doi: 10.1038/nature07746

% Variant in Population versus Prevalence of HLA Allele in Population

		Linear Regression Model			Logistic Regression Model	
Polymorphism	Epitope	r	p	(Pooled UK, r and p)	p	(Pooled UK, p)
I135X	B*51-TAFTIPSI	0.94	0.0002	(0.95 and 0.0004)	0.0001	(<0.0001)
S357X	B*07-GPSHKARVL	0.67	0.051	(0.65 and 0.082)	0.004	(0.005)
D260X	B*35-PPIPVGDIY	0.81	0.008	(0.84 and 0.009)	0.016	(0.003)
D312X	B*44-AEQATQDVKNW	0.30	0.436	(0.26 and 0.539)	0.080	(0.103)
I31V	B*51-LPPIVAKEI	0.84	0.017	(0.84 and 0.017)	0.001	(0.001)
R264X	B*27-KRWIILGLNK	0.85	0.003	(0.92 and 0.001)	0.003	(0.004)
L268X	B*27-KRWIILGLNK	0.93	0.0002	(0.94 and 0.0005)	0.001	(0.003)
A146X	B*57-AISPRTLNAW	0.38	0.317	(0.36 and 0.378)	0.048	(0.068)
I147X	B*57-ISPRTLNAW	0.87	0.002	(0.91 and 0.002)	0.007	(0.007)
A163X	B*5703-KAFSPEVIPMF	0.80	0.031	(0.80 and 0.031)	0.110	(0.110)
S165X	B*5703-KAFSPEVIPMF	0.46	0.303	(0.46 and 0.303)	0.364	(0.364)
I168V	B*5703-KAFSPEVIPMF	0.84	0.019	(0.84 and 0.019)	0.085	(0.085)
T242X	B*57/5801-TSTLQEQIAW	0.99	<0.0001	(0.83 and 0.010)	<0.0001	(0.0003)
I247X	B*5703-TSTLQEQIAW	0.91	0.004	(0.91 and 0.004)	0.041	(0.041)
All polymorphisms		0.69	<0.0001	(0.69 and p<0.0001)	<0.0001	(<0.0001)

Supplementary Table 1. Analyses of all polymorphisms using linear regression and logistic regression models, including data from acute and chronic UK cohorts analysed separately and together, showing correlation between HLA prevalence and variant frequency in the respective study population.

doi: 10.1038/nature07746

% Variant in HLA-Mismatched Population versus Prevalence of HLA Allele in Population

		Linear Regression Model		
Polymorphism	**Epitope**	**r**	**p**	**(Pooled UK, r and p)**
I135X	B*51-TAFTIPSI	0.91	0.0006	(0.93 and 0.001)
S357X	B*07-GPSHKARVL	0.49	0.181	(0.45 and 0.258)
D260X	B*35-PPIPVGDIY	0.81	0.008	(0.84 and 0.009)
D312X	B*44-AEQATQDVKNW	0.19	0.623	(0.12 and 0.783)
I31V	B*51-LPPIVAKEI	0.81	0.027	(0.81 and 0.027)
R264X	B*27-KRWIILGLNK	0.85	0.004	(0.90 and 0.003)
L268X	B*27-KRWIILGLNK	0.87	0.002	(0.91 and 0.002)
A146X	B*57-AISPRTLNAW	0.22	0.562	(0.21 and 0.616)
I147X	B*57-ISPRTLNAW	0.73	0.025	(0.82 and 0.012)
A163X	B*5703-KAFSPEVIPMF	0.66	0.108	(0.66 and 0.108)
S165X	B*5703-KAFSPEVIPMF	0.35	0.439	(0.35 and 0.439)
I168V	B*5703-KAFSPEVIPMF	0.87	0.011	(0.87 and 0.011)
T242X	B*57/5801-TSTLQEQIAW	0.60	0.090	(0.54 and 0.164)
I247X	B*5703-TSTLQEQIAW	0.87	0.012	(0.87 and 0.012)
All polymorphisms		0.60	<0.0001	(0.61 and p<0.0001)

Supplementary Table 2. Analyses of all polymorphisms using linear regression model, including data from acute and chronic UK cohorts analysed separately and together, showing correlation between HLA prevalence and variant frequency in the HLA-mismatched subset of the respective study population.

doi: 10.1038/nature07746

Phylogenetically corrected and uncorrected HLA-HIV amino acid polymorphism associations (Vancouver, Perth, Kumamoto, Durban, Gaborone cohorts)

		Uncorrected Fisher's	Phylogenetically corrected
Polymorphism	Epitope	p	p
I135X	B*51-TAFTIPSI	1.5×10^{-52}	2.9×10^{-45}
S357X	B*07-GPSHKARVL	6.7×10^{-24}	8.7×10^{-19}
D260X	B*35-PPIPVGDIY	3.8×10^{-17}	1.8×10^{-9}
D312X	B*44-AEQATQDVKNW	4.6×10^{-16}	3.1×10^{-12}
I31V	B*51-LPPIVAKEI	1.5×10^{-7}	9.2×10^{-3}
R264X	B*27-KRWIILGLNK	4.0×10^{-10}	2.8×10^{-9}
L268X	B*27-KRWIILGLNK	7.1×10^{-10}	3.4×10^{-10}
A146X	B*57-AISPRTLNAW	8.2×10^{-20}	1.7×10^{-9}
I147X	B*57-ISPRTLNAW	3.6×10^{-23}	7.5×10^{-6}
A163X	B*5703-KAFSPEVIPMF	1.3×10^{-21}	1.8×10^{-18}
S165X	B*5703-KAFSPEVIPMF	3.7×10^{-12}	5.7×10^{-6}
I168V	B*5703-KAFSPEVIPMF	ns	ns
T242X	B*57/5801-TSTLQEQIAW	3.9×10^{-64}	2.1×10^{-45}
I247X	B*5703-TSTLQEQIAW	5.7×10^{-7}	4.9×10^{-6}

Supplementary Table 3. Uncorrected Fisher exact test p values and p values following phylogenetic correlation of all 14 studied polymorphisms using all available data from 5 cohorts (n=1,373), comprising those study subjects in Vancouver, Perth, Kumamoto, Durban and Gaborone. The uncorrected p value was <0.05 for the V168X variant for this subset of the dataset.

doi: 10.1038/nature07746

Genbank numbers:

Durban sequences:

Gag:	FJ198407-FJ199088
Pol:	FJ199532-FJ199992

Gaborone Sequences:

Gag:	FJ497801-FJ497950
Pol:	FJ498244-FJ498543

London Sequences:

Gag:	FJ473452-FJ474070

Lusaka sequences:

Gag:	FJ606114-FJ606446

Kumamoto sequences:

Gag and Pol:	AB475154-AB476170

Oxford sequences:

Gag	FJ645274- FJ645344
	FJ645350- FJ645360
	FJ645409-FJ645410
Pol:	FJ645411- FJ645478
	FJ645483- FJ645488
	FJ645534- FJ645538

Barbados sequences:

Gag:	FJ645345- FJ645349
	FJ645361-FJ645408
Pol:	FJ645479-FJ645482
	FJ645489-FJ645533

Bayesian network analyses of resistance pathways against efavirenz and nevirapine

Koen Deforche[a], Ricardo J. Camacho[b], Zehave Grossman[c], Marcelo A. Soares[d], Kristel Van Laethem[a], David A. Katzenstein[e], P. Richard Harrigan[f], Rami Kantor[g], Robert Shafer[e], Anne-Mieke Vandamme[a] on behalf of the non-B workgroup

Objective: To clarify the role of novel mutations selected by treatment with efavirenz or nevirapine, and investigate the influence of HIV-1 subtype on nonnucleoside reverse transcriptase inhibitor (nNRTI) resistance pathways.

Design: By finding direct dependencies between treatment-selected mutations, the involvement of these mutations as minor or major resistance mutations against efavirenz, nevirapine, or coadministrated nucleoside analogue reverse transcriptase inhibitors (NRTIs) is hypothesized. In addition, direct dependencies were investigated between treatment-selected mutations and polymorphisms, some of which are linked with subtype, and between NRTI and nNRTI resistance pathways.

Methods: Sequences from a large collaborative database of various subtypes were jointly analyzed to detect mutations selected by treatment. Using Bayesian network learning, direct dependencies were investigated between treatment-selected mutations, NRTI and nNRTI treatment history, and known NRTI resistance mutations.

Results: Several novel minor resistance mutations were found: 28K and 196R (for resistance against efavirenz), 101H and 138Q (nevirapine), and 31L (lamivudine). Robust interactions between NRTI mutations (65R, 74V, 75I/M, and 184V) and nNRTI resistance mutations (100I, 181C, 190E and 230L) may affect resistance development to particular treatment combinations. For example, an interaction between 65R and 181C predicts that the nevirapine and tenofovir and lamivudine/emtricitabine combination should be more prone to failure than efavirenz and tenofovir and lamivudine/emtricitabine.

Conclusion: Bayesian networks were helpful in untangling the selection of mutations by NRTI versus nNRTI treatment, and in discovering interactions between resistance mutations within and between these two classes of inhibitors.

AIDS 2008, **22**:2107–2115

Keywords: Bayesian network, efavirenz, HIV-1 antiviral resistance, nevirapine

[a]Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium, [b]Molecular Biology Laboratory, Centro Hospitalar de Lisboa Ocidental, Lisbon, Portugal, [c]Chaim Sheba Medical Center, Ministry of Health, Tel-Aviv, Israel, [d]Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, [e]Division of Infectious Diseases, Stanford University, Stanford, California, USA, [f]Research Labs, BC Centre for Excellence in HIV/AIDS, Vancouver, Canada and [g]Division of Infectious Diseases, Brown University, Providence, Rhode Island, USA.

Correspondence to Anne-Mieke Vandamme, Rega Institute for Medical Research, Minderbroedersstraat 10, 3000 Leuven, Belgium.

E-mail: annemie.vandamme@uz.kuleuven.be
Received: 22 January 2008; revised: 9 June 2008; accepted: 30 June 2008.

DOI:10.1097/QAD.0b013e32830fe940

Introduction

Genotypic interpretation systems predict the therapy response for various drugs [1,2] based on the presence of mutations at positions associated with drug resistance. For many mutations observed after treatment failure in clinical isolates, their role is not sufficiently known and the impact of the large natural variation of HIV-1 is still debated. Different prevalences of known resistance-associated mutations and new mutations are seen in different subtypes [3–9], and with a few exceptions, these differences in prevalence could not be explained by different evolutionary distance because of different codon usage [10]. In previous work [11], we used Bayesian network learning to explain subtype differences for resistance development to nelfinavir from interactions between polymorphisms and resistance mutations. Such studies are useful to improve expert genotypic interpretation systems, in particular to make them more applicable to non-B subtypes by including resistance mutations that occur rarely in subtype B. Some rules for the Rega interpretation system [2] have been derived from Bayesian network learning (for example inclusion of 89I/V in RegaV7.1). By learning the relationship between resistance mutations, in-vitro experiments may be designed to study treatment-selected mutations in a relevant context of other resistance mutations or polymorphisms.

In recent years, treatment is finally reaching many HIV-infected individuals in low and middle-income countries [12]. At the end of 2005, around 1.3 million people in these countries were receiving antiretroviral treatment. Most countries have standardized on first-line and second-line treatments that were recommended by the WHO, and as a consequence the nonnucleoside reverse transcriptase inhibitors (nNRTIs) nevirapine (NVP), and to a lesser extent efavirenz (EFV) are predominantly used in first-line treatment regimens. Unlike the Western epidemic, which is dominated by HIV-1 subtype B, the large majority of the worldwide epidemic is caused by HIV-1 of diverse subtypes (and mainly subtype C). Therefore, a better understanding of mechanisms and mutations involved in antiviral resistance to these drugs is needed for these non-B subtypes.

In this work, we analyzed clinical data to study resistance pathways to nNRTIs. We investigated a possible impact of subtype and interactions with mutations conferring resistance to nucleoside analogue reverse transcriptase (NRTIs). Resistance development to NRTIs and nNRTIs is confounded by the use of combination therapy, which typically combines use of several inhibitors from the two different classes with resistance mutations in the same region. As we will show, Bayesian networks offer a direct benefit over other statistical approaches to determine whether treatment-associated mutations are related to NRTI or nNRTI resistance pathways, and to investigate possible interactions between NRTI and nNRTI resistance mutations.

Methods

Protease and partial reverse transcriptase (up to position 250) sequences of three HIV-1 populations were pooled from five clinical databases: Portugal, Belgium, Israel, Brazil and an international database containing sequences from subtypes other than subtype B (the non-B workgroup). The populations were defined by respectively 3837 sequences from nNRTI-naive patients, sequences from patients treated with only experience to EFV (462 sequences) or NVP (533 sequences) as nNRTI. For 10% of RT sequences, which covered RT starting around position 35, the missing sequence fragment was treated as missing values. At most, one treated sequence and one naive sequence per patient were included and identical sequences were removed. For each sequence, the detected polymorphic amino acids and nNRTI treatment-associated mutations, as well as known key NRTI mutations [13], with prevalence over 0.5%, and NRTI treatment experience were included in the data set for Bayesian network learning.

The analyses followed the method previously described [11], briefly summarized here. Subtyping was done on the nucleotide sequences using the Rega HIV-1 subtyping tool v2.0 [14]. Wild-type polymorphisms were detected based on a prevalence greater than 15% in untreated patients and treatment-associated mutations by testing for independence from treatment using a Cochran–Mantel–Haenszel test [correcting for multiple testing using Benjamini and Hochberg with a false discovery rate (FDR) of 0.05]. After stratifying the data sets for same subtype distribution in treated versus untreated patients, consensus Bayesian networks were learned from the data based on a nonparametric bootstrap [15]. A Bayesian network is a probabilistic model that describes statistical independencies between multiple variables. Using Bayesian network learning, a Bayesian network is searched that explains a maximum of the observed correlations in the data using a minimum number of direct influences.

The network structure was used to determine the role of new treatment-associated mutations. Mutations were considered novel for a drug, when not included in the International AIDS Society (IAS), USA list of resistance mutations of 2007 [13] or in one of the publicly available expert interpretation systems (Rega 6.4 HIVDB 4.2.9, or ANRS 2006.07) for that drug. Novel mutations whose prevalence was dependent on key nNRTI resistance mutations only were considered to be novel nNRTI resistance mutations, and similarly for NRTI resistance mutations. In addition, mutations whose prevalence was

determined by presence of other mutations rather than by treatment were considered minor resistance mutations, and mutations that were directly dependent on treatment and for which Conditional Probability Distributions showed a significant prevalence in the absence of other major mutations, were considered major mutations.

Results

Novel efavirenz, nevirapine, and nucleoside reverse transcriptase inhibitor resistance mutations

A comparison of sequences isolated from EFV and NVP-treated patients to sequences from nNRTI-naive patients, showed the association of 27 mutations with EFV treatment and 24 mutations with NVP treatment, excluding known NRTI resistance mutations (Fig. 1).

The consensus Bayesian networks that were learned from the data sets are shown in Figs 2 and 3. As in the EFV network, the novel minor mutations 28K, 90I, 196R, (currently not considered by expert genotypic interpretation systems) and 221Y (present only in the Rega 7.1 interpretation system, and added to this version based on findings in this manuscript) were associated with EFV treatment or with known EFV resistance mutations; they are potentially involved in resistance to EFV. Likewise, in the NVP network, the novel mutations 101H, 138Q, and 221Y were associated with NVP treatment or with known NVP resistance mutations and are thus potentially involved in resistance to NVP. The novel mutation 28K was mostly confined to subtype G, and within this subtype, was present in 25 out of 189 (13%) patients treated with EFV. The involvement of 28K in NRTI resistance through an interaction with mutations at position 215 (bootstrap support 38% in the NVP network) could not be excluded.

Despite the apparent interaction between mutation 188F and 190S, the mutation 188F is not likely to be directly involved in resistance, but instead a 'transitional' mutation in the two-nucleotide change required for the wild-type 188Y to mutate to 188L, as in the data sets, it appeared in 14 of 15 samples in a mixture with 188Y or 188L.

The novel mutation 31L was associated with 184V (bootstrap support 83%) and is therefore likely involved in lamivudine (3TC) resistance. Mutations 203K and 228H, and to a lesser extent mutation 228R, were associated with the thymidine analogue mutations (TAMs) 219Q/E, and thus likely involved in these TAM pathways of resistance to zidovudine (ZDV) and stavudine (D4T).

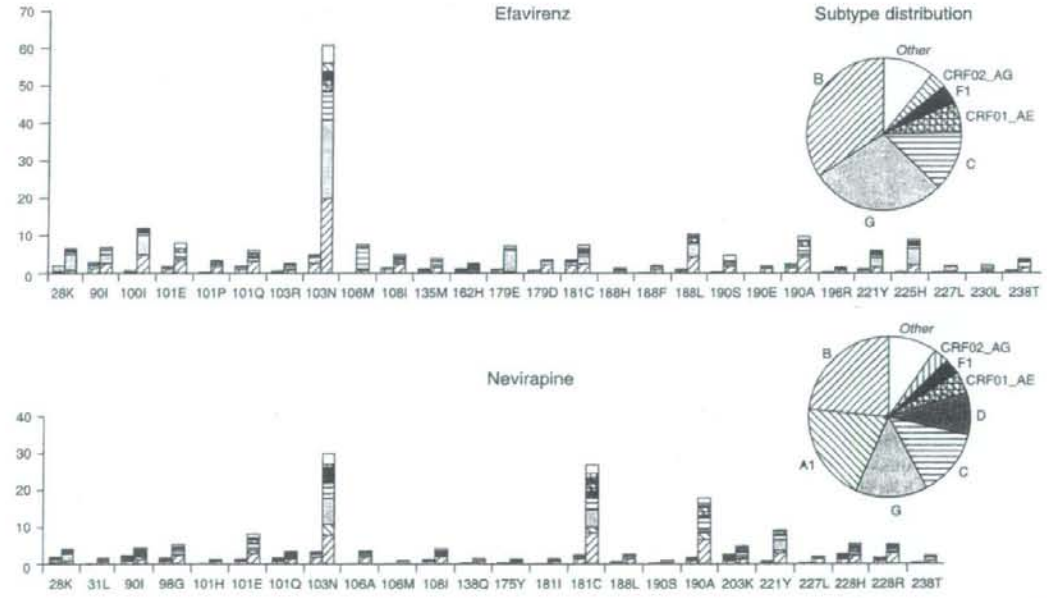

Fig. 1. Dataset prevalence (%) in sequences from nNRTI-naive (left bar) and nNRTI-treated (right bar) patients of mutations significantly associated with EFV or NVP treatment (FDR = 0.05). For both drugs, the data were stratified for the overall subtype distribution of the sequences, shown in the inset pies, to be identical for treated and untreated patients. Known NRTI resistance mutations that were associated with treatment are not shown. FDR, false discovery rate; NRTI, nucleoside reverse transcriptase inhibitor; nNRTI, nonnucleoside reverse transcriptase inhibitor; NVP, nevirapine; EFV, efavirenz.

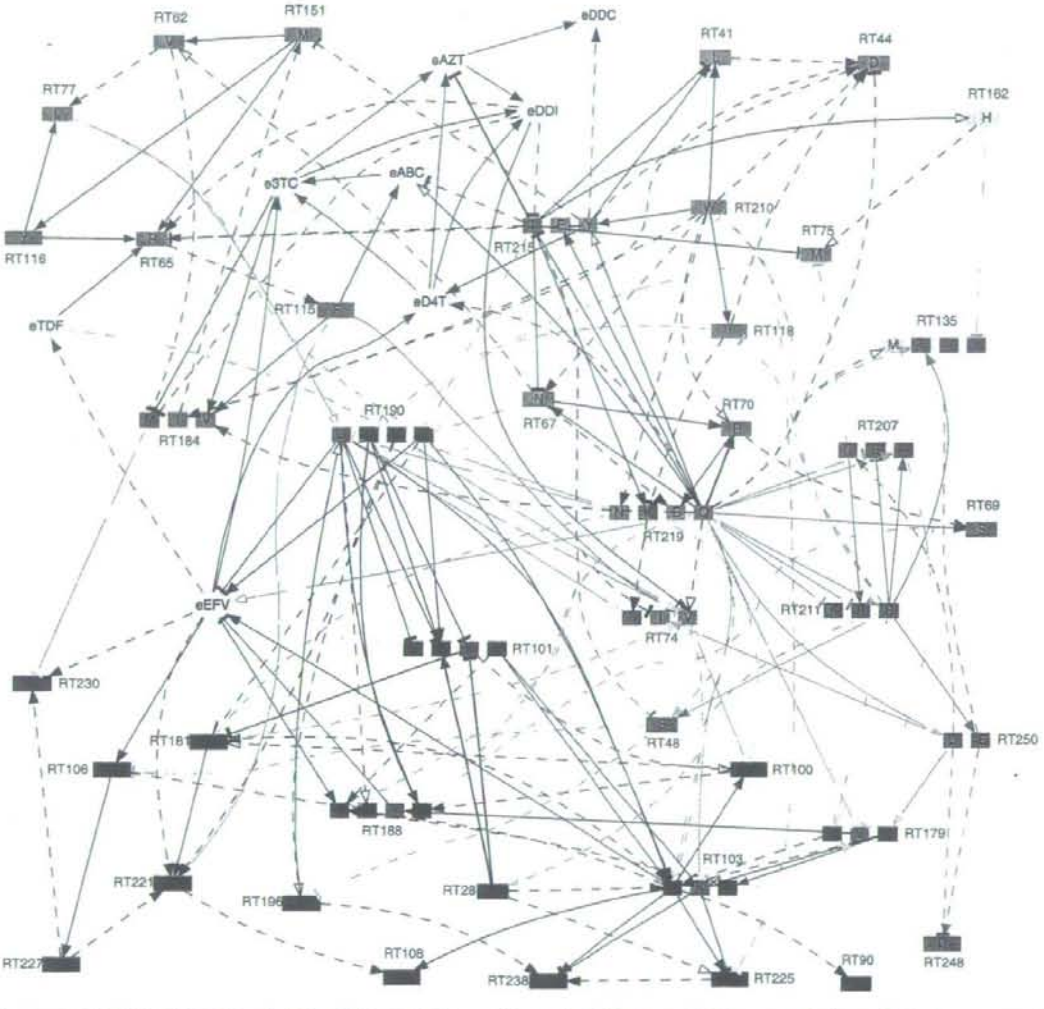

Fig. 2. Annotated Bayesian network learned for analyzing resistance to EFV, learned autonomously from clinical sequence data. An arc represents a direct dependency between the corresponding variables and thickness is proportional to bootstrap support. Solid arcs have bootstrap support over 65% and were considered robust. Dashed arcs have bootstrap support over 35% and lack of arcs was also considered robust. Arc color indicates whether it is a direct influence between resistance mutations (black), an influence from background polymorphisms on resistance mutations (blue), a direct influence between treatment reflecting bias in treatment combinations (purple), associations between background polymorphisms (green), or other interactions such as between nNRTI and NRTI resistance mutations (gray). Arc direction does not reflect cause and effect, nor order of accumulations of mutations, but may indicate a nonadditive multivariable effect, which requires analysis of the quantitative components of the Bayesian network. Arcs were colored according to their function to improve reading the graph, but coloring is only indicative. EFV, efavirenz; NRTI, nucleoside reverse transcriptase inhibitor; nNRTI, nonnucleoside reverse transcriptase inhibitor.

The involvement of 228R in nNRTI resistance through an interaction with 181Y (bootstrap support 61%) could however not be excluded.

The role for novel mutations 135M, 162H, and 175Y could not be clarified as robust associations were lacking for these mutations in the Bayesian networks.

Efavirenz resistance pathways

For resistance to EFV, next to the predominant use of 103N, HIV-1 escapes through pathways marked by mutations 188H/L or 190E/S/A in a minority of sequences, as indicated by the connectivity around the experience with EFV (eEFV) node in the network. These pathways are not exclusive, as combinations of

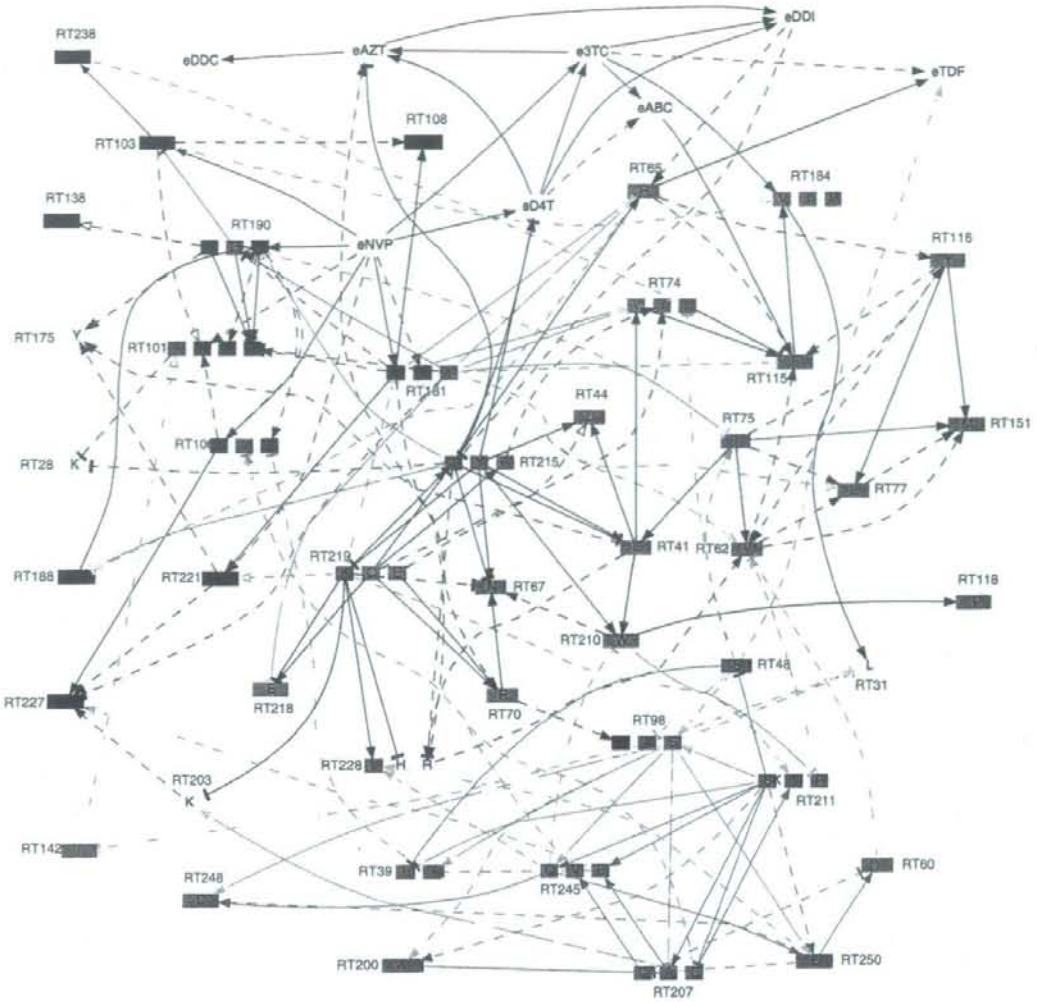

Fig. 3. Annotated Bayesian network learned for analyzing resistance to NVP. Legend as in Fig. 2. NRTI, nucleoside reverse transcriptase inhibitor; nNRTI, nonnucleoside reverse transcriptase inhibitor; NVP, nevirapine.

these major mutations were also commonly found, showing that they reinforce each other when present together. The 106M mutation, which was found almost exclusively in subtype C sequences, confirming earlier findings [16], was almost never observed without 103N or 190E/S/A. Even though it can therefore be considered a minor mutation, it was connected directly to the eEFV node because it was common to several different pathways. Mutation 227L showed a robust interaction with mutation 106M and was also found exclusively in subtype C.

The network indicated that the 103N pathway is further associated with mutations 90I, 100I, 101P, 108I, 181C, 221Y, 225H, 238T, and 190A. The 188L/F pathway was further associated with mutations 179D (in subtypes B and C), 179E (in subtype G), and the 103R mutation. The 190E/S/A pathway was further associated with mutations 28K, 101E/Q, 103N, 181C, and 221Y.

Several interactions between EFV and NRTI resistance mutations were indicated by the network. Mutation 190E, which is selected rarely by EFV, was indicated to strongly interact with NRTI resistance mutations 74V, 75M, and 219N. Mutation 100I showed a robust interaction with NRTI resistance mutations 74V and 219N, mutations 179D/E and 103N showed an interaction with NRTI resistance mutation 219Q, and mutation 230L showed a robust interaction with NRTI resistance mutations 184I/V.

Nevirapine resistance pathways

For resistance to NVP, the network suggested 103N, 190A, and 181C/I as major mutations, which were connected directly to the experience with NVP (eNVP) node. Similar to resistance to EFV, these major mutations do not mark exclusive pathways, but reinforce each other when present together. Mutation 106A was almost never seen without 103N or 190A, indicating that it was connected to the eNVP node as a minor mutation not specific to one particular pathway. As with 106M for resistance to EFV, it was further associated with mutation 227L.

The network indicated that the 103N pathway was further associated with mutations 190A/S, and 238T and possibly 108I. The 190A pathway was further associated with mutation 101E and possibly also mutations 101Q and 188L. The 181C pathway was further associated with mutation 221Y. The novel mutations 28K and 175Y could not be assigned clearly to be selected by NVP or an NRTI.

As for EFV, NRTI resistance mutations at position 74 and 75 interacted with NVP resistance mutations, but differently. Mutation 181C showed a strong interaction with NRTI resistance mutations 65R and 74V, and mutations 188L and 190A showed a strong interaction with NRTI resistance mutations 215Y/F.

Common features in both efavirenz and nevirapine Bayesian networks

The networks showed a remarkable resemblance of interactions between the resistance mutations that were in common for resistance to NVP and EFV. For example, both networks showed that 108I and 238T were associated with 103N, 101E was associated with 190A/S, 227L was associated with 106A/M, 221Y was associated with 181C, and 28K was associated with 101E.

The networks also confirmed current knowledge on the role of many resistance mutations in relation to NRTI treatment. For example, both networks indicated that 184V was the major resistance mutation selected by 3TC, 115F was the major resistance mutation selected by abacavir (ABC), 65R was the major resistance mutation selected by tenofovir (TDF), and 215F/Y were the major resistance mutations selected by ZDV. Also mutations from the multi-NRTI resistance 151 complex (62V, 77L, 116Y and 151M) clustered clearly together in both networks. Rare antagonistic effects between NRTI resistance mutations that have been reported before were also indicated, between 65R and 215Y/F, and between 151M and 215Y (the latter only in the EFV network).

Discussion

Novel resistance mutations

On the basis of higher prevalence in sequences from treated versus untreated patients, together with a robust association with a drug or a known resistance mutation, we confirmed the selection of many known mutations, but also identified selection of mutations currently not considered by expert interpretation system. In particular, mutations 28K, 90I, 196R, and 221Y are most likely involved in resistance to EFV, and mutations 101H, 138Q, and 221Y in resistance to NVP. Although the novel EFV mutations 28K and 90I were also associated with NVP treatment, they did not associate in a robust way with known NVP resistance mutations in the Bayesian network, probably owing to the small prevalence numbers. Of these novel mutations, most notably 221Y was selected at high levels by both NVP and EFV treatment, and at levels higher than some mutations currently included in the IAS list of resistance mutations such as 108I or 230L. Also, mutation 28K was selected in subtype G, more than most IAS listed mutations (for EFV, only with the exception of 103N and 100I, and for NVP only with the exception of 103N and 181C). The interaction of this mutation with 101E, indicated by both networks, may in part be explained by an association between these mutations in the p51/p66 interface of the reverse transcriptase heterodimer, as residue 28 in p51 and residue 101 in p66 are closely together with opposing charge and likely connected with a salt bridge (Frederico Gago and Joeri Auwerx, personal communication).

The clinical impact of most of these novel mutations on the activity of NVP or EFV may be limited, as they accumulate only after major resistance mutations (such as 103N or 181C) which on their own already reduce the activity of these low genetic barrier drugs such that a treatment change is necessary even in absence of the new mutations. Their selection by existing commonly used drugs may however impact cross-resistance with novel nNRTIs such as etravirine.

Some of the treatment-associated mutations were suggested to be involved in NRTI resistance rather than nNRTI resistance. The NVP treatment-associated 31L was robustly associated with 184V, and therefore likely involved in 3TC resistance. Given the high impact of 184V on replication capacity, perhaps this mutation partially restores this adverse effect of 184V. Other mutations likely involved in resistance to TAMs were 203K and 228H (and perhaps also 228R).

Of these novel mutations, mutations 28K, 101H, 138Q, and 196R were not previously reported as resistance mutations. Mutations 203K, 228H/R and 208Y were reported in other studies to be involved in NRTI resistance [17–20], and 90I and 221Y in nNRTI resistance [17,20,21], in agreement with our results. Contrary to our results, however, mutation 31L has been reported to be involved in nNRTI resistance [22] based on an association between the presence of this mutation and a reduced phenotypic susceptibility to nNRTIs, in univariable analyses in isolates without known nNRTI

resistance mutations. However, this association could not be confirmed in their multivariable analyses, suggesting that the univariable association with nNRTI phenotypic resistance was not causal. Multivariable analyses for association with phenotypic susceptibility (not only to nNRTIs but also to 3TC) are needed to confirm that mutation 31L is involved in resistance to 3TC rather than nNRTIs.

Also contrary to our results, mutations 228H/R have been reported to be involved in nNRTI resistance in two studies [17,21]. In contrast, we found that mutation 228H is involved in NRTI resistance (connected to mutations at position 219 with bootstrap support over 90%), whereas the role of 228R is less certain and perhaps dual. These conclusions were based on the association of these mutations with decreased susceptibility to nNRTI resistance, in a univariable analysis, and with an increased prevalence in nNRTI-treated patients compared with treatment-naive patients. It may not be excluded that the association found between presence of 228H/R and decreased susceptibility to nNRTIs may have been caused by the presence of other nNRTI mutations in the viral sequence, and no multivariable analyses supported their conclusions. As their data showed that in patients naive to nNRTIs, but experienced with NRTIs, the mutation 228H had a prevalence over 5%, compared with being virtually absent in treatment-naive patients, this is in agreement with a role for this mutation in NRTI resistance.

Subtype differences

One of the objectives was to look for subtype-dependent differences that could be explained by interactions with wild-type polymorphisms. Therefore, we included polymorphisms (some of which linked with subtype) in the analyses. In contrast with protease resistance mutations, very few interactions between background polymorphisms and nNRTI resistance mutations were found, indicating that observed differences are to be explained mostly by the impact of different codon usage on genetic barrier. The subtype dependence of 106M was previously explained in this way [8,16], as is the absence of 108I in subtype G (requiring two transitions), and the preference for 179E instead of 179D, also in subtype G [23]. Instead, robust interactions between polymorphisms and resistance mutations involved NRTI resistance mutations, but because data sets were not stratified according to subtype with respect to NRTI treatment, these associations are likely artifacts given that they were different in the EFV versus NVP network. The synergistic interaction between the polymorphism 135T and resistance mutation 103N [21] could not be confirmed in the Bayesian network analysis (although the association was present in the data set, Fisher's exact test, $P < 0.01$), and the Bayesian network could not indicate a possible cause for the confinement of the novel mutation 28K to subtype G.

Interactions between nonnucleoside reverse transcriptase inhibitor and nucleoside reverse transcriptase inhibitor resistance mutations

Interactions between nNRTI and NRTI resistance pathways have been repeatedly observed. For example, as early as in 1994, it was observed that in absence of ZDV, mutation 181C was the most prevalent NVP-selected mutation, whereas coadministration with ZDV prevented this mutation [24]. Using Bayesian network learning, it was possible to confirm that the mechanism for this observation is not an effect of ZDV directly, but rather an interaction of nNRTI resistance mutations with mutations at position 215, a major resistance position for resistance to ZDV [25]. The interaction between 190E and 74V or 75I was reported previously and verified with in-vitro experiments [26]. A number of novel interactions between NRTI and nNRTI resistance mutations were identified. The interaction between 184I/V and 230L may potentially be explained using the three-dimensional (3D) structure of the enzyme [27], by a direct steric interaction between these residues, which are closely located (<6 Å). The 184I/V mutations have been demonstrated to have a clinical effect due to lowered replication capacity [28], and have also been reported to increase reverse transcriptase fidelity [29]. Therefore, it may be interesting to investigate how the interaction between 230L and 184V influences these effects. The interactions between mutation 219N and mutations 190E and 100I, between mutations 100I and 74V, and between mutations 181C and 65R warrant more investigations. Some of these interactions could be involved in the re-sensitization by certain NRTI resistance mutations of susceptibility to nNRTIs [30,31] or vice versa. On the basis of the observed interactions and considering the difference in preferred mutations selected by NVP versus EFV, one could argue that certain treatment combinations are more likely to fail more quickly than other treatments. Most notably, the suggested interaction between 181I/C and 65R could indicate that a treatment including TDF and NVP will lead to a more rapid failure than a treatment including TDF and EFV.

Virological failure on combination therapy may be associated with resistance to one, two or all drugs in the combination. Typically, resistance to one drug will accelerate the development of resistance to the other drugs, as the inhibition of virus replication is weakened, and therefore the virus may actively replicate during selective pressure of the remaining active drugs in the therapy. Therefore, associated prevalence of NRTI and nNRTI resistance mutations may not necessarily imply a biochemical interaction. Still, some of the interactions that were found were previously described, or are plausible given the 3D structure of the enzyme. For interactions that involve mutations which are not the most common resistance mutations selected by specific drugs, a biological reason is the most likely explanation, in particular when the observed unconditional dependencies in the networks

were found highly robust and involved similar positions in both networks. However, as it cannot be excluded that the analyses were confounded, these interactions should be confirmed with in-vitro experiments.

Limitations

Our analysis was limited in two important ways. First, because only a fragment of reverse transcriptase is routinely sequenced, we were unable to find mutations outside this region that were involved in drug resistance development such as the mutations recently reported in the RNase H and connections domains [32]. Secondly, for computational reasons, only polymorphisms with a prevalence over 15% in nNRTI-naive patients were included. It is conceivable that polymorphisms with a lower prevalence may still explain some subtype dependencies that could not be explained (such as the occurrence of 28K in subtype G).

Conclusion

Bayesian network learning proved a valuable tool to untangle the simultaneous selection of mutations in reverse transcriptase by two classes of inhibitors, and contributed to identify novel mutations likely involved in EFV and NVP resistance. Unlike for protease, subtype dependencies in selection of resistance mutations are less common and seldom explained by interaction with polymorphic residues. The Bayesian networks indicated both known and novel interactions between nNRTI and NRTI resistance mutations that explain observed interactions seen in different treatment combinations.

Acknowledgements

The authors wish to thank E. Van Wijngaerden, M.E. Valadas, M.J. Aguas, L. Rosado, J. Vera, J. Poças, T. Batista, I. Soares, T. Branco, A. Mouzinho, E. Teófilo, K. Mansinho, T. Faria, and A. Abecasis for data collection and H. Blockeel and Y. Moreau for review of the manuscript.

K.D. was funded by a PhD grant of the Institute for the Promotion of Innovation through Sciences and Technology in Flanders (IWT). This work was supported by the AIDS Reference Laboratory of Leuven that receives support from the Belgian Ministry of Social Affairs through a fund within the Health Insurance System, and partially funded through the Katholieke Universiteit Leuven (Grant OT/04/43), the European Commission (VIROLAB project, EU IST STREP # 027446), and the Belgian Interuniversity Attractions Pole (IAP-P6/41).

K.D. conceived the study, performed the analyses and contributed to the manuscript. R.C., Z.G., K.V.L., M.A.S., D.A.K., P.R.H., R.K., R.S., A.-M.V., and the non-B workgroup contributed clinical and virological data. A.-M.V. supervised the work. All authors contributed to the design of the study, interpretation and discussion of the results, and to the manuscript.

The non-B workgroup consists of Rami Kantor of the Division of Infectious Diseases, Brown University, Providence, Rhode Island, USA; David A. Katzenstein, and Robert W. Shafer of the Division of Infectious Disease, Stanford University, Stanford, California, USA; Ricardo J. Camacho and Ana Patricia Carvalho of the Molecular Biology Laboratory, Centro Hospitalar de Lisboa Ocidental, Lisbon, Portugal; Brian Wynhoven and P. Richard Harrigan of the BC Centre for Excellence in HIV/AIDS, Vancouver, Canada; Patricia Cane of the Health Protection Agency Antiviral Susceptibility Reference Unit, Birmingham, UK; John Clarke and John Weber of the Wright Fleming Institute, Imperial College, St Mary's Hospital, London, UK; Sunee Sirivichayakul and Praphan Phanuphak of the Chulalongkorn University, Bangkok, Thailand; Macelo A. Soares of the Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; Amilcar Tanuri of the Universidade Federal do Rio de Janeiro, Rio de Janeiro – Rio de Janeiro, Brazil; Joke Snoeck and Anne-Mieke Vandamme of the Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium; Lynn Morris of the National Institute of Communicable Diseases, Johannesburg, South Africa; Hagit Rudich and Zehava Grossman of the Central Virology, PHL, Ministry of Health, Tel-Hashomer, Israel; Jonathan M Schapiro of the National Hemophilia Center, Tel-Hashomer, Israel; Rosangela Rodrigues and Luis F Brigido Instituto of the Adolfo Lutz, Sao Paulo, Brazil; Africa Holguin and Vincent Soriano of the Hospital Carlos III, Madrid, Spain; Koya Ariyoshi and Wataru Sugiura of the National Institute of Infectious Diseases, Tokyo, Japan; Maria Belen Bouzas and Pedro Cahn of the Fundación Huesped, Buenos Aires, Argentina; Deenan Pillay of the Department of Infection, University College London, London, United Kingdom; Terese L. Katzenstein and Louise Bruun Jørgensen of the Department of Infectious Diseases, Rigshospitalet, Copenhagen, Denmark.

References

1. Shafer RW. **Genotypic testing for human immunodeficiency virus type 1 drug resistance.** *Clin Microbiol Rev* 2002; **15**:247–277.

2. Van Laethem K, De Luca A, Antinori A, Cingolani A, Perno CF, Vandamme AM. **A genotypic drug resistance interpretation algorithm that significantly predicts therapy response in HIV-1 infected patients.** *Antivir Ther* 2002; **7**:1359–6535.
3. Frater AJ, Beardall A, Ariyoshi K, Churchill D, Galpin S, Clarke JR, *et al.* **Impact of baseline polymorphisms in RT and protease on outcome of highly active antiretroviral therapy in HIV-1-infected African patients.** *AIDS* 2001; **15**:1493–1502.
4. Parkin NT, Schapiro JM. **Antiretroviral drug resistance in non-subtype B HIV-1, HIV-2 and SIV.** *Antivir Ther* 2004; **9**:3–12.
5. Brindeiro PA, Brindeiro RM, Mortensen C, Hertogs K, De Vroey V, Rubini NPM, *et al.* **Testing genotypic and phenotypic resistance in human immunodeficiency virus type 1 isolates of clade B and other clades from children failing antiretroviral therapy.** *J Clin Microbiol* 2002; **40**:4512–4519.
6. Ariyoshi K, Matsuda M, Miura H, Tateishi S, Yamada K, Sugiura W. **Patterns of point mutations associated with antiretroviral drug treatment failure in CRF01_AE (subtype E) infection differ from subtype B infection.** *J Acquir Immune Defic Syndr* 2003; **33**:336–342.
7. Grossman Z, Vardinon N, Chemtob D, Afkan ML, Bentwich Z, Burke M, *et al.* **Genotypic variation of HIV-1 reverse transcriptase and protease: comparative analysis of clade C and clade B.** *AIDS* 2001; **15**:1453–1460.
8. Grossman Z, Istomin V, Averbuch D, Lorber M, Risenberg K, Levi I, *et al.*, Israel AIDS Multi-Center Study Group. **Genetic variation at NNRTI resistance-associated positions in patients infected with HIV-1 subtype C.** *AIDS* 2004; **18**:909–915.
9. Camacho RJ, Vandamme AM. **Antiretroviral resistance in different HIV-1 subtypes: impact on therapy outcomes and resistance testing interpretation.** *Curr Opin HIV AIDS* 2007; **2**:123–129.
10. van de Vijver DA, Wensing AMJ, Angarano G, Asjö B, Balotta C, Boeri E, *et al.* **The calculated genetic barrier for antiretroviral drug resistance substitutions is largely similar for different HIV-1 subtypes.** *J Acquir Immune Defic Syndr* 2006; **41**:352–360.
11. Deforche K, Silander T, Camacho R, Grossman Z, Soares MA, Van Laethem K, *et al.* **Analysis of HIV-1 pol sequences using Bayesian Networks: implications for drug resistance.** *Bioinformatics* 2006; **22**:2975–2979.
12. WHO. *Towards universal access by 2010: how WHO is working with countries to scale-up HIV prevention, treatment, care and support.* Geneva: WHO; 2006.
13. Johnson VA, Brun-Vézinet F, Bonaventura C, Günthard H, Kuritzkes DR, Pillay D, *et al.* **Update of the drug resistance mutations in HIV-1: 2007.** *Top HIV Med* 2007; **15**:119–125.
14. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, *et al.* **An automated genotyping system for analysis of HIV-1 and other microbial sequences.** *Bioinformatics* 2005; **21**:3797–3800.
15. Friedman N, Goldszmidt M, Wyner AJ. **Data analysis with Bayesian networks: a bootstrap approach.** In: *Proceedings of UAI–99.* San Francisco, CA, 1999. pp. 206–215.
16. Brenner B, Turner D, Oliveira M, Moisi D, Detorio M, Carobene M, *et al.* **A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to nonnucleoside reverse transcriptase inhibitors.** *AIDS* 2003; **17**:F1–F15.
17. Saracino A, Monno L, Scudeller L, Cibelli DC, Tartaglia A, Punzi G, *et al.* **Impact of unreported HIV-1 reverse transcriptase mutations on phenotypic resistance to nucleoside and non-nucleoside inhibitors.** *J Med Virol* 2006; **78**:9–17.
18. Svicher V, Sing T, Santoro MM, Forbici F, RodrÂguez-Barrios F, Bertoli A, *et al.* **Involvement of novel human immunodeficiency virus type 1 reverse transcriptase mutations in the regulation of resistance to nucleoside inhibitors.** *J Virol* 2006; **80**:7186–7198.
19. Rhee SY, Fessel WJ, Zolopa AR, Hurley L, Liu T, Taylor J, *et al.* **HIV-1 protease and reverse-transcriptase mutations: correlations with antiretroviral therapy in subtype B isolates and implications for drug-resistance surveillance.** *J Infect Dis* 2005; **192**:456–465.
20. Cane PA, Green H, Fearnhill E, Dunn D, and the UK collaborative group on HIV drug resistance. **Identification of accessory mutations associated with high-level resistance in HIV-1 reverse transcriptase.** *AIDS* 2007; **21**:447–455.
21. Ceccherini-Silberstein F, Svicher V, Sing T, Artese A, Santoro MM, Forbici F, *et al.* **Characterization and structural analysis of novel mutations in human immunodeficiency virus type 1 reverse transcriptase involved in the regulation of resistance to nonnucleoside inhibitors.** *J Virol* 2007; **81**:11507–11519.
22. Parkin NT, Gupta S, Chappey C, Petropoulos CJ. **The K101P and K103R/V179D mutations in human immunodeficiency virus type 1 reverse transcriptase confer resistance to nonnucleoside reverse transcriptase inhibitors.** *Antimicrob Agents Chemother* 2006; **50**:351–354.
23. Vergne L, Peeters M, Mpoudi-Ngole E, Bourgeois A, Liegeois F, Toure-Kane C, *et al.* **Genetic diversity of protease and reverse transcriptase sequences in nonsubtype-B human immunodeficiency virus type 1 strains: evidence of many minor drug resistance mutations in treatment-naive patients.** *J Clin Microbiol* 2000; **38**:3919–3925.
24. Richman DD, Havlir D, Corbeil J, Looney D, Ignacio C, Spector SA, *et al.* **Nevirapine resistance mutations of human immunodeficiency virus type 1 selected during therapy.** *J Virol* 1994; **68**:1660–1666.
25. Selmi B, Deval J, Alvarez K, Boretto J, Sarfati S, Guerreiro C, *et al.* **The Y181C substitution in 3′-azido-3′-deoxythymidine-resistant human immunodeficiency virus, type 1, reverse transcriptase suppresses the ATP-mediated repair of the 3′-azido-3′-deoxythymidine 5′-monophosphate-terminated primer.** *J Biol Chem* 2003; **278**:40464–40472.
26. Boyer PL, Gao HQ, Hughes SH. **A mutation at position 190 of human immunodeficiency virus type 1 reverse transcriptase interacts with mutations at positions 74 and 75 via the template primer.** *Antimicrob Agents Chemother* 1998; **42**:447–452.
27. Ren J, Milton J, Weaver KL, Short SA, Stuart DI, Stammers DK. **Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase.** *Structure* 2000; **8**:1089–1094.
28. Turner D, Brenner B, Wainberg MA. **Multiple effects of the M184V resistance mutation in the reverse transcriptase of human immunodeficiency virus type 1.** *Clin Diagn Lab Immunol* 2003; **10**:979–981.
29. Hsu M, Inouye P, Rezende L, Richard N, Li Z, Prasad VR, *et al.* **Higher fidelity of RNA-dependent DNA mispair extension by M184V drug-resistant than wild-type reverse transcriptase of human immunodeficiency virus type 1.** *Nucleic Acids Res* 1997; **25**:4532–4536.
30. Shulman N, Zolopa AR, Passaro D, Shafer RW, Huang W, Katzenstein D, *et al.* **Phenotypic hypersusceptibility to non-nucleoside reverse transcriptase inhibitors in treatment-experienced HIV-infected patients: impact on virological response to efavirenz-based therapy.** *AIDS* 2001; **15**:1125–1132.
31. Shulman NS, Bosch RJ, Mellors JW, Albrecht MA, Katzenstein DA. **Genetic correlates of efavirenz hypersusceptibility.** *AIDS* 2004; **18**:1781–1785.
32. Jones J, Taylor B, Wilkin TJ, Hammer SM. **Advances in antiretroviral therapy.** *Top HIV Med* 2007; **15**:48–82.

Parasite Immunology, 2008, **30**, 13–21 DOI: 10.1111/j.1365-3024.2007.00988.x

Recognition profiles of microsporidian *Encephalitozoon cuniculi* polar tube protein 1 with human immunoglobulin M antibodies

K. FURUYA,[1] M. OMURA,[1] S. KUDO,[2] W. SUGIURA[3] & H. AZUMA[4]

[1]Department of Parasitology, National Institute of Infectious Diseases, Tokyo, Japan, [2]Hokkaido Institute of Public Health, Sapporo, Japan, [3]AIDS Research Center, National Institute of Infectious Diseases, Tokyo, Japan, [4]Hokkaido Red Cross Blood Center, Sapporo, Japan

SUMMARY

Microsporidian Encephalitozoon cuniculi *has a unique organelle called a polar tube (PT), the extrusion of which is absolutely required to invade a host cell. We recently detected anti-*E. cuniculi *PT immunoglobulin (Ig) M antibodies in sera from many healthy individuals. The present one-dimensional (1-D) immunoblot analysis predominantly detected a band at 52 kDa in all of the examined human sera with anti-PT IgM. The use of mouse monoclonal antibody confirmed that the 52-kDa band detected in 1-D immunoblots was an antigen derived from the PT, which represents a glycoprotein nature. In addition, from changes in the immunoreactivity of the 52-kDa band before and after treatment with NaOH, we determined that the 24 human serum samples with anti-PT IgM activities could be roughly grouped into three types: (i) sera containing antibodies against only a saccharic determinant (n = 3); (ii) sera containing antibodies against only a proteinic determinant (n = 11); and (iii) sera showing dual recognition of saccharic and proteinic determinants (n = 10). Further two-dimensional (2-D) immunoblot analysis followed by proteomic analysis confirmed that human sera with anti-PT IgM reacted with* E. cuniculi *polar tube protein 1 (PTP1). Such circulating IgM antibodies may be important in the first line of defence against* E. cuniculi *infection.*

Keywords *antigenic determinant,* Encephalitozoon cuniculi, *human IgM antibody, Microsporidia, polar tube*

Correspondence: K. Furuya, Department of Parasitology, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan (e-mail: kfuruya@nih.go.jp).
Received: 17 July 2007
Accepted for publication: 14 September 2007

INTRODUCTION

Encephalitozoon cuniculi is a spore-forming obligate intracellular parasitic pathogen belonging to the phylum Microsporidia (1); it can also be a zoonotic parasite (2). Various animals can be naturally infected by *E. cuniculi* and humans can also be affected by this pathogen (3). However, most *E. cuniculi* infections occur in human immunodeficiency virus (HIV)-infected immunocompromised patients (2,4). A few cases of *E. cuniculi* infection have also been reported in transplant patients (2,5,6). Apart from an accidentally infected French individual who had severe keratoconjunctivitis (7), no symptomatic cases of infection with *E. cuniculi* among immunocompetent persons have been described (2,4). Thus, it is most unlikely that *E. cuniculi* causes microsporidiosis in immunocompetent persons (2,4). In humans, it can be regarded as an opportunistic pathogen (8,9).

We recently demonstrated immunoglobulin (Ig) M antibodies against the polar tubes (PTs) of *E. cuniculi* in about 36% of healthy people in Japan (10); by contrast, such IgM antibodies were poorly detected among HIV-positive persons. The rate of positivity for anti-PT IgM was significantly reduced with increasing age and decreasing CD4 cell count. These seroepidemiological results clearly indicate that circulating anti-PT IgM antibodies that are capable of strongly reacting with filaments extruded from germinated spores exist, thus suggesting that such antibodies may play a part in protective immunity.

The main objectives of this study were to provide further information about the immunoreactivity of human anti-PT IgM and to identify the corresponding determinants using one-dimensional (1-D) and two-dimensional (2-D) immunoblot analyses, and other biochemical and proteomic means.

MATERIALS AND METHODS

Serum samples

Twenty-four human serum samples were used for this study. These samples were collected from healthy individuals

under the ethical considerations described in a previous study (10).

As the positive control, serum samples from 11 infected rabbits were also used for this study. All the rabbits were sacrificed after approval from the breeding owners and were microbiologically confirmed to carry the pathogen in organs such as the brain and kidney. Isolated *E. cuniculi* spores from the organs were identified by polymerase chain reaction, as described before (11). The internal transcribed spacer gene sequence revealed that the *E. cuniculi* isolates were all classified into genotype 1.

In addition, 11 normal rabbit serum samples, which were serologically negative, were used as the negative control.

Mouse monoclonal antibody

A mouse monoclonal antibody against the soluble antigen of *E. cuniculi* PTs was generated using a hybridoma technique (12), which was tentatively named MAb2 for this study. After the second cloning, culture supernatants containing MAb2 were employed in the present study. In a later study, monoclonal antibodies from mouse ascitic fluids were available and were classified as IgE (manuscript in preparation). To detect IgE in this study, we used goat anti-mouse IgG + M + A (H + L) and rabbit anti-mouse IgG (H + L), which were labelled with enzyme, as documented below. All the reagents used could detect IgE, since they are also specific for light chains.

Preparation of purified spores of *Encephalitozoon cuniculi*

Encephalitozoon cuniculi spores of strain HF (genotype 1) were collected from the culture supernatants of infected RK 13 cells and purified using Percoll gradient centrifugation, as described previously (10). Finally, separated spores were filtered through Ultrafree-MC membranes (5·0 μm; Millipore Corporation, Bedford, MA).

Enzyme-linked immunosorbent assay (ELISA)

As described in a previous paper (10), anti-*E. cuniculi* PT antibodies were measured using a 96-well flat-bottom microplate coated with germinated spores. Twofold dilutions of each serum sample, starting a 1 : 50 dilution, were tested; bound antibodies were detected by the secondary antibody labelled with peroxidase (PO). Signals of PO bound to human or animal antibodies were visualized using an aminoethyl carbazole substrate kit (Zymed Laboratories Inc., San Diego, CA). The following PO conjugates were used as the secondary antibodies: rabbit anti-human IgM (μ-chain specific; QED Bioscience, San Diego, CA) for human IgM; goat anti-mouse IgG + M + A (H + L; Zymed Laboratories) for mouse monoclonal antibody. Also, protein A/G–PO conjugate (Prozyme Inc., San Leandro, CA) and protein A–PO conjugate (Kirkegaard & Perry Laboratories, Gaithersburg, MD) were used for detecting human IgG and rabbit IgG, respectively. Dilutions of the second reagents were adjusted between 1 : 3000 and 1 : 5000.

1-D polyacrylamide gel electrophoresis (1-D PAGE)

Purified *E. cuniculi* spores were suspended in an appropriate volume of Laemmli sample buffer (Bio-Rad Laboratories, Hercules, CA) containing 5% 2-mercaptoethanol (Bio-Rad) and heated at 95°C for 5 min. After centrifugation at about 7800 *g* for 5 min, pellets were treated twice more with the same buffer and the three supernatants were pooled. The protein fraction of these pooled supernatants was desalted using a NAP-5 column (GE Healthcare Bio-Sciences AB, Uppsala, Sweden). This soluble protein fraction was used as an antigen for 1-D immunoblot analysis.

Approximately 60 μg of the microsporidial soluble fraction was electrophoresed vertically using 15% polyacrylamide Ready mini-gels (Bio-Rad) under reducing conditions.

1-D immunoblot analysis

Electrophoresed spore proteins separated by 1-D PAGE were transferred electrically onto polyvinylidene fluoride (PVDF) membranes (immobilon transfer membranes; Millipore Corporation). Blotted membranes were cut into strips of about 4-mm width. Half of the strips were treated with a 0·055 M NaOH solution for 16 h at 40°C to remove carbohydrate chains from glycoproteins, according to the method described by Duk *et al.* (13). NaOH-treated strips were washed extensively with pure water.

All of the strips to be tested were treated with blocking buffer (SuperBlock; Pierce Chemical Company, Rockford, IL) and subsequently incubated with human sera, rabbit sera or a culture supernatant containing MAb2. Dilutions of human sera, rabbit sera and the supernatant to be tested were 1 : 100. Secondary antibodies labelled with alkaline phosphatase (ALP) were used to detect primary antibodies: goat anti-human IgM (μ-chain specific; Kirkegaard & Perry Laboratories) for human IgM; goat anti-rabbit IgM (μ-chain specific; Southern Biotechnology Associates, Birmingham, AL) for rabbit IgM; rabbit anti-mouse IgG (H + L; Zymed Laboratories) for mouse monoclonal antibody. The secondary antibodies were used at dilutions of 1 : 1000–1 : 3000. Reactions with the primary antibodies and the secondary antibodies were performed for 90 min at room temperature with shaking. After extensive washing with phosphate-buffered saline containing 0·05% Tween 20 (PBS-T), antibody binding was detected using a BCIP/

NBT phosphatase substrate kit (Kirkegaard & Perry Laboratories).

For each human serum, mouse monoclonal antibody and rabbit serum, the change in immunoreactivity between a NaOH-treated strip and an untreated strip was observed. In addition, the participation of a saccharic determinant in an immunological reaction was determined.

Analysis of lectin–carbohydrate interactions

Using some of the NaOH-treated or NaOH-untreated strips described above, the reactivities of various biotinylated lectins were tested. For this test, a biotinylated lectin kit I (BK-1000; Vector Laboratories, Buringame, CA), containing concanavalin A (ConA), *Dolichos biflorus* agglutinin, peanut agglutinin, *Ricinus communis* agglutinin I, soybean agglutinin, *Ulex europaeus* agglutinin I and wheat germ agglutinin, was used. Blots were incubated in a 20 μg/mL solution of each biotinylated lectin, in accordance with the manufacturer's instructions, followed by a 1 : 5000 dilution of streptavidin–ALP conjugate (Molecular Probes Inc., Eugene, OR). A BCIP/NBT phosphatase substrate kit (Kirkegaard & Perry Laboratories) was employed for colour development.

2-D polyacrylamide gel electrophoresis (2-D PAGE)

To an approximately 5×10^8 spore pellet, a 5% sodium dodecyl sulphate (SDS) solution was added, and the suspension was boiled for 5 min. Proteins were precipitated by the addition of trichloroacetic acid at a final concentration of 10%, washed once with cold 90% acetone, and dissolved in sample buffer containing 8 M urea, 2% CHAPS, 1% 1,4-dithiothreitol (DTT; Wako Pure Chemical Industries, Osaka, Japan), and 0·5% Pharmalyte (pH 3–10; Amersham Biosciences AB, Uppsala, Sweden). To remove protein aggregates, the protein solution was centrifuged and the supernatants were used as samples for 2-D PAGE. Protein concentrations were measured using a 2D-Quant kit (Amersham).

Isoelectric focusing was performed using commercially available Immobiline DryStrip gels (7 cm, pH 3–10; Amersham). Samples (about 50 μg of spore protein) were loaded by rehydration for 15 h in a solution containing 8 M urea, 2% CHAPS, 1% DTT and 0·5% Pharmalyte (pH 3–10). Isoelectric focusing was performed using an Ettan IPGphor II IEF system (Amersham) for 11 519 Vh. Voltage profiles were chosen according to the manual provided with the system. After consecutive equilibration of gels in a 50 mM Tris–HCl (pH 8·8) solution containing 6 M urea, 30% glycerol, 2% SDS and 1% DTT for 15 min, separation in the second dimension was performed using a mini-gel (10 × 10 cm) electrophoresis system with 12·5% Ready gels (Biocraft Inc., Tokyo, Japan), at 20 mA/gel.

N-terminal peptide sequencing

After 2-D PAGE, blotted proteins were stained with BODIPY FL-X, SE (Molecular Probes). Protein spots corresponding to those recognized by serum samples were excised and subjected to N-terminal sequence analysis in a Procise 494 protein sequencer (Applied Biosystems, Foster City, CA). Edman degradation was performed according to a standard program supplied by Applied Biosystems. The amino acid sequences obtained were compared to those of known proteins in the GenomeNet BLAST2 NR-AA database using the Web-accessible FASTA search program.

LC-MS/MS analysis

Spots assigned to immunogenic proteins were also excised from 2-D gels as described above. Gel pieces were subjected to in-gel digestion using the procedure published by Hellman *et al.* (14). Briefly, the protein in each gel plug was reduced with 100 mM DTT for 30 min at 50°C, alkylated with 100 mM iodoacetamide (Wako Pure Chemical) for 30 min at 37°C, and digested with 25 ng/μL trypsin (Wako Pure Chemical) overnight at 37°C. Peptides were extracted from the gel by diffusion in a 50% acetonitrile solution containing 5% formic acid, vacuum-dried, and then reconstituted in 20 μL of 0·1% trifluoroacetic acid. Extracted peptides were loaded on a high-performance liquid chromatography system (LC; MAGIC2002, Michrom BioResources Co. Ltd, Auburn, CA), eluted using an acetic acid–acetonitrile gradient, and directly injected into an LCQ ion trap mass spectrometer (MS; LCQ-Deca XP, Thermo Electron Co. Ltd, San Jose, CA) fitted with a nanoelectrospray ionization source, and operated in the positive ion mode. MS/MS spectra were acquired using data-dependent scanning, processed using the SEQUEST software program, and then searched against the NCBI NR database.

2-D immunoblot analysis

Spore proteins separated by 2-D PAGE were transferred onto PVDF membranes (ProBlot membranes, Applied Biosystems). Blotted membranes were treated with blocking buffer (SuperBlock) overnight at 4°C. Representative human sera were used as primary antibodies at a dilution of 1 : 100. ALP-labelled anti-human IgM (μ-chain specific; Kirkegaard & Perry Laboratories) was used as a secondary antibody at a dilution of 1 : 1000. Reactions with the primary and secondary antibodies were performed for 2 and 1 h, respectively, at room temperature with shaking. After extensive washing with

PBS-T, antibody binding was detected using a BCIP/NBT phosphatase substrate kit (Kirkegaard & Perry Laboratories).

As a positive control, we used a pool of 11 sera from infected rabbits at a dilution of 1 : 400. In addition, a pool of 11 sera from normal rabbits was used as a negative control at the same dilution. A culture supernatant containing MAb2 monoclonal antibody was also used at a dilution of 1 : 100 for epitope analysis. To detect these primary antibodies, ALP-labelled protein A (Zymed Laboratories) and ALP-labelled rabbit anti-mouse IgG (H + L; Zymed Laboratories) were used for rabbit IgG and the monoclonal antibody, respectively, at a dilution of 1 : 1000.

RESULTS

Measurements of anti-PT antibodies in the ELISA test

Twenty-four human serum samples used for this study revealed anti-PT IgM activities at titres between 1 : 200 and 1 : 1600 in the ELISA test. Figure 1(a) shows filaments that are strongly immunostained with IgM antibodies in the serum H-168. All samples except for H-197 had IgG antibody activity for extruded filaments below 1 : 50. However, the positive signal of H-197 was very weak, even at lower dilutions, probably because of the use of protein A/G conjugate as a secondary antibody.

Figure 1(b) shows filaments immunostained with IgG antibodies in a rabbit serum, which was used as a positive control for ELISA testing. The culture supernatant showed monoclonal antibody (MAb2) weakly immunostained filaments when it was tested by ELISA (data not shown).

Immunoreactivity of human sera on 1-D blots before and after treatment with NaOH

All 24 human serum samples generated clearly positive signals of an IgM antibody on a band migrating at 52 kDa in the 1-D immunoblot analysis without NaOH treatment (Figure 2; odd-numbered strips). Besides the positive signals for this 52-kDa band, which was MAb2-positive (strip P1), most sera also reacted with some bands below 52 kDa; however, the extent of staining was very weak, especially in bands below 28 kDa.

Using the 24 human serum samples, changes in IgM antibody binding were examined on 1-D blots before and after treatment with NaOH. When 1-D blots treated with NaOH were used for testing, the same human sera were divided into three groups based on the degree of immunostaining of the 52-kDa band: (i) 11 sera showed almost unchanged antibody binding (Figure 2; strips 6, 8 and 10); (ii) three sera showed no antibody binding (Figure 2; strips 2 and 4); and (iii) 10 sera showed significantly reduced antibody binding, resulting in the formation of faint or thin bands (Figure 2; strips 12, 14 and 16).

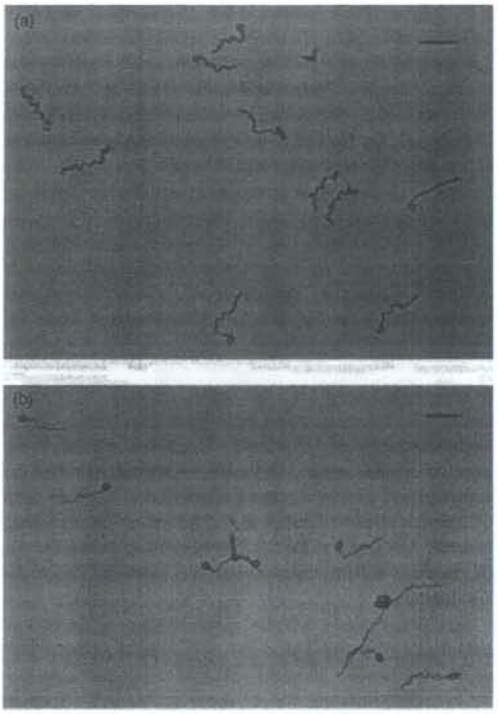

Figure 1 Immunostaining of filaments (PTs) that extruded from *E. cuniculi*-germinated spores with human IgM antibodies or rabbit IgG antibodies. (a) Positive results obtained by a 1 : 200 dilution of H-168 human serum. Note the strongly positive signals on filaments that extruded from the germinated spores. (b) Positive results obtained by a 1 : 400 dilution from a symptomatic rabbit with natural *E. cuniculi* infection. Note the positive signals on the spore wall and the filament. Bars, 10 μm.

Culture supernatant containing MAb2 strongly reacted with a band at 52 kDa (Figure 2; strip P1). A serum from a naturally infected rabbit generated IgM antibody-positive multibands between 28 and 96 kDa, and also a few bands around 19 kDa (Figure 2; strip P2). However, the rabbit serum did not show any reduced influence in IgM antibody-binding of the 52-kDa band, even after NaOH treatment (Figure 2; strip P3).

Reactivities of ConA and mouse monoclonal antibody on 1-D blots before and after NaOH treatment

When untreated 1-D blots were used, ConA was the only one of seven lectins that revealed clearly positive binding to

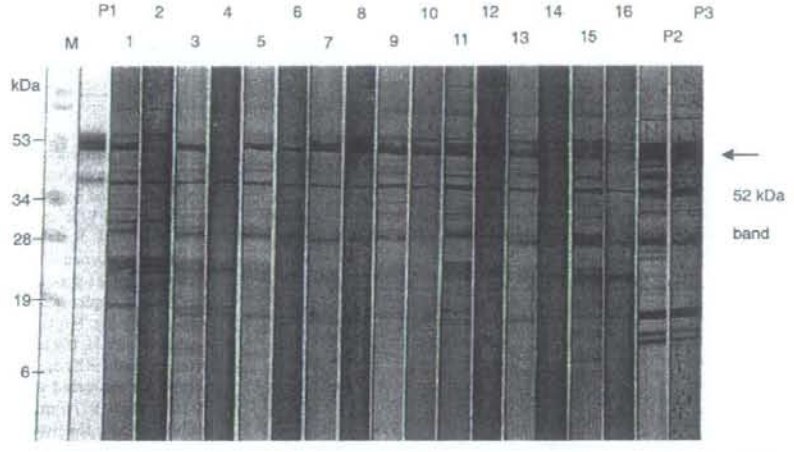

Figure 2 Binding profiles of human sera with anti-PT IgM antibody activities. *Encephalitozoon cuniculi* spore antigens were fractionated by 1-D PAGE and transferred to a PVDF membrane. Blotted strips before and after NaOH treatment were prepared and allowed to react with human serum samples followed by an anti-human IgM (μ-chain specific)-ALP conjugate. Odd-numbered strips and the P1- and P2-labelled strips were untreated, but even-numbered strips and the P3-labelled strip were treated with NaOH. Strips were reacted with each serum sample in pairs, starting from no. 1. M is a strip for markers of molecular weight (Prestained SDS-PAGE standards, Bio-Rad); P1 is a strip that reacted with MAb2; P2 and P3 are strips that reacted with a rabbit serum.

a 52-kDa band (data not shown). However, ConA did not bind at all to 1-D blots treated with NaOH (Figure 3; strip 2).

As seen on strip P1 in Figure 2 and strip 3 in Figure 3, MAb2 bound to the 52-kDa band on 1-D blots untreated with NaOH. However, MAb2-binding was not influenced by NaOH treatment at all (Figure 3; strip 4).

The location of the ConA-stained band on blotted strips was consistent with those bands detected by the monoclonal antibody (Figure 3). However, the band stained with ConA was always slightly thinner than bands detected by MAb2, suggesting that the PTP1 may be a mixture of different glycosylated and unglycosylated forms, which is likely to be slightly polydisperse in electrophoresis.

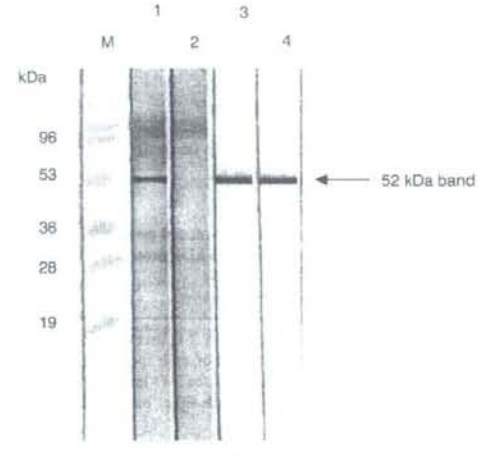

Figure 3 Binding profiles of ConA lectin and monoclonal antibody (MAb2) on 1-D blots. *Encephalitozoon cuniculi* spore antigens were fractionated by 1-D PAGE and transferred to a PVDF membrane. Blotted strips before and after NaOH treatment were prepared and allowed to react with ConA lectin and monoclonal antibody (MAb2). Odd-numbered strips were untreated, but even-numbered strips were treated with NaOH. Strip nos 1 and 2 and strip nos 3 and 4 were reacted with ConA and MAb2, respectively. M is a strip for markers of molecular weight (Prestained SDS-PAGE, Bio-Rad).

Identification of some proteins immunoreactive for human sera on 2-D blots

Eleven protein spots on a 2-D blot of *E. cuniculi* antigens were identified (Figure 4a, Table 1). Two of these spots, nos 1 and 2, had the same molecular size but different isoelectric points; these were identified as *E. cuniculi* polar tube protein (PTP), because both structures contained the peptide sequence ATALXSNAYGLTPGQQGMAQ by N-terminal amino acid sequencing. This has 90% sequence homology with the peptide sequence ATALCNAYGLTPGQQGMAQ of *E. cuniculi* genotype 1 (GenBank Accession No. AF310677) and was identified as PTP1 of *E. cuniculi* (SWISS-PRT

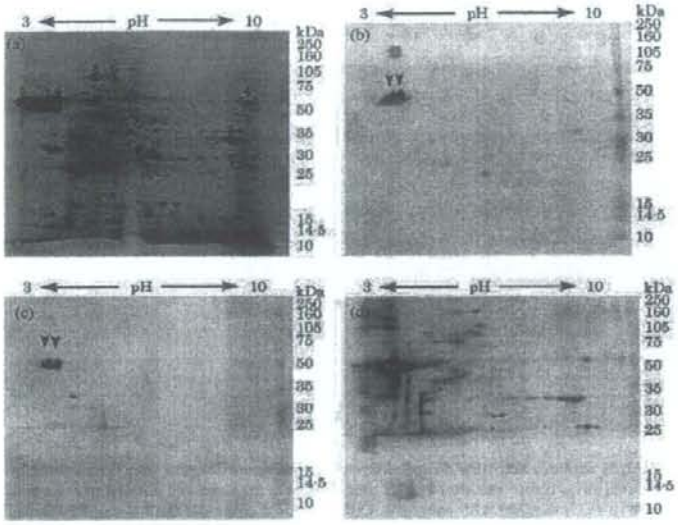

Figure 4 2-D immunoblot analysis. *Encephalitozoon cuniculi* spore antigens were fractionated by 2-D PAGE and transferred to PVDF membranes. (a) BODIPY FL-X staining: a blotted membrane is stained with the dye. Proteins extracted from 11 dye-stained spots (circles) identified by sequence analysis. (b–d) Immunostaining: blotted membranes incubated with H-168 serum (b), H-197 serum (c) and PRS (d). In the blotted membranes incubated with human sera (H-168 and H-197) containing anti-PT IgM antibodies, note the two acidic proteins (arrowheads) predominantly immunostained at 52 kDa; these proteins correspond to PTP1 spots capable of being stained with BODIPY FL-X, as seen in Figure 4(a).

Table 1 Human sera-immunoreactive spots in *E. cuniculi* identified by N-terminal amino acid sequencing (NTAAS) and LC-MS/MS

Spot no.	Protein	Accession no. (Swiss-Prot, NCBI)	Sequence coverage (%)	Method of identification	Immunoreactivity[a] with 168	197	PRS
1	Polar tube protein 1	sp. PTP1ENCCU	–	NTAAS	3+	3+	4+
2	Polar tube protein 1	sp. PTP1ENCCU	–	NTAAS	3+	3+	4+
3	Polar tube protein 2	gi: 19074274	21·3	LC-MS/MS	1+	0	2+
4	Translation elongation factor 1α	gi: 19074188	36·44	LC-MS/MS	0	0	1+
5	Spore wall protein 1	gi: 19074777	9·78	LC-MS/MS	0	0	2+
6	Hypothetical protein	gi: 19173268	42·86	LC-MS/MS	0	0	1+
7	Zinc finger protein	gi: 19074227	25·81	LC-MS/MS	1+	±	0
8	Similarity to HSP 70-related protein	gi: 19073931	29·55	LC-MS/MS	±	0	±
9	Heat shock-related 70 kDa protein	gi: 19173012	23·13	LC-MS/MS	0	0	±
10[b]	Phosphomannomutase	gi: 19173562	24·22	LC-MS/MS	0	0	0
11	6S proteasome ζ chain	gi: 19173680	17·67	LC-MS/MS	0	0	1+

[a]Immunoreactivity was examined by 2-D immunoblotting using human sera (H-168 and H-197) and rabbit sera (PRS – a pool of infected rabbit sera). The resultant intensity of each spot is scored as 0, ±, 1+, 2+, 3+ or 4+ (0, negative; ±, ambiguous; 1+ to 4+, graded positiveness). PRS was used as a positive control for 2-D immunoblot analysis.
[b]Spot no. 10 seems not to react with either human or rabbit samples, but was identified as a nonimmunogenic protein in this study.

database: PTP1ENCCU). This finding makes it clear that the 52-kDa PTP1 has more than one isoform.

H-168 and H-197 sera predominantly reacted with some spots that corresponded to PTP1 at 52 kDa (Figure 4b,c). These sera were examined as representative human sera with anti-PT IgM activities. H-168 serum also reacted with other protein spots, such as PTP2 and zinc finger protein (Figure 4b, Table 1). Such a positive reaction to zinc finger protein was not generated by a pool of sera from 11 naturally infected rabbits (PRS; Table 1).

When IgG antibodies in PRS were used as a positive control for 2-D immunoblot analysis, they reacted with spots that corresponded to PTP1, PTP2, spore wall protein 1, translation elongation factor 1α, hypothetical protein and 6S proteasome ζ chain in decreasing order of strength, as identified by LC-MS/MS analysis (Figure 4d, Table 1).

However, the rabbit IgG antibodies reacted very weakly with spots that corresponded to heat shock-related 70-kDa proteins (HSP70) and a protein with similarity to HSP70-related protein (Figure 4d, Table 1). Although a zinc finger protein and phosphomannomutase were also identified by LC-MS/MS analysis (Figure 4a), the corresponding spots were negative for rabbit IgG antibodies (Figure 4d, Table 1). IgG antibodies in a pool of 11 normal rabbit sera did not react with 2-D blots (data not shown).

Culture supernatant containing monoclonal antibody MAb2 reacted strongly with some spots of PTP1 at 52 kDa (data not shown).

DISCUSSION

All of the members of Microsporidia possess a unique, highly specialized structure, the PT, which is an organelle for invasion (15). *Encephalitozoon cuniculi* PT consists of three proteins: PTP1, PTP2 and PTP3 (16). The PTP1 is the major component of the PT and it is modified by O-linked mannose residues to which ConA binds (17).

In respect to antibodies against *Encephalitozoon* PT in immunocompetent persons, it has been reported that anti-*E. intestinalis* PT was demonstrated in 8% of Dutch blood donors and 5% of pregnant French women (18). Further study has demonstrated that carbohydrate moieties of the microsporidian PTPs are targeted by IgG in immunocompetent individuals (19). On the other hand, our study clearly indicates that the major activities of human anti-PT IgM antibodies are predominantly towards *E. cuniculi* PTP1, which is an acidic protein with an apparent molecular weight of 52 kDa in 2-D immunoblot analysis (Figure 4). Also, two chemically different antigenic sites can be found on the 52-kDa band in 1-D immunoblot analysis (Figure 2): proteinic determinant and saccharic determinant, as discussed below. In addition, although other few protein spots, such as PTP2 and zinc finger protein, weakly reacted with H-168 serum by 2-D immunoblotting, they did not appear to be effectively antigenic as compared to PTP1 (Table 1).

To analyse immunoreactivity of human anti-PT IgM with two different determinants, we used NaOH treatment capable of removing glycoepitopes from blotted antigens. Xu *et al.* (17) employed blots treated with 0·1 N NaOH at 50°C for 40 min to eliminate ConA binding to *E. hellem* PTP1. In our case, ConA binding to *E. cuniculi* PTP1 blotted onto a PVDF membrane was successfully eliminated by treating the blots with 0·05 M NaOH at 40°C for 16 h, conditions that were established by Duk *et al.* (13).

From the reactions of anti-PT IgM to the 52-kDa band after NaOH treatment, we could roughly group human sera into three categories when examined by 1-D immunoblotting: sera showing a completely negative profile; sera with a positive profile that was unmodified by NaOH treatment; and sera indicating a significantly reduced profile (Figure 2). Sera showing completely negative profiles are interpreted to have reacted with only a ConA-reactive determinant of PTP1 (Figure 2; strips 1–4). On the other hand, sera with a positive profile that was unmodified by NaOH treatment are interpreted to have reacted with only a proteinic determinant (Figure 2; strips 5–10). Interestingly, 41·3% (10/24) of the examined human serum samples belonged to the group with a significantly reduced profile. It was obvious that the significant reduction in binding after NaOH treatment resulted from the removal of the ConA-reactive determinant (Figure 2; strips 11–16), suggesting that this type of anti-PT IgM can recognize the saccharic determinant as well as the proteinic determinant. Thus, human anti-PT IgM could recognize the two determinants alone or in a pair, depending on serum used.

In a previous study (10) and also in this study, we confirmed that representative human sera with anti-PT IgM (H-168 and H-197) strongly immunostained the outer surface of filaments extruded from spores (Figure 1a). Our monoclonal antibody (MAb2) that could immunostain extruded PTs by ELISA definitely recognized the 52-kDa band by 1-D immunoblotting (Figures 2 and 3) and the PTP1 spots by 2-D immunoblotting (data not shown). Monoclonal and/or polyclonal antibodies raised to the purified *Encephalitozoon* PTP1 demonstrated reactivity with PTs by immunofluorescence and immunogold electron microscopy (17), indicating the presence of glycoconjugates in the PT. It has been proposed that the lipophilic regions of the PTPs adhere to the filament membrane and that the hydrophilic regions are orientated into the aqueous extrasporal environment (20). Thus, human anti-PT IgM-reactive determinants (constituted of saccharic and proteinic determinants and located on PTP1s) would be concentrated on the outside of the PTP coated on the surface of extruded filaments.

Regarding the specificity of anti-PT IgM in human sera, a previous study revealed that they did not react with extruded filaments from germinated *E. hellem* or *E. intestinalis*, indicating that human anti-PT IgM is species-specific (10). This finding was not in disagreement with previous findings that human sera containing anti-*E. intestinalis* PT IgG antibody activity did not immunostain filaments extruded from *E. cuniculi* or its PTP1 (19). One may consider that there is a cross-reactive relationship between microsporidian spores and fungal spores, because Microsporidia are phylogenetically related to fungi (21). However, apart from saccharic determinants, using an enzyme-linked immunosorbent assay, we found that MAb2 specific for *E. cuniculi* PTP1 did not cross-react with Laemmli sample buffer-extracted soluble antigens from the following fungi: *Candida albicans*, *Fusarium*