



(12) PATENTSKRIFT

Patent- og
Varemærkestyrelsen

- (51) Int.Cl.: C 07 K 7/06 C 07 K 14/155 C 12 N 15/48 G 01 N 33/536 G 01 N 33/569
(21) Patentansøgning nr: PA 1988 05133
(22) Indleveringsdag: 1988-09-15
(24) Løbedag: 1988-01-15
(41) Alm. tilgængelig: 1988-11-16
(45) Patentets meddelelse bkg. den: 2003-09-29
(86) International ansøgning nr: PCT/FR88/00025
(86) International indleveringsdag: 1988-01-15
(85) Videreførelsesdag: 1988-09-15
(30) Prioritet: 1987-01-16 US 003764 1987-02-11 FR 8701739 1987-04-15 FR 8705398

(73) Patenthaver: Institut Pasteur, 25-28 rue du Dr. Roux, 75724 Paris Cedex 15, Frankrig

(72) Opfinder: Luc Montagnier, 21, rue de Malabry, F-92350 Le-Plessis-Robinson, Frankrig
Denise Guetard, 4 B, rue Anselme-Payen, F-75015 Paris, Frankrig
Francois Clavel, 12103 Portree Drive, Rockville, MD 20852, USA
Pierre Sonigo, 23, rue Gutenberg, F-75015 Paris, Frankrig
Mireille Guyader, 68, rue Laugier, F-75017, Frankrig
Pierre Tiollais, 16, rue de la Glaciere, F-75013 Paris, Frankrig
Lisa Chakrabarti, 16, rue des 3 Portes, F-75005 Paris, Frankrig
Ronald Desrosiers, 13 Causeway Street, Udson, MA 01749, USA
Marc Alizon, 71, rue du Cardinal Lemoine, F-75005 Paris, Frankrig

(74) Fuldmægtig: Zacco Denmark A/S, Hans Bekkevolds Allé 7, 2900 Hellerup, Danmark

(54) Benævnelse: Peptider med immunologiske egenskaber svarende til HIV-2 virus, antigene og immunogene midler indeholdende sådanne peptider samt fremgangsmåde og kit til en vitro diagnose af HIV-2

(57) Sammendrag:

Peptider, der udviser immunologiske egenskaber fælles med peptidskelettet i virus af type HIV-2, især kappeglycoproteinet fra HIV-2 udviser endvidere en peptidstruktur fælles med peptidskelettet i SIV-virus, især kappeglycoproteinet. Peptiderne er anvendelige i diagnostiske midler og fremgangsmåder samt til fremstilling af vaccine.

Den foreliggende opfindelse angår peptider med immunologiske egenskaber, eventuelt immunogene egenskaber, der er fælles med antigener, der kan opnås i rensset form, ud fra virus, som er i stand til at fremprovokere lymphadenopatii, der kan udvikle sig til erhvervet immundefektsyndrom (AIDS) hos mennesker.

Opfindelsen angår nærmere bestemt hidtil ukendte peptider af den i krav 1 angivne art. Disse peptider kan genkendes af antistoffer, der induceres i mennesker af virus betegnet ved forkortelsen HIV (ifølge nomenklaturen defineret i NATURE). Peptiderne ifølge opfindelsen har immunogene egenskaber kan gøres immunogene in vivo, hvilken immunogenicitet kan manifestere sig ved en induktion in vivo af antistoffer, der genkender de antigener, som er karakteristiske for HIV-2 virus, samt (for så vidt angår visse af disse peptider) antigener karakteristiske for HIV-1. Opfindelsen angår endvidere et antigen middel af den i krav 23 angivne art eller et immunogent middel af den i krav 24 angivne art.

Opfindelsen angår også anvendelsen af disse peptider til diagnosticering in vitro af HIV-2-infektion i en biologisk væske.

Endelig angår opfindelsen et kit til in vitro diagnosticering af HIV-2-infektion i en biologisk væske.

Visse af de omhandlede peptider kan anvendes i metoder til diagnose in vivo af visse former for AIDS hos mennesker, og peptiderne kan endvidere anvendes i systemer eller "kits" til diagnose.

Et første retrovirus, som er benævnt LAV-1 eller HIV-1, er blevet isoleret og beskrevet i GB patentansøgning nr. 83/24.800 og EP patentansøgning nr. 84/401.834 indleveret den 14. september 1984. Dette virus er ligeledes blevet beskrevet af F. Barre Sinoussi et al i Science 220, 45-99, side 868-871.

Varianter af dette HIV-1 virus, som betegnes LAV, ELI og LAV MAL, er ligeledes blevet isoleret, karakteriseret og beskrevet i EP patentansøgning nr. 84/401.834.

- 5 HIV-1 virus og disses varianter besidder de følgende egenskaber:

Som foretrukne målceller har de humane Leu3 celler (eller T4-lymfocytter) og disses "immortaliserede" cellederivater.

- 10 De har en revers transcriptase-aktivitet, som nødvendiggør tilstedeværelsen af Mg^{2+} -ioner, og udviser en kraftig aktivitet overfor poly(adenylat-oligo-deoxythymidylase) poly(A)-oligo(dT)12-18).

De har en densitet på 1,16-1,17 bestemt i en saccharose-gradient.

15

De har en middeldiameter på 139 nm og en gennemsnitlig kernediameter på 41 nm.

- 20 Lysaterne af disse virus indeholder et protein benævnt p25 (kerneprotein), som ikke udviser immunologisk krydsning med proteinet p24 fra HTLV-1.

De indeholder et protein benævnt p42, som hører til deres kappe, og de indeholder desuden et kappe-glycoprotein benævnt gp110, som har en molekylvægt på 110 000.

25

- 30 Isoleringen og karakteriseringen af retrovirus, som hører til en bestemt klasse og som kun har et reduceret immunologisk slægtskab med sine ophav, er beskrevet i EP patentansøgning nr. 87/400.151.4. Disse retrovirus, som er blevet grupperet under betegnelsen HIV-2, er blevet isoleret i forbindelse med talrige afrikanske sygdomstilfælde, hvor der forekommer symptomer på en lymphadenopati eller et AIDS-tilfælde.

Retrovirus af typen HIV-2 udmærker sig, ligesom retrovirus af typen HIV-1, ved en tropisme overfor humane T4-lymfocytter og ved en cytopatogen effekt over for disse lymfocytter, når de formerer sig. Herved forårsages dels generaliserede og vedvarende polyadenopatier og dels AIDS.

5

Nærmere bestemt kan det siges, at retrovirus oprenset fra HIV-2 generelt besidder følgende egenskaber:

de foretrukne målceller for HIV-2 retrovirus er humane Leu3-celler (eller
10 T4-lymfocytter) og immortaliserede celler afledt af disse T4-lymfocytter;

de er cytotoxiske overfor humane T4-lymfocytter, de har en revers transcriptaseaktivitet, der nødvendiggør tilstedeværelsen af Mg^{2+} -ioner og udviser en kraftig aktivitet overfor poly(adenylat-oligodeoxythymidylase)
15 (poly(A)-oligo(dT)12-18);

de har en densitet på 1,16 bestemt i en saccharose-gradient;

de har en middeldiameter på 140 nm og en kerne med en middeldiameter
20 på 41 nm;

de kan dyrkes i permanente stammer af typen HUT eller som udtrykker T4-proteinet;

25 de inficerer ikke T8-lymfocytter;

lysater af disse virus indeholder et protein benævnt p26, der ikke udviser immunologisk overkrydsning med protein p24 fra HTLV-I virus eller HTLV-II virus;

30

disse lysater indeholder endvidere et protein benævnt p16, der ikke genkendes immunologisk af proteinet p19 fra HTLV-I eller HTLV-II virus i radioimmunopræcipitationsassay;

de indeholder endvidere et kappeglycoprotein med molekylvægt af størrelsesordenen 130.000-140.000, der ikke udviser immunologisk overkrydsning med gp110 fra HIV-1, men som til gengæld krydser immunologisk med kappeglycoproteinet gp140 fra STLV-III (virus isoleret fra aber);

disse lysater indeholder endvidere antigener, der kan mærkes med ³⁵S-cystein, hvis molekylvægt ligger mellem 32 000 og 42 000-45 000: de omfatter endvidere et antigen med molekylvægt af størrelsesordenen 36 000 og et antigen med en molekylvægt af størrelsesordenen 42 000, idet et af disse antigener (benævnt p36 og p42) sandsynligvis udgør et transmembran-glycoprotein fra HIV-2 virus,

det genomiske RNA i HIV-2 hybridiserer ikke med genomisk RNA fra HIV-1 under stringente betingelser;

under ikke-stringente betingelser hybridiserer det genomiske RNA fra HIV-2 hverken med env-genet eller med LTR, som støder op til genet, i HIV-1, eller med sekvenser fra pol-regionen i HIV-1 genomet;

under ikke-stringente betingelser hybridiserer det svagt med nucleotidsekvenserne i HIV-1 regionen.

Et andet retrovirus benævnt SIV-1, hvilken betegnelse erstatter den tidligere anvendte betegnelse STLV-III, er isoleret hos rhesusaber (M.D. Daniel et al. Science 228, 1201 (1985) N.L. Letwin et al, Science 230, 71 (1985) under betegnelsen "STLV-III_{mac}").

Et yderligere retrovirus benævnt "STLV-III_{AGM}" (eller SIV_{AGM}) er isoleret hos vilde grønne aber. Men i modsætning til de virus, der er til stede hos rhesusaben, synes tilstedeværelsen af "STLV-III_{AGM}" ikke at fremkalde sygdomme af typen AIDS hos afrikanske grønne aber.

En stamme af retrovirus SIV-1mac er deponeret ved CNCM den 7. februar 1986 under nr. I-521. Studier har vist, at retrovirus SIV-1 indeholder visse proteiner, der udviser et vist immunologisk slægtskab med strukturproteiner eller -glycoproteiner, der kan opnås under analoge betingelser ud fra HIV-2. Denne retrovirus SIV-1, hvis infektiøse karakter man har konstateret hos aber, blev betegnet STLVIII af de forskere, der isolerede det (jævnfør de bibliografiske referencer ovenfor).

Af sproglige årsager bliver disse virus i det følgende kun betegnet ved udtrykket SIV (udtrykket SIV er den engelske forkortelse for "Simian Immunodeficiency Virus" (immunodefektvirus hos aber) eventuelt efterfulgt af en forkortelse, der angiver den abeart, hvorfra de er opnået, f.eks. MAC eller mac for rhesusaber eller AGM for afrikanske grønne aber (forkortelse for African Green Monkey).

Ved at anvende de samme teknikker, som beskrevet ovenfor, har man konstateret, at man også ud fra SIV-1mac kan opnå:

et principalt kerneprotein benævnt p27, der har en molekylvægt af størrelsesordenen 27 kd,

et større kappeglycoprotein, benævnt gp140,

et protein benævnt p32, der sandsynligvis er et transmembrant protein, der sjældent observeres ved radioimmunopræcipitationsassay (RIPA), når virus forinden er mærket med ³⁵S-cystein, men som kan observeres ved immunoaftryksprøver (Western blots) i form af brede bånd.

Mere præcise undersøgelser er gennemført på de ovennævnte virus HIV-2 og SIV virus. Fortsatte studier af HIV-2 retrovirus har ligeledes ført til opnåelse af komplementære DNA-sekvenser (cDNA) til RNA i disses genomer. Den komplette nucleotidsekvens for cDNA for et repræsentativt retrovirus af klassen HIV-2 (HIV-2-ROD) er deponeret den 21. februar

1986 ved CNCM med nummeret I- 522 under referencebetegnelsen LAV-II ROD.

5 Denne nucleotidsekvens og de åbne læserammer som de indeholder er angivet på figur 1 A.

I øvrigt har fortsatte studier af andre retrovirus ligeledes muliggjort en kortlæggelse af disses fuldstændige nucleotidsekvenser. Dette gælder især for cDNA afledt af genomisk RNA fra SIV-virus.

10 Kloning og sekvensbestemmelse af SIV-1 mac virus har muliggjort en kortlæggelse af dets nucleotidsekvens og gennemførtes under følgende betingelser:

15 DNA fra HUT 78 celler inficeret med SIV-virus (isolat STLV-III mac 142-83 beskrevet af Daniel et al (1985) Science, 228, side 1201-1204), og partielt nedbrudt med restriktionsenzymet Sau3A blev klonet i BamHI-restriktionsstedet i bakteriophag λ -vektoren ELBL3 til opnåelse af et genomisk bibliotek. De 2 millioner rekombinante phager i det således sammensatte genomiske bibliotek blev undersøgt in situ under sikkerhedsbestemmelser P3 ved hjælp af sekvenser af HIV-2 virus stammende fra klonerne lambda-ROD4, lambda-ROD35 og E2 (Clevel et al (1986) Nature, 324, side 691) og blev nick-translateret.

25 Hybridiseringen gennemførtes ved 5xSSC ved 50 °C, og udvaskningerne ved 2xSSC ved 50 °C. En enkelt klon indeholdende samtlige virussekvenser opnåedes. Denne klon benævntes lambda-SIV-1. Indsætningen i phag lambda-SIV-1 måler 16,5 kb i alt og omfatter en integreret provirus, hvor der kun mangler de 250 første baser fra det venstrestillede LTR, mens det
30 højrestillede LTR er komplet.

Det integrerede provirus blev sekvensbestemt ved dideoxynucleotidmetoden efter subkloning af tilfældige fragmenter i phagen M13mp8. Der analyseredes 300 subkloner.

- 5 cDNA-fragmenter stammende fra klonen lambda-SIV-1 indsat i plasmidene pSIV1.1 og pSIV-1.2 blev deponeret ved CNCM den 15. april 1987 under numrene I-658 (pSIV-1.1) og I-659 (pSIV-1.2).

Resultaterne er angivet i de nedenfor beskrevne figurer.

10

Fig. 1B betegner nucleotidsekvensen af genomet fra SIV virus og de sekvenser, der er afledt deraf for de virale proteiner svarende til produkterne af generne gag, pol, env, Q, X, R, tat, art, F.

- 15 Fig. 3-11 og fig. 1C viser sammenligning af de teoretiske produkter af de virale gener og af LTR mellem HIV-2 og SIVmac (SIV-1).

- 20 Nucleinsyresekvenserne for cDNA fra SIV er anbragt over for nucleinsyresekvenserne i HIV-2 ROD virus for så vidt angår LTR-sekvensen (jvf. fig. 1C). Denne præsentation som man genkender fra det fuldstændige genom ved at sammenligne fig. 1B med fig. 3-11 muliggør at man kan finde frem til eller aflede de nucleinsyrer, der har essentielle strukturelementer fælles med de to virus.

- 25 Komparative studier, som ligeledes har muliggjort, at man opnår resultater for kerneproteiner, i det følgende benævnt gag-proteiner, samt kappeproteiner, i det følgende benævnt env-proteiner, er ligeledes beskrevet i den ovennævnte EP ansøgning nr. 87/400.151.4. Disse resultater viser, at kerneproteiner (gag-proteiner) i HIV-2 udviser mindre tydelige forskelle fra
30 celler fra virus HIV-1 end kappeproteinerne (env-proteinerne). Totalt set har env-proteinerne i HIV-2 vist sig at udvise ekstremt svage om ikke ikke-eksisterende immunologiske ligheder med de korresponderende env-proteiner i HIV-1 virus.

Derimod har komparative studier mellem strukturerne af cDNA-sekvenserne i HIV-2 og SIV-virus muliggjort, at man kan fastlægge visse fælles karakteristika, som viser sig på proteinniveauet.

5

Totalt set udviser proteinerne fra HIV-2 og SIV-1 vigtige immunologiske ligheder.

10 Det væsentligste kappeprotein fra HIV-2 har vist sig at være nærmere beslægtet med det væsentligste kappeglycoprotein fra SIV i immunologisk henseende end med det væsentligste kappeglycoprotein fra HIV-1.

Disse konstateringer viser sig ikke blot hvad angår molekylvægtene: 130-140 kd for de væsentligste glycoproteiner fra HIV-2 og SIV overfor ca. 110
15 kd for det væsentligste kappeglycoprotein fra HIV-1, men også hvad angår de immunologiske egenskaber, eftersom udtagne serumprøver fra patienter inficeret med HIV-2 og især antistoffer dannet mod gp140 fra HIV-2 genkender gp140 fra SIV-1mac, mens de samme serumprøver og de samme HIV-2-antistoffer ved tilsvarende forsøg ikke genkender gp110 fra
20 HIV-1. Men anti-HIV-1 serum, som aldrig har reageret med gp140 fra HIV-2, udfælder et protein på 26 kd mærket med ³⁵S-cystein, der er indeholdt i HIV-2 ekstrakter.

25 Det væsentligste kerneprotein fra HIV-2 synes at udvise en middelmolekylvægt på ca. 26 000, som ligger imellem molekylvægten for p25 fra HIV-1 og p27 fra SIV.

Disse observationer stammer fra forsøg gennemført på virusekstrakter opnået ud fra HIV-2 isoleret fra en af de ovennævnte patienter. Lignende
30 resultater er opnået med virusekstrakter af HIV-2 isoleret fra en anden patient.

Mere omfattende forsøg har fået opfinderne til at erkende en første klasse af peptider, der har aminosyresekvenser, der enten er identiske eller nært beslægtede med sekvenser indeholdt i det indre af strukturerne for gag-proteinerne for HIV-2 eller fra SIV, og endog fra HIV-1. Disse peptider er særligt anvendelige til diagnose af en infektion hos mennesker med HIV-2 virus eller en variant deraf.

På denne baggrund kan fremgangsmåden og midlet ifølge opfindelsen anvendes til diagnose in vitro af antistoffer rettet mod HIV-2 virus eller varianter deraf, især i biologiske prøver, navnlig serumprøver fra patienter, der har været udsat for en infektion med HIV-2 virus, idet visse af disse peptider muliggør en særlig tydelig diskrimination mellem infektioner forårsaget af HIV-2 virus og HIV-1 virus.

Disse omfattende studier har ligeledes ført til muligheden for at syntetisere peptider, der er immunogene eller kan gøres immunogene, og som udviser strukturelle karakteristika, der muliggør at de in vivo kan inducere dannelsen af antistoffer, der er i stand til at genkende env-proteiner både fra HIV-1 og HIV-2, og i det mindste for en del af disse peptiders vedkommende at binde sig både til HIV-1 virus og til HIV-2 virus, specielt med det formål at neutralisere disse. Anvendelsen af disse typer peptider er således særligt indiceret til fremstilling af aktive bestanddele af vacciner mod HIV-virus, og dermed imod AIDS.

For nedenfor at beskrive de aminosyrerester, der indgår i strukturen af de omhandlede peptider, anvender man for aminosyren den entydige internationale nomenklatur, der betegner hver naturlig aminosyre med et stort bogstav ifølge den nedenstående tabel:

M Methionin
L Leucin
I Isoleucin
V Valin

	F	Phenylalanin
	S	Serin
	P	Prolin
	T	Threonin
5	A	Alanin
	Y	Tyrosin
	H	Histidin
	Q	Glutamin
10	N	Asparagin
	K	Lysin
	D	Asparaginsyre
	E	Glutaminsyre
	C	Cystein
15	W	Tryptophan
	R	Arginin
	G	Glycin

20 Når en aminosyre p.g.a. sin position midt i aminosyrekæden, der er karakteristisk for et givet peptid, kan have flere betydninger, kan den enten være betegnet med en bindestreg "-", hvis dens betydning kan være vilkårlig, eller ved et lille bogstav, når denne aminosyre kan udvise et antal foretrukne betydninger, og dette antal imidlertid altid er større end 1. I sidstnævnte tilfælde er de mulige betydninger af det lille bogstav altid angivet i 25 forbindelse med det peptid, som det tilhører.

For at lette læsningen bliver disse peptider også betegnet med en forkortelse env efterfulgt af et tal, ved reference til de indeholdte aminosyrer efter omstændighederne, i env-proteinerne i visse typer af HIV-1, HIV-2 eller 30 SIV. Dette omtales nærmere nedenfor.

I de efterfølgende definitioner betegner grupperne X enten en fri NH₂-gruppe eller en amideret gruppe især med en eller to alkylgrupper inde-

holdende 1-5 carbonatomer, eller en peptidgruppe, der indeholder 1-5 aminosyrer, hvor den N-terminale aminosyre selv udviser en fri NH₂-gruppe eller amidgruppe, som angivet tidligere, og grupperne Z betegner enten en fri -OH-gruppe eller alkoxygruppe og indeholder så en alkylgruppe omfattende 1-5 carbonatomer, eller en peptidgruppe omfattende 1-5 aminosyrer, hvis C-terminale aminosyre selv udviser en fri -OH-gruppe eller alkoxygruppe, som angivet ovenfor, idet de grupper på 1-5 aminosyrer, der eventuelt indgår i X eller Z eller i begge er sådanne, hvis tilstedeværelse ikke er inkompatibel med bevarelsen i hovedsagen af de immunologiske, eventuelt immunogene egenskaber ved de peptider, som de er fjernet fra.

Peptiderne ifølge opfindelsen, som har immunologiske egenskaber fælles med antigener fra HIV-2 og for visse af dem endvidere med antigener fra HIV-1 eller varianter deraf, er kendetegnet ved, at de ligeledes har en peptidstruktur fælles med antigener fra SIV. Med fordel omfatter disse peptider normalt højst 40 aminosyrerester.

Foretrukne peptider er de følgende:

20

env1

XRV-AIEKYL-DQA-LN-WGCAFRQVCZ

env2

25 X-LE-AQI-QQEKMYELQKLNZ

env3

XELGDYKLVEITPIG-APT--KR-----Z

30

env4

X----VTV-YGVP-WK-AT--LFCA-Z

env5

X---QE--L-NVTE-F--W-NZ

env6

5 XL---S-KPCVKLTPLCV--Z

env7

X---N-S-IT--C-K----Z

10 env8

X-I---YC-P-G-A-L-C-N-TZ

env9

X-----A-C-----W--Z

15

env10

X-G-DPE-----NC-GEF-YCN-----NZ

env11

20 X-----C-IKQ-I-----G---YZ

Nærmere bestemt angår opfindelsen følgende peptider:

env1

25 XRV-AIEKYL-DQA-LN-WGCAFRQVCZ

env2

X-LE-AQIQQEKNMYELQKLNZ

30 env3

XELGDYKLVEITPIG-APT--KR-----Z

env4

13

X---VTV-YGVP-W--AT--LFCA-Z

env5

X---E--L-NVTE-F--W-NZ

5

env6

XL---S-KPCVKL-PLC---Z

env7

10 X--N-S-I--C-K---Z

env8

X-I---YC-P-G-A-L-C-N-TZ

15

env9

X----A-C-----W-Z

env10

Z-G-DPE-----NC-GEF-YC-----NZ

20

env11

X----C-I-Q-I-----G---YZ

25 Særligt fordelagtige peptider svarende til de foregående udviser de følgende formler:

env1

XRVT AIEKYLQDQARLNSWGCAFRQVCZ, eller

XRVT AIEKYLKDQAQLNAWGCAFRQVCZ

30

env2

XSLEQAQIQQEKNMYELQKLNSWZ, eller

XLLEEAQIQQEKNMYELQKLNSWZ

env3

XELGDYKLVEITPIGFAPTKEKRYSSAHZ, eller

5 XELGDYKLVEITPIGLAPTNVKRYTTG-Z

Man bemærker, at peptiderne env1, env2, env3 bekræfter den meget store lighed mellem HIV-2 og SIV-1. Faktisk er det første peptid inkluderet i HIV-2 genomet og det andet i SIV-1 genomet.

10

env4

XabsdVTVeYGVpFWogAThiLFCajZ,

hvor bogstaverne a-j og o kan have følgende betydninger:

15

a er C, E eller D

b er T, K, D, N eller I

c er Q eller L

d er Y eller W

20

e er F eller Y

f er T, V eller A

g er N eller E

h er I eller T

i er P eller T

25

j er T eller S

o er K eller R

env5

XabcoEdeLfnVTEgFhiWjNZ,

30

hvor bogstaverne a-j og o kan have de følgende betydninger:

a er D eller P

15

b er D eller N

c er Y eller P

d er I, V, I eller L

5 e er T, V, E eller A

f er V, G eller E eller -

g er A, N, G eller S

h er D eller N

i er A eller M

10 j er N, K eller E

o er Q eller S

env6

XLabcSdKPCVKLoPLCuefKZ,

15

hvori bogstaverne a-f, o og u kan have følgende betydninger:

a er F eller W

b er E eller D

20 c er T eller Q

d er I eller L

e er A, S eller T

f er M eller L

o er T eller S

25 u er V eller I

env7

XabCNxSylocdCeKfghiZ,

30 hvori bogstaverne a-i og x og y kan have følgende betydninger:

a er N eller T eller I

b er H eller S eller N

c er E eller Q
d er S, A eller C

e er D eller P
5 f er H, V eller D
g er Y eller S
h er W eller F
i er D eller E
X er T eller R
10 y er V eller A
o er T eller Q

env8

XalbcdYCxPeGfAgLhCiNjTZ,
15

hvori bogstaverne a-k og x kan have følgende betydninger:

a er A eller P
b er R eller P
20 c er F, I eller C
d er R eller H
e er P eller A
f er Y eller F
g er L eller I
25 h er R eller K
i er - eller N
j er D eller K
x er A eller T

30 env9

XwabcxyAdCefghizWjkZ,

hvori bogstaverne a-k og x-z kan have følgende betydninger:

- a er K eller - eller E
 b er R eller -
 5 c er P eller M eller I
 d er W eller H eller Y
 e er W eller N eller T eller R
 f er F eller I
 g er K eller S eller N eller G
 10 h er G eller R eller E
 i er - eller A eller T
 j er K eller N eller D eller S
 k er D eller A eller N eller K eller E
 w er N, D eller I
 15 x er R eller G eller K
 y er Q eller K eller R
 z er K eller E eller Q eller N

env10

- 20 XaGbDPEcdefghNCiGEFjYCokxlmnNZ,

hvor i bogstaverne a-n og x kan have følgende betydninger:

- a er K eller - eller G
 25 b er S eller G eller -
 c er V eller I
 d er A eller V eller T
 e er Y eller T eller M eller F
 f er M eller H
 30 g er W eller S
 h er T eller F
 i er R eller G
 j er L eller F

o er N eller K

k er M eller S

l er W eller Q eller K eller G

5 m er F eller L

n er L eller F

x er T eller S eller N

env11

10 XabcdwCeloQflxgyhizGjklYZ,

hvor i bogstaverne a-l og w-z kan have følgende betydninger:

a er R eller T eller S eller N

15 b er N eller I

c er Y eller T

d er A eller L eller V

e er H eller R

f er I eller F

20 g er T eller M

h er H eller Q eller A

i er K eller E

j er R eller K

k er N eller A

25 l er V eller M

w er P eller Q

x er N eller K

y er W eller V

z er V eller T eller K

30 o er K eller R

Det skal generelt bemærkes, at de aminosyrer, der har en entydig betydning (og som således er repræsenteret af et stort bogstav svarende det

internationale nomenklatur), og som indgår i de definitioner for peptiderne ifølge opfindelsen, der er angivet ovenfor, viser sig at korrespondere med identiske aminosyrer placeret i samme rækkefølge i de korresponderende sekvenser i mindst en HIV- eller SIV-1 virus.

5

Positionerne af disse sekvenser er understreget og fordelt inde i aminosyresekvenserne i env-proteinerne for henholdsvis HIV-2 ROD (CNCM nr. I-532) og HIV-1 BRU (CNCM nr. I-232) vist på fig. 2. Rækkefølgen af aminosyrerne i env- og gag-proteinerne for henholdsvis SIV-1mac (CNCM nr.

10 I-521) og af HIV-2 ROD vist på fig. 3 og fig. 4.

De ubrudte linier, der er vist på visse steder af disse sekvenser, skal understrege, at visse aminosyrer indeholdt i disse sekvenser er frivilligt udeladt på præsentationen for at muliggøre, at identiske aminosyrer bringes på linie (markeret med en stjerne), eller at to vertikale punkter på en og samme vertikale linie i de korresponderende proteinsekvenser fra på den ene side HIV-1 og HIV-2 og på den anden side SIV og HIV-2.

15

Bortset fra de ovennævnte peptider angår opfindelsen ligeledes peptider, der er modificeret ved indsætning og/eller udeladelse og/eller substitution af en eller flere aminosyrer, så længe de antigene eller immunogene egenskaber af disse peptider ikke modificeres, eller at genkendelsesegenskaberne for antigenet eller antistoffet hos disse peptider ikke er væsentligt modificeret.

20 25

I en særligt foretrukken udførelsesform angår opfindelsen peptider, der har immunologiske egenskaber fælles med peptidskelettet i kappeglycoproteinet fra virus af klasse HIV-2, idet disse peptider indeholder et antal aminosyrerester, der ikke overstiger 40.

30

Disse foretrukne peptider ifølge opfindelsen udviser følgende sekvenser:

env1

20

RVTAIEKYLQDQARLNSWGCAFRQVC
AIEKYLQDQ
RVSAIEKYLKDQAQLNAWGCAFRQVC
AIEKYLKDQ

5

env2
SLEQAQIQQEKNMYELQKLNWS
QIQQEKN
LLEEAQIQQEKNMYELQKLNWS

10

env3
ELGDYKLVEITPIGFAPTKEKRYSSAH
YKLVEITPIGFAPTKEK
ELGDYKLVEITPIGLAPTNVKRYTTG-
YKLVEITPIGLAPTNVK

15

env4
CTQYVTVFYGVPTWKNATIPLFCAT
VTVFYGVPTWKNAT
CIQYVTVFYGVPAWRNATIPLFCAT
VTVFYGVPAWRNAT

20

EKLWVTVYYGVPVWKEATTTLFCAS
VTVYYGVPVWKEAT
EDLWVTVYYGVPVWKEATTTLFCAS
VTVYYGVPVWKEAT
DNLWVTVYYGVPVWKEATTTLFCAS
VTVYYGVPVWKEAT

25

30

env5
DDYQEITL-NVTEAFDAWNN
L-NVTEAF

21

DDYSELAL-NVTESFDAWEN
L-NVTESF
PNPQEVVLVNVTENFNMWKN
LVNVTENF
5 PNPQEIELENVTEGFNMWKN
LENVTEGF
PNPQEIALENVTENFNMWKN
LENVTENF

10 env6
ETSIKPCVKLTPLCVAMK
ETSIKPCVKLSPLCITMR
DQSLKPCVKLTPLCVSLK
DQSLKPCVKLTPLCVTLN
15 PCVKLTPLCV

env7
NHCNTSVITESCD
NTSVIT
20 NHCNTSVIQECCD
NTSVIQ
TSCNTSVITQACP
NTSVIT
INCNTSVITQACP
25 NTSVIT
INCNTSAITQACP
NTSAIT

env8
30 YCAPPGYALLRC-NDT
YCAPAGFAILKCNNKT
YCAPAGFAILKCNDKK
YCAPAGFAILKCRDCK

env9

NKRPRQAWCWFKG-KWKD
NERPKQAWCRFGG-NWKE
5 N--MRQAHCNISRAKWNA
D--IRRAYCTINETEWDK
I--IGQAHCNISRAQWSK

env10

10 KGSDPEVAYMWTNCRGEFLYCNMTWFLN
NCRGEFLYCN
GG-DPEVTFMWTNCRGEFLYCKMNWFLN
NCRGEFLYCK
-GGDPEIVTHSFNCGGEFFYCNSTQLFN
15 NCGGEFFYCN
-GGDPEITTHSFNCRGEFFYCNSTKLFN
NCRGEFFYCN
-GGDPEITTHSFNCGGEFFYCNSTGLFN
NCGGEFFYCN

20

env11

RNYAPCHIKQIINTWHKVGRNVY
CHIKQII
RNYVPCHIRQIINTWHKVGKNVY
25 CHIRQII

TITLPCRQKQFINMWQEVGKAMY
CRIKQFI
SITLPCRQKQIINMWQKTCKAMY
30 CRIKQII
NITLQCRQKQIIMVAGR-KAIY
CRIKQII

Peptiderne ifølge opfindelsen kan med fordel fremstilles ved de klassiske metoder inden for peptidsyntese. Denne syntese kan gennemføres i homogen opløsning eller i fast fase.

- 5 F.eks. kan man anvende synteseteknikken i homogen opløsning beskrevet af HOUBEN-WEYL i "Methode der Organischen Chemie", Ed. E. Wunsch, vol. 15-I og II., THIEME, Stuttgart 1974.

10 Ved denne metode kondenserer man successivt to og to de successive aminoacylgrupper i den krævede rækkefølge eller man kondenserer aminoacylgrupper og foruddannede fragmenter, der allerede indeholder flere aminoacylgrupper i den rigtige rækkefølge, eller flere således tidligere fremstillede fragmenter, idet det er underforstået, at man forinden har taget højde for at beskytte alle de reaktive grupper på disse aminoacylgrupper eller fragmenter, bortset fra de aminogru-
15 per eller fragmenter, bortset fra de aminogru- og de carboxylgrupper i den anden eller vice-versa, som normalt skal indgå i dannelsen af peptidbindinger, især efter aktivering af carboxylgruppen ved velkendte metoder inden for peptidsyntese. Som variant kan man udnytte koblingsreaktioner, hvor der gøres brug af klassiske koblingsmidler af carbodiimidtypen, som f.eks. 1-ethyl-3-(3-dimethyl-amino-propyl)-carbodiimid.
20 Når den anvendte aminoacylgruppe udviser en supplerende syrefunktion, især når der er tale om glutaminsyre, kan disse være grupper være beskyttet, f.eks. med t-butylester-grupper.

25 Ved progressive synteser aminosyre efter aminosyre begynder syntesen fortrinsvis med en kondensation af den C-terminale aminosyre med den aminosyre, der svarer til det nabostillede aminoacyl i den ønskede sekvens og så fremdeles indtil den N-terminale aminosyre. Ved anden foretrukken teknik ifølge opfindelsen anvender man metoden beskrevet af
30 R.D. Merrifield, "Solid phase peptide synthesis" (J. Am. Soc., 45, 2149-2154).

For at fremstille en peptidkæde efter Merrifield-metoden anvender man en meget porøs polymer resin, hvorpå man fikserer den første C-terminale aminosyre i kæden. Denne aminosyre fikseres på resinets ved hjælp af sin carboxylgruppe, mens dens aminogruppe er beskyttet, f.eks. med en t-butyloxycarbonylgruppe.

Når den første C-terminale aminosyre således er fikseret på resin, fjerner man beskyttelsesgruppen på aminogruppen og ved at vaske resin med en syre.

Når beskyttelsesgruppen for aminogruppen er t-butyloxycarbonyl, kan den fjernes ved behandling af resin med trifluoreddikesyre.

Man kobler derpå den anden aminosyre, der leverer det andet aminoacyl i den tilstræbte sekvens ud fra den C-terminale aminoacyl, til den debeskyttede aminogruppe i den første C-terminale aminosyre fikseret til resin.

Fortrinsvis er carboxylgruppen i den anden aminosyre aktiveret, f.eks. med dicyclohexylcarbodiimid, og aminogruppen er beskyttet, f.eks. med t-butyloxycarbonyl.

Man opnår således den første del af den tilstræbte peptidkæde, der omfatter to aminosyrer, og hvis terminale aminogruppe er beskyttet. Som tidligere debeskytter man aminogruppen og man kan derpå fortsætte med fiksering af det tredje aminoacyl under analoge betingelser med dem, der anvendes ved additionen af den anden C-terminale aminosyre.

Man fikserer således den ene efter den anden af de aminosyrer, der skal udgøre peptidkæden, til aminogruppen, der hver gang forinden er debeskyttet, til den allerede dannede del af peptidkæden, som er bundet til resin.

Når hele den ønskede peptidkæde er dannet, fjerner man beskyttelsesgrupperne på de forskellige aminosyrer, der udgør peptidkæden, og man frigør peptidet fra resinen, f.eks. ved hjælp af flussyre.

- 5 Opfindelsen angår ligeledes vandopløselige oligomere af de ovennævnte monomere peptider. Oligomeriseringen kan fremkalde en forøgelse af immunogeniciteten af de monomere peptider ifølge opfindelsen. Uden at denne talangivelse skal opfattes som begrænsende, kan man ikke desto mindre nævne, at disse oligomere f.eks. kan indeholde 2-10 monomerenheder.

Monomerenhederne, der indgår i denne oligomer, udgøres enten alle af polypeptidet fra sekvens 1 eller af polypeptidet fra sekvens 2 eller af det ene eller det andet af disse polypeptider.

- 15 Man kan også til gennemførelse af oligomeriseringen anvende enhver sædvanlig polymerisationsteknik inden for peptidkemien, idet denne polymerisation gennemføres, indtil man opnår en oligomer eller polymer, der indeholder det krævede antal monomere enheder til opnåelse af den ønskede immunogenicitet.

En metode til oligomerisation eller polymerisation af den monomere består i at omsætte denne med et tværbindingsmiddel, såsom glutaraldehyd.

- 25 Man kan også anvende andre oligomerisations- eller koblingsmetoder, f.eks. den metode, der gør brug af successive koblinger af monomerenhederne ved hjælp af deres terminale carboxyl- og aminogrupeer i nærværelse af bifunktionelle homo- eller heterokoblingsmidler.
- 30 Man kan ligeledes til fremstilling af molekyler omfattende en eller flere enheder af de ovennævnte 17 aminosyrer anvende de genteknologiske metoder, der gør brug af mikroorganismer, der er transformeret med en given

nucleinsyre, og som omfatter de passende korresponderende nucleotidsekvenser.

I forbindelse med opfindelsen er der tilvejebragt nucleinsyrer, der indeholder en eller flere sekvenser afledt af cDNA-sekvensen for HIV-2 ROD-virus. Disse sekvenser, der er markeret med den nummerering, der figure-
5 rer på den ovenfor beskrevne sekvens, koder for visse interessante pepti-
der ifølge opfindelsen.

10	Sekvens, der koder for <u>env1</u> nucleotid 7850 - 7927
	- - - - <u>env2</u> - 8030 - 8095
	- - - - <u>env3</u> - 7601 - 7636
	- - - - <u>env4</u> - 6170 - 6247
	- - - - <u>env5</u> - 6294 - 6349
15	- - - - <u>env6</u> - 6392 - 6445
	- - - - <u>env7</u> - 6724 - 6763
	- - - - <u>env8</u> - 6794 - 6838
	- - - - <u>env9</u> - 7112 - 7162
	- - - - <u>env10</u> - 7253 - 7336
20	- - - - <u>env11</u> - 7358 - 7426

De korresponderende nucleinsyrer fra SIV-virus indeholder ligeledes en eller flere sekvenser afledt af cDNA fra SIV-1-virus. Disse sekvenser, der koder for peptiderne env1-env11 kan lokaliseres på fig. 3 ved sammenlig-
25 ning med de korresponderende sekvenser, der er beskrevet for HIV-2.

Når de immunogene midler er konjugeret med et bæremolekyle som angivet i krav 24 opnås dette ved kovalent kobling af peptiderne ifølge opfindelsen (eller de nævnte oligomere) til naturlige eller syntetiske fysiologisk
30 acceptable og ikke-toksiske bæremolekyler, ved hjælp af komplementære reaktive grupper, der henholdsvis bæres af bæremolekylet og peptidet. Eksempler på passende grupper er illustreret i det følgende:

Som eksempler på bærermolekyler eller makromolekylære bærere, der kan indgå i konstitutionen af konjugater ifølge opfindelsen, kan man nævnte naturlige proteiner, såsom tetanus anatoxin, ovalbumin, serumalbuminer, hæmocyaminer etc.

5

Som eksempler på syntetiske bærermolekyler kan man f.eks. nævne polylysiner eller poly(D-L-alanin)-poly(L-lysin).

10 I litteraturen nævnes andre typer anvendelige makromolekylære bærere, og disse udviser i reglen en molekylvægt på over 20 000.

For at syntetisere konjugaterne ifølge opfindelsen kan man anvende i og for sig kendte metoder, f.eks. som beskrevet af FRANTZ and ROBERTSON, *Infect. and Immunity*, 33, 193-198 (1981), eller P.E. KAUFFMAN, 15 *Applied and Environmental Microbiology*, (October 1981), vol. 42, nr. 4, 611-614 idet man anvender et passende peptid og bærermolekyle.

20 I praksis anvender man med fordel som koblingsmiddel f.eks. følgende forbindelser: glutaraldehyd, ethylchlorformiat, vandopløselige carbodiimider, såsom N-ethyl-N'-(3-dimethylaminopropyl)-carbodiimid, HCl, diisocyanater, bis-diazobenzidin, di- og trichlor-s-triaziner, cyanogenbromid samt de koblingsmidler, der er nævnt i *Scand. J. Immunol.*, (1978), vol. 8, p. 7-23 AVRAMEAS, TERNYNCK, GUESDON).

25 Man kan anvende en hvilken som helst koblingsproces, hvor der gøres brug af en eller flere reaktive grupper i peptidet og en eller flere reaktive grupper i bærermolekylerne. Med fordel drejer det sig om carboxyl- og aminogru-
pper, der kan give anledning til en koblingsreaktion i nærvær af
et koblingsmiddel af den type, der anvendes ved proteinsyntese, f.eks. 1-
30 ethyl-3-(3-dimethylaminopropyl)-carbodiimid, N-hydroxybenzotriazol etc.
Man kan også anvende glutaraldehyd, især når det drejer sig om indbyrdes at forbinde aminogru-
pper, der henholdsvis sidder på peptidet og bærermolekylet.

Peptiderne ifølge opfindelsen udviser antigeniske egenskaber. De kan således anvendes ved diagnostiske metoder til detektering af en infektion med HIV-2 virus.

5

Som nævnt tidligere har undersøgelser gjort det muligt at skelne mellem to grupper af peptider, der kan anvendes ved fremgangsmåder til detektering af antistoffer mod HIV-2 virus i humane biologiske væsker, især serum eller cephalospinalvæske.

10

En gruppe omfatter peptider, der nærmere bestemt svarer til de peptider, der er placeret i den transmembrane del og ved enden af den udvendige del af kappeproteinet. Disse proteiner er ovenfor betegnet env1, env2 og env3. De muliggør en specifik genkendelse af tilstedeværelsen af antistoffer mod HIV-2 og muliggør således en diskrimination mellem overståede eller tilstedeværende infektioner p.g.a. HIV hos en person, nærmere en bestemt en diskrimination mellem infektioner, som var provokeret af HIV-2 og infektioner, som var provokeret af HIV-1.

15

Opfindelsen angår ligeledes et antigen middel, der indeholder mindst et af de ovennævnte peptider eller mindst en oligomer af et sådant peptid, som er ejendommeligt ved, at det har evnen til at kunne genkendes af serum af human oprindelse, der indeholder antistoffer mod HIV-2.

20

Opfindelsen angår endvidere en fremgangsmåde til in vitro at diagnosticere et eller flere peptider ifølge opfindelsen for at kunne detektere antistoffer mod HIV-2 i biologiske væsker, især human serum.

25

Mere generelt omfatter in vitro diagnosemetoden følgende trin:

30

Den biologiske væske bringes i kontakt med peptiderne,

den eventuelle tilstedeværelse af et peptid-antistofkompleks detekteres ved fysiske eller kemiske metoder i den biologiske væske.

I en foretrukken udførelsesform for fremgangsmåden ifølge opfindelsen gennemføres detekteringen af antigen-antistofkomplekset ved hjælp af immunoenzymatiske prøver (af typen ELISA), immunofluorescensmetoder (af typen IFA), radioimmunologiske metoder (af typen RIA) eller radioimmunopræcipiteringsmetoder (af typen RIPA). Således angår opfindelsen ligeledes ethvert peptid ifølge opfindelsen, der er mærket ved hjælp af en passende markør, f.eks. en enzymatisk, fluorescerende eller radioaktiv markør.

Sådanne metoder omfatter f.eks. følgende trin:

15 afsætning af fastlagte mængder af et peptidholdigt middel ifølge opfindelsen i brøndene i en mikrotiterplade,

anbringelse af stigende fortyndinger af det serum, der skal diagnosticeres, i brøndene,

20

inkubering af mikrotiterpladen og gentagne skylninger af denne,

indføring i brøndene af et mærket antistof mod immunglobulinerne i blodet, hvor mærkningen af disse antistoffer kan være gennemført ved hjælp af et enzym valgt blandt enzymer, der er i stand til at hydrolysere et substrat ved at modificere absorptionen af dettes stråling, i det mindste ved en fastlagt bølgelængde,

25

detektering ved sammenligning med en kontrolprøve af den hydrolyserede mængde substrat.

30

Opfindelsen angår ligeledes kit til in vitro diagnose af tilstedeværelsen af antistoffer mod HIV-2 virus i et biologisk medium, som omfatter:

et peptidisk middel ifølge opfindelsen,

5 reagenser til tilvejebringelse af en sammensætning af mediet, der er hensigtsmæssig for gennemførelsen af den immunologiske reaktion,

reagenser, der muliggør en detektering af antigen-antistofkomplekset dannet ved den immunologiske reaktion. Sådanne reagenser kan ligeledes være forsynet med en markør eller være i stand til at kunne genkendes af
10 et mærket reagens. Dette er især tilfældet, når det ovennævnte peptidiske middel ikke er mærket.

- en biologisk referencevæsvæske fri for antistoffer, der genkendes af det
15 ovennævnte peptidiske middel.

Der kan dannes antistoffer mod peptiderne ifølge opfindelsen, og denne produktion er ikke begrænset til polyklonale antistoffer.

Den finder ligeledes anvendelse på et hvilket som helst monoklonalt antistof produceret af ethvert hybridom, der kan dannes ved klassiske metoder ud fra miltceller fra et dyr, især mus eller rotter, der er immuniseret med et af peptiderne ifølge opfindelsen, og celler fra en passende myelomacellelinie, og som kan udvælges pga. sin evne til at producere monoklonale antistoffer, der genkender det peptid, der oprindeligt blev anvendt til immuniseringen af dyret.
20
25

Opfindelsen angår endvidere immunogene midler til fremstilling af vacciner, hvis aktive bestanddel består af mindst et peptid ifølge opfindelsen eller en oligomer af et sådant peptid eller et peptid, der er konjugeret med
30 et bæremolekyle, og som er ejendommeligt ved, at de inducerer dannelsen af antistoffer mod disse peptider i tilstrækkelig mængde til også at inhibere proteiner fra HIV-2 retrovirus, ja endog selve HIV-2 retrovirus i association med en farmaceutisk acceptabel bærer.

De immunogene midler til fremstilling af vacciner omfatter fordelagtigt nærmere bestemt mindst et af de peptider, der ovenfor betegnedes env4, env5, env6, env7, env8, env9, env10, env11 eller blandinger heraf. Blandt disse peptider, der er velegnede til at udgøre aktive bestanddele i vacciner, er visse særligt foretrukne, fordi de udviser en basal aminosyrestruktur, der svarer til regioner i kappeglycoproteiner, der udviser en høj grad af konservering, ikke blot i HIV-2 og i SIV, men også i HIV-1 virus. Disse peptider, der er særligt foretrukne, er de peptider, der betegnes env4, og visse af peptiderne env5, env6 og env10.

I en foretrukken udførelsesform for opfindelsen vælges de immunogene peptider eller fragmenter af disse, der er velegnede som aktive bestanddele i vacciner, blandt de formler, der svarer til sekvenserne som i kappeglycoproteinerne fra HIV-2, SIV og HIV-1, udviser en aminosyrehomologi der er over 50%, og som hører til den eksterne del af kappen af virus, og som er fri for eller næsten fri for deletioner, og som indeholder cysteinrester, der fremmer stabiliteten af bindingerne og dannelsen af "hårnåle".

De efterfølgende peptider tilhører denne kategori af foretrukne peptider.

env4

XVTV-YGVP-W-ATZ

25 env5

XL-NVTE-FZ

env6

XKPCVKL-PLC-Z

30

env7

XN-S-I-Z

env10

XNC-GEF-YC-Z

env11

5 XC-I-Q-IZ

Fordelagtige farmaceutiske midler består af injicerbare opløsninger, suspensioner eller liposomer, der indeholder en effektiv dosis af mindst et produkt ifølge opfindelsen. Fortrinsvis tilvejebringes disse opløsninger, suspensioner eller liposomer i en steril isotonisk vandfase, fortrinsvis en fysiologisk saltopløsning eller glucoseopløsning.

Der anvendes nærmere bestemt sådanne suspensioner, opløsninger eller liposomer, der er i stand til at indgives ved intradermisk, intramuskulær eller subkutan injektion eller ved kopsætning.

Farmaceutiske midler, kan også indgives på anden måde, især oralt.

Farmaceutiske midler, der kan anvendes som vacciner, kan for at være effektive ved dannelsen af antistoffer mod HIV-2 virus f.eks. indgives i doser på 10-500 µg/kg af peptiderne ifølge opfindelsen, fortrinsvis 50-100 µg/kg.

Disse doser er kun anført som eksempler og skal ikke opfattes som begrænsende.

Som anført ovenfor kan de forskellige nærmere definerede peptider omfatte modifikationer, der ikke har til virkning, at de immunologiske egenskaber modificeres på fundamental vis. De ækvivalente peptider, der derved opnås, er omfattet af de efterfølgende patentkrav. Som eksempel på ækvivalente peptider kan man nævne peptider, hvis strukturer svarer til cDNA regioner i andre varianter af HIV-2, SIV eller HIV-1, når disse regioner bringes på linie under de betingelser, der svarer til de ovenfor angivne i

forbindelse med HIV-2 ROD, SIV og HIV-1 BRU. Som eksempler på andre af disse peptider kan man nævne peptider, hvis struktur svarer til de regioner i cDNA, der er deponeret ved CNCM, især med numrene I-502, I-642 (HIC-2 IRMO) og I-643 (HIV-2 EHO) samt i visse tilfælde varianter af HIV-1, der er deponeret ved CNCM med numrene I-232, I-240, I-241, I-550 og I-551.

Peptiderne ifølge opfindelsen kan også defineres ved de efterfølgende formler, hvori X, Z og bindestregerne "-" har de ovenfor angivne betydninger:

XRVAIEKYL-DQA-LN-WGCAFRQVCZ
XAIEKYL-DZ

15 X-LE-AQIQQEKNMYELOQKLSWZ
XQIQQEKNZ

XELGDYKLVEITPIG-APT--KR----Z
XYKLVEITPIG-APT--KRZ

20 X---VTV-YGVP-W--AT--LFCA-Z
XVTV-YGVP-W--ATZ

X---E--L-NVTE-F--W-NZ
25 XL-NVTE-FZ

XL---S-KPCVKL-PLC----Z
XKPCVKL-PLC-Z
XS-KPCVKL-PLC-Z

30 X--N-S-I--C-Z
XN-S-I-Z

XYC-P-G-A-L-C-N-TZ

X-----A-C-----W--Z

5 NKRPRQAWCWFKG-KWKD

X-G-DPE-----NC-GEF-YC-----NZ

X-----C-I-Q-I-----G---YZ

10

Opfindelsen angår ligeledes foruden de allerede beskrevne SIV peptider proteiner, som kodes for af cDNA fra SIV virus. De angår ligeledes proteiner fra enhver immunologisk virus, der i immunologisk henseende er nært beslægtet med SIV-1mac, specielt enhver virus, hvis kappeproteiner og -glycoproteiner udviser immunologisk overkrydsning, og hvis cDNA udviser

15

en homologiprocent på mindst 95%, fortrinsvis mest 98%.
Specielt angår opfindelsen kappeproteiner og -glycoproteiner, der kodes for af env-genet, og som er vist på fig. 3,

20

Aminosyrerne i de ovennævnte SIV-proteiner er vist på linie med aminosyresekvenserne fra de korresponderende proteiner fra HIV-2 virus; de vertikale punkter, der figurerer mellem de to sekvenser, svarer til aminosyrer, der er fælles for proteinerne fra de to virus.

25

De cDNA sekvenser der koder for de ovennævnte proteiner, er vist på fig. 1B.

Disse cDNA sekvenser lokaliseret ved den nummerering, der figurerer på de tidligere beskrevne sekvenser (figur 1B) er følgende:

30

- sekvens der koder for GAG, nucleotider 551-2068
- " " " " POL, " 1726-4893
- " " " " Q, " 4826-5467

	-	"	"	"	"	X,	"	5298-	5633
	-	"	"	"	"	R,	"	5637-	5939
	-	"	"	"	"	F,	"	8569-	9354
5	-	"	"	"	"	TAT-1,	"	5788-	6084
	-	"	"	"	"	ART-1,	"	6014-	6130
	-	"	"	"	"	TAT-2,	"	8296-	8391
	-	"	"	"	"	ART-2,	"	8294-	8548
	-	"	"	"	"	ENV,	"	6090-	8732

10

Opfindelsen angår således naturligvis de tidligere beskrevne proteiner, hvad enten de er opnået ud fra SIV virus, eller de er fremstillet af syntetisk vej, især ved hjælp af de metoder, der er nævnt ovenfor i forbindelse med syntese af mindre peptider.

15

Opfindelsen angår ligeledes anvendelsen af de foregående proteiner til diagnose af den eventuelle tilstedeværelse af antistoffer rettet mod HIV-2 proteiner, ja endog mod hele HIV-2 eller for visse af disses vedkommende anvendelsen til diagnose af en infektion pga. en HIV virus. Proteinerne ENV anvendes fortrinsvis til specifik diagnose af en infektion med HIV-2 eller en variant deraf, og sommetider til diagnose af en infektion af HIV-2 eller HIV-1.

20

25

Opfindelsen angår således ligeledes en in vitro diagnosemetode til detektering af antistoffer mod HIV-2 og eventuelt mod HIV-1 i biologiske væsker og især i human serum. Sådanne fremgangsmåder, der er anvendelige til anvendelse af de ovennævnte SIV-proteiner som diagnostiske proteiner, er beskrevet tidligere.

30

Opfindelsen angår således også kits til in vitro diagnose af tilstedeværelsen af antistoffer mod HIV-2 virus og også i tilfælde mod HIV-1 virus i et biologisk medium. Sådanne kits, der gør brug af de ovennævnte peptider, er ligeledes beskrevet ovenfor.

Endvidere angår opfindelsen immunogene midler til produktion af vacciner, hvis aktive bestanddel består af i det mindste en del af proteinet ENV fra SIV-virus, idet dette protein også kan være konjugeret med et bærer-
5 molekyle. Disse immunogene midler inducerer dannelsen af antistoffer mod dette peptid i tilstrækkelig mængde til at inhibere proteiner fra HIV-2 retrovirus eller endog selve HIV-2 retrovirus.

P a t e n t k r a v

1. Peptid, der udviser immunologiske egenskaber fælles med peptidskelettet i kappeglycoproteinet i virus af typen HIV-2, k e n d e t e g n e t ved, at
5 det ligeledes udviser en peptidstruktur, der er fælles med peptidskelettet i glycoproteinet fra SIV-1 virus.

2. Peptid ifølge krav 1, k e n d e t e g n e t ved, at antallet af aminosyrer ikke overstiger 40.
10

3. Peptid ifølge krav 2, k e n d e t e g n e t ved, at det udviser en af de al-
mene formler:

XRV-AIEKYL-DQA-LN-WGCAFRQVCZ

15 XAIEKYL-DZ

hvor X og Z er OH- eller NH₂-grupper, eller når de immunologiske egenskaber af peptider befriet for disse grupper ikke modificeres væsentligt, kan X og Z betegne grupper omfattende 1-5 aminosyrerester, og hver af
20 bindestregerne betegner en aminoacylrest valgt blandt sådanne, som muliggør, at det angivne peptid bevarer de immunologiske egenskaber fra en af følgende sekvenser:

RVTAIEKYLQDQARLNSWGCAFRQVC

25 AIEKYLQDQ

RVSAIEKYLKDQAQLNAWGCAFRQVC

AIEKYLKDQ

4. Peptid ifølge krav 2, k e n d e t e g n e t ved, at det udviser en af de al-
30 mene formler:

X-LE-AQIQQEKNMYELQKLNSWZ

XQIQQEKNZ

hvor X og Z er OH eller NH_2 -grupper, eller når de immunologiske egenskaber af peptider befriet for disse grupper ikke modificeres væsentligt, kan X og Z betegne grupper omfattende 1-5 aminosyrerester, og hver af
 5 bindestregerne betegner en aminoacylrest valgt blandt sådanne, som muliggør, at det angivne peptid bevarer de immunologiske egenskaber fra en af de følgende sekvenser:

SLEQAQIQQEKNMYELQKLNSW
 10 QIQQEKN
 LLEEAQIQQEKNMYELQKLNSW

5. Peptid ifølge krav 2, kendet ved, at det udviser en af de al-
 mene formler:

15
 XELGDYKLVEITPIG-APT--KR----Z
 XYKLVEITPIG-APT--KRZ

hvor X og Z er OH - eller NH_2 -grupper, eller når de immunologiske egenskaber af peptider befriet for disse grupper ikke modificeres væsentligt,
 20 kan X og Z betegne grupper omfattende 1-5 aminosyrerester, og hver af bindestregerne betegner en aminoacylrest valgt blandt sådanne, som muliggør, at det angivne peptid bevarer de immunologiske egenskaber fra en af de følgende sekvenser:

25
 ELGDYKLVEITPIGFAPTKEKRYSSAH
 YKLVEITPIGFAPTKEK
 ELGDYKLVEITPIGLAPTENVKRYTTG-
 YKLVEITPIGLAPTENVK

30
 6. Peptid ifølge krav 2, kendt ved, at det udviser en af de al-
 mene formler:

X---VTV-YGVP-W--AT--LFCA-Z
 XVTV-YGVP-W--ATZ

5 hvori X og Z er OH eller NH₂-grupper, eller når de immunologiske egenskaber af peptider befriet for disse grupper ikke modificeres væsentligt, kan X og Z betegne grupper omfattende 1-5 aminosyrerester, og hver af bindestregerne betegner en aminoacylrest valgt blandt sådanne, som muliggør, at det angivne peptid bevarer de immunologiske egenskaber fra en af de følgende sekvenser:

10

CTQYVTVFYGVPTWKNATIPLFCAT
 VTVFYGVPTWKNAT
 CIQYVTVFYGVPAWRNATIPLFCAT
 VTVFYGVPAWRNAT

15

EKLWVTVYYGVPVWKEATTTLFCAS
 VTVYYGVPVWKEAT

7. Peptid ifølge krav 6, k e n d e t e g n e t ved, at det udviser en af form-
 lerne:

20

CTQYVTVFYGVPTWKNATIPLFCAT
 VTVFYGVPTWKNAT
 CIQYVTVFYGVPAWRNATIPLFCAT
 VTVFYGVPAWRNAT

25

EKLWVTVYYGVPVWKEATTTLFCAS
 VTVYYGVPVWKEAT

EDLWVTVYYGVPVWKEATTTLFCAS
 VTVYYGVPVWKEAT

30

DNLWVTVYYGVPVWKEATTTLFCAS
 VTVYYGVPVWKEAT

8. Peptid ifølge krav 2, k e n d e t e g n e t ved, at det udviser en af de al-
 mene formler:

X---E--L-NVTE-F--W-NZ

XL-NVTE-FZ

- 5 hvori X og Z er OH eller NH₂-grupper, eller når de immunologiske egenskaber af peptider befriet for disse grupper ikke modificeres væsentligt, kan X og Z betegne grupper omfattende 1-5 aminosyrerester, og hver af bindestregerne betegner en aminoacylrest valgt blandt sådanne, som muliggør, at det angivne peptid bevarer de immunologiske egenskaber fra en
- 10 af de følgende sekvenser:

DDYQEITL-NVTEAFDAWNN

L-NVTE

DDYSELAL-NVTESFDAWEN

- 15 PNPQEVVLVNVTENFNMWKN

LVNVTE

9. Peptid ifølge krav 8, k e n d e t e g n e t ved, at det udviser en af form-
lerne:

20

DDYQEITL-NVTEAFDAWNN

L-NVTEAF

DDYSELAL-NVTESFDAWEN

L-NVTESF

- 25 PNPQEVVLVNVTENFNMWKN

LVNVTENF

PNPQEIELENVTEGFNMWKN

LENVTEGF

PNPQEIALENVTENFNMWKN

30

LENVTENF

10. Peptid ifølge krav 2, k e n d e t e g n e t ved, at det udviser en af de
almene formler:

XL---S-KPCVKL-PLC----Z

XKPCVKLTPLCVZ

XS-KPCVKLTPLCVZ

5

- hvor X og Z er OH eller NH₂-grupper, eller når de immunologiske egenskaber af peptider befriet for disse grupper ikke modificeres væsentligt, kan X og Z betegne grupper omfattende 1-5 aminosyrerester, og hver af bindestregerne betegner en aminoacylrest valgt blandt sådanne, som mulig-
- 10 liggør, at det angivne peptid bevarer de immunologiske egenskaber fra en af de følgende sekvenser:

LFETSIKPCVKLTPLCVAMK

LFETSIKPCVKLSPLCITMR

15 LWDQSLKPCVKLTPLCVSLK

KPCVKLTPLCV

KPCVKLSPLCI

SLKPCVKLTPLCV

- 20 11. Peptid ifølge krav 10, k e n d e t e g n e t ved, at det udviser en af de følgende strukturer:

LFETSIKPCVKLTPLCVAMK

LFETSIKPCVKLSPLCITMR

25 LWDQSLKPCVKLTPLCVSLK

LWDQSLKPCVKLTPLCVTLN

PCVKLTPLCV

KPCVKLSPLCI

- 30 12. Peptid ifølge krav 2, k e n d e t e g n e t ved, at det indeholder basisstrukturen:

X---N-S-I---C-Z

XN-S-I-Z

hvor X og Z er OH eller NH₂-grupper, eller når de immunologiske egenskaber af peptider befriet for disse grupper ikke modificeres væsentligt, kan X og Z betegne grupper omfattende 1-5 aminosyrerester, og hver af

5 bindestregerne betegner en aminoacylrest valgt blandt sådanne, som muliggør, at det angivne peptid bevarer de immunologiske egenskaber fra en af de følgende sekvenser:

10 NHCNTSVITESCD
 NTSVIT
 NHCNTSVIQECCD
 NTSVIQ
 TSCNTSVITQACP

15 NTSVIT

13. Peptid ifølge krav 12, k e n d e t e g n e t ved, at det udviser en af formierne:

20 NHCNTSVITESCD
 NTSVIT
 NHCNTSVIQECCD
 NTSVIQ
 TSCNTSVITQACP

25 NTSVIT
 INCNTSVITQACP
 NTSVIT
 INCNTSAITQACP
 NTSAIT

30

14. Peptid ifølge krav 2, k e n d e t e g n e t ved, at det har den almene formel:

XYC-P-G-A-L-C-N-TZ

- hvor X og Z er OH eller NH₂-grupper, eller når de immunologiske egenskaber af peptider befriet for disse grupper ikke modificeres væsentligt, kan X og Z betegne grupper omfattende 1-5 aminosyrerester, og hver af bindestregerne betegner en aminoacylrest valgt blandt sådanne, som muliggør, at det angivne peptid bevarer de immunologiske egenskaber fra en af de følgende sekvenser:

10 YCAPPGYALLRC-NDT
YCAPAGFAILKCNNKT

15. Peptid ifølge krav 14, kendes ved, at det har en af formlerne:

15 YCAPPGYALLRC-NDT
YCAPAGFAILKCNNKT
YCAPAGFAILKCNDKK
YCAPAGFAILKCRDKK

- 20 16. Peptid ifølge krav 2, kendes ved, at det har den almene formel:

25 X-----A-C-----W--Z

- hvor X og Z er OH eller NH₂-grupper, eller når de immunologiske egenskaber af peptider befriet for disse grupper ikke modificeres væsentligt, kan X og Z betegne grupper omfattende 1-5 aminosyrerester, og hver af bindestregerne betegner en aminoacylrest valgt blandt sådanne, som muliggør, at det angivne peptid bevarer de immunologiske egenskaber fra en af de følgende sekvenser:

NKRPRQAWCWFKG-KWKD

NERPKQAWCRFGG-NWKE
N--MRQAHCNISRAKWNA

5 17. Peptid ifølge krav 16, k e n d e t e g n e t ved, at det har en af formler-
ne:

NKRPRQAWCWFKG-KWKD
NERPKQAWCRFGG-KWKE
N--MRQAHCNISRAKWNA
10 D--IRRAYCTINETEWDK
I--IGQAHCNISRAQWSK

15 18. Peptid ifølge krav 2, k e n d e t e g n e t ved, at det udviser en af de
almene formler:

X-G-DPE-----NC-GEF-YC-----NZ
XNC-GEF-YC-Z

20 hvori X og Z er OH eller NH₂-grupper, eller når de immunologiske egen-
skaber af peptider befriet for disse grupper ikke modificeres væsentligt,
kan X og Z betegne grupper omfattende 1-5 aminosyrerester, og hver af
bindestregerne betegner en aminoacylrest valgt blandt sådanne, som mu-
liggør, at det angivne peptid bevarer de immunologiske egenskaber fra en
af de følgende sekvenser:

25 KGSDPEVAYMWTNCRGEFLYCNMTWFLN
NCRGEFLYCN
GG-DPEVTFMWTNCRGEFLYCKMNWFLN
NCRGEFLYCK
30 -GGDPEIVTHSFNCGGEFFYCNSTQLFN
NCGGEFFYCN

19. Peptid ifølge krav 18, kendet e g n e t ved, at det udviser en af følgende strukturer:

5 KGSDPEVAYMWTNCRGEFLYCNMTWFLN
 NCRGEFLYCN
 GG-DPEVTFMWTNCRGEFLYCKMNWFLN
 NCRGEFLYCK
 -GGDPEIVTHSFNCGGEFFYCNSTQLFN
 NCGGEFFYCN
 10 -GGDPEITTHSFNCRGEFFYCNSTSKLFN
 NCRGEFFYCN
 -GGDPEITTHSFNCGGEFFYCNSTGLFN
 NCGGEFFYCN

15 20. Peptid ifølge krav 2, kendet e g n e t ved, at det udviser en af de almene formler:

X-----C-I-Q-I-----G---YZ
 XC-I-Q-IZ

20

hvor X og Z er OH eller NH₂-grupper, eller når de immunologiske egenskaber af peptider befriet for disse grupper ikke modificeres væsentligt, kan X og Z betegne grupper omfattende 1-5 aminosyrerester, og hver af bindestregerne betegner en aminoacylrest valgt blandt sådanne, som muliggør, at det angivne peptid bevarer de immunologiske egenskaber fra en af de følgende sekvenser:

RNYAPCHIKQIINTWHKVGRNVY
 CHIKQII
 30 RNYVPCHIRQIINTWHKVGKNVY
 CHIRQII
 TITLPCRIKQFINMWQEVGKAMY
 CRIKQFI

21. Peptid ifølge krav 20, k e n d e t e g n e t ved, at det udviser en af strukturerne:

5 RNYAPCHIKQIINTWHKVGGRNVY
CHIKQII
RNYVPCHIRQIINTWHKVGKKNVY
CHIRQII
TITLPCRIKQFINMWQEVGKAMY
10 CRIKQFI
SITLPCRIKQIINMWQKTCKAMY
CRIKQII
NITLQCRIKQIIKMOVAGR-KAIY
CRIKQII

15

22. Peptid, der udviser peptidstruktur fælles med kappeglycoprotein peptidskelettet fra HIV-2, k e n d e t e g n e t ved, at det ligeledes udviser en peptidstruktur der er fælles med peptidskelettet i ENV aminosyresekvensen fra SIV-1 vist i figur 3.

20

23. Antigen middel indeholdende peptidet env ifølge krav 22 eller mindst et peptid ifølge krav 3, 4 og 5 eller mindst en oligomer af et sådant peptid, k e n d e t e g n e t ved, at det specifikt genkender tilstedeværelsen af antistoffer mod HIV-2.

25

24. Immunogen middel indeholdende hele eller en del af peptidet env ifølge krav 22 eller mindst et peptid eller mindst en oligomer af dette peptid eller et peptid ifølge krav 6-21 konjugeret med et bæremolekyle, i kombination med en farmaceutisk acceptabel bærer til fremstilling af vacciner, k e n d e t e g n e t ved, at det inducerer dannelsen af antistoffer mod disse peptider i tilstrækkelig mængde til effektivt at inhibere proteiner fra HIV-2 retrovirus, eller endog selve HIV-2 retrovirus.

30

25. Immunogent middel ifølge krav 24, k e n d e t e g n e t ved, at det indeholder peptider med sekvenser, som i kappeglycoproteiner fra HIV-2, SIV-1 og HIV-1 udviser en aminosyrehomologi på over 50%.
- 5 26. Immunogent middel ifølge krav 24 eller 25, k e n d e t e g n e t ved, at det indeholder mindst et peptid eller mindst en oligomer af et sådant peptid eller peptidet konjugeret med et bærermolekyle, hvilket peptid er valgt blandt env4, env5, env6 og env10.
- 10 27. Fremgangsmåde til in vitro diagnosticering af HIV-2 infektioner i en biologisk væske, k e n d e t e g n e t ved, at man bringer den biologiske væske i kontakt med mindst et peptid ifølge ethvert af kravene 1-5 eller et konjugat af disse peptider med et bærermolekyle, eller i kontakt med peptidet env ifølge krav 22, og detekterer den eventuelle tilstedeværelse af et antigen-antistofkompleks i den biologiske væske ved hjælp af fysiske eller
- 15 kemiske metoder.
28. Fremgangsmåde til in vitro diagnosticering af HIV-2 infektioner i en biologisk væske ifølge krav 27, k e n d e t e g n e t ved, at detekteringen af
- 20 det eventuelt dannede antigen-antistofkompleks gennemføres ved hjælp af immunoenzymatiske prøver (af typen ELISA), immunofluorescensprøver (af typen IFA), radioimmunologiske prøver (af typen RIA) eller radioimmunopræcipiteringsprøver (af typen RIPA).
- 25 29. Kit til in vitro diagnosticering af HIV-2 infektioner i en biologisk væske, k e n d e t e g n e t ved, at det omfatter:
- et peptidisk middel indeholdende et peptid ifølge ethvert af kravene 1-5, eller en blanding af sådanne peptider eller en konjugat af disse peptider
- 30 med et bærermolekyle, eller peptidet env ifølge krav 22,
- et reagens til sammensætning af et medium, der er gunstigt for gennemførelsen af en immunologisk reaktion,

et eller flere eventuelt mærkede reagenser til detektering af antigen-antistofkomplekset dannet ved den immunologiske reaktion, samt

- 5 en biologisk referencevæske, der er fri for antistoffer, som genkendes af det nævnte peptidiske

FIG. 1.A

HIV2.ROD

→ R
 GTCGCTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCTCCAGCACTAGCAGG
 TAGAGCCTGGGTGTTCCCTGCTAGACTCTCACCAGCACTTGGCCGGTGC TGGGCAGACGG
 CCCJACGCTTGCTTGCTTAAAAACCTCTTAATAAAGCTGCCAGTTAGAAGCAAGTTAAGT
 GTGTGCTCCCATCTCTCCTAGTCGCCGCCCTGGTCAITCGGTGTTACCTGAGTAACAAGA
 CCTGCTCTGTTAGGACCCTTCTTGCTTTGGGAAACCGAGGCAGGAAAAATCCCTAGCAGG
 TTGGCGCCTGAACAGGGACTTGAAGAAGACTGAGAAGTCTTGGAACACGGCTGAGTGAAG
 GCAGTAAGGGCGGCAGGAACAAACCACGACGGAGTGCTCCTAGAAAGCGCGGGCCGAGG
 CACCAAAGGCAGCGTGTGGAGCGGGAGGAGAAGACCCCTCCGGCTGACGGTAAGTACCTA
 CACCAAAAACGTAGCCGAAAGGGCTTGCTATCCTACCTTACATAGGTAGAAGATTGTG
 MetGlyAlaArgAsnSerValLeuArgGlyLysLysAlaAspIleLeuGluArgIle
 CGAGATGGGCGCGAGAAACTCCGTCTTGACAGGSAAAAAAGCAGATCAATTAGAAAGAA
 ArgLeuArgProGlyGlyLysLysLysTyrArgLeuLysHisIleValTrpAlaAlaAsn
 CAGGTACGGCCCGCGGAAAGAAAAAGTACAGGCTAAAAATATTTGTGTGGCAGCGAA
 LysLeuAspArgPheGlyLeuAlaGluSerLeuLeuGluSerLysGluGlyCysGlnLys
 TAAATTGCACAGATTCGGATTAGCAGAGACCTGTTCGACTCAAAGAGGGGTTGTC/AAA
 IleLeuThrValLeuAspProMetValProThrGlySerGluAsnLeuLysSerLeuPhe
 AATTCTTACAGTTTTAGATCCAATGGTACCGACAGGTTACAGAAAATTTAAAAAGTCTTT
 AsnThrValCysValIleTrpCysIleHisAlaGluGluLysValLysAspThrGluGly
 TAATACTGTCTGCGTCATTTGGTGCATACACGCAGAAGAGAAAAGTAAAGATACTGAAGG
 AlaLysGlnIleValArgArgHisLeuValAlaGluThrGlyThrAlaGluLysMetPro
 AGCAAAACAAATAGTGGGAGACATCTACTGGCAGAAACAGCAACTGCAGAGAAAATGCC

FIG. 1A

Ser Thr Ser Arg Pro Thr Ala Pro Ser Ser Glu Lys Gly Gly Asn Tyr Pro Val Gln His
 AAC CACA AGT AGACCAACAGCACCATCTACCGAGAAGGGAGGAAATTACCCAGTGC AAC A
 Val Gly Gly Asn Tyr Thr His Ile Pro Leu Ser Pro Arg Thr Leu Asn Ala Trp Val Lys
 TGTAGGGCGCAACTACACCCATATACCGCTGAGTCCCCGAACCCTAAATGCCTGGGTAAA
 1000
 Leu Val Glu Glu Lys Lys Phe Gly Ala Glu Val Val Pro Gly Phe Gln Ala Leu Ser Glu
 ATTAGTAGAGGAAAAAAGTTCCGGGCAGAACTAGTCCAGGATTT CAGGCACTCTCAGA
 Gly Cys Thr Pro Tyr Asp Ile Asn Gln Met Leu Asn Cys Val Gly Asp His Gln Ala Ala
 AGCCTGCACGCCCTATGATATCAACCAAATGCTTAATTGTGTGGGCGACCATCAAGCAGC
 1100
 Met Gln Ile Ile Arg Glu Ile Ile Asn Glu Ala Ala Glu Trp Asp Val Gln His Pro
 CATGCAGATAATCAGGGAGATTATCAATGAGGAAGCAGCAGAATGGGATGTGCAACATCC
 1200
 Ile Pro Gly Pro Leu Pro Ala Gly Gln Leu Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly
 AATACCAGGCCCTTACCAGCGGGGCGAGCTTAGAGAGCCAAGGGGATCTGACATAGCAGG
 Thr Thr Ser Thr Val Glu Glu Gln Ile Gln Trp Met Phe Arg Pro Gln Asn Pro Val Pro
 GACAACAAGCACAGTAGAAGAACAGATCCAGTGGATGTTTAGGCCACAAAATCCTGTACC
 1300
 Val Gly Asn Ile Tyr Arg Arg Trp Ile Gln Ile Gly Leu Gln Lys Cys Val Arg Met Tyr
 AGTAGGAAACATCTATAGAAGATGGATCCAGATAGGATTGCAGAAGTGTGTCAGGATGTA
 Asn Pro Thr Asn Ile Leu Asp Ile Lys Gln Gly Pro Lys Glu Pro Phe Gln Ser Tyr Val
 CAACCCGACCAACATCCTAGACATAAAACAGGGACCAAAGGAGCCGTTCCAAAGCTATGT
 1400
 Asp Arg Phe Tyr Lys Ser Leu Arg Ala Glu Gln Thr Asp Pro Ala Val Lys Asn Trp Met
 AGATAGATTCTACAAAAGCTTGAGGGCAGAACAAACAGATCCAGCAGTGAAGAATTGGAT
 1500
 Thr Gln Thr Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Leu Val Leu Lys Gly Leu
 GACCCAAACACTGCTAGTACAAAATGCCAACCCAGACTGTAAATTAGTGCTAAAAGGACT
 Gly Met Asn Pro Thr Leu Glu Glu Met Leu Thr Ala Cys Gln Gly Val Gly Gly Pro Gly
 AGGGATGAACCCTACCTTAGAAGAGATGCTGACCCGCTGT CAGGGGGTAGGTGGCCAGG
 1600
 Gln Lys Ala Arg Leu Met Ala Glu Ala Leu Lys Glu Val Ile Gly Pro Ala Pro Ile Pro
 CCAGAAAGCTAGATTAATGGCAGAGGCCCTGAAAGAGGT CATAGGACCTGCCCTATCCC
 Phe Ala Ala Ala Gln Gln Arg Lys Ala Phe Lys Cys Trp Asn Cys Gly Lys Glu Gly His
 ATTCCAGCAGCCCAGCAGAGAAAGGCATTTAAATGCTGGA ACTGTGGAAAGGAAGGGCA
 1700
 Ser Ala Arg Gln Cys Arg Ala Pro Arg Arg Gln Gly Cys Trp Lys Cys Gly Lys Pro Gly
 CTCGGCAAGACAATGCCGAGCACCTAGAAGGCAGGGCTGCTGGAAGTGTGGTAAGCCAGG
 1800
 Thr Gly Arg Phe Phe Arg Thr Gly Pro Leu Gly
 His Ile Met Thr Asn Cys Pro Asp Arg Gln Ala Gly Phe Leu Gly Leu Gly Pro Trp Gly
 ACACATCATGACAAACTGCCAGATAGACAGGCAGGTTTTTTTAGGACTGGGCCCTTGGGG
 Lys Glu Ala Pro Gln Leu Pro Arg Gly Pro Ser Ser Ala Gly Ala Asp Thr Asn Ser Thr
 Lys Lys Pro Arg Asn Phe Pro Val Ala Gln Val Pro Gln Gly Leu Thr Pro Thr Ala Pro
 AAAGAAGCCCCGCAACTTCCCCGTGGCCCAAGTTC CGCAGGGGCTGACACCAACAGCACC
 1900
 Pro Ser Gly Ser Ser Ser Gly Ser Thr Gly Glu Ile Tyr Ala Ala Arg Glu Lys Thr Glu
 Pro Val Asp Pro Ala Val Asp Leu Leu Glu Lys Tyr Met Gln Gln Gly Lys Arg Gln Arg
 CCCAGTGGATCCAGCAGTGGATCTACTGGAGAAATATATGCAGCAAGGGAAAAGACAGAG
 Arg Ala Glu Arg Glu Thr Ile Gln Gly Ser Asp Arg Gly Leu Thr Ala Pro Arg Ala Gly
 Glu Gln Arg Glu Arg Pro Tyr Lys Glu Val Thr Glu Asp Leu Leu His Leu Glu Gln Gly
 AGAGCAGAGAGAGAGACCATAACAAGGAAGTGACAGAGGACTTACTGCACCTCGAGCAGGG
 (fig. 1A-forts. 1)

GlyAspThrIleGlnGlyAlaThrAsnArgGlyLeuAlaAlaProGlnPheSerLeuTrp
 GluThrProTyrArgGluProProThrGluAspLeuLeuHisLeuAsnSerLeuPheGly
 GGAGACACCATACAGGGAGCCACCAACAGAGGACTTGCTGCACCTCAATTCTCTCTTTGG
 2100

LysArgProValValThrAlaTyrIleGluGlyGlnProValGluValLeuLeuAspThr
 LysAspGln
 AAAAGACCAGTAGTCACAGCATACATTGAGGGTCAGCCAGTAGAAGTCTTGTTAGACACA

GlyAlaAspAspSerIleValAlaGlyIleGluLeuGlyAsnAsnTyrSerProLysIle
 GGGGCTGACCACTCAATAGTAGCAGGAATAGAGTTAGGGAACAATTATAGCCCCAAAATA
 2200

ValGlyGlyIleGlyGlyPheIleAsnThrLysGluTyrLysAsnValGluIleGluVal
 GTAGGGGGAATAGGGGGATTTCATAAATACCAAGGAATATAAAAATGTAGAAATAGAAGTT

LeuAsnLysLysValArgAlaThrIleMetThrGlyAspThrProIleAsnIlePheGly
 CTAATAAAAAAGGTACGGGCCACCATAATGACAGGGCAGACCCCAATCAACATTTTTGGC
 2300

ArgAsnIleLeuThrAlaLeuGlyMetSerLeuAsnLeuProValAlaLysValGluPro
 AGAAATATTCTGACAGCCTTAGGCATGTCATTAATCTACCAGTCGCCAAAGTAGAGCCA
 2400

IleLysIleMetLeuLysProGlyLysAspGlyProLysLeuArgGlnTrpProLeuThr
 AAAAAATAATGCTAAAGCCAGGGAAAGATGGACCAAACTGAGACAATGGCCCTTAACA

LysGluLysIleGluAlaLeuLysGluIleCysGluLysMetGluLysGluGlyGlnLeu
 AAAGAAAAAATAGAAGCACTAAAAGAAATCTGTGAAAAAATGGAAAAAGAAGGCCAGCTA
 2500

GluGluAlaProProThrAsnProTyrAsnThrProThrPheAlaIleLysLysLysAsp
 GAGGAAGCACCTCCAATAATCCTTATAATACCCCCACATTTGCAATCAAGAAAAAGGAC

LysAsnLysTrpArgMetLeuIleAspPheArgGluLeuAsnLysValThrGlnAspPhe
 AAAAACAATGGAGGATGCTAATAGATTTTCAGAGAACTAAACAAGGTAACCTCAAGATTTT
 2600

ThrGluIleGlnLeuGlyIleProHisProAlaGlyLeuAlaLysLysArgArgIleThr
 ACAGAAATTCAGTTAGGAATTCACACCCAGCAGGGTTGGCCAAGAAGAGAAGAATTACT
 2700

ValLeuAspValGlyAspAlaTyrPheSerIleProLeuHisGluAspPheArgProTyr
 GTACTAGATGTAGGGGATGCTTACTTTTCCATACCACTACATGAGGACTTTAGACCATAT

ThrAlaPheThrLeuProSerValAsnAsnAlaGluProGlyLysArgTyrIleTyrLys
 ACTGCATTTACTCTACCATCAGTGAACAATGCAGAACCAGGAAAAAGATACATATATAAA
 2800

ValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnHisThrMetArgGlnVal
 GTCTTGCCACAGGGATGGAAGGGATCACCAGCAATTTTTCAACACACAATGAGACAGGTA

LeuGluProPheArgLysAlaAsnLysAspValIleIleIleGlnTyrMetAspAspIle
 TTAGAACCATTTCAGAAAAGCAAACAAGGATGTCATTATCATTCACTACATGGATGATATC
 2900

LeuIleAlaSerAspArgThrAspLeuGluHisAspArgValValLeuGlnLeuLysGlu
 TTAATAGCTAGTGACAGGACAGATTTAGAACATGATAGGGTAGTCTGCAGCTCAAGCAA
 3000

LeuLeuAsnGlyLeuGlyPheSerThrProAspGluLysPheGlnLysAspProProTyr
 CTTCTAAATGGCTTAGGATTTTCTACCCAGATGAGAAGTTCCAAAAAGACCCTCCATAC

HisTrpMetGlyTyrGluLeuTrpProThrLysTrpLysLeuGlnLysIleGlnLeuPro
 CACTGGATGGGCTATGAACTATGGCCAACTAAATGGAAGTTGCAGAAAATACAGTTGCC
 3100

GlnLysGluIleTrpThrValAsnAspIleGlnLysLeuValGlyValLeuAsnTrpAla
 CAAAAAGAAATATGGACAGTCAATGACATCCAGAAGCTAGTGGGTGTCCTAAATTGGGCA

(fig.1A- forts. 2)

AlaGlnLeuTyrProGlyIleLysThrLysHisLeuCysAspLeuIleArgGlyLysMet
 GCACAACCTCTACCCAGGGATAAAGACCCAAACACTTATGTAGGTTAATCAGAGGAAAAATG
 3200

ThrLeuThrGluGluValGlnTrpThrGluLeuAlaGluAlaGluLeuGluGluAsnArg
 ACACTCACAGAAGAAGTACACTGCACAGAATTAGCAGAAGCAGAGCTAGAAGAAAAACAGA
 3300

IleIleLeuSerGlnGluGlnGluGlyHisTyrTyrGlnGluGluLysGluLeuGluAla
 ATTATCCTAAGCCAGGAACAAGAGGGACACTATTACCAAGAAGAAAAAGAGCTAGAAGCA

ThrValClnLysAspGluGluAsnGluTrpThrTyrLysIleHisGlnGluGluLysIle
 AGAGTCCAAAAGGATCAAGAGAAATGACTGGACATATAAAATACACCAGGAAGAAAAAATT

LeuLysValGlyLysTyrAlaLysValLysAsnThrHisThrAsnGlyIleArgLeuLeu
 CTAAGAAGTAGGAAAATATGCAAAGGTGAAAAACACCCATACCAATGGAATCAGATTGTTA

AlaGlnValValGlnLysIleGlyLysGluAlaLeuValIleTrpGlyArgIleProLys
 GCACAGGTAGTTCAGAAAATAGGAAAAGAAGCACTAGTCATTGCGGGACGAATACCAAAA
 3500

PheHisLeuProValGluArgGluIleTrpGluGlnTrpTrpAspAsnTyrTrpGlnVal
 ITTCACCTACCAGTAGAGAGAGAAATCTGGGAGCAGTGGTGGGATAACTACTGGCAAGTG
 3600

ThrTrpIleProAspTrpAspPheValSerThrProProLeuValArgLeuAlaPheAsn
 ACATGGATCCCAGACTGGGACTTCGTGTCTACCCACCAGTGGTCAGGTTAGCGTTTAAC

LeuValGlyAspProIleProGlyAlaGluThrPheTyrThrAspGlySerCysAsnArg
 CTGGTAGGGGATCCTATACCAGGTGCAGAGACCTTCTACACAGATGGATCCTGCAATAGG
 3700

GlnSerLysGluGlyLysAlaGlyTyrValThrAspArgGlyLysAspLysValLysLys
 CAATCAAAAGAAGGAAAAGCAGGATATGTAACAGATAGAGGGAAAGACAAGGTAAGAAA

LeuGluGlnThrThrAsnGlnGlnAlaGluLeuGluAlaPheAlaMetAlaLeuThrAsp
 CTAGAGCAAACCTACCAATCAGCAAGCAGAACTAGAAGCCTTTGCGATGGCACTAACAGAC
 3800

SerGlyProLysValAsnIleIleValAspSerGlnTyrValMetGlyIleSerAlaSer
 TCGGGTCCAAAAGTTAATATTATAGTAGACTCACAGTATGTAATGGGGATCAGTGCAGC
 3900

GlnProThrGluSerGluSerLysIleValAsnGlnIleIleGluGluMetIleLysLys
 CAACCAACAGAGTCAGAAAGTAAAATAGTGAACCAGATCATAGAAGAAATGATAAAAAAG

GluAlaIleTyrValAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlnGluVal
 GAAGCAATCTATGTTGCATGGGTCCCAGCCCACAAAGGCATAGGGGAAACCAGGAAGTA
 4000

AspHisLeuValSerGlnGlyIleArgGlnValLeuPheLeuGluLysIleGluProAla
 GATCATTTAGTGAGTCAGGGTATCAGACAAGTGTTCCTCGGAAAAAATAGAGCCCGCT

GlnGluGluHisGluLysTyrHisSerAsnValLysGluLeuSerHisLysPheGlyIle
 CAGGAAGAACATCAAAAATATCATAGCAATGTAAGAAGACTGTCTCATAAATTTGGAATA
 4100

ProAsnLeuValAlaArgGlnIleValAsnSerCysAlaGlnCysGlnGlnLysGlyGlu
 CCCAATTTAGTGGCAAGGCAAATAGTAAACTCATGTGCCAATGTCAACAGAAAGGGGAA
 4200

AlaIleHisGlyGlnValAsnAlaGluLeuGlyThrTrpGlnMetAspCysThrHisLeu
 GCTATACATGGGCAAGTAAATCCAGAACTAGGCAGTGGCAAATGGACTGCACACATTTA

GluGlyLysIleIleIleValAlaValHisValAlaSerGlyPheIleGluAlaGluVal
 GAAGGAAAAGATCATTATAGTAGCAGTACATGTTGCAAGTGGATTTATAGAAGCAGAAGTC
 4300

IleProGlnGlnSerGlyArgGlnThrAlaLeuPheLeuLeuLysLeuAlaSerArgTrp
 ATCCACAGGAATCAGGAAGACAAAACAGCACTCTTCCTATTGAAACTGGCAAGTAGGTCG

(fig. 1A-forts. 3)

ProIleThrHisLeuHisThrAspAsnGlyAlaAsnPheThrSerGlnGluValLysMet
 CCAATAACACACTTGCATACAGATAATGGTGCCAACTTCACTTACAGGAGGTGAAGATC
 4400
 ValAlaTrpTrpIleGlyIleGluGlnSerPheGlyValProTyrAsnProGlnSerGln
 GTAGCATGGTGGATAGGTATAGAACAATCCTTTGGAGTACCTTACAATCCACAGAGCCAA
 4500
 GlyValValGluAlaMetAsnHisHisLeuLysAsnGlnIleSerGluThrIleValLeu
 GGAGTAGTAGAAGCAATGAATCACCATCTAAAAAACCAAATAAGTGAACAATAGTACTA
 MetAlaIleHisCysMetAsnPheLysArgArgGlyGlyIleGlyAspMetThrProSer
 ATGCCAATTCATTGCATGAATTTTAAAAGAAGGGGGGAATAGGGGATATGACTCCATCA
 4600
 GluArgLeuIleAsnMetIleThrThrGluGlnGluIleGlnPheLeuGlnAlaLysAsn
 GAAAGATTAATCAATATGATCACCACAGAACAAGAGATACAATTCCTCCAAGCCAAAAAT
 SerLysLeuLysAspPheArgValTyrPheArgGluGlyArgAspGlnLeuTrpLysGly
 TCAAATTAAGAAGATTTTCGGGTCTATTTTCAGAGAAGGCAGAGATCAGTTGTGAAAGGA
 4700
 ProGlyGluLeuLeuTrpLysGlyGluGlyAlaValLeuValLysValGlyThrAspIle
 CCTCGGGAAGTACTGTGGAAAGGAGAAGGAGCAGTCTTAGTCAAGGTAGGAACAGACATA
 4800
 LysIleIleProArgArgLysAlaLysIleIleArgAspTyrGlyGlyArgGlnGluMet
 MetGluGluAspLysArgTrp
 AAAATAATACCAAGAAGGAAAGCCAAGATCATCAGAGACTATGGAGGAAGACAAGAGATC
 AspSerGlySerHisLeuGluGlyAlaArgGluAspGlyGluMetAla
 IleValValProThrTrpArgValProGlyArgMetGluLysTrpHisSerLeuValLys
 GATAGTGGTTCCACCTGGAGGGTGCCAGGGAGGATGGAGAAATGGCATAGCCTTGTCAA
 4900
 TyrLeuLysTyrLysThrLysAspLeuGluLysValCysTyrValProHisHisLysVal
 GTATCTAAAATACAAAACAAAGGATCTAGAAAAGGTGTGCTATGTTCCCCACCATAAGGT
 GlyTrpAlaTrpTrpThrCysSerArgValIlePheProLeuLysGlyAsnSerHisLeu
 GGGATGGCATGGTGGACTTGCAGCAGGGTAATATTCCATTAAAAGGAAACAGTCATCT
 5000
 GluIleGlnAlaTyrTrpAsnLeuThrProGluLysGlyTrpLeuSerSerTyrSerVal
 AGAGATACAGGCATATTGGAACCTAACACCAGAAAAAGGATGGCTCTCCTCTTATTAGT
 5100
 ArgIleThrTrpTyrThrGluLysPheTrpThrAspValThrProAspCysAlaAspVal
 AAGAATAACTTGGTACACAGAAAAGTTCTGGACAGATGTTACCCAGACTGTGCAGATGT
 LeuIleHisSerThrTyrPheProCysPheThrAlaGlyGluValArgArgAlaIleArg
 CCTAATACATAGCACTTATTTCCCTTGCTTTACAGCAGGTCAGTAAGTAAGAAGAGCCATCAG
 5200
 GlyGluLysLeuLeuSerCysCysAsnTyrProArgAlaHisArgAlaGlnValProSer
 AGGGGAAAAGTTATTGTCTCTGCTGCAATTATCCCCGAGCTCATAGAGCCCAGGTACCGTC
 LeuGlnPheLeuAlaLeuValValValGlnGlnAsnAspArgProGlnArgAspSerThr
 MetThrAspProArgGluThrValPro
 ACTTCAATTTCTGGCCTTAGTGGTACTGCAACAAAATGACAGACCCAGAGACAGTAC
 5300
 ThrArgLysGlnArgArgAspTyrArgArgGlyLeuArgLeuAlaLysGlnAspSer
 ProGlyAsnSerGlyGluGluThrIleGlyGluAlaPheAlaTrpLeuAsnArgThrVal
 CACCAGGAAACAGCGCGGAAGAGACTATCGGAGAGGCCTTCGCCTGGCTAAACAGGACAG
 5400
 ArgSerHisLysGlnArgSerSerGluSerProThrProArgThrTyrPheProGlyVal
 GluAlaIleAsnArgGluAlaValAsnHisLeuProArgGluLeuIlePheGlnValTrp
 TAGAAGCCATAAACAGAGAAGCAGTGAATCACCTACCCGAGAAGTATTTTCCAGGTGT
 (fig.1A-forts: 4)

AlaGluValLeuGluIleLeuAla
 GlnArgSerTrpArgTyrTrpHisAspGluGlnGlyMetSerGluSerTyrThrLysTyr
 GGCAGAGGTCCTGCAGATACTGGCATGATGAACAAGGGATGTCAGAAAGTTACACAAAGT
 5500
 ArgTyrLeuCysIleIleGlnLysAlaValTyrMetHisValArgLysGlyCysThrCys
 ATAGATATTTGTGCATAATACAGAAAGCAGTGTACATGCATGTTAGGAAAGGGTGTACTT
 LeuGlyArgGlyHisGlyProGlyGlyTrpArgProGlyProProProProProProPro
 CCCTGGGGAGGGGACATGGCCAGGAGCGTGGAGACCAGGCCCTCCTCCTCCTCCCCCTC
 5600
 MetAlaGluAlaProThrGluLeuProProValAspGlyThrProLeu
 GlyLeuVal***
 CAGGTCTGGTCTAATGGCTGAAGCACCAACAGAGCTCCCCCGGTGGATGGGACCCCACT
 ArgGluProGlyAspGluTrpIleIleGluIleLeuArgGluIleLysGluGluAlaLeu
 GAGGGAGCCAGGGATGAGTGGATAATAGAAATCTTGAGAGAAATAAAGAAGAAGCTTT
 LysHisPheAspProArgLeuLeuIleAlaLeuGlyLysTyrIleTyrThrArgHisGly
 MetGlu
 AAAGCATTTTGACCCTCGCTTGCTAATTGCTCTTGGCAAATATATCTATACTAGACATGG
 5800
 AspThrLeuGluGlyAlaArgGluLeuIleLysValLeuGlnArgAlaLeuPheThrHis
 ThrProLeuLysAlaProGluSerSerLeuLysSerCysAsnGluProPheSerArgThr
 AGACACCCTTGAAGGCCCGAGAGAGCTCATTAAAGTCTGCAACGAGCCCTTTTCACGCA
 PheArgAlaGlyCysGlyHisSerArgIleGlyGlnThrArgGlyGlyAsnProLeuSer
 SerGluGlnAspValAlaThrGlnGluLeuAlaArgGlnGlyGluGluIleLeuSerGln
 CTTCAGAGCAGGATGTGGCCACTCAAGAATTGGCCAGACAAGGGGAGGAAATCCTCTCTC
 5900
 AlaIleProThrProArgAsnMetGln
 LeuTyrArgProLeuGluThrCysAsnAsnSerCysTyrCysLysArgCysCysTyrHis
 AGCTATACCGACCCTAGAAACATGCAATAACTCATGCTATTGTAAGCGATGCTGCTACC
 6000
 MetAsnGluArgAlaAsp
 CysGlnMetCysPheLeuAsnLysGlyLeuGlyIleCysTyrGluArgLysGlyArgArg
 ATIGTCAGATGTGTTTTCTAAACAAGGGGCTCGGGATATGTTATGAACGAAAGGGCAGAC
 GluGluGlyLeuGlnArgLysLeuArgLeuIleArgLeuLeuHisGlnThrSerGluTyr
 Met
 ArgArgThrProLysLysThrLysThrHisProSerProThrProAspLys
 GAAGAAGGACTCCAAAGAAAATAAGACTCATCGTCTCCTACACCAGACAAGTGAGTAT
 6100
 AspGluSerAlaAlaTyrCysHisPheIleSer
 MetAsnGlnLeuLeuIleAlaIleLeuLeuAlaSerAlaCysLeuValTyrCysThrGln
 GATGAATCAGCTGCTTATTGCCATTTTATTAGCTAGTGCTTGCTTAGTATATTGCACCCA
 TyrValThrValPheTyrGlyValProThrTrpLysAsnAlaThrIleProLeuPheCys
 ATATGTAAGTGTGTTTTCTATGGCGTACCCACGTGGAAAAATGCAACCATTCCCCTCTTTG
 6200
 AlaThrArgAsnArgAspThrTrpGlyThrIleGlnCysLeuProAspAspAspTyr
 TGCAACCAGAAATAGGGATACTTGGGGAACCATACAGTGCTTGCTGACAATGATGATTA
 6300
 GlnGluIleThrLeuAsnValThrGluAlaPheAspAlaTrpAsnAsnThrValThrGlu
 TCAGGAAATAACTTTGAATGTAACAGAGGCTTTTGATGCATGGAATAATACAGTAACAGA
 GlnAlaIleGluAspValTrpHisLeuPheGluThrSerIleLysProCysValLysLeu
 ACAAGCAATAGAAGATGTCTGGCATCTATTTCGAGACATCAATAAAACCATGTCTCAAAC
 6400
 *(fig.1A-forts. 5)

ThrProLeuCysValAlaMetLysCysSerSerThrGluSerSerThrGlyAsnAsnThr
 AACACCTTTATGCTAGCAATGAAATGCAGCAGCACACAGACCACCACACCGAACAACAG
 ThrSerLysSerThrSerThrThrThrThrThrProThrAspGlnGluGlnGluIleSer
 AACCTCAAAGAGCACAAGCACAACCACAACCACACCACAGACCAGGAGCAAGAGATAAG
 6500
 GluAspThrProCysAlaArgAlaAspAsnCysSerGlyLeuGlyGluGluGluThrIle
 TGAGGATACTCCATGCCGACGGCAGACAACCTGCTCAGGATTGGGAGAGGAAGAAACGAT
 6600
 AsnCysGlnPheAsnMetThrGlyLeuGluArgAspLysLysLysGlnTyrAsnGluThr
 CAATTGCCAGTTCAATATGACAGGATTAGAAAGAGATAAGAAAAAACAGTATAATGAAAC
 TrpTyrSerLysAspValValCysGluThrAsnAsnSerThrAsnGlnThrGlnCysTyr
 ATGGTACTCAAAGATGTGGTTTGTGAGACAAATAATAGCACAAATCAGACCCAGTGTTA
 6700
 MetAsnHisCysAsnThrSerValIleThrGluSerCysAspLysHisTyrTrpAspAla
 CATGAACCATIGCAACACATCAGTCATCACAGAATCATGTGACAAGCACTATTGGGATCC
 IleArgPheArgTyrCysAlaProProGlyTyrAlaLeuLeuArgCysAsnAspThrAsn
 TATAAGGTTTAGATACTGTGCACCACCGGGTTATGCCCTATTAAGATGTAATGATACCAA
 6800
 TyrSerGlyPheAlaProAsnCysSerLysValValAlaSerThrCysThrArgMetMet
 TTATTACAGGCTTTGCACCCAACCTGTTCTAAAGTAGTAGCTTCTACATGCACCAGGATGAT
 6900
 GluThrGlnThrSerThrTrpPheGlyPheAsnGlyThrArgAlaGluAsnArgThrTyr
 GAAACGCAAACCTCCACATGGTTTGGCTTTAATGCCACTAGAGCAGAGAATAGAACATA
 IleTyrTrpHisGlyArgAspAsnArgThrIleIleSerLeuAsnLysTyrTyrAsnLeu
 TATCTATTGGCATGGCAGAGATAATAGAACTATCATCAGCTTAAACAAATATTATAATCT
 7000
 SerLeuHisCysLysArgProGlyAsnLysThrValLysGlnIleMetLeuMetSerGly
 CAGTTTGCATTGTAAGAGGCCAGGGAATAAGACAGTGAAACAAATAATGCTTATGTCAGC
 HisValPheHisSerHisTyrGlnProIleAsnLysArgProArgGlnAlaTrpCysAsp
 ACATGTGTTTCACTCCACTACCAGCCGATCAATAAAAGACCCAGACAAGCATGGTGTG
 7100
 PheLysGlyLysTrpLysAspAlaMetGlnGluValLysGluThrLeuAlaLysHisPro
 GTTCAAAGGCAAATGGAAAGACGCCATGCAGGAGGTGAAGGAAACCTTGCAAAACATCC
 7200
 ArgTyrArgGlyThrAsnAspThrArgAsnIleSerPheAlaAlaProGlyLysGlySer
 CAGGTATAGAGGAACCAATGACACAAGGAATATTAGCTTTGCAGCGCCAGGAAAAGGCTC
 AspProGluValAlaTyrMetTrpThrAsnCysArgGlyGluPheLeuTyrCysAsnMet
 AGACCCAGAAGTAGCATACATGTGGACTAACTGCAGAGGAGAGTTTTCTACTGCAACAT
 7300
 ThrTrpPheLeuAsnTrpIleGluAsnLysThrHisArgAsnTyrAlaProCysHisIle
 GACTTGGTTCTCAATTGGATAGAGAATAAGACACACCCGCAATTATGCACCCGTGCCATAT
 LysGlnIleIleAsnThrTrpHisLysValGlyArgAsnValTyrLeuProProArgGlu
 AAAGCAAATAATTAACACATGGCATAAGGTAGGGAGAAATGTATATTTGCTCCCAGGGA
 7400
 GlyGluLeuSerCysAsnSerThrValThrSerIleIleAlaAsnIleAspTrpGlnAsn
 AGGGGAGCTGTCTGCAACTCAACAGTAACCAGCATAATTGCTAACATTGACTGGCAAAA
 7500
 AsnAsnGlnThrAsnIleThrPheSerAlaGluValAlaGluLeuTyrArgLeuGluLeu
 CAATAATCAGACAAACATTACCTTTAGTGCAGAGGTGGCAGAACTATACAGATTGCAGTT
 GlyAspTyrLysLeuValGluIleThrProIleGlyPheAlaProThrLysGluLysArg
 GCGAGATTATAAATTGGTAGAAATAACACCAATTGGCTTCCACCTACAAAAGAAAAAG
 7600
 (fig.1A-forts. 6)

TyrSerSerAlaHisGlyArgHisThrArgGlyValPheValLeuGlyPheLeuGlyPhe
 ATACTCTCTGCTCACGGGAGACATACAAGAGGTGTGTTCTGCTAGGGTTCCTGGGTTT

LeuAlaThrAlaGlySerAlaMetGlyAlaAlaSerLeuThrValSerAlaGlnSerArg
 TCTCGCAACAGCAGGTTCTGCAATGGGCGCGGCTCCCTGACCGTGTCCGGCTCAGTCCCC

7700

ThrLeuLeuAlaGlyIleValGlnGlnGlnGlnGlnLeuLeuAspValValLysArgGln
 GACTTTACTGGCCGGGATAGTGCAGCAACAGCAACAGCTGTTGGACGTGCTCAAGAGACA

7800

GlnGluLeuLeuArgLeuThrValTrpGlyThrLysAsnLeuGlnAlaArgValThrAla
 ACAAGAACTGTTGGCACTGACCGTCTGGGGAACGAAAAACCTCCAGGCAAGAGTCACTGC

IleGluLysTyrLeuGlnAspGlnAlaArgLeuAsnSerTrpGlyCysAlaPheArgGln
 TATAGAGAAGTACCTACAGGACCAGCGCGGCTAAATTCATGGGGATGTGCGTTTACACA

7900

ValCysHisThrThrValProTrpValAsnAspSerLeuAlaProAspTrpAspAsnMet
 AGTCTGCCACACTACTGTACCATGGGTTAATGATTCCCTTAGCACCTGACTGGGACAATAT

ThrTrpGlnGluTrpGluLysGlnValArgTyrLeuGluAlaAsnIleSerLysSerLeu
 GACGTGGCAGGAATGGGAAAAACAAGTCCGCTACCTGGAGGCAAATATCAGTAAAAGTTT

8000

GluGlnAlaGlnIleGlnGlnGluLysAsnMetTyrGluLeuGlnLysLeuAsnSerTrp
 AGAACAGGCACAAATTCAGCAAGAGAAAAATATGTATGAACTACAAAAATTAATAGCTG

8100

AspIlePheGlyAsnTrpPheAspLeuThrSerTrpValLysTyrIleGlnTyrGlyVal
 GGATATTTTTGGCAATTGGTTTTGACTTAACCTCCTGGGTCAAGTATATTCAATATGGAGT

LeuIleIleValAlaValIleAlaLeuArgIleValIleTyrValValGlnMetLeuSer
 Val

GCTTATAATAGTAGCAGTAATAGCTTTAAGAATAGTGATATATGTAGTACAAATGTTAAG

8200

AlaCysPheLeuPheProProArgLeuTyrProThrAsp
 ArgLeuArgLysGlyTyrArgProValPheSerSerProProGlyTyrIleGlnGlnIle
 GlyLeuGluArgAlaIleGlyLeuPheSerLeuProProProValIleSerAsnArgSer
 TAGGCTTAGAAAGGGCTATAGGCTGTTTTCTCTTCCCCCCCCGGTTATATCCAACAGAT

ProTyrProGlnGlyProGlyThrAlaSerGlnArgArgAsnArgArgArgArgTrpLys
 HisIleHisLysAspArgGlyGlnProAlaAsnGluGluThrGluGluAspGlyGlySer
 IleSerThrArgThrGlyAspSerGlnProThrLysLysGlnLysLysThrValGluAla
 CCATATCCACAAGGACCGGGACGCCAGCCAACGAAGAAACAGAAGAAGACCGTGGAAG

8300

GlnArgTrpArgGlnIleLeuAlaLeuAlaAspSerIleTyrThrPheProAspProPro
 AsnGlyGlyAspArgTyrTrpProTrpProIleAlaTyrIleHisPheLeuIleArgGln
 ThrValGluThrAspThrGlyProGlyArg

CAACGGTGGAGACAGATACTGGCCCTGGCCGATAGCATATATACATTTCTGATCCGCCA

8400

AlaAspSerProLeuAspGlnThrIleGlnHisLeuGlnGlyLeuThrIleGlnGluLeu
 LeuIleArgLeuLeuThrArgLeuTyrSerIleCysArgAspLeuLeuSerArgSerPhe
 GCTGATTCGCCTCTTGACCAGACTATACAGCATCTGCAGGGACTTACTATCCAGGAGCTT

ProAspProProThrHisLeuProGluSerGlnArgLeuAlaGluThr
 LeuThrLeuGlnLeuIleTyrGlnAsnLeuArgAspTrpLeuArgLeuArgThrAlaPhe
 CCTGACCTCCAACCTCATCTACCAGAATCTCAGAGACTGGCTGAGACTTAGAACAGCCTT

8500

LeuGlnTyrGlyCysGluTrpIleGlnGluAlaPheGlnAlaAlaAlaArgAlaThrArg
 MetGlyAlaSerGlySerLysLysHisSerArgProProArgGlyLeuGlnGlu
 CTTGCAATATGGGTGGGAGTGGATCCAAGAAGCATTCCAGGCCCGCCGGAGGGCTACAAG

(fig.1A- forts. 7)

FIG 1B

AGTCGCTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCTCCAGCACTAGCAG
 GTAGAGCCTGGGTGTTCCCTGCTAGACTCTCACCAGCACTGGCCGGTCTGGCCAGAGT
 GGCTCCACGCTTGCTTGCTTAAAGACCTCTTCAATAAAGCTGCCATTTAGAAGTAAGCTA
 GTGTGTGTTCCCATCTCTCCTAGTCGCCGCTGGTCAACTCGGTACTCGGTAATAAAAAG
 ACCCTGGTCTGTTAGGACCCTGGTCTGTTAGGACCCTTCTGCTTTGGGAAACCGAAGCA
 GGAAAATCCCTAGCAGATTGGCGCCCGAACAGGGACTTGAAGGAGAGTGAGAGACTCCTG
 AGTACGGCTGAGTGAAGGCAGTAAGGGCGGCAGGAACCAACCACGACGGAGTGCTCCTAG
 AAAGGCGCGGGTGGTACCAGACGGCGTGAGGAGCGGGAGAGAAGAGGCCTCCTGGTTG
 CAGGTAAGTGCAACACAAAAAGGAAATAGCTGTCTTTATCCAGGAAGGGATAATAAGAT
 GAGDMETGLYALAARGASNSERVALLEUSERGLYLYSLYSALAAASPLULEUGLU
 AGAGTGGGAGATGGGCGCGAGAACTCCGTCTTGTGAGGGAAGAAAGCAGATGAATTAGA
 LYSILEARGLEUARGPROGLYGLYLYSLYSLYSTRMETLEULYSHISVALVALTRPALA
 AAAAATTAGACTACGACCCGGCGGAAAGAAAAGTACATGTTGAAGCATGTAGTATGGGC
 ALAASNGLULEUASPARGPHEGLYLEUALAGLUSERLEULEUGLUASNLYSGLUGLYCYS
 AGCAAATGAATTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAACAAGAAGGATG
 GLNLYSILEUSERVALLEUALAPROLEUVALPROTHRGLYSERGLUASNLEULYSSER
 TCAAAAAATACTTTCCGGTCTTAGCTCCATTAGTGCCAACAGGCTCAGAAAATTTAAAAAG
 LEUTYRASNTHRVALCYSVALILETRPCYSILEHISALAGLUGLULYSVALLYSHISTHR
 CCTTATAATACTGTCTGCGTCATCTGGTGCATTACGCCAGAAGAGAAAGTGAACACAC
 GLUGLUALALYSGLNILEVALGLNARGHISLEUVALMETGLUTHRGLYTHRAGLUTHR
 TGAGGAAGCAAAACAGATAGTGCAGAGACACCTAGTGATGGAAACAGGAACAGCAGAAAC
 METPROLYSTHRSERARGPROTHRALAPROPHE SERGLYARGGLYGLYASNTYRPROVAL
 TATGCCAAAAACAAGTAGACCAACAGCACCATTTAGCGGCAGAGGAGGAATTACCCAGT
 GLNGLNILEGLYGLYASNTYRTHRHSLEUPROLEUSERPROARGTHRLEUASNALATRP
 ACAACAAATAGGTGGTAACTATACCCACCTACCATTAAGCCCGAGAACATTAATGCCTG
 VALLYSLEULEGLUGLULYSLYSPHEGLYALAGLUVALVALSERGLYPHEGLNALALEU
 GGTAATAATAAGAGGAGAAGAAATTTGGAGCAGAAGTAGTGTCAGGATTTACGGCACT
 SERGLUGLYCYSLEUPROTYRASPILEASNGLNMETLEUASNLCYSVALGLYASPHISGLN
 GTCAGAAGGCTGCCTCCCTATGACATTAATCAGATGTTAAATTGTGTGGGAGACCATCA
 ALAALAMETGLNILELEARGASPILELEASNGLUGLUALAALASPTRPASPLEUGLN
 AGCGGCTATGCAGATCATCAGAGATATTATAAATGAGGAGGCTGCAGATTTGGGACTTGCA
 HISPROGLNGLNALAPROGLNGLNGLYGLNLEUARGGLUPROSERGLYSERASPILEALA
 GCACCAACAACAAGCTCCACAACAAGGACAGCTTAGGGAGCCGTCAGGATCAGATATTGC
 GLYTHRTHRSETRHRVALGLUGLUGLNILEGLNTRPMETTYRARGGLNGLNASNPROILE
 AGGAACAACACTAGTACAGTAGAAGAACAATCCAGTGGATGTACAGACAACAGAACCCCAT
 1300

FIG. 1B-

PROVALGLYASNILETYRARGARGTRPILEGLNLEUGLYLEUGLNLYSCYSVALARGMET
 ACCAGTAGGCAACATTTACAGGAGATGGATCCAACTGGGGTTGCAAAAATGTGTCAGAAT
 TYRASNPROTHRASNILELEUASPVALLYSGLNGLYPROLYSGLUPROPHEGLNSERTYR
 GTATAACCCAAACAACATTTCTAGATGTAAAAACAAGGGCCAAAAGAGCCATTTACAGAGCTA
 1400
 VALASPARGPHEITYRLYSSERLEUARGALAGLUGLNTHRASPPODALAVALLYASNTRP
 TGTAGACAGGTTCTACAAAAGTTTAAGAGCAGAACAACAGATCCAGCAGTAAAGAATTG
 1500
 METTHRGLNTHRLEULEULEGLNASNALAASNPROASPCYSLYSLEUVALLEULYSGLY
 GATGACTCAAACACTGCTGATTCAAAAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGG
 LEUGLYTHRASNPROTHRLEUGLUGLUMETLEUTHRALACYSGLNGLYVALGLYGLYPRO
 GCTGGGTACGAATCCCACCCTAGAAGAAATGCTGACGGCCTGTCAAGGAGTAGGGGGGCC
 1600
 GLYGLNLYSALAARGLEUMETALAGLUALALEULYSGLUALALEUALAPRODALPROILE
 AGGACAGAAGGCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCACCAGGCCCAAT
 POLVALLEUGLULEUTRP
 PROPHEALAAALAGLNGLNLYSGLYPROARGLYSPTILELYSCYSTRPASNLYSGLY
 CCTTTTGCAGCAGCCCAACAGAAGGGACCAAGAAAGCCAATTAAGTGTGGAAATGTGG
 1700
 GLUGLYARGTHRLEUCYSLYSALAMETGLNSERPROLYSLYSTHRGLYMETLEUGLUMET
 LYSGLUGLYHISSERALAARGGLNLYSARGALAPROARGARGGLNGLYCYSTRPLYSCYS
 GAAGGAAGGACACTCTGCAAGGCAATGCAGAGCCCAAGAAGACAGGGATGCTGGAATG
 1800
 TRPLYSANGLYPROCYSTYRGLYGLNMETPROLYSGLNTHRGLYGLYPHEPHEARGPRO
 GLYLYSMETASPHISVALMETALALYSYSPROASNARGGLNALAGLYPHELEUGLYLEU
 TGGAAAAATGGACCATGTTATGGCCAAATGCCCAAACAGACAGGGGGTTTTTAGGCCCT
 TRPPROLEUGLYLYSGLUALAPROGLNPHEPROHISGLYSERSERALASERGLYALAAASP
 GLYPROTRPGLYLYSLSYSPROARGASNPHEPROMETALAGLNVALHISGLNGLYLEUTHR
 TGGCCCTTGGGAAAGAAGCCCCGCAATTTCCCATGGCTCAAGTGCATCAGGGGCTGAC
 1900
 ALAASNLYSERSERPROARGARGTHRSERCYSGLYSERALALYSGLULEUHSALALEUGLY
 PROTHRALA PROPROGLUGLUPROALAVALASPLEULEULYSASNTYRMETHISLEUGLY
 GCGAAGTCTCCCCAGAAGAACCAGCTGTGGATCTGCTAAAGAACTACATGCACTTGGG
 GLNALAALAGLUARGLYSGLNARGGLUALALEUGLNGLYGLYASPARGGLYPHEALAAAL
 LYSGLNGLNARGGLUSERARGGLYLYSPTYRLYSGLUVALTHRGLUASPLEULEUHS
 CAAGCAGCAGAGAGAAAGCAGAGGGAAGCCTTACAAGGAGGTGACAGAGGATTTGCTGCA
 2000
 PROGLNPHESERLEUTRARGARGPROVALVALTHRALAHISILEGLUGLYLNPROVAL
 LEUASN SERLEUPHEGLYGLYASPLN
 CCTCAATCTCTCTTTGGAGGAGACCAGTAGTCACTGCTCATATTGAAGGACAGCCTGTA
 2100
 GLUVALLEULEUASPTHRLYALAASPA SP SERILEVALTHRGLYILEGLULEUGLYPRO
 GAAGTATTATTAGATACAGGGGCTGATGATTCTATTGTAACAGGAATAGAGTTAGGTCCA
 HISTYRTHRPROLYSILEVALGLYGLYILEGLYGLYPHEILEASNTHRLYSGLUTYRLYS
 CATTATACCCAAAAAATAGTAGGAGGAATAGGAGGTTTTATTAATACTAAAGAATACAAA
 2200
 ASNYALGLUILEGLUVALLEUGLYLYSARGILELYSGLYTHRILEMETTHRGLYASPTH
 AATGTAGAAATAGAAGTTTTAGGCCAAAAGGATTAAGGGACAATCATGACAGGGGACACC
 PROILEASNILEPHEGLYARGASNLEULEUTHRALALEUGLYMETSERLEUASNLEUPRO
 CCGATTAACATTTTTGGTAGAAATTTACTAACAGCTCTGGGGATGCTCTCTAAATCTTCCC
 2300
 ILEALALYSVALGLUPROVALLYSERPROLEULYSPROGLYLYSASPLYPROLYSLEU
 ATAGCTAAGGTAGAGCCTGTAAGTCGCCCTTAAAGCCAGGAAAGGATGGACCAAAATTG
 2400
 LYSGLNTRPPROLEUSERLYSGLULYSILEVALALEUARGGLUILECYSGULYSMET
 AAGCAGTGGCCATTATCAAAAGAAAAGATAGTTGCATTAAGAGAAATCTGTGAAAAGATG

(fig. 13-forts. 1)

GLULYSASPLYGLNLEUGLUGLUALAPROPROTHRASNPROTYRASNTHRPROTHRPH
 GAAAAAGATGGTCAGTTGGAGGAAGCTCCCCGACCAATCCATATAACACCCCCACATTT
 2500
 ALAILELYSLYSLYASPLYASNLVSTRPARGMETLEUILEASPPHEARGGLULEUASN
 GCTATAAAGAAAAAGGATAAAAACAAATGGAGAATGCTGATAGATTTTAGGGAACTAAAT
 ARGVALTHRGLNASPPHE THRGLUVALGLNLEUGLYILEPROHISPROALAGLYLEUALA
 AGGGTCACTCAAGACTTTACGGAAGTCCAAT TAGGAAT ACCACACCCTGCAGGACTAGCA
 2600
 LYSARGLYSARGILETHRVALLEUASPILEGLYASPALATYRPHESERILEPROLEUASP
 AAAAGGAAAAAGGATTACAGTACTGGATATAGGTGACGCATATTTCTCTATACCTCTAGAT
 2700
 GLUGLUPHEARGGLN TYRTHRALAPHE THRLEUPROSERVALASNASNALAGLUPROGLY
 GAAGAATTTAGGCAGTACACTGCCTTTACTTTACCATCAGTAAATAATGCAGAGCCAGGA
 LYSARGTYRILETYRLYSVALLEUPROGLNGLYTRPLYSGLYSERPRDALAILEPHEGLN
 AAACGATACATTTATAAGGTTCTGCCTCAGGCATGGAAGGGGTCACCAGCCATCTTCCAA
 2800
 TYRTHRMETARGHISVALLEUGLUPROPHEARGLYSALAASNPROASPVALTHRLEUVAL
 TACACTATGAGACATGTGCTAGAACCCTTCAGGAAGGCAATCCAGATGTGACCTTAGTC
 GLN TYR METASPA S P I L E L E U I L E A L A S E R A S P A R G T H R A S P L E U G L U H I S A S P A R G V A L
 CAGTATATGGATGACATCTTAATAGCTAGTGACAGGACAGACCTGGAACATGACAGGGTA
 2900
 VALLEUGLNL EULYSGLULEUL EUASNSERILEGLYPHESERSERPROGLUGLULYSPHE
 GTTTTACAGTAAAAGAACTCTTAAATAGCATAGGGTTTTTCATCCCCAGAAGAGAAATTC
 3000
 GLNLYSASPPROPROPHEGLNTRPMETGLYTYRGLULEUTRPPROTHRLYSTRPLYSLEU
 CAAAAAGATCCCCCATTTCAATGGATGGGGTACGAATTGTGGCCGACAAAATGGAAGTTG
 GLNLYSILEGLULEUPROGLNARGGLUTHRTRP THRVALASNASPILEGLNLYSLEUVAL
 CAAAAGATAGAGTTGCCACAAGAGAGACCTGGACAGTGAATGATATACAGAAGTTAGTA
 3100
 GLYVALLEUASNTRPALAALAGLNILETYRPROGLYILELYSTHRLYSHISLEUCYSARG
 GGAGTATTAATTTGGGCAGCTCAAATTTATCCAGGTATAAAAACCAACATCTCTGTAGG
 LEUILEARGGLYLYSMETTHRLEUTHRGLUGLUALGLNTRP THRGLUMETALAGLUALA
 TTAATTAGAGGAAAAATGACTCTAACAGAGGAAGTTCACTGGACTGAGATGGCAGAAGCA
 3200
 GLUTYRGLUGLUASNLYSILEILELEUSERGLNGLUGLUGLYCYSTYR TYRGLNGLU
 GAATATGAGGAAAAATAAATAATTTCTCAGTCAGGAACAAGAAGGATGTTATTACCAAGAA
 3300
 SERLYS PROLEUGLUALATHRVALILELYSSERGLNASPASNGLNTRP SERTYRLYSILE
 AGCAAGCCATTAGAAGCCACGGTGATAAAGAGTCAGGACAATCAGTGGTCTTATAAAAT
 HISGLNGLUASPLYSILELEULYSVALGLYLYSPHEALALYSILELYSANTHRHISTHR
 CACCAAGAAGACAAAATACTGAAAGTAGGAAAATTTGCAAGATAAAGAATACACATACC
 3400
 ASNGLYVALARGLEULEUALAHISVALILEGLNLYSILEGLYLYSGLUALAILEVALILE
 AATGGAGTTAGACTATTAGCACATGTAATACAGAAAATAGGAAAGGAAGCAATAGTGATC
 TRPGLYGLNYALPROLYSPHEHISLEUPROVALGLULYSASPYALTRPGLUGLNTRPTRP
 TGGGCACAGGTCCC AAAATCCACTTACCAGTTGAGAAGGATGTATGGGAACAGTGGTGG
 3500
 THRASPTYRTRPGLNYALTHRTRP ILEPROGLUTRPASPPHE ILESERTHRPROPROLEU
 ACAGACTATTGGCAGGTAACCTGGATACCGGAATGGGATTTTCATCTCAACACCACCATTA
 3600
 VALARGLEUVALPHEASNLEUVALLYASPPROILEGLUGLYGLUGLUTHRTYRTRVAL
 GTAAGATTAGTCTTCAATCTAGTGAAGGACCCTATAGAGGGAGAAGAAACCTATTATGTA
 ASPGLYSERCYS SERLYSGLNSERLYSGLUGLYLYSALAGLYTYRILETHRASPARGGLY
 GATGGATCATGTAGTAAACAGTCAAAGAAGGAAAAGCAGGATATATCACAGACAGGGCC

(fig. 1B-forts. 2)

3700
 LYSASPLYSVALLYSVALLÉUGLUGLNTHRTHRASNGLNGLNALAGLULÉUGLUALAPHE
 AAAGACAAGGTAAGTGTAGAACAGACTACTAATCAACAAGCAGAATTGGAAGCATT
 LEUMETALALEUTHRASPSERGLYPROLYSALAASNILEILEVALASPSERGLNTRYVAL
 CTCATGGCATTGACAGACTCAGGGCCAAAGGCAAATATTATAGTAGACTCACAAATGT
 3800
 METGLYILEILETHRGLYCYSROTHRGLUSERGLUSERARGLEUVALASNGLNILEILE
 ATGGGAATAATAACAGGATGCCCTACAGAATCAGAGAGCAGGCTAGTTAACCAAATAA
 3900
 GLUGLUMETILELYSLYSTRGLUILETYRVALALATRPVALPROALAHISLYSGLYILE
 GAAGAAATGATCAAAAAGACAGAAATTTATGTGGCATGGGTACCAGCACACAAGGTATA
 GLYGLYASNGLNGLUILEASPHISLEUVALSERGLNGLYILEARGGLNVALLEUPHELEU
 GGAGAAACCAAGAAATAGACCACCTAGTTAGTCAAGGGATTAGACAAGTTCCTCTCTG
 4000
 GLULYSILEGLUPROALAGLNGLUGLUHISSELYSTYRHISSEASNILELYSGLULEU
 GAAAAGATAGAGCCAGCACAAGAAGAACATAGTAAATACCATAGTAACATAAAGAATTG
 VALPHELYSPHEGLYLEUPROARGLEUVALALALYSGLNILEVALASPTHRCYSASPLYS
 GTATTCAAATTTGGATTACCCAGACTAGTGGCCAAACAGATAGTAGACACATGTGATAAA
 4100
 CYSHISGLNLYSGLYLUALAILEHISGLYGLNVALASNSERASPLEUGLYTHRTRPGLN
 TGTCAATCAAAAAGGAGAAGCTATACATGGGCAGGTAAATTCAGACCTAGGGACTGGCAA
 4200
 METASPCYSTHRHISLEUGLUGLYLYSILEVALILEVALALAVLHISVALALASERGLY
 ATGGATTGTACCCATCTAGAGGGAAAAATAGTCATAGTTGCAGTACATGTAGCTAGTGA
 PHEILEGLUALAGLUVALILEPROGLNGLUTHRGLYARGGLNTHRALALEUPHELEULEU
 TTCATAGAAGCAGAAGTAATCCACAAGAACAGGAAGACAGACAGCCTATTCTGTTA
 4300
 LYSLEUALASERARGTRPPOILETHRHSLEUHISTRASPSNGLYALASNPHEALA
 AAATTGGCAAGCAGATGGCCTATTACACATCTGCACACAGATAATGGTGCTAACYTTGCT
 SERGLNGLUVALLYSMETVALALATRPTRPALAGLYILEGLUHISTHRPHEGLYVALPRO
 TCGCAAGAAGTAAAGATGGTTCATGGTGGCCAGGGATAGAGCACACCTTTGGGGTACCA
 4400
 TYRASNPROGLNSERGLNGLYVALVALGLUALAMETASNHISHISLEULYSANGLNILE
 TACAATCCACAGAGTCAGGGAGTAGTGAAGCAATGAATCACACCTGAAAAATCAAATA
 4500
 ASPARGILEARGGLUGLNALAASNSEVALGLUTHRILEVALLEUMETALAVLHISCVS
 GATAGAATCAGGGAACAAGCAATTCAGTAGAAACCATAGTATTAATGGCAGTTCATTGC
 METASNPHELYSARGARGGLYGLYILEGLYASPMETTHRPROALAGLUARGLEUILEASN
 ATGAATTTAAAAGAAGGGGAGGAATAGGGGATATGACTCCAGCAGAAAGATTAATTAAC
 4600
 NETILETHRTHRGLUGLNGLUILEGLNPHEGLNGLNSERLYSASNSERLYSPHELYSASN
 ATGATCACTACAGAACAAGAAATACAATTTCAACAATCAAAAACTCAAATTTAAAAAT
 PHEARGVALTYRTRYRARGGLUGLYARGASPLNLEUTRPLYSGLYPROGLYGLULEULEU
 TTTCCGGTCTATTACAGAGAAGGCAGAGATCAGCTGTGGAAGGGACCCGGTAGCTATTG
 4700
 TRPLYSGLYGLUGLYALAVLILELEULYSVALGLYTHRASPILELYSVALVALPROARG
 TGGAAAGGGGAAGGAGCAGTCATCTTAAGGTAGGAACAGACATTAAGGTAGTACCAGG
 4800
 ARGLYSALALYSILEILELYSASPTYRGLYGLYGLYLYSGLUMETASPSERSERSERHIS
 QMETGLUGLUGLULYSARGTRPILEVALVALPROTHR
 AGAAAGGCTAAAATTATCAAGATTATGGAGGAGGAAAAGAGATGGATAGTAGTCCAC
 METGLUASPTHRGLYGLUALAARGGLUYALALA
 TRPARGILEPROGLUARGLEUGLUARGTRPHISERLEUILELYSTYRLEULYSTYRYS
 ATGGAGGATACCGGAGAGGCTAGAGAGGTGCCATAGCCTATAAAATATTGAAATATA
 4900

(fig. 1B- forts. 3)

THRLYSASPLEUGLNLYSALACYSTYRVAL PROMISHISLYSVALGLYTRPALATRPTRP
 AACTAAAGATCTACAAAAGGCTTGCTATGTGCCCATCATAAGGTCGGATGGGCATGGTG
 THRCYSSERARGVALILEPHEPROLEUGLNGLUGLYSERHISLEUGLUVALGLNGLYTYR
 GACCTGCAGCAGAGTAATCTCCCACTACAGGAAGGAAGCCATTTAGAAGTACAAGGTA
 5000
 TRPASNLEUTHRPROGLUARGGLYTRPLEUSERTHRTYRALAVALARGILETHRTRPTYR
 TTGGAATTTGACACCAGAAAGAGGGTGGCTCAGTACTTATGCAGTGAGGATAACCTGGTA
 5100
 SERLYSASPHEPTRPTHRASPVALTHRPROGLUTYRALAASP ILELEULEUHISSERTHR
 CTCAAAGGACTTTTGGACAGATGTAACACCAGAATATGCAGATATTTTACTGCATAGCAC
 TYRPHEPROCYSPHETHRALAGLYGLUVALARGARGALALEARGGLYGLUARGLEULEU
 TTATTTCCCTTGCTTTACAGCGGGAGAAAGTGAGAAGGGCCATCAGGGGAGAACCCTGCT
 5200
 SERCYSCYSARGPHEPROARGALAHISLYSHISGLNVALPROSERLEUGLNTRYLEUALA
 GTCTTGCTGCAGGTTCCCAAGAGCTCATAAGCACAGGTACCAAGTCTACAGTACTTAGC
 LEUARGVALVALSERHISVALARGSERGLNGLYGLUASNPROTHRTRPLYSGLNTRPARG
 X METSERASPPROARGGLUARGILEPROPROGLYASNSEGLYGLU
 ACTGAGAGTAGTAAGTCATGTCAGATCCCAGGGAGAGAATCCCACCTGGAACAGTGGAG
 5300
 ARGASPASNARGARGSERLEUARGVALALALYSGLNASNSERARGGLYASPLYSGLNARG
 GLUTHRILEGLYGLUALAPHEGLUTRPLEUASNARGTHRVALGLUGLUILEASNARGGLU
 AAGAGACAATAGGAGAAGCCTTCGAGTGGCTAAACAGAACAGTAGAGGAGATAAACAGAG
 5400
 GLYGLYLYSPPROPROTHRGLUGLYALAA SNPHEPROGLYLEUALALYSVALLEUGLYILE
 ALAVALASNHISLEUPROARGGLULEUILEPHEGLNVALTRPGLNARGSERTRPGLUTYR
 AGGGGTAACCACCTACCGAGGGAGCTAATTTCCAGGTTTGGCAAAGGCTTGGGAAT
 LEUALA
 TRPHISASPLUGLNGLYMETSERGLNSERTYRTHRLYSTYRARGTYRLEUCYSLEUILE
 ACTGGCATCATGAACAAGGGATGTCACAAAGCTATACAAAATACAGATACTTGTGTTAA
 5500
 GLNLYSALALEUPHEMETHISCYSLYSLSGLYCYSARGCYSLEUGLYGLUGLYHISGLY
 TACAAAAGGCTTTATTTATGCATTGCAAGAAAGGCTGTAGATGTCTAGGGGAAGGACACC
 ALAGLYGLYTRPARGPROGLYPROPROPROPROPROPROGLYLEUALA R METGLU
 GGGCAGGGGGATGGAGACCAGGACCTCCTCCTCCTCCCCCTCCAGGACTAGCATAAATGG
 5600
 GLUARGPROPROGLUASNGLUGLYPROGLNARGGLUPROTRPASPLUTRPVALVALGLU
 AAGAAAGACCTCCAGAAAATGAAGGCCACAAGGGAACCATGGGATGAGTGGGTAGTGG
 5700
 VALLEULYSGLULEULYSGLUGLUALALEULYSHISPHASPPROARGLEULEUTHRALA
 AAGTTCGAAAGAACTGAAGAAGAAGCTTAAAGCATTTTGATCCTCGGCTCTAACC
 TAT1 METGLUTHRPROLEUARGGLUGLNGLUASNSE
 LEUGLYASNHISILETYRASNARGHISGLYASPTHREUGLUGLYALAGLYGLULEUILE
 CACTTGGTAATCATATCTATAATAGACATGGAGACACCCTTGAGGGAGCAGGAGAACTCA
 5800
 LEUGLUSERSERASNGLUARGSERSERTYRILESERGLUALAALAALAILEPROGLU
 ARGILELEUGLNARGALALEUPHEILEHISPHARGSERGLYCYSSERHISSERARGILE
 TTAGAATCCTCCAACGAGGGCTCTTCATACATTTCAGAAGCGGCTGCAGCCATTCCAGAA
 SERALAASNLEUGLYGLUGLUILELEUSERGLNLEUTYRARGPROLEUGLUALACYSTYR
 GLYGLNPROGLYGLYGLYASNPROLEUSERTHRIEPROPROSERARGSERMETLEU
 TCGGCCAACCTGGGGGAGAAATCCTCTCAACTATACCGCCCTTAGAAGCATGCTAT
 5900
 ASNTHRCYSTYRCYSLYSLSYSCYSTYRHISCYSGLNPHECYSPHELEULYSLSYGLY
 AACACATGCTATTGCAAAAAGTGTGCTACCATTGCCAGTTTTGTTTTCTAAAAAGGGC
 6000
 LEUGLYILESERTYRGLULYSSEHISARGARGARGARGTHRPROLYSLSALALYSALA
 ARTMETARGSERHISTHRLYGLUGLUGLULEUARGARGARGLEUARGLEU

(fig. 1B - forts: 4).

TTGGGGATAAGTTATCAGAAGTCACACAGGAGAAGAAGAACTCCGAAGAAGGCTAAGCCT
 ASNTHRSERSERALASERASNGLU
 ILEHISLFULEUMHISGLNTHRSERLYSTYRGLYLEUSERTRPLYSERALAALATYRARG
 ENV METGLYCYSLEUGLYASNGLNLEULEULEALA
 AATACATCTTCTGCATCAAAAGAGTAACTATGGGTGTCTYGGAAATCAGCTGCTTATCG
 6100
 HISLEULEU
 ILECYSSERLYSCYSLEUTRPILEILECYSILEGLNTRYVALTHRVALPHETYRGLYVAL
 CCATCTGCTTAAGYGTCTATGGATTATTTGTATTCAATATGTCACAGTCTTTATGCTG
 PROALATRPARGASNALATHRILEPROLEUPHECYSALATHRLYSASNARGASPTHTRP
 TACCAGCTTGGAGGAATGGACAATCCCTCTTCTGTGCAACCAAGAATAGGGATACTT
 6200
 GLYTHRHRGLNLCYSLEUPROASPASNASPASTYRSEGLULEUALALEUASNVALTHR
 CGGGAACAACCTCAGTGCTACCAGATAATGATGATTATTCAGAATTGGCCCTTAATGTTA
 6300
 GLUSERPHEASPALATRPLUASNTHRYVALTHRGLUGLNALALEGLUASPVALTRPGLN
 CAGAAAGCTTTGATGCTTGGGAGAATACAGTCACAGAACAGGCAATAGAGGACGTATGGC
 LEUPHEGLUTHRSERILELYSPROCYSVALLYSLEUSERPROLEUCYSILETHRMETARG
 AACTCTTTGAGACCTCAATAAAGCCTTGTGTAATAATATCCCCATTATGCATTACTATGA
 6400
 CYSASNLYSSERGLUTHRASPLYSTRPGLYLEUTHRLYSERSERTHRTHRHRALASER
 GATGCAATAAAGTGAGACAGATAAATGGGGATTGACAAATCATCAACAACACAGCAT
 THRTHRHRTHRHRTHRHRALALYSSEVALGLUTHRARGASPILEVALASNGLUTHRSER
 CAACAACAACAACAACAACAGCAAAATCAGTAGAGACAAGAGACATAGTCAATGAGACTA
 6500
 PROCYSVALVALHISASPASNCYSTHRGLYLEUGLUGLNGLUPROMETILESERCYSLYS
 GTCCTTGTGTAGTTATGATAAATGCACAGGCTTGAACAAGACCAATGATAAGCTGTA
 6600
 PHEASNMETHRGLYLEULYSARGASPLYSLYSLSYSLUTYRASNGLUTHRTRPTYRSE
 AATCAACATGACAGGGTTAAAAAGAGACAAGAAAAAGGAGTACAATGAACTTGCTACT
 ALAASPLEUVALCYSGLUGLNGLYASNERTHRGLYASNGLUSERARGCYSTYRMETASN
 CTGCAGACTCGGTTTGTGAACAAGGAATAGCACGTTAATGAAAGTAGATGTTACATGA
 6700
 HISCYASNTHRSERVALILEGLNGLUCYSCYASPLYASPTYRTRPASPALAILEARG
 ATCAGTGAATACTTCTGTTATCCAAGAGTGTGACAAAGATTATGGGATGCTATTA
 CYSARGTYRCYSALAPROPROGLTYRVALALEULEUARGCYSASNASPTHRASNTRYRSE
 GATGTAGATATTGTCACCTCCAGGTATGCTTTGCTTAGATGTAATGACACAAATTA
 6800
 GLYPHEMETPROASNCYSERLYSVALVALSERSERCYSTHRARGMETMETGLUTHR
 CAGGCTTATGCCTAACTGTTCTAAGGTAGTGGTCTTTCATGCACAAGGATGATGGAGA
 6900
 GLNTHRSERTHRTRPPHEARGPHEASNGLYTHRARGALAGLUASNARGTHRTRTYRILETYR
 CACAGACTTCTACTGGTTTCGGTTAATGGAAGTACAGCAGAAAATAGAACCTATATTT
 TRPHISGLYARGASPNARGTHRILEILESERLEUASNLYSHISTYRASNLEUTHRMET
 ACTGGCATGGTAGAGATAATAGGACTATAATTAGTCTAATAAGCATTATAATCTAACAA
 7000
 LYSCYSARGARGPROGLYASNLYSTHRVALLEUPROVALTHRILEMETSERALALEUVAL
 TGAATGTAGAAGACCAGGAAATAAGACAGTTTACCAGTCACCATTATGCTCTGCATTGG
 PHEHISERGLNPROVALASNGLUARGPROLYSGLNALATRPCYSARGPHEGLYGLYASN
 TTTTCCACTCACAACCAGTCAATGAGAGGCCAAAGCAGGCATGGTGTAGGTTTGGAGGAA
 7100
 TRPLYSGLUALALELYSGLUVALLYSGLNTHRILEVALLYSHISPROARGTYRTHRGLY
 ATTGGAAGGAGGCAATAAAGAGGCTGAAGCAGACCATTGTCAACATCCCAGGTATACTG
 7200
 THRASNASNTHRASPLYILEASNLEUTHRALAPROARGGLYGLYASPPROGLUVALTHR
 GAACTAACAACTACTGATAAATCAATTTGACGGCTCCTAGAGGAGGAGATCCGGAAGTTA

(fig. 1B - forts. 5)

PHEMETTRPTHRASNCSARGGLYLUPHELEUTYRCYSLYSMETASNTRPHELEUASN
 CCTTCATGTGGACAAATTGCAGAGGAGAGTTCTCTACTGTAAAATGAATTGGTTTCTAA
 7300
 TRPVALGLUASPARGSERLEUTHRTHRGLNLYSPROLYSGLUARGHISLYSARGASNTYR
 ATTGGGTAGAAGATAGGAGTCTAACTACCCAGAAGCCAAAGGAACGGCATAAAAAGGAATT
 VALPROCYSHISILEARGGLNILEILEASNTHRTRPHISLYSVALGLYLYSASNVALTYR
 ACGTACCATGTATATTAGACAAATAATCAACACTTGGCATAAAGTAGGCCAAAAATGTTT
 7400
 LEUPROPRDARGGLUGLYASPLEUTHRCYSASNSETRHRVALTHRSELEUILEALAAASN
 ATTTGCCCTCAAGAGAGGGAGACCTCACGTGTAACCTCCACAGTGACCAGTCTCATAGCAA
 7500
 ILEASNTRPTHRASPGLYASNGLNTHRSEIRETHRMETSERALAGLUVALALAGLLEU
 ACATAAATTGGACTGATGGAACCAACTAGTATCACCATGAGTGCAGAGGTGGCAGAAC
 TYRARGLEUGLULEUGLYASPTYRLYSLEUVALGLUILETHRPROILEGLYLEUALAPRO
 TGTATCGATTGGAATTGGGAGATTATAAATTAGTAGAAATCACTCCAATTGGCTTGGCCC
 7600
 THRASNVALLYSARGTYRTHRHRGLYGLYTHRSEARGASNLYSARGGLYVALPHEVAL
 CCACAAATGTAAGAGGTACACTACTGGTGGCACCTCAAGAAATAAAGAGGGGGTCTTTG
 LEUGLYPHELEUGLYPHELEUALATHRALAGLYSERALAMETGLYALAALASERLEUTHR
 TGCTAGGGTTCTTGGGTTTCTCGCAACGGCAGGTTCTGCAATGGCGCGGGCGTCTGGA
 7700
 VALTHRALAGLNSERARGTHRLEULEUALAGLYILEVALGLNGLNGLNGLNLEULEU
 CCGTGACCGCTCAGTCCCGGACTTTATTGGCTGGGATAGTGCAGCAACAGCAACAGCTGT
 7800
 ASPVALVALLYSARGGLNGLNGLULEULEUARGLEUTHRVALTRPGLYTHRLYSASNLEU
 TGGACGTGGTCAAGAGACAACAAGAATTGTTGCGACTGACCGTCTGGGGAACAAGAACC
 GLNTHRARGVALSERALAILEGLULYSTYRLEULYSASPGLNALAGLNUASNALATRP
 TCCAGACTAGGGTCTCTGCCATCGAGAAGTACTTAAAGGACCAGGCGCAGCTAAATGCTT
 7900
 GLYCYSALAPHEARGGLNVALCYSHISTHRTHRVALPROTRPPOASNALASERLEUTHR
 GGGGATGTGCGTTTAGACAAGTCTGTCACTACTGTACCATGGCCAAATGCAAGTCTAA
 PROASPTRPASNASNGLUTHRTRPGLNGLUTRPGUARGLYSYVALASPPHELEUGLUALA
 CACCAGATTGGAACAATGAGACTTGGCAAGAGTGGGAGCGGAAGGTTGACTTCTTGGAGG
 8000
 ASNILETHRALLALEULEUGLUGLUALAGLNIILEGLNGLNGLULYSASNMETTYRGLULEU
 CAAATATAACGGCCCTCTAGAAAGAGGCAACAATTCAACAAGAGAAGAACATGTATGAAT
 8100
 GLNLYSLEUASNSETRPASPVALPHEGLYASNTRPPHEASPLEUTHRSERTRPILEYL
 TACAAAAGTTGAATAGCTGGGATGTGTTTGGCAATTGGTTTACCTTACTTCTTGGATAA
 TYRILEGLNTRYGLYITILETYRILEILEVALGLYVALILELEULEUARGILEVALILETYR
 AGTATATACAATATGGAATTTATATAATTGTAGGAGTAATACTGTTAAGAATAGTGATCT
 8200
 ILEVALGLNMETLEUALAARGLEUARGGLNGLYTYRARGPROVALPHESESERPROPRO
 ATATAGTACAAATGCTAGCTAGGTTAAGACAGGGGTATAGGCCAGTGTCTCTTCCCCAC
 TAT2ARGPROILEPROASNARGILEARGL EUCYSGLNPROLYSLYSALA
 ART2VALASPPROTYPROTTHRGLYSERGLYSERALAASNGLNARGARGGLN
 SERTYRPHEN***THRHISTHRGLNGLNASPPROALALEUPROTHRLYSGLUGLYLYS
 CCTCTTATTTCCAGTAGACCCATACCCAACAGGATCCGGCTCTGCCAACCAAGAAGGCA
 8300
 LYSLSGLUTHRVALGLUALAVALALATHRALAPROGLYLEUGLYARG *TAT(fin)
 LYSARGARGARGTRPARGGGLNARGTRPGLNGLNLEULEUALALEUALAASPARGILETYR
 LYSGLYASPGLYGLYGLYSERGLYGLYASNSETRTRPPOTRPGLNILEGLUTYRILE
 AAAAAGGAGACGGTGGAGGCAGCGGTGGCAACAGCTCTGGCCTTGGCAGATAGAATATA
 8400

(fig. 1B - forts. 6)

SERPHEPROASPPROPROTHRASPTHRPROLEUASPLEUALATLEGLNGLNCEUGLNASN
 HISPHELEUILEARGGLNLFUILEARGLLEULEUTHRTRPLEUPHESERASNCYSARGTHR
 TTCATTTCCTGATCCGCCAACTGATACGCCCTTGACTTGGCTATTCCAGCAACTGCAGAA

LEUALAILEGLUSERILEPROASPPROPROTHRASNILEPROGLUALALEUCYSASPLEU
 LEULEUSERARGALATYRGLNILELEUGLNPROILEPHEGLNARGLEUSERALATHRYR
 CCTTGCTATCGAGAGCATACCAGATCCTCCAACCAATATTCCAGAGGCTCTCTGGCACCT

8500

F METGLYGLYALA

ARGARGILEARGARGSERPROGLNALA ● ART2 (fin)

GLYGLUPHEGLYGLUVALLEUARGLEUGLULEUTHRYRLEUGLNTRYRGLYTRPSERTYR
 ACGGAGAATTCCGAGAAGTCCTCAGGCTTGAACCTGACCTACCTACAATATGGGTGGAGCT

ILESERLYSLSYARGSERLYSPROPROGLUILECYASPARGASPSERCYSGLYARGVAL
 PHEGLNGLUALAVALGLNALAALAARGASPLEUARGGLNARGLEULEUARGALAARGGLY
 ATTTCCAAGAAGCGGCTCCAAGCCGCCAGAGATCTGCCAGAGAGACTCTTGGCGGGCCCGTG

8600

GLYARGASNTYRGLYARGLEUPHELYSGLYVALGLUASPGLYSERSERGLNSERLEUGLY
 GLULYSLEUTRPGLUALALEUGLNARGGLYGLYARGTRPILELEUALAILEPROARGARG
 GGGAGAAATTATGGGAGGCTCTTCAAAGGGGTGGAAGATGGATCCTCGCAATCCCTAGGA

8700

GLYLEUASPLYSGLYLEUSERSERLEUSERCYSGLUGLYGLNLYSTYRASNGLNGLYGLU
 ILEARGGLNGLYLEUGLULEUTHRLEULEU ●
 GGATTAGACAAGGCTTGAGCTCACTCTTGTGAGGGCCAAAAATACAATCAGGGAGAA

TYRMETASNTHRPROTRPARGASNPROALAGLUGLUARGLYSLSLEUPROTYRARGLYS
 TACATGAATACTCCATGGAGAAACCCAGCTGAAGAGAGGAAAAAATTACCATACAGAAA
 8800

GLNASNILEASPAPILEASPGLUGLUASPAASPLEUVALGLYILEPROVALGLUALA
 CAAAATATAGATGATATAGATGAGGAAGATGATGACTTGGTAGGGATACCAGTTGAGGCC

ARGVALPROLEUARGTHRMTSERTYRLYSLEUALAILEASPMTSERHISPHEILELYS
 AGAGTTCCCCTAAGAACAATGAGTTACAAATGGCAATAGATATGTCTCATTTTATAAAA
 8900

GLULYSGLYGLYLEUGLUGLYILETYRTRYSERALAARGARGHISARGILELEUASPILE
 GAAAGGGGGGACTGGAAGGGATTTATTACAGTGCAAGAAGACATAGAATCTTAGACATA
 9000

TYRLEUGLULYSGLUGLUGLYILEILEPROASPTRPGLNILEHISSERGLYPROGLYILE
 TACTTAGAAAAGGAAGAAGGCATCATACCAGATTGGCAGATACACTCCGGACCAGGAATT

ARGTYRLEULYSMETPHEGLYTRPLEUTRPLYSLEUILEPROVALASNYALSERASPGLU
 AGATACCTAAGATGTTTGGCTGGCTATGGAAATTAATCCCTGTAATGTATCAGATGAG
 9100

ALAGLNGLUASPGLUGLUHI STYRLEUVALHISPROALAGLNTHRSERGLNTRPASASP
 GCACAGGAGGATGAGGAGCATTATTTAGTGCACCCAGCTCAAACCTCCAGTGGGATGAC

PROTRPGLYGLUVALLEUALATRPLYSPEASPPROTHRLEUALATYRTHRYRGLUALA
 CCTTGGGAGAGGTTCTAGCATGGAAGTTTGATCCAACCTAGCCTACACTTATGAGGCA
 9200

TYRILEARGTYRPROGLUGLUPHEGLYSERLYSSERGLYLEUSERGLULYSGLUVALLYS
 TATATTAGATACCCAGAAGAGTTTGAAGCAAGTCAGGCTGTCAGAGAAAGAGGTTAAA
 9300

ARGARGLEUALAALAARGGLYLEULEUGLUMETALAASPARGLYSGLUTHRSE
 AGAAGGCTAGCCGCAAGAGGCCTTCTTGAAATGGCTGACAGGAAGGAACTAGCTGAGAC

AGCAGGGACTTCCACAAGGGGATGTGATGGGGAGGACTGGGGAGGAGCCGGTTGGGAA
 9400

CACCCACTTCTTGATGTATAAATATCACTGCATTTCCGCTCTGTATTCAGTCGCTCTGGC
 GAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCTCCAGCACTAGCAGGTAGAGCCTGGG
 9500

TGTTCCCTGCTAGACTCTCACCAGCACTTGGCCGGTGCTGGGCAGAGTGGCTCCACGCTT
 9600

(fig. 1B - forts. 7)

FIG. 1C

LTR sekvens
CIVET
versus
HIV-2 ROD

```
X      8960      8970      8980      8990      9000      9010
TGGAAGGGATTATTACAGTGCAAGAAGACATAGAATCTTAGACATATACTTAGAAAAGG
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TGGAAGGGATGTTTTACAGTGAAGAAGACATAAAATCTTAAATATATACTTAGAAAAGG
X      8950      8960      8970      8980      8990

      9020      9030      9040      9050      9060
AAGAAGGCATCATAACCAGATTGGCAGATACACTCCGGA---CCAGGAATTAGATACCTAA
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
AAGAAGGGATAATTGCAGATTGGCAGAACTACACTCATGGGCCAGGAGTAAGATACCCAA
      9010      9020      9030      9040      9050

      9080      9090      9100      9110      9120
AGATGTTTGGCTGGCTATGGAAATTAATCCCTGTAAATGTATCAGATGAGGCACAGGAGG
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TGTCTTTGGGTGGCTATGGAAGCTAGTACCAGTAGATGTCCCACAAGAAGGGGAGGACA
      9070      9080      9090      9100      9110

      9140      9150      9160      9170      9180
ATGAGGAGCATTATTTAGTGCACCCAGCTCAAACCTCCAGTGGGATGACCCCTGGGGAG
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CTGAGACTCACTGCTTAGTACATCCAGCACAAACAAGCAAGTTTGATGACCCGCATGGGG
      9130      9140      9150      9160      9170

      9200      9210      9220      9230      9240
AGGTTCTAGCATGGAAGTTTGATCCAACTCTAGCCTACACTTATGAGGCATATATTAGAT
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
AGACACTAGTCTGGGAGTTTGATCCCTTGCTGGCTTATAGTTACGAGGCTTTTATTCCGT
      9190      9200      9210      9220      9230

      9260      9270      9280      9290      9300
ACCCAGAAGAGTTTGAAGCAAGTCAGGCCTGTCAGAGAAAGAGGTTAAAAGAAGGCTAG
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
ACCCAGAGGAATTTGGGCACAAGTCAGGCCTGCCAGAGGAAGAGTGAAGGGCGAGACTGA
      9250      9260      9270      9280      9290

      9320      9330      9340      9350
CCGCAAGAGGCCTTCTTGAAATGGCT-GACAGGAAGGAACT-----
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
AAGCAAGAGGAATACCATTTAGTTAAAGACAGGAACAGCTATACTTGGTCAGGGCAGGAA
      9310      9320      9330      9340      9350
```

FIG. 1C

```

-----AGCTGAGACAGCAGGGACTTTCCACAAGGGGATGTCATG--GGGA
          9360      9370      9380      9390
          : : : : : : : : : : : : : : : : : : : : : : : : : : : :
GTAAC TAACAGAAACAGCTGAGACTGCAGGGACTTTCCAGAAGGGGCTGTAACCAAGGGA
          9370      9380      9390      9400      9410

          9400      9410      9420      9430      9440      9450
GGTACTGGGGAGGAGCCGGTTGGGAACACCCACTTTCTTGATGTATAAATATCACTGCAT
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
GGGACATGGGAGGAGCTGGTGGGGAACGCCCTCATATTCTCTGTATAAATATACCCGCTA
          9430      9440      9450      9460      9470

          9460      XX      10      20      30      40
TTCGCTCTGTA--TTCTGGAAGGGATTTATTACAGTGCAAGAAGACATAGAATCTTAGAC
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
GCTTGCATTGTACTTCTGGAAGGGATGTTTACAGTGAAAGAAGACATAAAATCTTAAAT
          9490      XX      10      20      30      40

          50      60      70      80      90
ATATACTTAGAAAAGGAAGAAGGCATCATACCAGATTGGCAGATACACTCCGGA---CCA
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
ATATACTTAGAAAAGGAAGAAGGGATAATTGCAGATTGGCAGAACTACACTCATGGCCA
          50      60      70      80      90      100

          110      120      130      140      150
GGAATTAGATACCTAAAGATGTTTGGCTGGCTATGGAAATTAATCCCTGTAATGTATCA
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
GGAGTAAGATACCCAATGTTCTTTGGGTGGCTATGGAAGCTAGTACCAGTAGATGTCCCA
          110      120      130      140      150      160

          170      180      190      200      210
GATGAGGCACAGGAGGATGAGGAGCATTATTTAGTGCACCCAGCTCAAACCTCCAGTGG
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
CAAGAAGGGGAGGACACTGAGACTCACTGCTTAGTACATCCAGCACAAACAAGCAAGTTT
          170      180      190      200      210      220

          230      240      250      260      270
GATGACCCCTGGGGAGAGGTTCTAGCATGGAAGTTTGATCCAACCTCTAGCCTACACTTAT
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
GATGACCCGCATGGGGAGACACTAGTCTGGGAGTTTGATCCCTTGCTGGCTTATAGTTAC
          230      240      250      260      270      280

          290      300      310
GAGGCATATATTAGATACCCAGAAGAGTTTGGGAAGCA
: : : : : : : : : :
GAGGCTTTTATTCCG
          290

```

(fig.1C. forts. 1)

FIG. 2 (HIV-2.P
(versus
(HIV-1.P

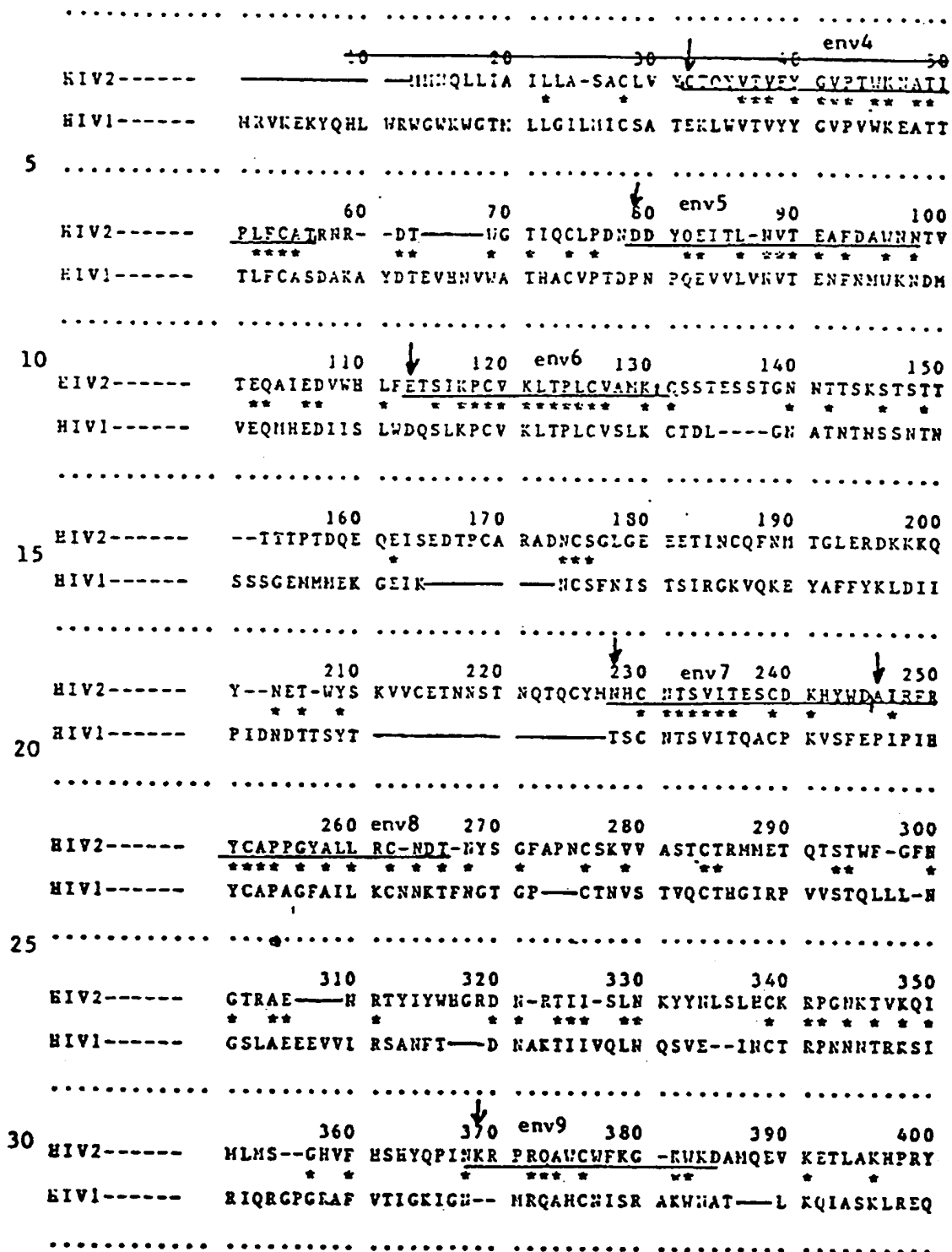


FIG. 2

```

                                410      ↓ 420  env10  430      440      450
HIV2-----  RGINDTRNIS  FAAPGKGS DP  EVAYEW TNCR  GEFLYCKNTV  FLH--WI--
* * * * *      * * * * *      * * * * *      * * * * *      * * * * *
HIV1-----  FGHHT--II  FKQSS--GGDP  EIVTHSFNCC  GEFFYCNSTQ  LFNSTWFNST
.....

                                460      ↓ 470  env11  480      490      500
5 HIV2-----  -----EN  KTRNYAPCH  IKQIINTWPK  VGRNVYLPPR  EGELSCNSTV
* * * * *      * * * * *      * * * * *      * * * * *      * * * * *
HIV1-----  WSTEGSNTE  GSDTITLPCR  IKQFINNWQE  VGRKAMYAPPI  SGQIRCSSNI
.....

                                510      520      530      540      ↓ 550
10 HIV2-----  TSIIANIDWQ  NNNQTNITFS  AEVAELYRL-  -ELGDYKLV  EITPIGFAPT
* * * * *      * * * * *      * * * * *      * * * * *      * * * * *
HIV1-----  TGLLLTRDGG  KNNNGSEIFR  PGGGDHRDNW  RSELYKYKVV  KIEPLGVAPT
.....

env3  560      570      580      590      600
HIV2-----  KEKRYSSAHC  RHTRGVFVLC  -FLGFLATA  GSAMGAAS-  LTVSAQSRTL
* * * * *      * * * * *      * * * * *      * * * * *      * * * * *
HIV1-----  KAKRE--VVQ.  REKRAVGI-G  ALFLGFLGAA  GSTMGARSMT  LTVQA--RQL
.....

                                610      620      630      ↓ 640  env1  650
HIV2-----  LAGIVQQQQQ  LLDVVKRQQE  LLRLTVWGTK  NLQARVTAIE  KYLQDQARLN
* * * * *      * * * * *      * * * * *      * * * * *      * * * * *
HIV1-----  LSGIVQQQNN  LLRAIEAQQH  LLQLTVWGIK  QLQARILAVE  RYLKDDQQLG
.....

                                660      670      680      690      700
20 HIV2-----  SWGCAFRQVC  HTTVPW-----  VNDLAPDWD  NHTWQEWKQ  VRYLEANISK
* * * * *      * * * * *      * * * * *      * * * * *      * * * * *
HIV1-----  IWGCSGKLIC  TTAVPWNASI  SRSLEQIIE  NHTWMEUDRE  INNYTSLIES
.....

                                ↓ 710  env2.  720      730      740      750
25 HIV2-----  SLEQAQIQQE  KNYELQKLN  SWDIFGNWFD  LTSWVKYIQY  GVLIIVAVIA
* * * * *      * * * * *      * * * * *      * * * * *      * * * * *
HIV1-----  LIEESQNQQE  KREQELLELD  KHASLWNWPN  ITNWLWYIKI  FIMIVGCLVG
.....

```

(fig.2 - forts. 1)

		760	770	780	790	800
	HIV2-----	LRIVIVVQ	MSLRKGYR	V-FSSPPGYI	QQIHKDRG	QPANEETEED
		**** *	* * * * *	*	**	* **
	HIV1-----	LRIVFAVLSI	VNRVRQGYSP	LSFQT-----	-----HLPTPRG	PDRPEGIEEE

5		810	820	830	840	850
	HIV2-----	GCSNGGDRYW	PWPIAYIHFL	IRQLIRLLT-	-----LYSIC	RDLLSRFLT
		** **	*	* *		****
	HIV1-----	GGERDRRSI	RLVNGSLA-L	IWDDLRSICL	FSYHRL-----	RDLLIVTRI

10		860	870	880	890	900
	HIV2-----	LQLIQNLRD	WLRLRTA-F	LQYGCWQIE	AFQ-----AAA	RATRETL-----
		*	* *	***	** *	* *
	HIV1-----	VELLC--RRG	WEALKYWNL	LQYWSQELKN	SAVSLLNATA	IAVAEGTDRV

		910	920	930	938	
	HIV2-----	-----AGACRG	LWRVLERIGR	GILAVPRRIR	QGAETALL	
		**		*****	** * **	
15	HIV1-----	IEVVQGACRA	-----	-IRHIPRRIR	QGLERILL	

(fig. 2 - forts. 2)

FIG. 3 (ENV-mac
(versus
(ENV-ROD

```

      10      20      30      40      50
MGCLGNOLLIAIC--SKCLWIIICIQYVTVFYGVPAWRNATIPLFCATKNRDTWGTTOCL
:   : : : : : :   : :   : : : : : : : : : : : : : : : : : : : : : : : :
MM----NQLLIAILLASACLVY-CTQYVTVFYGVPTWKNATIPLFCATRNRDTWGTIOCL
      10      20      30      40      50

      60      70      80      90      100     110
PDNDYSELALNVTESFDAWENTVTEQAIEDYWOLFETSISKPCVKLSPLCITMRCNKSET
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
PDNDYQEITLNVTEAFDAWNNTVTEQAIEDYVHWFETSISKPCVKLTPLCVAMKCSSTES
      60      70      80      90      100     110

      120     130     140     150     160     170
DKWGLTKSSTTTASTTTTTAKSVETRDIVNETS--PCVVHONCTGLEQEPHISCKFNM
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
STGNNTTSKST--STTTTTP-----T-DOEQEISEDTPCARADNCSGLGEEETINCOFNM
      120     130     140     150     160

      180     190     200     210     220     230
TGLKRDKKKEYNETHWYSADLVCEQNSTGNESRCYMNHCNTSVIQECCDKDYWDAIRCRY
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
TGLERDKKKQYNETHWYSKDVVCETNST-NQTQCYMNHCNTSVITESCDKHWDIAIRFRY
      170     180     190     200     210     220

      240     250     260     270     280     290
CAPPGYALLRCNDTNYSGFMPNCSKVVSSTRMNETQTSTWFRFNGTRAENRTYIYWHG
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
CAPPGYALLRCNDTNYSGFAPNCSKVVASTRMNETQTSTWFGFNGTRAENRTYIYWHG
      230     240     250     260     270     280

      300     310     320     330     340     350
RDNRTIISLNKHYNLTMKCRPGNKTVLPVTIMSAVHFS--OPVNERPKOAWCRFGGNW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RDNRTIISLNKYNLSLHCKRPGNKVKOINLMSGHVHSHYQPIPKRPROAWCHFKGKW
      290     300     310     320     330     340

      360     370     380     390     400
KEAIKEVKOTIVKHPRYTGTNNTDKINLTAPRGG-DPEVTFMWTNCRGEFLYCKMNWFLN
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
KOAMQEVKETLAKHPRYRGTNDTRNISFAAPGKGSDEPVAVMWTNCRGEFLYCNMTWFLN
      350     360     370     380     390     400

```

FIG. 3

FIG. 4 (GAG-mac
(versus
(GAG-ROD

```

      10      20      30      40      50
VQHKKEIAVFYPCGRDNKIEWEMGARNVLSGKKADELEKIRLRPGGKKKYMLKHVVWAAN
      :::::::::: :::::::::: ::::::::::: ; :: ::::
      MGARNVLRGKKADELERIRLRPGGKKKYRLKHIYWAAN
      10      20      30

      70      80      90      100      110
ELDRFGLAESLLENKEGCOKILSVLAPLVP TGSENLSKLYNTVCVIMCIHAEKVKHTEE
      ::::::::::: ::::::::::: : : ::::::::::: ::::::::::: : :
KLDRFGLAESLLESKEGCQKILTVDPMVPTGSENLSKLFNTVCVIMCIHAEKVKDTEG
40      50      60      70      80      90

      130      140      150      160      170
AKQIVQRHLMVETGTAETMPKTSRPTAPFSGRGGNYPVQOIGGNYTHLPLSPRTLNAWVK
      ::::: :::: ::::: : : ::::: : : ::::: :::: :::: :::: ::::
AKQIVRRHLVAETGTAEKMPSTSRPTAP SSEKGGNYPVOHVGGNYTHIPLSPRTLNAWVK
100      110      120      130      140      150

      190      200      210      220      230
LIEEKKFGAEVVSQFQALSEGCLPYDINQMLNCVGDHQAAMQIIRDIINEEAADWDLQHP
: ::::::::::: ::::::::::: ::::::::::: ::::::::::: ::::::::::: : :
LVEEKKFGAEVYPGFQALSEGCTPYDINQMLNCVGDHQAAMQIIREIINEEAADWVQHP
160      170      180      190      200      210

      250      260      270      280      290
QQAPQQ-GQLREPSGSDIAGTTSTVEEQIQHMYROONPIPVGNIYRRWIOLGLOKCVRMV
: ::::: :::: :::: :::: : : : : : : : : : : : : : : : : : : : : : :
IPGPLPAGQLREPRGSDIAGTTSTVEEQIQHMFROPNPVPGNIYRRWIOIGLOKCVRMV
220      230      240      250      260      270

      300      310      320      330      340      350
NPTNILDVKGQPKPEFOSYVDRFYKSLRAEQTPAVKNWMTQTL LIONANPDCKLVLKGL
      ::::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
NPTNILDIKGQPKPEFOSYVDRFYKSLRAEQTPAVKNWMTQTL LVONANPDCKLVLKGL
280      290      300      310      320      330

      360      370      380      390      400      410
GTNPTLEHLTACQVGGPGQKARLMAEALKEALAPAPIPFAAAQOKGPRKPIKCNCGK
: ::::::::::: ::::::::::: ::::::::::: : : ::::: : : ::::: : :
GMNPTLEHLTACQVGGPGQKARLMAEALKEVIGPAPIPFAAAQO---RKAFKCNCGK
340      350      360      370      380      390

```

FIG. 4

FIG. 5 (POL-mac
(versus
(POL-ROD

```

10      20      30      40      50
VLELWEGRTLCKAHOSPKKTGMLEMMKNGPCYGOMPQTGGFFRPWPLGKEAPQFPHGSS
      :: :: : : : : : : : : : :
TGRFFRTGPLGKEAPQLPRGPS
      10      20

70      80      90      100
ASGADANCSPRRTSCGS AKELHALGQAERKOREALOGGDRGF-----
      :: : : : : : : : : : : : : : :
SAGADTNSTPSGSSSGSTGEIYAAREKTERAERETIQGSDRGLTAPRAGGDTIOGATNRG
      30      40      50      60      70      80

110     120     130     140     150     160
-AAPOFSLWRRPVYTAHIEGQPVEVLLDTGADDSIVTGIELGPHYTPKIVGGIGGFINTK
      :: :: : : : : : : : : : : : : : : : : : : : : : : : : : :
LAAPOFSLWKRVPYTAHIEGQPVEVLLDTGADDSIVAGIELGNYSKIVGGIGGFINTK
      90      100     110     120     130     140

170     180     190     200     210     220
EYKNVEIEVLGKRIKGTINTGDPINIFGRNLLTALGMSLNLPVAKVEPVKSPKPKGKDG
      :: :: : : : : : : : : : : : : : : : : : : : : : : : : : :
EYKNVEIEVLNKKVRATINTGDPINIFGRNILTALGMSLNLPVAKVEPIKIMLKPKGKDG
      150     160     170     180     190     200

230     240     250     260     270     280
PKLKQWPLSKEKIVALREICEKMEKDGQLEEAPPTNPYNTPTFAIKKKDKNKNRMLIDFR
      :: :: : : : : : : : : : : : : : : : : : : : : : : : : : :
PKLRWPLTKIEALKEICEKMEKEGQLEEAPPTNPYNTPTFAIKKKDKNKNRMLIDFR
      210     220     230     240     250     260

290     300     310     320     330     340
ELNRVTQDFTEVQLGIPHPAGLAKRKRITVLDIGDAYFSIPLDEEFROYTAFTLPSVYNA
      :: :: : : : : : : : : : : : : : : : : : : : : : : : : : :
ELNKVTQDFTEIQLGIPHPAGLAKRKRITVLDVGDAYFSIPLHEDFRPYTAFTLPSVYNA
      270     280     290     300     310     320

350     360     370     380     390     400
EPGKRYIYKVLPOGHWKGSPIFQYTHRVLEPFRRKANPDVTLVQYMDILIASDRDLEH
      :: :: : : : : : : : : : : : : : : : : : : : : : : : : : :
EPGKRYIYKVLPOGHWKGSPIFQHTHROVLEPFRRKANKBVIIQYMDILIASDRDLEH
      330     340     350     360     370     380

```

FIG. 5

FIG. 6 (Q.mac
(versus
(Q.ROD

```

      10      20      30      40      50
MEEEKRWIVVPTWRIPERLERWHSLIKYLKYKTKDLOKACYVPHHKVGVWAWWTCSRVIFF
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
MEEDKRWIVVPTWRVPGRMEKWHSLVKYLKYKTKDLEKVCYVPHHKVGVWAWWTCSRVIFF
      10      20      30      40      50

      70      80      90     100     110
LQEGSHLEVQGYWNLTPERGWLSTYAVRITWYSKDFWTDVTPEYADILLHSTYFPCFTAG
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
LKGNSHLEIQAYWNLTPKGLSSYSVRITWYTEKFWTDVTPDCADVLHSTYFPCFTAG
      70      80      90     100     110

      130     140     150     160     170
EVRRAIRGERLLSCCRFPRAHKHQVPSLOYLALRVVSHV-RSQGENPTWKQWRRDNRSL
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
EVRRAIRGEKLLSCCNYPRAHRAQVPSLOFLALVVVQONDRPQRDSTTRKORRRDYRRGL
      130     140     150     160     170

      180     190     200     210
RYAKQNSRGDKQRGGKPPTEGANFPGLAKVLGILA
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RLAKQDSRSHKQRSSESPTPRTYFPGVAEVLILA
      190     200     210

```

FIG. 6

(R.mac
FIG. 7 (versus
(R.ROD

```
      10      20      30      40      50
ME---ERP PENEGP OREP WDEWVVEVLKELKEEALKHFOPRLLTALGNHIYNRHGDTLE
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
MAEAPTELPPVDGTPLEPGDEWIIIEILREIKEEALKHFOPRLLIALGKYIYTRHGDTLE
      10      20      30      40      50

      60      70      80      90      100
GAGELIRILORALFIHFRSGCShSRIGOPGGGNPLSTIPPSRSMML
:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
GARELIKVLORALFTHFRAGCGHSRIGQTRGGNPLSAIPTPRNMO
      70      80      90      100
```

FIG. 7

FIG. 9 (F.mac
 (versus
 (F.ROD

```

      10          20          30          40          50
MGGAI SKKRSKPPEICD-RDSCGRVGRNYGRLFK-GVEDGSSQSLGGLDKGLSSLCEGO
: :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :
MGASGSKKHSRPPRGLQERLLRARAGACGGYWNESGGEYSRFQE—GSDREQKSPSCEGR
      10          20          30          40          50

      60          70          80          90          100          110
KYNQGEYMNTPWNRNPAERKKLPYRKQNIODIDEEDDLVGIPEARVPLRTHSYKLAID
: :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :
QYQOGDFMNTPWKDPAAEREKNLYROONMDDVDSDDDDQVRVSVTPKVPLRPMTHRLAID
      60          70          80          90          100          110

      120          130          140          150          160          170
MSHFIKEKGGLEGIIYSARRHRILDIYLEKEEGIIPDWQI—HSGPGIRYLKMFGLWKL
: :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :
MSHLIKTRGGLEGMFYSERRHKILNIYLEKEEGIIADWQNYTH-GPGVRYPMFFGHLWKL
      120          130          140          150          160          170

      180          190          200          210          220          230
IPVNVSDAEQEDDEHYLVHPAQT SQWDDPWGEVLAMKFDPTLAYTYEAYIRYPEEFGSKS
: :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :  : :
VPVDVPOEGEDTETHCLVHPAQT SKFDDPHGETLVWEFDPLLAYS YEAFIRYPEEFGHKS
      180          190          200          210          220          230

      240          250          260
GLSEKEVKRRLAARGLLEMADRKETS
: :  : :  : :  : :  : :
GLPEEENKARLKARGIPFS
      240          250

```

FIG. 9

FIG. 10 (TAT.mac
 (versus
 (TAT.ROD

```

      10      20      30      40      50
METPLREQENSLESSNERSSYISEAAAAIPESANLGEIILSOLYRPLEACYNTCYCKKCC
:::::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
METPLKAPESLLKSCNEPFSRTSEODVATQELAROGEEIILSOLYRPLETCNNSCYCKRCC
      10      20      30      40      50

      70      80      90      100      110
YHCQFCFLKKGLGISYEKSHRRRRTPKKAKANTSSASNERP---IPNRIRLCOPKKAKKE
:::::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
YHCQMCFLNKGLGICYERKRRRRTPKKTKTHPSPT----PDKSISTRGDSOPTKKQKK
      70      80      90      100      110

      120      130
TVEAAVATAPGLGR
:::::  :  :  :  :  :  :
TVEATVETDTGPR
      120      130
  
```

FIG. 10

FIG. 11 (ART.mac
(versus
(ART.ROD

```
      10      20      30      40      50
MRSHTEEEELRRRLRLIHLHOTS KYGLSWKSAAYRHLLVDPYPTGSGSANORROKRRM
:   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
MNERADEEGLORKLRLIRLLMOTN-----PYPOGPGTASORRNRRRRM
      10      20      30      40

      70      80      90      100     110
RQRWOQLLALADRIYSFPDPPTDTPLDLAIOQLONLAIESIPDPPTNIPEALCOLRRIR
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
KORWROILALADSIYTFPDPADSPLDQTIQHLQGLTIOELPDPPTHLPESORLAET
      50      60      70      80      90      100
```

SPQA

FIG. 11