VACCINE FOR PREVENTION AND TREATMENT OF HIV-INFECTION

Field of the Invention

The present invention relates to novel HIV polypeptide constructs, to their use in medicine, to pharmaceutical compositions comprising them and to methods for their manufacture. The invention also relates to polynucleotides encoding the polypeptides. In particular, the invention relates to fusion proteins comprising HIV-1 Nef and HIV-1 Gag or fragments thereof, and to polynucleotides encoding them. More particularly, the invention relates to fusion proteins comprising HIV-1 Nef, HIV-1 Pol and HIV-1 Gag proteins or fragments thereof and to polynucleotides encoding them.

HIV-1 is the primary cause of the acquired immune deficiency syndrome (AIDS) which is regarded as one of the world's major health problems. There is a need for a vaccine for the prevention and/or treatment of HIV infection.

Background to the Invention

HIV-1 is an RNA virus of the family Retroviridiae. The HIV genome encodes at least nine proteins which are divided into three classes: the major structural proteins Gag, Pol and Env, the regulatory proteins Tat and Rev, and the accessory proteins Vpu, Vpr, Vif and Nef. The HIV genome exhibits the 5'LTR-gag-pol-env-LTR3' organization of all retroviruses.

The HIV envelope glycoprotein gp120 is the viral protein that is used for attachment to the host cell. This attachment is mediated by binding to two surface molecules of helper T cells and macrophages, known as CD4 and one of the two chemokine receptors CCR-5 or CXCR-4. The gp120 protein is first expressed as a larger precursor molecule (gp160), which is then cleaved post-translationally to yield gp120 and gp41. The gp120 protein is retained on the surface of the virion by linkage to the gp41 molecule, which is inserted into the viral membrane.

The gp120 protein is the principal target of neutralizing antibodies, but unfortunately the most immunogenic regions of the proteins (V3 loop) are also the most variable parts of the protein. Therefore, the use of gp120 (or its precursor gp160) as a vaccine antigen to elicit neutralizing antibodies is thought to be of limited use for a broadly protective vaccine. The gp120 protein does also contain epitopes that are recognized by cytotoxic T lymphocytes (CTL). These effector cells are able to eliminate virus-infected cells, and therefore constitute a second major antiviral immune mechanism. In contrast to the target regions of neutralizing antibodies some CTL epitopes appear to be relatively conserved among different HIV strains. For this reason gp120 and gp160 maybe useful antigenic components in vaccines that aim at eliciting cell-mediated immune responses (particularly CTL).

Non-envelope proteins of HIV-1 include for example internal structural proteins such as the products of the gag and pol genes and other non-structural proteins such as Rev, Nef, Vif and Tat (Green et al., New England J. Med, 324, 5, 308 et seq (1991) and Bryant et al. (Ed. Pizzo), Pediatr. Infect. Dis. J., 11, 5, 390 et seq (1992).

HIV Nef is an early protein,that is it is expressed early in infection and in the absence of structural protein.

The Nef gene encodes an early accessory HIV protein which has been shown to possess several activities. For example, the Nef protein is known to cause the down regulation of CD4, the HIV receptor, and MHC class I molecules from the cell surface, although the biological importance of these functions is debated. Additionally Nef interacts with the signal pathway of T cells and induces an active state, which in turn may promote more efficient gene expression. Some HIV isolates have mutations in this region, which cause them not to encode functional protein and are severely compromised in their replication and pathogenesis in vivo.

The Gag gene is translated as a precursor polyprotein that is cleaved by protease to yield products that include the matrix protein (p17), the capsid (p24), the nucleocapsid (p9), p6 and two space peptides, p2 and p1.

The Gag gene gives rise to the 55-kilodalton (kD) Gag precursor protein, also called p55, which is expressed from the unspliced viral mRNA. During translation, the N terminus of p55 is myristoylated, triggering its association with the cytoplasmic aspect of cell membranes. The membrane-associated Gag polyprotein recruits two copies of the viral genomic RNA along with other viral and cellular proteins that triggers the budding of the viral particle from the surface of an infected cell. After budding, p55 is cleaved by the virally encoded protease (a product of the pol gene) during the process of viral maturation into four smaller proteins designated MA (matrix [p17]), CA (capsid [p24]), NC (nucleocapsid [p9]), and p6.

In addition to the 3 major Gag proteins, all Gag precursors contain several other regions, which are cleaved out and remain in the virion as peptides of various sizes. These proteins have different roles e.g. the p2 protein has a proposed role in regulating activity of the protease and contributes to the correct timing of proteolytic processing.

The p17 (MA) polypeptide is derived from the N-terminal, myristoylated end of p55. Most MA molecules remain attached to the inner surface of the virion lipid bilayer, stabilizing the particle. A subset of MA is recruited inside the deeper layers of the virion where it becomes part of the complex which escorts the viral DNA to the nucleus. These MA molecules facilitate the nuclear transport of the viral genome because a karyophilic signal on MA is recognized by the cellular nuclear import machinery. This phenomenon allows HIV to infect non-dividing cells, an unusual property for a retrovirus.

The p24 (CA) protein forms the conical core of viral particles. Cyclophilin A has been demonstrated to interact with the p24 region of p55 leading to its incorporation into HIV particles. The interaction between Gag and cyclophilin A is essential because the disruption of this interaction by cyclosporin A inhibits viral replication.

The NC region of Gag is responsible for specifically recognizing the so-called packaging signal of HIV. The packaging signal consists of four stem loop structures located near the 5' end of the viral RNA, and is sufficient to mediate the incorporation of a heterologous RNA into HIV-1 virions. NC binds to the packaging signal through interactions mediated by two zinc-finger motifs. NC also facilitates reverse transcription.

The p6 polypeptide region mediates interactions between p55 Gag and the accessory protein Vpr, leading to the incorporation of Vpr into assembling virions. The p6 region also contains a so-called late domain which is required for the efficient release of budding virions from an infected cell.

The Pol gene encodes two proteins containing the two activities needed by the virus in early infection, the RT and the integrase protein needed for integration of viral DNA into cell DNA. The primary product of Pol is cleaved by the virion protease to yield the amino terminal RT peptide which contains activities necessary for DNA synthesis (RNA and DNA-dependent DNA polymerase activity as well as an RNase H function) and carboxy terminal integrase protein. HIV RT is a heterodimer of full-length RT (p66) and a cleavage product (p51) lacking the carboxy terminal RNase H domain.

RT is one of the most highly conserved proteins encoded by the retroviral genome. Two major activities of RT are the DNA Pol and Ribonuclease H. The DNA Pol activity of RT uses RNA and DNA as templates interchangeably and like all DNA polymerases known is unable to initiate DNA synthesis de novo, but requires a pre existing molecule to serve as a primer (RNA).

The RNase H activity inherent in all RT proteins plays the essential role early in replication of removing the RNA genome as DNA synthesis proceeds. It selectively degrades the RNA from all RNA - DNA hybrid molecules. Structurally the polymerase and ribo H occupy separate, non-overlapping domains with the Pol covering the amino two thirds of the Pol.

The p66 catalytic subunit is folded into 5 distinct subdomains. The amino terminal 23 of these have the portion with RT activity. Carboxy terminal to these is the RNase H Domain.

WO 03/025003 describes DNA constructs encoding HIV-1 p17/24 Gag, Nef and RT, wherein the DNA sequences may be codon optimized to resemble highly expressed human genes. These constructs are useful in DNA vaccines.

Fusion proteins containing multiple HIV antigens have been suggested as vaccine candidates for HIV, for example the Nef-Tat fusion as described in WO 99/16884. However, fusion proteins are not straightforward to produce; there can be difficulties in expressing them because they do not correspond to native proteins. There can be difficulties at the transcription level, or further downstream. Also, they may not be straightforward to formulate into a pharmaceutically acceptable composition. Notably, the majority of approaches to HIV vaccines that involve multiple antigens fused together, are DNA or live vector approaches rather than polypeptide fusion proteins.

Summary of the Invention

The present invention provides novel constructs for use in vaccines for the prophylaxis and treatment of HIV infections and AIDS.

In one aspect the invention provides a polypeptide which comprises Nef or an immunogenic fragment or derivative thereof, and p17 Gag and/or p24 Gag or immunogenic fragments or derivatives thereof, wherein when both p17 and p24 Gag are present there is at least one HIV antigen or immunogenic fragment between them.

In the constructs and compositions according to the invention as described herein, the Nef is preferably a full length Nef.

In the constructs according to the invention the p17 Gag and p24 Gag are preferably full length p17 and p24 respectively.

In one embodiment the polypeptide comprises both p17 and p24 Gag or immunogenic fragments thereof. In such a construct the p24 Gag component and p17 Gag component are separated by at least one further HIV antigen or immunogenic fragment, such as Nef and/or RT or immunogenic fragments or derivatives thereof.

Alternatively p17 or p24 Gag may be provided separately. Thus the invention also provides a composition comprising (i) a polypeptide which comprises Nef or an immunogenic fragment or derivative thereof and p17 Gag or an immunogenic fragment or derivative thereof, and (ii) p24 Gag or an immunogenic fragment or derivative thereof; or (i) a polypeptide which comprises Nef or an immunogenic fragment or derivative thereof and p24 Gag or an immunogenic fragment or derivative thereof, and (ii) p17 Gag or an immunogenic fragment or derivative thereof.

In another embodiment the polypeptide construct according to the invention further comprises Pol or a derivative of Pol such as RT or an immunogenic fragment or derivative thereof. Particular fragments of RT that are suitable for use in the invention are fragments in which the RT is truncated at the C terminus, preferably such that they lack the carboxy terminal RNase H domain. One such fragment lacking the carboxy terminal Rnase H domain is the p51 fragment described herein.

Preferably the RT or immunogenic fragment in the fusion proteins described herein is p66 RT or p51 RT.

The RT component of the fusion protein or composition according to the invention optionally comprises a mutation at position 592, or equivalent mutation in strains other than HXB2, such that the methionine is removed by mutation to another residue e.g. lysine. The purpose of this mutation is to remove a site which serves as an internal initiation site in prokaryotic expression systems.

The RT component also, or alternatively, comprises a mutation to remove the enzyme activity (reverse transcriptase). Thus K231 may be present instead of W.

In fusion proteins according to the invention which comprise p24 and RT, it may be preferable that the p24 precedes the RT in the construct because when the antigens are expressed alone in *E. coli* better expression of p24 than of RT is observed.

Preferred constructs according to the invention include the following:

1. p24 – RT – Nef – p17
2. p24 – RT* – Nef – p17
3. p24 – p51RT – Nef – p17
4. p24 – p51RT* - Nef – p17
5. p17 – p51RT – Nef
6. p17 – p51RT* - Nef
7. Nef – p17
8. Nef – p17 with linker
9. p17 – Nef
10. p17 – Nef with linker

* represents RT methionine$_{592}$ mutation to lysine

The linker included in the constructs listed above may be any short amino acid sequence for decreasing potential interactions between the two fusion partners that it links together. The linker may be for example from 4-10 amino acids in length. For example, it may be a 6 amino acid sequence such as the GSGGGP sequence described herein in the examples.

In another aspect the present invention provides a fusion protein of HIV antigens comprising at least four HIV antigens or immunogenic fragments, wherein the four antigens or fragments are or are derived from Nef, Pol and Gag. Preferably Gag is

present as two separate components which are separated by at least one other antigen in the fusion. Preferably the Nef is full length Nef. Preferably the Pol is p66 or p51RT. Preferably the Gag is p17 and p24 Gag. Other preferred features and properties of the antigen components of the fusion in this aspect of the invention are as described herein.

Preferred embodiments of this aspect of the invention are the four component fusions as already listed above:

1. p24 – RT – Nef – p17
2. p24 – RT* – Nef – p17
3. p24 – p51RT – Nef – p17
4. p24 – p51RT* - Nef – p17

The term "derived from" or "derivative" in relation to the HIV antigens included in the invention means that the antigens may have been altered in a limited way compared to their native counterparts. This includes point mutations which may change the properties of the protein for example by improving expression in prokaryotic systems or removing undesirable activity including undesirable enzyme activity. The point mutations described herein for RT are designed to achieve these things. However, the antigens must remain sufficiently similar to the native antigens such that they retain the antigenic properties desirable in a vaccine and thus they remain capable of raising an immune response against the native antigen. Whether or not a particular derivative raises such an immune response may be measured by a suitable immunological assay such as an ELISA (for antibody responses) or flow cytometry using suitable staining for cellular markers and cytokines (for cellular responses).

The polypeptide constructs of HIV antigens according to the invention are capable of being expressed in *in vitro* systems including prokaryotic systems such as *E. coli*. Advantageously they can be purified by conventional purification methods.

The fusions described herein are preferably soluble when expressed in a selected expression system, that is they are present in a substantial amount in the supernatant of a

crude extract from the expression system. The presence of the fusion protein in the crude extract can be measured by conventional means such as running on an SDS gel, coomassie staining and checking the appropriate band by densitometric measurement. Fusion proteins according to the invention are preferably at least 50% soluble, more preferably at least 70% soluble, most preferably 90% soluble or greater as measured by the techniques described herein in the Examples. Techniques to improve solubility of recombinantly expressed proteins are known, for example in prokaryotic expression systems solubility is improved by lowering the temperature at which gene expression is induced.

The fusion proteins described herein can be purified. In particular they can be purified while remaining soluble or significantly soluble.

Immunogenic fragments as described herein will contain at least one epitope of the antigen and display HIV antigenicity and are capable of raising an immune response when presented in a suitable construct, such as for example when fused to other HIV antigens or presented on a carrier, the immune response being directed against the native antigen. Typically the immunogenic fragments contain at least 20, preferably 50, more preferably 100 contiguous amino acids from the HIV antigen.

The invention provides in a further aspect polynucleotides encoding the polypeptides according to the invention.

Polynucleotides according to the invention may be used as polynucleotide vaccines. The polynucleotides may be present within any of a variety of delivery systems known to those of ordinary skill in the art, including nucleic acid expression systems such as plasmid DNA, bacterial and viral expression systems. Numerous gene delivery techniques are well known in the art, such as those described by Rolland, Crit. Rev. Therap. Drug Carrier Systems 15:143-198, 1998 and references cited therein. Appropriate nucleic acid expression systems contain the necessary DNA sequences for expression in the patient (such as a suitable promoter and terminating signal). When the

expression system is a recombinant live microorganism, such as a virus or bacterium, the gene of interest can be inserted into the genome of the live recombinant virus or bacterium. Inoculation and *in vivo* infection with this live vector will lead to *in vivo* expression of the antigen and induction of immune responses. Viruses and bacteria used for this purpose are for instance: poxviruses (e.g; vaccinia, fowlpox, canarypox, modified poxviruses e.g. Modified Virus Ankara (MVA)), alphaviruses (Sindbis virus, Semliki Forest Virus, Venezuelian Equine Encephalitis Virus), flaviviruses (yellow fever virus, Dengue virus, Japanese encephalitis virus), adenoviruses, adeno-associated virus, picornaviruses (poliovirus, rhinovirus), herpesviruses (varicella zoster virus, etc), morbilliviruses (e.g. measles), Listeria, Salmonella , Shigella, Neisseria, BCG. These viruses and bacteria can be virulent, or attenuated in various ways in order to obtain live vaccines. Such live vaccines also form part of the invention.

A preferred measles vector for use as a live vector according to the invention is the Schwartz strain or a strain derived therefrom.

A preferred adenovirus for use as a live vector is a low sero-prevalent human adenovirus such as Ad5 or Ad35 or a non-human originating adenovirus such as a non-human primate adenovirus such as a simian adenovirus. Such low sero-prevalent human or similar adenoviruses will have less than 60, typically less than 50% sero-prevelance in the population. Preferably the vectors are replication defective. Typically these viruses contain an E1 deletion and can be grown on cell lines that are transformed with an E1 gene. Preferred simian adenoviruses are viruses isolated from chimpanzee. In particular C68 (also known as Pan 9) (See US patent No 6083 716) and Pan 5, 6 and Pan 7 (WO 03/046124) are preferred for use in the present invention. These vectors can be manipulated to insert a heterologous polynucleotide according to the invention such that the polypeptides according to the invention maybe expressed. The use, formulation and manufacture of such recombinant adenoviral vectors is described in detail in WO 03/046142.

Thus, the Nef, p17 and p24 Gag and RT of a preferred vaccine according to the invention may be provided in the form of a polynucleotide encoding the desired polypeptide.

Polynucleotides according to the invention may be used to express the encoded polypeptides in a selected expression system. At least one of the HIV antigens, for example the RT, may be encoded by a codon optimized sequence in the polynucleotide, that is to say the sequence has been optimized for expression in a selected recombinant expression system such as *E. coli.*

In another aspect the invention provides a p51 RT polypeptide or derivative thereof or a polynucleotide encoding it, preferably codon optimized for expression in a suitable expression system, particularly a prokaryotic system such as *E. coli.*

The p51 RT polypeptide or polynucleotide may be used alone, or in combination with a polypeptide or polynucleotide construct according to the invention. Thus in a further aspect the invention provides a composition comprising (i) a polypeptide which comprises Nef or a fragment containing a Nef epitope and p17 Gag and/or p24 Gag, wherein when both p17 and p24 Gag are present there is at least one HIV antigen or immunogenic fragment between them and (ii) a p51 RT polypeptide. The invention further provides polynucleotides encoding these.

According to this embodiment (i) may be selected from for example:

1. Nef – p17
2. Nef – p17 with linker
3. p17 – Nef
4. p17 – Nef with linker

Preferably Nef is full length Nef. Preferably p17 is full length p17.

The polypeptides and polynucleotides according to the invention may be combined with other antigens or polynucleotides encoding other antigens. In particular, this may include HIV env proteins or fragments or derivatives thereof. Preferred forms of env are gp120,

gp140 and gp160. The env may be for example the envelope protein described in WO 00/07631 from an HIV-1 clade B envelope clone known as R2, or a fragment or derivative thereof. Thus the invention further provides a composition comprising any of the polypeptides or polypeptide compositions according to the invention, together with an HIV env protein or fragment or derivative thereof. Similarly the invention provides a composition comprising a polynucleotide or polynucleotides encoding a polypeptides or polypeptides according to the invention and a polynucleotide encoding an HIV env protein or fragment or derivative thereof.

The invention further provides methods of preparing the polypeptides described herein which method comprises expressing a polynucleotide encoding the polypeptide in a suitable expression system, particularly a prokaryotic system such as *E. coli* and recovering the expressed polypeptide. Preferably expression is induced at a low temperature, that is a temperature below 37°, to promote the solubility of the polypeptide.

The invention further provides a process for purifying a polypeptide as described herein, which process comprises:

i). providing a composition comprising the unpurified polypeptide;

ii). Subjecting the composition to at least two chromatographic steps;

iii). Optionally carboxyamidating the polypeptide;

iv) Performing a buffer exchange step to provide the protein in a suitable buffer for a pharmaceutical formulation.

The carboxyamidation may be performed between the two chromatographic steps. The carboxyamidation step may be performed using iodoacetimide.

In one example, the process according to the invention uses no more than two chromatographic steps.

The invention further provides pharmaceutical compositions and immunogenic compositions and vaccines comprising the polypeptides and polynucleotides according to the invention, in combination with a pharmaceutically acceptable adjuvant or carrier.

Vaccines according to the invention may be used for prophylactic or therapeutic immunization against HIV.

The invention further provides the use of the polypeptides and polypeptide compositions and the polynucleotides and polynucleotide compositions as described herein, in the manufacture of a vaccine for prophylactic or therapeutic immunization against HIV.

The vaccine of the present invention will contain an immunoprotective or immunotherapeutic quantity of the polypeptide and/or polynucleotide antigens and may be prepared by conventional techniques.

Vaccine preparation is generally described in New Trends and Developments in Vaccines, edited by Voller et al., University Park Press, Baltimore, Maryland, U.S.A. 1978. Encapsulation within liposomes is described, for example, by Fullerton, U.S. Patent 4,235,877. Conjugation of proteins to macromolecules is disclosed, for example, by Likhite, U.S. Patent 4,372,945 and by Armor et al., U.S. Patent 4,474,757.

The amount of protein in the vaccine dose is selected as an amount which induces an immunoprotective response without significant, adverse side effects in typical vaccinees. Such amount will vary depending upon which specific immunogen is employed and the vaccination regimen that is selected. Generally, it is expected that each dose will comprise 1-1000 μg of each protein, preferably 2-200 μg, most preferably 4-40 μg of the polypeptide fusion. An optimal amount for a particular vaccine can be ascertained by standard studies involving observation of antibody titres and other immune responses in subjects. Following an initial vaccination, subjects may receive a boost in about 4 weeks, and a subsequent second booster immunisation.

The proteins of the present invention are preferably adjuvanted in the vaccine formulation of the invention. Adjuvants are described in general in Vaccine Design – the Subunit and Adjuvant Approach, edited by Powell and Newman, Plenum Press, New York, 1995.

Suitable adjuvants include an aluminium salt such as aluminium hydroxide or aluminium phosphate, but may also be a salt of calcium, iron or zinc, or may be an insoluble suspension of acylated tyrosine, or acylated sugars, cationically or anionically derivatised polysaccharides, or polyphosphazenes.

In the formulation of the invention it is preferred that the adjuvant composition induces a preferential Th1 response. However it will be understood that other responses, including other humoral responses, are not excluded.

An immune response is generated to an antigen through the interaction of the antigen with the cells of the immune system. The resultant immune response may be broadly distinguished into two extreme catagories, being humoral or cell mediated immune responses (traditionally characterised by antibody and cellular effector mechanisms of protection respectively). These categories of response have been termed Th1-type responses (cell-mediated response), and Th2-type immune responses (humoral response).

Extreme Th1-type immune responses may be characterised by the generation of antigen specific, haplotype restricted cytotoxic T lymphocytes, and natural killer cell responses. In mice Th1-type responses are often characterised by the generation of antibodies of the IgG2a subtype, whilst in the human these correspond to IgG1 type antibodies. Th2-type immune responses are characterised by the generation of a broad range of immunoglobulin isotypes including in mice IgG1, IgA, and IgM.

It can be considered that the driving force behind the development of these two types of immune responses are cytokines, a number of identified protein messengers which serve to help the cells of the immune system and steer the eventual immune response to either a Th1 or Th2 response. Thus high levels of Th1-type cytokines tend to favour the

induction of cell mediated immune responses to the given antigen, whilst high levels of Th2-type cytokines tend to favour the induction of humoral immune responses to the antigen.

It is important to remember that the distinction of Th1 and Th2-type immune responses is not absolute. In reality an individual will support an immune response which is described as being predominantly Th1 or predominantly Th2. However, it is often convenient to consider the families of cytokines in terms of that described in murine CD4 +ve T cell clones by Mosmann and Coffman (*Mosmann, T.R. and Coffman, R.L. (1989) TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. Annual Review of Immunology, 7, p145-173*). Traditionally, Th1-type responses are associated with the production of the INF-γ and IL-2 cytokines by T-lymphocytes. Other cytokines often directly associated with the induction of Th1-type immune responses are not produced by T-cells, such as IL-12. In contrast, Th2- type responses are associated with the secretion of IL-4, IL-5, IL-6, IL-10 and tumour necrosis factor-β(TNF-β).

It is known that certain vaccine adjuvants are particularly suited to the stimulation of either Th1 or Th2 - type cytokine responses. Traditionally the best indicators of the Th1:Th2 balance of the immune response after a vaccination or infection includes direct measurement of the production of Th1 or Th2 cytokines by T lymphocytes *in vitro* after restimulation with antigen, and/or the measurement of the IgG1:IgG2a ratio of antigen specific antibody responses.

Thus, a Th1-type adjuvant is one which stimulates isolated T-cell populations to produce high levels of Th1-type cytokines when re-stimulated with antigen *in vitro*, and induces antigen specific immunoglobulin responses associated with Th1-type isotype.

Preferred Th1-type immunostimulants which may be formulated to produce adjuvants suitable for use in the present invention include and are not restricted to the following.

Monophosphoryl lipid A, in particular 3-de-O-acylated monophosphoryl lipid A (3D-MPL), is a preferred Th1-type immunostimulant for use in the invention. 3D-MPL is a well known adjuvant manufactured by Ribi Immunochem, Montana. Chemically it is often supplied as a mixture of 3-de-O-acylated monophosphoryl lipid A with either 4, 5, or 6 acylated chains. It can be purified and prepared by the methods taught in GB 2122204B, which reference also discloses the preparation of diphosphoryl lipid A, and 3-O-deacylated variants thereof. Other purified and synthetic lipopolysaccharides have been described (US 6,005,099 and EP 0 729 473 B1; Hilgers *et al.*, 1986, *Int.Arch.Allergy.Immunol.*, 79(4):392-6; Hilgers *et al.*, 1987, Immunology, 60(1):141-6; and EP 0 549 074 B1). A preferred form of 3D-MPL is in the form of a particulate formulation having a small particle size less than 0.2μm in diameter, and its method of manufacture is disclosed in EP 0 689 454.

Saponins are also preferred Th1 immunostimulants in accordance with the invention. Saponins are well known adjuvants and are taught in: Lacaille-Dubois, M and Wagner H. (1996. A review of the biological and pharmacological activities of saponins. Phytomedicine vol 2 pp 363-386). For example, Quil A (derived from the bark of the South American tree Quillaja Saponaria Molina), and fractions thereof, are described in US 5,057,540 and "Saponins as vaccine adjuvants", Kensil, C. R., *Crit Rev Ther Drug Carrier Syst*, 1996, 12 (1-2):1-55; and EP 0 362 279 B1. The haemolytic saponins QS21 and QS17 (HPLC purified fractions of Quil A) have been described as potent systemic adjuvants, and the method of their production is disclosed in US Patent No. 5,057,540 and EP 0 362 279 B1. Also described in these references is the use of QS7 (a non-haemolytic fraction of Quil-A) which acts as a potent adjuvant for systemic vaccines. Use of QS21 is further described in Kensil *et al.* (1991. J. Immunology vol 146, 431-437). Combinations of QS21 and polysorbate or cyclodextrin are also known (WO 99/10008). Particulate adjuvant systems comprising fractions of QuilA, such as QS21 and QS7 are described in WO 96/33739 and WO 96/11711. One such system is known as an Iscorn and may contain one or more saponins.

Another preferred immunostimulant is an immunostimulatory oligonucleotide containing unmethylated CpG dinucleotides ("CpG"). CpG is an abbreviation for cytosine-guanosine dinucleotide motifs present in DNA. CpG is known in the art as being an adjuvant when administered by both systemic and mucosal routes (WO 96/02555, EP 468520, Davis *et al., J.Immunol*, 1998, 160(2):870-876; McCluskie and Davis, *J.Immunol.*, 1998, 161(9):4463-6). Historically, it was observed that the DNA fraction of BCG could exert an anti-tumour effect. In further studies, synthetic oligonucleotides derived from BCG gene sequences were shown to be capable of inducing immunostimulatory effects (both in vitro and in vivo). The authors of these studies concluded that certain palindromic sequences, including a central CG motif, carried this activity. The central role of the CG motif in immunostimulation was later elucidated in a publication by Krieg, Nature 374, p546 1995. Detailed analysis has shown that the CG motif has to be in a certain sequence context, and that such sequences are common in bacterial DNA but are rare in vertebrate DNA. The immunostimulatory sequence is often: Purine, Purine, C, G, pyrimidine, pyrimidine; wherein the CG motif is not methylated, but other unmethylated CpG sequences are known to be immunostimulatory and may be used in the present invention.

In certain combinations of the six nucleotides a palindromic sequence is present. Several of these motifs, either as repeats of one motif or a combination of different motifs, can be present in the same oligonucleotide. The presence of one or more of these immunostimulatory sequences containing oligonucleotides can activate various immune subsets, including natural killer cells (which produce interferon γ and have cytolytic activity) and macrophages (Wooldrige et al Vol 89 (no. 8), 1977). Other unmethylated CpG containing sequences not having this consensus sequence have also now been shown to be immunomodulatory.

CpG when formulated into vaccines, is generally administered in free solution together with free antigen (WO 96/02555; McCluskie and Davis, *supra*) or covalently conjugated to an antigen (WO 98/16247), or formulated with a carrier such as aluminium hydroxide

((Hepatitis surface antigen) Davis *et al. supra* ; Brazolot-Millan *et al.*, *Proc.Natl.Acad.Sci.*, USA, 1998, 95(26), 15553-8).

Such immunostimulants as described above may be formulated together with carriers, such as for example liposomes, oil in water emulsions, and or metallic salts, including aluminium salts (such as aluminium hydroxide). For example, 3D-MPL may be formulated with aluminium hydroxide (EP 0 689 454) or oil in water emulsions (WO 95/17210); QS21 may be advantageously formulated with cholesterol containing liposomes (WO 96/33739), oil in water emulsions (WO 95/17210) or alum (WO 98/15287); CpG may be formulated with alum (Davis *et al. supra* ; Brazolot-Millan *supra*) or with other cationic carriers.

Combinations of immunostimulants are also preferred, in particular a combination of a monophosphoryl lipid A and a saponin derivative (WO 94/00153; WO 95/17210; WO 96/33739; WO 98/56414; WO 99/12565; WO 99/11241), more particularly the combination of QS21 and 3D-MPL as disclosed in WO 94/00153. Alternatively, a combination of CpG plus a saponin such as QS21 also forms a potent adjuvant for use in the present invention. Alternatively the saponin may be formulated in a liposome or in an Iscorn and combined with an immunostimulatory oligonucleotide.

Thus, suitable adjuvant systems include, for example, a combination of monophosphoryl lipid A, preferably 3D-MPL, together with an aluminium salt.
An enhanced system involves the combination of a monophosphoryl lipid A and a saponin derivative particularly the combination of QS21 and 3D-MPL as disclosed in WO 94/00153, or a less reactogenic composition where the QS21 is quenched in cholesterol containing liposomes (DQ) as disclosed in WO 96/33739. This combination may additionally comprise an immunostimulatory oligonucleotide.

A particularly potent adjuvant formulation involving QS21, 3D-MPL & tocopherol in an oil in water emulsion is described in WO 95/17210 and is another preferred formulation for use in the invention.

Another preferred formulation comprises a CpG oligonucleotide alone or together with an aluminium salt.

In a further aspect of the present invention there is provided a method of manufacture of a vaccine formulation as herein described, wherein the method comprises admixing a polypeptide according to the invention with a suitable adjuvant.

Particularly preferred adjuvant combinations for use in the formulations according to the invention are as follows:

i) 3D-MPL + QS21 in a liposome

ii) Alum + 3D-MPL

iii) Alum + QS21 in a liposome + 3D-MPL

iv) Alum + CpG

v) 3D-MPL + QS21 + oil in water emulsion

vi) CpG

Administration of the pharmaceutical composition may take the form of one or of more than one individual dose, for example as repeat doses of the same polypeptide containing composition, or in a heterologous "prime-boost" vaccination regime. A heterologous prime-boost regime uses administration of different forms of vaccine in the prime and the boost, each of which may itself include two or more administrations. The priming composition and the boosting composition will have at least one antigen in common, although it is not necessarily an identical form of the antigen, it may be a different form of the same antigen.

Prime boost immunisations according to the invention may be performed with a combination of protein and DNA-based formulations. Such a strategy is considered to be effective in inducing broad immune responses. Adjuvanted protein vaccines induce mainly antibodies and T helper immune responses, while delivery of DNA as a plasmid or a live vector induces strong cytotoxic T lymphocyte (CTL) responses. Thus, the

combination of protein and DNA vaccination will provide for a wide variety of immune responses. This is particularly relevant in the context of HIV, since both neutralising antibodies and CTL are thought to be important for the immune defence against HIV.

In accordance with the invention a schedule for vaccination may comprise the sequential ("prime-boost") or simultaneous administration of polypeptide antigens and DNA encoding the polypeptides. The DNA may be delivered as naked DNA such as plasmid DNA or in the form of a recombinant live vector, e.g. a poxvirus vector, an adenovirus vector, a measles virus vector or any other suitable live vector. Protein antigens may be injected once or several times followed by one or more DNA administrations, or DNA may be used first for one or more administrations followed by one or more protein immunisations.

A particular example of prime-boost immunisation according to the invention involves priming with DNA in the form of a recombinant live vector such as a modified poxvirus vector, for example Modified Virus Ankara (MVA) or an alphavirus, for example Venezuelian Equine Encephalitis Virus, or an adenovirus vector, or a measles virus vector, followed by boosting with a protein, preferably an adjuvanted protein.

Thus the invention further provides a pharmaceutical kit comprising:

a) a composition comprising a polypeptide comprising Nef or an immunogenic fragment or derivative thereof and p17 and/or p24 Gag or an immunogenic fragment or derivative thereof, wherein when both p17 and p24 Gag are present there is at least one HIV antigen or immunogenic fragment or derivative between them, together with a pharmaceutically acceptable excipient; and

b) a composition comprising a polynucleotide encoding one or more of Nef and Gag or an immunogenic fragment or derivative of Nef or Gag containing a Nef or Gag epitope present in the polypeptide of a), together with a pharmaceutically acceptable excipient.

Preferably the polypeptide of a) further comprises RT or an immunogenic fragment or derivative thereof such as p51RT.

In an alternative embodiment the pharmaceutical kit comprises:

a) a composition comprising a polynucleotide encoding a polypeptide comprising Nef or an immunogenic fragment or derivative thereof and p17 and/or p24 Gag or an immunogenic fragment or derivative thereof, wherein when both p17 and p24 Gag are present there is at least one HIV antigen or immunogenic fragment or derivative between them, together with a pharmaceutically acceptable excipient; and

b) a composition comprising a polypeptide comprising one or more of Nef and Gag or an immunogenic fragment or derivative of Nef or Gag containing a Nef or Gag epitope present in the polypeptide of a), together with a pharmaceutically acceptable excipient.

Preferably the polynucleotide of a) encodes a polypeptide which further comprises RT or an immunogenic fragment or derivative thereof such as p51RT.

Preferred polypeptides and polynucleotides for use in a prime/boost kit according to the invention are the polypeptides and polynucleotides as described herein. Thus, the protein component of a protein/DNA type prime boost approach may be any of the preferred fusion proteins described herein. Likewise, the DNA component may be a polynucleotide encoding any of the preferred proteins.

Thus for example, the p24 – RT – Nef – p17, p24 – RT* – Nef – p17, p24 – p51RT – Nef – p17, p24 – p51RT* - Nef – p17, p17 – p51RT – Nef or p17 – p51RT* - Nef fusions or any of the p17-Nef fusions as described herein may be provided in a prime boost kit wherein the priming composition comprises the fusion protein and the boosting composition comprises a polynucleotide encoding the fusion protein, or the priming

composition comprises the polynucleotide and the boosting composition comprises the fusion protein.

Both the priming composition and the boosting composition may be delivered in more than one dose. Furthermore the initial priming and boosting doses may be followed up with further doses which may be alternated to result in e.g. a DNA plasmid prime / protein boost / further DNA plasmid dose / further protein dose.

By codon optimisation it is meant that the polynucleotide sequence, is optimised to resemble the codon usage of genes in the desired expression system, for example a prokaryotic system such as *E. coli*. In particular, the codon usage in the sequence is optimised to resemble that of highly expressed *E. coli* genes.

The purpose of codon optimizing for expression in a recombinant system according to the invention is twofold: to improve expression levels of the recombinant product and to render expression products more homogeneous (obtain a more homogeneous expression pattern). Improved homogeneity means that there are fewer irrelevant expression products such as truncates. Codon usage adaptation to *E.coli* expression can also eliminate the putative "frame-shift" sequences as well as premature termination and/or internal initiation sites.

The DNA code has 4 letters (A, T, C and G) and uses these to spell three letter "codons" which represent the amino acids the proteins encoded in an organism's genes. The linear sequence of codons along the DNA molecule is translated into the linear sequence of amino acids in the protein(s) encoded by those genes. The code is highly degenerate, with 61 codons coding for the 20 natural amino acids and 3 codons representing "stop" signals. Thus, most amino acids are coded for by more than one codon - in fact several are coded for by four or more different codons.

Where more than one codon is available to code for a given amino acid, it has been observed that the codon usage patterns of organisms are highly non-random. Different

species show a different bias in their codon selection and, furthermore, utilisation of codons may be markedly different in a single species between genes which are expressed at high and low levels. This bias is different in viruses, plants, bacteria and mammalian cells, and some species show a stronger bias away from a random codon selection than others. For example, humans and other mammals are less strongly biased than certain bacteria or viruses. For these reasons, there is a significant probability that a viral gene from a mammalian virus expressed in *E. coli* , or a foreign or recombinant gene expressed in mammalian cells will have an inappropriate distribution of codons for efficient expression. It is believed that the presence in a heterologous DNA sequence of clusters of codons or an abundance of codons which are rarely observed in the host in which expression is to occur, is predictive of low heterologous expression levels in that host.

In the polynucleotides of the present invention, the codon usage pattern may thus be altered from that typical of human immunodeficiency viruses to more closely represent the codon bias of the target organism, e.g. *E. coli.*

There are a variety of publicly available programs useful for codon optimization, for example "CalcGene" (Hale and Thompson, Protein Expression and Purification 12: 185-189 (1998).

EXAMPLES

Example 1: Construction and expression of HIV-1 p24 – RT – Nef – p17 fusion F4 and F4 codon optimized (co)

1. F4 Non-codon-optimised

HIV-1 gag p24 (capsid protein) and p17 (matrix protein), the reverse transcriptase and Nef proteins were expressed in E.coli B834 strain (B834 (DE3) is a methionine auxotroph parent of BL21 (DE3)), under the control of the bacteriophage T7 promoter (pET expression system).

They were expressed as a single fusion protein containing the complete sequence of the four proteins. Mature p24 coding sequence comes from HIV-1 BH10 molecular clone, mature p17 sequence and RT gene from HXB2 and Nef gene from the BRU isolate.

After induction, recombinant cells expressed significant levels of the p24-RT-Nef-p17 fusion that amounted to 10% of total protein.

When cells were grown and induced at 22°C, the p24-RT-Nef-p17 fusion protein was confined mainly to the soluble fraction of bacterial lysates (even after freezing/thawing). When grown at 30°C, around 30% of the recombinant protein was associated with the insoluble fraction.

The fusion protein p24-RT-Nef-p17 is made up of 1136 amino acids with a molecular mass of approximately 129 kDa. The full-length protein migrates to about 130 kDa on SDS gels. The protein has a theoretical isoeleectric point (pI) of 7.96 based on its amino acid sequence, confirmed by 2D-gel electrophoresis.

Details of the recombinant plasmid:

name: pRIT15436 (or lab name pET28b/p24-RT-Nef-p17)

host vector: pET28b

replicon: colE1

selection: kanamycin

promoter: T7

insert: p24-RT-Nef-p17 fusion gene.

Details of the recombinant protein:

p24-RT-Nef-p17 fusion protein : 1136 amino acids.

N-term - p24: 232a.a. - hinge:2a.a. - RT: 562a.a. -hinge:2a.a. - Nef: 206a.a. - P17: 132a.a. - C-term

Nucleotide and amino-acid sequences:

Nucleotide sequence

atggttatcgtgcagaacatccaggggcaaatggtacatcaggccatatcacctagaact
ttaaatgcatgggtaaaagtagtagaagagaaggctttcagcccagaagtaatacccatg
ttttcagcattatcagaaggagccaccccacaagatttaaacaccatgctaaacacagtg
gggggacatcaagcagccatgcaaatgttaaaagagaccatcaatgaggaagctgcagaa
tgggatagagtacatccagtgcatgcagggcctattgcaccaggccagatgagagaacca
aggggaagtgacatagcaggaactactagtacccttcaggaacaaataggatggatgaca
aataatccacctatcccagtaggagaaatttataaaagatggataatcctgggattaaat
aaaatagtaagaatgtatagccctaccagcattctggacataagacaaggaccaaaagaa
ccttttagagactatgtagaccggttctataaaactctaagagccgagcaagcttcacag
gaggtaaaaaattggatgacagaaaccttgttggtccaaaatgcgaacccagattgtaag
actattttaaaagcattgggaccagcggctacactagaagaaatgatgacagcatgtcag
ggagtaggaggacccggccataaggcaagagttttgcatatg**ggccccattagccctat**
tgagactgtgtcagtaaaattaaagccaggaatggatggcccaaaagttaaacaatggcc
attgacagaagaaaaaataaaagcattagtagaaatttgtacagagatggaaaaggaagg
gaaaatttcaaaaattgggcctgaaaatccatacaatactccagtatttgccataaagaa
aaaagacagtactaaatggagaaaattagtagatttcagagaacttaataagagaactca
agacttctgggaagttcaattaggaataccacatcccgcagggttaaaaaagaaaaaatc
agtaacagtactggatgtgggtgatgcatatttttcagttcccttagatgaagacttcag
gaaatatactgcatttaccatacctagtataaacaatgagacaccagggattagatatca
gtacaatgtgcttccacagggatggaaaggatcaccagcaatattccaaagtagcatgac
aaaaatcttagagccttttagaaaacaaaatccagacatagttatctatcaatacatgga
tgatttgtatgtaggatctgacttagaaatagggcagcatagaacaaaaatagaggagct
gagacaacatctgttgaggtggggacttaccacaccagacaaaaaacatcagaaagaacc
tccattccttaaaatgggttatgaactccatcctgataaatggacagtacagcctatagt

gctgccagaaaaagacagctggactgtcaatgacatacagaagttagtggggaaattgaa
ttgggcaagtcagatttacccagggattaaagtaaggcaattatgtaaactccttagagg
aaccaaagcactaacagaagtaataccactaacagaagaagcagagctagaactggcaga
aaacagagagattctaaaagaaccagtacatggagtgtattatgacccatcaaaagactt
aatagcagaaatacagaagcaggggcaaggccaatggacatatcaaatttatcaagagcc
atttaaaaatctgaaaacaggaaaatatgcaagaatgaggggtgcccacactaatgatgt
aaaacaattaacagaggcagtgcaaaaaataaccacagaaagcatagtaatatggggaaa
gactcctaaatttaaactgcccatacaaaaggaaacatgggaaacatggtggacagagta
ttggcaagccacctggattcctgagtgggagtttgttaatacccctcctttagtgaaatt
atggtaccagttagagaaagaacccatagtaggagcagaaaccttctatgtagatggggc
agctaacagggagactaaattaggaaaagcaggatatgttactaatagaggaagacaaaa
agttgtcaccctaactgacacaacaaatcagaagactgagttacaagcaatttatctagc
tttgcaggattcgggattagaagtaaacatagtaacagactcacaatatgcattaggaat
cattcaagcacaaccagatcaaagtgaatcagagttagtcaatcaaataatagagcagtt
aataaaaaaggaaaaggtctatctggcatgggtaccagcacacaaaggaattggaggaaa
tgaacaagtagataaattagtcagtgctggaatcaggaaagtgctagctatgggtggca
agtggtcaaaaagtagtgtggttggatggcctactgtaagggaaagaatgagacgagctg
agccagcagcagatggggtgggagcagcatctcgagacctggaaaaacatggagcaatca
caagtagcaatacagcagctaccaatgctgcttgtgcctggctagaagcacaagaggagg
aggaggtgggttttccagtcacacctcaggtacctttaagaccaatgact
tacaaggcagctgtagatcttagccacttttttaaaagaaaaggggggactggaagggcta
attcactcccaacgaagacaagatatccttgatctgtggatctaccacacacaaggctac
ttccctgattggcagaactacacaccagggccaggggtcagatatccactgacctttgga
tggtgctacaagctagtaccagttgagccagataaggtagaagaggccaataaaggagag
aacaccagcttgttacaccctgtgagcctgcatggaatggatgaccctgagagagaagtg
ttagagtggaggtttgacagccgcctagcatttcatcacgtggcccgagagctgcatccg
gagtacttcaagaactgcaggcctatgggtgcgagagcgtcagtattaagcggggaga
attagatcgatgggaaaaaattcggttaaggccaggggaaagaaaaaatataaattaaa
acatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttaga
aacatcagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatc
agaagaacttagatcattatataatacagtagcaaccctctattgtgtgcatcaaaggat
agagataaaagacaccaaggaagctttagacaagatagaggaagagcaaaacaaaagtaa
gaaaaaagcacagcaagcagcagctgacacaggacacagcaatcaggtcagccaaaatta
ctaa [SEQ ID NO:1]

p24 sequence is in bold
Nef sequence is underlined
Boxes: nucleotides introduced by genetic construction

Amino-Acid sequence

MVIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATP 50
QDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREP 100
RGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPTS 150
ILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCK 200
TILKALGPAATLEEMMTACQGVGGPGHKARVLHMGPISPIETVSVKLKPG 250
MDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKK 300
KDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAY 350
FSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMT 400
KILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLT 450
TPDKKHQKEPPFLKMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLN 500
WASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELELAENREILKEPVH 550
GVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHTNDV 600
KQLTEAVQKITTESIVIWGKTPKFKLPIQKETWETWWTEYWQATWIPEWE 650
FVNTPPLVKLWYQLEKEPIVGAETFYVDGAANRETKLGKAGYVTNRGRQK 700
VVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSES 750
ELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVLAMGGK 800
WSKSSVVGWPTVRERMRRAEPAADGVGAASRDLEKHGAITSSNTAATNAA 850
CAWLEAQEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQ 900
RRQDILDLWIYHTQGYFPDWQNYTPGPGVRYPLTFGWCYKLVPVEPDKVE 950
EANKGENTSLLHPVSLHGMDDPEREVLEWRFDSRLAFHHVARELHPEYFK 1000
NCRPMGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAV 1050
NPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKD 1100
TKEALDKIEEEQNKSKKKAQQAAADTGHSNQVSQNY 1136
[SEQ ID NO:2]

P24 sequence: amino-acids 1-232 (in bold)
RT sequence: amino-acids 235-795
Nef sequence: amino-acids 798-1002
P17 sequence: amino-acids 1005-1136
Boxes:amino-acids introduced by genetic construction
K (Lysine): instead of Tryptophan (W). Mutation introduced to remover enzyme activity.

Expression of the recombinant protein:

In pET plasmid, the target gene (p24-RT-Nef-p17) is under control of the strong bacteriophage T7 promoter. This promoter is not recognized by E.coli RNA polymerase and is dependent on a source of T7 RNA polymerase in the host cell. B834 (DE3) host

cell contains a chromosomal copy of the T7 RNA polymerase gene under lacUV5 control and expression is induced by the addition of IPTG to the bacterial culture.

Pre-cultures were grown, in shake flasks, at 37°C to mid-log phase (A620:0.6) and then stored at 4°C overnight (to avoid stationary phase cultures). Cultures were grown in LBT medium supplemented with 1% glucose and 50 µg/ml kanamycin. Addition of glucose to the growth medium has the advantage to reduce the basal recombinant protein expression (avoiding cAMP mediated derepression of lacUV5 promoter)

Ten ml of cultures stored overnight at 4°C were used to inoculate 200 ml of LBT medium (without glucose) containing kanamycin. Cultures were grown at 30°C and 22°C and when O.D.620 reached 0.6, IPTG was added (1mM final). Cultures were incubated for further 3, 5 and 18 hours (overnight). Samples were collected before and after 3, 5 and 18 hours induction.

Extract preparation was as follows:
Cell pellets were suspended in breaking buffer* (at a theoretical O.D. of 10) and disrupted by four passages in French press (at 20.000psi or 1250 bars). Crude extracts (T) were centrifuged at 20.000g for 30 min to separate the soluble (S) and insoluble (P) fractions.
*Breaking buffer: 50mM Tris-HCL pH 8.0, 1mM EDTA, 1mM DTT + protease inhibitors cocktail (Complete/Boerhinger).

SDS-PAGE and Western Blot analysis:
Fractions corresponding to insoluble pellet (P), supernatant (S) and crude extract (T) were run on 10 % reducing SDS-PAGE. p24-RT-Nef -p17recombinant was detected by Coomassie blue staining and on Western blot (WB).

Coomassie staining: p24-RT-Nef-p17 protein appears as:
one band at ± 130 kDa (fitting with calculated MW)

MW theoretical: 128.970 Daltons

MW apparent: 130 kDa

Western blot analysis:

Reagents = - Monoclonal antibody to RT (p66/p51)
Purchased from ABI (Advanced Biotechnologies)
dilution: 1/5000
-Alkaline phosphatase-conjugate anti-mouse antibody
dilution: 1/7500

Expression level: - Very strong p24-RT-Nef-p17 specific band after 20h induction at 22°C, representing up to 10% of total protein (See Figure 1A).

Recombinant protein "solubility":

"Fresh" cellular extracts (T,S,P fractions): With growth/induction at 22°C/20h, almost all p24-RT-Nef-p17 fusion protein is recovered in the soluble fraction of cellular extract (Figure 1A). With growth/induction at 30°C/20h, around 30% of p24-RT-Nef-p17 protein is associated with the insoluble fraction (Figure 1A).

"Freezing/thawing" (S2, P2 fractions):
Soluble (S1) fraction (20h induction at 22°C) conserved at -20°C. Thawed and centrifuged at 20.000g/30 min : S2 and P2 (resuspended in 1/10 vol.)

Breaking buffer with DTT : almost all p24-RT-Nef-p17 fusion protein still soluble (only 1-5 % precipitated) (see Figure 1B)

Breaking buffer without DTT: 85-90 % of p24-RT-Nef-p17 still soluble (Figure 1B)

Figures:
Figure 1A - Coomassie staining and western blot.

Figure 1B – p24-RT-Nef-p17 solubility assay

The F4 protein was purified using purification method I in Example 7.

The cell growth and induction conditions and cellular extracts preparation for the examples which follow are as described in Example 1 unless other conditions are specified (e.g. temperature, composition of breaking buffer).

2. F4 codon-optimised

The following polynucleotide sequence is codon optimized such that the codon usage resembles the codon usage in a highly expressed gene in *E.coli*. The amino acid sequence is identical to that given above for F4 non-codon optimized.

Nucleotide sequence for F4co:

```
atggtcattgttcagaacatacagggccaaatggtccaccaggcaattagtccgcgaact
cttaatgcatgggtgaaggtcgtggaggaaaaggcattctccccggaggtcattccgatg
ttttctgcgctatctgagggcgcaacgccgcaagaccttaataccatgcttaacacggta
ggcgggcaccaagccgctatgcaaatgctaaaagagactataaacgaagaggccgccgaa
tgggatcgagtgcacccggtgcacgccggcccaattgcaccaggccagatgcgcgagccg
cgcgggtctgatattgcaggaactacgtctacccttcaggagcagattgggtggatgact
aacaatccaccaatcccggtcggagagatctataagaggtggatcatactgggactaaac
aagatagtccgcatgtattctccgacttctatactggatatacgccaaggcccaaaggag
ccgttcagggactatgtcgaccgattctataagacccttcgcgcagagcaggcatcccag
gaggtcaaaaattggatgacagaaactcttttggtgcagaatgcgaatccggattgtaaa
acaattttaaaggctctaggaccggccgcaacgctagaagagatgatgacggcttgtcag
ggagtcggtggaccggggcataaagcccgcgtcttacacatgggcccgatatctccgat
agaaacagtttcggtcaagcttaaaccagggatggatggtccaaaggtcaagcagtggcc
gctaacggaagagaagattaaggcgctcgtagagatttgtactgaaatggagaaggaagg
caagataagcaagatcgggccagagaacccgtacaatacaccggtatttgcaataaagaa
aaaggattcaacaaaatggcgaaagcttgtagattttagggaactaaacaagcgaaccca
agacttttgggaagtccaactagggatcccacatccagccggtctaaagaagaagaaatc
ggtcacagtcctggatgtaggagacgcatattttagtgtaccgcttgatgaggacttccg
aaagtatactgcgtttactataccgagcataaacaatgaaacgccaggcattcgctatca
gtacaacgtgctcccgcagggctggaaggggtctccggcgatatttcagagctgtatgac
aaaaatacttgaaccattccgaaagcagaatccggatattgtaatttaccaatacatgga
cgatctctatgtgggctcggatctagaaattgggcagcatcgcactaagattgaggaact
gaggcaacatctgcttcgatggggcctcactactcccgacaagaagcaccagaaggagcc
```

gccgttcctaaagatgggctacgagcttcatccggacaagtggacagtacagccgatagt
gctgcccgaaaaggattcttggaccgtaaatgatattcagaaactagtcggcaagcttaa
ctgggcctctcagatttacccaggcattaaggtccgacagctttgcaagctactgagggg
aactaaggctctaacagaggtcatcccattaacggaggaagcagagcttgagctggcaga
gaatcgcgaaattcttaaggagccggtgcacggggtatactacgacccctccaaggacct
tatagccgagatccagaagcaggggcagggccaatggacgtaccagatatatcaagaacc
gtttaagaatctgaagactgggaagtacgcgcgcatgcgaggggctcatactaatgatgt
aaagcaacttacggaagcagtacaaaagattactactgagtctattgtgatatggggcaa
gaccccaaagttcaagctgcccatacagaaggaaacatgggaaacatggtggactgaata
ttggcaagctacctggattccagaatgggaatttgtcaacacgccgccacttgttaagct
ttggtaccagcttgaaaaggagccgatagtaggggcagagaccttctatgtcgatggcgc
cgcgaatcgcgaaacgaagctaggcaaggcgggatacgtgactaataggggccgccaaaa
ggtcgtaacccttacggataccaccaatcagaagactgaactacaagcgatttaccttgc
acttcaggatagtggcctagaggtcaacatagtcacggactctcaatatgcgcttggcat
tattcaagcgcagccagatcaaagcgaaagcgagcttgtaaaccaaataatagaacagct
tataaagaaagagaaggtatatctggcctgggtccccgctcacaagggaattggcggcaa
tgagcaagtggacaagctagtcagcgctgggattcgcaaggttcttgcgatgggggta
agtggtctaagtctagcgtagtcggctggccgacagtccgcgagcgcatgcgacgcgccg
aaccagccgcagatggcgtgggggcagcgtctagggatctggagaagcacggggctataa
cttccagtaacacggcggcgacgaacgccgcatgcgcatggttagaagcccaagaagagg
aagaagtagggtttccggtaactccccaggtgccgttaaggccgatgacc
tataaggcagcggtggatctttctcacttccttaaggagaaaggggggctggagggctta
attcacagccagaggcgacaggatattcttgatctgtggatttaccatacccaggggtac
tttccggactggcagaattacaccccggggccaggcgtgcgctatcccctgactttcggg
tggtgctacaaactagtcccagtggaacccgacaaggtcgaagaggctaataagggcgag
aacacttctcttcttcacccggtaagcctgcacgggatggatgacccagaacgagaggtt
ctagaatggaggttcgactctcgacttgcgttccatcacgtagcacgcgagctgcatcca
gaatatttcaagaactgccgcccaatgggcgccagggccagtgtacttagtggcggaga
actagatcgatgggaaaagatacgcctacgcccgggggcaagaagaagtacaagcttaa
gcacattgtgtgggcctctcgcgaacttgagcgattcgcagtgaatccaggcctgcttga
gacgagtgaaggctgtaggcaaattctggggcagctacagccgagcctacagactggcag
cgaggagcttcgtagtctttataataccgtcgcgactctctactgcgttcatcaacgaat
tgaaataaaggatactaaagaggcccttgataaaattgaggaggaacagaataagtcgaa
aaagaaggcccagcaggccgccgccgacaccgggcacagcaaccaggtgtcccaaaacta
ctaa

[SEQ ID NO:3]

p24 sequence is in bold

Nef sequence is underlined

Boxes: nucleotides introduced by genetic construction

The procedures used in relation to F4 non-codon optimized were applied for the codon-optimised sequence.

Details of the recombinant plasmid:

name:	pRIT15513 (lab name: pET28b/p24-RT-Nef -p17)
host vector:	pET28b
replicon:	colE1
selection:	kanamycin
promoter:	T7
insert:	p24-RT-Nef-p17 fusion gene, codon-optimized

The F4 codon-optimised gene was expressed in *E.coli* BLR(DE3) cells, a recA$^-$ derivative of B834(DE3) strain. RecA mutation prevents the putativve production of lambda phages.

Pre-cultures were grown, in shake flasks, at 37°C to mid-log phase (A_{620}:0.6) and then stored at 4°C overnight (to avoid stationary phase cultures).

Cultures were grown in LBT medium supplemented with 1% glucose and 50 µg/ml kanamycin. Addition of glucose to the growth medium has the advantage to reduce the basal recombinant protein expression (avoiding cAMP mediated derepression of lacUV5 promoter).

Ten ml of cultures stored overnight at 4°C were used to inoculate 200 ml of LBT medium (without glucose) containing kanamycin. Cultures were grown at 37°C and when $O.D._{620}$ reached 0.6, IPTG was added (1mM final). Cultures were incubated for further 19 hours (overnight), at 22°C. Samples were collected before and 19 hours induction.

Extract preparation was as follows:

Cell pellets were resuspended in sample buffer (at a theoretical O.D. of 10), boiled and directly loaded on SDS-PAGE.

SDS-PAGE and Western Blot analysis:

Crude extracts samples were run on 10 % reducing SDS-PAGE.

p24-RT-Nef -p17 recombinant protein is detected by Coomassie blue staining (Figure 2) and on Western blot.

Coomassie staining: p24-RT-Nef-p17 protein appears as:

one band at ± 130 kDa (fitting with calculated MW)

MW theoretical: 128.967 Daltons

MW apparent: 130 kDa

Western blot analysis:

Reagents =
- Rabbit polyclonal anti RT (rabbit PO3L16) dilution: 1/10.000
- Rabbit polyclonal anti Nef-Tat (rabbit 388) dilution 1/10.000
- Alkaline phosphatase-conjugate anti- rabbit antibody . dilution: 1/7500

After induction at 22°C over 19 hours, recombinant BLR(DE3) cells expressed the F4 fusion at a very high level ranging from 10-15% of total protein.

In comparison with F4 from the native gene, the F4 recombinant product profile from the codon-optimised gene is slightly simplified. The major F4-related band at 60 kDa, as well as minor bands below, disappeared (see Figure 2). Compared to the B834(DE3) recombinant strain expressing F4, the BLR(DE3) strain producing F4co has the following advantages: higher production of F4 full-length protein, less complex band pattern of recombinant product.

Example 2: Construction and expression of P51 RT (truncated, codon-optimised RT)

The RT/p66 region between amino acids 428-448 is susceptible to E.coli proteases. The P51 construct terminates at Leu 427 resulting in the elimination of RNaseH domain (see RT sequence alignment in Figure 3).

The putative E.coli "frameshift" sequences identified in RT native gene sequence were also eliminated (by codon-optimization of p51 gene).

p51 synthetic gene design/construction:

The sequence of the synthetic p51 gene was designed according to *E.coli* codon usage. Thus it was codon optimized such that the codon usage resembles the codon usage in a highly expressed gene in *E.coli*. The synthetic gene was constructed as follows: 32 oligonucleotides were assembled in a single-step PCR. In a second PCR the full-length assembly was amplified using the ends primers and the resulting PCR product was cloned into pGEM-T intermediate plasmid. After correction of point errors introduced during gene synthesis, the p51 synthetic gene was cloned into pET29a expression plasmid. This recombinant plasmid was used to transform B834 (DE3) cells.

Recombinant protein characteristics:

P51 RT nucleotide sequence

```
atgagtactggtccgatctctccgatagaaacagtttcggtcaagcttaaaccagggatg   60
gatggtccaaaggtcaagcagtggccgctaacggaagagaagattaaggcgctcgtagag  120
atttgtactgaaatggagaaggaaggcaagataagcaagatcgggccagagaacccgtac  180
aatacaccggtatttgcaataaagaagaaggattcaacaaaatggcgaaagcttgtagat  240
tttagggaactaaacaagcgaacccaagacttttgggaagtccaactaggtatcccacat  300
ccagccggtctaaagaagaagaaatcggtcacagtcctggatgtaggagacgcatatttt  360
agtgtaccgcttgatgaggacttccgaaagtatactgcgtttactataccgagcataaac  420
aatgaaacgccaggcattcgctatcagtacaacgtgctcccgcagggctggaaggggtct  480
```

ccggcgatatttcagagctctatgacaaaaatacttgaaccattccgaaagcagaatccg 540
gatattgtaatttaccaatacatggacgatctctatgtgggctcggatctagaaattggg 600
cagcatcgcactaagattgaggaactgaggcaacatctgcttcgatggggcctcactact 660
cccgacaagaagcaccagaaggagccgccgttcctaaagatgggctacgagcttcatccg 720
gacaagtggacagtacagccgatagtgctgcccgaaaaggattcttggaccgtaaatgat 780
attcagaaactagtcggcaagcttaactgggcctctcagatttacccaggcattaaggtc 840
cgacagctttgcaagctactgaggggaactaaggctctaacagaggtcatcccattaacg 900
gaggaagcagagcttgagctggcagagaatcgcgaaattcttaaggagccggtgcacggg 960
gtatactacgacccctccaaggaccttatagccgagatccagaagcaggggcagggccaa 1020
tggacgtaccagatatatcaagaaccgtttaagaatctgaagactgggaagtacgcgcgc 1080
atgcgaggggctcatactaatgatgtaaagcaacttacggaagcagtacaaaagattact 1140
actgagtctattgtgatatggggcaagaccccaaagttcaagctgcccatacagaaggaa 1200
acatgggaaacatggtggactgaatattggcaagctacctggattccagaatgggaattt 1260
gtcaacacgccgccgctggtaaaactg[aggcctgctagc]taa 1302

[SEQ ID NO:4]

Boxes: amino-acids introduced by genetic construction

Amino-acid sequence:

M[ST]GPISPIETVSVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPY 60
NTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYF 120
SVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNP 180
DIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFL[K]MGYELHP 240
DKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLT 300
EEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYAR 360
MRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQKETWETWWTEYWQATWIPEWEF 420
VNTPPLVKL[RPAS] 433

[SEQ ID NO:5]

Boxes: amino-acids introduced by genetic construction.

K (Lysine): instead of Tryptophan (W). Mutation introduced to remover enzyme activity.

Length, Molecular Weight, Isoelectric Point (IP):

433 AA, MW : 50.3 kDa,, IP: 9.08

p51 expression in B834(DE3) cells:

P51 expression level and recombinant protein solubility were evaluated, in parallel to RT/p66 production strain.

p51 expression level:

Induction condition: cells grown / induced at 37°C (+1mM IPTG), during 5 hours.

Breaking buffer: 50 mM Tris/HCl, pH:7.5, 1mM EDTA, +/- 1mM DTT.

Western blot analysis:

Reagents: - rabbit polyclonal anti RT (rabbit PO3L16) (dilution: 1/10,000)

- Alkaline phosphatase-conjugate anti-rabbit antibody (dilution: 1/7500)

Cellular fractions corresponding to crude extracts (T), insoluble pellet (P) and supernatant (S) were run on 10 % reducing SDS-PAGE.

As illustrated on Coomassie stained gel and Western Blot (Figure 4) very high expression of P51 (15-20% of total protein) was observed, higher than that observed for P66.

For both p51 and p66 proteins (after 5h induction at 37°C), 80% of the recombinant products were recovered in the soluble fraction (S1) of cellular extracts (See Figure 4). When expressed at 30°C, 99% of recombinant proteins were associated with the soluble fraction (data not shown).

The p51 Western Blot pattern was multiband, but less complex than that observed for P66.

Solubility assay

Solubility assay: Freezing/thawing of Soluble (S1) fraction (5h induction, 37°C) prepared under reducing (breaking buffer with DTT) and non-reducing conditions. After thawing, S1 samples were centrifuged at 20.000g/30 minutes, generating S2 and P2 (p2 is resuspended in 1/10 vol.).

After freezing/thawing of soluble fractions (S1), prepared under reducing as well as non-reducing conditions, 99% of p51 and p66 are still recovered in soluble (S2) fraction. Only 1% is found in the precipitate (P2). This is shown in Figure 5.

Example 3: Construction and expression of p17-Nef and Nef-p17 with or without linker

The double fusion proteins were constructed with and without linkers. The linkers aimed to decrease potential interactions between the two fusion partners and are as follows:

Nef-GSGGGP-P17 and p17-GSGGGP-Nef

Recombinant plasmids construction:

- **pET29a/Nef-p17 expression vector:**

Nef-p17 fusion gene was amplified by PCR from the F4 recombinant plasmid. The PCR product was cloned into the intermediate pGEM-T cloning vector and subsequently into the pET29a expression vector.

- **pET28b/p17-Nef expression vector:**

Nef gene was amplified by PCR from the F4 recombinant plasmid. The PCR product was cloned into the intermediate pGEM-T cloning vector and subsequently into the pET28b/p17 expression vector, as a C-terminal in frame fusion with the p17 gene.

- **pET29a/Nef-linker-p17 and pET28b/p17-linker-Nef expression vector:**

A 18 bp DNA fragment coding for the hexapeptide linker (GSGGGP) was inserted between Nef and p17 fusion partners, by site-directed mutagenesis (using the "GeneTailor Site-Directed Mutagenesis System", Invitrogen).

Recombinant protein characteristics:

- **Length, Molecular Weight, Isoelectric Point (IP)**

Nef-p17 (named NP): 340 AA, MW: 38.5 kDa, IP:7.48

Nef-GSGGGP-P17 (named NLP): 346 AA, MW:38.9 kDa, IP: 7.48

p17-Nef (named PN): 342 AA, MW: 38.7 kDa, IP: 7.19

p17-GSGGGP-Nef (named PLN): 348 AA, MW: 39.1kDa, IP: 7.19

- **Amino-acid sequences and polynucleotide sequences:**

Nef-p17 nucleotide sequence

```
Atgggtggcaagtggtcaaaaagtagtgtggttggatggcctactgtaagggaaagaatg    60
Agacgagctgagccagcagcagatggggtgggagcagcatctcgagacctggaaaaacat   120
Ggagcaatcacaagtagcaatacagcagctaccaatgctgcttgtgcctggctagaagca   180
Caagaggaggaggaggtgggttttccagtcacacctcaggtacctttaagaccaatgact   240
Tacaaggcagctgtagatcttagccacttttaaaagaaaaggggggactggaagggcta    300
Attcactcccaacgaagacaagatatccttgatctgtggatctaccacacacaaggctac   360
Ttccctgattggcagaactacacaccagggccaggggtcagatatccactgacctttgga   420
Tggtgctacaagctagtaccagttgagccagataaggtagaagaggccaataaaggagag   480
Aacaccagcttgttacaccctgtgagcctgcatggaatggatgaccctgagagagaagtg   540
Ttagagtggaggtttgacagccgcctagcatttcatcacgtggcccgagagctgcatccg   600
Gagtacttcaagaactgcaggcctatgggtgcgagagcgtcagtattaagcgggggagaa   660
Ttagatcgatgggaaaaaattcggttaaggccaggggaaagaaaaaatataaattaaaa    720
Catatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttagaa   780
Acatcagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatca   840
Gaagaacttagatcattatataatacagtagcaaccctctattgtgtgcatcaaaggata   900
Gagataaaagacaccaaggaagctttagacaagatagaggaagagcaaaacaaaagtaag   960
Aaaaaagcacagcaagcagcagctgacacaggacacagcaatcaggtcagccaaaattac  1020
Taa                                                            1023
```

[SEQ ID NO:6]

Nef-p17 (NP)

```
MGGKWSKSSVVGWPTVRERMRRAEPAADGVGAASRDLEKHGAITSSNTAATNAACAWLEA    60
QEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQRRQDILDLWIYHTQGY   120
FPDWQNYTPGPGVRYPLTFGWCYKLVPVEPDKVEEANKGENTSLLHPVSLHGMDDPEREV   180
```

LEWRFDSRLAFHHVARELHPEYFKNCRP**MGARASVLSGGELDRWEKIRLRPGGKKKYKLK** 240
HIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRI 300
EIKDTKEALDKIEEEQNKSKKKAQQAAADTGHSNQVSQNY 340
[SEQ ID NO:7]

Box: amino-acids introduced by genetic construction.
Nef sequence is in bold.

P17-Nef nucleotide sequence:

Atgggtgcgagagcgtcagtattaagcgggggagaattagatcgatgggaaaaaattcgg 60
Ttaaggccaggggaaagaaaaaatataaattaaaacatatagtatgggcaagcagggag 120
Ctagaacgattcgcagttaatcctggcctgttagaaacatcagaaggctgtagacaaata 180
Ctgggacagctacaaccatcccttcagacaggatcagaagaacttagatcattatataat 240
Acagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagct 300
Ttagacaagatagaggaagagcaaaacaaaagtaagaaaaaagcacagcaagcagcagct 360
Gacacaggacacagcaatcaggtcagccaaaattacctcgacaggcctatgggtggcaag 420
Tggtcaaaaagtagtgtggttggatggcctactgtaagggaaagaatgagacgagctgag 480
Ccagcagcagatggggtgggagcagcatctcgagacctggaaaaacatggagcaatcaca 540
Agtagcaatacagcagctaccaatgctgcttgtgcctggctagaagcacaagaggaggag 600
Gaggtgggttttccagtcacacctcaggtacctttaagaccaatgacttacaaggcagct 660
Gtagatcttagccactttttaaaagaaaaggggggactggaagggctaattcactcccaa 720
Cgaagacaagatatccttgatctgtggatctaccacacacaaggctacttccctgattgg 780
Cagaactacacaccagggccaggggtcagatatccactgacctttggatggtgctacaag 840
Ctagtaccagttgagccagataaggtagaagaggccaataaaggagagaacaccagcttg 900
Ttacaccctgtgagcctgcatggaatggatgaccctgagagagaagtgttagagtggagg 960
Tttgacagccgcctagcatttcatcacgtggcccgagagctgcatccggagtacttcaag 1020
Aactgctaa 1029
[SEQ ID NO:8]

P17-Nef (PN)

MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI 60
LGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAA 120
DTGHSNQVSQNYLDRP**MGGKWSKSSVVGWPTVRERMRRAEPAADGVGAASRDLEKHGAIT** 180
SSNTAATNAACAWLEAQEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQ 240
RRQDILDLWIYHTQGYFPDWQNYTPGPGVRYPLTFGWCYKLVPVEPDKVEEANKGENTSL 300
LHPVSLHGMDDPEREVLEWRFDSRLAFHHVARELHPEYFKNC 342
[SEQ ID NO:9]

Box: amino-acids introduced by genetic construction.
p17 sequence is in bold.

Nef-linker-p17 nucleotide sequence:

```
Atgggtggcaagtggtcaaaaagtagtgtggttggatggcctactgtaagggaaagaatg    60
Agacgagctgagccagcagcagatggggtgggagcagcatctcgagacctggaaaaacat   120
Ggagcaatcacaagtagcaatacagcagctaccaatgctgcttgtgcctggctagaagca   180
Caagaggaggaggaggtgggttttccagtcacacctcaggtacctttaagaccaatgact   240
Tacaaggcagctgtagatcttagccacttttaaaagaaaaggggggactggaagggcta   300
Attcactcccaacgaagacaagatatccttgatctgtggatctaccacacacaaggctac   360
Ttccctgattggcagaactacacaccagggccaggggtcagatatccactgacctttgga   420
Tggtgctacaagctagtaccagttgagccagataaggtagaagaggccaataaaggagag   480
Aacaccagcttgttacaccctgtgagcctgcatggaatggatgaccctgagagagaagtg   540
Ttagagtggaggtttgacagccgcctagcatttcatcacgtggcccgagagctgcatccg   600
Gagtacttcaagaactgcaggcctggatccggtggcggccctatgggtgcgagagcgtca   660
Gtattaagcgggggagaattagatcgatgggaaaaaattcggttaaggccaggggaaag   720
Aaaaaatataaattaaaacatatagtatgggcaagcagggagctagaacgattcgcagtt   780
Aatcctggcctgttagaaacatcagaaggctgtagacaaatactgggacagctacaacca   840
Tcccttcagacaggatcagaagaacttagatcattatataatacagtagcaaccctctat   900
Tgtgtgcatcaaaggatagagataaaagacaccaaggaagctttagacaagatagaggaa   960
Gagcaaaacaaaagtaagaaaaaagcacagcaagcagcagctgacacaggacacagcaat  1020
Caggtcagccaaaattactaa                                         1041
```

[SEQ ID NO:10]

Nef-linker-p17 (NLP)

```
MGGKWSKSSVVGWPTVRERMRRAEPAADGVGAASRDLEKHGAITSSNTAATNAACAWLEA    60
QEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQRRQDILDLWIYHTQGY   120
FPDWQNYTPGPGVRYPLTFGWCYKLVPVEPDKVEEANKGENTSLLHPVSLHGMDDPEREV   180
LEWRFDSRLAFHHVARELHPEYFKNCRPGSGGGPMGARASVLSGGELDRWEKIRLRPGGK   240
KKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLY   300
CVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAADTGHSNQVSQNY                 346
```

[SEQ ID NO:11]

Hexapeptide linker

Box: amino-acids introduced by genetic construction.

P17-linker-Nef nucleotide sequence:

```
Atgggtgcgagagcgtcagtattaagcgggggagaattagatcgatgggaaaaaattcgg    60
Ttaaggccaggggaaagaaaaaatataaattaaaacatatagtatgggcaagcagggag    120
Ctagaacgattcgcagttaatcctggcctgttagaaacatcagaaggctgtagacaaata   180
Ctgggacagctacaaccatcccttcagacaggatcagaagaacttagatcattatataat   240
Acagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagct   300
```

```
Ttagacaagatagaggaagagcaaaacaaaagtaagaaaaaagcacagcaagcagcagct  360
Gacacaggacacagcaatcaggtcagccaaaattacctcgacaggcctggatccggtggc  420
Ggtcctatgggtggcaagtggtcaaaaagtagtgtggttggatggcctactgtaagggaa  480
Agaatgagacgagctgagccagcagcagatggggtgggagcagcatctcgagacctggaa  540
Aaacatggagcaatcacaagtagcaatacagcagctaccaatgctgcttgtgcctggcta  600
Gaagcacaagaggaggaggaggtgggttttccagtcacacctcaggtacctttaagacca  660
Atgacttacaaggcagctgtagatcttagccacttttttaaaagaaaaggggggactggaa  720
Gggctaattcactcccaacgaagacaagatatccttgatctgtggatctaccacacacaa  780
Ggctacttccctgattggcagaactacacaccagggccaggggtcagatatccactgacc  840
Tttggatggtgctacaagctagtaccagttgagccagataaggtagaagaggccaataaa  900
Ggagagaacaccagcttgttacaccctgtgagcctgcatggaatggatgaccctgagaga  960
Gaagtgttagagtggaggtttgacagccgcctagcatttcatcacgtggcccgagagctg 1020
Catccggagtacttcaagaactgctaa                                   1047
```

[SEQ ID NO:12]

P17-linker-Nef (PLN)

```
MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI   60
LGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAA  120
DTGHSNQVSQNYLDRPGSGGGPMGGKWSKSSVVGWPTVRERMRRAEPAADGVGAASRDLE  180
KHGAITSSNTAATNAACAWLEAQEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLE  240
GLIHSQRRQDILDLWIYHTQGYFPDWQNYTPGPGVRYPLTFGWCYKLVPVEPDKVEEANK  300
GENTSLLHPVSLHGMDDPEREVLEWRFDSRLAFHHVARELHPEYFKNC              348
```

[SEQ ID NO:13]

Hexapeptide linker
Box: amino-acids introduced by genetic construction.

Comparative expression of Nef-p17 , p17-Nef fusions, with and w/o linkers:

The four recombinant strains were induced at 30°C over 3 hours, in parallel to F4 and Nef producing strains. Crude extracts were prepared and analyzed by Coomassie stained gel and Western blotting.

Western blot analysis:

Reagents: - *rabbit polyclonal anti RT (rabbit PO3L16) (dilution: 1/10,000)*

- *Alkaline phosphatase-conjugate anti-rabbit antibody (dilution: 1/7500)*

As illustrated in Figure 6, Nef-p17 and p17-Nef fusions, with and w/o linker, are expressed at a high level (10% total proteins).

In the Western blot: the four double fusion constructs present a multi-band pattern, but less complex than what was observed for F4. When expressed alone, the Nef and p17 proteins present single band patterns.

Strains expressing Nef-p17 (NP) and p17-Nef (PN) fusions, without linker peptide, were further analysed (solubility assays, see below).

Nef-p17 and p17-Nef solubility assay:

Nef-p17 and p17-Nef proteins were induced, in parallel to F4 and Nef producing strains.
Induction condition: cells grown / induced at 30°C (+1mM IPTG), over 3 hours.
Breaking buffer: 50mM Tris/HCl pH:8, 50 mM NaCl, 1mM EDTA

Fresh cellular extracts:

Cellular extracts were prepared (under non-reducing conditions) and fractions corresponding to crude extracts (T), insoluble pellet (P), and supernatant (S1) were analyzed on Coomassie stained gel and Western blot.

As illustrated in Figure 7 on Coomassie stained gel and Western blot, almost all Nef-p17, p17-Nef, as well as Nef proteins are recovered in the soluble fraction (S) of cellular extracts. For F4 construct: 5-10% of recombinant protein already recovered in the pellet fraction.

Conclusions:

All double fusion constructs tested are highly expressed (> 10% of total protein). P17-

Nef and Nef-p17 fusion proteins are more soluble than F4. Both present a less complex WB pattern.

Example 4: Construction and expression of p24-RT*-Nef-p17 (F4*)

F4* is a mutated version of the F4 (p24-RT/p66-Nef-p17) fusion where the Methionine at position 592 is replaced by a Lysine. This methionine is a putative internal transcriptional "start" site, as supported by N-terminal sequencing performed on a Q sepharose eluate sample of F4 purification experiment. Indeed, the major F4-related small band at 62 kDa present in the Q eluate sample starts at methionine 592.

Methionine is replaced by a lysine: RMR → RKR. The RKR motif is naturally present in clade A RT sequences.

The impact of this mutation on CD4-CD8 epitopes was evaluated:

- one HLA-A3 CTL epitope (A* 3002) is lost, but 9 other HLA-A3 epitopes are present in the RT sequence.
- No helper epitope identified in this region.

Recombinant protein characteristics:

N-term - **p24: 232a.a.** - **hinge:2a.a.** - **RT: 562a.a.** -**hinge:2a.a.** - **Nef: 206a.a.** -

- **P17: 132a.a.** - *C-term*

- **Length, Molecular Weight, Isoelectric Point (IP):**

  1136 AA, 129 kDa, IP: 8.07

- **Nucleotide sequence:**

```
atggttatcgtgcagaacatccaggggcaaatggtacatcaggccatatcacctagaact
ttaaatgcatgggtaaaagtagtagaagagaaggctttcagcccagaagtaatacccatg
ttttcagcattatcagaaggagccaccccacaagatttaaacaccatgctaaacacagtg
```

gggggacatcaagcagccatgcaaatgttaaaagagaccatcaatgaggaagctgcagaa
tgggatagagtacatccagtgcatgcagggcctattgcaccaggccagatgagagaacca
aggggaagtgacatagcaggaactactagtacccttcaggaacaaataggatggatgaca
aataatccacctatcccagtaggagaaatttataaaagatggataatcctgggattaaat
aaaatagtaagaatgtatagccctaccagcattctggacataagacaaggaccaaaagaa
ccttttagagactatgtagaccggttctataaaactctaagagccgagcaagcttcacag
gaggtaaaaaattggatgacagaaaccttgttggtccaaaatgcgaacccagattgtaag
actattttaaaagcattgggaccagcggctacactagaagaaatgatgacagcatgtcag
ggagtaggaggacccggccataaggcaagagttttgcatatg**ggccccattagccctat**
tgagactgtgtcagtaaaattaaagccaggaatggatggcccaaaagttaaacaatggcc
attgacagaagaaaaaataaaagcattagtagaaatttgtacagagatggaaaaggaagg
gaaaatttcaaaaattgggcctgaaaatccatacaatactccagtatttgccataaagaa
aaaagacagtactaaatggagaaaattagtagatttcagagaacttaataagagaactca
agacttctgggaagttcaattaggaataccacatcccgcagggttaaaaaagaaaaaatc
agtaacagtactggatgtgggtgatgcatatttttcagttcccttagatgaagacttcag
gaaatatactgcatttaccatacctagtataaacaatgagacaccagggattagatatca
gtacaatgtgcttccacagggatggaaaggatcaccagcaatattccaaagtagcatgac
aaaaatcttagagccttttagaaaacaaaatccagacatagttatctatcaatacatgga
tgatttgtatgtaggatctgacttagaaatagggcagcatagaacaaaaatagaggagct
gagacaacatctgttgaggtggggacttaccacaccagacaaaaaacatcagaaagaacc
tccattccttaaaatgggttatgaactccatcctgataaatggacagtacagcctatagt
gctgccagaaaaagacagctggactgtcaatgacatacagaagttagtggggaaattgaa
ttgggcaagtcagatttacccagggattaaagtaaggcaattatgtaaactccttagagg
aaccaaagcactaacagaagtaataccactaacagaagaagcagagctagaactggcaga
aaacagagagattctaaaagaaccagtacatggagtgtattatgacccatcaaaagactt
aatagcagaaatacagaagcaggggcaaggccaatggacatatcaaatttatcaagagcc
atttaaaaatctgaaaacaggaaaatatgcacgtaaacgcggtgcccacactaatgatgt
aaaacaattaacagaggcagtgcaaaaataaccacagaaagcatagtaatatggggaaa
gactcctaaatttaaactgcccatacaaaaggaaacatgggaaacatggtggacagagta
ttggcaagccacctggattcctgagtgggagtttgttaataccctcctttagtgaaatt
atggtaccagttagagaaagaacccatagtaggagcagaaaccttctatgtagatggggc
agctaacagggagactaaattaggaaaagcaggatatgttactaatagaggaagacaaaa
agttgtcaccctaactgacacaacaaatcagaagactgagttacaagcaatttatctagc
tttgcaggattcgggattagaagtaaacatagtaacagactcacaatatgcattaggaat
cattcaagcacaaccagatcaaagtgaatcagagttagtcaatcaaataatagagcagtt
aataaaaaaggaaaaggtctatctggcatgggtaccagcacacaaaggaattggaggaaa
tgaacaagtagataaattagtcagtgctggaatcaggaaagtgcta**gctatg**ggtggca
agtggtcaaaaagtagtgtggttggatggcctactgtaagggaaagaatgagacgagctg
agccagcagcagatggggtgggagcagcatctcgagacctggaaaaacatggagcaatca
caagtagcaatacagcagctaccaatgctgcttgtgcctggctagaagcacaagaggagg
aggaggtgggttttccagtcacacctcaggtacctttaagaccaatgact
tacaaggcagctgtagatcttagccacttttttaaaagaaaaggggggactggaagggcta
attcactcccaacgaagacaagatatccttgatctgtggatctaccacacacaaggctac

ttccctgattggcagaactacacaccagggccaggggtcagatatccactgacctttgga
tggtgctacaagctagtaccagttgagccagataaggtagaagaggccaataaaggagag
aacaccagcttgttacaccctgtgagcctgcatggaatggatgaccctgagagagaagtg
ttagagtggaggtttgacagccgcctagcatttcatcacgtggcccgagagctgcatccg
gagtacttcaagaactgcaggcctatgggtgcgagagcgtcagtattaagcgggggaga
attagatcgatgggaaaaaattcggttaaggccaggggaaagaaaaaatataaattaaa
acatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttaga
aacatcagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatc
agaagaacttagatcattatataatacagtagcaaccctctattgtgtgcatcaaaggat
agagataaaagacaccaaggaagctttagacaagatagaggaagagcaaaacaaaagtaa
gaaaaaagcacagcaagcagcagctgacacaggacacagcaatcaggtcagccaaaatta
ctaa

[SEQ ID NO:14]

p24 sequence is in bold
Nef sequence is underlined
Boxes: nucleotides introduced by genetic construction

• **Amino-Acid sequence**

MVIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATP 50
QDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREP 100
RGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPTS 150
ILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQNANPDCK 200
TILKALGPAATLEEMMTACQGVGGPGHKARVLHMGPISPIETVSVKLKPG 250
MDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAIKK 300
KDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAY 350
FSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMT 400
KILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLT 450
TPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLN 500
WASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELELAENREILKEPVH 550
GVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARKRGAHTNDV 600
KQLTEAVQKITTESIVIWGKTPKFKLPIQKETWETWWTEYWQATWIPEWE 650
FVNTPPLVKLWYQLEKEPIVGAETFYVDGAANRETKLGKAGYVTNRGRQK 700
VVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSES 750
ELVNQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVLAMGGK 800
WSKSSVVGWPTVRERMRRAEPAADGVGAASRDLEKHGAITSSNTAATNAA 850
CAWLEAQEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQ 900
RRQDILDLWIYHTQGYFPDWQNYTPGPGVRYPLTFGWCYKLVPVEPDKVE 950
EANKGENTSLLHPVSLHGMDDPEREVLEWRFDSRLAFHHVARELHPEYFK 1000

NCRPMGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAV 1050
NPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKD 1100
TKEALDKIEEEQNKSKKKAQQAAADTGHSNQVSQNY 1136

[SEQ ID NO:15]

P24 sequence: amino-acids 1-232 (in bold)
RT sequence: amino-acids 235-795
Nef sequence: amino-acids 798-1002
P17 sequence: amino-acids 1005-1136
Boxes:amino-acids introduced by genetic construction
K (Lysine) : instead of Methionine (1st codon)

K (Lysine) : instead of Tryptophan (W). Mutation introduced to remover enzyme activity.

F4* expression in B834(DE3) cells:

F4* recombinant strain was induced at 22°C during 18h, in parallel to F4 non-mutated construct. Crude extracts were prepared and analyzed by Coomassie stained gel and Western blotting.

As illustrated in Figure 8, F4* was expressed at a high level (10% total protein), slightly higher compared to F4 and the small 62 kDa band disappeared.

Western blot analysis:

Reagents: *- pool 3 Mabs anti p24 (JC13.1, JC16.1, IG8.1.1)(dilution 1/5000)*

- rabbit polyclonal anti RT (rabbit PO3L16) (dilution: 1/10 000)

- rabbit polyclonal anti Nef-Tat (rabbit 388) (dilution 1/10 000)

-Alkaline phosphatase-conjugate anti-rabbit antiboby (dilution: 1/7500)

-Alkaline phosphatase-conjugate anti-mouse antiboby (dilution: 1/7500)

Example 5: Construction and expression of F3 and F3* (mutated F3)

F3 (p17-p51-Nef) and F3* (p17-p51*-Nef) in which the putative internal Methionine initiation site replaced by Lysine.

F3 and F3* fusions could be used in combination with p24.

Recombinant plasmids construction:

F3: The sequence encoding p51 was excized (as ScaI and StuI DNA fragment) from pET29a/p51 expression plasmid and ligated into pET28b/p17-Nef plasmid , at the StuI site (located between p17 and Nef gene), as an in frame fusion with p17 and Nef sequences. The resulting fusion construct p17-p51-Nef is named F3.

F3*: Mutation of the putative internal methionine initiation site was achieved using the "Gene Tailor Site-Directed Mutagenesis system" (Invitrogen), generating F3* construct.

F3 and F3* plasmids were used to transform B834 (DE3) cells.

Recombinant protein characteristics:

N-term **p17: 134a.a. - hinge:2a.a. – p51/p51*: 426a.a. -hinge:2a.a. - Nef: 206a.a.** ***C-term***

- **Length Molecular Weight, Isoelectric Point (IP)**

  770 AA, 88.5 kDa, IP:8.58

- **Nucleotide sequence (for F3*)**

```
atgggtgcgagagcgtcagtattaagcgggggagaattagatcgatgggaaaaaattcgg   60
ttaaggccaggggaaagaaaaatataaattaaaacatatagtatgggcaagcagggag   120
ctagaacgattcgcagttaatcctggcctgttagaaacatcagaaggctgtagacaaata   180
ctgggacagctacaaccatcccttcagacaggatcagaagaacttagatcattatataat   240
```

acagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagct 300
ttagacaagatagaggaagagcaaaacaaaagtaagaaaaaagcacagcaagcagcagct 360
gacacaggacacagcaatcaggtcagccaaaattacctcgacaggact**GGTCCGATCTCT** 420
CCGATAGAAACAGTTTCGGTCAAGCTTAAACCAGGGATGGATGGTCCAAAGGTCAAGCAG 480
TGGCCGCTAACGGAAGAGAAGATTAAGGCGCTCGTAGAGATTTGTACTGAAATGGAGAAG 540
GAAGGCAAGATAAGCAAGATCGGGCCAGAGAACCCGTACAATACACCGGTATTTGCAATA 600
AAGAAGAAGGATTCAACAAAATGGCGAAAGCTTGTAGATTTTAGGGAACTAAACAAGCGA 660
ACCCAAGACTTTTGGGAAGTCCAACTAGGTATCCCACATCCAGCCGGTCTAAAGAAGAAG 720
AAATCGGTCACAGTCCTGGATGTAGGAGACGCATATTTTAGTGTACCGCTTGATGAGGAC 780
TTCCGAAAGTATACTGCGTTTACTATACCGAGCATAAACAATGAAACGCCAGGCATTCGC 840
TATCAGTACAACGTGCTCCCGCAGGGCTGGAAGGGGTCTCCGGCGATATTTCAGAGCTCT 900
ATGACAAAAATACTTGAACCATTCCGAAAGCAGAATCCGGATATTGTAATTTACCAATAC 960
ATGGACGATCTCTATGTGGGCTCGGATCTAGAAATTGGGCAGCATCGCACTAAGATTGAG 1020
GAACTGAGGCAACATCTGCTTCGATGGGGCCTCACTACTCCCGACAAGAAGCACCAGAAG 1080
GAGCCGCCGTTCCTAAAGATGGGCTACGAGCTTCATCCGGACAAGTGGACAGTACAGCCG 1140
ATAGTGCTGCCCGAAAAGGATTCTTGGACCGTAAATGATATTCAGAAACTAGTCGGCAAG 1200
CTTAACTGGGCCTCTCAGATTTACCCAGGCATTAAGGTCCGACAGCTTTGCAAGCTACTG 1260
AGGGGAACTAAGGCTCTAACAGAGGTCATCCCATTAACGGAGGAAGCAGAGCTTGAGCTG 1320
GCAGAGAATCGCGAAATTCTTAAGGAGCCGGTGCACAGGGTATACTACGACCCCTCCAAG 1380
GACCTTATAGCCGAGATCCAGAAGCAGGGGCAGGGCCAATGGACGTACCAGATATATCAA 1440
GAACCGTTTAAGAATCTGAAGACTGGGAAGTACGCGCGCAAACGAGGGGCTCATACTAAT 1500
GATGTAAAGCAACTTACGGAAGCAGTACAAAAGATTACTACTGAGTCTATTGTGATATGG 1560
GGCAAGACCCCAAAGTTCAAGCTGCCCATACAGAAGGAAACATGGGAAACATGGTGGACT 1620
GAATATTGGCAAGCTACCTGGATTCCAGAATGGGAATTTGTCAACACGCCGCCGCTGGTA 1680
AAACTG**aggcct**ATGggtggcaagtggtcaaaaagtagtgtggttggatggcctactgta 1740
agggaaagaatgagacgagctgagccagcagcagatggggtgggagcagcatctcgagac 1800
ctggaaaaacatggagcaatcacaagtagcaatacagcagctaccaatgctgcttgtgcc 1860
tggctagaagcacaagaggaggaggaggtgggttttccagtcacacctcaggtaccttta 1920
agaccaatgacttacaaggcagctgtagatcttagccacttttttaaaagaaaaggggggа 1980
ctggaagggctaattcactcccaacgaagacaagatatccttgatctgtggatctaccac 2040
acacaaggctacttccctgattggcagaactacacaccagggccaggggtcagatatcca 2100
ctgacctttggatggtgctacaagctagtaccagttgagccagataaggtagaagaggcc 2160
aataaaggagagaacaccagcttgttacaccctgtgagcctgcatggaatggatgaccct 2220
gagagagaagtgttagagtggaggtttgacagccgcctagcatttcatcacgtggcccga 2280
gagctgcatccggagtacttcaagaactgctaa 2213

[SEQ ID NO:16]

P17: sequence in bold
P51: sequence in capital letter
Nef: sequence in small letter
Boxes: nucleotides introduced by genetic construction

- **Amino-Acid sequence (for F3)**

```
MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQI   60
LGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAA  120
DTGHSNQVSQNYLDRTGPISPIETVSVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEK  180
EGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKK  240
KSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSS  300
MTKILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQK  360
EPPFLKMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLL  420
RGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQ  480
EPFKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQKETWETWWT  540
EYWQATWIPEWEFVNTPPLVKLRPMGGKWSKSSVVGWPTVRERMRRAEPAADGVGAASRD  600
LEKHGAITSSNTAATNAACAWLEAQEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGG  660
LEGLIHSQRRQDILDLWIYHTQGYFPDWQNYTPGPGVRYPLTFGWCYKLVPVEPDKVEEA  720
NKGENTSLLHPVSLHGMDDPEREVLEWRFDSRLAFHHVARELHPEYFKNC            770
```

[SEQ ID NO:17]

P17 sequence: amino-acids 1-134 (in bold)
P51 sequence: amino-acids 137-562
Nef sequence: amino-acids 565-770
Boxes:amino-acids introduced by genetic construction

Methionine 494 replaced by Lysine (K) in F3' construct
K(Lysine) instead of Tryptophan (W). Mutation introduced to remover enzyme activity.

F3 expression in B834(DE3) cells:

F3 expression level and recombinant protein solubility were evaluated, in parallel to F4 (p24-p66-Nef-p17) and p17-Nef (F2) production strains.

Induction condition: cells grown at 37°C / induced at 30°C (+1mM IPTG), during 3h.

Breaking buffers: F4 : 50mMTris/HCl pH:8.0, 50mM NaCl, 1mM EDTA, +/- 1mM DTT

F2: 50mMTris/HCl pH:8.0, 50mM NaCl, 1mM EDTA, without DTT

F3: 50mMTris/HCl pH:7.5, 50mM NaCl, 1mM EDTA, +/- 1mM DTT

Western blot analysis:

reagents - rabbit polyclonal anti RT (rabbit PO3L16) (dilution: 1/10 000)

- rabbit polyclonal anti Nef-Tat (rabbit 388) (dilution 1/10 000)

-Alkaline phosphatase-conjugate anti-rabbit antibody (dilution: 1/7500)

"Fresh" cellular extracts

Cellular fractions corresponding to crude extracts (T), insoluble pellet (P) and supernatant (S) were analyzed on 10% reducing SDS-PAGE. As illustrated in Figure 9, the F3 fusion protein is expressed at a high level (10% total protein). Almost all F3 is recovered in the soluble fraction (S) of cellular extracts, although 5-10% of F4 product are already associated with the pellet fraction. The WB pattern is simplified compared to F4.

F3* expression in B834(DE3) cells:

F3* recombinant strain was induced at 37°C over 3h, in parallel to F3 non-mutated constructed. Crude cellular extracts were prepared and analyzed by Coomassie stained gel and Western blotting. As illustrated in Figure 10, the F3* fusion protein is expressed at a very high level (10-20% total protein). There was a simplified WB pattern compared to F3; a very faint band at +/- 32kDa (detected on WB only) had disappeared.

Example 6: Construction and expression of F4(p51) and F4(p51)*

RT/p51 was used in the F4 fusion construct (in place of RT/p66).

F4(p51) = p24-p51-Nef-p17

F4(p51)* = p24-p51*-Nef-p17 - Mutated F4(p51): putative internal Methionine initiation site (present in RT portion) replaced by Lysine, to further simplify the antigen pattern.

Recombinant plasmids construction:

F4(p51): The sequence encoding p51 was amplified by PCR from pET29a/p51

expression plasmid. Restriction sites were incorporated into the PCR primers (NdeI and StuI at the 5' end. AvrII at the 3' end of the coding sequence). The PCR product was cloned into pGem-T intermediate plasmid and sequenced. pGem-T/p51 intermediate plasmid was restricted by NdeI and AvrII and the p51 fragment was ligated into pET28b/p24-RT/p66-Nef-p17 expression plasmid restricted by NdeI and NheI (resulting in the excision of RT/p66 sequence). Ligation was performed by combining digestion reactions in appropriate concentrations, in the presence of T4 DNA ligase. Ligation product was used to transform DH5α E.coli cells. Verification of insertion of p51 into the correct translational reading frame (in place of RT/p66 in the f4 fusion) was confirmed by DNA sequencing. The resulting fusion construct p24-RT/p51-Nef-p17 is named F4(p51).

F4(p51)*: Mutation of the putative internal methionine initiation site (present in RT/p51) was achieved with "GeneTailor Site-Directed Mutagenesis system" (Invitrogen), generating F4(p51)* construct.

F4(p51) and F4(p51)* expression plasmids were used to transform B834(DE3) cells.

Recombinant proteins characteristics:

N-term **p24: 232a.a.** - **hinge:4a.a.** – **p51/51*: 426a.a.** -**hinge:3a.a.** - **Nef: 206a.a.** - **hinge:2a.a.** - **p17: 132a.a.** - *C-term*

- **Length, Molecular Weight, Isoelectric Point (IP):**
  1005 AA, 114.5 kDa, IP: 8.47

- **Nucleotide sequence (for F4(p51)*)**

```
Atggttatcgtgcagaacatccaggggcaaatggtacatcaggccatatcacctagaact   60
Ttaaatgcatgggtaaaagtagtagaagagaaggctttcagcccagaagtaatacccatg  120
Ttttcagcattatcagaaggagccaccccacaagatttaaacaccatgctaaacacagtg  180
Gggggacatcaagcagccatgcaaatgttaaaagagaccatcaatgaggaagctgcagaa  240
Tgggatagagtacatccagtgcatgcagggcctattgcaccaggccagatgagagaacca  300
Aggggaagtgacatagcaggaactactagtacccttcaggaacaaataggatggatgaca  360
Aataatccacctatcccagtaggagaaatttataaaagatggataatcctgggattaaat  420
```

Aaaatagtaagaatgtatagccctaccagcattctggacataagacaaggaccaaaagaa 480
Ccttttagagactatgtagaccggttctataaaactctaagagccgagcaagcttcacag 540
Gaggtaaaaaattggatgacagaaaccttgttggtccaaaatgcgaacccagattgtaag 600
Actattttaaaagcattgggaccagcggctacactagaagaaatgatgacagcatgtcag 660
GgagtaggaggacccggccataaggcaagagttttgCATATGaggcctGGTCCGATCTCT 720
CCGATAGAAACAGTTTCGGTCAAGCTTAAACCAGGGATGGATGGTCCAAAGGTCAAGCAG 780
TGGCCGCTAACGGAAGAGAAGATTAAGGCGCTCGTAGAGATTTGTACTGAAATGGAGAAG 840
GAAGGCAAGATAAGCAAGATCGGGCCAGAGAACCCGTACAATACACCGGTATTTGCAATA 900
AAGAAGAAGGATTCAACAAAATGGCGAAAGCTTGTAGATTTTAGGGAACTAAACAAGCGA 960
ACCCAAGACTTTTGGGAAGTCCAACTAGGTATCCCACATCCAGCCGGTCTAAAGAAGAAG 1020
AAATCGGTCACAGTCCTGGATGTAGGAGACGCATATTTTAGTGTACCGCTTGATGAGGAC 1080
TTCCGAAAGTATACTGCGTTTACTATACCGAGCATAAACAATGAAACGCCAGGCATTCGC 1140
TATCAGTACAACGTGCTCCCGCAGGGCTGGAAGGGGTCTCCGGCGATATTTCAGAGCTCT 1200
ATGACAAAAATACTTGAACCATTCCGAAAGCAGAATCCGGATATTGTAATTTACCAATAC 1260
ATGGACGATCTCTATGTGGGCTCGGATCTAGAAATTGGGCAGCATCGCACTAAGATTGAG 1320
GAACTGAGGCAACATCTGCTTCGATGGGGCCTCACTACTCCCGACAAGAAGCACCAGAAG 1380
GAGCCGCCGTTCCTAAAGATGGGCTACGAGCTTCATCCGGACAAGTGGACAGTACAGCCG 1440
ATAGTGCTGCCCGAAAAGGATTCTTGGACCGTAAATGATATTCAGAAACTAGTCGGCAAG 1500
CTTAACTGGGCCTCTCAGATTTACCCAGGCATTAAGGTCCGACAGCTTTGCAAGCTACTG 1560
AGGGGAACTAAGGCTCTAACAGAGGTCATCCCATTAACGGAGGAAGCAGAGCTTGAGCTG 1620
GCAGAGAATCGCGAAATTCTTAAGGAGCCGGTGCACAGGGTATACTACGACCCCTCCAAG 1680
GACCTTATAGCCGAGATCCAGAAGCAGGGGCAGGGCCAATGGACGTACCAGATATATCAA 1740
GAACCGTTTAAGAATCTGAAGACTGGGAAGTACGCGCGCAAACGAGGGGCTCATACTAAT 1800
GATGTAAAGCAACTTACGGAAGCAGTACAAAAGATTACTACTGAGTCTATTGTGATATGG 1860
GGCAAGACCCCAAAGTTCAAGCTGCCCATACAGAAGGAAACATGGGAAACATGGTGGACT 1920
GAATATTGGCAAGCTACCTGGATTCCAGAATGGGAATTTGTCAACACGCCGCCGCTGGTA 1980
AAACTGgccctaGCTATGggtggcaagtggtcaaaaagtagtgtggttggatggcctact 2040
Gtaagggaaagaatgagacgagctgagccagcagcagatggggtgggagcagcatctcga 2100
Gacctggaaaaacatggagcaatcacaagtagcaatacagcagctaccaatgctgcttgt 2160
Gcctggctagaagcacaagaggaggaggaggtgggttttccagtcacacctcaggtacct 2220
Ttaagaccaatgacttacaaggcagctgtagatcttagccacttttaaaagaaaagggg 2280
Ggactggaagggctaattcactcccaacgaagacaagatatccttgatctgtggatctac 2340
Cacacacaaggctacttccctgattggcagaactacacaccagggccaggggtcagatat 2400
Ccactgacctttggatggtgctacaagctagtaccagttgagccagataaggtagaagag 2460
Gccaataaaggagagaacaccagcttgttacaccctgtgagcctgcatggaatggatgac 2520
Cctgagagagaagtgttagagtggaggtttgacagccgcctagcatttcatcacgtggcc 2580
CgagagctgcatccggagtacttcaagaactgcAGGCCTATGGGTGCGAGAGCGTCAGTA 2640
TTAAGCGGGGGAGAATTAGATCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAGAAA 2700
AAATATAAATTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAAT 2760
CCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCC 2820
CTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGT 2880
GTGCATCAAAGGATAGAGATAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAG 2940
CAAAACAAAAGTAAGAAAAAAGCACAGCAAGCAGCAGCTGACACAGGACACAGCAATCAG 3000

GTCAGCCAAAATTACtaa 3018

[SEQ ID NO:18]

P24: sequence in bold
P51: sequence in capital letter
Nef: sequence in small letter
P17: sequence underlined
Boxes: nucleotides introduced by genetic construction

- **Amino-Acid sequence (for F4(p51)*)**

```
MVIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTV   60
GGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMT  120
NNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQ  180
EVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVLHMRPGPIS  240
PIETVSVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPVFAI  300
KKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDED  360
FRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQY  420
MDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQP  480
IVLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELEL  540
AENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARKRGAHTN  600
DVKQLTEAVQKITTESIVIWGKTPKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLV  660
KLALAMGGKWSKSSVVGWPTVRERMRRAEPAADGVGAASRDLEKHGAITSSNTAATNAAC  720
AWLEAQEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIHSQRRQDILDLWIY  780
HTQGYFPDWQNYTPGPGVRYPLTFGWCYKLVPVEPDKVEEANKGENTSLLHPVSLHGMDD  840
PEREVLEWRFDSRLAFHHVARELHPEYFKNCRPMGARASVLSGGELDRWEKIRLRPGGKK  900
KYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYC  960
VHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAADTGHSNQVSQNY                1005
```

[SEQ ID NO:19]

P24: amino-acids 1-232
P51: amino-acids 237-662
Nef: amino-acids 666-871
P17: amino-acids 874-1005
K (Lysine) : instead of Methionine (internal "start" codon)
K (Lysine) : instead of Tryptophan (W). Mutation introduced to remove

enzyme activity.

F4(p51) expression in B834(DE3) cells:

F4(p51) expression level and recombinant protein solubility were evaluated, in parallel to F4 expressing strain.

Induction condition: cells grown at 37°C / induced at 22°C (+1mM IPTG), over 19h.
Breaking buffer: 50mMTris/HCl pH:7.5, 1mM EDTA, 1mM DTT
Western blot analysis:
reagents - rabbit polyclonal anti RT (rabbit PO3L16) (dilution: 1/10 000)
- rabbit polyclonal anti Nef-Tat (rabbit 388) (dilution 1/10 000)
-Alkaline phosphatase-conjugate anti-rabbit antiboby (dilution: 1/7500)

Cellular fractions corresponding to crude extracts (T), insoluble pellet (P) and supernatant (S) were analyzed on 10% reducing SDS-PAGE.

As illustrated in Figure 11, F4(p51) was expressed at a high level (10% of total protein), similar to F4. Almost all F4(p51) is recovered in the soluble fraction (S) of cellular extracts. Upon detection with an anti-Nef-tat reagent, F4(p51) the WB pattern was shown to be simplified (reduction of truncated products below +/- 60kDa).

F4(p51)* expression in B834(DE3) cells:

F4(p51)* recombinant strain was induced at 22°C over 18h, in parallel to F4(p51) non-mutated construct, F4 and F4*. Crude cellular extracts were prepared and analyzed by Coomassie stained gel and Western blotting. As illustrated in Figure 12 high expression of F4(p51) and F4(p51)* fusions was observed, representing at least 10% of total protein. WB pattern: reduction of truncated products below +/- 60kDa. In addition, for

F4(p51)* construct, the 47kDa band (due to internal start site) has disappeared.

Example 7: Purification of F4, F4(p51)* and F4* - Purification Method I

The fusion protein F4, comprising the 4 HIV antigens p24-RT-Nef-p17, was purified from a *E. coli* cell homogenate according to purification method I, which comprises the following principal steps:

- Ammonium sulfate precipitation of F4
- SO3 Fractogel cation-exchange chromatography (positive mode)
- Octyl sepharose hydrophobic interaction chromatography (positive mode)
- Q sepharose FF anion-exchange chromatography (positive mode)
- Superdex 200 gel filtration chromatography in presence of SDS
- Dialysis and concentration

Additionally, the F4(p51)* fusion protein (RT replaced by the codon optimized p51 carrying an additional mutation Met592Lys) and the F4* protein (F4 carrying an additional Met592Lys mutation) were purified using the same purification method I.

Protein quantification

- Total protein was determined using the Lowry assay. Before measuring the protein concentration all samples are dialyzed overnight against PBS, 0.1% SDS to remove interfering substances (urea, DTT). BSA (Pierce) was used as the standard.

SDS-PAGE and western blot

- Samples were prepared in reducing or non-reducing SDS-PAGE sample buffer (+/- β–mercaptoethanol) and heated for 5 min at 95°C.
- Proteins were separated on 4-20% SDS-polyacrylamide gels at 200 V for 75 min using pre-cast Novex Tris-glycine gels or Criterion gels (Bio-Rad), 1 mm thick.
- Proteins were visualized with Coomassie-blue R250.

- For the western blots (WB), the proteins were transferred from the SDS-gel onto nitrocellulose membranes (Bio-Rad) at 4°C for 1.5 h at 100 V or overnight at 30 V.
- F4 was detected using monoclonal antibodies against the different antigens, anti-p24, anti-Nef-Tat, anti-RT (sometimes a mixture of anti-p24 and anti Nef-Tat was used to detect a maximum number of protein bands).
- Alkaline-phosphatase conjugated anti-mouse or anti-rabbit antibodies were bound to the primary antibodies and protein bands were visualized using BCIP and NBT as the substrates.

anti-E. coli western blot

- 5 µg protein (Lowry) were separated by SDS-PAGE and transferred onto nitrocellulose membranes as above.
- Residual host cell proteins were detected using polyclonal anti-E. coli antibodies. Protein bands were visualized with the alkaline-phosphatase reaction as above.

Purification Method I

Method I comprises a precipitation by ammonium sulfate and four chromatographic steps:

- E.coli cells were homogenized in 50mM Tris buffer at pH 8.0 in the presence of 10mM DTT, 1mM PMSF, 1mM EDTA at OD50 (~360 ml). 2 Rannie passages were applied at 1000 bars.
- Cells debris and insoluble material were removed by centrifugation at 14400 × g for 20 min.
- Ammonium sulfate (AS) was added from a 3.8M stock solution to the clarified supernatant to a final concentration of 1.2M. Proteins were precipitated for ~2 hours at room temperature (RT) and then pelleted by centrifugation (10 min at 14400 × g). The pellet was resuspended in 8M urea, 10mM DTT in 10mM phosphate buffer at pH 7.0.
- The antigen was captured on a SO3 Fractogel column (Merck) in the presence of 8M urea and 10mM DTT at pH 7.0 in phosphate buffer. The column was washed to elute non-bound protein followed by a pre-elution step with 170mM NaCl to

remove bound host cell proteins (HCP). F4 was then eluted with 460mM NaCl, 8M urea, 10mM DTT in phosphate buffer at pH 7.0.

- The SO3 eluate was 2fold diluted with 10mM phosphate buffer, pH 7, and loaded onto a Octyl sepharose column (Amersham Biosciences) in the presence of 4M urea, 1mM DTT, 230mM NaCl in phosphate buffer at pH 7.0. Following a washing step (equilibration buffer) bound F4 was eluted with 8M urea, 1mM DTT in 25mM Tris buffer at pH 8.0.
- The Octyl eluate was diluted and adjusted to pH 9.0 and F4 was then bound to an Q sepharose column (Amersham Bioscience) in the presence of 8M urea at pH 9.0 (25mM Tris). Unbound protein was washed off (8M urea, 25mM Tris at pH 9.0) and a pre-elution step (90mM NaCl in 8M urea, 25mM Tris, pH 9.0) removed HCP and F4-degradation products. F4 was desorped from the column with 200mM NaCl, 8M urea in Tris buffer at pH 9.0.
- An aliquot of the Q eluate was spiked with 1% SDS and dialyzed against PBS buffer containing 0.1% SDS and 1mM DTT to remove the urea prior to injecting the sample onto the gel filtration column (prep grade Superdex 200, two 16 × 60 cm columns connected in a row). The relevant fractions were pooled after in-process SDS-PAGE analysis.
- Samples were dialyzed twice at RT in dialysis membranes (12-14 kDa cut-off) overnight against 1 l 0.5M Arginine, 10mM Tris, 5mM Glutathione, pH 8.5.

The sequential purification steps are shown in the flowchart below.

Purification Flowsheet

360 ml homogenate OD50 (Rannie)

50 mM Tris pH 8.0, 1 mM PMSF, 10mM DTT, 2 mM EDTA

↓

Clarification

20 min centrifugation at 14400 × g

↓

Ammonium sulfate precipitation

1.2 M AS, 2 h at RT, centrifugation 14400 × g, 10 min

↓

pellet resuspended in 8 M urea, 10 mM PO_4, 10mM DTT, pH 7.0

↓

(+) SO3 Fractogel EMD 650 (M) chromatography

pH 7.0, 8 M urea, 10mM DTT, pre-elution at 170 mM NaCl, elution 460 mM NaCl

↓

2 × dilution to pH 7.0, 4 M urea, 5 mM DTT, 230 mM NaCl

↓

(+) Octyl sepharose chromatography

pH 7.0, 4 M urea 230 mM NaCl, elution 8 M urea, 20 mM Tris pH 8.0

↓

~2 × dilution, adjustment to pH 9.0 (NaOH)

↓

(+) Q Sepharose FF chromatography

Tris pH 9.0, 8 M urea, pre-elution 90 mM NaCl, elution 200mM NaCl

↓

addition of **1% SDS**

↓

dialysis

→ TBS, 0.1% SDS, pH 8.5

↓

Superdex 200 gel filtration chromatography 16 × 120 cm

TBS, 0.1% SDS, pH 8.5

↓

IPA SDS-PAGE

↓

pool/concentration/dialysis

→ formulation compatible buffer

IPA – In process analysis

All buffers contain **1 mM DTT** if not otherwise specified.

Results Purification of F4

SDS-PAGE/Western Blot Follow-up of The Purification Process

Figure 13 shows the SDS gel and the anti-p24/anti-Nef-Tat western blot of the F4-containing fractions collected during the purification of F4.

The E. coli homogenate is shown in Fig.13, lane 2, with F4 estimated to represent about 10% of the total proteins (density scans of Coomassie blue stained SDS-gels). After centrifugation, the soluble fraction of F4 was recovered in the clarified supernatant (lane 3). The ammonium sulfate precipitation step eliminated many impurities (lane 4) and reduced the proteic charge for the subsequent chromatographic step. Additionally, the 8M urea used to resuspend the precipitate dissociated complexes of F4 with HCP and allowed both complete capture of F4 by and quantitative elution from the SO3 resin. The SO3 eluate shown in lane 5 was considerably enriched in F4 but the heterogeneous pattern remained principally unchanged. The hydrophobic Octyl sepharose column mainly removed low molecular weight (LMW) HCP and F4-degradation products (lane 6), thereby simplifying the F4 pattern. The Q sepharose chromatography further simplified the F4 pattern and removed many impurities (lane 7). Final purity in terms of *E.coli* impurities was obtained after this step. In fact, no host cell proteins were detected in the Q eluate by anti-*E. coli* western blot analysis. The purified F4 thus produced is referred to as F4Q. The Superdex 200 column separated LMW F4-degradation products from the full length F4 improving F4 homogeneity in the Superdex 200 eluate (lane 8). The term F4S may be used to refer to F4 purified according to the full scheme of method I.

An anti-*E. coli* western blot was done of the same fractions collected during the purification of F4. The absence of visible bands on the anti-*E. coli* western blot indicated HCP contamination below 1% in the Q eluate and in the Superdex eluate.

F4 and Protein Recovery

F4 recovery at each step of the purification process was estimated from SDS-PAGE and western blot analysis. To estimate F4 recovery from SDS-gels, the sample volumes loaded onto the SDS-gels corresponded to the volumes of the different fractions collected during the purification.

Table 1 displays the protein recovery in the F4-containing fractions.

Table 1: Protein recovery in the F4-positive fractions collected during the purification process (360 ml homogenate). The protein concentration was determined with the Lowry assay.

Purification Step	Protein (mg)	Step Recovery (%)	Cum. Recovery (%)
homogenate	6500	100	100
clarified homogenate	4641	71	71
resuspended AS precipitate	728	16	11
SO3 eluate	247	34	3.8
Octyl sepharose eluate	129	52	2.0
Q sepharose eluate	74	57	1.1
Superdex 200	36	49	0.6

The table shows the amount of protein in the homogenate and the soluble material, including F4, recovered in the supernatant after the clarification step. The AS-precipitation step removed a great amount of HCP and only a slight loss of F4 was observed on the SDS-gel. The SO3 chromatography additionally removed many impurities and the SDS-gel indicated a high recovery of F4. In contrast, the ~50% protein recovery measured with both the Octyl sepharose and the Q sepharose columns were also

accompanied by losses of F4. Protein recovery after the gel filtration chromatography was about 50%. The SDS-gel shows that many LMW-protein bands (F4-degradation bands) were removed, concomitantly reducing F4 recovery.

F4 Yield

Table 1 above shows that about 36 mg purified F4 could be obtained from 360 ml homogenate at OD50. Therefore, 1 l homogenate at OD 50 should yield about 100 mg purified F4. Since ODs of 70 – 90 were achieved during the fermentation process, the yield per liter fermenter would accordingly be in the range of 140 to 180 mg F4.

Results Purification of F4(p51)*

The F4(p51)* fusion construct was purified using purification method I described above without modifications.

SDS-PAGE/Western Blot Follow-up of The Purification Process

Figure 14 shows the SDS gel and the anti-p24/anti-Nef-Tat western blot of the F4(p51)*-containing fractions collected during the purification of F4(p51*).

The SDS-gel and the western blot demonstrate that the F4(p51)* fusion protein globally behaved similarly to F4 at the ammonium sulfate precipitation step as well as during the chromatographic steps. Purified F4(p51)* had a heterogeneity pattern similar to purified F4.

An anti *E. coli* western blot indicated that HCP contamination was below 1% in both the Q eluate and the Superdex eluate.

Yield

About 25% of F4(p51)* were lost in the insoluble fraction of the homogenate. Additionally, because the purification method was not adapted to this protein, losses were observed at the chromatographic steps. Therefore the overall recovery of F4(p51)* was reduced to about 25 mg per liter homogenate (OD50). Extrapolated to 1 litre culture at OD 177, the yield would accordingly be in the range of 85 mg F4(p51)*.

Results Purification of F4*

The F4* fusion construct was purified using purification method I described above without modifications.

SDS-PAGE/Western Blot Follow-up of The Purification Process

Figure 15 shows the SDS gel and the anti-p24/anti-Nef-Tat western blot of the F4*-containing fractions collected during the purification of F4*.

As with F4(p51)* it can also be noted that F4* globally behaved quite similarly to F4 during the purification procedure. The protein was recovered in the expected fractions as shown by the SDS-gel and the western blot. An anti-*E. coli* western blot also demonstrated elimination of most HCP already after the Q sepharose column.

Yield

The global recovery was about 17 mg purified F4* obtained from 465 ml homogenate OD50. Extrapolated to 1 l culture at OD 140, the yield would accordingly be in the range of 100 mg F4*.

In summary, the three fusion proteins F4, F4(p51)* and F4* were purified employing purification method I. The SDS gel in Figure 16 compares the three purified proteins showing the different level of heterogeneity of the constructs after the Q sepharose step and after elimination of LMW bands by the Superdex 200 column.

Example 8: Purification of F4 and F4co (codon optimized) – Purification Method II

Purification Method II

A simplified purification procedure, method II as compared to method I, was also developed. Method II consists of only 2 chromatographic steps and a final dialysis/diafiltration for buffer exchange. Notably, a CM hyperZ chromatographic column (BioSepra) was introduced to replace the clarification step, the ammonium sulfate precipitation and the SO3 chromatography of method I (Example 7). Method II was used to purify both F4 and full-codon optimized F4 ("F4co"). For F4co, two different forms of method II were performed, one involving carboxyamidation and one not. The purpose of the carboxyamidation step was to prevent oxidative aggregation of the protein. This carboxyamidation is performed after the 1st chromatographic step (CM hyperZ).

- E.coli cells (expressing F4 or F4co) were homogenized in 50mM Tris buffer at pH 8.0 in the presence of 10mM DTT, at OD90. 2 Rannie passages were applied at 1000 bars.
- 8M urea were added to the homogenate before application to the CM hyperZ resin (BioSepra) equilibrated with 8M urea in phosphate buffer at pH 7. Antigen capture was done in a batch mode. The resin was then packed in a column, unbound proteins were washed off with the equilibration buffer and bound host cell proteins (HCP) were removed by a pre-elution step with 120mM NaCl. F4co was then eluted with 360mM NaCl, 8M urea, 10mM DTT in phosphate buffer at pH 7.0.
- To control oxidative aggregation of the fusion protein, the cysteine groups of F4co can be carboxyamidated with idoacetamide. Therefore, optionally, 50 mM iodoacetamide was added to the CM hyperZ eluate and carboxyamidation was done for 30 min at room temperature in the dark.
- The CM hyperZ eluate was then adequately diluted (about 5–8 fold) and adjusted to pH 9.0. F4co or F4coca was then bound to a Q sepharose column (Amersham Bioscience) in the presence of 8M urea in Tris buffer at pH 9.0. Unbound protein was washed off with the equilibration buffer and a pre-elution step with 90mM

NaCl (only with non-carboxyamidated protein) in the same buffer removed bound HCP. F4co was desorped from the column with 200mM NaCl, 8M urea in Tris buffer at pH 9.0.

- Samples were dialyzed twice at RT in dialysis membranes (12-14 kDa cut-off) overnight against 1 l 0.5M Arginine, 10mM Tris buffer, 10mM Glutathione (only added to the non-carboxyamidated protein), pH 8.5. Alternatively, buffer exchange was accomplished by diafiltration against 10 sample volumes of the same buffer using a tangential-flow membrane with 30 or 50 kDa cut-off.
- Finally, the dialyzed product was sterile filtered through a 0.22 µm membrane.

The sequential purification steps are shown in the flowchart below.

Purification Flowsheet

Homogenate OD90 (Rannie)
50mM Tris pH 8.0, 10mM DTT
↓
addition of **8M urea,** adjusting to **pH 7.0**
↓
(+) CM hyperZ chromatography
pH 7.0, 8M urea, 10mM DTT, pre-elution at 120mM NaCl, elution 360mM NaCl
↓
optional carboxyamidation: addition of 50mM iodoacetamide, 30 min at RT
↓
dilution and **adjusting** to pH 9.0, 8M urea
↓
(+) Q Sepharose FF chromatography
Tris pH 9.0, 8 M urea, pre-elution, elution with NaCl*
↓
dialysis/diafiltration
→ phosphate buffer, 0.5M Arginine, pH 8.5 (10mM Glutathione)
↓

Sterile filtration

All buffers contained DTT if F4co was not carboxyamidated and Glutathione in the purified bulk. Reducing agents were omitted once the protein was carboxyamidated. *NaCl – for F4co this was 200mM NaCl, for F4coca elution was by gradient of NaCl. This step can be further optimized for F4coca by pre-eluting with 60mM NaCl and eluting with 100mM NaC; and for F4co by eluting with 100mM NaCl (no pre-elution step needed).

Results: Purification of F4co

Figure 17 shows a SDS gel of the F4-containing fractions collected during the purification of F4co and the purification of carboxyamidated F4co ("F4coca").

The CM hyperZ resin completely captured F4co from the crude homogenate (lane 1) in the presence of 8M urea and quantitative elution was achieved with 360mM NaCl. The CM hyperZ eluate shown in lane 2 was considerably enriched in F4co. After appropriate dilution and adjustment of the sample to pH 9, F4co or F4coca was bound to a Q sepharose column. F4co or F4coca was then specifically eluted with 200mM NaCl as shown in lane 3. This chromatography not only removed remaining host cell proteins but also DNA and endotoxins. To bring the purified material in a formulation-compatible buffer, the Q sepharose eluate was dialyzed against 10mM Tris buffer, 0.5M Arginine, 10mM Glutathione pH 8.5 in a dialysis membrane with 12-14 kDa cut-off. Glutathione was omitted with the carboxyamidated protein.

Purification of both F4co and F4coca yielded about 500 mg purified material per L of culture OD130. This was in a similar range as observed before with the non-codon-optimized F4.

As described above, two different purification methods (I and II) have been developed to purify the different F4 constructs. Figure 18 compares the different purified bulks that were obtained.

The SDS gel in Figure 18 clearly illustrates the distinct pattern of the two different proteins, F4 and F4co. Whereas F4 presented several strong low molecular weight (LMW) bands, only faint bands were visible with the codon-optimized F4co. Method I and method II produce a very similar F4co pattern. Anti-*E. coli* western blot analysis confirmed the purity of the purified proteins indicating host cell protein contamination below 1% in all the preparations.

Example 9: Immunogenicity of F4 in mice

Formulation:

Adjuvant formulation 1B:

To prepare Adjuvan t formulation 1 B, A mixture of lipid (such as phosphatidylcholine either from egg-yolk or synthetic) and cholesterol and 3 D- MPL in organic solvent, is dried down under vacuum (or alternatively under a stream of inert gas). An aqueous solution (such as phosphate buffered saline) is then added, and the vessel agitated until all the lipid is in suspension. This suspension is then microfluidised until the liposome size is reduced to about 100 nm, and then sterile filtered through a 0.2 µm filter. Extrusion or sonication could replace this step.

Typically the cholesterol:phosphatidylcholine ratio is 1:4 (w/w), and the aqueous solution is added to give a final cholesterol concentration of 5 to 50 mg/ml. .

The liposomes have a defined size of 100 nm and are referred to as SUV (for small unilamelar vesicles). If this solution is repeatedly frozen and thawed the vesicles fuse to form large multilamellar structures (MLV) of size ranging from 500nm to 15 µm.
The liposomes by themselves are stable over time and have no fusogenic capacity.

QS21 in aqueous solution is added to the liposomes to reach a final 3 D- MPL and QS21 concentrations of 100 μg/ml.

Formulation 2A: **3 De acylated monophoshphoryl lipid A and QS21 in an oil in water emulsion;**

Preparation of oil in water emulsion can be made by following the protocol as set forth in WO 95/17210. In detail the emulsion contains: 5% Squalene 5% tocopherol 2.0% tween 80; the particle size is 180 nm.

Preparation of Oil in water emulsion (2 fold concentrate)

Tween 80 is dissolved in phosphate buffered saline (PBS) to give a 2% solution in the PBS. To provide 100 ml two fold concentrate emulsion 5g of DL alpha tocopherol and 5ml of squalene are vortexed to mix thoroughly. 90ml of PBS/Tween solution is added and mixed thoroughly. The resulting emulsion is then passed through a syringe and finally microfluidised by using an M110S microfluidics machine. The resulting oil droplets have a size of approximately 180 nm.

Sterile bulk emulsion is added to PBS to reach a final concentration of 500 μl of emulsion per ml (v/v). 3 D- MPL is then added to reach a final concentration of 100μg. QS21 is then added to reach a final concentration of 100μg per ml. Between each addition of component, the intermediate product is stirred for 5 minutes

F4Q not codon optimized, purified according to purification method I, was diluted in a phosphate/Arginine buffer pH 6.8. The dilution was mixed with two different concentrated adjuvants (adjuvants 2A and 1B) in order to obtain a final formulation of 40μg/dose of 500μl of F4 in presence of 290 (for adjuvant 2A) - 300 (for adjuvant 1B) mM Argnine, 50μg MPL and 50μg QS21. 100μl of each formulation were injected in mice.

Mouse immunogenicity studies were performed to evaluate the cellular and humoral immune responses to the four antigens found within F4 (p24, p17, RT and Nef).

Due to the complexicity of the F4 antigen, eight strains of mice, each with a different genetic background, were immunised twice at day 0 and day 21 with 8μg of adjuvanted F4 protein prepared as described above, in a 100 μl volume. Serum and spleen samples were collected 14 days following the last immunisation (day 35) for analysis of the humoral and cellular responses to each of the four components of F4 (p24, p17, RT and Nef), as well as F4.

Total antibody responses were characterised by ELISAs specific for p24, p17, RT, Nef and F4. The following table, Table 2, summarises where antigen specific humoral responses were observed in each strain. The results indicate the presence or absence of antibodies compared to control animals immunized with adjuant alone. The results presented are a compilation from two separate but identical experiments. In the table, 2A refers to antigen formulated with 3D-MPL and QS21 in an oil in water emulsion and 1B refers to antigen formulated with 3D-MPL, QS21 and cholesterol containing liposomes.

Table 2

mouse strain	p17	p24	Nef	RT	F4
CB6F1	+/- +2A -1B	+	+	+	+
Balb/c	- +/-2A -1B	+	+	+	+
C3H	-	-	-	-	-
DBA	-	+	+	+	+
CBA	-	-	+/- +2A -1B	+ +2A +/-1B	+
129Sv	-	+	+	+	+
B6D2F1	+/- +2A -1B	+	+	+	+
OF1	+	+	+	+	+

+ =presence of antibodies - = absence of antibodies

OF1 mice mounted antibody responses to all four F4 components. The responses observed are shown in Figure 19. +/- indicates that the response observed was weak or only observed with one of the two adjuvant. For example, B6D2F1 mice p17 responses: +/- overall with a +2A and –1B means that there was a response with 2A (not weak) and none with 1B. Balb/c mice p17 responses: - overall, with a +/- 2A and a – 1B, here the +/- means that the response with adjuvant 2A was weak.

.

Cellular responses were characterised by flow cytometry staining for CD4 and CD8, IFNγ and IL-2 expression (intracellular cytokine staining for IFNγ and IL-2 expression), following restimulation of spleen cells with p24, p17, RT or Nef specific peptides, using peptide library pools of 15 mers with ll mer overlap. CD4 responses were the dominant cellular response observed. The following table, Table 3, summarises where antigen specific CD4+IL-2+ responses were observed for each mouse strain. Again, this is shown as presence or absence of a response.

Table 3

mouse strain	p17	p24	Nef	RT
CB6F1	- +/-2A -1B	+	+	+
Balb/c	-	+/- weak	+/- weak	+
C3H	+	+	-	+
DBA	+	+	+	+
CBA	+	+ -2A +1B	-	+
129Sv	+	+	-	+
B6D2F1	-	+	+	+
OF1	-	-	-	+

+ = presence of CD4+IL-2+ - = absence of CD4+IL-2+

DBA mice mounted CD4 responses to all four F4 components. The CD4+IL-2+ and CD4+IFNγ+ responses observed for this mouse strain are shown in Figure 20.

In summary, F4 formulated in either of the two adjuvant formulations is able to promote humoral and cellular responses to p24, p17, RT and Nef. This shows that each region of F4 is immunogenic in an *in vivo* situation.

CLAIMS

1. A polypeptide which comprises Nef or an immunogenic fragment or derivative thereof, and p17 Gag and/or p24 Gag or immunogenic fragments or derivatives thereof, wherein when both p17 and p24 Gag are present there is at least one HIV antigen or immunogenic fragment between them.

2. The polypeptide according to claim 1 further comprising Pol or RT or an immunogenic fragment or derivative thereof.

3. The polypeptide according to claim 1 or claim 2 wherein the RT or immunogenic fragment is a fragment in which RT is truncated at the C terminus such that it lacks the carboxy terminal RNase H domain.

4. The polypeptide according to claim 3 wherein the RT fragment is the p51 fragment.

5. The polypeptide according to any one of claims 2 to 4 wherein the RT comprises a mutation at position 592 to replace the methionine by another residue e.g. lysine.

6. The polypeptide according to any one of claims 1 to 4 wherein the Nef is full length Nef.

7. A polypeptide selected from one of the following:

1. p24 – RT – Nef – p17
2. p24 – RT* – Nef – p17
3. p24 – p51RT – Nef – p17
4. p24 – p51RT* - Nef – p17
5. p17 – p51RT – Nef
6. p17 – p51RT* - Nef

7. Nef – p17

8. Nef – p17 with linker

9. p17 – Nef

10. p17 – Nef with linker

* represents RT methionine$_{592}$ mutation to lysine

8. A process for purifying a polypeptide according to any one of claims 1 to 7 which process comprises:

i). providing a composition comprising the unpurified polypeptide;

ii). Subjecting the composition to at least two chromatographic steps;

iii). Optionally carboxyamidating the polypeptide;

iv) Performing a buffer exchange step to provide the protein in a suitable buffer for a pharmaceutical formulation.

9. The process according to claim 8 wherein there are no more than two chromatographic steps.

10. The process according to claim 8 or claim 9 wherein the carboxyamidation is performed between the two chromatographic steps.

11. A composition comprising (i) a polypeptide which comprises Nef or an immunogenic fragment or derivative thereof and p17 Gag or an immunogenic fragment or derivative thereof, and (ii) p24 Gag or an immunogenic fragment or derivative; or (i) a polypeptide which comprises Nef or an immunogenic fragment or derivative thereof and p24 Gag or an immunogenic fragment or derivative thereof, and (ii) p17 Gag or an immunogenic fragment or derivative thereof.

12. The composition according to claim 11 further comprising RT or an immunogenic fragment or derivative thereof.

13. A composition comprising (i) a polypeptide which comprises Nef or an immunogenic fragment or derivative thereof and p17 Gag and/or p24 Gag or immunogenic fragments or derivatives thereof, wherein when both p17 and p24 Gag are present there is at least one HIV antigen or immunogenic fragment between them and (ii) a p51 RT polypeptide.

14. A composition according to any one of claims 11 to 13 wherein (i) is selected from

1. Nef – p17
2. Nef – p17 with linker
3. p17 – Nef
4. p17 – Nef with linker

15. A composition according to any one of claims 12 to 14 wherein the RT or immunogenic fragment thereof is a fragment in which RT is truncated at the C terminus such that it lacks the carboxy terminal RNase H domain.

16. The composition according to claim 15 wherein the RT fragment is the p51 fragment.

17. The composition according to any one of claims 11 to 16 wherein the RT comprises a mutation at position 592 such that the methionine is replaced by another residue e.g. lysine.

18. The composition according to any one of claims 11 to 17 wherein the Nef is full length Nef.

19. A polypeptide which is a fusion of HIV antigens comprising at least four HIV antigens or immunogenic fragments or derivatives, wherein the four antigens or fragments are or are derived from Nef, Pol and Gag.

20. The polypeptide according to claim 19 wherein the Gag is present as two separate components which are separated by at least one other antigen in the fusion.

21. The polypeptide according to claim 19 or 20 wherein the Nef is full length Nef.

22. The polypeptide according to any one of claims 19 to 21 wherein the Pol is p66 or p51RT.

23. The polypeptide according to any one of claims 19 to 22 wherein the Gag is p17 and p24 Gag.

24. A polynucleotide or polynucleotides encoding a polypeptide or composition of polypeptides or fusion protein according to any one of claims 1 to 23.

25. A p51 RT polypeptide or derivative thereof or a polynucleotide encoding it, preferably codon optimized for expression in *E. coli.*

26. A polypeptide encoded by the polynucleotide according to claim 25.

27. A pharmaceutical composition comprising a polypeptide or polynucleotide or composition of polypeptides or polynucleotides according to any previous claim or a polypeptides purified according to any previous claim, together with a pharmaceutically acceptable carrier or adjuvant.

28. A pharmaceutical composition according to claim 27 wherein the adjuvant is a Th1 inducing adjuvant such as QS21 or 3D-MPL or a combination of QS21 and 3D-MPL.