5-chimera, **5** (31.0 kDa), **E**; b5-IFN-chimera, **6** (30.7 kDa).

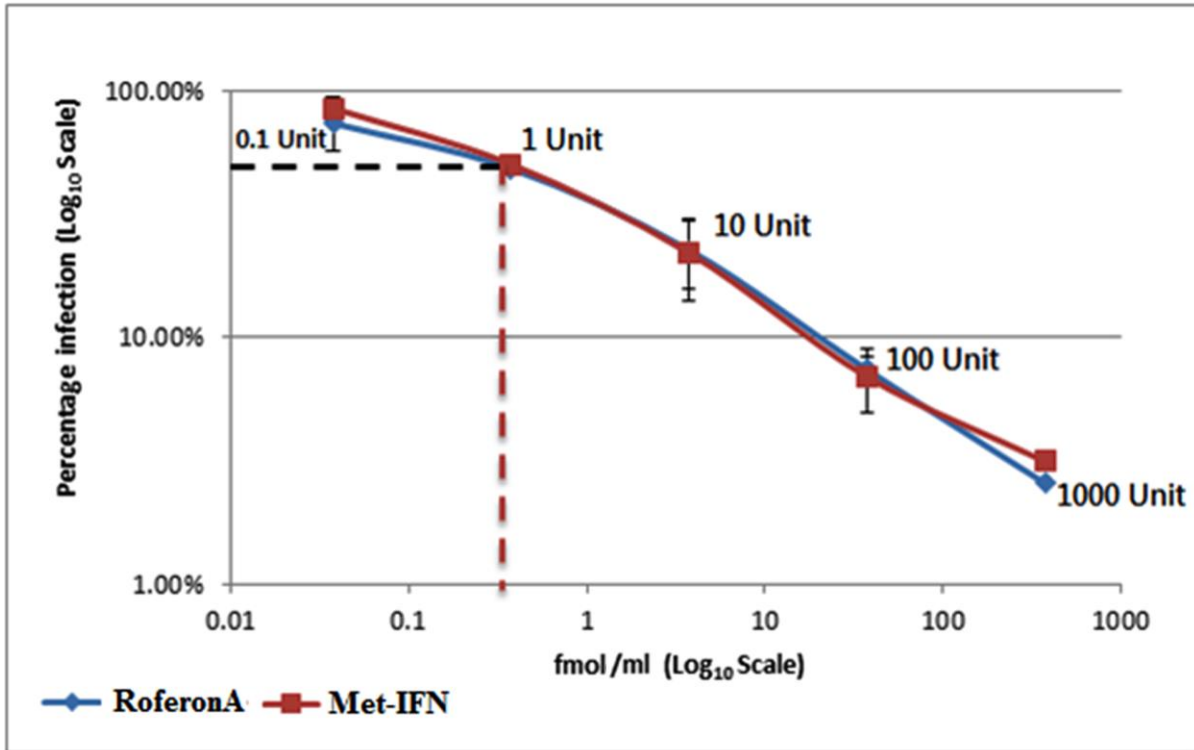*2.4 Antiviral activities of the various derivatives of interferon*

The SDS-PAGE profiles of all the five interferon derivatives (**2-6** Figure 1) are shown in Figure 6. With the availability of 4 new derivatives **(3-6)** their antiviral activities were assayed using pseudotyped viral particles expressing enhanced green fluorescent protein (eGFP) in the THP-1 cell line [18]. This assay is based on the infection of target cells by VSV-G pseudotyped HIV-1 vector, carrying vesicular stomatitis virus G glycoprotein (VSV-G) in the envelop and eGFP as reporter gene. . The number of eGFP expressing cells was used as a measure of viral infection. Cells were pre-incubated with interferon or its derivatives and anti-viral activity was determined by the test molecules' ability to reduce viral infection in the target cells A comparison of IC_{50} values and relative bioactivities (%), calculated on a molar basis, is given in Table 1 which shows that **Met-IFN** (**2**, Table 1 and Figure 7) prepared in the present work had the same activity as a reference standard In the light of the profile in Figure 7, the same five concentrations (0.0374, 0.374, 3.74, 37.4 and 374 pM) which gave a satisfactory spread of points across the curve in Figure 7 were selected for each construct. The anti-viral activity profile of each interferon derivative in

Figure 8 shows that for four of the constructs (**2, 4, 5, 6; figure 1**) the experimental points, as expected, were satisfactorily spread over the curves, however, in the case of **Met-Phe-IFN** (**3**), the points were clustered towards the lower half of the curve, which indirectly points to the fact that this construct differs from others in having higher activity. This feature is also indicated by the observation that **3** at 0.0374 pM showed similar inhibitory effect (ca.52%) as **Met-IFN** (**3**) at 0.374 pM (ca. 49.4%). This trend was present at all concentrations. With the proviso that points in the curve for **3** are not broadly scattered, the comparison of the IC_{50} values in Table 1 indicates that this construct is around 8-fold more active than its comparator **2.**
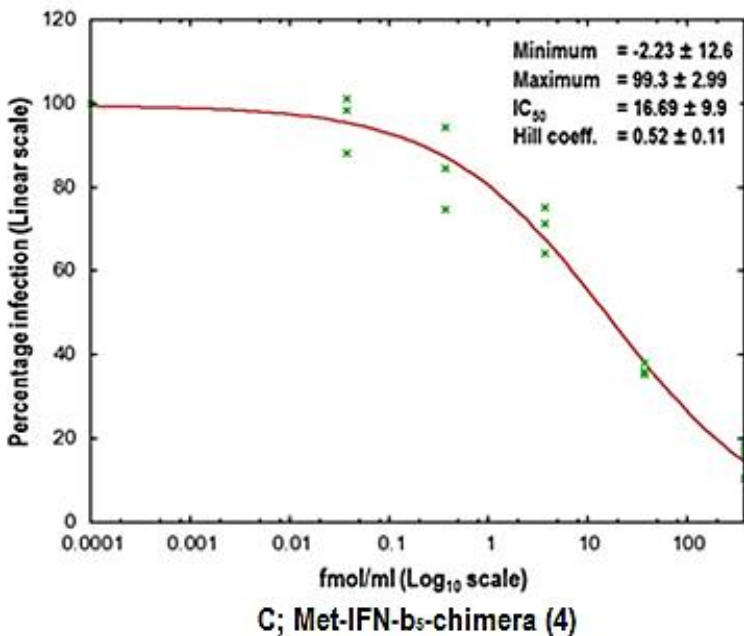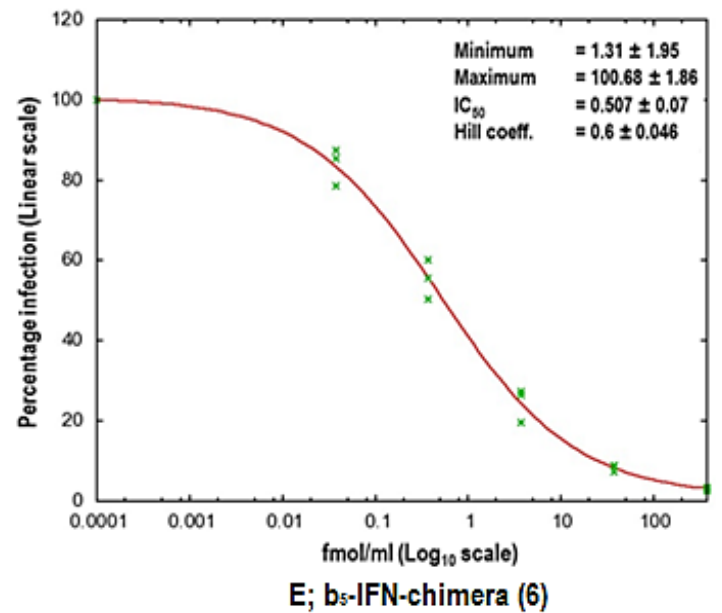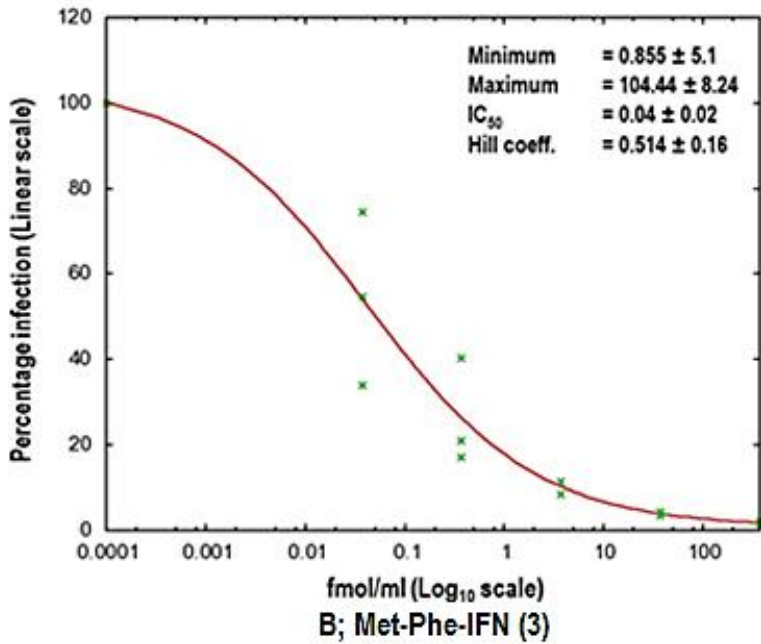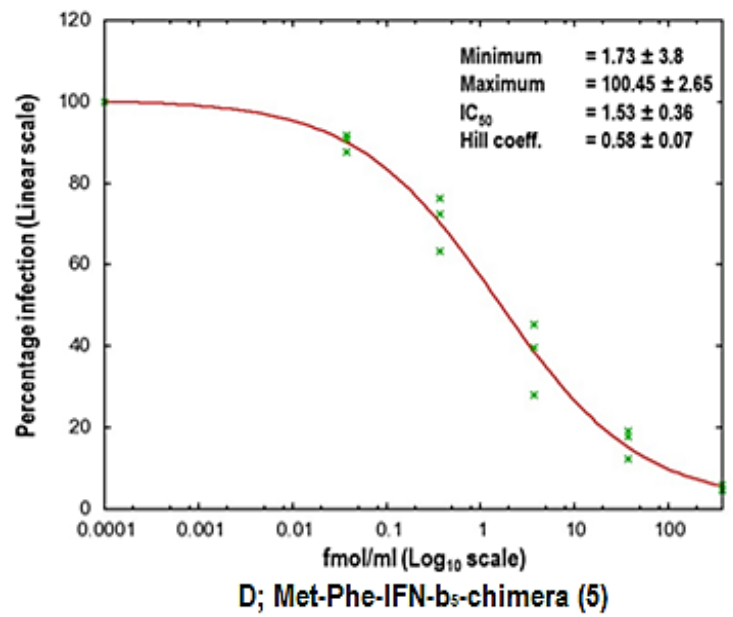
 The fusion of cytochrome b_5 at the amino-terminal of interferon (**b₅-IFN**-chimera, **6**, Figure 1) was advantageous over fusion at the carboxy- terminal (**Met-IFN-b₅-chimra, 4,** Figure 1), since the former had 75% bioactivity, relative to the control, compared to 3% for the latter. This shows that an amino terminal, shielded by the b_5 domain in **b₅-IFN**-**chimera** enhances the biological activity of the interferon domain. This trend of enhancing the biological activity by a stable N-terminal appendage is further shown by the fact that the relative bioactivity of **Met-Phe-IFN-b₅-chimera** (increased to 33% from 1.3% for its comparator). The only difference between the two molecules is the presence after the initiator Met of Phe, which protects the N-terminal against the action of a peptidase. The nearly 8-fold increase in activity for construct **2** versus **3** parallels 11 fold that of **4** versus **5;** the difference in both cases is the replacement of N-terminal Met for Met-Phe, which improves activity in both cases.

Table 1. Relative bioactivities of recombinant interferon and its derivatives prepared in this study and those reported in literature. Numbers in parenthesis refer to structures in Figure 1.The IC_{50}values were determined using the online software "IC_{50} toolkit (ic50.tk)" http://www.ic50.tk/ and counter checked by the GraphPad prism7 ® software. International units were calculated as described in the Experimental Section. International units (IU) for the constructs **g, h and i** were calculated from % bioactivity, relative to interferon α-2b, reported in [13] and for **j and k** in [15].

Sr. No.	Protein	Molecular Weight (kDa)	IC_{50} (pM) (calculated from the software)*	IU/mol	Relative bioactivity (calculated on molar basis) (%)
a.	RoferonA	19	-	2.66×10^{15}	100
b.	**Met-IFN (2)**	19	0.313 ± 0.13	2.64×10^{15}	100
c.	**Met-Phe-IFN (3)**	19.5	0.04 ± 0.02	2.21×10^{16}	837
d.	**Met-IFN-b$_5$-chimera (4)**	30.8	16.69 ± 9.9	0.53×10^{14}	2.0
e.	**Met-Phe-IFN-b$_5$-chimera (5)**	31	1.53 ± 0.36	6.983×10^{14}	26.4
f.	**b$_5$-IFN-chimera (6)**	30.7	0.507 ± 0.07	1.604×10^{15}	60.75
g.	PEG-Intron	31	-	9.3×10^{14}	35
h.	PegaSys	59	-	1.86×10^{14}	7
i.	rHSA/IFNα2b	86	-	3.72×10^{13}	1.4
j.	Albuferon beta	87	-	7.2×10^{13}	2.7
k.	IFN-beta	20	-	6.6×10^{14}	25

Figure 7. Comparative anti-viral activity of Met-IFN (present work) **and RoferonA.** A ten-fold serially diluted **Met-IFN** from 374 to 0.0374 fmol/ml was titrated for antiviral activity. Each point represents the average value from three independent experiments. Error bars represent standard deviation values. The units of RoferonA used for comparison are shown next to the symbols.

A; Met-IFN (2)

Minimum	= -0.75 ± 5.8
Maximum	= 101.59 ± 5.4
IC_{50}	= 0.313 ± 0.13
Hill coeff.	= 0.497 ± 0.107

D; Met-Phe-IFN-b₅-chimera (5)

Minimum	= 1.73 ± 3.8
Maximum	= 100.45 ± 2.65
IC_{50}	= 1.53 ± 0.36
Hill coeff.	= 0.58 ± 0.07

B; Met-Phe-IFN (3)

Minimum	= 0.855 ± 5.1
Maximum	= 104.44 ± 8.24
IC_{50}	= 0.04 ± 0.02
Hill coeff.	= 0.514 ± 0.16

E; b₅-IFN-chimera (6)

Minimum	= 1.31 ± 1.95
Maximum	= 100.68 ± 1.86
IC_{50}	= 0.507 ± 0.07
Hill coeff.	= 0.6 ± 0.046

C; Met-IFN-b₅-chimera (4)

Minimum	= -2.23 ± 12.6
Maximum	= 99.3 ± 2.99
IC_{50}	= 16.69 ± 9.9
Hill coeff.	= 0.52 ± 0.11

15

Figure 8. Anti-viral activity of various interferon derivatives. Ten-fold serially diluted derivative from 374 to 0.0374 fmol/ml was titrated for antiviral activity. The graphs were plotted between percentage infection (linear scale) and conc. (fmol/ml, Log_{10} scale of conc.) of each interferon derivative, using online software programme "IC_{50} toolkit (ic50.tk)" http://www.ic50.tk/, showing the data points of three sets of readings. The programme draws the line, among three sets of readings to give the best sigmoid curve, showing the general trend to calculate the minimum and maximum Y-asymptotes, IC_{50}, and Hill-coefficient of each interferon derivative along with their standard errors.

According to Huang *et al*. [15] a fusion protein between human serum albumin and interferon α-2b (rHSA/IFNα2b) has 1.4% relative bioactivity when compared on molar basis to unmodified interferon. A second fusion construct of interferon β and albumin was estimated from the IC_{50} values to have 10% of the antiviral activity compared to interferon β [17]. Of all the fusion proteins reported in the literature PEG-intron (12-kDa polyethylene glycol moiety conjugated to interferon α-2b) was the most active having 35% of the activity of interferon α-2b (Table 1). Our comparison of the relative bioactivities of the molecules that we have developed has shown that fusion of interferon with cytochrome b_5 at its amino terminal, (**b_5-IFN-chimera, 6,** Figure 1), gives the most active fusion protein; 75% of the activity of interferon α-2b. It is interesting to note that the molecular mass of PEG-intron is very similar to that of **b_5-IFN-chimera.** Furthermore, a simple addition of Met-Phe to the N-terminal of **IFN** (**1a** Figure 1) produces the highest specific activity antiviral protein with 8 fold enhancement over the native cytokine.

In Table 1 calculated specific activities are stated in international units per mole (mol) as well per mg of the proteins. The relative potency of the constructs, based on their activities, in decreasing order is:

Met-Phe-IFN > **Met-IFN**=RoferonA> **b_5-IFN-chimera**> **Met-Phe-IFN-b_5-chimera**> **Met-IFN-b_5-chimera**

3. Conclusions

The N-terminal heterogeneity found in the expression of **Met-IFN** (**1a**, Figure1) in *E. coli* was mimicked by the fusion construct **Met-IFN-b_5-chimera** (**4** Figure 1) which also gave three species. These were the primary translation product (MCDLPQT…), another from which the N-terminal Met was removed and the third arising from the acetylation of the latter. It is reasonable to assume that these post-translation modifications occur within the bacterial cells when the protein is unfolded and is, somehow, the property of the interferon N-terminal sequence. The removal of Met by, methionine aminopeptidase is expected because of the small size of the next residue. Acetylation of the newly formed N-terminal of Cys is unexpected and such a modification has not been reported previously. However, once the IFN N-terminus is modified to contain a Phe residue following the initiator Met, the heterogeneity is abolished as illustrated by **Met-Phe-IFN** and **Met-Phe-IFN-b_5-chimera** (**3 and 5** Figure 1). The latter modification has the further advantage of enhancing the antiviral activity by about 8 fold relative to their unmodified counterparts; compare the activity data of entry **b** (**Met-IFN**) with **c** (**Met-Phe-IFN**) and **d** (**Met-IFN-b_5-chimera**) with

e (**Met-Phe-IFN-b₅-chimera)** in Table 1. We cannot tell whether the enhancement of the activity is due to stabilization of the N-terminal of the cytokine, by having a peptidase resistant motif, or that the Met-Phe sequence improves receptor binding. However, that the former may be the case is indicated by the comparison of the relative activities of proteins carrying the interferon domain either at the N- or C-terminal. Construct **6** containing the b₅ appendage at the N-terminal is nearly 2.2 times more active than IFN with b₅ at the C-terminal. Such a behavior is likely explained by b₅ protecting the interferon from degradation initiated from its N-terminus.

4. Materials and Methods

4.1 Construction of Interferon derivatives

Interferon derivatives were engineered using two plasmids: **pET-*met-ifn*,** which contained the human interferon α-2b gene, cloned in the pET-21a expression vector [1] and a pT7-7 derivative, harbouring the complete gene for the membrane-bound form of human cytochrome b₅ (a gift from Mrs Monika Akhtar, Biological Sciences, University of Southampton, UK). T7-promoter and ChimR2 primers (Table S1, supplementary) were used for PCR amplification of the soluble domain of cytochrome b₅. As illustrated in Figure S1, fusion of interferon and b₅ domains were achieved by placing a XhoI site in between the two domains thus making it possible to stitch the interferon DNA either N- or C-terminal to the cytochrome b₅ DNA. This feature added two extra residues i.e. Leu Glu, at the junction of the two domains in the protein sequence (see **4-6** Figure 1). The derivatives containing Phe as the second residue, for example **3** and **5** (Figure 1), were constructed using Phe-F1 primer (Table S2, Supplementary). pTZ-57 R/T vector was used for cloning PCR amplified DNA fragments and the pET-21a vector was used for expression in this study. The derivatives generated in this study are tabulated in Table S3, while Figures S1 A and B (Supplementary) summarize the generalized approach adopted to generate the chimeric interferon derivatives.

4.2 Expression and purification of interferon derivatives

E. coli BL21 codon plus® harboring the appropriate plasmids were grown and inclusion bodies isolated as described previously [1]. It was found that solubilization of inclusion bodies at pH 11 had a beneficial effect on the proper folding and yield of native **IFN** as well as **Met-Phe-IFN** (**1a** and **3** Figure 1 respectively). Following is illustrative of the refolding procedure used for these constructs.

Inclusion bodies were suspended in a small volume of solubilization buffer (8 M urea, 50 mM Tris-HCl pH 8.0, 50 mM glycine), the mixture sonicated on ice, pH adjusted to pH-11.5 with NaOH to give a clear solution and volume made up to 5-7 mg of protein/ml using the solublization buffer. To the latter 4 mM DTT was added and the mixture incubated at 37 °C for 30 minutes to allow complete reduction of disulfide bridges, while the pH remained 11.5.

25 ml of the solubilized protein above (5 mg of protein /ml) was transferred to the refolding sink (250 ml, 100 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.5 mM cysteine, 5 mM cystine, 0.1 mM PMSF) at a rate of 2.5 ml per hour; the final protein concentration in the refolding sink at the end of 10 hours was 0.5 mg/ml and the urea concentration was 0.8 M ±0.2 M. During the latter operation, the reaction was gently stirred and maintained at 4^0C; after 10 hours all the thiols had been lost but the pH remained at 8.0. The refolded sample was dialyzed against 20 mM Tris-HCl pH 8.0, with 4 buffer changes, each after 6 hours, and then the protein was quantified by measuring absorbance at 280 nm.

4.3 Q-sepharose column purification

Q-FF pre-filled 1 ml or 5 ml Q-sepharose columns (G.E, Healthcare®) were used depending on the amount of protein being purified. All the solutions were passed through the column with the help of a peristaltic pump (P1- G.E, Healthcare®). The column (washed with 10 column volumes of distilled water) was equilibrated with 5 column volumes of 20 mM Tris-HCl pH 8.0. Refolded protein solution was passed through Whatman filter paper 1 by gravitational flow. This solution was loaded on the pre-equilibrated column at the rate of 1 ml/min and the flow through was collected in bulk in a flask. At the completion of protein loading, the column was washed again with the same buffer (20 mM Tris–HCl pH 8.0) until the OD at 280 nm of the flow-through was zero. Protein was eluted with a step-wise gradient of 0.1 to 1 M NaCl. 5 ml of each concentration was passed through the column. The protein profile in the collected fractions was monitored by SDS-PAGE and the fractions containing the desired protein combined, dialyzed against 0.1% TFA, in case of **IFN** and **Met-Phe-IFN** or 5 mM Tris-HCl pH 8.0, in case of all chimeras. The dialyzed protein was lyophilized and stored at -20˚C. The proteins were analyzed by HPLC and mass spectrometry. The yield of purified product varied amongst different constructs, ranging from 30 to 60 percent of the starting inclusion bodies. For the identification of properly refolded proteins, tryptic digestion of various interferon constructs was performed as described in [1] and MALDI mass spectrometric analysis of fragment containing correct disulphide bridges was performed in linear positive mode as described previously [19].

4.4 Bioassays for anti-viral activity of IFN and its derivatives

In order to measure interferon activity, we measured inhibition of VSV-G pseudotyped HIV-1 derived vector particles encoding GFP. The particles were generated in HEK-293T cells (described in Supplementary Materials) , using a three plasmid system comprising pMDG (encoding VSV-G envelope), p8.91 (encoding HIV-1 Gag-pol, Tat and Rev) and pCSGW (HIV-1 genome plasmid encoding GFP) [18]. THP-1 cells were infected with the pseudotyped particles in the presence or absence of the drug and infected cells measured by flow cytometry for GFP expression 72 hours later as detailed below.

THP-1 cells were grown to 70-80% confluence in RPMI -1640 medium containing 10% FBS. For each of the test sample a 12 well plate was used and the respective interferon sample (dissolved in solubilizing buffer comprising: NaCl, 7.2 mg/ml; Tween 20, 0.2 mg/ml; ammonium acetate 0.77mg/ml) was ten-fold serially diluted with DMEM, starting from 374 to 0.0374 fmol/ml (1000 U/ml to 0.1 U/ml in case of RoferonA), in 5 consecutive wells and one well left untreated with medium only. The aforementioned concentrations were used making the assumption that the specific activities of the test compounds on per mole basis will be similar to that of RoferonA (1.4×108 Units/mg or 2.66×1015 IU/mol) and the duplicate dilutions were made in next 6 wells. In each well 10^5 cells were seeded and incubated overnight at 37 °C with 5% CO_2 to provide ample time for the induction of antiviral state; the volume at this stage was 1000-900 µl. Next day 0.1 ml of the supernatant, containing VSV-G pseudotyped HIV-1 vector with a multiplicity of infection (MOI) of 0.3 (multiplicity of infection: ratio of infectious agent to infection target: in this case the conditions that infected 33% of the target cells), was added to each well along with 1µl of 10 mg/ml polybrene solution (final conc. 10µg/ml). Culture was incubated and cells harvested 72 hours post infection by centrifugation at 1200 x g for 5 minutes. The cells were fixed with para-formaldehye by suspending the pellet in 4% para-formaldehyde in PBS. The number of GFP positive cells per 10000 counted cells was determined by FACS-analysis. The decrease in the number of GFP positive cells on drug treatment was taken as measure of the antiviral activity of interferon and its derivatives.

The activity units of interferon derivatives were determined by comparison with the activity of a standard interferon (RoferonA/interferon α-2b), determined in parallel, as shown in Figures 6 (also see Figure S5, supplementary material). The IC_{50} of a test derivative was taken to be equal to the number of units of RoferonA which produced 50% inhibition. For example, Figure 6 shows that the inhibition caused by 0.0071 ng (0.374 fmol) of Met-IFN, prepared in the present study, was

the same as 1 unit of RoferonA. Therefore, 0.374 fmol of Met-IFN was equal to 1 unit of interferon (giving a specific activity, 1.4×10^8 IU/mg and 2.66×10^{15} IU/mol) and its IC_{50} equal to 0.374fmol/ml (that is 0.374 pM). Similar equivalence was used to calculate the activity, in Units, of other derivatives and their IC_{50}values obtained using an online web-resource "IC_{50} Tool Kit" and the "GraphPad Prism7®" software (Fig.7 and Table 1).

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: FA did the experimental work and helped in the preparation of the manuscript, QAG did the mass spectrometric analysis, its interpretation and graphics, NR provided the molecular biology expertise, GJT provided the reagents for pseudotyped virus and supervised its use for interferon bioassay as well as edited the manuscript and MA designed the project and wrote the paper.

1

REFERENCES

[1] F. Ahsan, A. Arif, N. Mahmood, Q.A. Gardner, N. Rashid, M. Akhtar, Characterization and bioassay of post-translationally modified interferon alpha-2b expressed in Escherichia coli, J Biotechnol 184 (2014) 11-16.

[2] A. Isaacs, J. Lindenmann, Virus interference. I. The interferon, Proceedings of the Royal Society of London. Series B-Biological Sciences 147(927) (1957) 258-267.

[3] S. Pestka, [1] Interferon from 1981 to 1986, in: P. Sidney (Ed.), Methods in Enzymology, Academic Press 1986, pp. 3-14.

[4] J.M. Harris, R.B. Chess, Effect of pegylation on pharmaceuticals, Nat Rev Drug Discov 2(3) (2003) 214-221.

[5] S. Balan, J.-w. Choi, A. Godwin, I. Teo, C.M. Laborde, S. Heidelberger, M. Zloh, S. Shaunak, S. Brocchini, Site-specific PEGylation of protein disulfide bonds using a three-carbon bridge, Bioconjugate Chemistry 18(1) (2006) 61-76.

[6] B. Spolaore, S. Raboni, A.A. Satwekar, A. Grigoletto, A. Mero, I.M. Montagner, A. Rosato, G. Pasut, A. Fontana, Site-specific transglutaminase-mediated conjugation of interferon alpha-2b at glutamine or lysine Residues, Bioconjug Chem 27(11) (2016) 2695-2706.

[7] G.M. Subramanian, M. Fiscella, A. Lamouse-Smith, S. Zeuzem, J.G. McHutchison, Albinterferon alpha-2b: a genetic fusion protein for the treatment of chronic hepatitis C, Nat Biotechnol 25(12) (2007) 1411-1419.

[8] V.G. Bain, K.D. Kaita, E.M. Yoshida, M.G. Swain, E.J. Heathcote, A.U. Neumann, M. Fiscella, R. Yu, B.L. Osborn, P.W. Cronin, W.W. Freimuth, J.G. McHutchison, G.M. Subramanian, A phase 2 study to evaluate the antiviral activity, safety, and pharmacokinetics of recombinant human albumin-interferon alfa fusion protein in genotype 1 chronic hepatitis C patients, J Hepatol 44(4) (2006) 671-678.

[9] A. Arif, Studies on the preparation of interferon α-2b and removal of its N-terminal methionine using methionine aminopeptidases, School of Biological Sciences, University of the Punjab, Lahore, 2014.

[10] B. Polevoda, F. Sherman, Nα-terminal acetylation of eukaryotic proteins, Journal of Biological Chemistry 275(47) (2000) 36479-36482.

[11] Q.A. Gardner, H. Younas, N. Rashid, J.N. Wright, M. Akhtar, Inventory of 'slow exchanging' hydrogen atoms in human proinsulin and its derivatives: observations on the mass spectrometric analysis of deuterio-proteins in D(2)O, Biochim Biophys Acta 1794(8) (2009) 1224-33.

[12] F. Aslam, Q.A. Gardner, H. Zain, M.S. Nadeem, M. Ali, N. Rashid, M. Akhtar, Studies on the expression and processing of human proinsulin derivatives encoded by different DNA constructs, Biochim Biophys Acta 1834(10) (2013) 2116-2123.

[13] F. Kratz, Albumin as a drug carrier: design of prodrugs, drug conjugates and nanoparticles, Journal of controlled release 132(3) (2008) 171-183.

[14] K. Park, Albumin: a versatile carrier for drug delivery, J Control Release 157(1) (2012) 3.

[15] Y.-S. Huang, Z. Chen, Z.-Y. Yang, T.-Y. Wang, L. Zhou, J.-B. Wu, L.-F. Zhou, Preparation and characterization of a potent, long-lasting recombinant human serum albumin-interferon-α2b fusion protein expressed in Pichia pastoris, European Journal of Pharmaceutics and Biopharmaceutics 67(2) (2007) 301-308.

[16] V.K. Rustgi, Albinterferon alfa-2b, a novel fusion protein of human albumin and human interferon alfa-2b, for chronic hepatitis C, Curr Med Res Opin 25(4) (2009) 991-1002.

[17] C. Sung, B. Nardelli, D.W. LaFleur, E. Blatter, M. Corcoran, H.S. Olsen, C.E. Birse, O.K. Pickeral, J. Zhang, D. Shah, An IFN-β-albumin fusion protein that displays improved pharmacokinetic and pharmacodynamic properties in nonhuman primates, Journal of Interferon & Cytokine Research 23(1) (2003) 25-36.

[18] C. Besnier, Y. Takeuchi, G. Towers, Restriction of lentivirus in monkeys, Proc Natl Acad Sci U S A 99(18) (2002) 11920-11925.

[19] Q.A. Gardner, H. Younas, M. Akhtar, Studies on the regioselectivity and kinetics of the action of trypsin on proinsulin and its derivatives using mass spectrometry, Biochim Biophys Acta 1834(1) (2013) 182-190.