Predictions of linear T-cell and B-cell epitopes in proteins encoded by HIV-1, HIV-2 and SIV_{MAC} and the conservation of these sites between strains

Markéta J.J.M. Zvelebil⁺, Michael J.E. Sternberg^{+*}, John Cookson° and Anthony R.M. Coates[†]

†Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, °Hill Computer Centre and †Department of Medical Microbiology, The London Hospital Medical College, Turner Street, London El 2AD, England

Received 13 October 1988

An important consideration in the design of vaccines to prevent HIV-1 infection effective against different strains is the amino acid sequence conservation of antigenic determinants. Even one amino acid change can destroy the antigenicity of a site for the antibody or T-cell receptor. The comparisons of predicted T- and B-cell epitopes between human HIV-1, HIV-2 and monkey SIV_{MAC} AIDS viruses are presented. The three major gene products (env,gag and pol) were examined. A number of epitopes were identical between strains of HIV-1. Our analysis highlights the problem of designing an effective HIV-1 and HIV-2 vaccine and also the problem of testing human vaccines in monkey models.

AIDS; HIV-1; HIV-2; SIV_{MAC}; Vaccine design; Antigenic epitope

1. INTRODUCTION

There is no effective vaccine for the prevention of AIDS (Acquired Immune Deficiency Syndrome). In designing a vaccine for this disease, consideration should be given to epitope sequence variation between strains of the human immunodeficiency virus type-1 [1-4] because a successful vaccine should be effective against a range of isolates. Furthermore, if epitopes were identified which are shared between human immunodeficiency virus type-1 (HIV-1) and type-2 (HIV-2) [5] a common HIV vaccine for these two strains could be feasible. SIV_{MAC} causes an AIDS-like disease in rhesus macaque monkeys [6];

Correspondence (present) address: M.J.J.M. Zvelebil, Biomolecular Computing Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, England

Present address: Biomolecular Computing Laboratory, Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, England

therefore, if a section of the chain is conserved between HIV-1, and SIV_{MAC} it could lead to testing of the effect of HIV-1 peptides in monkeys. In this paper we use the nucleotide-derived protein sequences of HIV-1 strains, HIV-2 and SIV_{MAC} to predict areas of sequence homology which contain antigenic B-cell and T-cell determinants. We consider the three major gene products: gag, pol and env.

Research has concentrated on the development of an envelope vaccine [7–11] because certain retroviral envelope proteins [12] do confer antibody-mediated protection against viral challenge in animals. So far HIV-1 envelope vaccines [13] have failed to protect chimpanzees against infection. We have suggested [14,15] that the search for a vaccine should not only include env, but also gag and possibly pol gene products. This proposal is based upon the observation [16] that T-cell epitopes of the core protein of hepatitis-B virus induce T-cell help of B lymphocytes. This results in the production of antibody against the surface antigen. Furthermore, preliminary ex-

	10 20 30 40	50 60	70 80	90 100
BIV-1 VARIANCE	* * * * * * * *	v v v v	VVVV V V	** ***
HIV-1 SIV HIV-1	NGARASVLSGGELDRNER I RERPGGREKKELERH I VNAS RELERPA			** * * * * * *
SIV	HOARNSVESCEE ELECTROCKER TRUMP VINASKEELER F			
HIV-2	MCARNSVERGKKADELER I RERFOCKKKYREKH I VMAANKEDRYC	Laeslleskegegki ltvi	DPHVPTG\$ENLKSLFWTVCVII	IC I HABEKVKDTEGA
HIV-1 HIV-2		*** *** ** *		
B-EPITOPE HIV-1	**16***14***9** ***6**			++17+++5++1
B-EPITOPE SIV	++16+++14+++9++ +++6++ +++2++ +++9++ ++16++ ++3+++++12++++5+++ ++16++	++13++		+++3 ++11+
B-EPITOPE HIV-2	**3*** **12** **5*** **16**	++14++		**2***
T-CELL/DB HIV-1		****		
T-CELL/DB SIV T-CELL/DB SIV-2				
T-CELL/RT BIV-1 T-CELL/RT BIV				
T-CELL/RT BIV-2				
	110 120 130 140	150 160	170 180	190 200
SIV-1 VARIANCE SIV-1_SIV	V V VVV V VVVVVV V	v	· · · · · · · · · · · · · · · · · · ·	
MIV-1	LDKIEEE ONKSKKKAQQAAADTGESSQVSQNYPIVONIQGON			
8 IV	KOIVORELVHETGTAETHPKTSRFTAPFSGRGRYPV QQIGGRY	TELPLEPRTLEAMVKLIE	KKFGAEVVSGFQALSEGCLFYI	LINGHLINCACOBONY
HIV-2 HIV-1_HIV-2	KOIVERELVAETGTAE MIPSTERPTAPSEERGGETPV QUVOGNY	TEIPLSPRTLMANVKLVE	KKFGAEVVPGFQALBEGCTPYI	INGHLHCVGDBQAA
H14-1_H14-2	, , ,			
B-EPITOPE BIV-1 B-EPITOPE SIV	0+++1++ 3+++2++ +	++13	++ +5++	
B-EPITOPE HIV-2	++17++		6+++	
T-CELL/DB BIV-1				
T-CELL/DB SIV				
T-CELL/DB HIV-2				
T-CELL/RT HIV-1	-			
T-CELL/RT SIV				
T-CELL/RT HIV-2				
HIV-1 VARIANCE	210 220 230 240	250 260	270 280 V	290 300
EIV-1 VARIANCE EIV-1_SIV	··· · · · · · · · · · · · · · · · · ·		· · · · · · · · · · · · · · · · · · ·	
MIV-1_SIV BIV-1	MONERETIME ERRENDRY VERGPIA PROMIEPROSDIR GTTST	LQEQIGM THEP1PVQ	V V	ILDINGSPREPFRDY
MIV-1 SIV	MONIKETINEEAARMORVEPVHAGFIAFGGMEPRGSDIAGTTST MOIIRDINEEAARMORVEPVHAGFIAFGGDIAGTTST MOIIRDINEEAARMORVERIF GELFANGLERFRGSDIAGTTST	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	ILD IRGGPREPFOSY
MIV-1_SIV MIV-1 SIV	MONERETIME ERRENDRY VERGPIA PROMIEPROSDIR GTTST	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	ILD IRGGPREPFOSY
MIV-1_SIV BIV-1 SIV BIV-2	MONIKETINEEAARMORVEPVHAGFIAFGGMEPRGSDIAGTTST MOIIRDINEEAARMORVEPVHAGFIAFGGDIAGTTST MOIIRDINEEAARMORVERIF GELFANGLERFRGSDIAGTTST	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	ILD IRGGPREPFOSY
BIA-1 ⁻ BIA-5 BIA-1 BIA-1 BIA-1 ⁻ SIA	MOMERATIME EARLEMONVER VERAGE LAR COURSE PASSO LAGTTST MOITMOITMEETH MORE AND A COURSE PASSO LAGTTST MOITMOITMEETH MORE AND A COURSE PASSO LAGTTST MOITMEITMEENAMENOUS PASSO LAGTTST HOTTEL THE LARLEMOND PASSO LAGTTST HOTTEL THE LARLEMOND PASSO LAGTTST HOTTEL THE LARLEMOND PASSO LAGTTST	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	ILD IRGSPREPFRDY ILDVKGSPREPFGSY ILD IRGSPREPFGSY
B-EPITOPE BIV-1 BIV-1_BIV-2 BIV-1_BIV-2	MOMERATIME EARL MORVEY VERGE LAPOGRAFE PAGED LAGTTST MOITRE ITME EARL MODEL OF THE COORTINE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGE PAGE PAGE PAGE PAGE PAGE PAGE	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	ILD IROGPREP FROY LLDVKGGPREP FGSY LLD IROGPREP FGSY +15+11++ +++6++
B-F5110bE SIA-1 BIA-1 BIA-5 BIA-1 SIA BIA-1 SIA	MOMERATIME EARLEMONVER VERAGE LAR COURSE PASSO LAGTTST MOITMOITMEETH MORE AND A COURSE PASSO LAGTTST MOITMOITMEETH MORE AND A COURSE PASSO LAGTTST MOITMEITMEENAMENOUS PASSO LAGTTST HOTTEL THE LARLEMOND PASSO LAGTTST HOTTEL THE LARLEMOND PASSO LAGTTST HOTTEL THE LARLEMOND PASSO LAGTTST	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	ILDIRGSPREPFRDY ILDVRGSPREPFGSY ILDIRGSPREPFGSY ++++++++++++++++++++++++++++++++++++
B-EPITOPE BIV-1 BIV-1_BIV-2 BIV-1_BIV-2	MOMERATIME EARL MORVEY VERGE LAPOGRAFE PAGED LAGTTST MOITRE ITME EARL MODEL OF THE COORTINE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGE PAGE PAGE PAGE PAGE PAGE PAGE	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	ILD IROGPREP FROY LLDVKGGPREP FGSY LLD IROGPREP FGSY +15+11++ +++6++
MIV-1 SIV BIV-1 SIV BIV-2 BIV-1 HIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-2 T-CELL/DB MIV-1	MOMERATIME EARL MORVEY VERGE LAPOGRAFE PAGED LAGTTST MOITRE ITME EARL MODEL OF THE COORTINE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGE PAGE PAGE PAGE PAGE PAGE PAGE	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	LIDIRGOPKEPFRDY LIDVKGOPKEPFGSY LIDIKGOPKEPFGSY +15+11++ +++6++ ++8+++
RIV-1 RIV-1 SIV RIV-2 RIV-1-RIV-2 B-EPITOPE RIV-1 B-EPITOPE RIV-1 T-CELL/DB RIV-1 T-CELL/DB RIV-1	MOMERATIME EARL MORVEY VERGE LAPOGRAFE PAGED LAGTTST MOITRE ITME EARL MODEL OF THE COORTINE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGE PAGE PAGE PAGE PAGE PAGE PAGE	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	LLD IRGGPKEPFRDY LLDVGGPKEPFGSY +15+11++ ++6++
MIV-1 SIV BIV-1 SIV BIV-2 BIV-1 HIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-2 T-CELL/DB MIV-1	MOMERATIME EARL MORVEY VERGE LAPOGRAFE PAGED LAGTTST MOITRE ITME EARL MODEL OF THE COORTINE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGE PAGE PAGE PAGE PAGE PAGE PAGE	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	LIDIRGOPKEPFRDY LIDVKGOPKEPFGSY LIDIKGOPKEPFGSY +15+11++ +++6++ ++8+++
RIV-1 SIV RIV-2 RIV-2 B-EPITOPE RIV-2 B-EPITOPE RIV-2 T-CELL/OB RIV-1 T-CELL/OB RIV-2	MOMERATIME EARL MORVEY VERGE LAPOGRAFE PAGED LAGTTST MOITRE ITME EARL MODEL OF THE COORTINE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGE PAGE PAGE PAGE PAGE PAGE PAGE	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	LD IRQUPREP FROY LLD/WGGPREP FGSY +15+11++ ++6++ ++8+++
HIV-1 SIV BIV-2 BIV-2 BIV-1_BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/DB SIV-2 T-CELL/DB TV-2	HOULKETINEENARMONVERVHAGIJAFOORMERASDIAGTET HOUIMDINEENARMONVERTP GPLFAGGLARFGSDIAGTET HOUIMDINEENARMONVERTP GPLFAGGLARFGSDIAGTET ++12+ ++13++ ++13++ ++15++ ++7+++	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	LYKNNI LLGLUK VYMYSPTS: LYKNNI GLGLGKCVYMYSPTNI LYKNNI GLGLGKCVYMYYNPTNI LYKNNI GLGLGKCVYMYNPTNI	LID INGGEREP PROFY LID INGGEREP PROFY LID INGGEREP PROFY +15+11++ +++6++ ++8+++
RIV-1 SIV RIV-2 RIV-2 B-EPITOPE RIV-2 B-EPITOPE RIV-2 T-CELL/OB RIV-1 T-CELL/OB RIV-2	MOMERATIME EARL MORVEY VERGE LAPOGRAFE PAGED LAGTTST MOITRE ITME EARL MODEL OF THE COORTINE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGED LAGTTST MOITRE ITME EARL MOVE PER GET PAGE PAGE PAGE PAGE PAGE PAGE PAGE PAGE	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	I YERNI QI GLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE I YERNI QLGLQKCVENYEPTE	LD IRQUPREP FROY LLD/WGGPREP FGSY +15+11++ ++6++ ++8+++
HIV-1 SIV BIV-1 SIV BIV-2 BIV-1 BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/DB BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1	MGMLKETINEEAAANDAWEN VHAGE IAPOGRAEPAGSD IAGTEST MGIIRDIINEEAAANDAWEN VHAGE IAPOGRAEPAGSD IAGTEST MGIIRDIINEEAAANDAWEN IAPOGRAEPAGSD IAGTEST WATER AND	T.QEQIGNN THEPPIPVGE VEEQIGNNYEQQIPIPVGE VEEQIGNNFEPQHPVPVG	LYKNNI LLGLUK VYMYSPTS: LYKNNI GLGLGKCVYMYSPTNI LYKNNI GLGLGKCVYMYYNPTNI LYKNNI GLGLGKCVYMYNPTNI	LD IRQUPREP FROY LDOWGOPREP FROY LLDOWGOPREP FROY +15+11++ +++6++ ++8+++
RIV-1 SIV RIV-2 RIV-2 RIV-2 B-EPITOPE RIV-2 B-EPITOPE RIV-2 T-CELL/DB RIV-1 T-CELL/DB RIV-1 T-CELL/DB RIV-2 T-CELL/RI RIV-1 T-CELL/RI RIV-1	HOGHLEFTHEEAARHONNEFVHAGFIAFOGNEFSGSDIAGTTST HQUIRDITHEEAARHONNEFVHAGFIAFOGNEFSGSDIAGTTST HQITALTHEEAARHONNEFTF GFLFAGLREFSGSDIAGTTST ++12++ ++13++ ++13++ ++13++ ++17+++	T.CKGIGIGHM THEP I PVO T.CKGIGHM THEP I PVO TVEGIGHMT THE HIP I PVO TVEGIGHM THE HIP I PVO	LYKNNI LLGLUK VYMYSPTS: LYKNNI GLGLGKCVYMYSPTNI LYKNNI GLGLGKCVYMYYNPTNI LYKNNI GLGLGKCVYMYNPTNI	LD IRQUPREP FROY LDOWGOPREP FROY LLDOWGOPREP FROY +15+11++ +++6++ ++8+++
RIV-1 SIV RIV-2 RIV-1 RIV-2 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-1 R-EPITOPE RIV-1 T-CELL/OB RIV-1 T-CELL/OB RIV-1 T-CELL/RT RIV-1 T-CELL/RT RIV-1 T-CELL/RT RIV-1	HOULKETINEEAANDNUMEN VALAGI IAPOGREP ROSD LAGTTEF HOULTED INEEAANDNUMENOO AP COORDIAE PROSD LAGTTEF HOULTED INEEAANDNUMEN OF GPLFAGGLEF ROSD LAGTTEF HOULTE INEEAANDNUMEN OF HOULTE HOUSE LAGTTEF HOULTE INEEAANDNUMEN OF HOULTE HOUSE LAGTTEF HOULTE HEELAANDNUMEN OF HOULTE HOUSE LAGTTEF HOULTE HOULTE HOUSE LAGTTEF HOULTE HOULTE HOUSE LAGTTEF HOULT	T.CKGIGIGHM THEP I PVO T.CKGIGHM THEP I PVO TVEGIGHMT THE HIP I PVO TVEGIGHM THE HIP I PVO	U. ITERNIT LIGHTEN VANITATES INTERNIT CALLEGE CONTRIBER OF CALLEGE CONTR	LID INGGEREPRONY LID INGGEREPRONY LID INGGEREPRONY +15+11++ +++6+- ++8+++
RIV-1 SIV RIV-1 SIV RIV-1 SIV RIV-2 RIV-2 B-EPITOPE RIV-1 B-EPITOPE RIV-2 T-CELL/OB RIV-1 T-CELL/OB RIV-1 T-CELL/OB RIV-1 T-CELL/RT RIV-1 T-CELL/RT RIV-1 T-CELL/RT RIV-2 EIV-1 VARIANCE	HOGHLEFTHEEAARHONNEFVHAGE JAPOGUREPAGED LAGTEST HOGHLEFTHEEAARHONNEFVHAGE JAPOGUREPAGED LAGTEST HOGHTALTHEEAARHONNEFTP GPLFAGGLREPAGED LAGTEST ++12++ ++13++ ++13++ ++13++ ++13++ ++13++	TLOGGIGHM THREFIFMEN VERGIGHMTAR GHIPVPV VERGIGHMTAR GHIPVPV VERGIGHMTAR GHIPVPV A 350 360	V V ITYRNIT ILGLINK VPMPYSPTS ITYRNIT GLEGGCVMPYMPTN ITYRNIT GLEGGCVMPYMPTN ITYRNIT GLEGGCVMPYMPTN ITYRNIT GLEGGCVMPYMPTN ITYRNIT GLEGGCVMPYMPTN ITYRNIT GLEGGCVMPYMPTN ITYRNIT GLEGGCVMPTMPTN ITYRNIT GLEGGCVMPTMPTMPTN ITYRNIT GLEGGCVMPTMPTN ITYRNIT GLEGGCVMPTMPTMPTN ITYRNIT GLEGGCVMPTMPTMPTMPTN ITYRNIT GLEGGCVMPTMPTMPTMPTMPTMPTMPTMPTMPTMPTMPTMPTMPTM	**************************************
HIV-1 SIV BIV-2 BIV-1 BIV-2 BIV-2 BIV-1 BIV-2 B-EPITOPE BIV-2 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB BIV-2 T-CELL/R BIV-1 T-CELL/R BIV-1 T-CELL/R BIV-1 T-CELL/R BIV-1	HOULKETINEEAANNAVERVALASIAFOOMEPANSDIAGTTEE HOITEDINEEAANNAVEROO AP COORDEPANSDIAGTTEE	Tuggiciam Thereisve Tuggiciam Thereisve Vezgiciam Thereisve Vezgic	U V VIIVENI ILGUME VANNYSPTS: ITERNI GLOGOCYMNISPTM ITERNI GLOGOCY	LID INGGPREP PRODY LID INGGPREP POSY LID INGGPREP POSY +15+11++ +++6++ ++8+++ 390 400 vvvvvvv
HIV-1 SIV BIV-2 BIV-1 BIV-2 BIV-2 BIV-1 BIV-2 B-EPITOPE BIV-2 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB BIV-2 T-CELL/R BIV-1 T-CELL/R BIV-1 T-CELL/R BIV-1 T-CELL/R BIV-1	HOULKETIME HARMONWEN VHAD IAPOGREP HOSD LAGTES HOUISENGED LAGTES HOUISENGE HOUISENGED LAGTES HOUISENGED LAGTES HOUISENGED LAGTES HOUISENGE HOUISENGED LAGTES HOUISENGED LAGTES HOUISENGE HOUIS	TUESDIAN THEFFIFUR VECTOMER PROFIT VECTOMER PROFIT 350 360 VILLEBURACQUEGROUS	U V V V V V V V V V V V V V V V V V V V	11D INGGPKEPFEDY LID INGGPKEPFEDY LID INGGPKEPFGSY +15+11++ +++6++ +18+++ 390 400 VVVVVVVV NEMORRI IVECPHCE NEMORRI IVECPHCE OCKEPERFI INCHMCC
RIV-1 SIV RIV-1 SIV RIV-2 RIV-2 RIV-2 B-EPITOPE RIV-1 B-EPITOPE RIV-2 T-CELL/OB RIV-1 T-CELL/OB RIV-1 T-CELL/OB RIV-1 T-CELL/RT RIV-1 T-CELL/RT RIV-1 T-CELL/RT RIV-1 SIV RIV-1 SIV RIV-1 SIV RIV-1 SIV RIV-2	HOULKETIME HARMONWEN VHAD IAPOGREP HOSD LAGTES HOUISENGED LAGTES HOUISENGE HOUISENGED LAGTES HOUISENGED LAGTES HOUISENGED LAGTES HOUISENGE HOUISENGED LAGTES HOUISENGED LAGTES HOUISENGE HOUIS	TUESDIAN THEFFIFUR VECTOMER PROFIT VECTOMER PROFIT 350 360 VILLEBURACQUEGROUS	U V V V V V V V V V V V V V V V V V V V	11D INGGPKEPFEDY LID INGGPKEPFEDY LID INGGPKEPFGSY +15+11++ +++6++ +18+++ 390 400 VVVVVVVV NEMORRI IVECPHCE NEMORRI IVECPHCE OCKEPERFI INCHMCC
HIV-1 SIV BIV-2 BIV-1 BIV-2 BIV-2 BIV-1 BIV-2 B-EPITOPE BIV-2 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB BIV-2 T-CELL/R BIV-1 T-CELL/R BIV-1 T-CELL/R BIV-1 T-CELL/R BIV-1	HOSELETINE HARMONUM VERAS LAPOSMET MASS LAGTET MOLITAL INVESTMENT AND AND LASTES HOSELET MOLITAGE HARMONUM PROPERTY OF LAPOSMET MOLITAGE HARMONUM PROPERTY OF LAPOSMET MOLITAGE HARMONUM PROPERTY HARMONUM PROPERTY HARMONUM PROPERTY HARMONUM PROPERTY LANGUAGE PARTY NAME AND	TUEBOLING STATE PER PER PER PER PER PER PER PER PER PE	U V V V V V V V V V V V V V V V V V V V	11D INGGPKEPFEDY LID INGGPKEPFEDY LID INGGPKEPFGSY +15+11++ +++6++ +18+++ 390 400 VVVVVVVV NEMORRI IVECPICE NEMORRI IVECPICE OCKEPERFI INCINCE
EIV-1 SIV EIV-2 EIV-2 EIV-1 EIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-1 T-CELL/OB BIV-1 T-CELL/OB BIV-1 T-CELL/RT BIV-1 T-CELL/RT SIV T-CELL/RT SIV T-CELL/RT SIV-1 EIV-1 SIV EIV-1 EIV-1 EIV-1 EIV-2 EIV-1-BIV-2	HOULKETIME HARMONWEN VHAD IAPOGREP HOSD LAGTES HOUISENGED LAGTES HOUISENGE HOUISENGED LAGTES HOUISENGED LAGTES HOUISENGED LAGTES HOUISENGE HOUISENGED LAGTES HOUISENGED LAGTES HOUISENGE HOUIS	TUEBOLING STATE PER PER PER PER PER PER PER PER PER PE	U V V V V V V V V V V V V V V V V V V V	11D INGGPKEPFEDY LID INGGPKEPFEDY LID INGGPKEPFGSY +15+11++ +++6++ +18+++ 390 400 VVVVVVVV NEMORRI IVECPICE NEMORRI IVECPICE OCKEPERFI INCINCE
RIV-1 SIV RIV-2 RIV-1 RIV-2 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2	MODELETINE HARMONUM PRAGI IAPOGREP MOSD LAGTES MODIFICATION AND THE MODI	TUEBOLING STATE PER PER PER PER PER PER PER PER PER PE	ITERNIT LIGILIK TVANYEPPS ITERNIT GLIGGECVANYEPPK ITERNIT GLIGGECVANYEPK ITERNIT GLIGGECVANYEPK ITERNIT GLIGGECVANYEPK ITERNIT GLIGGECVANYEPK ITERNIT GLIGGECVANYEPK ITERNIT GLIGGECVANYEPK ITERNIT GL	11D INGGPREPPRDY 11D INGGPREPPGSY 11D INGGREGO Q QRAFFINGENECG Q QRAFFINGENECG
EIV-1 SIV EIV-2 EIV-2 EIV-1 EIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-1 T-CELL/OB BIV-1 T-CELL/OB BIV-1 T-CELL/RT BIV-1 T-CELL/RT SIV T-CELL/RT SIV T-CELL/RT SIV-1 EIV-1 SIV EIV-1 EIV-1 EIV-1 EIV-2 EIV-1-BIV-2	HOULKETIME HARMONYMEN VALAGI JAPOGMEP MOSD LAGTTET HOULKETIME HARMONYMEN VALAGI JAPOGMEP MOSD LAGTTET HOULKETIME HARMONYMEN P GPLPAGGLEF MOSD LAGTTET HOULKETIME HARMONYMEN P GPLPAGGLEF MOSD LAGTTET HOULKETIME HARMONYMEN P GPLPAGGLEF MOSD LAGTTET HOULKET LAGTTE MOSD LAGTTET HOULKET LAGTTE MOSD LAGTTET	TUEBOLING STATE PER PER PER PER PER PER PER PER PER PE	U V V V V V V V V V V V V V V V V V V V	11D INGGPKEPFEDY LID INGGPKEPFEDY LID INGGPKEPFGSY +15+11++ +++6++ +18+++ 390 400 VVVVVVVV NEMORRI IVECPICE NEMORRI IVECPICE OCKEPERFI INCINCE
EIV-1 SIV EIV-2 EIV-2 EIV-1 EIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-1 T-CELL/DB BIV-1 T-CELL/DB BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 EIV-1 SIV EIV-1 SIV EIV-1 BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-1 B-EPITOPE BIV-1	HOGHLETTHEEAARHOUNES VERAE JAPOGUREP SOSD LAGTEST MQILRETHEEAARHOUNES VERAE HQITAZITHEEAARHOUNES TP GRIPAGULEP SOSD LAGTEST HQITAZITHEEAARHOUNES TP GRIPAGULEP SOSD LAGTEST ++12++ ++13++ ++13++ ++13++ ++13++ ++13++ 310 320 330 340 V V V V V V V V V V V V V V V V V V V	TUEBOLING STATE PER PER PER PER PER PER PER PER PER PE	TYRNIT LLGLIKE VYMYSPTS: ITRINIT GLOGGECVIMYSPTM: ITRINIT GLOGGECVIMYSP	11D INGGPREPPRDY 11D INGGPREPPGSY 11D INGGREGO Q QRAFFINGENECG Q QRAFFINGENECG
RIV-1 SIV RIV-2 RIV-1 RIV-2 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RI	HOULKETINEERARMONUME VARAE IAPOQUEE MOSDIAGTEE HOUINDINEERARMONUMET P GPLFAGGLAEF MOSDIAGTEE HOUINDINEERARMONUMET P GPLFAGGLAEF MOSDIAGTEE ++12+ ++13+ ++13+ ++13++ ++13++ ++14+ 310 320 330 340 V V V V V V V V V V V V V V V V V V	TUEBOLING STATE PER PER PER PER PER PER PER PER PER PE	TYRNIT LLGLIKE VYMYSPTS: ITRINIT GLOGGECVIMYSPTM: ITRINIT GLOGGECVIMYSP	11D INGGPREPPRDY 11D INGGPREPPGSY 11D INGGREGO Q QRAFFINGENECG Q QRAFFINGENECG
RIV-1 SIV RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RI	HOUREFINEERARMONUME VARAE IAPOQUEE MOSDIAGTEE HOUIDS INVERENAMENDAMENO AP COORDER MOSDIAGTEE HOUIDS INVERENAMENDAMENO AP COORDER MOSDIAGTEE HOUIDS INVERNAMENDAMEN APPLICATION	TLEBOLIA CONVERT DEVI	JYOU SEE CONTROL OF S	SID INGGEREP FROM LIDVEGOPREP FROM LIDVE
EIV-1 SIV EIV-2 EIV-2 EIV-1 EIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-1 T-CELL/DB BIV-1 T-CELL/DB BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-2 EIV-1 VARIANCE EIV-1 SIV EIV-1 EIV-1 EIV-2 B-EPITOPE BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-1 EIV-1 EIV-1 EIV-2 B-EPITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/DB BIV-1	HOSELETTIME PARK MONVER VERAS LAPOGREP SOSD LAGTET MODIFIED LIMITED LIMITED LAPOGREP SOSD LAGTET HOST LATER AND AVERAGE POR CONTROLLER SOSD LAGTET HOST LIMITED LAW WITH A CONTROLLER SOSD LAGTET HOST LIMITED LAW WITH A CONTROLLER SOSD LAGTET HOST LIMITED LATER AND LAGTE HOST LAGTEN HOST LIMITED LATER AND LAGTE HOST LAGTEN HOST LAGTEN HOST LAGTEN HOST LAGTEN VORT YKEL BAROZD PAVENMENT TILL OGRAMP DOEN LIALAPAN VORT YKEL BAROZD PAVEN TILL OGRAMP DOEN LIALAPAN VORT YKEL BAROZD PAVEN TILL TILL TILL TILL TILL TILL TILL TIL	LOGGIGNM THREFIFM LOGGIGNM THREFIFM VERGIGNMTRAPGHEVPV VERGIGNMTRAPGHEVPV VERGIGNMTRAPGHEVPV VILENTARGEVERGENGUM TLENTARGEVERGENGUM TLENTARGEVERGENGUM TLENTARGEVERGENGUM TLENTARGEVERGENGUM	JYON TILGLINK VANNYPPS: IYRANI GLOGGOVANNYNPH IYRANI GLOGGOVANNYNP	11D INGGPREPPRDY 11D INGGPREPPGSY 11D INGGREGO Q QRAFFINGENECG Q QRAFFINGENECG
RIV-1 SIV RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RIV-2 RIV-1 RI	HOUREFINEERARMONUME VARAE IAPOQUEE MOSDIAGTEE HOUIDS INVERENAMENDAMENO AP COORDER MOSDIAGTEE HOUIDS INVERENAMENDAMENO AP COORDER MOSDIAGTEE HOUIDS INVERNAMENDAMEN APPLICATION	TLEBOLIA CONVERT DEVI	JYOU SEE CONTROL OF S	SID INGGEREP FROM LIDVEGOPREP FROM LIDVE
SIV-1 SIV SIV-1 SIV-1 SIV-1 SIV-2 B-EPITOPE BIV-2 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB BIV-1 T-CELL/RT BIV-2 T-CELL/RT BIV-2 EIV-1 VARIANCE BIV-1 SIV BIV-1 SIV-2 BIV-1 SIV-2 BIV-1 SIV-2 BIV-1 SIV-2 BIV-1 SIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1	HOULKETHEEAARMONYEVHAGIJAPOOMIEPASSIJAGTEV HOIIMSIJABAANONORPOO AP COORDEPASSIJAGTEV HOIIMSIJABAANONORPIP GPLPAGOLAEPASSIJAGTEV HOIIMSIJABAANONORPIP GPLPAGOLAEPASSIJAGTEV HOIIMSIJABAANONORPIP GPLPAGOLAEPASSIJAGTEV HOIIMSIJABAANONORPIP GPLPAGOLAEPASSIJAGTEV HOIIMSIJABAANONORPIP GPLPAGOLAEPASSIJAGTEV HOIIMSIJABAANONORPIP GPLPAGOLAEPASSIJAGAANONORTILIKALGPAN VORTYKALAARGYDPANVANNETYTLINGMANPOCKVILKALGPAN VORTYKALAARGYDPANVANNETYTLINGMANPOCKVILKALGPAN VORTYKALAARGYDPANVANNETYTLINGMANPOCKVILKALGPAN VORTYKALAARGYDPANVANNETYTLINGMANPOCKVILKALGPAN VORTYKALAARGYDPANVANNETYTLINGMANPOCKVILKALGPAN VORTYKALAARGYDPANVANNETYTLINGMANPOCKVILKALGPAN VORTYKALAARGYDPANVANNETYTLINGMANPOCKVILKALGPAN VORTYKALAARGYDPANVANNETYTLINGMANPOCKVILKALGPAN VORTYKALAARGYDPANVANNETYTLINGMANPOCKVILKALGPAN VORTYKALAARGYDPANVANNETYTLINGMANPOCKVILKALGAPAN	TLEEDIGHM THREF I FVO VERGI CHRISTORIO FI I FVO VERGI CHRISTORIO FI I FVO VERGI CHRISTORIO FI FVO VERGI CHRISTORIO FI FVO VILLE ENTRACQUI VIGINI FI VILLE ENTRACQUI VIGINI FI VILLE ENTRACQUI VIGINI FI VILLE ENTRACQUI FI FVO VILLE ENTRACQUI FVO VILLE ENTRACQUI FVO VILLE ENTRACQUI FVO VILLE ENTRACQUI FI FVO VILLE ENTRACQUI FVO VILLE ENTRA	JYOU TICHUR VENEYERS ITERNIT GLOCKOWNING PRINT ITERNIT GLOCKOWNING PRI	SID INGGPREP PRODY LID INGGPREP PROSY LID INGGPREP PROSY +15+11++ +++6++ ++8+++ 390 400 vvvvvvv NFRINGRI I VICEPICG GOKGP RIFT I INCIDING Q GREAT PICHICO Q GREAT PICHICO CONCERNE INFORMATION +++20++
EIV-1 SIV EIV-2 EIV-2 EIV-1 EIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-1 T-CELL/DB BIV-1 T-CELL/DB BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-2 EIV-1 VARIANCE EIV-1 SIV EIV-1 SIV EIV-2 B-EPITOPE BIV-1 T-CELL/DB BIV-1	MOREKETINEERAANDAVEN VRAGE JAPOGREP MOSD LAGTEN HQIIRDI JIREERAANDAVEN OA 20 GOOGLEP MOSD LAGTEN HQIIRDI JIREERAANDAVEN OA 20 GOOGLEP MOSD LAGTEN H13+++13+++13++ ++13++ ++13++ ++13++ ++13++ ++13++ ++13++ ++13++ ++14++ VORFYKZLAARGANG VKHNOTETILVGHAMP OCKT LIKALGPAN VKH	TLEBOLINA CHEEP I PVO	TYRNIT ILGLIK VINITYPES TYRNIT GLECOKYMYTHPEN TYRNIT GLECOKYMYTHPE	**************************************
SIV-1 SIV SIV-1 SIV-1 SIV-1 SIV-2 B-EPITOPE BIV-2 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB BIV-1 T-CELL/RT BIV-2 T-CELL/RT BIV-2 EIV-1 VARIANCE BIV-1 SIV BIV-1 SIV-2 BIV-1 SIV-2 BIV-1 SIV-2 BIV-1 SIV-2 BIV-1 SIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1	HOULKETIME HARMONEW VAND IAPOGMEN BOSD LAGTEN HOUITED IMPERANCH DIRECT AND CONCLETE STATES HOUITED IAPOGMEN BOSD LAGTEN HOUITED IMPERANCH DIRECT AND CONCLETE STATES HOUITE IMPERANCH DURCH PRODUCT AND CONCLETE STATES HOUITED IAPOCH AND CONCLETE STATES HOUITED IA	TLEEDIGHM THREF I FVO VERGI CHRISTORIO FI I FVO VERGI CHRISTORIO FI I FVO VERGI CHRISTORIO FI FVO VERGI CHRISTORIO FI FVO VILLE ENTRACQUI VIGINI FI VILLE ENTRACQUI VIGINI FI VILLE ENTRACQUI VIGINI FI VILLE ENTRACQUI FI FVO VILLE ENTRACQUI FVO VILLE ENTRACQUI FVO VILLE ENTRACQUI FVO VILLE ENTRACQUI FI FVO VILLE ENTRACQUI FVO VILLE ENTRA	TYRNIT ILGLIK VINITYPES TYRNIT GLOCKOWNITYPEN TYRNIT GLOCKOWNITYPE	**************************************

Fig.1. The alignment of gag HIV-1, HIV-2 and SIV_{MAC} sequences. The '*' on the top row indicates identities between HIV-1 and SIV_{MAC}. '*' on the bottom row indicates identities between HIV-1 and HIV-2. A '+' denotes the core prediction for the B epitopes where the number gives the rank peak height. A '-' illustrates predicted positions for T epitopes by both the DeLisi and Berzofsky (BD) method and the Rothbard and Taylor (RT) method. A 'v' denotes an inter-strain variable position and a 'g' indicates where it was necessary to insert a gap into the alignment.

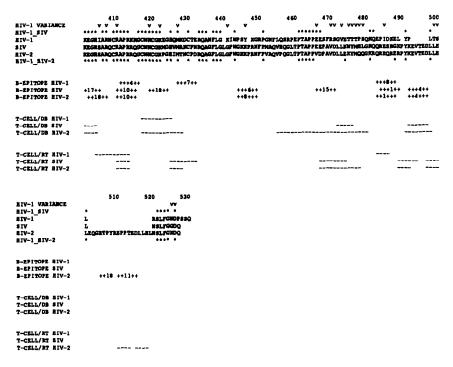


Fig.1 (contd).

periments [17,18] suggest the core antigen of hepatitis-B may induce protection in chimpanzees against viral challenge. Whether any of this applies to an AIDS vaccine is unknown, but T-cell-mediated immunity against HIV-1 env and gag proteins, both helper and cytotoxic, has been recorded [19-21]. In addition antibody against HIV-1 core (gag) antigen disappears, in many patients, prior to the onset of the disease, but it is not clear whether this indicates that an anti-gag immune response prevents the onset of the disease [22] or whether it merely accompanies progression towards AIDS.

In this paper, we have performed epitope predictions for T- and B-cell sites. T-cells may recognize different [23] parts of the chain from B cells but there is evidence that protective B epitopes of influenza virus overlap with T epitopes [24].

2. MATERIALS AND METHODS

The studies were carried out on the HIV-1_{LAV} virus, the simian virus (SIV_{MAC}) and HIV-2_{ROD}. The sequences for the HIV-1 virus were obtained from the Protein Information Resource Databank (see [15] for further details). The sequences

for other HIV-1 strains, SIV_{MAC}, HIV-2 were obtained from the Los Alamos National Database [25].

2.1. Sequence variation between the HIV-1 strains

Ten gag, 13 env and 8 pol HIV-1 sequences were aligned by the method of Barton and Sternberg [26] to investigate interstrain variation. In figs 1-3 a 'v' denotes inter-strain variation for HIV-1 and a 'g' denotes the introduction of a gap into the alignment. In tables 1 and 2, %SI denotes the identity between HIV-1 isolates within a predicted epitope.

2.2. Sequence variation between HIV-1, SIV_{MAC} and HIV-2

A multiple alignment of HIV-1, SIV_{MAC}, and HIV-2 sequences was obtained [26]. In figs 1-3 an identity between HIV-1 and SIV_{MAC} is indicated by a * on the top row, while identities between HIV-1 and HIV-2 are indicated by a * on the bottom row of the alignment.

2.3. Principle of analyses

Predictions of B-cell and T-cell epitopes (see sections 2.4 and 2.5) were performed on the HIV-1, HIV-2 and SIV_{MAC} sequences (figs 1-3). These figures provide information to guide vaccine design based on many different criteria. For example, peptides might be selected on sequence conservation or on regions of high variation. In this paper we analyse conservation of sequences and of predicted antigenic epitopes. Predicted epitopes were taken to be at least 12 residues long as peptides of this length have frequently been shown to contain an epitope [27-29]. This analysis, of course, considers only linear epitopes.

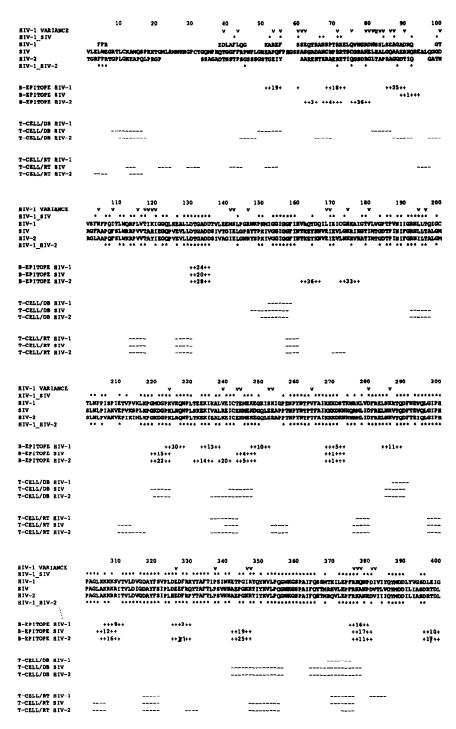


Fig.2 (contd).

V-1 VALUE V		410 420 430 440	450 460 470 480	490 500
The content of the		* * * * *	*** ** * * * * *	V VVV
### 1	BIV-1	QHRTKI EELAQHLLAMGLTTPDKKHQKEPPFLMMGYELHPDKI	wtvop i vlpendantvnd i oklvgklanasqi ypgi kvrqi	CKLLAGTKALTEVIPLT
### 1-1 PATIONS #IV-1 ### 1-2-11700 ## 11-2	RIV-2	EHDRYVLOLKELLNGLGFSTPDERFORDPP YHMIGYELMP TRI	eklori glegke i wtvnd i gklygyl nnaaglypgi ktrei	CRL IRGINATE TEEVONT
### 1-201-100 STV-1 1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1	# [V-1_H [V-2			
### 12 12 12 12 12 12 12 1	B-EPITOPE BIV-1	++14++ +++6++	++31++	
T-CELL/DS SIV-1 T-CELL/DS SIV-2 T-CELL/DS SIV-				
T-CELL/SS SIV-2	S-EPITOPE BIV-2	12**		
T-CELL/PR SIV-1 T-CELL/PR SIV-1 T-CELL/PR SIV-2 T-CELL/PR SIV-				
T-CELL/FF SIV-1 T-CELL				
T-CELL/EF SIV-1 S10 520 530 540 550 540 570 580 590 600 SIV-1 (MAILMEE) S-217072 SIV-1 SIV-1 (MAILMEE) SIV-1 (
### 1				
### 1310 \$20 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$30 \$40 \$				
SIV-1 MAILINEE				
SIV-1 MAILINEE		E10 E20 E20 E40		500 600
### ### ### ### #### #### #### #### ####		v v v	AA AAA A AAAA	VV V
SIV				
### ### ##############################		emaraeyzenki ilsqeqegcyyqeskplratvi ksqdnqns	ykinged kilkvokfakikntethgvrllanvigkighd	INIOCONSKEETS AFKO
3-PITTORE SIV-2 3-PITTORE SIV-2 7-CELL/OB SIV-1 7-CELL/OB SIV-1 7-CELL/OB SIV-2 8-PITTORE SIV-1 8-PITTORE SIV-1 8-PITTORE SIV-1 8-PITTORE SIV-2 7-CELL/OB SIV-		ETATAL DEDAKT I DEGLESSI I GERELEKI VORAL SAMI	INTEGER RICKYGRIAKYRWIHINGIRLDMOYYORIGRD	TAIMENIAN AT A
3-PITTORE SIV-2 3-PITTORE SIV-2 7-CELL/OB SIV-1 7-CELL/OB SIV-1 7-CELL/OB SIV-2 8-PITTORE SIV-1 8-PITTORE SIV-1 8-PITTORE SIV-1 8-PITTORE SIV-2 7-CELL/OB SIV-				
### ### ##############################	B-EPITOPE RIV-1	++15++ ++21++ ++29++		
T-CELL/NT SIV-1 SIV-			*** 26***	+++
T-CELL/NT SIV-1 SIV-				
T-CELL/RT SIV-1 T-CELL/RT SIV-2 SIV-1 VARIANCE SIV-1 SIV-1 SIV SIV-1 SIV SIV-1 SI				
T-CELL/RT BIV-2 610 620 630 640 550 660 670 680 690 700 BIV-1 VANIANCE BIV-1 WENTERT THOUTH PERFEVENT PLANEAUTH FOR SERVING STREET REACTIVE THOUGHT FOR SERVING STREET FOR SERVING TO SERVING TO SERVING THOUGHT FOR SERVING TH				
T-CELL/RT BIV-2 610 620 630 640 550 660 670 680 690 700 BIV-1 VANIANCE BIV-1 WENTERT THOUTH PERFEVENT PLANEAUTH FOR SERVING STREET REACTIVE THOUGHT FOR SERVING STREET FOR SERVING TO SERVING TO SERVING THOUGHT FOR SERVING TH				
### ##################################				
### 1 VANIANCE ### 19 V V V V V V V V V V V V V V V V V V				
### 1 VANIANCE ### 19 V V V V V V V V V V V V V V V V V V				
### 1 VANIANCE ### 19 V V V V V V V V V V V V V V V V V V				
No.	RIV-1 VARIANCE		V VV V VV VVVV	v v
SIV	8IV-1_5IV			
B-EFITOPE BIV-1 B-EFITOPE BIV-1 B-EFITOPE BIV-1 T-CELL/DB BIV-1 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T-CELL/RT BIV-1 T-CELL/RT BIV-2 T-CELL/RT BIV-1 B-EFITOPE BIV-1 B-EFITOPE BIV-1 B-EFITOPE BIV-1 T-CELL/RT BIV-1	SIV	VWEQUATDYWQVTWIPEMDFISTPPLVALVFNLVRDPIEGEET	YYYDGSCSKQSKEGKAGYITDRGKDKVKVLEQTTNQQAEL	eaflmaltdsgpkan I I
B-EPITOPE BIV-1 B-EPITOPE BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-				
S-EPITOPE RIV-1	-			
S-EPITOPE RIV-1				
### ##################################		++32++		
T-CELL/RT SIV-1 T-CELL/RT SIV-1 T-CELL/RT SIV-2 710 720 730 740 750 760 770 780 790 800 RIV-1 VARIANCE RIV-1-SIV RIV-1-SIV RIV-1-SIV RIV-1-SIV RIV-1-SIV-2		***		
T-CELL/RT SIV-1 T-CELL/RT SIV-1 T-CELL/RT SIV-2 710 720 730 740 750 760 770 780 790 800 RIV-1 VARIANCE RIV-1-SIV RIV-1-SIV RIV-1-SIV RIV-1-SIV RIV-1-SIV-2				
T-CELL/RT BIV-1 T-CELL/RT BIV-2 T-CELL/RT BIV-2 T10	T-CELL/DB HIV-1 T-CELL/DB SIV			
T-CELL/RT SIV 710 720 730 740 750 760 770 780 790 800 RIV-1 VARIANCE RIV-1 750 750 760 770 780 790 800 RIV-1 SIV V V V V V V V V V V V V V V V V V V				
T-CELL/RT SIV 710 720 730 740 750 760 770 780 790 800 RIV-1 VARIANCE RIV-1 750 750 760 770 780 790 800 RIV-1 SIV V V V V V V V V V V V V V V V V V V				
TIV-1 VARIANCE	T-CELL/RT BIV-1			
### ##################################	T-CELL/RT HIV-2	****		
### ##################################				
SIV		710 720 730 740	750 760 770 780	790 800
TOSQTALGI IOAGO DIRESERIANG LIBOLI DIREMVILANOPARRIGI GONICONDIAUSAGI RAYUFLOG IDIAGOEREKYESHIRANARADIPELPVAREI SIV-1 VOSQTYMOI IARAGO PERSERIANG LIBOLI DIREMVILANOPARRIGI GONICONDIAUSAGI RAYUFLEK IERAGERESKYESHI IRELVEK NIGLERIAVARII SIV-1			······ · · · · · · · · · · · · · · · ·	
## ## ## ## ## ## ## ## ## ## ## ## ##	8 (V-1		ekgiggnegydklysagirkylflogidkagdehekyeshi	
B-EPITOPE BIV-1 +++1++ ++20++ ++26++ +++7++ B-EPITOPE SIV ++14++ ++16++ ++16++ ++23++ ++7++ T-CELL/DB BIV-1	#1V-2	vdsqyvng i basqpte se sk i vng i i emikkea i yvanvpa	hkgi ggngevdelvsog i rovlflek i epage ehekyesh	ykelsekpcipnlvarqi
B-EPITOPE SIV +-14++ +-16++ +-	H1A-1_H1A-5			
B-EPITOPE SIV +-14++ +-16++ +-	3	444144 349044		
7-CELL/08 BIV-1 7-CELL/08 SIV	B-EPITOPE SIV	++14++ ++16++	++8+++	
7-CELL/D8 SIV	B-EPITOPE BIV-2	++20++ ++23++	++7+++	
7-CELL/D8 SIV	T-CELL/DB BIV-1			
7-CELL/R7 BIV-1	T-CELL/DB SIV	******		
T-CELL/RT SIV	T-CELL/DB RIV-2			
T-CELL/RT SIV	T-CELL/RT BIV-1			
1	T-CELL/RT SIV			
	A-CETP\ML HIA-5			

Fig.2 (contd).

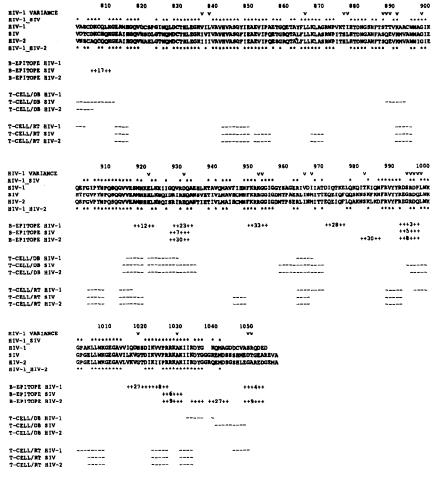


Fig.2. The alignment of pol HIV-1, HIV-2 and SIV_{MAC}.

The identity between HIV-1 and HIV-2 was calculated for positions in the HIV-1 predicted epitope irrespective of whether an epitope was predicted in HIV-2 (denoted as HIV-1 %1-2). Similarly HIV-1 %1-S denotes the identity between HIV-1 and SIV_{MAC}. Other measures of sequence identity between epitopes were considered but these did not markedly alter the results. Tables 1 and 2 list all B and T epitopes with %SI of 100%. Epitopes with 1 or 2 sequence variable positions are reported in the text.

2.4. B epitopes

B-cell epitopes were predicted by the algorithm of Hopp and Woods [30] which searches for a local maximum in a hydrophilicity profile smoothed over 6 residues. Hydrophilic peaks were selected in order of decreasing hydrophilicity. The 6 residue predicted B-epitope core (denoted by a '+' in the figures) was extended on either side by 3 residues to obtain a 12 residue long polypeptide (called the predicted epitope). In our previous study [14,15] we concentrated on B epitopes that did not occur within a predicted secondary structure. Because of the sequence variation between HIV-1, HIV-2 and SIV_{MAC}, dif-

ferent predictions of secondary structure were obtained (results not shown). Given the limited accuracy of secondary structure predictions, a structural restriction on B epitopes would unduly complicate the analysis.

2.5. T epitopes

Two algorithms to predict T-cell epitopes were applied, and the results denoted by a '-' in the figures. (a) The method of DeLisi and Berzofsky [23] calculates the amphipathicity of a section 11 residues long and regions with an α -helix periodicity are identified as possible T-cell epitopes. (b) The procedure of Rothbard and Taylor [29] scans for a linear pattern composed of a charged residue or a glycine, followed by 2 or 3 hydrophobic amino acids and terminated by a charged or polar residue.

The predicted T epitope regions can be longer or shorter than the 12 residues defined for the B epitopes. In this paper, the core prediction was extended towards the N- and C-termini to a polypeptide of minimum 12 residues or to as long as T epitopes positions were predicted by either method and overlapped by at least one position.

	10 20	30 40	50 60	70 90	90 100
HIV-1 VARIANCE HIV-1 SIV	***********************	v v vvv	, v vv v	* * * * *	v v v v
H1V- 1	MRVKERYQHLMMGWMG7WLLGI	LMICSATERLEVIVYYGV	PVMKEATTTLPCASDAKAYD	TEVHIVNIATRACVPTDPHPQEVV	
51V 81V-2	MGCLGNQLLI AILLISV MMQLLI AILLASA		Panrhati Plecation Penghati Plecation	OTHERTIGELPONDEYSELA DTHERTIGELPONDEYGEIT	
HIV-1 HIV 2	•				
B-EPITOPE SIV	***6**		++26 ++15	**	++2
B-EPITOPE HIV 2			++14	++ ++15++	
T-CELL/DB HIV-1					
T-CELL/DB SIV					
T-CELL/DB RIV-2					
T-CELL/RT BIV-1					
T-CELL/RT SIV					
T-CELL/RT HIV-2					
	110 120	130 140	150 160	170 180	190 200
HIV-1 VARIANCE	VV		**************************************		
BIV-1_SIV BIV-1	VEGHEDI ISLKOQSLKPCVKLTP		DLONATUTHE SUTH SE SCEN		INCSPHISTS INCK
SIV	TEGALEDVINGLE ETS I KPCVKLSP				
MIV-2	Tega i edvmilpets i kpcvkltp	LCVANKCESTES S	TGWYTTSKSTSTTTTTYYTDQ	eqe i sedtpcaradocoglgree	T INCOMMETCLEROK
BIA-1 BIA-5	** ** * *******	••• ••	• • • •	•	••••

B-EPITOPE HIV-1 B-EPITOPE SIV	2+++15++ ++23++	++4+++	++14++	+++5+	++1
B-EPITOPE HIV-2	++23++		++1	0++ ++23++	++1+
T-CELL/DB BIV-1 T-CELL/DB SIV					
T-CELL/DB BIV-2					
T-CELL/RT HIV-1					
T-CELL/RT SIV T-CELL/RT SIV-2					
- Cabb, R1 817 2					
	210 220	230 240	250 260	270 280	290 300
HIV-1 VARIANCE HIV-1 SIV	AAAAA A AA AAAAA	v vvv vvv	. * * * *		. * *
HIV-1_SIV HIV-1	VORTA FFYKLDI IF IDEDT	V VVV VVV SYTLTSCHTSVITQAC	PRVEPEPIPIRTCAPAGFAI	LECKERTY STOP CTEVETYO	V V
HIV-1_SIV HIV-1 SIV	VONEYA PPYKLDIIPIDIDTT KREYMETWYSAOLVCEQGMSTGM	STILTSCHISVITQAC	Prvepep ip i sycapagpa Dedynda i rcrycappgyat	LECHNITY BOYON CTWV STAGE LECHOTY Y SOPHIPM CS IVVV SSC LECHOTY Y SOPHIPM CS IVVV SSC	THE INPOVE TOLLLE
HIV-1_SIV HIV-1	VORTA FFYKLDI IF IDEDT	V VVV VVV SYTLTSCHTSVITQAC SSRCYMHSCHTSVITESC GTQCYMHSCHTSVITESC	Prvepep i pi i stcapagpai Dedyndas ncrtcappgyas Desyndas aprycappgyas	LECHNITY BOYON CTWV STAGE LECHOTY Y SOPHIPM CS IVVV SSC LECHOTY Y SOPHIPM CS IVVV SSC	THE THE THE PERSON
HIV-1_SIV HIV-1 SIV HIV-2 HIV-1_HIV-2	VOREYA FFYKLDI IF IDNOTT KREYNEZWYSADLVCEGMISTEN KNOYNETGYSKDVVCETMIST N	SYTLTSCHIEVITQAC SYTLTSCHIEVITQAC SEACHMECHTEVIQEC GTQCHMECHTEVITESC	Prvepep i pi i stcapagpai Dedyndas ncrtcappgyas Desyndas aprycappgyas	LECHTHYSTRY CTHVSTVQC LECHTHYSTWHESKVVSSC LECHTHYSTVANCSKVVASTC	THE TRY VETGLLIE THE TYPE TWEET
HIV-1_SIV HIV-1 SIV HIV-2 HIV-1_HIV-2 B-EPITOPE HIV-1	VOREYA FFYRLDI IP IDROTT KREYNETWY SADLVCEGON STON KNOTH ETGY SKUVVCETHINST II	V VVV VVV STILTGUTSVITAGE SENCIPHISCUTSVIGEO GTQCTHHISCUTSVITESO ***********************************	PRVETEP I PI I RTCAPAGFAI DEDYMDA I RCRTCAPPGYAI DERYMDA I RPRTCAPPGYAI	LECHTHYSTRY CTHVSTVQC LECHTHYSTWHESKVVSSC LECHTHYSTVANCSKVVASTC	THE THE THE PERSON
HIV-1_SIV HIV-1 SIV HIV-2 HIV-1_HIV-2	VOREYA FFYKLDI IF IDNOTT KREYNEZWYSADLVCEGMISTEN KNOYNETGYSKDVVCETMIST N	V VVV VVV STILTGUTSVITAGE SENCIPHISCUTSVIGEO GTQCTHHISCUTSVITESO ***********************************	PRVEFEFFFFEETAPAGFAI DEDYNDAIRCRYCAPPOYAI DENYMDAIRCRYCAPPOYAI	LECHTHYSTRY CTHVSTVQC LECHTHYSTWHESKVVSSC LECHTHYSTVANCSKVVASTC	THE TRY VETGLLIE THE TYPE TWEET
HIV-1_SIV HIV-1 SIV HIV-2 HIV-1_HIV-2 B-EPITOPE HIV-1 B-EPITOPE SIV	VQRAYA FFYKLDIF IDMOTT KRAYMETWY SADLVCHOGOSPTON KNOYMETWY SKRAYVCETHSEE N 10++ ++28 +++	STILTCHESVITOR SPRINGHESCHESVITOR SPRINGHESCHESVITO	PRVEFEFFFFEETAPAGFAI DEDYNDAIRCRYCAPPOYAI DENYMDAIRCRYCAPPOYAI	LECHTHYSTRY CTHVSTVQC LECHTHYSTWHESKVVSSC LECHTHYSTVANCSKVVASTC	THE TRY VETGLLIE THE TYPE TWEET
HIV-1 SIV HIV-1 SIV HIV-2 HIV-1 HIV-2 B-EFITOPE HIV-1 B-EFITOPE SIV B-EFITOPE HIV-2 T-CELL/DB HIV-1	VQRAYA FFYKLDIF IDMOTT KRAYMETWY SADLVCHQQSBFTGR KRAYMETWY SKRAYVCETHWSF N 10++ ++28 +++	STILTCHESVITOR SPRINGHESCHESVITOR SPRINGHESCHESVITO	PRVEFEFFFFEETAPAGFAI DEDYNDAIRCRYCAPPOYAI DENYMDAIRCRYCAPPOYAI	LECHTHYSTRY CTHVSTVQC LECHTHYSTWHESKVVSSC LECHTHYSTVANCSKVVASTC	THE TRY VETGLLIE THE TYPE TWEET
RIV-1 SIV BIV-2 RIV-2 RIV-1_RIV-2 B-EPITOPE RIV-1 B-EPITOPE RIV-2 T-CELL/DB RIV-1 T-CELL/DB SIV	VQRAYA FFYKLDIF IDMOTT KRAYMETWY SADLVCHQQSBFTGR KRAYMETWY SKRAYVCETHWSF N 10++ ++28 +++	STILTCHESVITOR SPRINGHESCHESVITOR SPRINGHESCHESVITO	PRVEFEFFFFEETAPAGFAI DEDYNDAIRCRYCAPPOYAI DENYMDAIRCRYCAPPOYAI	LECHTHYSTRY CTHVSTVQC LECHTHYSTWHESKVVSSC LECHTHYSTVANCSKVVASTC	THE TRY VETGLLIE THE TYPE TWEET
HIV-1 SIV HIV-1 SIV HIV-2 HIV-1 HIV-2 B-EFITOPE HIV-1 B-EFITOPE SIV B-EFITOPE HIV-2 T-CELL/DB HIV-1	VQRAYA FFYKLDIF IDMOTT KRAYMETWY SADLVCHQQSBFTGR KRAYMETWY SKRAYVCETHWSF N 10++ ++28 +++	STILTCHESVITOR SPRINGHESCHESVITOR SPRINGHESCHESVITO	PRVEFEFFFFEETAPAGFAI DEDYNDAIRCRYCAPPOYAI DENYMDAIRCRYCAPPOYAI	LICHTHYSTOP CTHVSTVQC LICHTHYSTWIPESKVVSSC LICHTHYSTVANCSKVVASTC	THE TRY VETGLLIE THE TYPE TWEET
BIV-1 BIV-1 BIV-2 BIV-2 BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/DB BIV-2	VQRAYA FFYKLDIF IDMOTT KRAYMETWY SADLVCHQQSBFTGR KRAYMETWY SKRAYVCETHWSF N 10++ ++28 +++	STILTCHESVITOR SPRINGHESCHESVITOR SPRINGHESCHESVITO	PRVEFEFFFFEETAPAGFAI DEDYNDAIRCRYCAPPOYAI DENYMDAIRCRYCAPPOYAI	LICHTHYSTOP CTHVSTVQC LICHTHYSTWIPESKVVSSC LICHTHYSTVANCSKVVASTC	THE TRY VETGLLIE THE TYPE TWEET
BIV-1 SIV BIV-1 SIV BIV-2 BIV-1 BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB SIV T-CELL/AT BIV-1 T-CELL/AT SIV	VOUNT A PPENDLUIT FORMET REST METWYSADLVCESSMESS IN NOTHER CYS RUVVCESSMESS II 10++ ++28 +++	STILTCHESVITOR SPRINGHESCHESVITOR SPRINGHESCHESVITO	PRVEFEFFFFEETAPAGFAI DEDYNDAIRCRYCAPPOYAI DENYMDAIRCRYCAPPOYAI	LICHTHYSTOP CTHVSTVQC LICHTHYSTWIPESKVVSSC LICHTHYSTVANCSKVVASTC	THE TRY VETGLLIE THE TYPE TWEET
BIV-1 SIV BIV-2 BIV-2 BIV-2 BIV-1 BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB BIV-2 T-CELL/TB BIV-2	VOREYA FFYELDIF DINDTE RENY HEFTYRADIV COGNIFICAN KNOTHETOY SKOV VCETHN 87 H 10++ ++28 +++ ++2	STILTCHESVITOR SPRINGHESCHESVITOR SPRINGHESCHESVITO	PRVSFPEP IS LETCARAGES ENTERPRESE IS LETCARAGES ORIETMOA IS PRICAP GUAL **** **** **** **** ****	LICHTHYSTOP CTHVSTVQC LICHTHYSTWIPESKVVSSC LICHTHYSTVANCSKVVASTC	THE TRY VETGLLIE THE TYPE TWEET
BIV-1 SIV BIV-1 SIV BIV-2 BIV-1 BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB SIV T-CELL/AT BIV-1 T-CELL/AT SIV	VOUNT A PPENDLUIT FORMET REST METWYSADLVCESSMESS IN NOTHER CYS RUVVCESSMESS II 10++ ++28 +++	STILTCHESVITOR SPRINGHESCHESVITOR SPRINGHESCHESVITO	PRVSFPEP IS LETCARAGES ENTERPRESE IS LETCARAGES ORIETMOA IS PRICAP GUAL **** **** **** **** ****	LICHTHYSTOP CTHVSTVQC LICHTHYSTWIPESKVVSSC LICHTHYSTVANCSKVVASTC	THE TRY VETGLLIE THE TYPE TWEET
BIV-1 SIV BIV-1 SIV BIV-2 BIV-1 BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB SIV T-CELL/AT BIV-1 T-CELL/AT SIV	VORUM V VV VVV VVVVV VORUM PROTEST IN THE THE TABLE V V V V V V V V V V V V V V V V V V V	V VVV SVV STILLENGTSVICOS SERVINGSVICOS SERV	PRIVATE P P P P P P P P P P P P P P P P P P P	LICHTHYSTOP CTHVSTVQC LICHTHYSTWIPESKVVSSC LICHTHYSTVANCSKVVASTC	TROIRPOWETQLLIS TROOR TOTAL TROOR TROTE ++17
BIV-1 SIV BIV-1 SIV BIV-2 BIV-1 BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/DB SIV T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-2 BIV-1 VARIANCE	VOREYA FFYELDIE FORDYT REKE THE THY SADD VCGETHE FF H 10++ ++28 +++ ++2 110	**************************************	PRIVATE PER LECCA MARTA DESCRIPTION INCREMENTATION OF THE CONTROL OF T	LECTRICTH STORY CTWYSTOC LECTRIC STORY STORY LECTRIC STORY STORY LECTRIC STORY STORY LECTRIC STORY L	TROIRPOWETQLLIS TROOR TOTAL TROOR TROTE ++17
BIV-1 SIV BIV-2 BIV-1 BIV-2 B-EFITOPE BIV-2 B-EFITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB BIV-1 T-CELL/OB BIV-2 T-CELL/AT BIV-1 T-CELL/AT BIV-1 T-CELL/AT BIV-1 T-CELL/AT BIV-1	VOREYA FFYELDIF DIRECT REKY HEFYENDIV GETHER FY HE 10++ ++28 +++ ++2 310 320 VV 320	STELTHORISTINGOUS CONTROL OF THE CON	PRVSFPEP IF I FETCAN AGENT ORIENTIAL AREA AGENT	LIACINSTITUTOR CTWESTOC LIACINSTITUTOR CTWESTOC LIACIN	72012FVWTOLLES 7200ETWIPS 7200ETWIPS ++17
BIV-1 SIV BIV-1 SIV BIV-2 BIV-1 BIV-2 BIV-1 BIV-2 B-EFITOPE BIV-1 B-EFITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/BB BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 BIV-1 VARIANCE BIV-1 SIV	VORVEYA FFYELDIE FORDYT REKT THE THY SADOVCESTEN F7 H 10++ ++28 +++ ++2 310 320 VV V V V V V V V V V V V V V V V V V V	SETELTHORISMIT ON SETEL STATE OF SETEL STATE OF SETEL SET	PRVERFER FERCEMONTAL DESTRUCTION AND ACCEPTANCE OF THE CONTROL OF	LIACINSTITUTOR CTWEFFOR LIACIN	7300 400 7300 400 7300 400 7300 400 7300 400 7300 400
BIV-1 SIV BIV-1 SIV BIV-2 BIV-1 BIV-2 BIV-1 BIV-2 B-EFITOPE BIV-1 B-EFITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/DB BIV-1 T-CELL/BB SIV T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 SIV BIV-1 SIV BIV-1 SIV BIV-1 SIV BIV-1	VOICE A FFYELD LE FORDY REST HEFFYE ADD LYCEGOR FFOR RECT HEFFY SELVENCE HIRST IN 10++ +28 +++ +22 310 320 GSLAREEEVE REAMFTOMARTI LY GTRAREETY THING ROMETISA	STITUTE CHEST VICES STITUTE C	PRVERFE FERTCANAPA DRIPTION AFTA DRIPTION AFTA DRIPTION AFTA 2++ 350 360 WS IR TOPGRORAFYTICK TO EPVELMENUTES OPWER EPVELMENUTES OPWER	LIACINSTRUCTOR CTWESTOC LIACINSTRUCTORY OF THE CONTROL OF THE CONT	7300 400 7300 400 7300 400 7300 400 7300 400 7300 400
BIV-1 SIV BIV-1 SIV BIV-2 BIV-1 BIV-2 BIV-1 BIV-2 B-EFITOPE BIV-1 B-EFITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/BB BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 BIV-1 VARIANCE BIV-1 SIV	VORVEYA FFYELDIE FORDYT REKT THE THY SADOVCESTEN F7 H 10++ ++28 +++ ++2 310 320 VV V V V V V V V V V V V V V V V V V V	STITUTE CHEST VICES STITUTE C	PRVERFE FERTCANAPA DRIPTION AFTA DRIPTION AFTA DRIPTION AFTA 2++ 350 360 WS IR TOPGRORAFYTICK TO EPVELMENUTES OPWER EPVELMENUTES OPWER	LIACINSTITUTOR CTWEFFOR LIACIN	230 400 250 Representation of the second sec
BIV-1 SIV BIV-2 BIV-2 BIV-2 BIV-2 B-EPITOPE BIV-2 B-EPITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/DB BIV-2 T-CELL/DB BIV-2 T-CELL/AT SIV T-CELL/AT SIV-2 BIV-1 VARIANCE BIV-1 SIV BIV-2 BIV-1 SIV BIV-2 BIV-1_BIV-2	VOUR YA FFYELD IT FORDTY KRY METWYSADLVCE GORST KRYNTETY'S RUVVCE THESE R 10++ +28 +++ ++2 310 320 VV VV VV VV VV GSLAREEVVI RAMITTORATY IV GTRAERRYYIYMIG ROMPTIISL GTRAERRYYIYMIG ROMPTIISL	STELTHOUTSVICES SERVISHENISVICES GROCTHHENISVITES 4++ ++5 ++1 330 340 QUANGSVILINGTERHINTER REFERVILINGERGERGERGY	PRVERFE FETCAMORA DESTINATE FOR THE STATE OF	LICHINETYMPROP CTWYSTQC LICCIOTRYSORPHICS EVVAST LICCIOTRYSORPHICS EV	TROIRPOWERGILLS TROSERY THE TRUTH ++17 ++17 350 400 Y GVVV GVVV KIL REGE ONNE E IVERIFATTOTHERT TLAKEFATHOTHUTE
BIV-1 SIV BIV-2 BIV-2 BIV-2 B-EPITOPE BIV-2 B-EPITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/DB BIV-2 T-CELL/DB BIV-2 T-CELL/AT SIV T-CELL/AT SIV-1 T-CELL/AT SIV-1 BIV-1 SIV BIV-1 SIV BIV-1 SIV-1 BIV-1 SIV-2 BIV-1 BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-1	VOREYA FFYELDIF DIRBYT FREY HEFTYREDIF DIRBYT FREY HEFTYREDIV CETHWEF H 10++ ++28 ++++ ++28 ++++ ++28 ++++++++	STELTHORESTIGATE SERVINGENCESSAGE CONTROL SERVINGENCE CONTROL SERV	PRVSFPF FF FF CAPAGES DESTINATION FOR THE STATE OF THE S	LICOMENTY WORD CONVEYOR LICCOMENTY WORD CONVEYOR 370 380 370 380 370 380 370 380 370 380 370 380 370 380	TRIGITPOWETQLLES TRIGITS TWO THE TRIGITS ++17 350 400 V 9VVV 9VVV TOL REOF ONIK REVENERATIONS TO THANKER A VACTOURA
BIV-1 SIV BIV-1 BIV-2 BIV-1 BIV-2 BIV-1 BIV-2 B-EFITOPE BIV-1 B-EFITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB BIV-1 T-CELL/AT BIV-1 T-CELL/AT BIV-1 T-CELL/AT BIV-1 T-CELL/AT BIV-2 BIV-1 SIV BIV-1 SIV BIV-1 BIV-2 BIV-1 BIV-2 BIV-1 BIV-2	VORVEYA FFYELDIEF FORDY VORVEYA FFYELDIEF FORDY EKKY METWYSADAVCESHMEY KMCTHETCYSKDVVCESHMEY 10++ ++28 +++ ++2 310 320 V GSLAEREVVIRSAMFTUNANTIEV GYRAERETYIWNG ROMETIEL GYRAERETYIWNG ROMETIEL GYRAERETYIWNG ROMETIEL	STITUTEMENTAL STATE OF THE STAT	PRVSFPF FF FF CAPAGES DESTINATION FOR THE STATE OF THE S	LIACINSTITUTOR CITYUSTOC LIACINSTITUTORIUS ENVASTO LIACINSTITUTORIUS ENVASTO LIACINSTITUTORIUS ENVASTO LIACINSTITUTORIUS ENVASTO 370 380 370 380 370 380 370 380 370 380 370 380 370 380	TRIGITPOWETQLLES TRIGITS TWO THE TRIGITS ++17 350 400 V 9VVV 9VVV TOL REOF ONIK REVENERATIONS TO THANKER A VACTOURA
BIV-1 SIV BIV-1 BIV-2 BIV-1 BIV-2 BIV-1 BIV-2 BIV-1 BIV-2 B-EPITOPE BIV-1 T-CELL/OB SIV-1 T-CELL/OB SIV-1 T-CELL/AT BIV-1 T-CELL/AT BIV-1 T-CELL/AT BIV-2 BIV-1 SIV BIV-1 SIV BIV-1 BIV-2 BIV-1 BIV-2 BIV-1 BIV-2 BIV-1 BIV-2 BIV-1 BIV-2 B-EPITOPE BIV-1	VOREYA FFYELDIF DIRBYT FREY HEFTYREDIF DIRBYT FREY HEFTYREDIV CETHWEF H 10++ ++28 ++++ ++28 ++++ ++28 ++++++++	STELTHORESTIGATE SERVINGENCESSAGE CONTROL SERVINGENCE CONTROL SERV	PRVSFPF FF FF CAPAGES DESTINATION FOR THE STATE OF THE S	LICOMENTY WORD CONVEYOR LICCOMENTY WORD CONVEYOR 370 380 370 380 370 380 370 380 370 380 370 380 370 380	TRIGITPOWETQLLES TRIGITS TWO THE TRIGITS ++17 350 400 V 9VVV 9VVV TOL REOF ONIK REVENERATIONS TO THANKER A VACTOURA
BIV-1 SIV BIV-2 BIV-2 BIV-2 B-EPITOPE BIV-2 B-EPITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/DB BIV-2 T-CELL/DB BIV-2 T-CELL/AT SIV T-CELL/AT SIV-1 T-CELL/AT SIV-1 BIV-1 SIV BIV-1 SIV BIV-1 SIV-1 BIV-1 SIV-2 BIV-1 BIV-2 B-EPITOPE BIV-1 B-EPITOPE BIV-1	VOREYA FFYELDIF DIRBYT FREY HEFTYREDIF DIRBYT FREY HEFTYREDIV CETHWEF H 10++ ++28 ++++ ++28 ++++ ++28 ++++++++	STELTHORESTIGATE SERVINGENCESTIGATE CONTROL THE STELL 4++ ++5 4++ ++5 330 340 QUANGS VILING TREPRINETE SERVIL THE SERVICE THE SERVIL THE SERVICE	PRVSFPF FF FF CAPAGES DESTINATION FOR THE STATE OF THE S	LICOMENTY WORD CONVEYOR LICCOMENTY WORD CONVEYOR 370 380 370 380 370 380 370 380 370 380 370 380 370 380	TRIGITPOWETQLLES TRIGITS TWO THE TRIGITS ++17 350 400 V 9VVV 9VVV TOL REOF ONIK REVENERATIONS TO THANKER A VACTOURA
BIV-1 SIV BIV-1 SIV-1 BIV-2 BIV-2 BIV-1 BIV-2 BIV-1 B-EFITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/DB BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 T-CELL/RT BIV-1 SIV BIV-1 SIV BIV-1 SIV-1 BIV-2 B-EFITOPE BIV-2 T-CELL/RT BIV-1 T-CELL/RT BIV-1 BIV-1 BIV-1 BIV-1 BIV-2 B-EFITOPE BIV-2 T-CELL/DB BIV-1	VOREYA FFYELDIF DIRBYT FREY HEFTYREDIF DIRBYT FREY HEFTYREDIV CETHWEF H 10++ ++28 ++++ ++28 ++++ ++28 ++++++++	STELTHORESTIGATE SERVINGENCESTIGATE CONTROL THE STELL 4++ ++5 4++ ++5 330 340 QUANGS VILING TREPRINETE SERVIL THE SERVICE THE SERVIL THE SERVICE	PRVSFPF FF FF CAPAGES DESTINATION FOR THE STATE OF THE S	LICOMENTY WORD CONVEYOR LICCOMENTY WORD CONVEYOR 370 380 370 380 370 380 370 380 370 380 370 380 370 380	390 400 390 400 900 900 12. REOF CHIEFT 13. SALEP ANGTHUTE 14. SALEP ANGTHUTE 15. SALEP ANGTHUTE 16. SALEP ANGTHUTE 16. SALEP ANGTHUTE
BIV-1 SIV BIV-1 SIV BIV-2 BIV-2 BIV-2 BIV-2 BIV-1 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB SIV T-CELL/RT SIV T-CELL/RT SIV-1 T-CELL/RT SIV-1 T-CELL/RT SIV-1 SIV-1 BIV-1 SIV-1 SIV-2 BIV-1 BIV-2 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB BIV-1 T-CELL/OB BIV-1 T-CELL/OB BIV-1	VOREYA FFYELDIF DIRBYT FREY HEFTYREDIF DIRBYT FREY HEFTYREDIV CETHWEF H 10++ ++28 ++++ ++28 ++++ ++28 ++++++++	330 340 330 340 330 340 330 340 340	PRVSFPF FF FF CAPAGES DESTINATION FOR THE STATE OF THE S	JACOMETRICATE CTRESTON LACINITES AT A PART OF THE PROPERTY OF	390 400 390 400 900 900 12. REOF CHIEFT 13. SALEP ANGTHUTE 14. SALEP ANGTHUTE 15. SALEP ANGTHUTE 16. SALEP ANGTHUTE 16. SALEP ANGTHUTE
BIV-1 SIV BIV-2 B-EPITOPE BIV-2 B-EPITOPE BIV-2 B-EPITOPE BIV-2 T-CELL/DB BIV-1 T-CELL/DB BIV-1 T-CELL/RT BIV-2 T-CELL/RT BIV-2 BIV-1 SIV BIV-1 BIV-2 T-CELL/DB BIV-1 T-CELL/DB BIV-1 T-CELL/DB BIV-2 T-CELL/DB BIV-1 T-CELL/DB BIV-1	VOREYA FFYELDIF DIRBYT FREY HEFTYREDIF DIRBYT FREY HEFTYREDIV CETHWEF H 10++ ++28 ++++ ++28 ++++ ++28 ++++++++	330 340 330 340 330 340 330 340 340	PRVSFPF FF FF CAPAGES DESTINATION FOR THE STATE OF THE S	JACOMETRICATE CTRESTON LACINITES AT A PART OF THE PROPERTY OF	390 400 390 400 900 900 12. REOF CHIEFT 13. SALEP ANGTHUTE 14. SALEP ANGTHUTE 15. SALEP ANGTHUTE 16. SALEP ANGTHUTE 16. SALEP ANGTHUTE
BIV-1 SIV BIV-1 SIV BIV-2 BIV-2 BIV-2 BIV-2 BIV-1 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB SIV T-CELL/RT SIV T-CELL/RT SIV-1 T-CELL/RT SIV-1 T-CELL/RT SIV-1 SIV-1 BIV-1 SIV-1 SIV-2 BIV-1 BIV-2 B-EPITOPE BIV-2 T-CELL/OB BIV-1 T-CELL/OB BIV-1 T-CELL/OB BIV-1 T-CELL/OB BIV-1	VOREYA FFYELDIET FORDYT REWY HET YES HELD VEGEN FORM REWY HET YES HAVVUET HET H 10++ +28 +++ +28 +++ +	330 340 330 340 330 340 340 340 340	PRVEREP PERCEAPORA DRETHOAL RETCAPPORA	JACOMETRICATE CTRESTON LACINITES AT A PART OF THE PROPERTY OF	390 400 390 400 900 900 12. REOF CHIEFT 13. SALEP ANGTHUTE 14. SALEP ANGTHUTE 15. SALEP ANGTHUTE 16. SALEP ANGTHUTE 16. SALEP ANGTHUTE

Fig.3 (contd).

HIV-1 VARIANCE HIV-1 SIV HIV-1 SIV HIV-2 HIV-1_HIV-2	TIIFNOS SOCOPEIVTESS KINLTAP EGGOPEVTENNI HISPAAFGKGSDPEVANNI	Ticggeffych Stolfh St Tichgeflychomwplh Tichgeflychwtwflh	rnedreltighe Wedreltighe Wienk	TEGEDTITLPCRI ERENRWYV PCHI TERWYA PCEI	MANAMATER TO THE TOTAL T	MAXTE 1805 MAXTE 1805 MAXTE 1805	Pachelale I Ivni Pichelale Itali Inchen Ilaifi
B-EPITOPE BIV-1 B-EPITOPE SIV B-EPITOPE BIV-2	++ 23++ ++11++	++7+++	++ 22 ++ + ++7	++27++ +2+++ +++		**11+* **10**	
T-CELL/DB BIV-1 T-CELL/DB BIV-2						-	
T-CELL/RT RIV-1 T-CELL/RT BIV-2					-		
HIV-1 VARIANCE HIV-1_SIV HIV-1 SIV HIV-2 HIV-1_KIV-2	510 520 VVVVVV V V V 5 COMMINGSE IFFFGGOMB, HWTGGMGTSITMERTVAEL DRGMMGTHITT FAEVAEL	ANTERGOAKTAE 145 I OMMUSETAKAKAKAK 145 I	GVAPTKAKRRYVQ GLAPTKVKRYTTG GFAPTKEKRYSSA	REXIDANG I CAL OTS INVESTIGAT OTS INVESTIGATION	PLOF LGANGE THE PLOF LATROSAME	WARNITANG MWRITANG MWRITANG	INATITIVE I AGGOGG INATITIVE I AGGOGG INGETRE I AGGOMI
B-EPITOPE BIV-1 B-EPITOPE BIV-2 B-EPITOPE BIV-2	\$** **	12++	******	#++ ++27++			*****
T-CELL/DB BIV-1 T-CELL/DB BIV-2							
T-CELL/RT RIV-1 T-CELL/RT SIV T-CELL/RT HIV-2							
HIV-1 VARIANCE HIV-1 SIV HIV-1 SIV HIV-2 HIV-1_HIV-2	510 520 VV LIERATERAGEILIGETVMGTMM LLOVVKROGELLELTVMGTMM LLOVVKROGELERLTVMGTMM	LOTRVSA I EKYLKDOAOLI	ingeschlietta Angeafroventt	VPWHASWSHKSLE VPW PHASLT	Q I MKHPYTMPE HD F PDMRH B TMQE HE R	e innytslih KVDFLeanit	slieesqnqqe Alleeaqiqqe
B-EPITOPE HIV-1 B-EPITOPE HIV-2	**17** **16**	+++9++ ++17++			++13 ++30		++20++ +
T-CELL/DB HIV-1 T-CELL/DB HIV-2							
T-CELL/RT HIV-1 T-CELL/RT SIV T-CELL/RT HIV-2							
RIV-1 VARIANCE RIV-1 SIV RIV-1 SIV RIV-2 RIV-1_BIV-2	710 720 V V V V KINGGEL, ELD INNA LARRIYEN KOMY EL GELM SMOVT GENETO KOMY EL GELM SMOV I PERMITON	TEMPERATION AND THE PROPERTY OF THE PROPERTY O	V V V V VGLRIVFAVLSIV LRIVIVIVIVOML	HRVRQGYSPL ARLRQGYRPVF65	SF QTHLETP SF QTHIETQQQ PPGYIQQIHIHKQ	RGPDRPEGIE PALPTKEGKR	EDGGSNGGDRYN
B-EPITOPE BIV-1 B-EPITOPE SIV B-EPITOPE BIV-2	++2++			++1 8+ +		+7++ + ++3++ +++ ++2+	•
T-CELL/DB RIV-1 T-CELL/DB RIV-2							
T-CELL/RT BIV-1 T-CELL/RT SIV T-CELL/RT BIV-2							

Fig.3 (contd).

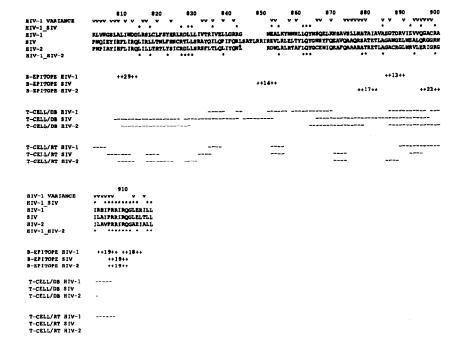


Fig.3. The alignment of env HIV-1, HIV-2 and SIV_{MAC}.

Table 1 The predicted B epitopes with %SI = 100 in HIV-1

HIV-1		ні	HIV-2		SIV _{MAC}	
Epitope numbers	Epitope sequence	Epitope numbers	HIV-1 %1-2	Epitope numbers	HIV-1 %1-S	
Pol						
128-139	ALLDTGADDTVL	128-139	67	128-139	67	
265-276	FAIKKKDSTKWR	265-276	83	265-276	75	
303-314	GLKKKKSVTVLD	301-313	67	301-312	67	
421-432	PDKKHQKEPPFL	416-427	58	419-430	58	
	_	420-431	58			
507-518	LAENREILKEPV	503-514	50	503–51 à	33	
746~757	ISGNEQVDKLVS		75		67	
968-979	IATDIQTKELQK		33		25	
Gag						
230-241	QMREPRGSDIAG	229-240	92	229-240	83	
288-299	LDIRQGPKEPFR	288-299	83	288-299	75	
Env						
294-305	STQLLINGSLAE		50	299-310	50	

The epitope numbers denote the positions of the predicted epitopes in the figures. The HIV-1 epitope sequence is given. If an epitope is found in corresponding regions in HIV-2 or SIV_{MAC} then their position numbers are given. %HIV-1 is defined in procedures

Table 2 The predicted T epitopes with %SI = 100% in HIV-1

HIV-1		HIV-2		SIV _{MAC}	
Epitope numbers	Epitope sequence	Epitope numbers	HIV-1 %1-2	Epitope numbers	HIV-1 %1-S
Pol					
123-134	QLKEALLDTGAD	123-134	75	123-143	75
150-161	KMIGGIGGFIKV	150-161	67	147-160	67
293-304	EVQLGIPHPAGL	293-304	92	293-304	100
314-325	DVGDAYFSVPLD	314-325	83	314-325	83
502-513	EAELELAENREI		67		58
615-626	IPEWEFVNTPPL	615-626	75	615626	75
704-715	QYALGIIQAQPD		50	706-717	50
796-807	VAKEIVASCDKC	797-808	58	799-810	75
808-819	QLKGEAMHGQVD	808-819	67	808-819	67
840851	AVHVASGYIEAE	840-851	92	840851	92
901-912	QEFGIPYNPQSQ	901-912	83	901-912	75
Gag					
347~359	LEEMMTACQGVG	346-357	92	346-357	92
Env					
42- 53	VPVWKEATTTLF		67		58
111-124	LWDQSLKPCVKLTP	103-116	71	103-116	64

3. RESULTS

The HIV-2 and SIV_{MAC} sequences are far more homologous to one another than either are to HIV-1. We pay particular attention to regions where B and T epitopes overlap because it has been shown that protective epitopes in influenza virus contain overlapping B and T epitopes [31]. All residue numbers refer to the alignment numbers used in the figures.

3.1. Analysis of the gag gene

Fig.1 illustrates the alignment and prediction of the gag gene product. 50% of residues are conserved between HIV-1, SIV_{MAC}, and HIV-2. Ten B epitopes were located that had 2 or less HIV-1 inter-strain variable positions. Two B epitopes were found in which there was total sequence identity between the thirteen HIV-1 isolates (100% SI). Both regions have corresponding HIV-2 and SIV_{MAC} predicted B epitopes with reasonable sequence conservation to HIV-1. Part of the 288-299 B epitope region (LDIRQGPKEPFR) has also previously been identified as a promising candidate for vaccine studies [14,15]. There are four B epitopes with one inter-strain variable position

(15-26, 158-169, 290-301, 480-492). All but 480-492 have overlapping HIV-2 and SIV_{MAC} predicted B epitopes. Four B epitopes had two inter-strain variable positions (20-31, 95-106, 99-113, 308-419). Only 20-31 and 308-319 had a HIV-2 B epitope associated with them while 20-31, 95-106 and 308-319 a SIV_{MAC} B epitope. Only the region 20-31 has a HIV-2 and SIV_{MAC} sequence with high sequence identity to HIV-1. Thus, although gag contains 2 B epitopes that have no inter-strain variability, they are not identical to either HIV-2 or SIV_{MAC}. Therefore if one residue difference disrupts the antigenicity of the epitopes, it is not possible to produce a HIV-1 B epitope that will be effective against HIV-2 infection or that could be tested in SIV_{MAC}. However, if peptides shorter than 12 residues are considered then there is an epitope (15-26 in HIV-1) that has 8 residues for which there is 100% sequence identity with SIV_{MAC}.

Only one T epitope (347–358) has no inter-strain variation. In HIV-2 and SIV_{MAC} there are corresponding T epitopes in this region and the sequence homology to HIV-1 is high. Eight T epitopes are found that have one inter-strain variable position (73–84, 156–167, 170–181,

198-209, 215-226, 273-284, 335-346, 365-376) and further 8 T epitopes are located that have 2 inter-strain variable positions (96-107, 187-198, 207-218, 260-271, 291-307, 311-322, 415-426, 479-491). Of the 17 T epitopes with less than three sequence variations between HIV-1 strains, 13 are in the p24 protein.

The B epitope in position 288-299 has a T epitope region overlapping at positions 291 to 307 which has two inter-strain variable positions. The only T epitope with no inter-strain variation (347-358) has no overlapping predicted B epitope.

3.3. Analysis of the pol gene

The pol sequence is as well conserved as the gag, 49% identity between the three aligned sequences and there are a considerable number of B and T epitopes found with no inter-strain variation. Seven B epitopes are found with 100% SI. But none have HIV-1 %1-2 and HIV-1 %1-S of 100%. However, one HIV-1 B epitope (1018–1029) has 8 residues that are fully conserved to the predicted B epitope in SIV_{MAC}. Ten B epitopes were located with one inter-strain variable position (220–231, 230–241, 281–292, 322–333, 499–510, 522–533, 914–925, 925–936, 1012–1023, 1018–1029). Six of these had overlapping B epitopes in HIV-2 and only 4 in SIV_{MAC}.

A total of 11 T epitopes are identified in the pol product that has no inter-strain variation. One of these, epitope 293–304 has 100% sequence identity to SIV_{MAC}, which also has a T epitope predicted. Between HIV-1 and HIV-2 there is only one residue difference and also a T epitope is predicted in the same position. This T epitope prediction overlaps with a B epitope prediction (303–314) that also has no inter-strain variable position. Further 10 T epitopes are found that have only one inter-strain variable position (273–284, 284–295, 478–489, 512–524, 683–694, 749–760, 780–791, 910–921, 1019–1030, 1026–1039).

3.4. Analysis of the env gene

There is a low homology (28% identity) between the three env sequences. One B epitope is found that has no inter-strain variation. However, this B epitope has no corresponding prediction in HIV-2 and although there is an overlapping B epitope prediction in SIV_{MAC} the sequence homology is very low. No B epitopes with only one inter-strain

variable position were found. There are five B epitopes with two inter-strain variable positions (100-111, 516-527, 539-550, 546-558, 628-639).

Only two T epitopes that are conserved between the HIV-1 isolates are found in the env product. One (42–53) has no corresponding T epitope prediction in either HIV-2 or SIV_{MAC}. The other (111–124) has overlapping T epitope predictions in HIV-2 and SIV_{MAC} but the sequence homology between these and HIV-1 is low. Three T epitopes with one inter-strain variable position (37–48, 521–532, 586–600) and nine T epitopes with two inter-strain variable positions (283–297, 527–538, 535–546, 544–556, 599–610, 606–617, 627–638, 643–654, 764–757) were located. None showed high sequence homology to HIV-2 or SIV_{MAC}.

4. DISCUSSION AND CONCLUSIONS

Antigen prediction techniques were used to analyse the potential effect of HIV epitope sequence variability on AIDS vaccine design. This study, being based on predictions, is intended as a guide for experimental vaccine design. For the Bepitopes, the predictions with the higher rank order of hydrophilicity (figs 1-3) will be more accurate [30]. The highest peak is nearly always antigenic [30]. For example, the env B epitope 787-800 is antigenic [32] and contains two predicted epitopes of rank 1 and 3. The accuracy of T epitope predictions has been less well studied. However, two predicted T epitopes [15] have subsequently been shown [33] to be part of experimentally verified T-cell sites (peptides 105-117, 465-480). Env, gag and pol each code for several protein products that result from proteolytic cleavage of the initial polyprotein gene product. If a predicted epitope includes a cleavage site then this epitope might not be antigenic. However, in cells infected with a related retrovirus (equine infectious anemia virus), uncleaved and partially cleaved polyproteins have been shown to be antigenic [34]. Thus peptides that span protease cleavage sites might be antigenic. Accordingly we have included them in this study but the location of these cleavage sites should be considered in the interpretation of experimental results.

We addressed three specific problems. Firstly, we investigated the likelihood of finding a single HIV-1 vaccine which might protect an individual

against a range of HIV-1 isolates. In other microorganisms neutralising antibody against surface antigens, such as envelope, confers protection and accordingly the most promising candidates for de novo protection against HIV infection are the env proteins. In env there was only one predicted B epitope (294-305) which was found to be conserved between the 13 isolates considered. The hydrophilicity is ranked 17th, which indicates a low accuracy of prediction. In contrast, the B epitope 787-800, which has the highest rank, is poorly conserved between HIV-1 strains and only 5 out of its 14 residues were found in all the isolates which were studied. This suggests that env from one HIV-1 strain is unlikely to induce protective antibody-mediated immunity against a range of wild HIV-1 strains unless the 294-305 epitope is effective as a vaccine. In gag two B epitopes are conserved between all strains but in pol seven B epitopes are conserved. In viruses, a major component of the T-cell response is directed at internal proteins. In HIV-1 two T epitopes are fully conserved in all strains for env, one for gag and 11 for pol. If a polypeptide from a single strain is to be effective against all HIV-1 isolates in an outbred population such as man, then this analysis indicated T epitopes from pol as the most promising candidate. The analysis highlights the need in animal studies to challenge with the same strain of virus which was used to immunize. A problem with this approach might occur if epitope variation were to occur early after infection.

Secondly, we assessed the feasibility of using HIV-1 antigenic peptides in a monkey AIDS model, by searching for sections of the chain which are conserved between HIV-1_{LAV} and SIV_{MAC}. In env and gag, no predicted sites were 100% identical between HIV-1 and SIV_{MAC}. Pol had only one conserved site. Thus it is unlikely that HIV-1 peptides will confer protection against SIV_{MAC} challenge in the monkey if complete epitope sequence conservation is required.

Thirdly, the prospect for a single HIV-1 vaccine which would protect an individual against HIV-2 was addressed. In env, gag and pol no predicted B-cell and T-cell sites were 100% identical between HIV-1 and HIV-2, and so a single effective HIV-1/HIV-2 vaccine is unlikely to emerge.

Apart from T epitope based predictions in pol, a cocktail of peptides from different viral strains

may be required as an effective vaccine against HIV-1. It is unlikely that this cocktail will be effective against HIV-2 and so additional peptides based on the HIV-2 virus will need to be introduced. The HIV-1 vaccines would be difficult to test in rhesus macaque monkeys because of considerable sequence variation between HIV-1 and SIV_{MAC}.

Acknowledgements: The MRC, AFRC and the Imperial Cancer Research Fund provided financial support. We thank Dr Geoff Barton and Dr D. Sia for advice.

REFERENCES

- [1] Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. and Alizon, A. (1985) Cell 40, 9-17.
- [2] Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C. and Wong-Staal, F. (1985) Nature 313, 277-284.
- [3] Sanchez-Pescador, R., Power, M.D., Barr, P.J., Steimer, K.S., Stempien, M.M., Brown-Shimer, S.L., Gee, W.W., Renard, A., Randolph, A., Levy, J.A., Dina, D. and Luciw, P.A. (1985) Science 227, 484-492.
- [4] Meusing, M.A., Smith, D.H., Cabradilla, C.D., Benton, C.V., Lasky, L.A. and Capon, D.J. (1985) Nature 313, 450-458.
- [5] Clavel, F., Guyader, M., Guetard, D., Salle, M., Montagnier, L. and Alison, M. (1986) Nature 324, 691-695.
- [6] Daniel, M.D., Letvin, N.L., King, N.W., Kannagi, M., Sehgal, P.K., Hunt, R.D., Kanki, P.J., Essex, M. and Desrosiers, R.C. (1985) Science 228, 1201-1204.
- [7] Chanh, T.C., Dreesman, G.R., Kanda, P., Linette, G.P., Sparrow, J.T., Ho, D.D. and Kennedy, R.C. (1986) EMBO J. 5, 3065-3071.
- [8] Wang, J.J.G., Steel, S., Wisniewolski, R. and Wang, C.Y. (1986) Proc. Natl. Acad. Sci. USA 83, 6159-6163.
- [9] Crowl, R., Ganguly, K., Gordon, M., Conroy, R., Schaber, M., Kramer, R., Shaw, G., Wong-Staal, F. and Reddy, E.P. (1985) Cell 41, 979-986.
- [10] Pauletti, D., Simmonds, R., Dreesman, G.R. and Kennedy, R.C. (1985) Anal. Biochem. 151, 540-546.
- [11] Robson, B., Fishleigh, R.V. and Morrison, C.A. (1987) Nature 325, 395.
- [12] Hunsmann, G., Schneider, J. and Schulz, A. (1981) Virology 113, 602-612.
- [13] Hu, S.-L., Fultz, P.N., McClure, H.M., Eichberg, J.W., Thomas, E.K., Zarling, J., Singhai, M.C., Kosowski, S.G., Swenson, R.B., Anderson, D.C. and Todaro, G. (1987) Nature 328, 721-723.
- [14] Coates, A.R.M., Cookson, J., Barton, G.J., Zvelebil, M.J. and Sternberg, M.J.E. (1987) Nature 326, 549-550.

- [15] Sternberg, M.J.E., Barton, G.J., Zvelebil, M.J., Cookson, J. and Coates, A.R.M. (1987) FEBS Lett. 218, 231-237.
- [16] Milich, D.R., McLachlan, A., Thornton, G.B. and Hughes, J.L. (1987) Nature 329, 547-549.
- [17] Murray, K., Bruce, S.A., Hinnen, A., Wingfield, P., Van Erd, P.M.C.A., De Reus, A. and Schellekens, H. (1984) EMBO J. 3, 645-650.
- [18] Tabor, E. and Geret, R.J. (1984) Lancet 1, 172.
- [19] Zagury, D., Leonard, R., Fouchard, M., Reveil, B., Bernard, J., Ittelé, D., Cattan, A., Zirimwabagabo, L., Kalumbu, M., Justin, W., Salaun, J.J. and Goussard, B. (1987) Nature 326, 249-250.
- [20] Zarling, J.M., Morton, W., Moran, P.A., McClure, J., Kosowski, S.G. and Hu, S.-L. (1986) Nature 323, 344-346.
- [21] Walker, B.D., Chakrabarti, S., Moss, B., Paradis, T.J., Flynn, T., Durno, A.G., Blumberg, R.S., Kaplan, J.C., Hirsch, M.S. and Schooley, R.T. (1987) Nature 328, 345-348.
- [22] Weber, J.N., Clapham, P.R., Weiss, R.A., Parker, D., Roberts, C., Duncan, J., Weller, I., Carne, C., Tedder, R.S., Pinching, A.J. and Cheingsong-Popov, R. (1987) Lancet 1, 119-122.
- [23] DeLisi, C. and Berzofsky, J.A. (1985) Proc. Natl. Acad. Sci. USA 82, 7048-7052.
- [24] Mills, K.H., Skehel, J.J. and Thomas, D.B. (1986) J. Exp. Med. 163, 1477-1490.

- [25] Myers, G., Rabson, A.B., Josephs, S.F., Smith, T.F. and Wong-Staal, F. (eds) (1987) in: Human Retroviruses and AIDS. A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences. Theor. Biol. Biophys. Group T-10, New Mexico.
- [26] Barton, G.J. and Sternberg, M.J.E. (1987) J. Mol. Biol. 198, 327-337.
- [27] Hackett, C.J., Dietzchold, B., Gerhard, W., Ghrist, B., Knorr, R., Gillessen, H. and Melchers, F. (1983) J. Exp. Med. 158, 294-302.
- [28] Allen, P.M., Strydom, D.J. and Unanue, E.R. (1984) Proc. Natl. Acad. Sci. USA 81, 2489-2493.
- [29] Rothbard, J.B. and Taylor, W.R. (1988) EMBO J. 7, 93-100.
- [30] Hopp, T.P. and Woods, K.R. (1981) Proc. Natl. Acad. Sci. USA 78, 3824-3828.
- [31] Thomas, D.B., Skehel, J.J., Mills, K.H. and Graham, C.M. (1987) Eur. J. Immunol. 17, 133-136.
- [32] Chanh, T.C., Dreesman, G.R., Kaula, P., Linette, G.P., Sparrow, J.T., Ho, D.D. and Kennedy, R.C. (1986) EMBO J. 5, 3605-3671.
- [33] Cease, K.B., Margalit, H., Cornette, J.L., Putney, S.D., Robey, W.G., Ouyang, C., Streicher, H.Z., Fischinger, P.J., Gallo, R.C., DeLisi, C. and Berzofsky, J.A. (1987) Proc. Natl. Acad. Sci. USA 84, 4249-4253.
- [34] Henderson, L.E., Sowder, R.C., Smythers, G.W. and Oroszlan, S. (1987) J. Virol. 61, 1116-1124.