US 20020022033A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2002/0022033 A1

Alizon et al. (43) Pub. Date: Feb. 21, 2002

(54) VARIANT OF LAV VIRUSES

(75) Inventors: **Marc Alizon**, Paris (FR); **Pierre Sonigo**, Paris (FR); **Simon Wain-Hobson**, Montigny Les Bretonneux (FR); **Luc Montagnier**, Le Plessis Robinson (FR)

Correspondence Address:
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER LLP
1300 I STREET, NW
WASHINGTON, DC 20005 (US)

(73) Assignee: **Institut Pasteur**

(21) Appl. No.: **09/834,627**

(22) Filed: **Apr. 16, 2001**

Related U.S. Application Data

(60) Division of application No. 09/048,580, filed on Mar. 27, 1998, now Pat. No. 6,284,454, which is a division of application No. 08/154,397, filed on Nov. 18, 1993, now Pat. No. 5,773,602, which is a continuation of application No. 07/988,530, filed on Dec. 10, 1992, now abandoned, which is a continuation of application No. 07/656,797, filed on Feb. 19, 1991, now abandoned, which is a division of application No. 07/038,330, filed on Apr. 13, 1987, now Pat. No. 5,030,714.

(30) **Foreign Application Priority Data**

Jun. 23, 1986 (EP) 86401380.0

Publication Classification

(51) **Int. Cl.**[7] **C12Q 1/68**; G01N 33/53; C07H 21/04; A61K 39/12; C07K 5/00; C07K 7/00; C07K 16/00; C07K 17/00; A61K 38/00; C12N 5/00; C12N 5/02; C12P 21/08; C12N 1/00

(52) **U.S. Cl.** **424/186.1**; 536/23.72; 536/24.3; 536/24.32; 435/6; 530/324; 435/7.1; 435/810; 435/975; 530/387.9; 435/325

(57) ABSTRACT

A variant of a LAV virus, designated LAV_{MAL} and capable of causing AIDS. The cDNA and antigens of the LAV_{MAL} virus can be used for the diagnosis of AIDS and pre-AIDS.

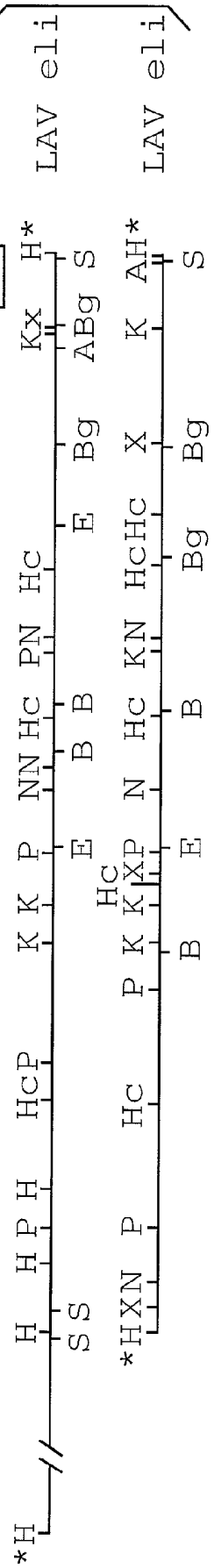

LAV eli
LAV eli
gag
pol
Q
S
env
F
*H
H
H P H
HcP
K K P
NN Hc PN Hc
E B B E Bg
Hc
*H X N P
S S
Hc
P K K X P
N Hc KN HcHc X
B E B Bg Bg
Kx H*
ABg S
K AH*
S

FIG. 1A

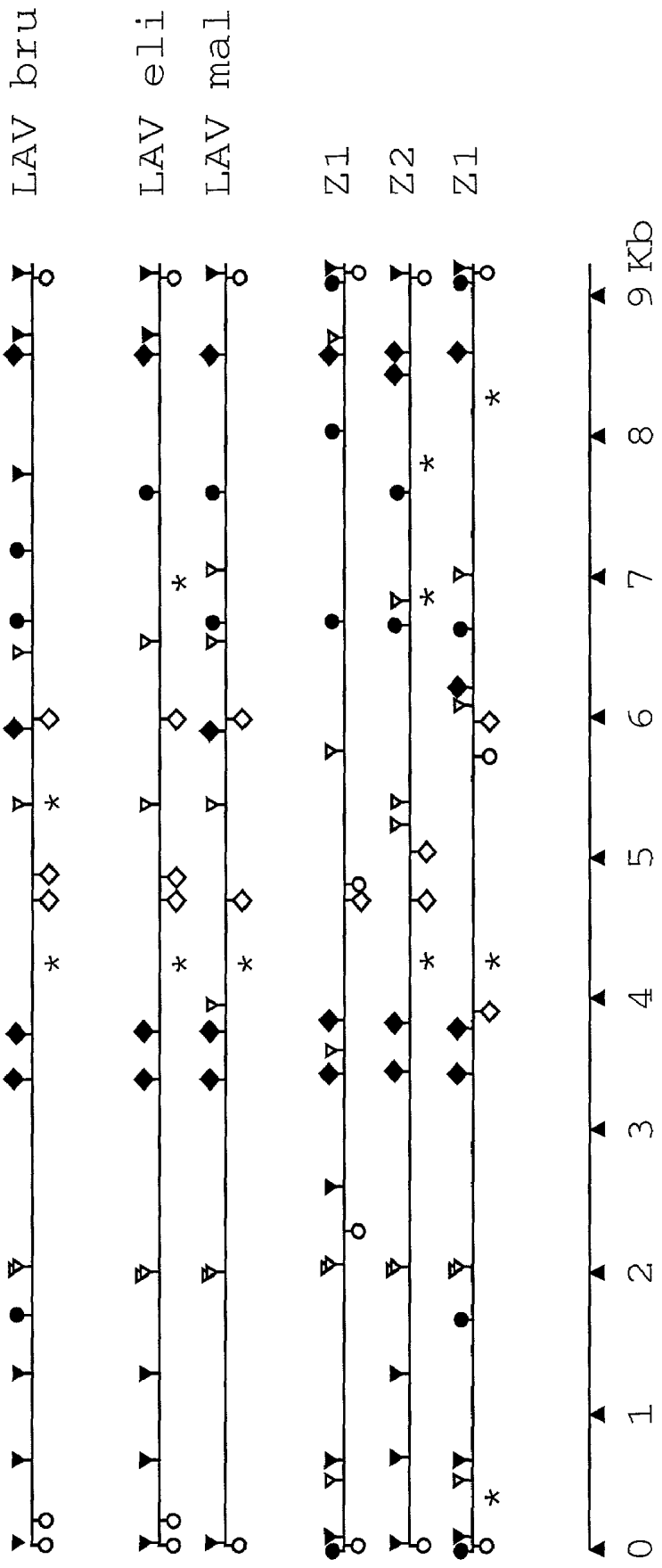


FIG. 1B

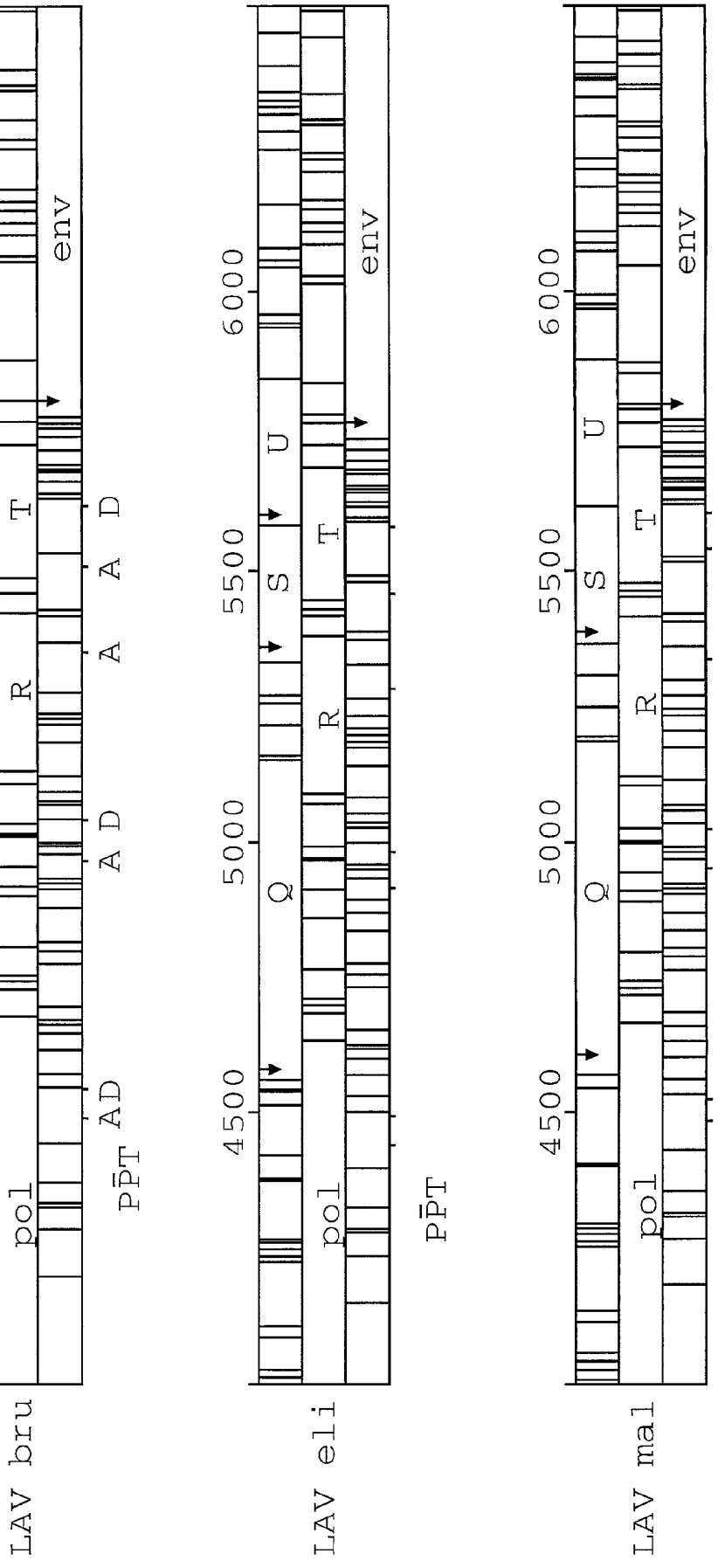

LAV bru
LAV eli
LAV mal
4500
5000
5500
6000
pol
env
Q
R
S
T
U
A
D
PPT
AD

FIG. 2

GAG

```
                10         20         30         40         50         60         70         80
LAV BRU  MGARASVLSG GELDRWEKIR LRPGGKKKYK LKHIVWASRE LERFAVNPGL LETSEGCRQI LGQLQPSLQT GSEELRSLYN
ARV 2                   K
LAV MAL              K  A               R    L           L        C   Q   ME   ST K      IK
LAV ELI              K  K               R              Y L            K   I     AI    T
                                                                        p25
                90         100        110        120        130        140        150        160
LAV BRU  TVATLYCVHQ RIEIKDTKEA LDKIEEEQNK SKKKAQQAAA ------DTGH SSQVSQNYPI VQNIQGQMVH QAISPRTLNA
ARV 2                 DV       E                    ----AAG  N                L
LAV MAL               DV            I     RQ T      AQQAAAA KN   S           A    I
LAV ELI           K G DV       E M                    ------   N N             L

               170        180        190        200        210        220        230        240
LAV BRU  WVKVVEEKAF SPEVIPMFSA LSCGATPQDL NTMLNTVGGH QAAMQMLKET INEEAAEWDK VHPVHAGPIA PGQMREPRGS
ARV 2
LAV MAL      I                          M   I            D        D             P
LAV ELI      I

               250        260        270        280        290        300        310        320
LAV BRU  DIAGTTSTLQ EQIGWMTNNP PIPVGEIYKR WIILGLNKIV RMYSPTSILD IRQGPKEPFR DYVDRFYKTL RAEQASQEVK
ARV 2                                                                                      D
LAV MAL                  S        D                    V                      F         T
LAV ELI               A  S                  V          V                                D
```

FIG. 3A-1

```
                                                                                p13
                330        340        350        360        370        380        390        400
LAV BRU  NWMTETLLVQ NANPDCKTIL KALGPAATLE EMMTAACQGVG GPGHKARVLA EAMSQVTNS- ATIMMQRGNF RNQRKIVKCF
ARV 2                                                                 P-  N             T
LAV MAL                          G                  S             A   T  A        KG - RI
LAV ELI                          Q                  S             A   V T A        KGP   I

               410        420        430        440        450        460        470        480
LAV BRU  NCGKEGHIAR NCRAPRKKGC WKCGKEGHQM KDCTERQANF LGKIWPSYKG RPGNFLQSRP EPTAPPFLQS RPEPTAPPEE
ARV 2             K             R  R                                                ----  --------
LAV MAL         L                                      H                            ----  --------A
LAV ELI           K             R       L              R    H                       ----  --------A

                490        500        510
LAV BRU  SFRSGVETTT PSQKQEPIDK ELYPLTSLRS LFGNDPSSQ
ARV 2       F E K
LAV MAL    GF E IK-       QK         A  K       QL
LAV ELI    GF E I -       QK            K        L
```

FIG. 3A-2

```
CENTRAL REGION:  Q
                 10         20         30         40         50         60         70         80
LAV BRU  MENRWQVMIV WQVDRMRIRT WKSLVKHHMY VSGKARGWFY RHHYESPHPR ISSEVHIPLG DARLVITTYW GLHTGERDWH
ARV 2                                    I K  K          T    V           K              E
LAV MAL                       H            K  KN         R  K V              VR      Q   K
LAV ELI                     K              K  NR            K            E     K          E

                 90         100        110        120        130        140        150        160
LAV BRU  LGQGVSIEWR KKRYSTQVDP ELADQLIHLY YFDCFSDSAI RKALLGHIVS PRCEYQAGHN KVGSLQYLAL AALITPKKIK
ARV 2         A       K        G        H       E    KN I  YR                                T
LAV MAL    H        Q      L   D                E     Q I          D                  T   A   TR
LAV ELI              R        G       M        E       I  D                          T   A  Q

                170        180        190
LAV BRU  PPLPSVTKLT EDRWNKPQKT KGHRGSHTMN GH
ARV 2          K
LAV MAL        R            Q
LAV ELI        R            Q  R
```

FIG. 3B-1

```
                R
                10         20         30         40         50         60         70         80
LAV BRU  MEQAPEDQGP QREPHNEWTL ELLEELKNEA VRHFPRIWLH GLGQHIYETY GDTWAGVEAI IRILQQLLFI HFRIGCRHSR
ARV 2                    Y             R         P    S   Y                                   Q
LAV MAL       A                        Q              S                E         S             Q
LAV ELI       A          Y   A         S              S                V                       Q

                90
LAV BRU  IGVTQQRRAR -NGASRS
ARV 2      II       R
LAV MAL    I R      -  S
LAV ELI    IIR      -  S

             S (tat)

                10         20         30         40         50         60         70
LAV BRU  MEPVDPRLEP WKHPGSQPKT ACTTCYCKKC CFHCQVCFTT KALGISYGRK KRRQRRRPPQ GSQTHQVSLS KQ
ARV 2          N           R    NN          YA   R  G               A   D     A
LAV MAL   D    N     N     R  P NK        Y   M  I  G                    N A  DP P E
LAV ELI   D    N     N     R  P NK H      Y  P   LN  G             G     G A   PIP
```

FIG. 3B-2

POL

```
                10         20         30         40         50         60         70         80
LAV BRU  FFREDLAFLQ GKAREFSSEQ TRANSPTFSS EQTRANSPTR RELQVWGRDN NSLSEAGADR QGTVSFNFPQ ITLWQRPLVT
ARV 2                              ---  ---------      GE
LAV MAL      N   P        P        ---  ---------S     R   G - KT   T  E   I    S          V
LAV ELI      N   P     G L PK      ---  ---------S     R     -  P  KT  E                      A

                90         100        110        120        130        140        150        160
LAV BRU  IKIGGQLKEA LLDTGADDTV LEEMSLPGRW KPKMIGGIGG FIKVRQYDQI LIEICGHKAI GTVLVGPTPV NIIGRNLLTQ
ARV 2     R                          N   K                      PV
LAV MAL  VRV                        IN   K                           K      I              M
LAV ELI                             N   K                      P     Q

               170        180        190        200        210        220        230        240
LAV BRU  IGCTLNFPIS PIETVPVKLK PGMDGPKVKQ WPLTEEKIKA LVEICTEMEK EGKISKIGPE NPYNTPVFAI KKKDSTKWRK
ARV 2
LAV MAL                             R                T    KD       L
LAV ELI                                              T     D        R          I

               250        260        270        280        290        300        310        320
LAV BRU  LVDFRELNKR TQDFWEVQLG IPHPAGLKKK KSVTVLDVGD AYFSVPLDED FRKYTAFTIP SINNETPGIR YQYNVLPQGW
ARV 2                                                          K
LAV MAL    N
LAV ELI                                                                 S
```

FIG. 3C-1

```
                330        340        350        360        370        380        390        400
LAV BRU  KGSPAIFQSS MTKILEPFRK QNPDIVIYQY MDDLYVGSDL EIGQHRTKIE ELRQHLLRWG LTTPDKKHQK EPPFLWMGYE
ARV 2                                                                       F
LAV MAL                    T K  E                                  E   K   F
LAV ELI                        EM                              K  E       F R

                410        420        430        440        450        460        470        480
LAV BRU  LHPDKWTVQP IVLPEKDSWT VNDIQKLVGK LNWASQIYPG IKVRQLCKLL RGTKALTEVI PLTEEAELEL AENREILKEP
ARV 2                M                           A     K
LAV MAL              Q  D E                            K         A    DIV    A
LAV ELI            S  K    E         N  ER

                490        500        510        520        530        540        550        560
LAV BRU  VHGVYYDPSK DLIAEIQKQG QGQWTYQIYQ EPFKNLKTGK YARTRGAHTN DVKQLTEAVQ KITTESIVIW GKTPKFKLPI
ARV 2      E          V                               M                   VS          I
LAV MAL                                      QY            IKS                   AQ             R
LAV ELI                        H                         M           A     R S         R    R
```

FIG. 3C-2

```
                570        580        590        600        610        620        630        640
LAV BRU  QKETWETWWT EYWQATWIPE WEFVNTPPLV KLWYQLEKEP IVGAETFYVD GAASRETKLG KAGYVTNRGR QKVVTLTDTT
ARV 2          A  M                                            N             D        SIA
LAV MAL        A                              T               N    K        D        S  E
LAV ELI           A                                  I        N             D        P

                650        660        670        680        690        700        710        720
LAV BRU  NQKTELQAIH LALQDSGLEV NIVTDSQYAL GIIQAQPDKS ESELVNQIIE QIIKKEKVYL AWVPAHKGIG GNEQVDKLVS
ARV 2                                                    S
LAV MAL                    S                          I          Q D     S
LAV ELI           N

                730        740        750        760        770        780        790        800
LAV BRU  AGIRKVLFLD GIDKAQDEHE KYHSNWRAMA SDFNLPPVVA KEIVASCDKC QLKGEAMHGQ VDCSPGIWQL DCTHLEGKVI
ARV 2             N        E                                                                  I
LAV MAL  S                 E                    I
LAV ELI Q                  E       N

                810        820        830        840        850        860        870        880
LAV BRU  LVAVHVASGY IEAEVIPAET GQETAYFLLK LAGRWPVKTI HTDNGSNFTS TTVKAACWWA GIKQEFGIPY NPQSQGVVES
ARV 2
LAV MAL  I                           I          VV            AA         N
LAV ELI                              VV            AA
```

FIG. 3D-1

```
                890        900        910        920        930        940        950        960
LAV BRU  MNKELKKIIG QVRDQAEHLK TAVQMAVFIH NFKRKGGIGG YSAGERIVDI IATDIQTKEL QKQITKIQNF RVYYRDSRDP
ARV 2      N                                                                              KK
LAV MAL               E                                    I M
LAV ELI                                          RR          I                 I

                970        980        990       1000       1010
LAV BRU  LWKGPAKLLW KGEGAVVIQD NSDIKVVPRR KAKIIRDYGK QMAGDDCVAS RQDED
ARV 2
LAV MAL  I                                                   G G
LAV ELI  I                   K          V
```

FIG. 3D-2

```
ENV
                                     SP                            OMP
                  10         20         30    ↓    40         50         60         70         80
LAV BRU   MRVK---EKY QHLWRWGWKW GTMLLGILMI CSATEKLWVT VYYGVPVWKE ATTTLFCASD AKAYDTEVHN VWATHACVPT
ARV 2      K  GTRRN         --- - L   M                                      R
LAV MAL      REIQRN   NW    --- -M    M  T   IA D                             S E      I
LAV ELI     ARGIERNC  NW K  --- -I       T    ADN                             S E  A   I

                  90        100        110        120        130        140        150        160
LAV BRU   DPNPQEVVLV NVTENFNMWK NDMVEQMHED IISLWDQSLK PCVKLTPLCV SLKCTDL-CN ATNTNSSNTN SSSGEMMME-
ARV 2              C             N     Q                         T N    - K        --- ----NWKE I
LAV MAL          IE E    G       N                               T N  NVN T  V GTNACS  RTNA LK  I
LAV ELI          IA E            N                               T N S E--L RN GTMG NV TTEEKG----

                 170        180        190        200        210        220        230        240
LAV BRU   KGEIKNCSFN ISTSIRGKVQ KEYAFFYKLD IIPIDNDTTS -----YTLTS CNTSVITQAC PKVSFEPIPI HYCAPAGFAI
ARV 2                 T    D I    N L RN   VV    AS T TNYTN R IN   R                      T
LAV MAL   -  V        TPVGSD R  -   T  N   LVQ  DSDN  ----S R IN               T D
LAV ELI   ---M       VT VLKD K  QV  L  R    V     SST -NSTSN R IN    A

                 250        260        270        280        290        300        310        320
LAV BRU   LKCNKKTFNG TGPCTNVSTV QCTHGIRPVV STQLLLNGSL AEEEVVIRSA NFTDNAKTII VQLNQSVEIN CTRPNNNTRK
ARV 2                K                  I                      D    N           E A
LAV MAL       D K     EI K            K                   IM   E  L   T N       ET T       G    R
LAV ELI      RD K                                          I   E  L N   N   AH  E  K T  A  YQ   Q
```

FIG. 3E-1

```
                330        340        350        360        370        380        390        400
LAV BRU  SIRIQRGPGR AFVTIGK-IG NMRQAHCNIS RAKWNATLKQ IASKLREQFG NNKT-IIFKQ SSGGDPEIVT HSFNCGGEFF
ARV 2      Y --       W T RI   DI K         Q  N  E   VK             - V N          M      R
LAV MAL  G HF--   Q  LY T I-V  DI R Y T N ETE DK  Q  V V  GSLL- -   K   NS         T      R
LAV ELI  RTP -- L Q SLY TKS-RS IIG          Q SK  Q  V R  GTLL- -   I K  P         T

                410        420        430        440        450        460        470        480
LAV BRU  YCNSTQLFNS TWFNSTWSTE CSNNTEGSDT ITLPCRIKQF INMWQEVGKA MYAPPISGQI RCSSNITGLL LTRDGGNN--
ARV 2       T     N   -----RLN RTEG K N    I       I                    C    S                T -V
LAV MAL     TSK       Q NGARL- -  S  STGS          I      KT           A V  N L      I        NSSD
LAV ELI     TSG       NI A NNI  TES NSTNTN   Q     I  K VARGR-  I     ERN  L                I --

                490        500        510        520        530        540        550        560
LAV BRU  NNGSEIFRPG GGDMRDNWRS ELYKYKVVKI EPLGVAPTKA KRRVVQREKR AVGI-GALFL GFLGAAGSTM GARSMTLTVQ
ARV 2    T DT V                      I       I                        V  M                V L
LAV MAL   SDN TL           I          R                    E      I L-  M                A L
LAV ELI   STN T                       Q          R         E      I L-  M                  V
```

FIG. 3E-2

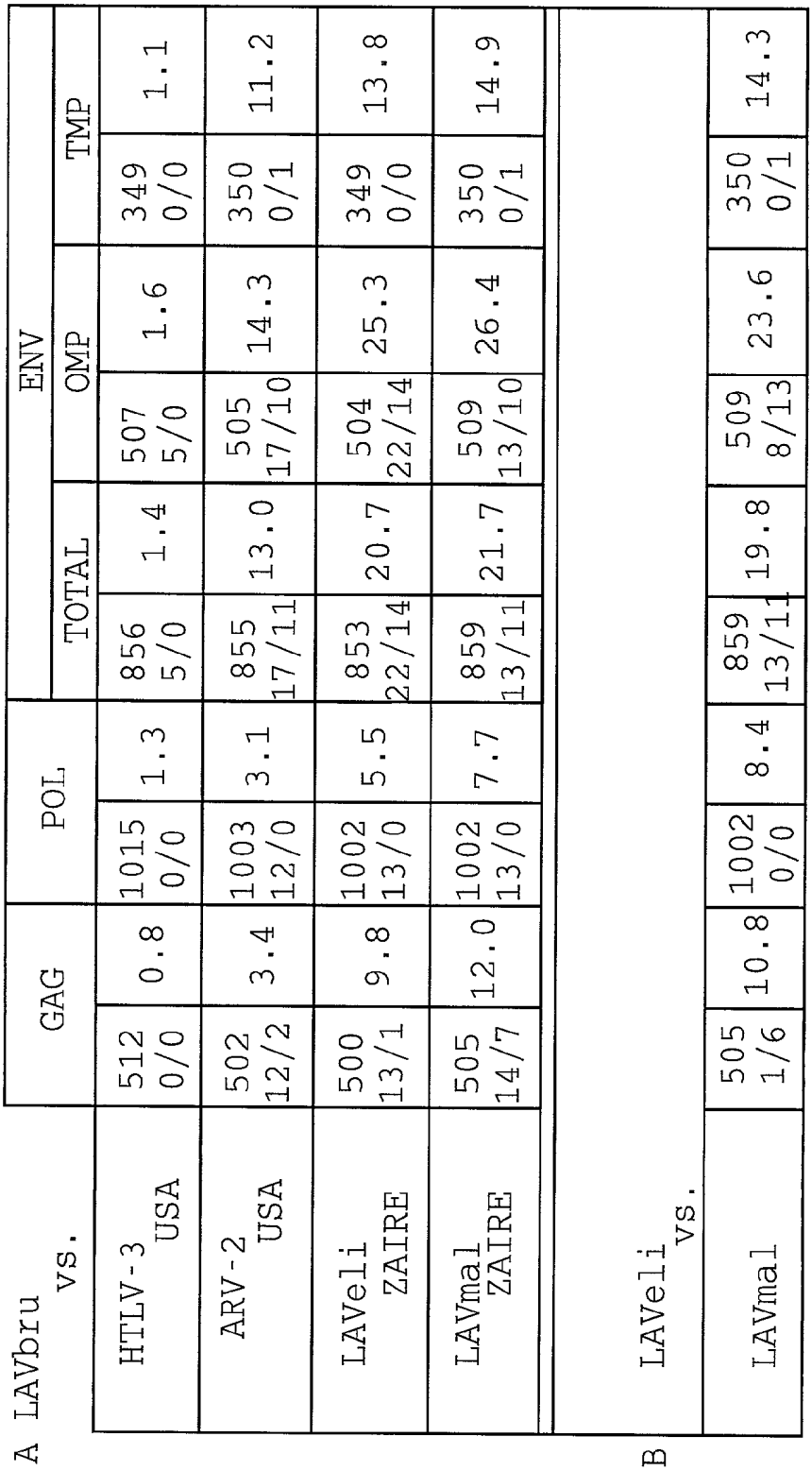
```
                570        580        590        600        610        620        630        640
LAV BRU  ARQLLSGIVQ QQNNLLRAIE AQQHLLQLTV WGIKQLQARI LAVERYLKDQ QLLGIWGCSG KLICTTAVPW NASWSNKSLE
ARV 2                                               W          R
LAV MAL                                             W          Q   R   M        H    F     S    R  D
LAV ELI      M                                                                     H    N     S    R  N

                650        660        670        680        690        700        710        720
LAV BRU  QIWNNMTWME WDREINNYTS LINSLIEESQ NQQEKNEQEL LELDKWASLW NWFNITNWLW YIKIFIMIVG GLVGLRIVFA
ARV 2    D  D     Q  E   D   N T YT L                                 S
LAV MAL  D        Q  EK  S   G I YN       I      K                   S SK      R   IV     I    I
LAV ELI  E  Q        E   D   G   Y        T      K                   S  Q           I    I

                730        740        750        760        770        780        790        800
LAV BRU  VLSIVNRVRQ GYSPLSFQTH LPTPRGP-DR PEGIEEEGGE RDRDRSIRLV NGSLALIWDD LRSLCLFSYH RLRDLLLIVT
ARV 2                      R   V    -    D              V    D F     E           R          AA
LAV MAL     L            L  L        P              QG G        FS         N                 A
LAV ELI     L               L   A    -      T        G    V  L   FS                     I  AV

                810        820        830        840        850        860        870
LAV BRU  RIVELLGRRG WEALKYWWNL LQYWSQELKN SAVSLLNATA IAVAEGTDRV IEVVQGACRA IRHIPRRIRQ GLERILL
ARV 2     T  I  K           S      I          W         T          A R Y    L  H          L
LAV MAL                   L        G        I    T        C       IG RFG    L          F  A
LAV ELI              DI   L             R    S  FD I              II R     VLN            S
```

FIG. 3F-1

F

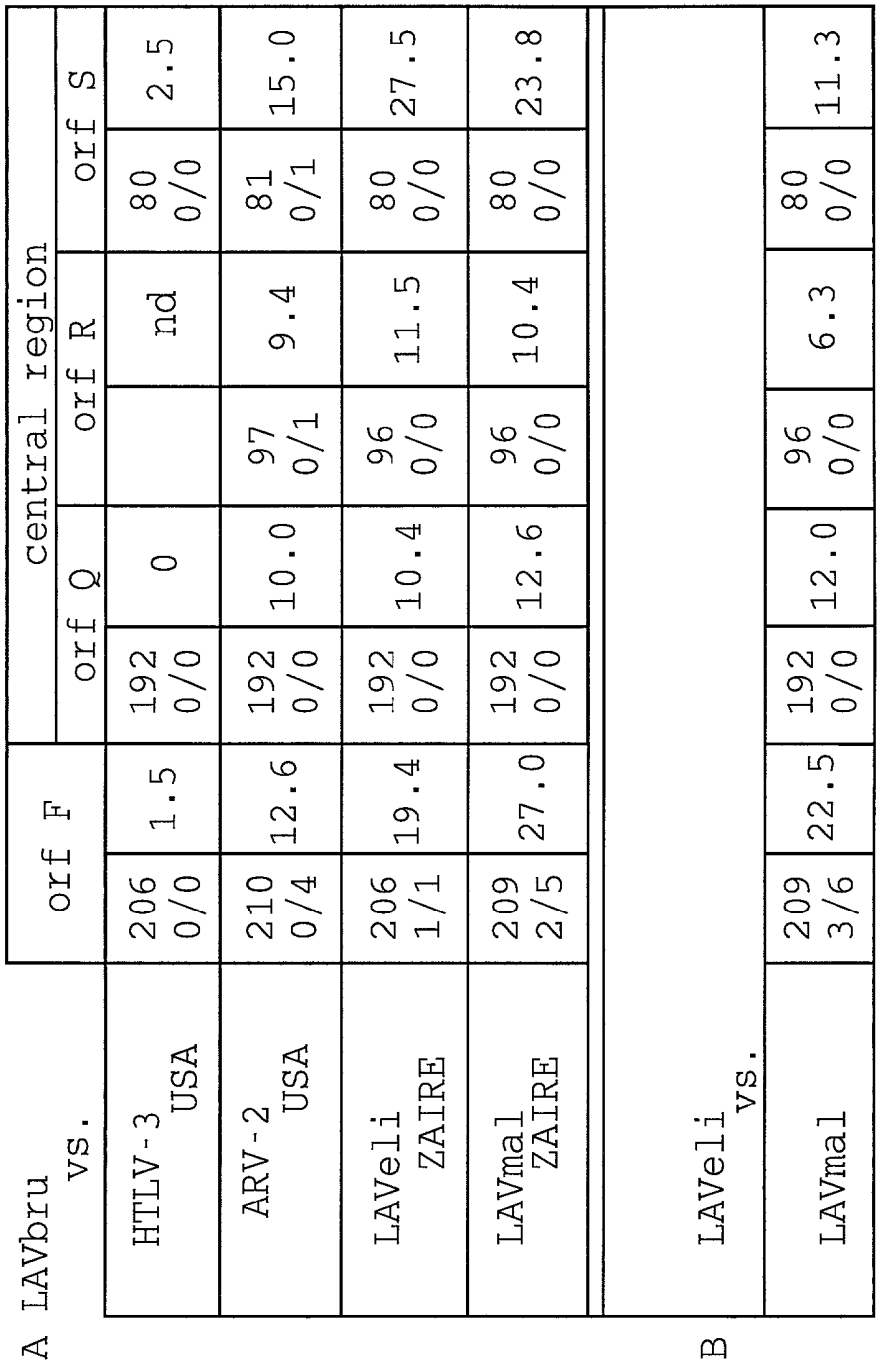
```
                10         20         30         40         50         60         70         80
LAV BRU  MGGKWSKSSV VGWPTVRERM R----RAEPA ADGVGAASR- ----DLEKUG AITSSNTAAT NAACAWLEAQ EE-EEVGFPV
ARV 2          R M G  SAI      RAEP           V  - ----                    D          -
LAV MAL          I     KI   I  ---- TP T ET    V QD AVSQ  D C   AA  SP  N   S --- PP   E
LAV ELI          I     AI   I  ---- TM         V  - ----                S   D          SD

                90        100        110        120        130        140        150        160
LAV BRU  TPQVPLRRHT YKAAVDLSHF LKEKGGLEGL IHSQRRQDIL DLWIYUTQGY FPDWQNYTPC PGVRYPLTFG WCYKLVPVEP
ARV 2    R             L  I               W     E                          I          F
LAV MAL  R           G F             D   VW PK  E      V                   I F        F    HS
LAV ELI  R           E L                 W KK  E      V N   I               I          E    D

               170        180        190        200        210
LAV BRU  DKVEEANKGE NTSLLHPVSL HGMDDPEREV LEWRFDSRLA FHHVARELHP EYFKNC
ARV 2    E      E    N     M      E A K    V     K      M         Y D
LAV MAL  EE     E    NC    I Q    E A      K K   S   LR R  Q      Y D
LAV ELI  QE   DTE   TN     ICQ    E    Q   K   N      E K  M     FY  -
```

FIG. 3F-2



A		GAG		POL		ENV					
						TOTAL		OMP		TMP	
LAVbru	vs.	512 0/0	0.8	1015 0/0	1.3	856 5/0	1.4	507 5/0	1.6	349 0/0	1.1
HTLV-3	USA	502 12/2	3.4	1003 12/0	3.1	855 17/11	13.0	505 17/10	14.3	350 0/1	11.2
ARV-2	USA	500 13/1	9.8	1002 13/0	5.5	853 22/14	20.7	504 22/14	25.3	349 0/0	13.8
LAVeli	ZAIRE	505 14/7	12.0	1002 13/0	7.7	859 13/11	21.7	509 13/10	26.4	350 0/1	14.9
LAVmal	ZAIRE										
B											
LAVeli	vs.	505 1/6	10.8	1002 0/0	8.4	859 13/11	19.8	509 8/13	23.6	350 0/1	14.3
LAVmal											

FIG. 4A



A LAVbru vs.	orf F		central region				orf S	
			orf Q		orf R			
HTLV-3 USA	206 0/0	1.5	192 0/0	0		nd	80 0/0	2.5
ARV-2 USA	210 0/4	12.6	192 0/0	10.0	97 0/1	9.4	81 0/1	15.0
LAVeli ZAIRE	206 1/1	19.4	192 0/0	10.4	96 0/0	11.5	80 0/0	27.5
LAVmal ZAIRE	209 2/5	27.0	192 0/0	12.6	96 0/0	10.4	80 0/0	23.8

B LAVeli vs.								
LAVmal	209 3/6	22.5	192 0/0	12.0	96 0/0	6.3	80 0/0	11.3

FIG. 4B

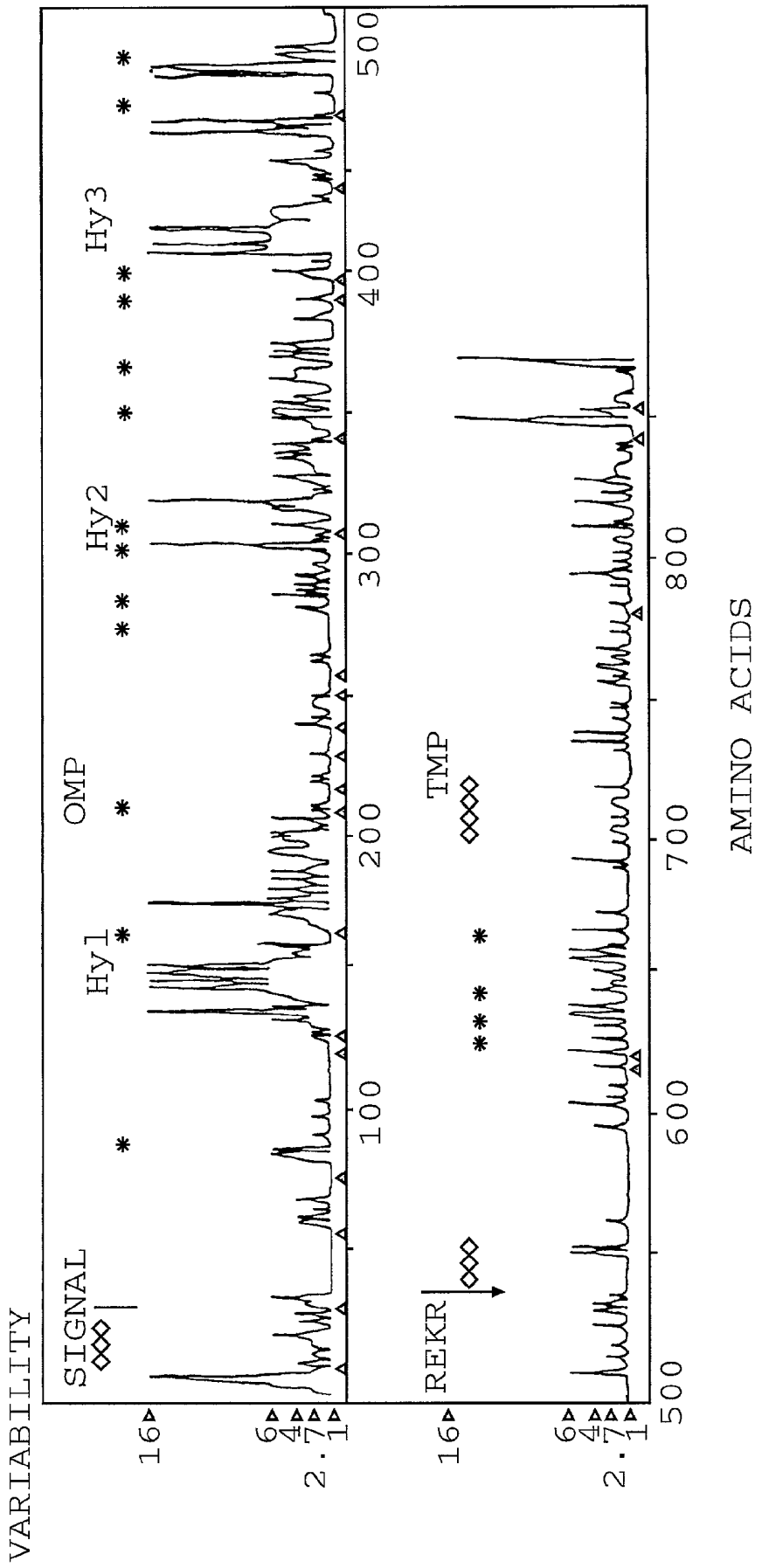

VARIABILITY
SIGNAL
Hy1
OMP
Hy2
Hy3
TMP
REKR
16
6
4
7
2.1
100
200
300
400
500
600
700
800
500
AMINO ACIDS

FIG. 5

```
GAG

a
                                  120

             K    A    Q    Q    A    A    A                                      D    T
LAV.BRU     AAA  GCA  CAG  CAA  GCA  GCA  GCT   -    -    -    -    -             GAC  ACA

             K    A    Q    Q    A    A    A    A    A    -    -    -    -       G    T
ARV 2       AAG  GCA  CAG  CAA  GCA  GCA  GCT  GCA  GCT                          GGC  ACA

             K    T    Q    Q    A    A    A    A    Q    Q    A    A    A    A    T
LAV.MAL     AAG  ACA  CAG  CAG  GCA  GCA  GCT  GCA  GAG  GAG  GCA  GCA  GCT  GCC  ACA

             X    A    Q    Q    A    A    A                                      D    T
LAV.ELI     AAG  GCA  CAG  CAA  GCA  GCA  GCT   -    -    -    -    -    -        GAC  ACA
```

FIG. 6A-1

b

LAV.BRU

460 470 480

	G	N	F	L	Q	S	R	P	E	P	T	A	P	P	F	L	Q	S	R	P	E	P	T	A	P	P	E	E
GGG	AAT	TTT	CTT	CAP	AGC	AGA	CCA	GAG	CCA	ACA	CCC	CCA	CCA	TTT	CTT	CAG	AGC	AGA	CCA	GAG	CCA	ACA	GCC	CCA	CCA	GAA	GAG

ARV 2

G	N	F	L	Q	S	R	P	E	P	T	A	P	P															E	E
GGG	AAT	TTT	CTT	CAG	AGC	AGA	CCA	GAG	CCA	ACA	GCC	CCA	CCA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GAA	GAG

LAV.MAL

G	N	F	L	Q	S	R	P	E	P	T	A	P	P															A	E
GGG	AAT	TTC	CTT	CAG	AGC	AGA	CCA	GAG	CCA	ACA	GCC	CCA	CCA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GCA	GAG

LAV.ELI

G	N	F	L	Q	S	R	P	E	P	T	A	P	P															A	E
GGG	AAC	TTT	CTC	CAA	AGC	AGA	CCA	GAG	CCA	ACA	GCC	CCA	CCA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	GCA	GAG

FIG. 6A-2

```
c
                         20                                        30
                R    M    R                       R    A    E    P    A
LAV.BRU        AGA  ATG  AGA   -    -    -    -  CGA  GCT  GAG  CCA  GCA

                R    M    R    R    A    E    P    R    A    E    P    A
ARV 2          AGA  ATG  AGA  CGA  GCT  GAG  CCA  CGA  GCT  GAG  CCA  GCA

                R    I    R                       R    T    P    P    T
LAV.MAL        AGA  ATA  AGA   -    -    -    -  CGA  ACT  CCC  CCA  ACA
                               -    -    -    -
                R    I    R                       R    T    P    P    T
LAV.ELI        AGA  ATA  AGA                     AGA  ACT  AAT  CCA  GCA

d
                                         40
                V    G    A    A    S    R                             D
LAV.BRU        GTG  GGA  GCA  GCA  TCT  CGA   -    -    -    -    -  GAC

                V    G    A    V    A    R                             D
ARV 2          GTG  GGA  GCA  GTA  TCT  CGA   -    -    -    -    -  GAC

                V    G    A    V    S    R    D    A    V    S    Q    D
LAV.MAL        GTA  GGA  GCA  GTA  TCT  CAA  GAT  GCA  GTA  TCT  CAA  GAT

                V    G    A    V    S    R                             D
LAV.ELI        GTA  GGA  GCA  GTA  TCT  CGA   -    -    -    -    -  GAC
```

FIG. 6A-3

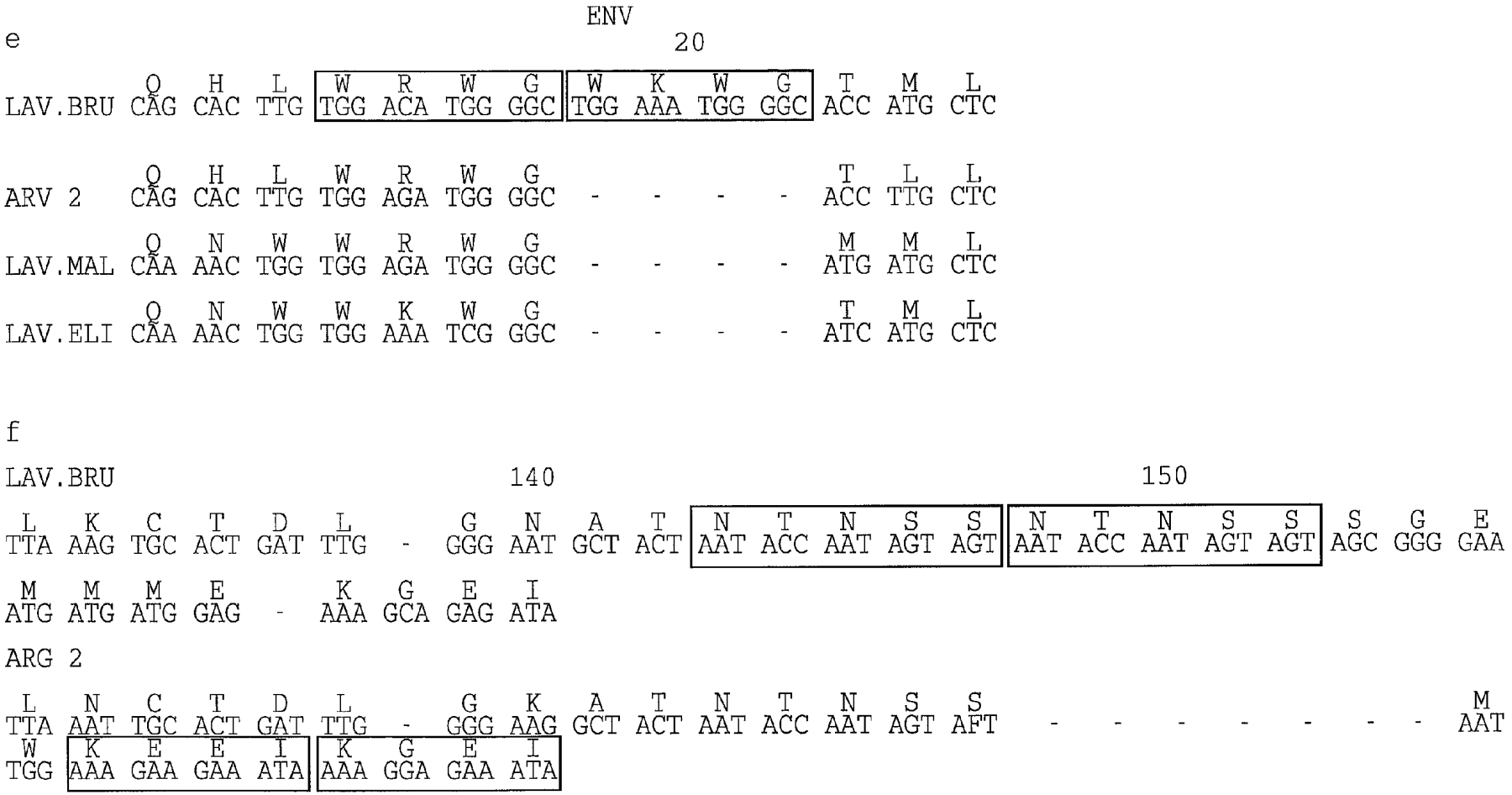

ENV
e
20
LAV.BRU Q H L W R W G W K W G T M L
CAG CAC TTG TGG ACA TGG GGC TGG AAA TGG GGC ACC ATG CTC
ARV 2 Q H L W R W G T L L
CAG CAC TTG TGG AGA TGG GGC - - - - ACC TTG CTC
LAV.MAL Q N W W R W G M M L
CAA AAC TGG TGG AGA TGG GGC - - - - ATG ATG CTC
LAV.ELI Q N W W K W G T M L
CAA AAC TGG TGG AAA TCG GGC - - - - ATC ATG CTC
f
LAV.BRU 140 150
L K C T D L G N A T N T N S S N T N S S S G E
TTA AAG TGC ACT GAT TTG - GGG AAT GCT ACT AAT ACC AAT AGT AGT AAT ACC AAT AGT AGT AGC GGG GAA
M M M E K G E I
ATG ATG ATG GAG - AAA GCA GAG ATA
ARG 2
L N C T D L G K A T N T N S S M
TTA AAT TGC ACT GAT TTG - GGG AAG GCT ACT AAT ACC AAT AGT AFT - - - - - - - AAT
W K E E I K G E I
TGG AAA GAA GAA ATA AAA GGA GAA ATA

FIG. 6B-1

LAV.MAL

```
 L   N   C   T   N   V   N   G   T   A   V   N   G   T   N   A   G   S   N   R   T   N   A   E
TTA AAC TGC ACT AAT GTG AAT GGG ACT GCT GTG AAT GGG ACT AAT GCT GGG AGT AAT AGG ACT AAT GCA GAA

 L   K   M   E   I       G   E   V
TTG AAA ATG GAA ATT -   GGA GAA GTG
```

LAV.ELI

```
 L   N   C   S   D   E           L   R   N   N   G   T   M   G   N   N   V   T   T   E   E   K
TTA AAC TGT AGT GAT GAA   -   - TTG AGG AAC AAT GGC ACT ATG GGG AAC AAT GTC ACT ACA GAG GAG AAA

 G                               M
GGA   -   -   -   -   -   -   - ATG
```

FIG. 6B-2

g

```
                                  200
          D   N   D   T   T   S                           Y   T   L
LAV.BRU   GAT AAT GAT ACT ACC AGC -   -   -   -   -   -   TAT ACG TTG

          D   N   A   S   T   T   T   N   Y   T   N   Y   R   L
ARV 2     GAT AAT GCT AGT ACT ACT ACT AAC AAC TAT AAC TAT AGG TTG

          D   D   S   D   N   S   S                       Y   R   L
LAV.MAL   GAT GAT AGT GAT AAT AGT AGT -   -   -   -   -   TAT AGG TTA

          D   N   D   S   S   T       N   S   T   N       Y   R   L
LAV.EIL   GAC AAT GAT AGT AGT ACC -   AAT AGT ACC AAT TAT AGG TTA
```

h

```
LAV.BRU                   410                                 420                 430

 C   N   S   T   Q   L   F   N   S   T   W   F   N   S   T   W   S   T   E   G   S   N   N   T   E   G
TGT AAT TCA ACA CAA CTG TTT AAT AGT ACT TGG TTT AAT AGT ACT TGG AGT ACT GAA GGG TCA AAT AAC ACT GAA GGA

 S   D   T   I
AGT GAC ACA ATC

ARV 2

 C   N   T   T   Q   L   F   N   N   T   W                   R   L   N   H   T   K   G   T   K   G
TGT AAT ACA ACA CAA CTG TTT AAT AAT ACA TGG                 AGG TTA AAT CAC ACT GAA GGA ACT AAA GGA

 N   D   T   I
AAT GAC ACA ATC
```

FIG. 6B-3

LAV.MAL

```
 C   N   T   S   K   L   F   N   S   T   W   Q   N   N   G   A   R   L           S   N   S   T   E   S
TGT AAT ACA TCA AAA CTG TTT AAT AGT ACA TGG CAG AAT AAAT GGT GCA AGA CTA  -   -  AGT AAT AGC ACA GAG TCA

 T   G   S   I
ACT GGT AGT ATC
```

LAV.ELI

```
 C   N   T   S   G   L   F   N   S   T   W   N   I   S   A   W   N   N   I   T   E   S   N   N   S   T
TGT AAT ACA TCA GGA CTG TTT AAT AGT ACA TGG AAT ATT AGT GCA TGG AAT AAT ATT ACA GAG TCA AAT AAT AGC ACA

 N   T   N   I
AAC ACA AAC ATC
```

FIG. 6B-4

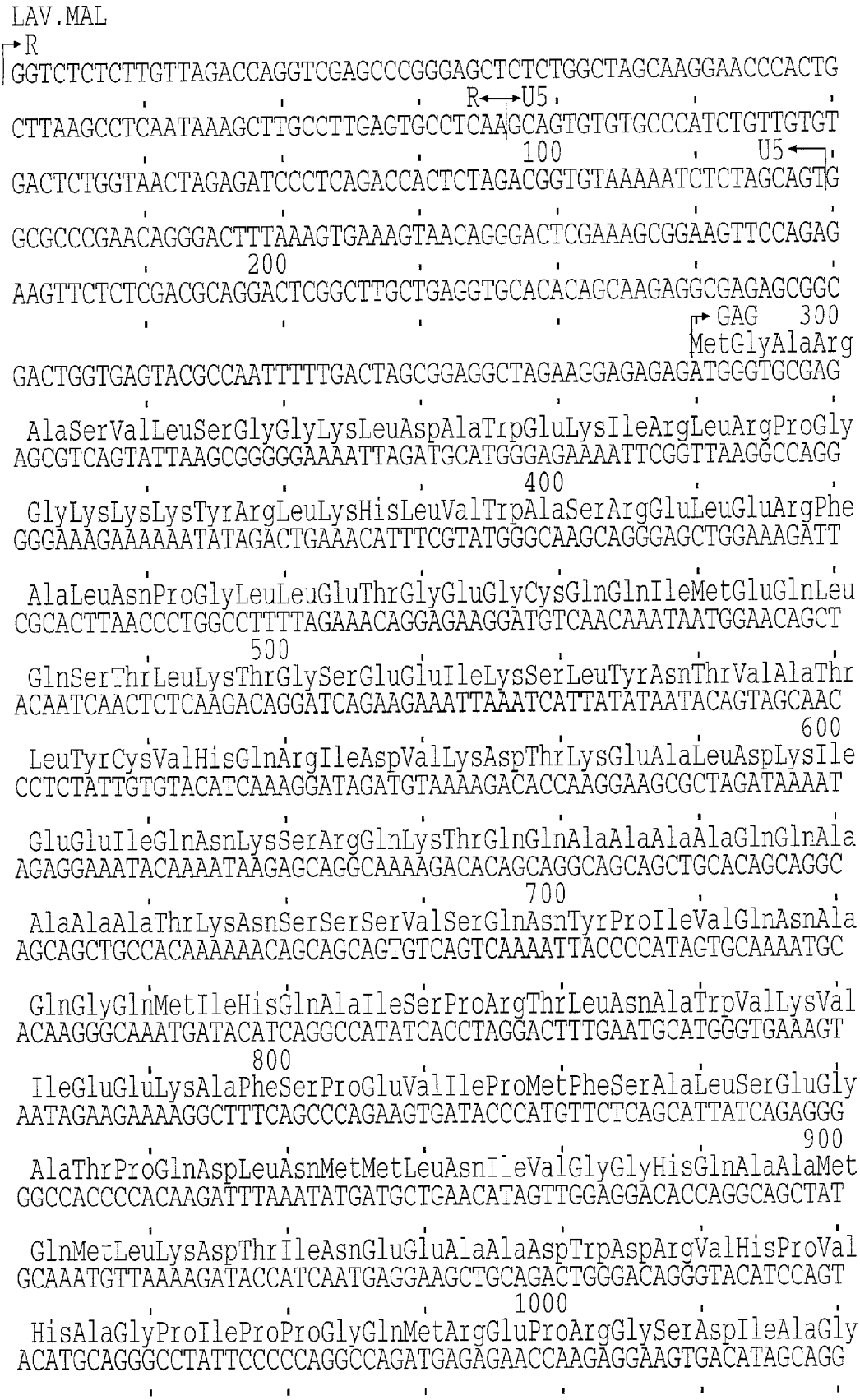

LAV.MAL
R
GGTCTCTCTTGTTAGACCAGGTCGAGCCCGGGAGCTCTCTGGCTAGCAAGGAACCCACTG
R
U5
CTTAAGCCTCAATAAAGCTTGCCTTGAGTGCCTCAAGCAGTGTGTGCCCATCTGTTGTGT
100
U5
GACTCTGGTAACTAGAGATCCCTCAGACCACTCTAGACGGTGTAAAAATCTCTAGCAGTG
GCGCCCGAACAGGGACTTTAAAGTGAAAGTAACAGGGACTCGAAAGCGGAAGTTCCAGAG
200
AAGTTCTCTCGACGCAGGACTCGGCTTGCTGAGGTGCACACAGCAAGAGGCGAGAGCGGC
GAG
300
MetGlyAlaArg
GACTGGTGAGTACGCCAATTTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAG
AlaSerValLeuSerGlyGlyLysLeuAspAlaTrpGluLysIleArgLeuArgProGly
AGCGTCAGTATTAAGCGGGGGAAAATTAGATGCATGGGAGAAAATTCGGTTAAGGCCAGG
400
GlyLysLysLysTyrArgLeuLysHisLeuValTrpAlaSerArgGluLeuGluArgPhe
GGGAAAGAAAAAATATAGACTGAAACATTTCGTATGGGCAAGCAGGGAGCTGGAAAGATT
AlaLeuAsnProGlyLeuLeuGluThrGlyGluGlyCysGlnGlnIleMetGluGlnLeu
CGCACTTAACCCTGGCCTTTTAGAAACAGGAGAAGGATGTCAACAAATAATGGAACAGCT
500
GlnSerThrLeuLysThrGlySerGluGluIleLysSerLeuTyrAsnThrValAlaThr
ACAATCAACTCTCAAGACAGGATCAGAAGAAATTAAATCATTATATAATACAGTAGCAAC
600
LeuTyrCysValHisGlnArgIleAspValLysAspThrLysGluAlaLeuAspLysIle
CCTCTATTGTGTACATCAAAGGATAGATGTAAAAGACACCAAGGAAGCGCTAGATAAAAT
GluGluIleGlnAsnLysSerArgGlnLysThrGlnGlnAlaAlaAlaAlaGlnGlnAla
AGAGGAAATACAAAATAAGAGCAGGCAAAAGACACAGCAGGCAGCAGCTGCACAGCAGGC
700
AlaAlaAlaThrLysAsnSerSerSerValSerGlnAsnTyrProIleValGlnAsnAla
AGCAGCTGCCACAAAAAACAGCAGCAGTGTCAGTCAAAATTACCCCATAGTGCAAAATGC
GlnGlyGlnMetIleHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysVal
ACAAGGGCAAATGATACATCAGGCCATATCACCTAGGACTTTGAATGCATGGGTGAAAGT
800
IleGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGly
AATAGAAGAAAAGGCTTTCAGCCCAGAAGTGATACCCATGTTCTCAGCATTATCAGAGGG
900
AlaThrProGlnAspLeuAsnMetMetLeuAsnIleValGlyGlyHisGlnAlaAlaMet
GGCCACCCCACAAGATTTAAATATGATGCTGAACATAGTTGGAGGACACCAGGCAGCTAT
GlnMetLeuLysAspThrIleAsnGluGluAlaAlaAspTrpAspArgValHisProVal
GCAAATGTTAAAAGATACCATCAATGAGGAAGCTGCAGACTGGGACAGGGTACATCCAGT
1000
HisAlaGlyProIleProProGlyGlnMetArgGluProArgGlySerAspIleAlaGly
ACATGCAGGGCCTATTCCCCCAGGCCAGATGAGAGAACCAAGAGGAAGTGACATAGCAGG

FIG. 7A

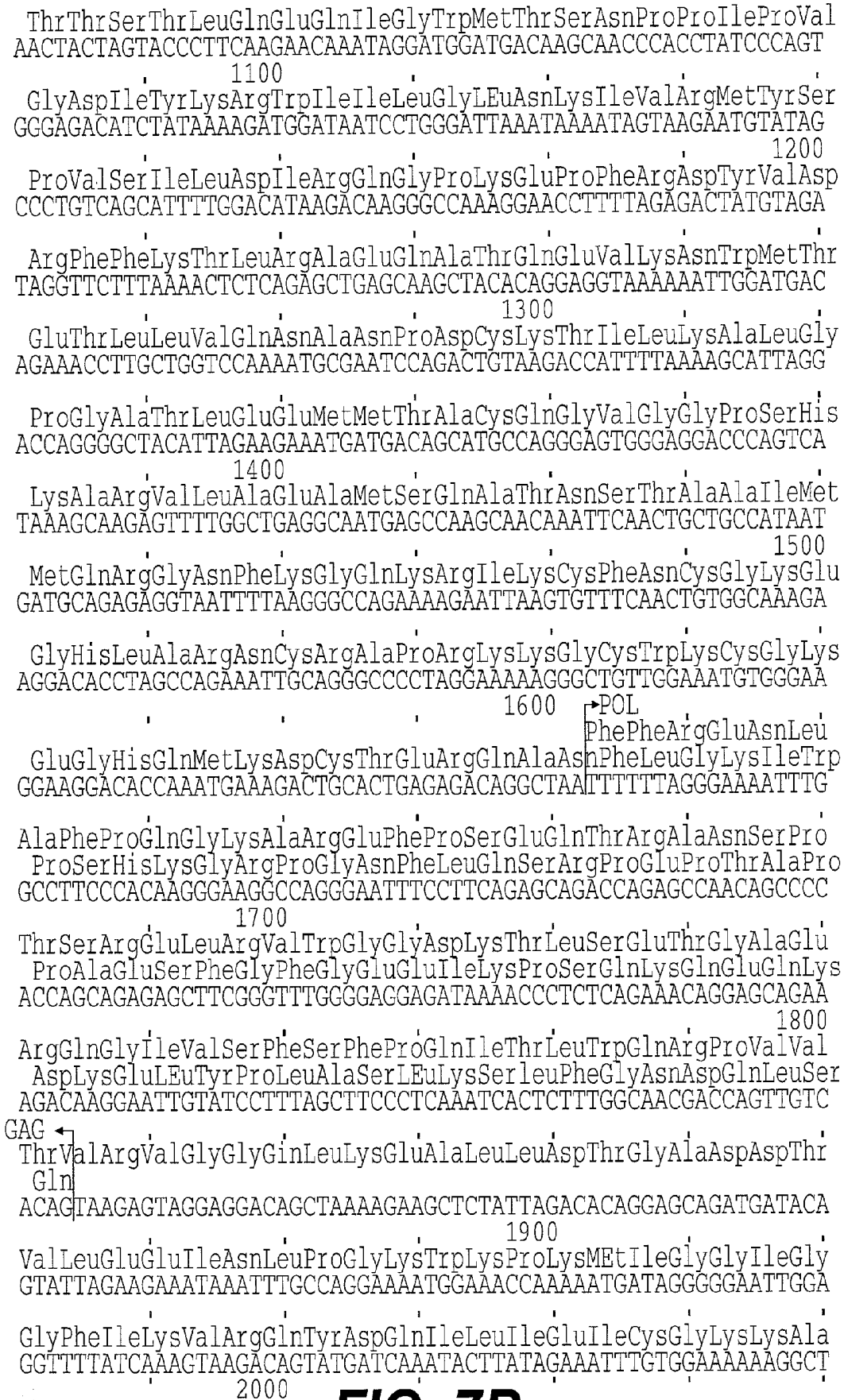

ThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrSerAsnProProIleProVal
AACTACTAGTACCCTTCAAGAACAAATAGGATGGATGACAAGCAACCCACCTATCCCAGT
1100
GlyAspIleTyrLysArgTrpIleIleLeuGlyLEuAsnLysIleValArgMetTyrSer
GGGAGACATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAG
1200
ProValSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAsp
CCCTGTCAGCATTTTGGACATAAGACAAGGGCCAAAGGAACCTTTTAGAGACTATGTAGA
ArgPhePheLysThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThr
TAGGTTCTTTAAAACTCTCAGAGCTGAGCAAGCTACACAGGAGGTAAAAAATTGGATGAC
1300
GluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGly
AGAAACCTTGCTGGTCCAAAATGCGAATCCAGACTGTAAGACCATTTTAAAAGCATTAGG
ProGlyAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHis
ACCAGGGGCTACATTAGAAGAAATGATGACAGCATGCCAGGGAGTGGGAGGACCCAGTCA
1400
LysAlaArgValLeuAlaGluAlaMetSerGlnAlaThrAsnSerThrAlaAlaIleMet
TAAAGCAAGAGTTTTGGCTGAGGCAATGAGCCAAGCAACAAATTCAACTGCTGCCATAAT
1500
MetGlnArgGlyAsnPheLysGlyGlnLysArgIleLysCysPheAsnCysGlyLysGlu
GATGCAGAGAGGTAATTTTAAGGGCCAGAAAAGAATTAAGTGTTTCAACTGTGGCAAAGA
GlyHisLeuAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLys
AGGACACCTAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGGAA
1600
POL
PhePheArgGluAsnLeu
GluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrp
GGAAGGACACCAAATGAAAGACTGCACTGAGAGACAGGCTAATTTTTTAGGGAAAATTTG
AlaPheProGlnGlyLysAlaArgGluPheProSerGluGlnThrArgAlaAsnSerPro
ProSerHisLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPro
GCCTTCCCACAAGGGAAGGCCAGGGAATTTCCTTCAGAGCAGACCAGAGCCAACAGCCCC
1700
ThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeuSerGluThrGlyAlaGlu
ProAlaGluSerPheGlyPheGlyGluGluIleLysProSerGlnLysGlnGluGlnLys
ACCAGCAGAGAGCTTCGGGTTTGGGGAGGAGATAAAACCCTCTCAGAAACAGGAGCAGAA
1800
ArgGlnGlyIleValSerPheSerPheProGlnIleThrLeuTrpGlnArgProValVal
AspLysGluLEuTyrProLeuAlaSerLEuLysSerleuPheGlyAsnAspGlnLeuSer
AGACAAGGAATTGTATCCTTTAGCTTCCCTCAAATCACTCTTTGGCAACGACCAGTTGTC
GAG
ThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThr
Gln
ACAGTAAGAGTAGGAGGACAGCTAAAAGAAGCTCTATTAGACACAGGAGCAGATGATACA
1900
ValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMEtIleGlyGlyIleGly
GTATTAGAAGAAATAAATTTGCCAGGAAAATGGAAACCAAAAATGATAGGGGGAATTGGA
GlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLysLysAla
GGTTTTATCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAAGGCT
2000

FIG. 7B

```
IleGlyThrIleLeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThr
ATAGGTACAATATTGGTAGGACCTACACCTGTCAACATAATTGGACGAAATATGTTGACT
                                                          2100
GlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeu
CAGATTGGTTGTACTTTAAATTTTCCAATTAGTCCTATTGAGACTGTACCAGTAAAATTA

LysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrGluGluLysIleLys
AAGCCAGGGATGGATGGCCCAAGGGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAA
                                  2200
AlaLeuThrGluIleCysLysAspMetGluLysGluGlyLysIleLeuLysIleGlyPro
GCATTAACAGAAATTTGTAAAGATATGGAAAAGGAAGGAAAAATTTTAAAAATTGGGCCT

GluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArg
GAAAATCCATACAATACTCCAGTATTTGCCATAAAGAAAAAAGACAGCACTAAATGGAGA
          2300
LysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeu
AAATTAGTGAATTTCAGAGAGCTTAATAAAAGAACTCAAGATTTTTGGGAAGTTCAATTA
                                                          2400
GlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGly
GGAATACCACATCCTGCTGGGTTGAAAAAGAAAAAATCAGTCACAGTATTGGATGTGGGG

AspAlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIle
GATGCATATTTTTCAGTCCCTTTAGATGAAGATTTCAGGAAGTATACTGCATTCACTATA
                                  2500
ProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValIleProGlnGly
CCCAGTATTAATAATGAGACACCAGGGATTAGATATCAGTACAATGTGCTACCACACGGA

TrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArg
TGGAAAGGATCACCAGCAATATTCCAGAGTAGCATGACAAAAATCTTAGAACCCTTTAGA
          2600
ThrLysAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAsp
ACAAAAAATCCAGAAATAGTCATATACCAATACATGGATGATTTGTATGTAGGGTCTGAT
                                                          2700
LeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGluHisLeuLeuLysTrp
TTAGAAATAGGACAACATAGAACAAAAATAGAGGAACTAAGAGAACATCTATTGAAATGG

GlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyr
GGATTTACCACACCAGACAAAAAGCATCAGAAAGAACCCCCATTTCTTTGGATGGGGTAT
                                  2800
GluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProAspLysGluSerTrp
GAACTCCACCCTGACAAATGGACAGTGCAGCCTATACAACTGCCAGACAAGGAAAGCTGG

ThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrPro
ACTGTCAATGATATACAGAAATTGGTGGGAAAACTAAATTGGGCAAGTCAGATTTATCCA
          2900
GlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIle
GGAATTAAAGTAAAGCAATTATGTAAACTCCTTAGGGGAGCAAAAGCACTAACAGACATA
                                                          3000
ValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGlu
GTACCATTAACTGCAGAGGCAGAATTAGAATTGGCAGAGAACAGGGAAATTCTAAAAGAA
```

FIG. 7C

```
ProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGln
CCAGTGCATGGGGTATATTATGACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAG
                                    3100
GlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyrLysAsnLeuLysThrGly
GGGCAAGGTCAATGGACATATCAAATATACCAAGAGCAATATAAAAATCTGAAAACAGGG

LysTyrAlaArgIleLysSerAlaHisThrAsnAspValLysGlnLeuThrGluAlaVal
AAGTATGCAAGAATAAAGTCTGCCCACACTAATGATGTAAAACAATTAACAGAAGCAGTG
          3200
GlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThrProLysPheArgLeuPro
CAAAAGATAGCCCAAGAAAGCATAGTAATATGGGGAAAAACTCCTAAATTTAGACTACCC
                                                    3300
IleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrpGlnAlaThrTrpIlePro
ATACAAAAAGAAACATGGGAGGCATGGTGGACAGAATATTGGCAAGCCACCTGGATCCCT

GluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGlu
GAATGGGAGTTTGTCAATACTCCTCCCCTAGTAAAACTATGGTACCAGTTAGAAACAGAA
                    3400
ProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLys
CCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCTAATAGAGAAACTAAAAAG

GlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValValSerLeuThrGluThr
GGAAAAGCAGGATATGTTACTGACAGAGGAAGACAAAAGGTTGTCTCCTTAACTGAAACA
          3500
ThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlySerGlu
ACAAATCAGAAGACTGAATTACAAGCAATCCACTTAGCTTTACAGGATTCAGGATCAGAA
                                                    3600
ValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLys
GTAAACATAGTAACAGACTCACAGTATGCATTAGGGATTATTCAAGCACAACCAGATAAA

SerGluSerGluIleValAsnGlnIleIleGluGlnLeuIleGlnLysAspLysValTyr
AGTGAATCAGAGATTGTTAATCAAATAATAGAGCAATTAATACAGAAGGACAAGGTCTAC
                    3700
LeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuVal
CTGTCATGGGTACCAGCACACAAAGGGATTGGAGGAAATGAACAAGTAGATAAATTAGTC

SerSerGlyIleArgLysValLeuPheLeuAspGlyIleAspLysAlaGlnGluGluHis
AGCAGTGGAATCAGAAAGGTACTATTTTTAGATGGGATAGATAAGGCTCAAGAAGAACAT
          3800
GluLysTyrHisSerAsnTrpArgAlaMetAlaSerAspPheAsnLeuProProIleVal
GAAAAATATCACAGCAATTGGAGAGCAATGGCTAGTGACTTTAATCTACCACCTATAGTA
                                                    3900
AlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGluAlaMetHisGly
GCGAAGGAAATAGTAGCCAGCTGTGATAAATGTCAACTAAAAGGGGAAGCCATGCATGGA

GlnValAspCysSerProGlyIleTrpGlnLeuAspCysThrHisLeuGluGlyLysIle
CAAGTAGACTGTAGTCCAGGGATATGGCAATTAGATTGCACACATCTAGAAGGAAAAATA
                    4000
IleIleValAlaValHisValAlaSerGlyTyrIleGluAlaGluValIleProAlaGlu
ATCATAGTAGCAGTCCATGTAGCCAGTGGATATATAGAAGCAGAAGTTATCCCAGCAGAA

ThrGlyGlnGluThrAlaTyrPheIleLeuLysLeuAlaGlyArgTrpProValLysVal
ACAGGACAGGAGACAGCATACTTTATACTAAAATTAGCAGGAAGATGGCCAGTAAAAGTA
          4100
```

FIG. 7D

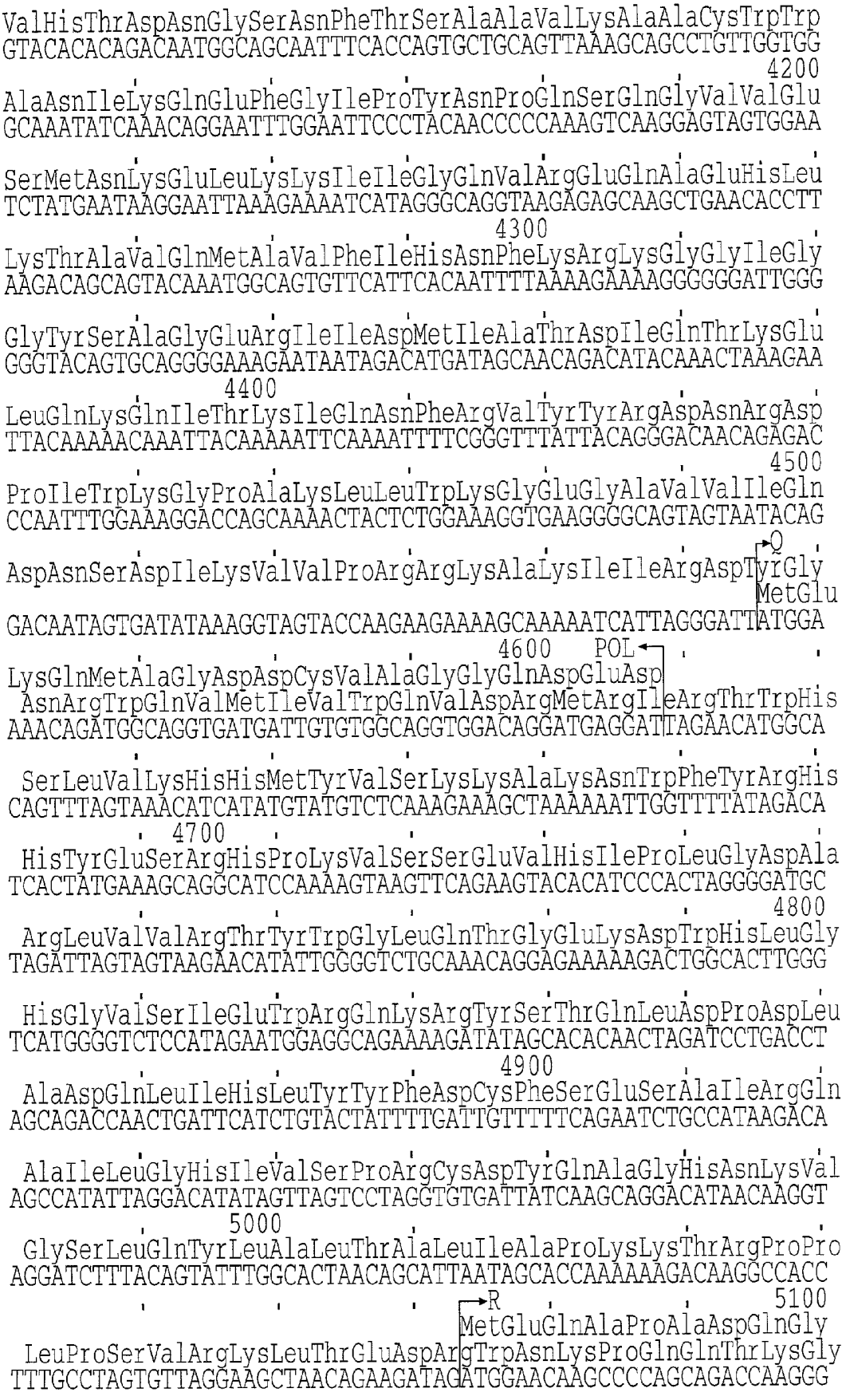

ValHisThrAspAsnGlySerAsnPheThrSerAlaAlaValLysAlaAlaCysTrpTrp
GTACACACAGACAATGGCAGCAATTTCACCAGTGCTGCAGTTAAAGCAGCCTGTTGGTGG
4200
AlaAsnIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGlnGlyValValGlu
GCAAATATCAAACAGGAATTTGGAATTCCCTACAACCCCCAAAGTCAAGGAGTAGTGGAA
SerMetAsnLysGluLeuLysLysIleIleGlyGlnValArgGluGlnAlaGluHisLeu
TCTATGAATAAGGAATTAAAGAAAATCATAGGGCAGGTAAGAGAGCAAGCTGAACACCTT
4300
LysThrAlaValGlnMetAlaValPheIleHisAsnPheLysArgLysGlyGlyIleGly
AAGACAGCAGTACAAATGGCAGTGTTCATTCACAATTTTAAAAGAAAAGGGGGGATTGGG
GlyTyrSerAlaGlyGluArgIleIleAspMetIleAlaThrAspIleGlnThrLysGlu
GGGTACAGTGCAGGGGAAAGAATAATAGACATGATAGCAACAGACATACAAACTAAAGAA
4400
LeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArgAspAsnArgAsp
TTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGGACAACAGAGAC
4500
ProIleTrpLysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAlaValValIleGln
CCAATTTGGAAAGGACCAGCAAAACTACTCTGGAAAGGTGAAGGGGCAGTAGTAATACAG
Q
AspAsnSerAspIleLysValValProArgArgLysAlaLysIleIleArgAspTyrGly
MetGlu
GACAATAGTGATATAAAGGTAGTACCAAGAAGAAAAGCAAAAATCATTAGGGATTATGGA
4600 POL
LysGlnMetAlaGlyAspAspCysValAlaGlyGlyGlnAspGluAsp
AsnArgTrpGlnValMetIleValTrpGlnValAspArgMetArgIleArgThrTrpHis
AAACAGATGGCAGGTGATGATTGTGTGGCAGGTGGACAGGATGAGGATTAGAACATGGCA
SerLeuValLysHisHisMetTyrValSerLysLysAlaLysAsnTrpPheTyrArgHis
CAGTTTAGTAAACATCATATGTATGTCTCAAAGAAAGCTAAAAAATTGGTTTTATAGACA
4700
HisTyrGluSerArgHisProLysValSerSerGluValHisIleProLeuGlyAspAla
TCACTATGAAAGCAGGCATCCAAAAGTAAGTTCAGAAGTACACATCCCACTAGGGGATGC
4800
ArgLeuValValArgThrTyrTrpGlyLeuGlnThrGlyGluLysAspTrpHisLeuGly
TAGATTAGTAGTAAGAACATATTGGGGTCTGCAAACAGGAGAAAAAGACTGGCACTTGGG
HisGlyValSerIleGluTrpArgGlnLysArgTyrSerThrGlnLeuAspProAspLeu
TCATGGGGTCTCCATAGAATGGAGGCAGAAAAGATATAGCACACAACTAGATCCTGACCT
4900
AlaAspGlnLeuIleHisLeuTyrTyrPheAspCysPheSerGluSerAlaIleArgGln
AGCAGACCAACTGATTCATCTGTACTATTTTGATTGTTTTTCAGAATCTGCCATAAGACA
AlaIleLeuGlyHisIleValSerProArgCysAspTyrGlnAlaGlyHisAsnLysVal
AGCCATATTAGGACATATAGTTAGTCCTAGGTGTGATTATCAAGCAGGACATAACAAGGT
5000
GlySerLeuGlnTyrLeuAlaLeuThrAlaLeuIleAlaProLysLysThrArgProPro
AGGATCTTTACAGTATTTGGCACTAACAGCATTAATAGCACCAAAAAAGACAAGGCCACC
R
5100
MetGluGlnAlaProAlaAspGlnGly
LeuProSerValArgLysLeuThrGluAspArgTrpAsnLysProGlnGlnThrLysGly
TTTGCCTAGTGTTAGGAAGCTAACAGAAGATAGATGGAACAAGCCCCAGCAGACCAAGGG

FIG. 7E

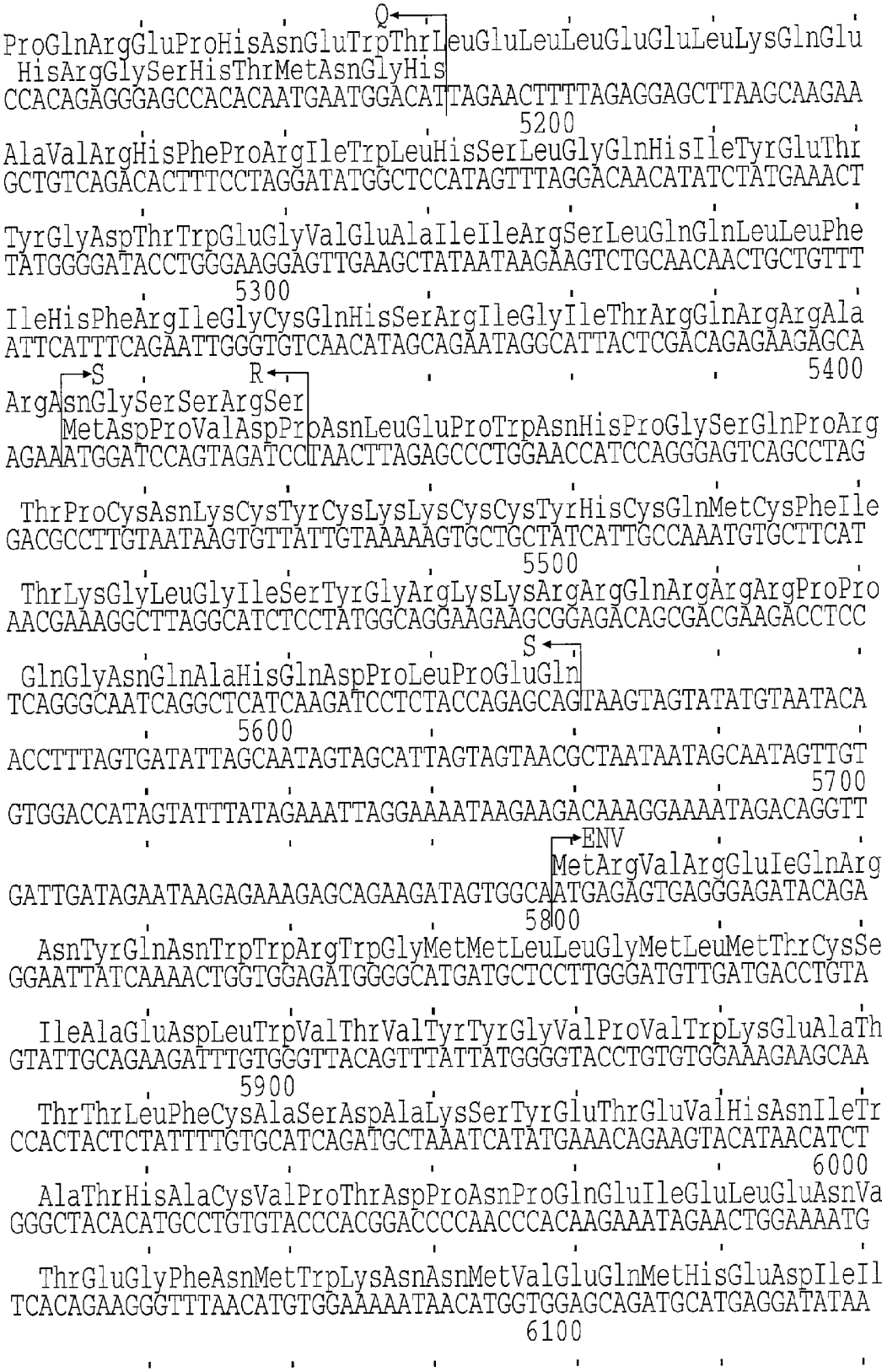

Q
ProGlnArgGluProHisAsnGluTrpThrLeuGluLeuLeuGluGluLeuLysGlnGlu
HisArgGlySerHisThrMetAsnGlyHis
CCACAGAGGGAGCCACACAATGAATGGACATTAGAACTTTTAGAGGAGCTTAAGCAAGAA
5200
AlaValArgHisPheProArgIleTrpLeuHisSerLeuGlyGlnHisIleTyrGluThr
GCTGTCAGACACTTTCCTAGGATATGGCTCCATAGTTTAGGACAACATATCTATGAAACT
TyrGlyAspThrTrpGluGlyValGluAlaIleIleArgSerLeuGlnGlnLeuLeuPhe
TATGGGGATACCTGGGAAGGAGTTGAAGCTATAATAAGAAGTCTGCAACAACTGCTGTTT
5300
IleHisPheArgIleGlyCysGlnHisSerArgIleGlyIleThrArgGlnArgArgAla
ATTCATTTCAGAATTGGGTGTCAACATAGCAGAATAGGCATTACTCGACAGAGAAGAGCA
S
R
5400
ArgAsnGlySerSerArgSer
MetAspProValAspProAsnLeuGluProTrpAsnHisProGlySerGlnProArg
AGAAATGGATCCAGTAGATCCTAACTTAGAGCCCTGGAACCATCCAGGGAGTCAGCCTAG
ThrProCysAsnLysCysTyrCysLysLysCysCysTyrHisCysGlnMetCysPheIle
GACGCCTTGTAATAAGTGTTATTGTAAAAAGTGCTGCTATCATTGCCAAATGTGCTTCAT
5500
ThrLysGlyLeuGlyIleSerTyrGlyArgLysLysArgArgGlnArgArgArgProPro
AACGAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGACCTCC
S
GlnGlyAsnGlnAlaHisGlnAspProLeuProGluGln
TCAGGGCAATCAGGCTCATCAAGATCCTCTACCAGAGCAGTAAGTAGTATATGTAATACA
5600
ACCTTTAGTGATATTAGCAATAGTAGCATTAGTAGTAACGCTAATAATAGCAATAGTTGT
5700
GTGGACCATAGTATTTATAGAAATTAGGAAAATAAGAAGACAAAGGAAAATAGACAGGTT
ENV
MetArgValArgGluIleGlnArg
GATTGATAGAATAAGAGAAAGAGCAGAAGATAGTGGCAATGAGAGTGAGGGAGATACAGA
5800
AsnTyrGlnAsnTrpTrpArgTrpGlyMetMetLeuLeuGlyMetLeuMetThrCysSer
GGAATTATCAAAACTGGTGGAGATGGGGCATGATGCTCCTTGGGATGTTGATGACCTGTA
IleAlaGluAspLeuTrpValThrValTyrTyrGlyValProValTrpLysGluAlaThr
GTATTGCAGAAGATTTGTGGGTTACAGTTTATTATGGGGTACCTGTGTGGAAAGAAGCAA
5900
ThrThrLeuPheCysAlaSerAspAlaLysSerTyrGluThrGluValHisAsnIleTrp
CCACTACTCTATTTTGTGCATCAGATGCTAAATCATATGAAACAGAAGTACATAACATCT
6000
AlaThrHisAlaCysValProThrAspProAsnProGlnGluIleGluLeuGluAsnVal
GGGCTACACATGCCTGTGTACCCACGGACCCCAACCCACAAGAAATAGAACTGGAAAATG
ThrGluGlyPheAsnMetTrpLysAsnAsnMetValGluGlnMetHisGluAspIleIle
TCACAGAAGGGTTTAACATGTGGAAAAATAACATGGTGGAGCAGATGCATGAGGATATAA
6100

FIG. 7F

```
  SerLeuTrpAspGlnSerLeuLysProCysValLysLeuThrProLeuCysValThrLeu
TCAGTTTATGGGATCAAAGCCTAAAACCATGTGTAAAGCTAACCCCACTCTGTGTCACTT

  AsnCysThrAsnValAsnGlyThrAlaValAsnGlyThrAsnAlaGlySerAsnArgThr
TAAACTGCACTAATGTGAATGGGACTGCTGTCAATGGGACTAATGCTGGGAGTAATAGGA
          6200
  AsnAlaGluLeuLysMetGluIleGlyGluValLysAsnCysSerPheAsnIleThrPro
CTAATGCAGAATTGAAAATGGAAATTGGAGAAGTGAAAAACTGCTCTTTCAATATAACCC
                                                        6300
  ValGlySerAspLysArgGlnGluTyrAlaThePheTyrAsnLeuAspLeuValGlnIle
CAGTAGGAAGTGATAAAAGGCAAGAATATGCAACTTTTTATAACCTTGATCTAGTACAAA

  AspAspSerAspAsnSerSerTyrArgLeuIleAsnCysAsnThrSerValIleThrGln
TAGATGATAGTGATAATAGTAGTTATAGGCTAATAAATTGTAATACCTCAGTAATTACAC
                                  6400
  AlaCysProLysValThrPheAspProIleProIleHisTyrCysAlaProAlaGlyPhe
AGGCTTGTCCAAAGGTAACCTTTGATCCAATTCCCATACATTATTGTGCCCCAGCTGGTT

  AlaIleLeuLysCysAsnAspLysLysPheAsnGlyThrGluIleCysLysAsnValSer
TTGCAATTCTAAAGTGTAATGATAAGAAGTTCAATGGAACGGAAATATGTAAAAATGTCA
          6500
  ThrValGlnCysThrHisGlyIleLysProValValSerThrGlnLeuLeuLeuAsnGly
GTACAGTACAATGTACACATGGAATTAAGCCAGTGGTGTCAACTCAACTGCTGTTAAATG
                                                        6600
  SerLeuAaGluGluGluIleMetIleArgSerGluAsnLeuThrAspAsnThrLysAsn
GCAGTCTAGCAGAAGAAGAGATAATGATTAGATCTGAAAATCTCACAGACAATACTAAAA

  IleIleValGlnLeuAsnGluThrValThrIleAsnCysThrArgProGlyAsnAsnThr
ACATAATAGTACAGCTTAATGAAACTGTAACAATTAATTGTACAAGGCCTGGAAACAATA
                                  6700
  ArgArgGlyIleHisPheGlyProGlyGlnAlaLeuTyrThrThrGlyIleValGlyAsp
CAAGAAGAGGCATACATTTCGGCCCAGGGCAAGCACTCTATACAACAGGGATAGTAGGAG

  IleArgArgAlaTyrCysThrIleAsnGluThrGluTrpAspLysThrLeuGlnGlnVal
ATATAAGAAGAGCATATTGTACTATTAATGAAACAGAATGGGATAAAACTTTACAACAGG
          6800
  AlaValLysLeuGlySerLeuLeuAsnLysThrLysIleIlePheAsnSerSerSerGly
TAGCTGTAAAACTAGGAAGCCTTCTTAACAAAACAAAAATAATTTTTAATTCATCCTCAG
                                                        6900
  GlyAspProGluIleThrThrHisSerPheAsnCysArgGlyGluPhePheTyrCysAsn
GAGGGGACCCAGAAATTACAACACACAGTTTTAATTGTAGAGGGGAATTTTTCTACTGTA

  ThrSerLysLeuPheAsnSerThrTRpGlnAsnAsnGlyAlaArgLeuSerAsnSerThr
ATACATCAAAACTGTTTAATAGTACATGGCAGAATAATGGTGCAAGACTAAGTAATAGCA
                                  7000
  GluSerThrGlySerIleThrLeuProCysArgIleLysGlnIleIleAsnMetTrpGln
CAGAGTCAACTGGTAGTATCACACTCCCATGCAGAATAAAACAAATTATAAATATGTGGC

  LysThrGlyLysAlaMetTryAlaProProIleAlaGlyValIleAsnCysLeuSerAsn
AGAAAACAGGAAAAGCTATGTATGCCCCTCCCATCGCAGGAGTCATCAACTGTTTATCAA
          7100
  IleThrGlyLeuIleLeuThrArgAspGlyGlyAsnSerSerAspAsnSerAspAsnGlu
ATATTACAGGGCTGATATTAACAAGAGATGGTGGAAATAGTAGTGACAATAGTGACAATG
                                                        7200
```

FIG. 7G

```
  ThrLeuArgProGlyGlyGlyAspMetArgAspAsnTrpIleSerGluLeuTyrLysTyr
AGACCTTAAGACCTGGAGGAGGAGATATGAGGGACAATTGGATAAGTGAATTATATAAAT

  LysValValArgIleGluProLeuGlyValAlaProThrLysAlaLysArgArgValVal
ATAAAGTAGTAAGAATTGAACCCCTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGG
                                  7300
  GluArgGluLysArgAlaIleGlyLeuGlyAlaMetPheLeuGlyPheLeuGlyAlaAla
TGGAAAGAGAAAAAAGAGCAATAGGACTAGGAGCCATGTTCCTTGGGTTCTTGGGAGCAG

  GlySerThrMetGlyAlaAlaSerLeuThrLeuThrValGlnAlaArgGlnLeuLeuSer
CAGGAAGCACGATGGGCGCAGCGTCACTAACGCTGACGGTACAGGCCAGACAGTTACTGT
          7400
  GlyIleValGlnGlnGlnAsnAsnLeuLeuArgAlaIleGluAlaGlnGlnHisLeuLeu
CTGGTATAGTGCAACAGCAAAACAATTTGCTGAGGGCTATAGAGGCGCAACAGCATCTGT
                                                  7500
  GlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaValGluArgTyr
TGCAACTCACGGTCTGGGGCATTAAACAGCTCCAGGCAAGAGTCCTGGCTGTGGAAAGAT

  LeuGlnAspGlnArgLeuLeuGlyMetTrpGlyCysSerGlyLysHisIleCysThrThr
ACCTACAGGATCAACGGCTCCTAGGAATGTGGGGTTGCTCTGGAAAACACATTTGCACCA
                                  7600
  PheValProTrpAsnSerSerTrpSerAsnArgSerLeuAspAspIleTrpAsnAsnMet
CATTTGTGCCTTGGAACTCTAGTTGGAGTAATAGATCTCTAGATGACATTTGGAATAATA

  ThrTrpMetGlnTrpGluLysGluIleSerAsnTyrThrGlyIleIleTyrAsnLeuIle
TGACCTGGATGCAGTGGGAAAAAGAAATTAGCAATTACACAGGCATAATATACAACTTAA
          7700
  GluGluSerGlnIleGlnGlnGluLysAsnGluLysGluLeuLeuGluLeuAspLysTrp
TTGAAGAATCGCAAATCCAGCAAGAAAAGAATGAAAAGGAATTATTGGAATTGGACAAGT
                                                  7800
  AlaSerLeuTrpAsnTrpPheSerIleSerLysTrpLeuTrpTyrIleArgIlePheIle
GGGCAAGTTTGTGGAATTGGTTTAGCATATCAAAATGGCTGTGGTATATAAGAATATTCA

  IleValValGlyGlyLeuIleGlyLeuArgIleIlePheAlaValLeuSerLeuValAsn
TAATAGTAGTAGGAGGCTTAATAGGTTTAAGAATAATTTTTGCTGTGCTTTCTTTAGTAA
                                  7900
  ArgValArgGlnGlyTyrSerProLeuSerLeuGlnThrLeuLeuProThrProArgGly
ATAGAGTTAGGCAGGGATACTCACCTCTGTCGTTGCAGACCCTCCTCCCAACACCGAGGG

  ProProAspArgProGluGlyIleGluGluGluGlyGlyGluGlnGlyArgGlyArgSer
GACCACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGCAAGGCAGAGGCAGAT
          8000
  IleArgLeuValAsnGlyPheSerAlaLeuIleTrpAspAspLeuArgAsnLeuCysLeu
CAATTCGATTGGTGAACGGATTCTCAGCACTTATCTGGGACGACCTGAGGAACCTGTGCC
                                                  8100
  PheSerTyrHisArgLeuArgAspLeuLeuLeuIleAlaThrArgIleValGluLeuLeu
TCTTCAGTTACCACCGCTTGAGAGACTTACTCTTAATTGCAACGAGGATTGTGGAACTTC

  GlyArgArgGlyTrpGluAlaLeuLysTyrLeuTrpAsnLeuLeuGlnTyrTrpGlyGln
TGGGACGCAGGGGGTGGGAAGCCCTCAAATATCTGTGGAATCTCCTGCAATATTGGGGTC
                                  8200
```

FIG. 7H

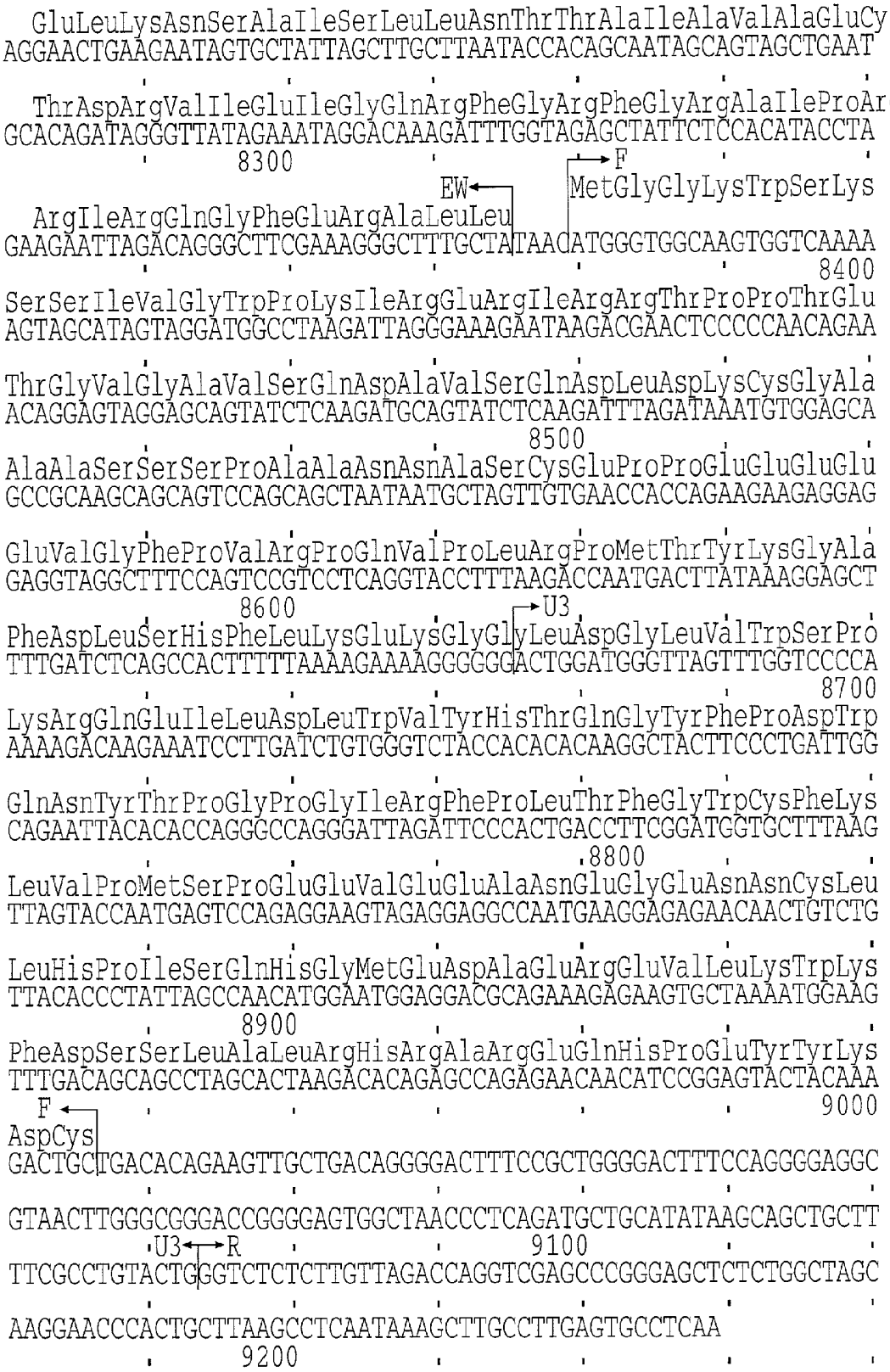

GluLeuLysAsnSerAlaIleSerLeuLeuAsnThrThrAlaIleAlaValAlaGluCys
AGGAACTGAAGAATAGTGCTATTAGCTTGCTTAATACCACAGCAATAGCAGTAGCTGAAT
ThrAspArgValIleGluIleGlyGlnArgPheGlyArgPheGlyArgAlaIleProArg
GCACAGATAGGGTTATAGAAATAGGACAAAGATTTGGTAGAGCTATTCTCCACATACCTA
8300
F
EW
MetGlyGlyLysTrpSerLys
ArgIleArgGlnGlyPheGluArgAlaLeuLeu
GAAGAATTAGACAGGGCTTCGAAAGGGCTTTGCTATAACATGGGTGGCAAGTGGTCAAAA
8400
SerSerIleValGlyTrpProLysIleArgGluArgIleArgArgThrProProThrGlu
AGTAGCATAGTAGGATGGCCTAAGATTAGGGAAAGAATAAGACGAACTCCCCCAACAGAA
ThrGlyValGlyAlaValSerGlnAspAlaValSerGlnAspLeuAspLysCysGlyAla
ACAGGAGTAGGAGCAGTATCTCAAGATGCAGTATCTCAAGATTTAGATAAATGTGGAGCA
8500
AlaAlaSerSerSerProAlaAlaAsnAsnAlaSerCysGluProProGluGluGluGlu
GCCGCAAGCAGCAGTCCAGCAGCTAATAATGCTAGTTGTGAACCACCAGAAGAAGAGGAG
GluValGlyPheProValArgProGlnValProLeuArgProMetThrTyrLysGlyAla
GAGGTAGGCTTTCCAGTCCGTCCTCAGGTACCTTTAAGACCAATGACTTATAAAGGAGCT
8600
U3
PheAspLeuSerHisPheLeuLysGluLysGlyGlyLeuAspGlyLeuValTrpSerPro
TTTGATCTCAGCCACTTTTTAAAAGAAAAGGGGGGACTGGATGGGTTAGTTTGGTCCCCA
8700
LysArgGlnGluIleLeuAspLeuTrpValTyrHisThrGlnGlyTyrPheProAspTrp
AAAAGACAAGAAATCCTTGATCTGTGGGTCTACCACACACAAGGCTACTTCCCTGATTGG
GlnAsnTyrThrProGlyProGlyIleArgPheProLeuThrPheGlyTrpCysPheLys
CAGAATTACACACCAGGGCCAGGGATTAGATTCCCACTGACCTTCGGATGGTGCTTTAAG
8800
LeuValProMetSerProGluGluValGluGluAlaAsnGluGlyGluAsnAsnCysLeu
TTAGTACCAATGAGTCCAGAGGAAGTAGAGGAGGCCAATGAAGGAGAGAACAACTGTCTG
LeuHisProIleSerGlnHisGlyMetGluAspAlaGluArgGluValLeuLysTrpLys
TTACACCCTATTAGCCAACATGGAATGGAGGACGCAGAAAGAGAAGTGCTAAAATGGAAG
8900
PheAspSerSerLeuAlaLeuArgHisArgAlaArgGluGlnHisProGluTyrTyrLys
TTTGACAGCAGCCTAGCACTAAGACACAGAGCCAGAGAACAACATCCGGAGTACTACAAA
F
9000
AspCys
GACTGCTGACACAGAAGTTGCTGACAGGGGACTTTCCGCTGGGGACTTTCCAGGGGAGGC
GTAACTTGGGCGGGACCGGGGAGTGGCTAACCCTCAGATGCTGCATATAAGCAGCTGCTT
U3
R
9100
TTCGCCTGTACTGGGTCTCTCTTGTTAGACCAGGTCGAGCCCGGGAGCTCTCTGGCTAGC
AAGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCCTCAA
9200

FIG. 7I

VARIANT OF LAV VIRUSES

BACKGROUND OF THE INVENTION

[0001] The present invention relates to a virus capable of inducing lymphadenopathies (hereinafter "LAS") and acquired immuno-depressive syndromes (hereinafter "AIDS"), to antigens of this virus, particularly in a purified form, and to a process for producing these antigens, particularly antigens of the envelope of this virus. The invention also relates to polypeptides, whether glycosylated or not, produced by the virus and to DNA sequences which code for such polypeptides. The invention further relates to cloned DNA sequences hybridizable to genomic RNA and DNA of the lymphadenopathy associated virus (hereinafter "LAV") of this invention and to processes for their preparation and their use. The invention still further relates to a stable probe including a DNA sequence which can be used for the detection of the LAV virus of this invention or related viruses or DNA proviruses in any medium, particularly biological, and in samples containing any of them.

[0002] An important genetic polymorphism has been recognized for the human retrovirus which is the cause of AIDS and other diseases like LAS, AIDS-related complex (hereinafter "ARC") and probably some encephalopathies (for review, see Weiss, 1984). Indeed all of the isolates, analyzed until now, have had distinct restriction maps, even those recovered at the same place and time [Benn et al., 1985]. Identical restriction maps have only been observed for the first two isolates which were designated LAV [Alizon et al., 1984] and human T-cell lymphotropic virus type 3 (hereinafter "HTLV-3"[Hahn et al., 1984] and which appear to be exceptions. The genetic polymorphism of the AIDS virus was better assessed after the determination of the complete nucleotide sequence of LAV [Wain-Hobson et al., 1985], HTLV-3 [Ratner et al., 1985; Muesing et al., 1985] and a third isolate designated AIDS-associated retrovirus (hereinafter "ARV2") [Sanchez-Pescador et al., 1985]. In particular, it appeared that, besides the nucleic acid variations responsible for the restriction map polymorphism, isolates could differ significantly at the protein level, especially in the envelope (up to 13% of difference between ARV and LAV), by both amino acids substitutions and reciprocal insertions-deletions [Rabson and Martin, 1985].

[0003] Nevertheless, such differences did not go so far as to destroy the immunological similarity of such isolates as evidenced by the capabilities of their similar proteins, (e.g., core proteins of similar nature, such as the p25 proteins, or similar envelope glycoproteins, such as the 110-120 kD glycoproteins) to immunologically cross-react. Accordingly, the proteins of any of said LAV viruses can be used for the in vitro detection of antibodies induced in vivo and present in biological fluids obtained from individuals infected with the other LAV variants. Therefore, these viruses are grouped together as a class of LAV viruses (hereinafter "LAV-1 viruses").

SUMMARY OF THE INVENTION

[0004] In accordance with this invention, a new virus has been discovered that is responsible for diseases clinically related to AIDS and that can be classified as a LAV-1 virus but that differs genetically from known LAV-1 viruses to a much larger extent than the known LAV-1 viruses differ from each other. The new virus is basically characterized by the cDNA sequence which is shown in **FIGS. 7A** to **7I**, and this new virus is hereinafter generally referred to as "LAV-$_{MAL}$".

[0005] Also in accordance with this invention, variants of the new virus are provided. The RNAs of these variants and the related cDNAs derived from said RNAs are hybridizable to corresponding parts of the cDNA of LAV_{MAL} The DNA of the new virus also is provided, as well as DNA fragments derived therefrom hybridizable with the genomic RNA of LAV_{MAL}, such DNA and DNA fragments particularly consisting of the cDNA or cDNA fragments of LAV_{MAL} or of recombinant DNAs containing such cDNA or cDNA fragments.

[0006] DNA recombinants containing the DNA or DNA fragments of LAV or its variants are also provided. It is of course understood that fragments which would include some deletions or mutations which would not substantially alter their capability of also hybridizing with the retroviral genome of LAV_{MAL} are to be considered as forming obvious equivalents of the DNA or DNA fragments referred to hereinabove.

[0007] Cloned probes are further provided which can be made starting from any DNA fragment according to the invention, as are recombinant DNAs containing such fragments, particularly any plasmids amplifiable in procaryotic or eucaryotic cells and carrying said fragments. Using cloned DNA containing a DNA fragment of LAV_{MAL} as a molecular hybridization probe—either by marking with radionucleotides or with fluorescent reagents—LAV virion RNA may be detected directly, for example, in blood, body fluids and blood products (e.g., in antihemophylic factors such as Factor VIII concentrates). A suitable method for achieving such detection comprises immobilizing LAV_{MAL} on a support (e.g., a nitrocellulose filter), disrupting the virion and hybridizing with a labelled (radiolabelled or "cold" fluorescent- or enzyme-labelled) probe. Such an approach has already been developed for Hepatitis B virus in peripheral blood (according to Scotto J. et al. Hepatology (1983), 3, 379-384).

[0008] Probes according to the invention can also be used for rapid screening of genomic DNA derived from the tissue of patients with LAV related symptoms to see if the proviral DNA or RNA present in their tissues is related to LAV_{MAL}. A method which can be used for such screening comprises the following steps: extraction of DNA from tissue, restriction enzyme cleavage of said DNA, electrophoresis of the fragments and Southern blotting of genomic DNA from tissues and subsequent hybridization with labelled cloned LAV provival DNA. Hybridization in situ can also be used. Lymphatic fluids and tissues and other non-lymphatic tissues of humans, primates and other mammalian species can also be screened to see if other evolutionary related retroviruses exist. The methods referred to hereinabove can be used, although hybridization and washings would be done under non-stringent conditions.

[0009] The DNA according to the invention can be used also for achieving the expression of LAV viral antigens for diagnostic purposes, as well as for the production of a vaccine against LAV. Fragments of particular advantage in that respect will be discussed later. The methods which can be used are multifold:

[0010] a) DNA can be transfected into mammalian cells with appropriate selection markers by a variety of techniques, such as calcium phosphate precipitation, polyethylene glycol, protoplast-fusion, etc.

[0011] b) DNA fragments corresponding to genes can be cloned into expression vectors for *E. coli*, yeast or mammalian cells and the resultant proteins purified.

[0012] c) The provival DNA can be "shot-gunned" (fragmented) into procaryotic expression vectors to generate fusion polypeptides.

[0013] Recombinants, producing antigenically competent fusion proteins, can be identified by simply screening the recombinants with antibodies against LAV_{MAL} antigens. Particular reference in this respect is made to those portions of the genome of LAV_{MAL} which, in the figures, are shown to belong to open reading frames and which encode the products having the polypeptidic backbones shown.

[0014] Different polypeptides which appear in **FIGS. 7A** to **7I** are still further provided. Methods disclosed in European application 0 178 978 and in PCT application PCT/EP 85/00548, filed Oct. 18, 1985, are applicable for the production of such peptides from LAV_{MAL}. In this regard, polypeptides are provided containing sequences in common with polypeptides comprising antigenic determinants included in the proteins encoded and expressed by the LAV_{MAL} genome. Means are also provided for the detection of proteins of LAV_{MAL} particularly for the diagnosis of AIDS or pre-AIDS or, to the contrary, for the detection of antibodies against LAV_{MAL} or its proteins, particularly in patients afflicted with AIDS or pre-AIDS or more generally in asymtomatic carriers and in blood-related products. Further provided are immunogenic polypeptides and more particularly protective polypeptides for use in the preparation of vaccine compositions against AIDS or related syndroms.

[0015] Yet further provided are polypeptide fragments having lower molecular weights and having peptide sequences or fragments in common with those shown in **FIGS. 7A** to **7I**. Fragments of smaller sizes can be obtained by resorting to known techniques, for instance, by cleaving the original larger polypeptide by enzymes capable of cleaving it at specific sites. By way of examples may be mentioned the enzyme of *Staphylococcyus aureus* V8, α-chymotrypsine, "mouse sub-maxillary gland protease" marketed by the Boehringer company, Vibrio alginolyticus chemovar iophaqus collagenase, which specifically recognizes the peptides Gly-Pro, Gly-Ala, etc.

[0016] Other features of this invention will appear in the following disclosure of data obtained starting from LAV_{MAL}, in relation to the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] **FIGS. 1A and 1B** provide comparative restriction maps of the genomas of LAV_{MAL} as compared to LAV_{ELI} (Applicants' related new LAV virus which is the subject of their copending application, filed herewith) and LAV_{BRU} (a known LAV isolate deposited at the Collection Nationale des Cultures de Micro-organismes (hereinafter "CNCM") of the Pasteur Institute, Paris, France under No. I-232 on Jul. 15, 1983);

[0018] **FIG. 2** shows comparative maps setting forth the relative positions of the open reading frames of the above genomas;

[0019] FIGS. **3A-3F** (also designated generally hereinafter "**FIG. 3**") indicate the relative correspondance between the proteins (or glycoproteins) encoded by the open reading frames, whereby amino acid residues of protein sequences of LAV_{MAL} are in vertical alignment with corresponding amino acid residues (numbered) of corresponding or homologous proteins or glycoproteins of LAV_{BRU}, as well as LAV_{ELI} and ARV 2.

[0020] FIGS. **4A-4B** (also designated generally hereinafter "**FIG. 4**") provide tables quantitating the sequence divergence between homologous proteins of LAV_{BRU}, LAV_{ELI} and LAV_{MAL};

[0021] **FIG. 5** shows diagrammatically the degree of divergence of the different virus envelope proteins;

[0022] **FIGS. 6A and 6B** ("**FIG. 6**" when consulted together) render apparent the direct repeats which appear in the proteins of the different AIDS virus isolates.

[0023] FIGS. **7A-7I** show the full nucleotidic sequences of LAV_{MAL}.

DETAILED DESCRIPTION OF THE INVENTION

[0024] Characterization and Molecular Cloning of an African Isolate.

[0025] The different AIDS virus isolates concerned are designated by three letters of the patients name, LAV_{BRU} referring to the prototype AIDS virus isolated in 1983 from a French homosexual patient with LAS and thought to have been infected in the USA in the preceding years [Barré-Sinoussi et al., 1983]. LAV_{MAL} was recovered in 1985 from a 7-year old boy from Zaire. Related LAV_{ELI} was recovered in 1983 from a 24-year old woman with AIDS from Zaire. Recovery and purification of the LAV_{MAL} virus were performed according to the method disclosed in European Patent Application 84 401834/138 667 filed on Sep. 9, 1984.

[0026] LAV_{MAL} is indistinguishable from the previously characterized isolates by its structural and biological properties in vitro. Virus metabolic labelling and immune precipitation by patient MAL sera, as well as reference sera, showed that the proteins of LAV_{MAL} had the same molecular weight (hereinafter "MW") as, and cross-reacted immunologically with those of, prototype AIDS virus (data not shown) of the LAV-1 class.

[0027] Reference is again made to European Application 178 978 and International Application PCT/EP 85/00548 as concerns the purification, mapping and sequencing procedures used herein. See also the discussion under the headings "Experimental Procedures" and "Significance of the Figures" hereinafter.

[0028] Primary restriction enzyme analysis of LAV_{MAL} genome was done by southern blot with total DNA derived from acutely infected lymphocytes, using cloned LAV_{BRU} complete genome as probe. Overall cross-hybridization was observed under stringent conditions, but the restriction profile of the Zairian isolate was clearly different. Phage lambda clones carrying the complete viral genetic information were

obtained and further characterized by restriction mapping and nucleotide sequence analysis. A clone (hereinafter "M-H11") was obtained by complete HindIII restriction of DNA from LAV_{MAL}-infected cells, taking advantage of the existence of a unique HindIII site in the long terminal repeat (hereinafter "LTR"). M-H11 is thus probably derived from unintegrated viral DNA since that species was at least ten times more abundant than integraded provirus.

[0029] **FIG. 1B** gives a comparaison of the restriction maps derived from the nucleotide sequences of LAV_{ELI}, LAV_{MAL} and prototype LAV_{BRU}, as well as from three other Zairian isolates (hereinafter "Z1", "Z2", and "Z3" respectively) previously mapped for seven restriction enzymes [Benn et al., 1985]. Despite this limited number, all of the profiles are clearly different (out of the 23 sites making up the map of LAV_{BRU}, only seven are present in all six maps presented), confirming the genetic polymorphism of the AIDS virus. No obvious relationship is apparent between the five Zairian maps, and all of their common sites are also found in LAV_{BRU}.

[0030] Conservation of the Genetic Organization.

[0031] The genetic organization of LAV_{MAL} as deduced from the complete nucleotide sequences of its cloned genome is identical to that found in other isolates, i.e., 5' gag-pol-central region-env-F3'. Most noticeable is the conservation of the "central region" (**FIG. 2**), located between the pol and env genes, which is composed of a series of overlapping open reading frames (hereinafter "orf") previously designated Q, R, S, T, and U in the ovine lentivirus visna [Sonigo et al., 1985]. The product of orf S (also designated "tat") is implicated in the transactivation of virus expression [Sodroski et al., 1985; Arya et al., 1985]; the biological role of the product of orf Q (also designated "sor" or "Horf A") is still unknown [Lee et al., 1986 Kang et al., 1986]. Of the three other orfs, R, T, and U, only orf R is likely to be a seventh viral gene, for the following reasons: the exact conservation of its relative position with respect to Q and S (**FIG. 2**), the ponstant presence of a possible splice acceptor and of a consensus AUG initiator codon, its similar codon usage with respect to viral genes, and finally the fact that the variation of its protein sequence within the different isolates is comparable to that of gag, pol and Q (see **FIG. 4**).

[0032] Also conserved are the sizes of the U3, R and U5 elements of the LTR (data not shown), the location and sequence of their regulatory elements such as TATA box and AATAAA polyadenylation signal, and their flanking sequences, i.e., primer binding site (hereinafter "PBS") complementary to 3' end of $tRNA^{LYS}$ and polypurine tract (hereinafter "PPT"). Most of the genetic variability within the LTR is located in the 5' half of U3 (which encodes a part of orf F) while the 3' end of U3 and R, which carry most of the cis-acting regulatory elements, promoter, enhancer and trans-activating factor receptor [Rosen et al., 1985], as well as the U5 element, are well-conserved.

[0033] Overall, it clearly appears that this Zairian isolate, LAV_{MAL} is the same type of retrovirus as the previously sequenced isolates of American or European origin.

[0034] Variability of the Viral Proteins

[0035] Despite their identical genetic organization, the LAV_{ELI} and LAV_{MAL} shows substantial differences in the primary structure of their proteins. The amino acid sequences of LAV_{ELI} and LAV_{MAL} proteins are presented in FIGS. **3A-3F**, aligned with those of LAV_{BRU} and ARV 2. Their divergence was quantified as the percentage of amino acids substitutions in two-by-two alignments (**FIG. 4**). The number of insertions and deletions that had to be introduced in each of these alignments has also been scored.

[0036] Three general observations can be made. First, the protein sequences of the LAV_{ELI} and LAV_{MAL} are more divergent from LAV_{BRU} than are those of HTLV-3 and ARV 2 (**FIG. 4A**); similar results are obtained if ARV 2 is taken as reference (not shown). The range of genetic polymorphism between isolates of the AIDS virus is considerably greater than previously observed. Second, our two sequences confirm that the envelope is more variable than the gag and pol genes. Here again, the relatively small difference observed between the env of LAV_{BRU} and HTLV-3 appears as an exception. Third, the mutual divergence of the LAV_{ELI} and LAV_{MAL} (**FIG. 4B**) is comparable to that between LAV_{BRU} and either of them; as far as we can extrapolate from only three sequenced isolates from the USA and Europe and two (LAV_{ELI} and LAV_{MAL}) from Africa, this is indicative of a wider evolution of the AIDS virus in Africa. gag and pol: Their greater degree of conservation compared to the envelope is consistent with their encoding important structural or enzymatic activities. Of the three mature gag proteins, the p25 which was the first recognized immunogenic protein of LAV [Barré-Sinoussi et al., 1983] is also the better conserved (**FIG. 3**). In gag and pol, differences between isolates are principally due to point mutations, and only a small number of insertional or deletional events is observed. Among these, we must note the presence in the overlapping part of gag and pol of LAVBRU of an insertion of 12 amino acids (AA) which is encoded by the second copy of a 36 bp direct repeat present only in this isolate and in HTLV-3. This duplication was omitted because of a computing error in the published sequence of LAV_{BRU} (position 1712, Wain-Hobson et al., 1985) but was indeed present in the HTLV-3 sequences [Ratner et al., 1985; Muesing et al., 1985].

[0037] env: Three segments can be distinguished in the envelope glycoprotein precursor [Allan et al., 1985; Montagnier et al., 1985; DiMarzoVeronese et al., 1985]. The first is the signal peptide (positions 1-33 in **FIG. 3**), and its sequence appears as variable; the second segment (pos. 34-530) forms the outer membrane protein (hereinafter "OMP" or "gp110") and carries most of the genetic variations, and in particular almost all of the numerous reciprocal insertions and deletions; the third segment (531-877) is separated from the OMP by a potential cleavage site following a constant basic stretch (Arg-Glu-Lys-Arg) and forms the transmembrane protein (hereinafter "TMP" or "gp41") responsible for the an**30** chorage of the envelope glycoprotein in the cellular membrane. A better conservation of the TMP than the OMP has also been observed between the different murine leukemia viruses (hereinafter "MLV") [Koch et al., 1983] and could be due to structural constraints.

[0038] From the alignment of **FIG. 3** and the graphical representation of the envelope variability shown in **FIG. 5**, we clearly see the existence of conserved domains, with little or no genetic variation, and hypervariable domains, in which even the alignment of the different sequences is very difficult, because of the existence of a large number of mutations and of reciprocal insertions and deletions. We

have not included the sequence of the envelope of the HTLV-3 isolate since it so close to that of LAV_{BRU} (cf. **FIG. 4**), even in the hypervariable domains, that it did not add anything to the analysis. While this graphical representation will be refined by more sequence data, the general profile is already apparent, with three hypervariable domains (Hy1, 2 and 3) all being located in the OMP and separated by three well-conserved stretches (residues 37-130, 211-289, and 488-530 of **FIG. 3** alignment) probably associated with important biological functions.

[**0039**] In spite of the extreme genetic variability, the folding pattern of the envelope glycoprotein is probably constant. Indeed the position of virtually all of the cysteine residues is conserved within the different isolates (**FIGS. 3 and 5**), and the only three variable cysteines fall either in the signal peptide or in the very C-terminal part of the TMP. The hypervariable domains of the OMP are bounded by conserved cysteines, suggesting that they may represent loops attached to the common folding pattern. Also the calculated hydropathic profiles [Kyte and Doolittle, 1982] of the different envelope proteins are remarkably conserved (not shown).

[**0040**] About half of the potential N-glycosylation sites, Asn-X-Ser/Thr, found in the envelopes of the Zairian isolates map to the same positions in LAV_{BRU} (17/26 for LAV_{ELI} and 17/28 for LAV_{MAL}). The other sites appear to fall within variable domains of env, suggesting the existence of differences in the extent of envelope glycosylation between different isolates.

[**0041**] Other viral proteins: Of the three other identified viral proteins, the p27 encoded by orf F, 3' of env [Allan et al., 1985b] is the most variable (**FIG. 4**). The proteins encoded by orfs Q and S of the central region are remarkable by their absence of insertions/deletions. Surprisingly, a high frequency of amino acids substitutions, comparable to that observed in env, is found for the product of orf S (trans-activating factor). On the other hand, the protein encoded by orf Q is no more variable than gag. Also noticeable is the lower variation of the proteins encoded by the central regions of LAV_{ELI} and LAV_{MAL}.

[**0042**] With the availability of the complete nucleotide sequence from five independant isolates, some general features of the AIDS virus' genetic variability are now emerging. Firstly, its principal cause is point mutations which very often result in amino acid substitutions and which are more frequent in the 3' part of the genome (orf S, env and orf F). Like all RNA viruses, the retroviruses are thought to be highly subject to mutations caused by errors of the RNA polymerases during their replication, since there is no proof-reading, of this step [Holland et al., 1982; Steinhauer and Holland, 1986].

[**0043**] Another source of genetic diversity is insertions/deletions. From the **FIG. 3** alignments, insertional events seem to be implicated in most of the cases, since otherwise deletions should have occurred in independent isolates at precisely the same locations. Furthermore, upon analyzing these insertions, we have observed that they most often represent one of the two copies of a direct repeat (**FIG. 6**). Some are perfectly conserved like the 36 bp repeat in the gag-pol overlap of LAV_{BRU} (**FIG. 6**-*a*); others carry point mutations resulting in amino acid substitutions, and as a consequence, they are more difficult to observe, though clearly present, in the hypervariable domains of env (cf. **FIG. 6**-*g* and -*h*). As noted for point mutations, env gene and orf F also appear as more susceptible to that form of genetic variation than the rest of the genome. The degree of conservation of these repeats must be related to their date of occurrence in the analyzed sequences: the more degenerated, the more ancient. A very recent divergence of LAV_{BRU} and HTLV-3 is suggested by the extremely low number of mismatched AA between their homologous proteins. However, one of the LAV repeats (located in the Hyl domain of env, **FIG. 6**-*f*) is not present in HTLV-3, indicating that this generation of tandem repeats is a rapid source of genetic diversity. We have found no traces of such a phenomenon, even when comparing very closely related viruses, such as the Mason-Pfizer monkey virus (hereinafter "MPMV") [Sonigo et al., 1986], and an immunosuppressive simian virus (hereinafter "SRV-1") [Power et al., 1986]. Insertion or deletion of one copy of a direct repeat have been occasionally reported in mutant retroviruses [Shimotohno and Temin, 1981; Darlix, 1986], but the extent to which we observe this phenomenon is unprecedented. The molecular basis of these duplications is unclear, but could be the copy-choice phenomenon, resulting from the diploidy of the retroviral genome [Varmus and Swanstrom, 1984; Clark and Mak, 1983]. During the synthesis of the first-strand of the viral DNA, jumps are known to occur from one RNA molecule to another, especially when a break or a stable secondary structure is present on the template; an inaccurate re-initiation on the other RNA template could result in the generation (or the elimination) of a short direct repeat.

[**0044**] Genetic variability and subsequent antigenic modifications have often been developed by microorganisms as a means for avoiding the host's immune response, either by modifying their epitopes during the course of the infection, as in trypanosomes [Borst and Cross, 1982], or by generating a large repertoire of antigens, as observed in influenza virus [Webster et al., 1982]. As the human AIDS virus is related to animal lentiviruses [Sonigo et al., 1985; Chiu et al., 1985], its genetic variability could be a source of antigenic variation, as can be observed during the course of the infection by the ovine lentivirus visna (Scott et al., 1979; Clements et al., 1980] or by the equine infectious anemia virus (hereinafter "EIAV") [Montelaro et al., 1984]. However, a major discrepancy with these animal models is the extremely low, and possibly nonexistant, neutralizing activity of the sera of individuals infected by the AIDS virus, whether they are healthy carriers, displaying minor symptoms, or afflicted with AIDS [Weiss et al., 1985; Clavel et al., 1985]. Furthermore, even for the visna virus the exact role of antigenic variation in the pathogenesis is unclear [Thormar et al., 1983; Lutley et al., 1983]. We rather believe that genetic variation represents a general selective advantage for lentiviruses by allowing an adaptation to different environments, for example by modifying their tissue or host tropisms. In the particular case of the AIDS virus, rapid genetic variations are tolerated, especially in the envelope. This could allow the virus to become adapted to different "micro-environments" of the membrane of their principal target cells, namely the T4 lymphocytes. These "micro-environments" could result from the immediate vicinity of the virus receptor to polymorphic surface proteins, differing either between individuals or between clones of lymphocytes.

[0045] Conserved Domains in the AIDS Virus Envelope

[0046] Since the proteins of most of the isolates are antigenically cross-reactive, the genotypic differences do not seem to affect the sensitivity of actual diagnostic tests, based upon the detection of antibodies to the AIDS virus and using purified virions as antigens. They nevertheless have to be considered for the development of the "second-generation" tests, that are expected to be more specific, and will use smaller synthetic or genetically-engineered viral antigens. The identification of conserved domains in the highly immunogenic envelope glycoprotein and the core structural proteins (gag) is very important for these tests. The conserved stretch found at the end of the OMP and the beginning of the TMP (490-620, **FIG. 3**) could be a good candidate, since a bacterial fusion protein containing this domain was well-detected by AIDS patients' sera [Chang et al., 1985].

[0047] The envelope, specifically the OMP, mediates the interaction between a retrovirus and its specific cellular receptor [DeLarco and Todaro, 1976; Robinson et al., 1980]. In the case of the AIDS virus, in vitro binding assays have shown the interaction of the envelope glycoprotein gp110 with the T4 cellular surface antigen [McDougal et al., 1986], already thought to be closely associated with the virus receptor [Klatzmann et al., 1984; Dagleish et al., 1984]. Identification of the AIDS virus envelope domains that are responsible for this interaction (receptor-binding domains) appears to be fundamental for understanding of the host-viral interactions and for designing a protective vaccine, since an immune response against these epitopes could possibly elicit neutralizing antibodies. As the AIDS virus receptor is at least partly formed of a constant structure, the T4 antigen, the binding site of the envelope is unlikely to be exclusively encoded by domains undergoing drastic genetic changes between isolates, even if these could be implicated in some kind of an "adaptation". One or several of the conserved domains of the OMP (residues 37-130, 211-289, and 488-530 of **FIG. 3** alignment), brought together by the folding of the protein, must play a part in the virus-receptor interaction, and this can be explored with synthetic or genetically-engineered peptides derived from these domains, either by direct binding assays or indirectly by assaying the neutralizing activity of specific antibodies raised against them.

African AIDS Viruses

[0048] Zaire and the neighbouring countries of Central Africa are considered as an area endemic with the AIDS virus infection, and the possibility that the virus has emerged in Africa has became a subject of intense controversy (see Norman, 1985). From the present study, it is clear that the genetic organization of Zairian isolates is the same as that of american isolates, thereby indicating a common origin. The very important sequence differences observed between the proteins are consistent with a divergent evolutionary process. In addition, the two African isolates are mutually more divergent than the American isolates already analyzed; as far as that observation can be extrapolated, it suggests a longer evolution of the virus in Africa and is also consistent with the fact that a larger fraction of the population is exposed than in developed countries.

[0049] A novel human retrovirus with morphology and biologocal properties (cytopathogenicity, T4 tropism) similar to those of LAV, but nevertheless clearly genetically and antigenically distinct from it, was recently isolated from two patients with AIDS originating from Guinea Bissau, West-Africa [Clavel et al., 1986]. In neighboring Senegal, the population was seemingly exposed to a retrovirus also distinct from LAV but apparently non-pathogenic [Barin et al., 1985; Kanki et al., 1986]. Both of these novel African retroviruses seem to be antigenically related to the simian T-cell lymphotropic virus (hereinafter "STLV-III") shown to be widely present in healthy African green monkeys and other simian species [Kanki et al. 1985]. This raises the possibility of a large group of African primate lentiviruses,. ranging from the apparently non-pathogenic simian viruses to the LAV-type viruses. Their precise relationship will only be known after their complete genetic characterization, but it is already very likely that they have evolved from a common progenitor. The important genetic variability we have observed between isolates of the AIDS virus in Central Africa is probably a hallmark of this entire group and may account for the apparently important genetic divergence between its members (loss of cross-antigenicity in the envelopes). In this sense, the conservation of the tropism for the T4 lymphocytes suggests that it is a major advantage aquired by these retroviruses.

Experimental Procedures

[0050] Virus Isolation

[0051] LAV_{MAL} was isolated from the peripheral blood lymphocytes of the patient as described [Barré-Sinoussi et al., 1983]. Briefly, the lymphocytes were fractionated and co-cultivated with phytohaemagglutinin-stimulated normal human lymphocytes in the presence of interleukin 2 and anti-alpha interferon serum. Viral production was assessed by cell-free reverse transcriptase (hereinafter "RT") activity assay in the cultures and by electron microscopy.

[0052] Molecular Cloning

[0053] Normal donor lymphocytes were acutely infected (10^4 cpm of RT activity/10^6 cells) as described [Barré-Sinoussi et al., 1983], and total DNA was extracted at the beginning of the RT activity peak. A lambda library using the L47-1 vector [Loenen and Brammar, 1982] was constructed by partial HindIII digestion of the DNA as already described [Alizon et al., 1984]. DNA from infected cells was digested to completion with HindIII, and the 9-10 kb fraction was selected on 0.8% low melting point agarose gel and ligated into L47-1 HindIII arms. About 2.10^5 plaques for LAV_{MAL}, obtained by in vitro packaging (Amersham), were plated on *E. coli* LA101 and screened in situ under stringent conditions, using the 9 kb SacI insert of the clone lambda J19 [Alizon et al., 1984] carrying most of the LAV_{BRU} genome as probe. Clones displaying positive signals were plaque-purified and propagated on *E. coli* C600 recBC, and the recombinant phage M-H11 carrying the complete genetic information of LAV_{MAL} was further characterized by restriction mapping.

[0054] Nucleotide Sequence Strategy

[0055] Viral fragments derived from M-H11 were sequenced by the dideoxy chain terminator procedure [Sanger et al., 1977] after "shotgun" cloning in the M13mp8 vector [Messing and Viera, 1982] as previously described [Sonigo et al., 1985]. The viral genome of LAV_{MAL} is 9229

nucleotides long as shown in FIGS. **7A-7I**. Each nucleotide of LAV_{MAL} was determined from more than independent clones on average.

SIGNIFICANCE OF THE FIGURES

[**0056**] **FIG. 1** contains an analysis of AIDS virus isolates, showing:

[**0057**] A/ Restriction maps of the inserts of phage lambda clones derived from cells infected with LAV_{ELI} (hereinafter "E-H12") and with LAV_{MAL} (M-H11). The schematic genetic organization of the AIDS virus has been drawn above the maps. The LTRs are indicated by solid boxes. Restriction sites are indicated as follows: A:AvaI; B:BamHI; Bg:BglII; E:EcoRI; H:HindIII; Hc:HincII; K:KpnI; N:NdeI; P:PstI; S:SacI; and X:XbaI. Asterisks indicate the HindIII cloning sites in lambda L47-1 vector.

[**0058**] B/ A comparison of the sites for seven restriction enzymes in six isolates: the prototype AIDS virus LAV_{BRU}, LAV_{MAL}; and Z1, Z2 and Z3. Restriction sites are represented by the following symbols vertically aligned wih the symbols in **FIG. 1**A: ●: BgIII; * :EcoRI; ▽:HincII; ▼:HindIII; ♦:KpnI; ◊:NdeI; and ○:SacI.

[**0059**] **FIG. 2** shows the genetic organization of the central region in AIDS virus isolates. Stop codons in each phase are represented as vertical bars. Vertical arrows indicate possible AUG initiation codons. Splice acceptor (A) and donor (D) sites identified in subgenomic viral mRNA [Muesing et al., 1985] are shown below the graphic of LAV_{BRU}, and corresponding sites in LAV_{ELI} and LAV_{MAL} are indicated. PPT indicates the repeat of the polypurine tract flanking the 3'LTR. As observed in LAV_{BRU} [Wain-Hobson et al., 1985], the PPT is repeated 256 nucleotides 5' to the end of the pol gene in both the LAV_{ELI} and LAV_{MAL} sequences, but this repeat is degenerated at two positions in LAV_{ELI}.

[**0060**] **FIG. 3** shows an alignment of the protein sequences of four AIDS virus isolates. Isolate LAV_{BRU} [Wain-Hobson et al., 1985] is taken as reference ; only differences with LAV_{BRU} are noted for ARV 2 [Sanchez-Pescador et al., 1985] and the two Zairian isolates LAV_{MAL} and LAV_{ELI}. A minimal number of gaps (-) were introduced in the alignments. The NH_2-termini of $p25^{gag}$ and $p18^{gag}$ are indicated [Sanchez-Pescador, 1985]. The potential cleavage sites in the envelope precursor [Allan et al., 1985a; diMarzoVeronese, 1985] separating the signal peptide (hereinafter "SP"), OMP and TMP are indicated as vertical arrows; conserved cysteines are indicated by black circles and variable cysteines are boxed. The one letter code for each amino acid is as follows: A:Ala C:Cys; D:Asp; E:Glu; F:Phe; G:Gly; H :His; I:Ile; K:Lys; L:Leu; M:Met; N:Asn; P:Pro; Q:Gln; R:Arg; S:Ser; T:Thr; V:Val; W:Trp; Y:Tyr.

[**0061**] **FIG. 4** shows a quantitation of the sequence divergence between homologous proteins of different isolates. Part A of each table gives results deduced from two-by-two alignments using the proteins of LAV_{BRU} as reference, part B, those of LAV_{ELI} as reference. Sources: Muesing et al., 1985 for HTLV-3; Sanchez-Pescador et al., 1985 for ARV 2 and Wain-Hobson et al., 1985 for LAV_{BRU}. For each case in the tables, the size in amino acids of the protein (calculated from the first methionine residue or from the beginning of the orf for pol) is given at the upper left part. Below are given the number of deletions (left) and insertions (right) necessary for the alignment. The large numbers in bold face represent the percentage of amino acids substitutions (insertions/deletions being excluded). Two by two alignments were done with computer assistance [Wilburg and Lipman, 1983], using a gag penalty of 1, K-tuple of 1, and window of 20, except for the hypervariable domains of env, where the number of gaps was made minimum, and which are essentially aligned as shown in **FIG. 3**. The sequence of the predicted protein encoded by orf R of HTLV-3 has not been compared because of a premature termination relative to all other isolates.

[**0062**] **FIG. 5** shows the variability of the AIDS virus envelope protein. For each position x of the alignment of env (**FIG. 3**), variability V(x) was calculated as: V(x)=number of different amino-acids at position x/frequency of the most abundant amino acid at position x. Gaps in the alignments are considered as another amino acid. For an alignment of 4 proteins, V(x) ranges from 1 (identical AA in the 4 sequences) to 16 (4 different AA). This type of representation has previously been used in a compilation of the AA sequence of immunoglobulins variable regions [Wu and Kabat, 1970]. Vertical arrows indicate the cleavage sites ; asterisks represent potential N-glysosylation sites (N-X-S/T) conserved in all three four solates; black triangles represent conserved cysteine residues. Black lozanges mark the three major hydrophobic domains: OMP, TMP and SP; and the hypervariable domains: Hy1, 2 and 3.

[**0063**] **FIG. 6** shows the direct repeats in the proteins of different AIDS virus isolates. These examples are derived from the aligned sequences of gag (a, b), F (c,d) and env (e, f, g, h) shown in **FIG. 3**. The two elements of the direct repeat are boxed, while degenerated positions are underlined.

[**0064**] FIGS. **7A-7I** show the complete cDNA sequence of LAV_{MAL} of this invention.

[**0065**] The invention thus pertains more specifically to the proteins, glycoproteins and other polypeptides including the polypeptidic structures shown in the FIGS. **1-7**. The first and last amino acid residues of these proteins, glycoproteins and polypeptides carry numbers computed from a first amino acid of the open-reading frames concerned, although these numbers do not correspond exactly to those of the LAV_{MAL} proteins concerned, rather to the corresponding proteins of the LAV_{BRU} or sequences shown in **FIGS. 3A, 3B** and **3**C. Thus a number corresponding to a "first amino acid residue" of a LAV_{MAL} protein corresponds to the number of the first amino-acyl residue of the corresponding LAV_{BRU} protein which, in any of **FIGS. 3A, 3B** or **3**C, is in direct alignment with the corresponding first amino acid of the LAV_{MAL} protein. Thus the sequences concerned can be read from FIGS. **7A-7I** to the extent where they do not appear with sufficient clarity from FIGS. **3A-3**F.

[**0066**] The preferred protein sequences of this invention extend between the corresponding "first" and "last" amino acid residues. Also preferred are the protein(s)- or glycoprotein(s)-portions including part of the sequences which follow:

[0067] OMP or gp110 proteins, including precursors:

[0068] 1 to 530

[0069] OMP or gp110 without precursor:

[0070] 34-530

[0071] Sequence carrying the TMP or gp41 protein:

[0072] 531-877, particularly

[0073] 680-700

[0074] well conserved stretches of OMP:

[0075] 37-130,

[0076] 211-289 and

[0077] 488-530

[0078] well conserved stretch found at the end of the OMP and the beginning of TMP:

[0079] 490-620.

[0080] Proteins containing or consisting of the "well conserved stretches" are of particular interest for the production of immunogenic compositions and (preferably in relation to the stretches of the env protein) of vaccine compositions against the LAV-1 viruses.

[0081] The invention concerns more particularly all the DNA fragments which have been more specifically referred to in the drawings and which correspond to open reading frames. It will be understood that one skilled in the art will be able to obtain them all, for instance by cleaving an entire DNA corresponding to the complete genome of LAV_{MAL}, such as by cleavage by a partial or complete digestion thereof with a suitable restriction enzyme and by the subsequent recovery of the relevant fragments. The DNA disclosed above can be resorted to also as a source of suitable fragments. The techniques disclosed in PCT application for the isolation of the fragments which can then be included in suitable plasmids are applicable here too. Of course, other methods can be used, some of which have been examplified in European Application No. 178,978, filed Sep. 17, 1985. Reference is for instance made to the following methods:

[0082] a) DNA can be transfected into mammalian cells with appropriate selection markers by a variety of techniques, such as calcium phosphate precipitation, polyethylene glycol, protoplast-fusion, etc.

[0083] b) DNA fragments corresponding to genes can be cloned into expression vectors for *E. coli*, yeast- or mammalian cells and the resultant proteins purified.

[0084] c) The provival DNA can be "shot-gunned" (fragmented) into procaryotic expression vectors to generate fusion polypeptides. Recombinants, producing antigenically competent fusion proteins, can be identified by simply screening the recombinants with antibodies against LAV antigens.

[0085] The invention further refers to DNA recombinants, particularly modified vectors, including any of the preceding DNA sequences adapted to transform corresponding microorganisms or cells, particularly eucaryotic cells such as yeasts, for instance *Saccharomyces cerevisiae*, or higher eucaryotic cells, particularly cells of mammals, and to permit expression of said DNA sequences in the corresponding microorganisms or cells. General methods of that type have been recalled in the abovesaid PCT international patent aplication PCT/EP 85/00548, filed Oct. 18, 1985.

[0086] More particularly the invention relates to such modified DNA recombinant vectors modified by the abovesaid DNA sequences and which are capable of transforming higher eucaryotic cells particularly mammalian cells. Preferably, any of the abovesaid sequences are placed under the direct control of a promoter contained in said vectors and recognized by the polymerases of said cells, such that the first nucleotide codons expressed correspond to the first triplets of the above-defined DNA sequences. Accordingly, this invention also relates to the corresponding DNA fragments which can be obtained from the genome of LAV_{MAL} or its cDNA by any appropriate method. For instance, such a method comprises cleaving said LAV_{MAL} genome or its cDNA by restriction enzymes preferably at the level of restriction sites surrounding said fragments and close to the opposite extremities respectively thereof, recovering and identifying the fragments sought according to sizes, if need be checking their restriction maps or nucleotide sequences (or by reaction with monoclonal antibodies specifically directed against epitopes carried by the polypeptides encoded by said DNA fragments), and further if need be, trimming the extremities of the fragment, for instance by an exonucleolytic enzyme such as Ba131, for the purpose of controlling the desired nucleotid-sequences of the extremities of said DNA fragments or, conversely, repairing said extremities with Klenow enzyme and possibly ligating the latter to synthetic polynucleotide fragments designed to permit the reconstitution of the nucleotide extremities of said fragments. Those fragments may then be inserted in any of said vectors for causing the expression of the corresponding polypeptide by the cell transformed therewith. The corresponding polypeptide can then be recovered from the transformed cells, if need be after lysis thereof, and purified by methods such as electrophoresis. Needless to say, all conventional methods for performing these operations can be resorted to.

[0087] The invention also relates more specifically to cloned probes which can be made starting from any DNA fragment according to this invention, thus to recombinant DNAs containing such fragments, particularly any plasmids amplifiable in procaryotic or eucaryotic cells and carrying said fragments. Using the cloned DNA fragments as a molecular hybridization probe—either by labelling with radionucleotides or with fluorescent reagents—LAV virion RNA may be detected directly in the blood, body fluids and blood products (e.g. of the antihemophylic factors such as Factor VIII concentrates) and vaccines (e.g., hepatitis B vaccine). It has already been shown that whole virus can be detected in culture supernatants of LAV producing cells. A suitable method for achieving that detection comprises immobilizing virus on a support (e.g., a nitrocellulose filter), disrupting the virion and hybridizing with labelled (radiolabelled or "cold" fluorescent- or enzyme-labelled) probes. Such an approach has already been developed for Hepatitis B virus in peripheral blood [SCOTTO J. et al. Hepatology (1983), 3, 379-384].

[0088] Probes according to the invention can also be used for rapid screening of genomic DNA derived from the tissue

of patients with LAV related symptoms, to see if the proviral DNA or RNA present in host tissue and other tissues can be related to that of LAV_{MAL}.

[0089] A method which can be used for such screening comprises the following steps: extraction of DNA from tissue, restriction enzyme cleavage of said DNA, electrophoresis of the fragments and Southern blotting of genomic DNA from tissues, subsequent hybridization with labelled cloned LAV proviral DNA. Hybridization in situ can also be used.

[0090] Lymphatic fluids and tissues and other non-lymphatic tissues of humans, primates and other mammalian species can also be screened to see if other evolutionnary related retrovirus exist. The methods referred to hereinabove can be used, although hybridization and washings would be done under non-stringent conditions.

[0091] The DNAs or DNA fragments according to the invention can be used also for achieving the expression of viral antigens of LAV_{MAL} for diagnostic purposes.

[0092] The invention relates generally to the polypeptides themselves, whether synthesized chemically, isolated from viral preparations or expressed by the different DNAs of the invention, particularly by the ORFs or fragments thereof in appropriate hosts, particularly procaryotic or eucaryotic hosts, after transformation thereof with a suitable vector previously modified by the corresponding DNAS.

[0093] More generally, the invention also relates to any of the polypeptide fragments (or molecules, particularly glycoproteins having the same polypeptidic backbone as the polypeptides mentioned hereinabove) bearing an epitope characteristic of a protein or glycoprotein of LAV_{MAL}, which polypeptide or molecule then has N-terminal and C-terminal extremities respectively either free or, independently from each other, covalently bonded to amino acids other than those which are normally associated with them in the larger polypeptides or glycoproteins of the LAV virus, which last mentioned amino acids are then free or belong to another polypeptidic sequence. Particularly, the invention relates to hybrid polypeptides containing any of the epitope-bearing-polypeptides which have been defined more specifically hereinabove, recombined with other polypeptides fragments normally foreign to the LAV proteins, having sizes sufficient to provide for an increased immunogenicity of the epitope-bearing-polypeptide yet, said foreign polypeptide fragments either being immunogenically inert or not interfering with the immunogenic properties of the epitope-bearing-polypeptide.

[0094] Such hybrid polypeptides, which may contain from 5 up to 150, even 250 amino acids, usually consist of the expression products of a vector which contained ab initio a nucleic acid sequence expressible under the control of a suitable promoter or replicon in a suitable host, which nucleic acid sequence had however beforehand been modified by insertion therein of a DNA sequence encoding said epitope-bearing-polypeptide.

[0095] Said epitope-bearing-polypeptides, particularly those whose N-terminal and C-terminal amino acids are free, are also accessible by chemical synthesis according to technics well known in the chemistry of proteins.

[0096] The synthesis of peptides in homogeneous solution and in solid phase is well known. In this respect, recourse may be had to the method of synthesis in homogeneous solution described by Houbenweyl in the work entitled "Methoden der Organischen Chemie" (Methods of Organic Chemistry) edited by E. WUNSCH., vol. 15-I and II, THIEME, Stuttgart 1974. This method of synthesis consists of successively condensing either the successive amino acids in twos, in the appropriate order or successive peptide fragments previously available or formed and containing already several amino-acyl residues in the appropriate order respectively. Except for the carboxyl and aminogroups which will be engaged in the formation of the peptide bonds, care must be taken to protect beforehand all other reactive groups borne by these amino-acyl groups or fragments. However, prior to the formation of the peptide bonds, the carboxyl groups are advantageously activated, according to methods well known in the synthesis of peptides. Alternatively, recourse may be had to coupling reactions bringing into play conventional coupling reagents, for instance of the carbodiimide type, such as 1-ethyl-3-(3-dimethyl-amino-propyl)-carbodiimide. When the amino acid group used carries an additional amine group (e.g., lysine) or another acid function (e.g., glutamic acid), these groups may be protected by carbobenzoxy or t-butyloxycarbonyl groups, as regards the amine groups, or by t-butylester groups, as regards the carboxylic groups. Similar procedures are available for the protection of other reactive groups. for example, an —SH group (e.g., in cysteine) can be protected by an acetamidomethyl or paramethoxybenzyl group.

[0097] In the case of a progressive synthesis, amino acid by amino acid, the synthesis starts preferably with the condensation of the C-terminal amino acid with the amino acid which corresponds to the neighboring aminoacyl group in the desired sequence and so on, step by step, up to the N-terminal amino acid. Another preferred technique which can be used is that described by R. D. Merrifield in "Solid Phase Peptide Synthesis"[J. Am. Chem. Soc., 45, 2149-2154]. In accordance with the Merrifield process, the first C-terminal amino acid of the chain is fixed to a suitable porous polymeric resin, by means of its carboxylic group, the amino group of the amino acid then being protected, for example by a t-butyloxycarbonyl group. When the first C-terminal amino acid is thus fixed to the resin, the protective group of the amine group is removed by washing the resin with an acid, i.e., trifluoroacetic acid, when the protective group of the amine group is a t-butyloxycarbonyl group. Then, the carboxylic group of the second amino acid, which is to provide the second amino-acyl group of the desired peptidic sequence, is coupled to the deprotected amine group of the C-terminal amino acid fixed to the resin. Preferably, the carboxyl group of this second amino acid has been activated, for example by dicyclohexyl-carbodiimide, while its amine group has been protected, for example by a t-butyloxycarbonyl group. The first part of the desired peptide chain, which comprises the first two amino acids, is thus obtained. As previously, the amine group is then deprotected, and one can further proceed with the fixing of the next amino-acyl group and so forth until the whole peptide sought is obtained. The protective groups of the different side groups, if any, of the peptide chain so formed can then be removed. The peptide sought can then be detached from the resin, for example by means of hydrofluoric acid, and finally recovered in pure form from the acid solution according to conventional procedures.

[0098] As regards the peptide sequences of smallest size bearing an epitope or immunogenic determinant, and more particularly those which are readily accessible by chemical synthesis, it may be required, in order to increase their in vivo immunogenic character, to couple or "conjugate" them covalently to a physiologically acceptable and non-toxic carrier molecule. By way of examples of carrier molecules or macromolecular supports which can be used for making the conjugates according to the invention can be mentioned natural proteins, such as tetanic toxoid, ovalbumin, serum-albumins, hemocyanins, etc. Synthetic macromolecular carriers, for example polysines or poly(D-L-alanine)-poly(L-lysine)s, can be used too. Other types of macromolecular carriers that can be used, which generally have molecular weights higher than 20,000, are known from the literature. The conjugates can be synthesized by known processes such as are described by Frantz and Robertson in "Infect. and Immunity", 33, 193-198 (1981) and by P. E. Kauffman in "Applied and Environmental Microbiology", October 1981 Vol. 42, No. 4, pp. 611-614. For instance, the following coupling agents can be used : glutaric aldehyde, ethyl chloroformate, water-soluble carbodiimides such as(N-ethyl-N'(3-dimethylamino-propyl) carbodiimide, HCl), diisocyanates, bis-diazobenzidine, di- and trichloro-s-triazines, cyanogen bromides and benzaquinone, as well as the coupling agents mentioned in "Scand. J. Immunol.", 1978, vol. 8, pp. 7-23 (Avrameas, Ternynck, Guesdon).

[0099] Any coupling process can be used for bonding one or several reactive groups of the peptide, on the one hand, and one or several reactive groups of the carrier, on the other hand. Again coupling is advantageously achieved between carboxyl and amine groups carried by the peptide and the carrier or vice-versa in the presence of a coupling agent of the type used in protein synthesis, e.g., 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide, N-hydroxybenzotriazole, etc. Coupling between amine groups respectively borne by the peptide and the carrier can also be made with glutaraldehyde, for instance, according to the method described by BOQUET, P. et al. (1982) Molec. Immunol., 19, 1441-1549, when the carrier is hemocyanin.

[0100] The immunogenicity of epitope-bearing-peptides can also be reinforced by oligomerisation thereof, for example in the presence of glutaraldehyde or any other suitable coupling agent. In particular, the invention relates to the water soluble immunogenic oligomers thus obtained, comprising particularly from 2 to 10 monomer units.

[0101] The glycoproteins, proteins and other polypeptides (generally designated hereinafter as "antigens" of this invention) whether obtained by methods, such as are disclosed in the earlier patent applications referred to above, in a purified state from LAV_{MAL} virus preparations or—as concerns more particularly the peptides—by chemical synthesis, are useful in processes for the detection of the presence of anti-LAV antibodies in biological media, particularly biological fluids such as sera from man or animal, particularly with a view of possibly diagnosing LAS or AIDS.

[0102] Particularly the invention relates to an in vitro process of diagnosis making use of an envelope glycoprotein or of a polypeptide bearing an epitope of this glycoprotein of LAV_{MAL} for the detection of anti-LAV antibodies in the serums of persons who carry them. Other polypeptides—particular those carrying an epitope of a core protein—can be used too.

[0103] A preferred embodiment of the process of the invention comprises:

[0104] depositing a predetermined amount of one or several of said antigens in the cups of a titration microplate;

[0105] introducing increasing dilutions of the biological fluid, to be diagnosed (e.g., blood serum, spinal fluid, lymphatic fluid, and cephalo-rachidian fluid), into these cups;

[0106] incubating the microplate;

[0107] washing carefully the microplate with an appropriate buffer;

[0108] adding into the cups specific labelled antibodies directed against blood immunoglobulins and

[0109] detecting the antigen-antibody-complex formed, which is then indicative of the presence of LAV antibodies in the biological fluid.

[0110] Advantageously the labelling of the anti-immunoglobulin antibodies is achieved by an enzyme selected from among those which are capable of hydrolysing a substrate, which substrate undergoes a modification of its radiation-absorption, at least within a predetermined band of wavelenghts. The detection of the substrate, preferably comparatively with respect to a control, then provides a measurement of the potential risks, or of the effective presence, of the disease.

[0111] Thus, preferred methods of immuno-enzymatic and also immunofluorescent detections, in particular according to the ELISA technique, are provided. Titrations may be determinations by immunofluorescence or direct or indirect immuno-enzymatic determinations. Quantitative titrations of antibodies on the serums studied can be made.

[0112] The invention also relates to the diagnostic kits themselves for the in vitro detection of antibodies against the LAV virus, which kits comprise any of the polypeptides identified herein and all the biological and chemical reagents, as well as equipment, necessary for peforming diagnostic assays. Preferred kits comprise all reagents required for carrying out ELISA assays. Thus preferred kits will include, in addition to any of said polypeptides, suitable buffers and anti-human immunoglobulins, which anti-human immunoglobulins are labelled either by an immunofluorescent molecule or by an enzyme. In the last instance, preferred kits also comprise a substrate hydrolysable by the enzyme and providing a signal, particularly modified absorption of a radiation, at least in a determined wavelength, which signal is then indicative of the presence of antibody in the biological fluid to be assayed with said kit.

[0113] It can of course be of advantage to use several proteins or polypeptides not only of LAV_{MAL}, but also of LAV_{ELI} together with homologous proteins or polypeptides of earlier described viruses, such as LAV_{BRU}, HTLV-3, ARV 2, etc.

[0114] The invention also relates to vaccine compositions whose active principle is to be constituted by any of the antigens, i.e., the hereinabove disclosed polypeptides of LAV_{MAL}, particularly the purified gp110 or immunogenic fragments thereof, fusion polypeptides or oligopeptides in association with a suitable pharmaceutically or physiologi-

cally acceptable carrier. A first type of preferred active principle is the gp110 immunogen of said immunogens. Other preferred active principles to be considered in that fields consist of the peptides containing less than 250 amino acid units, preferably less than 150, particularly from 5 to 150 amino acid residues, as deducible for the complete genome of LAV_{MAL} and even more preferably those peptides which contain one or more groups selected from Asn-X-Thr and Asn-X-Ser as defined above. Preferred peptides for use in the production of vaccinating principles are peptides (a) to (f) as defined above. By way of example, there may be mentioned that suitable dosages of the vaccine compositions are those which are effective to elicit antibodies in vivo, in the host, particularly a human host. Suitable doses range from 10 to 500 micrograms of polypeptide, protein or glycoprotein per kg, for instance 50 to 100 micrograms per kg.

[0115] The different peptides according to this invention can also be used themselves for the production of antibodies, preferably monoclonal antibodies specific for the respective different peptides. For the production of hybridomas secreting said monoclonal antibodies, conventional production and screening methods can be used. These monoclonal antibodies, which themselves are part of the invention, provide very useful tools for the identification and even determination of relative proportions of the different polypeptides or proteins in biological samples, particularly human samples containing LAV or related viruses.

[0116] The invention further relates to the hosts (procaryotic or eucaryotic cells) which are transformed by the above mentioned recombinants and which are capable of expressing said DNA fragments.

[0117] Finally the invention also concerns vectors for transforming eucaryotic cells of human origin, particularly lymphocytes, the polymerase of which are capable of recognizing the LTRs of LAV. Particularly said vectors are characterized by the presence of a LAV LTR therein, said LTR being then active as a promoter enabling the efficient transcription and translation in a suitable host of a DNA insert coding for a determined protein placed under its controls.

[0118] Needless to say, the invention extends to all variants of genomes and corresponding DNA fragments (ORFS) having substantially equivalent properties, all of said genomes belonging to retroviruses which can be considered as equivalents of LAV_{MAL}. It must be understood that the claims which follow are also intended to cover all equivalents of the products (glycoproteins, polypeptides, DNAS, etc.) whereby an equivalent is a product, e.g., a polypeptide, which may distinguish from a product defined in any of said claims, say through one or several amino acids, while still having substantially the same immunological or immunogenic properties. A similar rule of equivalency shall apply to the DNAs, it being understood that the rule of equivalency will then be tied to the rule of equivalency pertaining to the polypeptides which they encode.

[0119] It will also be understood that all the literature referred to hereinbefore and hereinafter and all patent applications and patents not specifically identified herein but which form counterparts of those specifically designated herein, must be considered as incorporated herein by reference.

[0120] It should further be mentioned that the invention further relates to immunogenic compositions that contain preferably one or more of the polypeptides, which are specifically identified above and which have the amino acid sequences of LAV_{MAL} that have been identified, or peptidic sequences corresponding to previously defined LAV proteins. In this respect, the invention relates more particularly to the particular polypeptides which have the sequences corresponding more specifically to the LAV_{BRU} sequences which have been referred to earlier, i.e., the sequences extending between the following first and last amino acids, of the LAV_{BRU} proteins themselves, i.e., the polypeptides having sequences contained in the LAV_{BRU} OMP or LAV_{BRU} TMP or sequences extending over both, particularly those extending from between the following positions of the amino acids included in the env open reading frame of the LAV_{BRU} genome,

[0121] 1-530

[0122] 34-530

[0123] and more preferably

[0124] 531-877, particularly 680-700,

[0125] 37-130

[0126] 211-289

[0127] 488-530

[0128] 490-620.

[0129] These different sequences can be used for any of the above defined purposes and in any of the compositions which have been disclosed.

[0130] Finally the invention also relates to the different antibodies which can be formed specifically against the different peptides which have been disclosed herein, particularly to the monoclonal antibodies which recognize them specifically. The corresponding hybridomas which can be formed starting from spleen cells previously immunized with such peptides which are fused with appropriate myeloma cells and selected according to standard procedures also form part of the invention.

[0131] Phage λ clone E-H12 derived from LAV_{ELI} infected cells has been deposited at the CNCM under No. I-550 on May 9, 1986. Phage clone M-H11 derived from LAV_{MAL} infected cells has been deposited at the CNCM under No. I-551 on May 9, 1986.

1. The virus LAV_{MAL} comprising RNA corresponding to the cDNA of FIGS. 7A-7I.

2. The CDNA of FIGS. 7A-7I.

3. A DNA recombinant comprising the cDNA of claim 2.

4. A probe containing a nucleic acid sequence hybridizable with RNA of said LAV_{MAL} virus of claim 1.

5. A method for identifying the presence in a host tissue of LAV virus which comprises hybridizing RNA obtained from said tissue with said probe of claim 4.

6. The method of claim 5, wherein said probe can hybridize with RNA from said LAV_{MAL} virus to identify said LAV_{MAL} virus.

7. A peptide or fragment thereof whose amino acid sequence is encoded by an open reading frame of a cDNA sequence of the LAV_{MAL} virus of claim 1.

8. The peptide of claim 7 encoded by a cDNA sequence from amino-acyl residue 37 to amino-acyl residue 130, or from amino-acyl residue 211 to amino-acyl residue 289, or from amino-acyl residue 488 to amino-acyl residue 530 of FIGS. **3**A-**3**F and **7**A-**7**I.

9. The peptide of claim 7 encoded by a cDNA sequence from amino-acyl residue 490 to amino-acyl residue 620 or from amino-acyl residue 680 to amino-acyl residue 700 of FIGS. **3**A-**3**F and **7**A-**7**I.

10. The peptide of claim 7 which comprises a protein or glycoprotein whose amino acid sequence is encoded by all or part of one of the following cDNA sequences of FIGS. **3**A-**3**F and **7**A-**7**I:

OMP or gp110 proteins, including precursors: 1 to 530;

OMP or gp110 without precursor: 34-530; and

TMP or gp41 protein: 531-877.

11. The peptide of claim 10 encoded by all or part of one of the following cDNA sequences of FIGS. **3**A-**3**F and **7**A-**7**I: 37-130, 211-289, 488-530, 490-620 or 680-700.

12. A method for the in vitro detection of the presence of an antibody directed against a LAV virus in a human body fluid, which comprises: contacting said body fluid with an antigen obtained from said virus LAV_{MAL} of claim 1, said antigen consisting of a peptide or a fragment thereof whose amino acid sequence is encoded by an open reading frame of a cDNA sequence of FIGS. **7**A-**7**I; and then detecting the immunological reaction between said antigen and said antibody.

13. The method of claim 12 wherein said antigen detects said LAV_{MAL} virus of claim 1.

14. The method of claim 12 which comprises the steps of:

a) depositing a predetermined amount of said antigen into a cup of a titration microplate;

b) introducing increasing dilutions of said body fluid into said cup;

c) incubating said microplate;

d) washing the microplate with a buffer;

e) adding into said cup a labelled antibody directed against blood immunoglobulins; and then

f) determining whether an antigen-antibody-complex has formed in said cup which is indicative of the presence of a LAV antibody in said body fluid.

15. A diagnostic kit for the in vitro detection of antibodies against a LAV virus, which kit comprises: an antigen consisting of a peptide of claim 7.

16. The kit of claim 15 wherein the antigen consists of a peptide of said LAV_{MAL} virus of claim 1, encoded by the open reading frame of a cDNA sequence of said LAV_{MAL} virus.

17. An immunogenic composition comprising: an antigen of the LAV_{MAL} virus of claim 1 or an immunogenic peptide or fragment thereof encoded by RNA of said virus; and a physiologically acceptable carrier.

18. The immunogenic composition of claim 17 wherein said peptide is the gp110 envelope glycoprotein or a fragment thereof.

19. The immunogenic composition of claim 17 wherein the peptide comprises a protein or glycoprotein hwose amino acid sequence is encoded by all or part of one of the following cDNA sequences of **FIGS. 3***a*-**3**F and **7**A-**7**I:

OMP or gp110 proteins, including precursors: 1 to 530;

OMP or gp110 without precursor: 34-530; and

TMP or gp41: 531-877.

20. The composition of claim 19 wherein the protein or glycoprotein is encoded by all or part of one of the following cDNA sequences of FIGS. **3**A-**3**F and **7**A-**7**I: 37-130, 211-289, 488-530, 490-620 or 680-700.

21. An antibody formed against a peptide of claim 7.

22. A cell transformed with a DNA recombinant of claim **3**.

* * * * *

专利名称(译)	变异或LAV病毒		
公开(公告)号	US20020022033A1	公开(公告)日	2002-02-21
申请号	US09/834627	申请日	2001-04-16
[标]申请(专利权)人(译)	巴斯德研究所		
申请(专利权)人(译)	巴斯德研究所		
当前申请(专利权)人(译)	巴斯德研究所		
[标]发明人	ALIZON MARC SONIGO PIERRE WAIN HOBSON SIMON MONTAGNIER LUC		
发明人	ALIZON, MARC SONIGO, PIERRE WAIN-HOBSON, SIMON MONTAGNIER, LUC		
IPC分类号	A61K38/00 A61K39/12 A61K39/21 C07H21/04 C07K5/00 C07K7/00 C07K16/00 C07K17/00 C12N1/00 C12N5/00 C12N5/02 C12N7/00 C12P21/08 C12Q1/68 C12Q1/70 G01N33/53 G01N33/537 G01N33/543		
CPC分类号	A61K38/00 C12N7/00 C12N2740/16021 C12N2740/16051 C12Q1/702 C12Q1/703 G01N33/56988		
优先权	1986401380 1986-06-23 EP		
其他公开文献	US6635752		
外部链接	Espacenet USPTO		

摘要(译)

LAV病毒的一种变体，命名为LAVMAL，能够引起艾滋病。 LAVMAL病毒的cDNA和抗原可用于诊断AIDS和艾滋病前期。

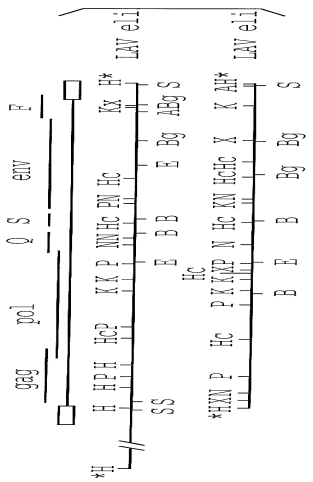


FIG. 1A