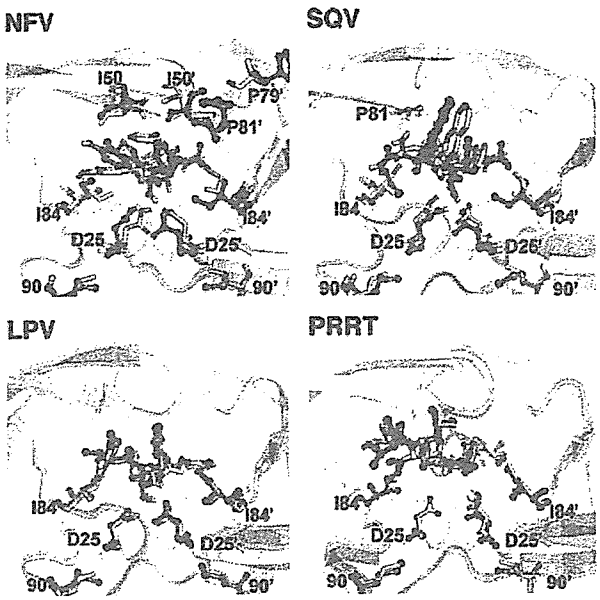

Figure 4. Structure around the active site. Each PR in the L90M model is shown in green cartoon, and each of the ligands and important residues is shown in green ball-and-stick representation. The superimposed gray cartoons and sticks represent the structure of the WT model.

regions occur. Dislocations of the side chains of D25/D25′ and rotations of I84/I84′ also occur, as in the L90M PR/NFV complex. These conformational changes have also been observed in the crystal structure of G48V/L90M PR/SQV complex (PDB code: 1FB7[20]) when compared with WT PR/SQV complex (PDB code: 1HXB[42,49]). In addition, a comparison of the structure of L90M PR in complex with SQV and the crystal structure of G48V/L90M PR/SQV complex shows that the 80's loop, where L90M causes conformational changes, is located at similar positions in the two structures (Supporting Information Figure S3). In the complex with each of these two inhibitors, L90M mutation induces common effects: dislocations of the 25th residues and rotations of the side chains of the 84th residues. In addition, L90M PR decreases the binding energy with each of NFV and SQV, which would reflect the positional shift of the inhibitors. In contrast, in the L90M PR/LPV complex, conformational changes hardly occur at the active site. Dislocations of the 25th residues and rotations of the side chains of the 84th residues hardly appear. Energetically, LPV exhibits the same binding affinity with L90M and WT PRs. PRRT also exhibits the same affinity with L90M and WT PRs, although conformations at both of the residues near I50/I50′ and P2′-P4′ subunits of PRRT are greatly changed. In the PRRT model, no rotation of the side chains of 84I/84′I occurs, while the distance between the two side chains of D25 and D25′ is changed. These results indicate that rotations of the side chains of I84/I84′ are involved in the resistance due to L90M. Consequently, we can conclude that the mechanism of resistance due to L90M is rotations of the side chains of the 84th residues due to dislocations of the side chains of the 25th residues, which are initiated by changes in the interactions between the 90th and the 25th residues. These rotations change the shapes of the active sites, and the change decreases the interactions between PR and ligands (Figure 4). There is still the question of why rotation of the 84th side chains occurs when L90M PR is bound with NFV or SQV but does not occur when L90M PR is bound with LPV and PRRT. The answer to this question is that the rotations is due to not only to dislocations of side chains of D25/D25′ but also to the geometry of the ligand. The shift in side chains of 25D/25′D occurs when L90M PR is bound with NFV, SQV, and PRRT. Focusing on the P1/P1′ subsites of those ligands, NFV and SQV each contain a dodecahydroisoquinoline ring, which is a rigid and bulky functional group, and PRRT has a ring of PRO. These rings are located near D25/D25′ and assist the dislocation of the side chains of D25/D25′ because of their rigidity. Moreover, the size of the rings is responsible for the rotation of the side chains of I84/I84′. Rotation of the side chains of the 84th residues occurs when L90M PR is bound with NFV or SQV. In contrast, rotation hardly occurs despite the side chain dislocations of D25/D25′ when L90M PR is bound with PRRT. The ring size of PRRT is smaller than those of NFV and SQV and makes no unfavorable collision with side chains of I84/I84′. Consequently, the size and flexibility of P1/P1′ subsites of the ligand are closely related to the resistance due to L90 M. We speculate that a single L90M mutation has little effect on the binding affinity with a ligand that has a linear group or a small ring at its P1/P1′ subsite.

We further investigated the interactions between the ligands and each amino acid residue of PRs by performing fragment molecular orbital (FMO) calculations.[60] In the FMO scheme, the total system is divided into fragments and calculations are carried out in parallel, which makes it possible to adopt the ab initio MO calculation for a large molecule like a protein. The single point energy of each model was calculated at the FMO–HF/6-31G level using the ABINIT-MP program[61] The model structures were constructed by the following two steps. First, the average structure was calculated on the basis of 1000 coordinates acquired during the last 500 ps of MD simulation. Next, energy minimization was executed on the average structure. One amino acid residue or one inhibitor was set as a single fragment. It was confirmed from the computational results shown in Figure 5 that each of the ligands indeed interacts with the active site residues or their neighboring residues. Notably, LPV and PRRT interact with only several active site residues. That is, the residues they interact with are quite limited compared with those with which NFV and SQV interact. Furthermore, LPV shows no significant difference between its interactions in WT and L90M PRs. LPV has highly specific interactions with D29 and D25′, whose mutations inactivate the function of the PR.[2,62] Figure 5 also indicates that NFV and SQV show noticeable loss of interaction energies with several residues in both the WT and L90M models. In particular, NFV has unfavorable contact with K45, R87, R8′, D29′, and D30′, and SQV has unfavorable contact with D25′ and D29′. In contrast, LPV shows little energetical loss in the interaction with protein residues in both models. Accordingly, we speculate that this specificity and the little energetical loss are also reasons why L90M mutation has little effect on the binding of LPV.

Last, we investigated whether simulations can provide the correct order in terms of potency of the inhibitors. In the

(60) Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *312*, 319.
(61) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2002**, *351*, 475.
(62) Ishima, R.; Torchia, D. A.; Lynch, S. M.; Gronenborn, A. M.; Louis, J. M. *J. Biol. Chem.* **2003**, *278*, 43311.

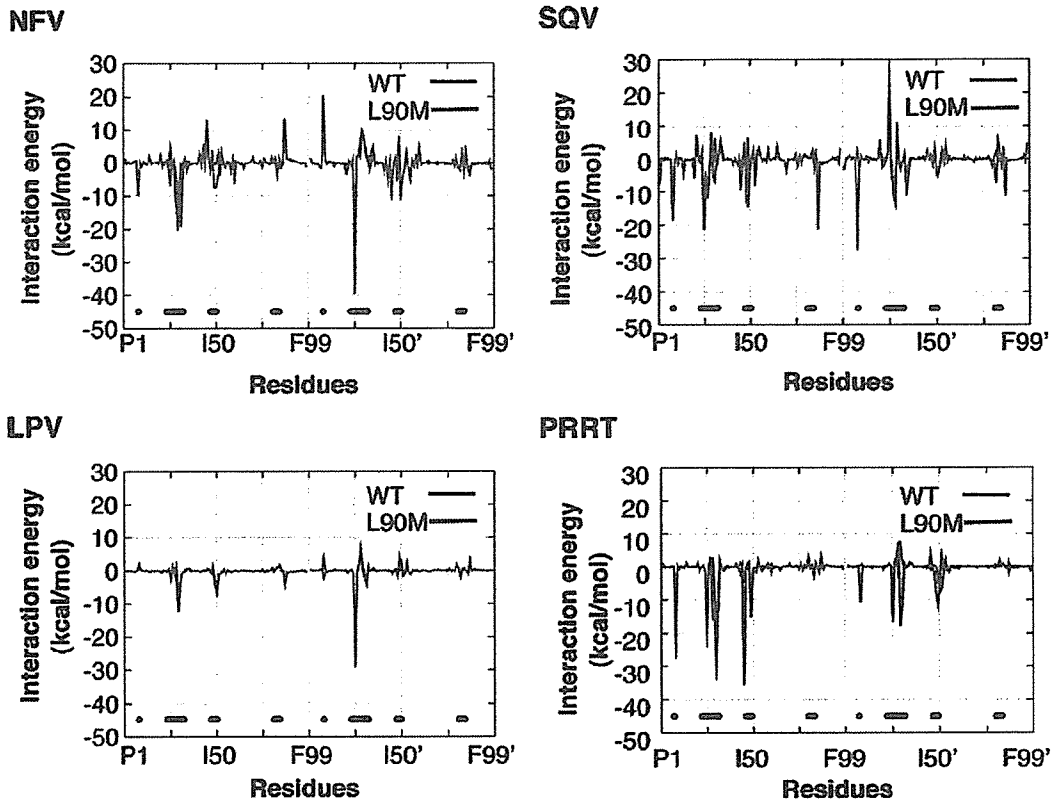

Figure 5. Interresidue interaction energies between PR and ligand calculated by FMO–HF/6-31G. Black lines below indicate the location of the active site residues (R8, L23-V32, I47-I50, P81-I84, R8′, L23′-V32′, I47′-I50′, P81′-I84′).

comparison of ΔG, the calculated inhibitory order for NFV, SQV, and LPV is not compatible with that determined by experiments[63] (Supporting Information Table S4). In contrast, when compared with experimental ΔH, the order in calculated ΔG for the inhibitors is consistent with the experimental measurements. That is because our MM/PBSA calculations do not include entropic terms. Hence, the incorporation of an entropic term will enable us to accurately predict the potency of a new drug by MD simulations.

LPV is one of the most promising drugs for AIDS treatment as shown in the present study. However, as the number of mutations increases, the efficacy of LPV decreases. For example, according to Virologic phenotypic assays, patient-derived HIV-1 confers 20-fold resistance against LPV. The PR sequence of this HIV-1 includes some drug resistant-related mutations (L10I, K20R, M36I, R41K, M46I, F53L, Q61N, L63P, A71V, T74S, V82T, N88S, L90M, I93L). Results of our additional simulation of this resistant PR with LPV have indicated a decrease in inhibitory efficacy ($\Delta G_b = -59.5$ kcal/mol, $\Delta\Delta G_b = +2.3$ kcal/mol) (Supporting Information Figure S4). These mutations decrease the number of hydrogen bonds between LPV and this clinically derived PR (Supporting Information Table S5). Furthermore, the mutations change the conformations at the flap and 80's loop regions (Supporting Information Figure S5). A design to remove the collisions at these regions will further enhance the efficacy of LPV. It should be emphasized that most of the mutated residues are located at the nonactive site of PR. Thus, to create more potent drugs, it is important to clarify the roles of the drug-resistant related nonactive site residues.

On the basis of the findings obtained in this study, we suggest the following strategy for the design of HIV-1 PR inhibitors. First, inhibitors should not contain a large ring such as a dodecahydroisoquinoline ring at P1/P1′ subsites; a linear chemical group is favorable. Second, to remove the collisions at the 80's loop and the flap region, functional groups at P2P1/P1′P2′ subsites of inhibitors should be in the same size as those of PRRT. Third, inhibitors should interact only with limited PR residues such as D25/D25′ and D29/D29′. Finally, inhibitors should not make unnecessary contact with any residues even in WT PR.

In summary, the mechanism of resistance due to the nonactive site mutation L90M has been clarified through theoretical calculations. The 90th residue of HIV-1 PR is located at the dimer interface and has no direct contact with ligand chemicals. The simulations demonstrate that the nonactive site mutation affects conformation of the binding cavity and ligand-binding affinity at the active site. The results of the present study have revealed the drug resistance mechanism of nonactive site mutation and provide a clue for designing a promising drug to reduce the drug resistance due to nonactive site mutation. Adaptive drugs,[64–67] which have the ability to inhibit several variants of a targeting enzyme, are needed in anti-HIV therapy.

(63) Yanchunas, J., Jr.; Langley, D. R.; Tao, L.; Rose, R. E.; Friborg, J.; Colonno, R. J.; Doyle, M. L. *Antimicrob. Agents Chemother.* **2005**, *49*, 3825.

(64) Velazquez-Compoy, A.; Freire, E. *J. Cell. Biochem.* **2001**, *S37*, 82.
(65) Freire, E. *Nat. Biotechnol.* **2002**, *20*, 25.
(66) Ohtaka, H.; Schon, A.; Freire, E. *Biochemistry* **2003**, *42*, 13659.
(67) Ohtaka, H.; Freire, E. *Prog. Biophys. Mol. Biol.* **2005**, *88*, 193.

Some reviews have suggested strategies for the design of adaptive inhibitors for HIV-1 PR.[64,66-69] These strategies, however, do not give sufficient consideration to the structural effects due to nonactive site mutations. The findings of this work should be useful for producing practical adaptive drugs.

Acknowledgment. This work was supported by a Health and Labor Science Research Grant for Research on HIV/AIDS from the Ministry of Health and Labor of Japan and by JSPS Research Fellowships for Young Scientists and a grant-in-aid for JSPS Fellows. A part of this work was also supported by a grant from the Japan Science and Technology Agency.

(68) King, N. M.; Prabu-Jeyabalan, M.; Nalivaika, E. A.; Schiffer, C. A. *Chem. Biol.* **2004**, *11*, 1333.
(69) Prabu-Jeyabalan, M.; Nalivaika, E. A.; Schiffer, C. A. *Structure* **2002**, *10*, 369.

Supporting Information Available: RMSD plots during MD simulations, determination of protonation states of catalytic aspartates in the SQV and/or LPV complex models, a list of hydrogen bond networks in each model, RMSF plot of main chain atoms N, Cα and C in each model, a list of differences between RMSD values (Å) of the main chain atoms in WT and L90M models, comparison of the computed L90M model and the crystal structure (1FG7), comparison of computed ΔG with experimental ΔG and ΔH, results of analyses of the MD simulations of clinically derived HIV-1 PR in complex with LPV, and a complete list of author citations with more than 10 authors. This material is available free of charge via the Internet at http://pubs.acs.org.

JA060682B

ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Feb. 2006, p. 694–701 Vol. 50, No. 2
0066-4804/06/$08.00+0 doi:10.1128/AAC.50.2.694–701.2006

Discordances between Interpretation Algorithms for Genotypic Resistance to Protease and Reverse Transcriptase Inhibitors of Human Immunodeficiency Virus Are Subtype Dependent

Joke Snoeck,[1] Rami Kantor,[2] Robert W. Shafer,[2] Kristel Van Laethem,[1] Koen Deforche,[1] Ana Patricia Carvalho,[3] Brian Wynhoven,[4] Marcelo A. Soares,[5] Patricia Cane,[6] John Clarke,[7] Candice Pillay,[8] Sunee Sirivichayakul,[9] Koya Ariyoshi,[10] Africa Holguin,[11] Hagit Rudich,[12] Rosangela Rodrigues,[13] Maria Belen Bouzas,[14] Françoise Brun-Vézinet,[15] Caroline Reid,[16] Pedro Cahn,[14] Luis Fernando Brigido,[13] Zehava Grossman,[12] Vincent Soriano,[11] Wataru Sugiura,[10] Praphan Phanuphak,[9] Lynn Morris,[8] Jonathan Weber,[7] Deenan Pillay,[17] Amilcar Tanuri,[5] Richard P. Harrigan,[4] Ricardo Camacho,[3] Jonathan M. Schapiro,[18] David Katzenstein,[2] and Anne-Mieke Vandamme[1]*

Rega Institute for Medical Research, Leuven, Belgium[1]; Stanford University, Stanford, California[2]; Hospital Egas Moniz, Lisbon, Portugal[3]; BC Center for Excellence in HIV/AIDS, Vancouver, Canada[4]; Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil[5]; Health Protection Agency, Salisbury, United Kingdom[6]; Wright Fleming Institute, London, United Kingdom[7]; National Institute for Communicable Diseases, Johannesburg, South Africa[8]; Chulalongkorn University, Bangkok, Thailand[9]; National Institute of Infectious Diseases, Tokyo, Japan[10]; Hospital Carlos III, Madrid, Spain[11]; Ministry of Health, Tel Aviv, Israel[12]; Instituto Adolfo Lutz, Sao Paulo, Brazil[13]; Fundación Huesped, Buenos Aires, Argentina[14]; Laboratory of Virology, Bichat, Claude Bernard Hospital, Paris, France[15]; Bayer Health Care-Diagnostics, Toronto, Canada[16]; HPA Antiviral Susceptibility Reference Unit, Birmingham, United Kingdom[17]; and National Hemophilia Center, Sheba Medical Center, Tel Aviv, Israel[18]

Received 5 May 2005/Returned for modification 25 August 2005/Accepted 26 November 2005

The major limitation of drug resistance genotyping for human immunodeficiency virus remains the interpretation of the results. We evaluated the concordance in predicting therapy response between four different interpretation algorithms (Rega 6.3, HIVDB-08/04, ANRS [07/04], and VGI 8.0). Sequences were gathered through a worldwide effort to establish a database of non-B subtype sequences, and demographic and clinical information about the patients was gathered. The most concordant results were found for nonnucleoside reverse transcriptase (RT) inhibitors (93%), followed by protease inhibitors (84%) and nucleoside RT inhibitor (NRTIs) (76%). For therapy-naive patients, for nelfinavir, especially for subtypes C and G, the discordances were driven mainly by the protease (PRO) mutational pattern 82I/V + 63P + 36I/V for subtype C and 82I + 63P + 36I + 20I for subtype G. Subtype F displayed more discordances for ritonavir in untreated patients due to the combined presence of PRO 20R and 10I/V. In therapy-experienced patients, subtype G displayed a lot of discordances for saquinavir and indinavir due to mutational patterns involving PRO 90 M and 82I. Subtype F had more discordance for nelfinavir attributable to the presence of PRO 88S and 82A + 54V. For the NRTIs lamivudine and emtricitabine, CRF01_AE had more discordances than subtype B due to the presence of RT mutational patterns 65R + 115 M and 118I + 215Y, respectively. Overall, the different algorithms agreed well on the level of resistance scored, but some of the discordances could be attributed to specific (subtype-dependent) combinations of mutations. It is not yet known whether therapy response is subtype dependent, but the advice given to clinicians based on a genotypic interpretation algorithm differs according to the subtype.

Genotyping for the assessment of anti-human immunodeficiency virus (HIV) drug resistance is often used in the management of individual patient therapy. Currently, it is recommended in European as well as American guidelines (17, 38). In several retrospective and prospective studies, genotyping proved beneficial in optimizing treatment for individual patients (5, 10, 16, 23, 25, 31, 37).

* Corresponding author. Mailing address: Rega Institute for Medical Research, Minderbroedersstraat 10, 3000 Leuven, Belgium. Phone: 32 16332160. Fax: 32 16332131. E-mail: annemie.vandamme@uz.kuleuven.ac.be.

Although genotyping is commonly used, there are still many uncertainties with respect to the value of genotype in the assignment of a new regimen. The current genotypic assays are not always able to report all drug resistance mutations among non-B subtypes (11, 18, 19, 24). Regardless of subtype, genotyping is not sensitive to mutations that are present as a minor variant in the population (22, 40). Genotyping results also differ depending on the laboratory where they are performed. Quality control studies indicate that mutations, even present as a pure variant, are often underestimated (32).

However, separate from the quality and sensitivity issues, the interpretation of genotypic results is still not standardized.

Several interpretation algorithms have been designed to aid in this, but they may differ in the prediction of therapy response and/or drug susceptibility. Studies were performed mainly on subtype B viruses, and even within this subtype, differences have been detected (6, 21, 29, 34, 35, 36).

Non-B subtypes are a challenge for these systems, since algorithms for these subtypes were designed using genotype, phenotype, and therapy response information that was largely derived from experience with subtype B. Recent analyses suggest that non-B viruses can develop specific mutations that differ from those identified in subtype B under the same treatment pressure (1, 20). For example, in CRF01_AE but not in subtype B viruses, V75M seems to be significantly associated with stavudine treatment (2) and, in subtype C but not in subtype B, V106M is a signature substitution of patients treated with efavirenz (4). There is a continuing controversy about the impact of secondary protease mutations (positions 36, 71, 77, etc.) which evolve in subtype B following protease exposure and are relatively frequent in untreated patients with non-B subtypes. It has been suggested that some of these can affect the susceptibility to certain protease inhibitor (PI) therapies in B and non-B subtypes (14, 28).

Although some short-term studies suggest little difference in therapy response in patients carrying non-B subtypes from that of patients infected with subtype B (12), other studies showed a significant difference in responses to treatment for different subtypes (8, 13). However, current studies have included a limited number of subjects. Potential differences can be due to differences in drug resistance. It is therefore important to know how the current drug resistance interpretation systems perform on different subtypes, and first of all, we need to know what the subtype-dependant discrepancies between the systems are.

Comparisons between these interpretation systems have already been made for subtype B strains; however, the subtype dependency of resistance assessment by these interpretations systems has not yet been determined (6, 21, 29, 34, 35, 36). In this study, we investigated four frequently used interpretation systems across a large number of non-B sequences to determine whether discordance between the systems was dependent on the viral subtype.

MATERIALS AND METHODS

Sequences. Sequences of HIV-1 protease (positions 1 to 99) and reverse transcriptase (RT) (positions 1 to 240) were collected from the published literature and from 14 laboratories in 12 countries through the non-B workgroup, a worldwide effort to establish a database of non-subtype B sequences (20). Three separate analyses were performed based on the treatment history of the patient at the time of sequencing: PI analysis, nucleoside RT inhibitor (NRTI) analysis, and nonnucleoside RT inhibitor (NNRTI) analysis. A sequence was included in the respective analysis either if the patient was reported to have had no previous exposure to a drug in that class or if the patient was being treated with a drug in that class at the time of sequencing, thus separating the analyses according to drug class exposure. In this way, sequences from patients that had drug exposure from a particular class in the past but were not at the time of sequencing taking a drug from that class were excluded. The treatment data gathered for this database were therapy history, with start and stop dates for a treatment, the regimens in the therapy, and the doses of the separate antivirals. Sequences were excluded when there was no therapy history.

Subtyping. Subtyping was performed by phylogenetic analysis using the subtyping tool developed by de Oliveira et al. separately for protease and reverse transcriptase sequences (7). Briefly, sequences are first analyzed using pure subtypes as a reference; in a second step, known circulating recombinant forms are added to the alignment. To detect recombination, bootscanning was performed using a sliding window of 400 nucleotides that was advanced 20 nucleotides at a time. Recombinants were included only if they were CRF01_AE or CRF02_AG since we had sufficient data for only these two circulating recombinant forms.

Algorithms. Four publicly available algorithms were applied on each of the sequences: Agence Nationale de Recherche sur le SIDA (ANRS) July 2004 (http://www.sante.gouv.fr/htm/actu/36_vih_2.htm) (25), HIV RT and Protease Sequence Database (HIVDB) August 2004 (http://hivdb.stanford.edu) (33), Rega Institute (Rega) version 6.3 (http://www.kuleuven.be/rega/cev/pdf/ResistanceAlgorithm6_3.pdf) (39), and Bayer Health Care-Diagnostics (VGI) version 8 (30) (formerly Visible Genetics).

Mutations considered. In all statistical analyses (see below), we scored all mutations that are included in one of the algorithms we used in the analyses: 18 NRTI resistance positions, i.e., 41, 44, 62, 65, 67, 69, 70, 74, 75, 77, 115, 116, 118, 151, 184, 210, 215, and 219; 16 NNRTI resistance positions, i.e., 98, 100, 101, 103, 106, 108, 179, 181, 188, 190, 225, 227, 230, 234, 236, and 318; and 23 PI resistance positions, i.e., 10, 20, 24, 30, 32, 33, 36, 46, 47, 48, 50, 53, 54, 60, 63, 71, 73, 77, 82, 84, 88, 90, and 93. For most positions, more than one mutant amino acid can be scored. All mixtures at resistance positions were scored as mutants.

Scoring of discordances—statistical analyses and data mining. The algorithm specification interface at the web site for the Stanford HIV drug resistance database (http://hivdb.stanford.edu) was used to apply the interpretation algorithms to each sequence (3). We assigned three levels of resistance: susceptible (S), intermediate (I), and resistant (R). For HIVDB, which assigns five levels of resistance, we obtained three by pooling the two highest and two lowest categories.

Interpretations were considered concordant if each of the algorithms assigned the same level of resistance to a sequence for a particular drug. We considered the algorithms to be fully discordant if one of them scored the sequence S for a particular drug, and another one scored it as R. Interpretations were considered partially discordant when, among the scores of the different systems, both S and I or both R and I were found for the same drug. The numbers of fully discordant (counted as 1) and partially discordant (counted as 0.5) strains were added to compute the proportion of discordant strains.

Statistical analyses were performed to see whether the number of discordances were drug and subtype dependent. We performed a one-way analysis of variance (ANOVA) with Tukey's confidence intervals to check for differences between different drugs and different subtypes. Differences between only subtype B and each of the other subtypes have been analyzed in this study.

The data mining program Weka, version 3.4.4 (http://www.cs.waikato.ac.nz/~ml/weka/), was used to identify mutational patterns that were responsible for the observed discordances, thereby also identifying the algorithms that caused the discordances. We used this tool to build binary decision trees with which it tries to predict all observed discordances. To evaluate the predictive power of the decision trees, we performed a 10-fold cross-validation. In this method, the data set is split 10-fold and the predictive performance for every subset is evaluated for a decision tree trained on the other subsets.

We built a model for each drug in which we found a statistically significant effect of subtype on discordance. We included all subtypes in the model and tried to predict discordances (three levels, concordant, discordant, and partially discordant). For each leaf in the resulting tree that predicted discordance, we calculated the subtype distribution. Fisher exact tests were performed to analyze whether a rule in the decision tree explained significantly more discordances for a particular subtype.

RESULTS

Subtype distribution. We obtained protease and/or reverse transcriptase sequences from 5,030 patients. The subtype distribution for each analysis (PI, NRTI, or NNRTI) is shown in Table 1. In total, we obtained 6,916 (3,926 from naive and 2,990 from treated patients) sequences for PI analyses, 5,689 (2,331 naive and 3,358 treated) for NRTI analyses, and 5,557 (4,208 naive and 1,349 treated) for NNRTI analyses. Twelve protease and five RT sequences were filtered out due to suspected recombination or were untypable. The majority of the sequences were of a non-B subtype except for the PI-treated and NRTI-treated class, where the prevalence of subtype B was 82% and 66%, respectively. Subtypes H, J, and K were excluded because of a limited number of sequences.

TABLE 1. Subtype distribution for sequences in the analysis groups PI, NRTI, and NNRTI

Subtype	No. of sequences for[a]:					
	PI		NRTI		NNRTI	
	Naive	Treated	Naive	Treated	Naive	Treated
A	363	35	318	105	217	206
B	1,661	2,467	632	2,224	2,139	585
C	672	201	644	339	805	178
D	260	37	201	89	159	131
F	126	80	79	107	140	46
G	128	87	63	158	144	77
CRF01_AE	207	36	132	251	291	92
CRF02_AG	509	47	262	85	313	34

[a] A sequence was included in the analysis if there was no previous exposure to a drug in that class or the patient was being treated with a drug in that class at the time of sequencing.

Discordances. Overall, the different interpretation systems agreed well on the level of resistance. Eighty-four percent of the sequences had concordant results for PIs. In only 6% of the cases, the algorithms gave full discordant results; most of the observed differences were due to partial discordances (10%). For NRTIs, 76% of the sequences gave concordant results and 8% were fully discordant. The most concordant results, 93%, were found for NNRTI. Only 1% of the sequences caused full discordances. The results for each drug are shown in Fig. 1.

The concordance was significantly higher for therapy-naive patients than for treatment-experienced patients ($P < 0.0001$) for all drug classes.

Protease inhibitor analysis. The number of discordances seemed to be drug and subtype dependent for therapy-naive patients as well as treated patients (Tables 2 and 3).

In therapy-naive patients, results for nelfinavir were discordant in 1.8% of the sequences, while for lopinavir, this was 0.3% and for tipranavir, this was 0%. When considering the results for a single drug, the proportion of sequences displaying full or partial discordances was subtype dependent. Concerning specific subtypes in therapy-naive patients, discordances were observed for ritonavir (subtype F, $P < 0.01$) and nelfinavir (subtypes G and C) (Table 2).

In treated patients, the results were different. The highest level of discordance was obtained for amprenavir (50%), whereas 36% of the sequences were scored as discordant for lopinavir and 14% for nelfinavir. Tipranavir gave still the least discordant results; only 2% of the sequences were causing discordances between algorithms. Compared to subtype B, more discordances were observed for nelfinavir in subtype F and for indinavir and saquinavir in subtype G ($P < 0.01$), while less discordances were observed for amprenavir in subtypes C and D and for atazanavir in subtype C ($P < 0.01$) (Table 3).

Nonnucleoside reverse transcriptase inhibitor analysis. For therapy-naive patients, no differences could be found between drugs, while for treated patients, efavirenz scored the most discordances (11%), followed by delavirdine and nevirapine (5%).

The proportion of sequences displaying full or partial discordances was subtype dependent in this drug class except for delavirdine and nevirapine in naive patients. But no specific subtypes were found that had differences in the resistance interpretation compared to subtype B.

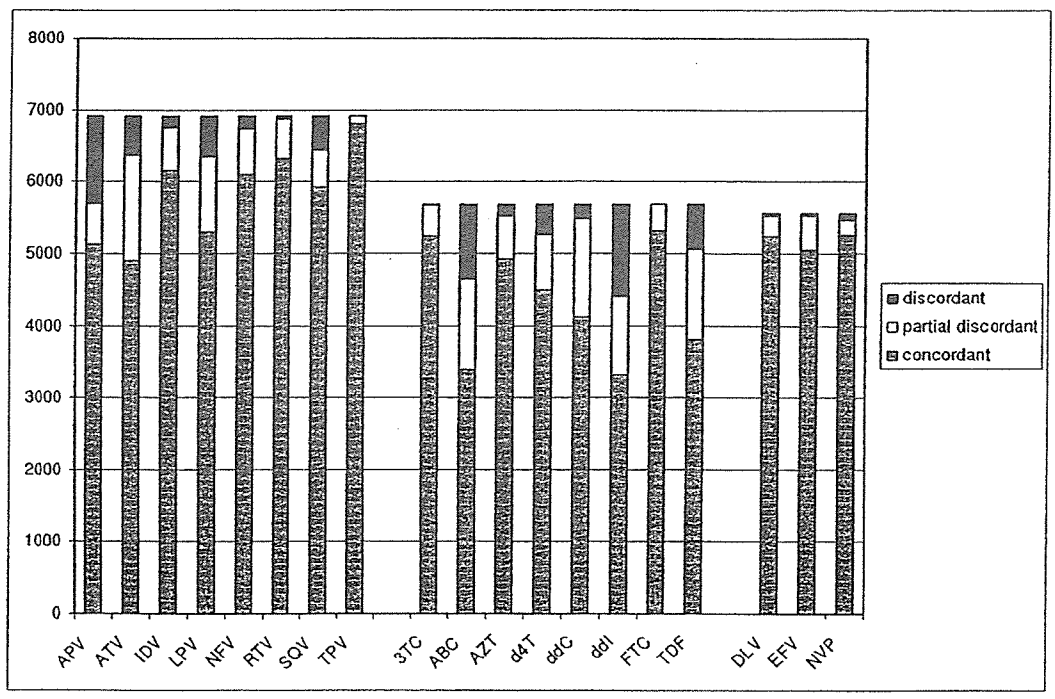

FIG. 1. Graphic representation of the number of discordant sequences per drug class. Gray bars represent the number of sequences for which concordant predictions were made by the four algorithms, white bars represent the number of sequences with partial discordance, and black bars represent sequences with discordant predictions.

TABLE 2. Interalgorithm discordances between genotypic drug resistance interpretation for sequences obtained from therapy-naive patients infected with HIV-1

Drug	Discordances (%)[a]	*P* value for subtype dependency[b]	Subtypes[c]
Protease inhibitors			
Nelfinavir	1.8	<0.01	G and C more than B
Atazanavir	1.1	<0.01	
Ritonavir	1.1	<0.01	F more than B
Amprenavir	0.6	NS	
Indinavir	0.5	NS	
Saquinavir	0.4	NS	
Lopinavir	0.3	NS	
Tipranavir	0	NS	
Nonnucleoside reverse transcriptase inhibitors			
Delavirdine	5	NS	
Nevirapine	5	NS	
Efavirenz	5	<0.01	
Nucleoside reverse transcriptase inhibitors			
Zidovudine	1.6	<0.01	
Zalcitabine	1.2	<0.01	
Stavudine	1	<0.01	C less than B
Abacavir	0.7	NS	
Didanosine	0.6	NS	
Tenofovir	0.4	NS	
Lamivudine	0.2	NS	
Emtricitabine	0.1	NS	

[a] Percentage of sequences that had discordant results between genotypic interpretation algorithms.
[b] One-way ANOVA was used to evaluate whether the number of discordances was subtype dependent (*P* of <0.05 was considered significant). NS, not significant.
[c] If the number of discordances was subtype dependent, Tukey's confidence intervals were used for a pairwise analysis to look for subtypes that caused significantly fewer or more discordances than subtype B. Although the percentage of discordances for some drugs was significantly subtype dependent, this did not always relate to a specific subtype that displayed significantly more or less discordances then subtype B.

TABLE 3. Interalgorithm discordances between genotypic drug resistance interpretation for sequences obtained from therapy-experienced patients infected with HIV-1

Drug	Discordances (%)[a]	*P* value for subtype dependency[b]	Subtypes[c]
Protease inhibitors			
Amprenavir	50	<0.01	C and D less than B
Atazanavir	42	<0.01	C less than B
Lopinavir	36	<0.01	
Saquinavir	24	<0.01	G more than B
Indinavir	15	<0.01	G more than B
Nelfinavir	14	<0.01	F more than B
Ritonavir	9	<0.01	
Tipranavir	2	NS	
Nonnucleoside reverse transcriptase inhibitors			
Efavirenz	11	<0.01	
Delavirdine	5	<0.01	
Nevirapine	5	<0.01	
Nucleoside reverse transcriptase inhibitors			
Didanosine	54	<0.01	C and D less than B
Abacavir	49	<0.01	
Tenofovir	37	<0.01	G, A, C, and D less than B
Zalcitabine	26	<0.01	C less than B
Stavudine	23	<0.01	
Zidovudine	13	<0.01	
Lamivudine	7	<0.01	CRF01_AE more than B
Emtricitabine	5	<0.01	CRF01_AE more than B

[a] Percentage of sequences that had discordant results between genotypic interpretation algorithms.
[b] One-way ANOVA was used to evaluate whether the number of discordances was subtype dependent (*P* value of <0.05 was considered significant). NS, not significant.
[c] If the number of discordances was subtype dependent, Tukey's confidence intervals were used for a pairwise analysis to look for subtypes that caused significantly fewer or more discordances than subtype B. Although the percentage of discordances for some drugs was significantly subtype dependent, this did not always relate to a specific subtype that displayed significantly more or fewer discordances then subtype B.

Nucleoside reverse transcriptase inhibitor analysis. In 1.6% of the sequences, zidovudine (AZT) was responsible for most of the discordances in therapy-naive patients; didanosine (ddI) was responsible for most of the discordances in treated patients (54%). The difference between drugs in this class was significant for both therapy-naive (Table 2) and therapy-experienced (Table 3) patients.

For zidovudine, zalcitabine, and stavudine in the naive population, the number of discordances was associated with subtype ($P < 0.01$). For only stavudine, subtype C was found to display less discordances than subtype B.

The number of discordances was significantly associated with subtype for all drugs in therapy-experienced patients ($P < 0.01$). For lamivudine and emtricitabine, CRF01_AE seemed to display significantly more discordances than subtype B. Subtypes C and D had fewer discordant interpretations for didanosine, and subtype C had also fewer for zalcitabine. For tenofovir, a lot of non-B subtypes had fewer discordant results than subtype B. This was the case for subtypes A, C, D, and G.

Mutational features of the subtype dependency. The results have been summarized in Table 4.

In therapy-naive patients among non-B subtype viruses, subtypes C and G showed partial discordances with respect to saquinavir susceptibility.

For subtype C, the most frequent pattern that caused partial discordances was a combination of protease (PRO) 82V/I + 63P + 36V/I. This pattern significantly explained more partial discordances for subtype C than for subtype B ($P < 0.0001$). This seemed due to the HIVDB interpretation algorithm. All subtype C sequences displaying this pattern also had the PRO 93L mutation. This mutation is taken into account for only nelfinavir by the HIVDB algorithm, which scores this pattern as intermediate, while all other algorithms score these sequences susceptible.

Two rules were discovered in the tree for subtype G that explained significantly more discordances than subtype B. One was a rule very similar to that for subtype C, PRO 82I + 63P + 36I ($P = 0.04$), and the other rule was PRO 82I + 63mt (any mutation) + 20I ($P = 0.01$). In practice, these rules cover the same sequences, as all subtype G sequences with the first pattern also harbor a mutation at position PRO 20 and all

TABLE 4. Mutations at least partially responsible for the subtype dependent behavior of genotypic interpretation algorithms for a drugs and algorithms responsible for the observed discordances

Drug[a]	Subtype	Mutation patterns (score)[b]	Algorithm responsible[c]
Naive population			
Nelfinavir	C	82I/V + 63P + 36I/V (SISS)	HIVDB (all sequences also 93L, taken into account by only HIVDB)
	G	82I + 63P + 36I (SISS) and 82I + 63mt + 20I (SISS)	HIVDB (high weight for 82I)
Ritonavir	F	20R + 10V/I (nrSIS)	Rega (all sequences also 36I, three secondary PI mutations scored as I by only Rega)
Treated population			
Saquinavir	G	90M + 82I (SRIR)	ANRS (does not score this as resistant)
Indinavir	G	90M + 82I + 54V (RRSI) and	HIVDB and ANRS (all sequences also 36I, pattern scored as R by HIVDB and ANRS)
		90M + 82I + 71T + 20I (RISI)	Rega (L90M not scored as R)
Nelfinavir	F	88S (RRSI) and	Rega (scores this as S)
		82A + 54V (IRRR)	ANRS (all sequences also 36I, not scored as R by ANRS)
Lamivudine	CRF01_AE	65R + 151M (IRRI)	ANRS and VGI (do not have a rule for the presence of both)
Emtricitabine	CRF01_AE	118I + 215Y (SSInr)	Rega (all sequences also 41L and 67N, 67N scored only by Rega)

[a] Only drugs for which the subtype dependence was proven and for which we found subtypes that displayed significantly more or fewer discordances than subtype B are shown. As explained in the text, the decision trees for the drugs where subtype B displayed more discordances were often too complex. Those are not included in this table.

[b] Positions at which mutations are responsible for discordances as revealed by data mining analysis. The order of the scores is shown alphabetically according to the algorithm name (ANRS, HIVDB, Rega, and VGI). Only the scoring patterns that accounted for most of the discordances (>85%) are shown. nr, no rule available for the drug.

[c] Algorithm(s) responsible for the observed discordances. Some information is provided in parentheses as to why these algorithms cause a discordance.

sequences with the second pattern also harbor a mutation at position PRO 36. Again, these discordances were due to the HIVDB algorithm, which is the only one that takes into account mutations at position PRO 20 and gives a rather high weight for the PRO 82I mutation for nelfinavir.

For ritonavir, subtype F caused more discordances than subtype B. We found a rule, PRO 20R + 10V/I, in the decision tree explaining significantly more subtype F partial discordances than those observed in subtype B. An example of the Weka decision tree with subsequent statistical analyses is shown in Fig. 2. Those subtype F sequences all had the PRO 36I mutation and thus harbored three secondary PI mutations. The Rega algorithm scores this as intermediate for ritonavir, while all other algorithms score this as susceptible.

For NRTIs, subtype B gave a lot of discordant interpretations. The rule predictive for this discordance in the decision tree was any mutation at RT 215, but this was not significant ($P = 0.07$). When examining the data, we found that the dis-

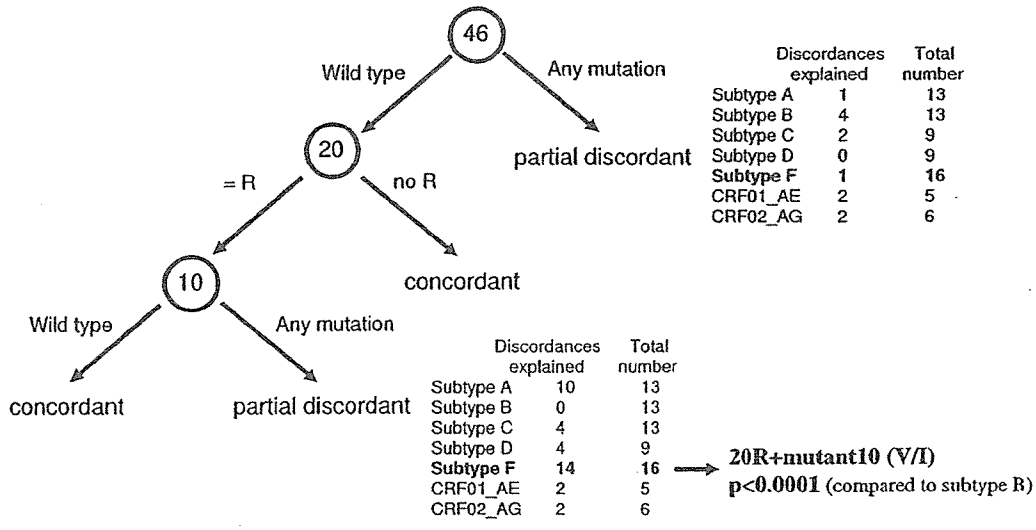

FIG. 2. Representation of the Weka decision tree for ritonavir in our untreated population. In the circles, the amino acid position is represented and, along the arrows, the mutation present is shown. R, arginine. We found that subtype F displayed more discordance. In the Weka decision tree, two rules were found, i.e., (i) any mutations at position PRO 46 and (ii) 20R + mutant 10. We calculated the number of discordances found that were explained by these rules and compared these numbers for subtype F and subtype B. Only the second rule explained significantly more discordances for subtype F than for subtype B.

cordances for stavudine were due to the ANRS system, which scores the presence of a mutation at 215 by itself as intermediately susceptible; all the other systems score this as susceptible. We found that subtype B more often had a mutation at this position than did subtype C, although this was not significant.

For the PI saquinavir in therapy-experienced patients, the full discordances observed in subtype G sequences could be attributed to mutations PRO 90 M + 82I. This was due to the ANRS interpretation system, which does not score this as resistant (as HIVDB and VGI did) if PRO 82I is present. Only PRO 82A is taken into account by ANRS.

For indinavir, subtype G also displayed more discordances than subtype B, apparently due to PRO 90 M + 82I + 54V, which was scored as resistant by HIVDB and ANRS because all these samples also had the PRO 36I mutation. Another rule predictive for discordance was PRO 90 M + 82I + 71T + 20I. The Rega system scores this pattern as susceptible, since the PRO 90 M mutation by itself is not scored as resistant by this algorithm.

Subtype F causes more discordances for nelfinavir in treated patients. The PRO 88S mutation was partially responsible for these discordances. The Rega algorithm considers these isolates to be susceptible, while the score from other algorithms was at least intermediate resistant. The partial discordances for subtype F are explained by PRO 82A + 54V. All these sequences had also PRO 36I, which is not considered resistant by ANRS relative to the other algorithms.

Subtype B displayed a lot of discordances for amprenavir. In fact, the decision tree incorporated subtype in this model. The resulting rule was PRO 90 M + 54V + 20R + 82A. All these sequences had an additional PRO 36I mutation, which is not included in the amprenavir rules of the Rega algorithm. This mutation pattern scored as intermediate for this system, while for the other algorithms, the additional PRO 36I mutation is responsible for the resistant score.

For atazanavir, subtype B caused a lot of discordances. The decision tree was very complex, and no clear rule had a high coverage and was predictive for the observed discordances in all subtypes. The atazanavir rules incorporate a number of mutations also observed for other PIs. Patients harboring a subtype B virus are probably treated with protease inhibitors more often and for a longer time, since subtype B has dominated since the beginning of the epidemic in countries where treatment was available and subsequently has been subject to drug selective pressure earlier. In these sequences, the large background of PI resistance mutations probably causes the discordances observed for atazanavir.

For lamivudine and emtricitabine (FTC), CRF01_AE scored more discordances than subtype B. For lamivudine resistance interpretation, this was caused by RT 65R + 151 M ($P < 0.05$). ANRS scores the presence of both mutations separately as intermediate but does not provide a rule for the presence of both of them, while the Rega algorithm for example scores this combination as resistant.

For emtricitabine, no clear rules were found in the tree, although it seemed that RT 41L + 67N + 118I + 215Y caused most of the partial discordances observed for CRF01_AE. The Rega algorithm is the only one that scores the RT 67N mutation for FTC. VGI does not provide rules for FTC.

For didanosine, tenofovir, and zalcitabine, subtype B had a lot more discordant interpretations than a number of non-B subtypes. The decision trees were very complex and also for these drugs, no clear rules could be deduced.

DISCUSSION

HIV genotypic information has led to an improved understanding of mutations in *pol*, which is associated with virological failure. Although resistance genotyping still has some limitations, it is often used to guide therapy start or change. One of the major problems is the interpretation of genotypic results. The knowledge on which such interpretation systems are built is based mainly on subtype B data. Considering the possible differences in therapy response in other subtypes, it would be interesting to verify whether our genotypic interpretation systems are equally valid for all subtypes. A first approach is to map discrepancies in drug resistance interpretation algorithms between subtypes and to identify which mutational patterns are responsible for such discrepancies. Such patterns can then further be investigated by, for example, in vitro mutagenesis and measuring the associated phenotype, taking into account that virus replication under drug selective pressure not only is a matter of protease and RT mutations but also is determined by the whole viral genome.

In this study, performed on sequences obtained from 5,030 patients, we investigated subtype-dependant discrepancies between four commonly used interpretation systems (Rega 6.3, HIVDB-08/04, ANRS [07/04], and VGI 8.0). The versions analyzed were the ones available to us at the time of analysis. In the meantime, updates have become available for all of these systems. None of these systems include subtype-dependant rules.

We did find drug- and subtype-dependent differences in the drug susceptibility/therapy response predictions of commonly used interpretation algorithms. We also identified mutational patterns that seemed to be partially responsible for the observed discordances.

Concordance was the lowest in the interpretation of therapy-experienced sequences, which means that it is less clear which mutations are really important for resistance development. This may explain some of the differences seen between algorithms in predicting treatment outcome (6). For lopinavir especially, the pathway towards resistance is unclear, which explains the high number of discordant results between the interpretation systems found in therapy-experienced patients (26, 27).

Our analyses revealed that the proportion of discordances between commonly used algorithms is subtype dependent for many drugs, in naive as well as in therapy-experienced patients. Concordance was higher in naive patients. However, non-B subtype sequences and subtype B sequences overall had equal numbers of resistance mutations. Both groups had mostly "wild-type" sequences. Therefore, the higher number of concordances is probably due to a larger agreement on what is a wild-type sequence.

In naive patients, discordances were found for nelfinavir (subtypes C and G). Incidentally, it is known that the pathway towards resistance for nelfinavir differs for subtypes C and G from that for subtype B. The PRO D30N mutation is not the

preferred one as in subtype B; it seems that, rather, the PRO L90M is selected (15) (P. Gomes, I. Diogo, M. F. Gonves, et al., Abstr. 9th Conf. Retrovir. Opportunistic Infect., abstr. 46, 2002). We found mutational patterns that partially explained these discordances. Those were mostly due to combinations of secondary PI mutations, which are often present as a polymorphism in non-B subtypes. Some algorithms include these mutations in their rules, while others do not. The PRO 93L mutation for example, is included by only HIVDB and not by the other systems. This mutation was present in all subtype C sequences with the pattern PRO 82I/V + 63P + 36I/V. Similarly for subtype G, the PRO 20I mutation is incorporated by only HIVDB.

For subtype F and ritonavir, the pattern PRO 20R + 10V/I also included the PRO 36I mutation. Three secondary PI mutations are scored as intermediate by only the Rega Algorithm.

For NNRTIs, we did not find any subtype-dependent discordances in resistance scoring, although some differences in resistance development have already been reported for subtype C under efavirenz treatment (2).

For NRTIs, only in naive patients did we find that the proportion of discordances is subtype dependent for stavudine. Subtype C had significantly less discordances than subtype B due to a mutation on RT 215 that occurred more frequently in subtype B sequences.

For PI resistance in treated patients, a lot of discordances are observed for subtype G in predicting resistance for saquinavir and indinavir and in subtype F for nelfinavir resistance prediction. The patterns observed here are related to a single algorithm that scores this differently. Differences often occur due to the presence of the PRO 36I mutation, which is present as a polymorphism in non-B subtypes. This mutation often triggers the switch to score an isolate as intermediate, while other systems do not take into account the substitution and consider the isolate to be susceptible. Apparently, there is no agreement on the role of some of these polymorphic resistance mutations in PI resistance.

For amprenavir and atazanavir, subtype B displayed a lot of discordances for treated patients. The decision trees for these drugs were very complex. The tree for amprenavir included subtype as a node, so a rule, PRO 90 M + 54V + 20R + 82A, could be deduced. For atazanavir, no clear rule was found. These two drugs are only recently being used in clinical practice, and the pathway towards resistance is not fully understood yet. The presence of a number of PI mutations, instead of some clear rules, is mostly used in the algorithms.

For lamivudine and emtricitabine in treated patients, CRF01_AE scored more discordances than subtype B. Although resistance for both drugs are predicted by the same rules in the algorithms, different mutation patterns are found in the decision trees. For lamivudine resistance interpretation, this was caused by RT 65R + 151 M. For emtricitabine, this was RT 41L + 67N + 118I + 215Y (although not statistically supported).

Tipranavir has a low number of discordances for naive patients as well as treated patients. This is mainly due to the limited amount of information that is available on resistance towards this drug (9). All algorithms are based on the same available information and thus predict the same level of resistance.

The four evaluated algorithms, in fact, belong to two different models. The Stanford algorithm assigns a score to each of the observed mutations and uses the sum to decide on the level of resistance, allowing complex patterns of mutations to be taken into account. The VGI, ANRS, and Rega algorithms are restrained to specific rules that describe specific mutational patterns. Therefore, the discordance for complex patterns is especially inevitable since both models use different ways to take these into account.

This study is not intended to draw conclusions on the validity of the different algorithms, but rather to identify mutation patterns that result in divergence between the algorithms, among different subtypes. The mutations and particularly the patterns of polymorphisms in non-B subtypes that are associated with viral resistance warrant further in vitro studies and ultimately need to be confirmed by clinical observation. We acknowledge, as a limitation of this study, the absence of measures of either in vitro or clinical resistance, which are phenotype and therapy outcome, respectively. However, the mutation patterns associated with discordance between the algorithms may identify the sequences of interest in larger datasets, obtained prospectively, and linked to viral load and/or CD4 data to correlate treatment outcomes.

In conclusion, the different algorithms agreed quite well on the level of resistance scored. However, where there are differences, in many cases these can be attributed to specific subtype-dependent combinations of mutations. The mutations found here should further be investigated as to whether they contribute to differences in resistance and therapy response between different subtypes. Our expertise in interpretation of genotypic resistance will increase with a scale-up of treatment to include millions of individuals with non-subtype B virus infections.

REFERENCES

1. **Abecasis, A. B., K. Deforche, J. Snoeck, L. Bacheler, P. McKenna, P. Carvalho, P. Gomes, R. Camacho, and A.-M. Vandamme.** 2005. Protease mutation M89I/V is linked to therapy failure in patients infected with the HIV-1 non-B subtypes C, F or G. AIDS **19**:1799–1806.
2. **Ariyoshi, K., M. Matsuda, H. Miura, S. Tateishi, K. Yamada, and W. Sugiura.** 2003. 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. **33**:336–342.
3. **Betts, B. J., and R. W. Shafer.** 2003. Algorithm specification interface for human immunodeficiency virus type 1 genotypic interpretation. J. Clin. Microbiol. **41**:2792–2794.
4. **Brenner, B., D. Turner, M. Oliveira, D. Moisi, M. Detorio, M. Carobene, R. G. Marlink, J. Schapiro, M. Roger, and M. A. Wainberg.** 2003. A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to non-nucleoside reverse transcriptase inhibitors. AIDS **17**:F1–F5.
5. **Cohen, C. J., S. Hunt, M. Sension, C. Farthing, M. Conant, S. Jacobson, J. Nadler, W. Verbiest, K. Hertogs, M. Ames, A. R. Rinehart, and N. M. Graham.** 2002. A randomized trial assessing the impact of phenotypic resistance testing on antiretroviral therapy. AIDS **16**:579–588.
6. **De Luca, A., A. Cingolani, S. Di Giambenedetto, M. P. Trotta, F. Baldini, M. G. Rizzo, A. Bertoli, G. Liuzzi, P. Narciso, R. Murri, A. Ammassari, C. F. Perno, and A. Antinori.** 2003. Variable prediction of antiretroviral treatment outcome by different systems for interpreting genotypic human immunodeficiency virus type 1 drug resistance. J. Infect. Dis. **187**:1934–1943.
7. **de Oliveira, T., K. Deforche, S. Cassol, M. O. Salminen, D. Paraskevis, C. Seebregts, J. Snoeck, E. J. van Rensburg, A. M. J. Wensing, D. A. M. C. van de Vijver, C. A. Boucher, R. Camacho, and A.-M. Vandamme.** 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics **21**:3797–3800.
8. **De Wit, S., R. Boulmé, B. Poll, J.-C. Schmit, and N. Clumeck.** 2004. Viral load and CD4 cell response to protease inhibitor-containing regimens in subtype B versus non-B treatment-naive HIV-1 patients. AIDS **18**:2330–2331.

9. **Doyon, L., S. Tremblay, L. Bourgon, E. Wardrop, and M. G. Cordingley.** 2005. Selection and characterization of HIV-1 showing reduced susceptibility to the non-peptidic protease inhibitor tipranavir. Antivir. Res. 68:27–35.
10. **Durant, J., P. Clevenbergh, P. Halfon, P. Delgiudice, S. Porsin, P. Simonet, N. Montagne, C. A. Boucher, J. M. Schapiro, and P. Dellamonica.** 1999. Drug-resistance genotyping in HIV-1 therapy: the VIRADAPT randomised controlled trial. Lancet 353:2195–2199.
11. **Fontaine, E., C. Riva, M. Peeters, J.-C. Schmit , E. Delaporte, K. Van Laethem, K. Van Vaerenbergh, J. Snoeck, E. Van Wijngaerden, E. De Clercq, E. M. Van Ranst, and A.-M. Vandamme.** 2002. Evaluation of two commercial kits for the detection of genotypic drug-resistance on a panel of human immunodeficiency virus type-1 subtypes A-J. J. Acquir. Immune Defic. Syndr. 28:254–258.
12. **Frater, A. J., A. Beardall, K. Ariyoshi, D. Churchill, S. Galpin, J. R. Clarke, J. N. Weber, and M. O. McClure.** 2001. Impact of baseline polymorphisms in RT and protease on outcome of highly active antiretroviral therapy in HIV-1-infected African patients. AIDS 15:1493–1502.
13. **Frater, A. J., D. T. Dunn, A. J. Beardall, K. Ariyoshi, J. R. Clarke, M. O. McClure, and J. N. Weber.** 2002. Comparative response of African HIV-1-infected individuals to highly active antiretroviral therapy. AIDS 16:1139–1146.
14. **Gonzalez, L. M. F., R. M. Brindeiro, M. Tarin, A. Calazans, M. A. Soares, S. Cassol, and A. Tanuri.** 2003. In vitro hypersusceptibility of human immunodeficiency virus type 1 subtype C protease to lopinavir. Antimicrob. Agents Chemother. 47:2817–2822.
15. **Grossman, Z., E. E. Paxinos, D. Averbuch, S. Maayan, N. T. Parkin, D. Engelhard, M. Lorber, V. Istomin, Y. Shaked, E. Mendelson, D. Ram, C. J. Petropoulos, and J. M. Schapiro.** 2004. Mutation D30N is not preferentially selected by human immunodeficiency virus type 1 subtype C in the development of resistance to nelfinavir. Antimicrob. Agents Chemother. 48:2159–2165.
16. **Haubrich, R., and L. M. Demeter.** 2001. Clinical utility of resistance testing: retrospective and prospective data supporting use and current recommendations. J. Acquir. Immune Defic. Syndr. 26:S51–S59.
17. **Hirsch, M. S., F. Brun-Vezinet, C. Bonaventura, B. Conwy, D. R. Kuritzkes, R. T. D'Aquila, L. M. Demeter, S.-M. Hammer, V. A. Johnson, C. Loveday, J. W. Mellors, D. M. Jacobsen, and D. D. Richman.** 2003. Antiretroviral drug resistance testing in adults infected with human immunodeficiency virus type 1: 2003 recommendations of an International AIDS Society–USA Panel. Clin. Infect. Dis. 37:113–128.
18. **Holguin, A., K. Hertogs, and V. Soriano.** 2003. Performance of drug resistance assays in testing HIV-1 non-B subtypes. Clin. Microbiol. Infect. 9:323–326.
19. **Jagodzinski, L. L., J. D. Cooley, M. Weber, and N. L. Michael.** 2003. Performance characteristics of human immunodeficiency virus type 1 (HIV-1) genotyping systems in sequence-based analysis of subtypes other than HIV-1 subtype B. J. Clin. Microbiol. 41:998–1003.
20. **Kantor, R., D. Katzenstein, B. Efron, P. Carvalho, B. Wynhoven, P. Cane, J. R. Clarke, S. Sirivichayakul, M. A. Soares, J. Snoeck, C. Pillay, H. Rudich, R. Rodrigues, A. Holguin, K. Ariyoshi, P. Weidle, M. B. Bouzas, P. Cahn, W. Sugiura, V. Soriano, L. F. Brigido, Z. Grossman, L. Morris, A. M. Vandamme, A. Tanuri, P. Phanuphak, J. Weber, D. Pillay, P. R. Harrigan, R. Camacho, J. M. Schapiro, and R. W. Shafer.** 26 April 2005. Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. PLOS Med. 2:e112. [Epub ahead of print.]
21. **Kijak, G. H., A. E. Rubio, S. E. Pampuro, C. Zala, P. Cahn, R. Galli, J. S. Montaner, and H. Salomon.** 2003. Discrepant results in the interpretation of HIV-1 drug-resistance genotypic data among widely used algorithms. HIV Med. 4:72–78.
22. **Korn, K., H. Reil, H. Walter, and B. Schmidt.** 2003. Quality control trial for human immunodeficiency virus type 1 drug resistance testing using clinical samples reveals problems with detecting minority species and interpretation of test results. J. Clin. Microbiol. 41:3559–3565.
23. **Loveday, C., D. Dunn, H. Green, A. R. Rinehart, and P. McKenna on behalf of the ERA Steering Committee.** 2003. A randomized controlled trial of phenotypic resistance testing in addition to genotypic resistance testing: the ERA trial. Antivir. Ther. 8(Suppl. 1):S188.
24. **Maes, B., Y. Schrooten, J. Snoeck, I. Derdelinckx, M. Van Ranst, A. M. Vandamme, and K. Van Laethem.** 2004. Performance of Viroseq HIV-1 genotyping system in routine practice at a Belgian clinical laboratory. J. Virol. Methods 119:45–49.
25. **Meynard, J. L., M. Vray, L. Morand-Joubert, E. Race, D. Descamps, G. Peytavin, S. Matheron, C. Lamotte, S. Guiramand, D. Costagliola, F. Brun-Vezinet, F. Clavel, P. M. Girard, and the Narval Trial Group.** 2002. Phenotypic or genotypic resistance testing for choosing antiretroviral therapy after treatment failure: a randomized trial. AIDS 16:727–736.
26. **Monno, L., A. Saracino, L. Scudeller, G. Pastore, S. Bonora, A. Cargnel, G. Carosi, and G. Angarano.** 2003. HIV-1 phenotypic susceptibility to lopinavir (LPV) and genotypic analysis in LPV/r-naive subjects with prior protease inhibitor experience. J. Acquir. Immune Defic. Syndr. 33:439–447.
27. **Parkin, N. T., C. Chappey, and C. J. Petropoulos.** 2003. Improving lopinavir genotype algorithm through phenotype correlations: novel mutation patterns and amprenavir cross-resistance. AIDS 17:955–961.
28. **Perno, C. F., A. Cozzi-Lepri, F. Forbici, A. Bertoli, M. Violin, M. Stella Mura, G. Cadeo, A. Orani, A. Chirianni, C. De Stefano, C. Balotta, A. d'Arminio Monforte, and the Italian Cohort Naive Antiretrovirals Study Group.** 2004. Minor mutations in HIV protease at baseline and appearance of primary mutation 90M in patients for whom their first protease-inhibitor antiretroviral regimens failed. J. Infect. Dis. 189:1983–1987.
29. **Ravela, J., B. J. Betts, F. Brun-Vezinet, A.-M. Vandamme, D. Descamps, K. Van Laethem, K. Smith, J. M. Schapiro, D. L. Winslow, C. Reid, and R. W. Shafer.** 2003. HIV-1 protease and reverse transcriptase mutation patterns responsible for discordances between genotypic drug resistance interpretation algorithms. J. Acquir. Immune Defic. Syndr. 33:8–14.
30. **Reid, C. L., R. Bassett, S. Day, B. Larder, V. De Gruttola, and D. L. Winslow.** 2002. A dynamic rules-based interpretation system derived by an expert panel is predictive of virological failure. Antivir. Ther. 7:S121.
31. **Sarmati, L., E. Nicastri, M. A. Montano, L. Dori, A. R. Buonomini, G. d'Ettorre, F. Gatti, S. G. Parisi, V. Vullo, and M. Andreoni.** 2004. Decrease of replicative capacity of HIV isolates after genotypic guided change of therapy. J. Med. Virol. 72:511–516.
32. **Schuurman, R., D. Brambilla, T. de Groot, D. Huang, S. Land, J. Bremer, L. Benders, C. A. Boucher, and the ENVA Working Group.** 2002. Underestimation of HIV type 1 drug resistance mutations: results from the ENVA-2 genotyping proficiency program. AIDS Res. Hum. Retrovir. 18:243–248.
33. **Shafer, R. W., R. J. Duane, B. J. Betts, Y. Xi, and M. J. Gonzales.** 2000. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res. 28:346–348.
34. **Stürmer, M., H. W. Doerr, and W. Preiser.** 2003. Variety of interpretation systems for human immunodeficiency virus type 1 genotyping: confirmatory information or additional confusion? Curr. Drug Targets Infect. Disord. 3:255–262.
35. **Stürmer, M., H. W. Doerr, S. Staszewski, and W. Preiser.** 2003. Comparison of nine resistance interpretation systems for HIV-1 genotyping. Antivir. Ther. 8:55–60.
36. **Torti, C., E. Quiros-Roldan, W. Keulen, L. Scudeller, S. Lo Caputo, C. A. Boucher, F. Castelli, F. Mazzotta, P. Pierotti, A. M. Been-Tiktak, G. Buccolieri, M. De Gennaro, G. Carosi, C. Tinelli, and the GenPherex Study Group of the MaSTeR Cohort.** 2003. Comparison between rules-based human immunodeficiency virus type 1 genotype interpretations and real or virtual phenotype: concordance analysis and correlation with clinical outcome in heavily treated patients. J. Infect. Dis. 188:194–201.
37. **Tural, C., L. Ruiz, C. Holtzer, J. Schapiro, P. Viciana, J. Gonzales, P. Domingo, C. A. Boucher, C. Rey-Joly, B. Clotet, and the Havana Study Group.** 2002. The clinical utility of HIV-1 genotyping and expert advice: the Havana trial. AIDS 16:209–218.
38. **Vandamme, A. M., A. Sonnerborg, M. Ait-Khaled, J. Albert, B. Asjo, L. Bacheler, D. Banhegyi, C. A. Boucher, F. Brun-Vezinet, R. Camacho, P. Clevenbergh, N. Clumeck, N. Dedes, A. De Luca, R. W. Doerr, J. L. Faudon, G. Gatti, J. Gerstoft, W. W. Hall, A. Hatzakis, N. S. Hellmann, A. Horban, J. D. Lundgren, D. J. Kempf, D. Miller, V. Miller, T. W. Myers, C. Nielsen, M. Opravil, L. Palmisano, C. F. Perno, A. N. Phillips, D. Pillay, T. Pumarola, L. Ruiz, M. O. Salminen, J. M. Schapiro, B. Schmidt, J.-C. Schmit, R. Schuurman, E. Shulse, V. Soriano, S. Staszewski, S. Vella, R. Ziermann, and L. Perrin.** 2004. Updated European recommendations for the clinical use of HIV drug resistance testing. Antivir. Ther. 9:829–848.
39. **Van Laethem, K., A. De Luca, A. Antinori, A. Cingolani, C. F. Perno, and A.-M. Vandamme.** 2002. A genotypic drug resistance algorithm that significantly predicts therapy response in HIV-1 infected patients. Antivir. Ther. 7:123–129.
40. **Van Laethem, K., K. Van Vaerenbergh, J.-C. Schmit, S. Sprecher, P. Hermans, V. De Vroey, R. Schuurman, T. Harrer, M. Witvrouw, E. Van Wijngaerden, L. Stuyver, M. Van Ranst, J. Desmyter, E. De Clercq, and A.-M. Vandamme.** 1999. Phenotypic assays and sequencing are less sensitive than point mutation assays for detection of resistance in mixed HIV-1 genotypic populations. J. Acquir. Immune Defic. Syndr. 22:107–118.

Available online at www.sciencedirect.com
ScienceDirect
Infection, Genetics and Evolution xxx (2007) xxx–xxx
Infection, Genetics and Evolution
www.elsevier.com/locate/meegid

Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors

K. Deforche [a,*], R. Camacho [b], Z. Grossman [c], T. Silander [d], M.A. Soares [e], Y. Moreau [f], R.W. Shafer [g], K. Van Laethem [a,h], A.P. Carvalho [b], B. Wynhoven [i], P. Cane [j], J. Snoeck [a], J. Clarke [k], S. Sirivichayakul [l], K. Ariyoshi [m], A. Holguin [n], H. Rudich [c], R. Rodrigues [o], M.B. Bouzas [p], P. Cahn [p], L.F. Brigido [o], V. Soriano [n], W. Sugiura [m], P. Phanuphak [l], L. Morris [q], J. Weber [k], D. Pillay [r], A. Tanuri [e], P.R. Harrigan [i], J.M. Shapiro [s], D.A. Katzenstein [g], R. Kantor [g], A.-M. Vandamme [a]

[a] *Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium*
[b] *Virology Laboratory, Hospital Egas Moniz, Lisbon, Portugal*
[c] *Chaim Sheba Medical Center, Ministry of Health, Tel-Aviv, Israel*
[d] *Helsinki Institute for Information Technology, Helsinki, Finland*
[e] *Departamento de Genética, Universidade Federal do Rio de Janeiro, Brazil*
[f] *ESAT, Katholieke Universiteit Leuven, Leuven, Belgium*
[g] *Division of Infectious Diseases, Brown University, Providence RI, USA*
[h] *AIDS Reference Laboratory, University Hospitals, Leuven, Belgium*
[i] *BC Centre for Excellence in HIV/AIDS, Vancouver, Canada*
[j] *Reference and Microbiology Division, Health Protection Agency, Porton Down, UK*
[k] *Department of GUM & Communicable Diseases, Wright Fleming Institute, London, UK*
[l] *Department of Medicine, Chulalongkorn University, Bangkok, Thailand*
[m] *Department of Pathology, National Institute of Infectious Diseases, Tokyo, Japan*
[n] *Department of Infectious Diseases, Hospital Carlos III, Madrid, Spain*
[o] *Laboratorio de Retrovirologia, Instituto Adolfo Lutz, São Paulo, Brazil*
[p] *Fondación Huesped, Buenos Aires, Argentina*
[q] *AIDS Unit, National Institute for Communicable Diseases, Johannesburg, South Africa*
[r] *Antiviral Susceptibility Reference Unit, Health Protection Agency, Birmingham, UK*
[s] *National Hemophilia Center, Sheba Medical Center, Tel Aviv, Israel*

Received 17 February 2006; received in revised form 8 September 2006; accepted 11 September 2006

Abstract

Interpretation of Human Immunodeficiency Virus 1 (HIV-1) genotypic drug resistance is still a major challenge in the follow-up of antiviral therapy in infected patients. Because of the high degree of HIV-1 natural variation, complex interactions and stochastic behaviour of evolution, the role of resistance mutations is in many cases not well understood. Using Bayesian network learning of HIV-1 sequence data from diverse subtypes (A, B, C, F and G), we could determine the specific role of many resistance mutations against the protease inhibitors (PIs) nelfinavir (NFV), indinavir (IDV), and saquinavir (SQV). Such networks visualize relationships between treatment, selection of resistance mutations and presence of polymorphisms in a graphical way. The analysis identified 30N, 88S, and 90M for nelfinavir, 90M for saquinavir, and 82A/T and 46I/L for indinavir as most probable major resistance mutations. Moreover we found striking similarities for the role of many mutations against all of these drugs. For example, for all three inhibitors, we found that the novel mutation 89I was minor and associated with mutations at positions 90 and 71. Bayesian network learning provides an autonomous method to gain insight in the role of resistance mutations and the influence of HIV-1 natural variation.

* Corresponding author. Tel.: +32 16 332160; fax: +32 16 332131.
E-mail address: koen.deforche@gmail.com (K. Deforche).

1567-1348/$ – see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.meegid.2006.09.004

Please cite this article in press as: Deforche, K. et al., Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors, Infect. Genet. Evol. (2007), doi:10.1016/j.meegid.2006.09.004

We successfully applied the method to three protease inhibitors. The analysis shows differences with current knowledge especially concerning resistance development in several non-B subtypes.

Keywords: HIV; Protease; Nelfinavir; Indinavir; Saquinavir

1. Introduction

Human Immunodeficiency Virus (HIV) escapes the inhibitory effect of antiretroviral drugs by selection of mutations that increase resistance against those drugs. To obtain an effective therapy, it is thus necessary to use antiretroviral drugs for which the virus remains susceptible. Genotypic interpretation systems predict the susceptibility or therapy response for various drugs (Shafer, 2002; Van Laethem et al., 2002), based on the presence of mutations at positions associated with drug resistance. Unfortunately, the role of many resistance mutations remains unsufficiently known, as well as the role of HIV-1 natural variation. This variation within the HIV main group is reflected in a subtype system with 9 identified subtypes and 16 Circulating Recombinant Forms (CRFs). In addition, unclassified strains and new recombinants are increasingly reported. Different prevalences of known resistance-associated mutations and new mutations are seen in different subtypes (Frater et al., 2001; Grossman et al., 2001; Brindeiro et al., 2002; Ariyoshi et al., 2003; Parkin and Schapiro, 2004). With a few exceptions, these differences in prevalence could not be explained by different genetic barriers because of different codon usage (Turner et al., 2004). In previous work, we used Bayesian network (BN) learning to demonstrate how polymorphisms may influence how drug-associated mutations get selected. These explained some notable subtype differences that have been observed for resistance development against nelfinavir (Deforche et al., 2006).

In this work we present the application of Bayesian network learning to study development of resistance against three protease inhibitors (PIs): nelfinavir (NFV), indinavir (IDV), and saquinavir (SQV). Results were compared in the context of cross-resistance within the class of PIs.

A Bayesian network (BN) is a probabilistic model that describes statistical independencies between multiple variables. In this work, we learn Bayesian networks from observations of the variables. In this way, the best Bayesian network is searched that explains a maximum of the observed correlations in the data using a minimum number of *direct influences*. Dependencies are visualized in a directed acyclic graph and form the qualitative component of the BN. In this graph, each node corresponds to a variable, and a directed arc (arrow) between nodes represents a direct influence. Mathematically, a Bayesian network provides a refactoring of the Joint Probability Distribution (JPD) of the data, using Bayes' rule. As a BN simplifies the JPD, it provides an effective model that summarizes statistical properties of the data.

Within the study of drug resistance, one often refers to a mutation that is selected as a first mutation as a *major mutation* (Shafer, 2002; Johnson et al., 2004). Similarly, a *minor mutation* further increases resistance only in presence of other mutations, or compensates for a possible fitness impact of other mutations, and is therefore selected only in presence of these other mutations. Although these concepts are not rigorously defined, conditional independencies in the networks allow us to identify major and minor mutations, in agreement with these definitions.

2. Materials and methods

Data was derived from five clinical databases: Portugal, Belgium, Israel, Brazil and an international database containing sequences from subtypes other than subtype B. In total we had access to 4911 sequences. Protease (PRO) and partial reverse transcriptase (RT) HIV-1 sequences from protease inhibitor (PI) naive patients and from patients treated with only experience to NFV, IDV, or SQV as only PI, either unboosted or boosted with ritonavir, were trimmed to the first 350 amino-acids. At most one treated sequence and one naive sequence per patient were included and identical sequences were removed. RT inhibitor experienced patients were included in the PI naive patient population, since no resistance to RT inhibitors is expected in the protease gene.

The analysis followed closely the method described in Deforche et al. (2006). Subtyping was done using a phylogenetic analysis (de Oliveira et al., 2005). We identified wild type polymorphisms based on a prevalence greater than 10% in untreated patients and determined treatment associated mutations by testing for independence from treatment using a Cochran–Mantel–Haenszel χ^2 test, stratifying in each combination of subtype and database. The statistical analysis was corrected for multiple comparisons using Benjamini & Hochberg with a False Discovery Rate of 0.05. The data sets for Bayesian network were also stratified for an equal ratio of treated and untreated sequences within each combination of subtype and database, and included next to treatment experience, Boolean variables indicating presence of each treatment associated mutation and presence of polymorphic amino acids. Bayesian network learning was done by searching using a simulated annealing heuristic for the most probable network structure using a Bayesian scoring metric. A non-parametric bootstrap was performed by resampling from the sequences, to assess the robustness of network features.

In the final networks, we do not show the obvious strong antagonistic direct influences between different amino acids at single residue. Only network features (presence or absence of arcs) with a bootstrap higher than 65% were considered robust, and only robust arcs are shown. To reduce the

Please cite this article in press as: Deforche, K. et al., Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors, Infect. Genet. Evol. (2007), doi:10.1016/j.meegid.2006.09.004

complexity of the graphs, polymorphic positions that did not directly influence any treatment-associated mutations were omitted. Arcs were colored according to their function to improve reading the graph, but this coloring is only indicative. For each drug, known resistance mutations are those that are defined for that drug in either the International AIDS Society list of resistance mutations of 2005 (Johnson et al., 2005) or that are included in the resistance score in at least one of the latest versions of public resistance interpretation systems ANRS 2004.09, REGA 6.2 or HIVDB 2004.12 (Kantor et al., 2001).

To interpret the Bayesian network in the context of antiretroviral resistance, we considered the meaning of an arc between two mutations that was derived in Deforche et al. (2006). A major mutation is unconditionally dependent on

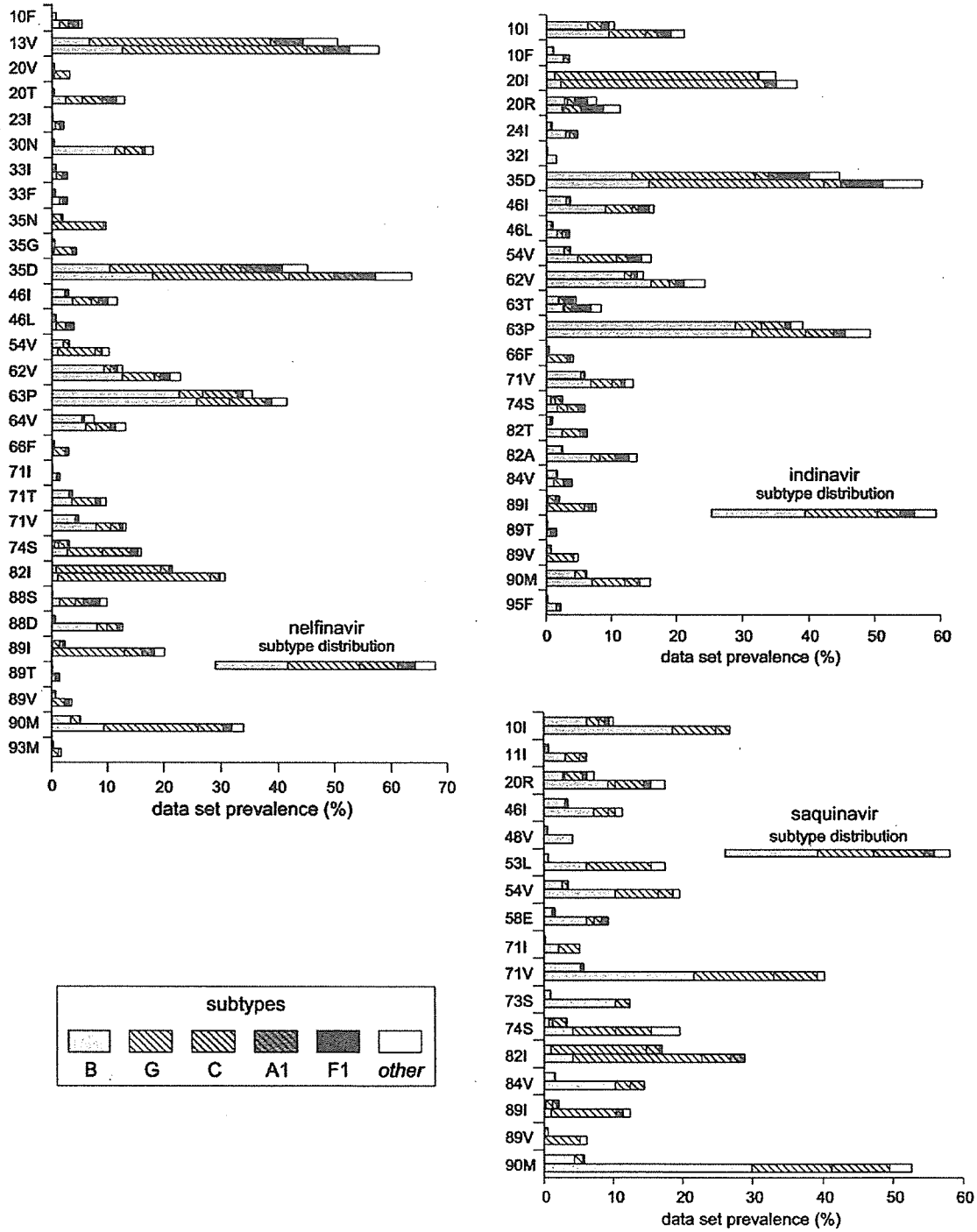

Fig. 1. Dataset prevalence (%) of NFV, IDV, and SQV treatment-associated mutations in sequences from untreated (top bar) and treated (bottom bar) patients. For each drug, the data was stratified for the overall subtype distribution of the sequences to be identical for treated and untreated patients.

Please cite this article in press as: Deforche, K. et al., Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors, Infect. Genet. Evol. (2007), doi:10.1016/j.meegid.2006.09.004

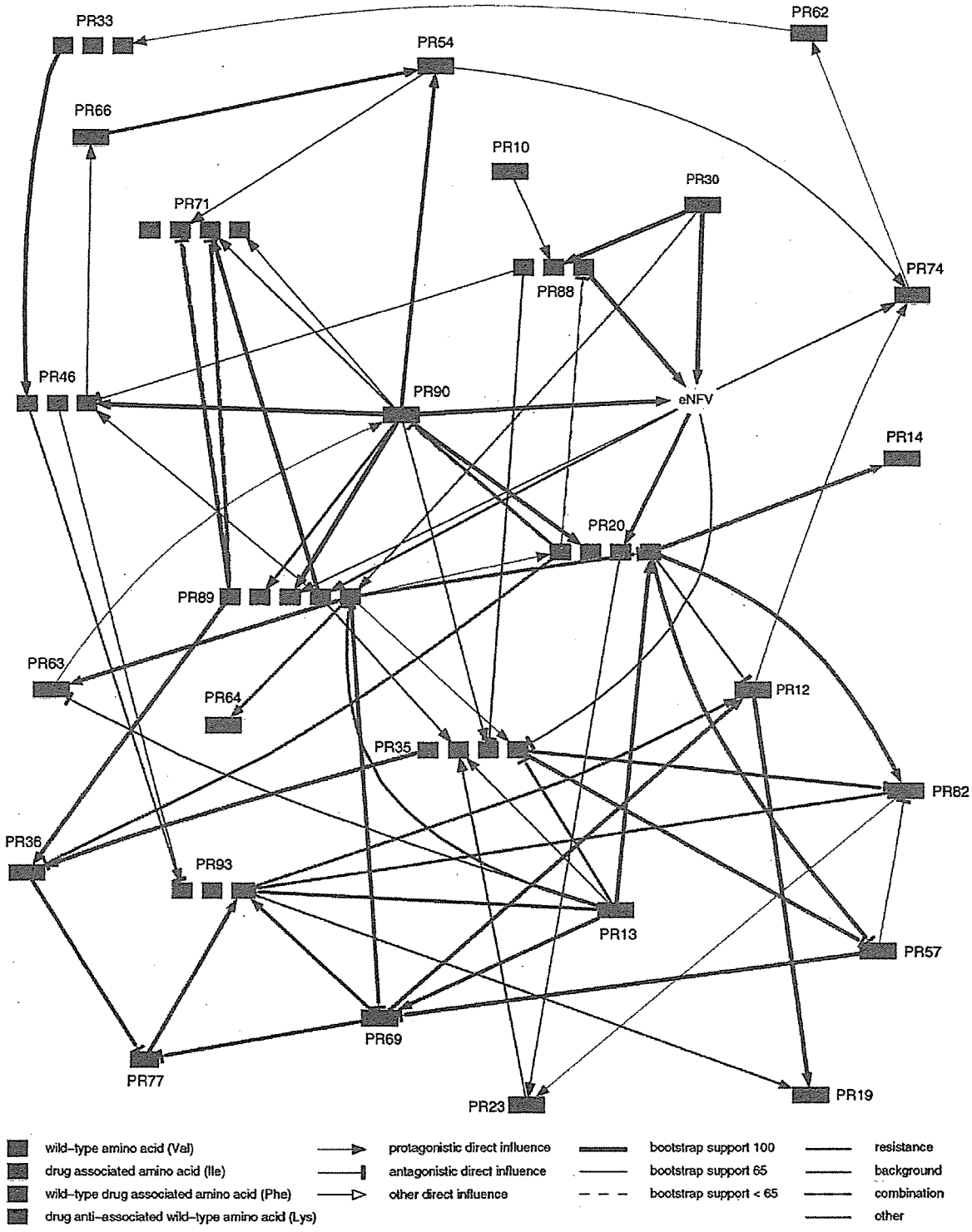

Fig. 2. Annotated NFV experience Bayesian network showing direct influences between NFV-associated mutations, polymorphisms and NFV treatment (eNFV). An arc represents a direct dependency between the corresponding variables and thickness is proportional to bootstrap support. Arc color indicates whether it is a direct influence between NFV-associated mutations (black), an influence from background polymorphisms on NFV-associated mutations (blue), or a combination of these (blue–black dashed) or merely an association between background polymorphisms (green). An antagonistic arc with a wild type was treated the same as a synergistic arc with mutations at this position. Arc direction has no causal meaning, but may indicate a non-additive multivariate effect. Unconditional dependencies with treatment with bootstrap support between 35% and 65% are shown dashed.

Please cite this article in press as: Deforche, K. et al., Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors, Infect. Genet. Evol. (2007), doi:10.1016/j.meegid.2006.09.004

treatment, which is indicated in the networks as the robust presence of an arc between the mutation and treatment. In this respect, an antagonistic arc with a wild type was treated the same as a synergistic arc with mutations at this position. Similarly, a minor mutation is expected to be conditionally independent from treatment but dependent on other resistance mutations, and thus indicated by the robust absence of an arc between the mutation and treatment. As was discussed in Deforche et al. (2006), a minor mutation may still be connected to treatment, when the cost is lower to connect to treatment instead of all the resistance mutations it is associated with. Where appropriate, we used the multivariate effect implied by arc directions as well to narrow down the list of major mutations.

3. Results

3.1. Subtyping

The subtype could be determined for 85% of the sequences. The overall subtype distribution of the sequences was subtype G (31%), B (27%), C (24%), A1 (12%), D (3%), F1 (3%) and other subtypes (<1%). The subtype distribution was different for the untreated and each of the NFV, IDV and SQV treated populations. As a result the subtype distributions for each analysis were slightly different (Fig. 1).

3.2. Treatment-associated mutations

The data used to determine mutations associated with treatment with each of the drugs, included 479 (NFV), 539 (IDV), and 97 (SQV) sequences from patients with experience with that drug as sole PI.

Fig. 1 shows the prevalence of treatment associated mutations in naive and treated patients, that were identified for each of the three drugs, using a χ^2 statistical analysis.

The most notable discordances with known resistance mutations were the novel mutations 20V, 33I/F, 35D/G/N, 62V, 64V, 66F, 74S, 89I/T/V and 93M for NFV; 35D, 62V, 63T, 66F, 74S, 89I/T/V and 95F for IDV; and 11I, 58E, 74S, 82I, and 89I/V for SQV. These mutations, except for 20V, 35N, and 89T have been previously described to be associated with PI experience in different studies (Wu et al., 2003; Svicher et al., 2005), but not with specific inhibitors. Some of these novel mutations were associated with treatment by all three drugs (74S and 89I/V) or by two drugs (35D, 58E, 66F, 89T, and 95F). The selection of some of these mutations was more pronounced than selection of mutations that have been widely accepted as resistance mutations.

At the same time we did not find selection of mutations 82A/F/T/S or 84V by NFV, even though they are considered important for NFV resistance by all algorithms.

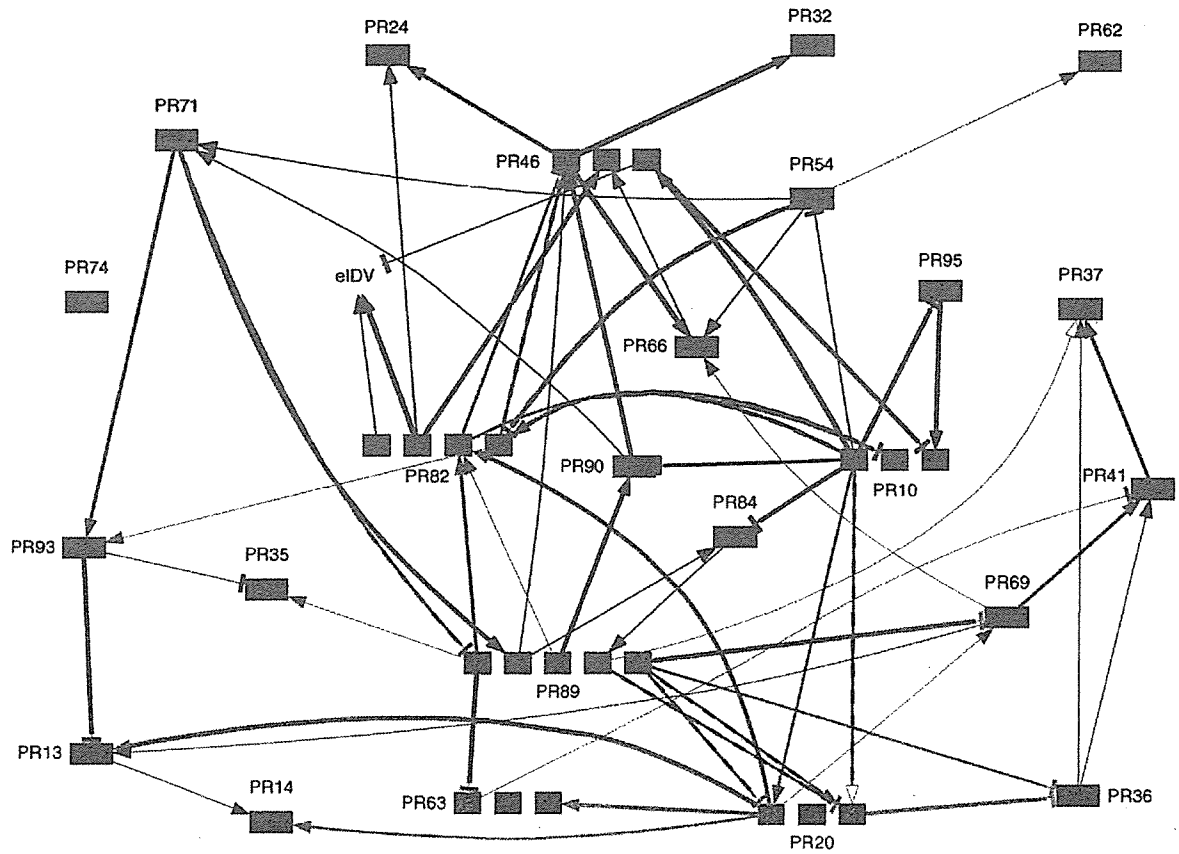

Fig. 3. Annotated IDV experience Bayesian network showing direct influences between IDV-associated mutations, polymorphisms and IDV treatment (eIDV). Legend as in Fig. 2.

Please cite this article in press as: Deforche, K. et al., Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors, Infect. Genet. Evol. (2007), doi:10.1016/j.meegid.2006.09.004

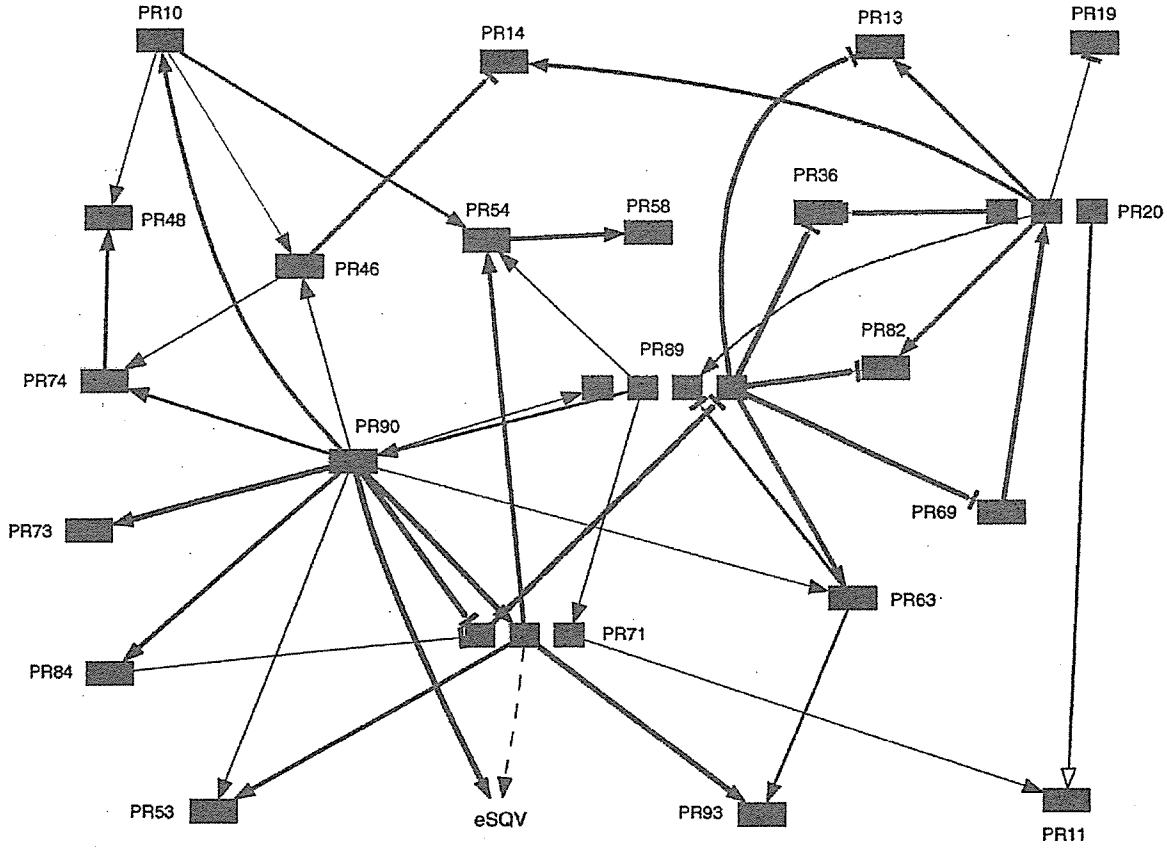

Fig. 4. Annotated SQV experience Bayesian network showing direct influences between SQV-associated mutations, polymorphisms and SQV treatment (eSQV). Legend as in Fig. 2.

3.3. Bayesian network learning

The data sets for Bayesian network learning included 340 (NFV), 288 (IDV), and 31 (SQV) sequences from patients on treatment, and respectively 967, 925, and 716 sequences from PI naive patients. Because of the stratification, the overall ratios of treated to untreated sequences were the same within every combination of data set and subtype.

The Bayesian network learning discovered many robust interactions between the variables in our data sets: the fraction of arcs with bootstrap support over 65% increased with available data and ranged from 44% (for the SQV network) to 68% (for the IDV network). The networks are shown in Figs. 2–4. For resistance against NFV, the network indicated 30N, 88S and 90M as major mutations, since they show a robust unconditional dependence on treatment, and analysis of the non-additive multivariable effect implied by arc directions at the treatment node, indicated that these three mutations occur mostly independently. The amount of selection of one of these major mutations is different for different subtypes (see Fig. 1). In subtype B, 30N is selected most. In subtypes C and G, 90M is selected most. Finally, in subtypes Al and Fl, 88S is selected most. For resistance against IDV, the network indicated 82A/T and 46I/L as major mutations. Mutation 84V is also selected by IDV in our data set, but we find it to be selected only after accumulation of mutations 82A/T or 46I/L, and 10I/F. The results for SQV were less conclusive, as more network features (presence or absence of arcs) were not robust because of the low amount of data. The network indicated 90M as major mutations but could not exclude 71V as additional major mutation (which was also unconditionally dependent on treatment in the most probable network, but with bootstrap support 47%), and 46I, 48V, 53L, 58E, 73S, 74S, or 89V as alternative major mutations, ordered by likelihood. These mutations were indicated by the most probable network as conditionally independent from treatment but this independence was not robust. Mutation 48V only occurred in subtype B, and the most probable network indicated that it appeared only after accumulation of mutations 90M and 10I or 74S in our data set.

For most minor mutations, that are conditionally independent from treatment, the networks suggest their role in more detail by indicating robust interactions with other resistance mutations in whose presence they are selected, and thus contribute to a selective advantage of the virus. The network for NFV shows that minor mutations 20T/V, 35N, 46I/L, 54V, 71I/T/V, and 89I/T/V, and the polymorphisms 63P and 89L directly influence a major mutation, while minor mutations 10F, 23I, 33I/F, 66F, or 93M are further away in the resistance pathway. Similarly, for IDV, mutations 10I/F, 24I, 32I, 54V, 66F, or 90M directly influence a major mutation, and mutations 20R, 71V, 74S, 84V, 89I/T/V, and 95F are further away in the resistance pathway.

Please cite this article in press as: Deforche, K. et al., Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors, Infect. Genet. Evol. (2007), doi:10.1016/j.meegid.2006.09.004

The networks indicate robust interactions between polymorphisms and resistance mutations, which may explain subtype differences. For example, the NFV network indicated a protagonistic interaction of the 89L polymorphism on development of the 30N mutation, explaining the higher prevalence of 30N in subtype B (Grossman et al., 2004; Abecasis et al., 2005). This effect was recently confirmed in *in vitro* experiments (Calazans et al., 2005). Similarly, both the IDV and SQV networks suggested a protagonistic interaction of the 93L polymorphism on development of the 71V mutation.

There are striking similarities when comparing the networks for different drugs, especially when considering arcs with bootstrap below 65% as well (which are not shown). All three networks indicated interactions between resistance mutations 90M and 89I, 90M and 71V, and 90M and 54V, and between the polymorphism 89L/M on mutations at position 71. In addition, interactions between polymorphisms and resistance mutations found in two networks were interactions of polymorphism T12S on mutation 74S, L63P on mutation 90M, I93L on mutation 90M, I93L on mutation 71V, and L10I on mutations 54V and 90M.

4. Discussion

Based on higher prevalence in sequences from treated versus untreated patients, we confirmed the selection of many known mutations by antiviral drugs, but also identified selection of novel mutations. As can be seen from Fig. 1, selection of these novel mutations was often more pronounced or sometimes exclusively in non-B subtypes.

The low level of selection of mutations at position 82 and 84 during NFV treatment is confirmed in other data sets (Kantor et al., 2001). In Shafer (2002) it is argued that mutations at position 82 have no phenotypic effect on their own for resistance against NFV, but contribute to resistance together with other mutations. A possible explanation for this discrepancy may be that selection of these mutations depends on the presence of other mutations that are not commonly present in untreated patients, or that are not selected by NFV, but are more common in patients exposed to other PI treatment.

The learned Bayesian networks indicated major mutations, largely in agreement with current knowledge (Johnson et al., 2004), with some exceptions. For NFV we found that 88S has a different role as 88D, and should be considered a major mutation, and may be more important than 30N or 90M in subtypes A1 and F1. Published phenotypic data supports this finding by indicating a phenotypic fold change in EC_{50} of 8.9 for 88S alone (Kantor et al., 2001). The IDV network indicated that 84V is not a major mutation, while it is widely considered so. As it is documented that it rarely develops as a first mutation, but only appears in isolates that already have a 90M (Shafer, 2002), this discrepancy is explained by a discordant definition of a major mutation. Similarly, according to our semantics, the SQV network indicates that 48V is not a major resistance mutation for SQV, since it virtually never occurred without mutation 90M. This is not due to its low prevalence in SQV failing sequences, which was comparable to the prevalence of the major IDV mutation 82T. However, the SQV dataset was rather small to make final conclusions.

The power of Bayesian network learning to find robust (in)dependencies in the data, depends on the sample size, the number of variables, and the actual number of independencies in the data set. It has previously been observed that resistance against IDV is less structured than resistance development against SQV (Beerenwinkel et al., 2004), which may explain why a similar amount of robust dependencies were observed for resistance against IDV as for SQV, despite the fact that the IDV data set was several times larger.

The biological role of minor mutations is to further increase resistance, and/or to compensate for a loss in replication capacity caused by the major mutation. Minor mutations that only improve replication capacity that was compromised by a resistance mutation in the virus should develop in the context of the same resistance mutations regardless of the inhibitor used. Indeed, these mutations may even develop in absence of the inhibitor, to improve replication capacity compromised by other resistance mutations (van Maarseveen et al., 2006). The similarities observed in the networks for different drugs, could thus indicate that mutations 10I, 12S, 54V, 63P, 71V, 89I, and 93L improve replication capacity compromised by other mutations, although their role in increasing resistance cannot be excluded.

The method of identifying possible resistance mutations by considering mutations associated with treatment in a cross-sectional data set can be confounded by drift. Drift may be the reason for a higher prevalence of a mutation in the treated population, and this is more likely for polymorphisms that occur frequently in the untreated population. Even after stratifying in combinations of database and subtype, we cannot exclude this effect of drift. At the same time, the Bayesian networks could not be used in most cases to reliably determine the role of these polymorphic resistance mutations, since they mostly ignored the relative low amount of variation associated with treatment while explaining the larger amount of variation at these polymorphisms in association with other polymorphisms. Mutations 10I and 63P however show similar linkage in the networks for different PIs, indicating their role. The interactions we found between 63P and the major mutation 90M are in agreement with earlier reports on the role of L63P (Martinez-Picado et al., 1999; Sune et al., 2004). The Bayesian networks could not clarify in a consistent way the role of other polymorphisms that we found to be associated with treatment (13V, 20I, 35D, 36I, 62V, 82I), despite the possibly important clinical implications.

For the analysis of SQV and IDV, data from both boosted and unboosted regimens were combined. The effect of boosting is suggested to primarily increase the genetic barrier to develop a clinically relevant level of resistance, by increasing the intracellular concentration of the drug. Whether this changes the patterns of drug resistance mutations has not been investigated yet. If using boosted regimen changes resistance pathways, then this would have blurred the analysis only yielding lower bootstrap confidence per pathway.

Please cite this article in press as: Deforche, K. et al., Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors, Infect. Genet. Evol. (2007), doi:10.1016/j.meegid.2006.09.004

A difficult question in resistance interpretation is how to score presence at baseline of minor mutations. These generally do not have an effect on resistance on their own, and may even represent a fitness penalty with respect to the wild type. Thus, without considering further evolution, the virus remains fully susceptible to the drug. However, some of these mutations improve the fitness or resistance impact of the corresponding major mutation. Therefore, the presence of these mutations will speed up the selection and increase the impact of these major mutations. Other minor mutations do not directly influence the major mutation, and thus do not have the same clinical significance when present at baseline. Therefore, we predict that for NFV, in absence of major mutations 30N, 88S, or 90M, presence of mutations 20T/V, 35N, 46I/L, 54V, 71I/T/V, or 89I/T/V, or of polymorphisms 63P or 89L, should impact clinical outcome to a greater extent than mutations 10F, 23I, 33I/F, 66F, or 93M. Similarly, for IDV, in absence of major mutations 82A/T and 46I/V, presence of mutations 10I/F, 24I, 32I, 54V, 66F, or 90M should have a higher impact on clinical outcome than mutations 20R, 71V, 74S, 84V, 89I/T/V, and 95F.

The power of Bayesian network learning lies in its sound mathematical foundation to distill likely direct interactions (which in many cases could be causalities) from the many observed associations between different residues. Bayesian network learning has previously successfully been applied to amino acid sequence data, in the context of secondary structure prediction (Klingler and Brutlag, 1994), but also in the field of HIV Drug Resistance (Beerenwinkel et al., 2004). In the latter analysis, Bayesian network models were constrained to trees with a special structure of the Conditional Probability Distributions (CPDs). In this way, the models described ordered accumulation of mutations. In contrast, we use unconstrained Bayesian networks, and added information on background polymorphisms to the analysis. As a consequence, both antagonistic and synergistic interactions between treatment associated mutations and polymorphisms were learned, without the prior assumption of a strict ordered accumulation of mutations.

5. Conclusions

We applied Bayesian network learning to HIV-1 protease sequence data and exposure to protease inhibitors to learn many aspects of resistance development against three protease inhibitors. We used the structure of the network to infer hypotheses about the role of resistance mutations. Our analysis confirmed current knowledge, especially for resistance development in subtype B viruses. Our analysis shows an important impact of polymorphisms on resistance development that could explain subtype differences in resistance development. Our results may suggest new *in vitro* experiments, to confirm the hypothesised role of novel resistance mutations, or be used to update genotypic resistance interpretation systems.

Acknowledgments

The authors wish to thank Ana Abecasis and Esmeralda AJM Soares for data collection. Koen Deforche was funded by a Ph.D. grant of the Institute for the Promotion of Innovation through Sciences and Technology in Flanders (IWT). Yves Moreau is a post-doctoral researcher with the FWO-Vlaanderen; his research is supported by KULeuven GOA-Mefisto-666 and GOA-Ambiorics, Belspo IUAP V-22, and EU FP6 NoE Biopattern. 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 funded through FWO-Vlaanderen grant G.0266.04, and by the Katholieke Universiteit Leuven through grant OT/04/43.

References

Abecasis, A.B., Deforche, K., Snoeck, J., Bacheler, L.T., Kenna, P.M., Carvalho, A.P., Gomes, P., Camacho, R.J., Vandamme, A.-M., 2005. Protease mutation M89I/V is linked to therapy failure in patients infected with the HIV-1 non-B subtypes C, F or G. AIDS 19 (16), 1799–1806.

Ariyoshi, K., Matsuda, M., Miura, H., Tateishi, S., Yamada, K., Sugiura, W., 2003. 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. 33 (3), 336–342.

Beerenwinkel, N., Rahnenfiihrer, J., Daumer, M., Hoffmann, D., Kaiser, R., Selbig, J., Lengauer, T., 2004. Learning multiple evolutionary pathways from cross-sectional data. In: RECOMB, ACM Press, pp. 36–44.

Brindeiro, P.A., Brindeiro, R.M., Mortensen, C., Hertogs, K., De Vroey, V., Rubini, N.P.M., Sion, F.S., De Sa, C.A.M., Machado, D.M., Succi, R.C.M., Tanuri, A., 2002. 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. 40 (12), 4512–4519.

Calazans, A., Brindeiro, R., Brindeiro, P., Verli, H., Arruda, M.B., Gonzalez, L.M.F., Guimaraes, J.A., Diaz, R.S., Antunes, O.A.C., Tanuri, A., 2005. Low accumulation of L90M in protease from subtype F HIV-1 with resistance to protease inhibitors is caused by the L89M polymorphism. J. Infect. Dis. 191 (11), 1961–1970.

de Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., Snoeck, J., van Rensburg, E.J., Wensing, A.M.J., van de Vijver, D.A., Boucher, C.A., Camacho, R., Vandamme, A.-M., 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21 (19), 3797–3800.

Deforche, K., Silander, T., Camacho, R., Grossman, Z., Soares, M.A., Van Laethem, K., Kantor, R., Moreau, Y., Vandamme, A.-M., 2006. On behalf of the Non-B Workgroup. Analysis of HIV-1 pol sequences using Bayesian Networks: implications for drug resistance. Bioinformatics, in press, http://bioinformatics.oxfordjournals.org/cgi/reprint/btl508v1.

Frater, A.J., Beardall, A., Ariyoshi, K., Churchill, D., Galpin, S., Clarke, J.R., Weber, J.N., McClure, M.O., 2001. Impact of baseline polymorphisms in RT and protease on outcome of highly active antiretroviral therapy in HIV-1-infected African patients. AIDS 15 (12), 1493–1502.

Grossman, Z., Paxinos, E.E., Averbuch, D., Maayan, S., Parkin, N.T., Engelhard, D., Lorber, M., Istomin, V., Shaked, Y., Mendelson, E., Ram, D., Petropoulos, C.J., Schapiro, J.M., 2004. Mutation D30N is not preferentially selected by human immunodeficiency virus type 1 subtype C in the development of resistance to nelfinavir. Antimicrob. Agents Chemother. 48 (6), 2159–2165.

Grossman, Z., Vardinon, N., Chemtob, D., Alkan, M.L., Bentwich, Z., Burke, M., Gottesman, G., Istomin, V., Levi, I., Maayan, S., Shahar, E., Schapiro, J.M., 2001. Genotypic variation of HIV-1 reverse transcriptase and protease: comparative analysis of clade C and clade B. AIDS 15 (12), 1453–1460.

Johnson, V.A., Brun-Vénizet, F., Onaventura, C., Conway, B., D'Aquila, R.T., Demeter, L.M., Kuritzkes, D.R., Pillay, D., Schapiro, J.M., Telenti, A., Richman, D.D., 2004. Update of the drug resistance mutations in HIV-1: 2004. Top HIV Med. 12 (4), 119–123.

Johnson, V.A., Brun-Vénizet, F., Onaventura, C., Conway, B., Kuritzkes, D.R., Pillay, D., Schapiro, J.M., Telenti, A., Richman, D.D., 2005. Update

Please cite this article in press as: Deforche, K. et al., Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors, Infect. Genet. Evol. (2007), doi:10.1016/j.meegid.2006.09.004

of the Drug Resistance Mutations in HIV-1: Fall 2005. Top HIV Med. 13 (4), 125–131.

Kantor, R., Machekano, R., Gonzales, M.J., Dupnik, K., Schapiro, J.M., Shafer, R.W., 2001. Human immunodeficiency virus reverse transcriptase and protease sequence database: an expanded data model integrating natural language and sequence analysis programs. Nucleic Acids Res. 29 (1), 296–299.

Klingler, T.M., Brutlag, D.L., 1994. Discovering structural correlations in α-helices. Protein Sci. 3 (10), 1847–1857.

Martinez-Picado, J., Savara, A.V., Sutton, L., D'Aquila, R.T., 1999. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. J. Virol. 73 (5), 3744–3752.

Parkin, N.T., Schapiro, J.M., 2004. Antiretroviral drug resistance in non-subtype B HIV-1, HIV-2 and SIV. Antivir. Ther. 9 (1), 3–12.

Shafer, R.W., 2002. Genotypic testing for human immunodeficiency virus type 1 drug resistance. Clin. Microbiol. Rev. 15 (2), 247–277.

Sune, C., Brennan, L., Stover, D.R., Klimkait, T., 2004. Effect of polymorphisms on the replicative capacity of protease inhibitor-resistant HIV-1 variants under drug pressure. Clin. Microbiol. Infect. 10 (2), 119–1119.

Svicher, V., Ceccherini-Silberstein, F., Erba, F.u., Santoro, M., Gori, C., Bellocchi, M.C., Giannella, S., Trotta, M.P., Monforte, A.d., Antinori, A., Perno, C.F., 2005. Novel human immunodeficiency virus type 1 protease mutations potentially involved in resistance to protease inhibitors. Antimicrob. Agents Chemother. 49 (5), 2015–2025.

Turner, D., Brenner, B., Moisi, D., Detorio, M., Cesaire, R., Kurimura, T., Mori, H., Essex, M., Maayan, S., Wainberg, M.A., 2004. Nucleotide and amino acid polymorphisms at drug resistance sites in non-B-subtype variants of human immunodeficiency virus type 1. Antimicrob. Agents Chemother. 48 (8), 2993–2998.

Van Laethem, K., De Luca, A., Antinori, A., Cingolani, A., Perno, C.F., Vandamme, A.-M., 2002. A genotypic drug resistance interpretation algorithm that significantly predicts therapy response in HIV-1 infected patients. Antivir. Ther. 7 (2), 1359–6535.

van Maarseveen, N.M., de Jong, D., Boucher, C.A.B., Nijhuis, M., 2006. An increase in viral replicative capacity drives the evolution of protease inhibitor-resistant human immunodeficiency virus type 1 in the absence of drugs. J. Acquir. Immune Defic. Syndr. 42 (2), 162–168.

Wu, T.D., Shiffer, C.A., Gonzales, M.J., Taylor, J., Kantor, R., Chou, S., Israelski, D., Zolopa, A.R., Fessel, W.J., Shafer, R.W., 2003. Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. J. Virol. 77 (8), 4836–4847.

Please cite this article in press as: Deforche, K. et al., Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors, Infect. Genet. Evol. (2007), doi:10.1016/j.meegid.2006.09.004