Available online at www.sciencedirect.com
ScienceDirect
Infection, Genetics and Evolution 7 (2007) 382–390
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
Available online 28 November 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

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.
© 2007 Published by Elsevier B.V.

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

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

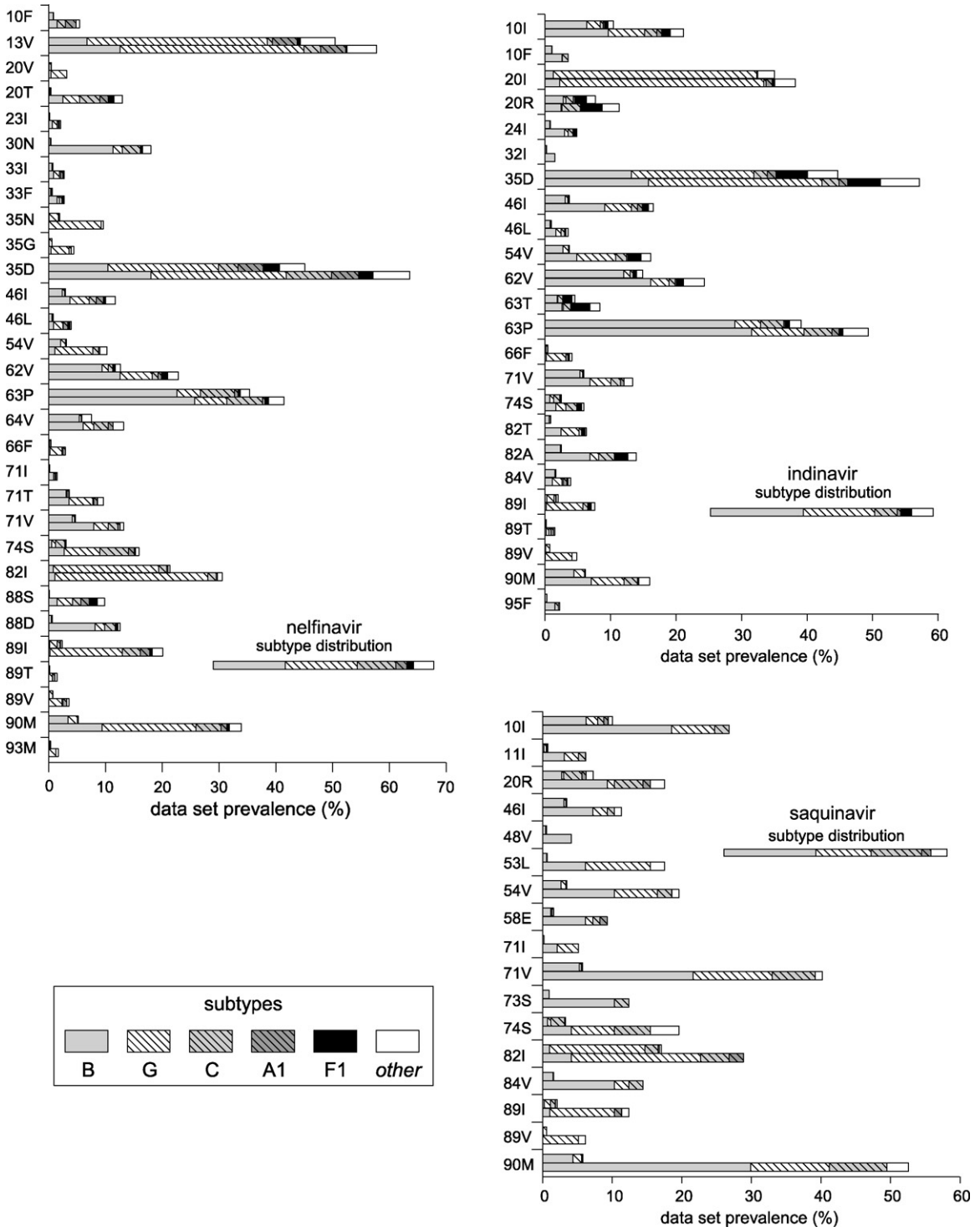


Fig. 1. Dataset prevalence (%) of NFV, IDV, and SQV treatment-associated mutations in sequences from untreated (top bar) and treated (bottom bar) patients. For each drug, the data was stratified for the overall subtype distribution of the sequences to be identical for treated and untreated patients.

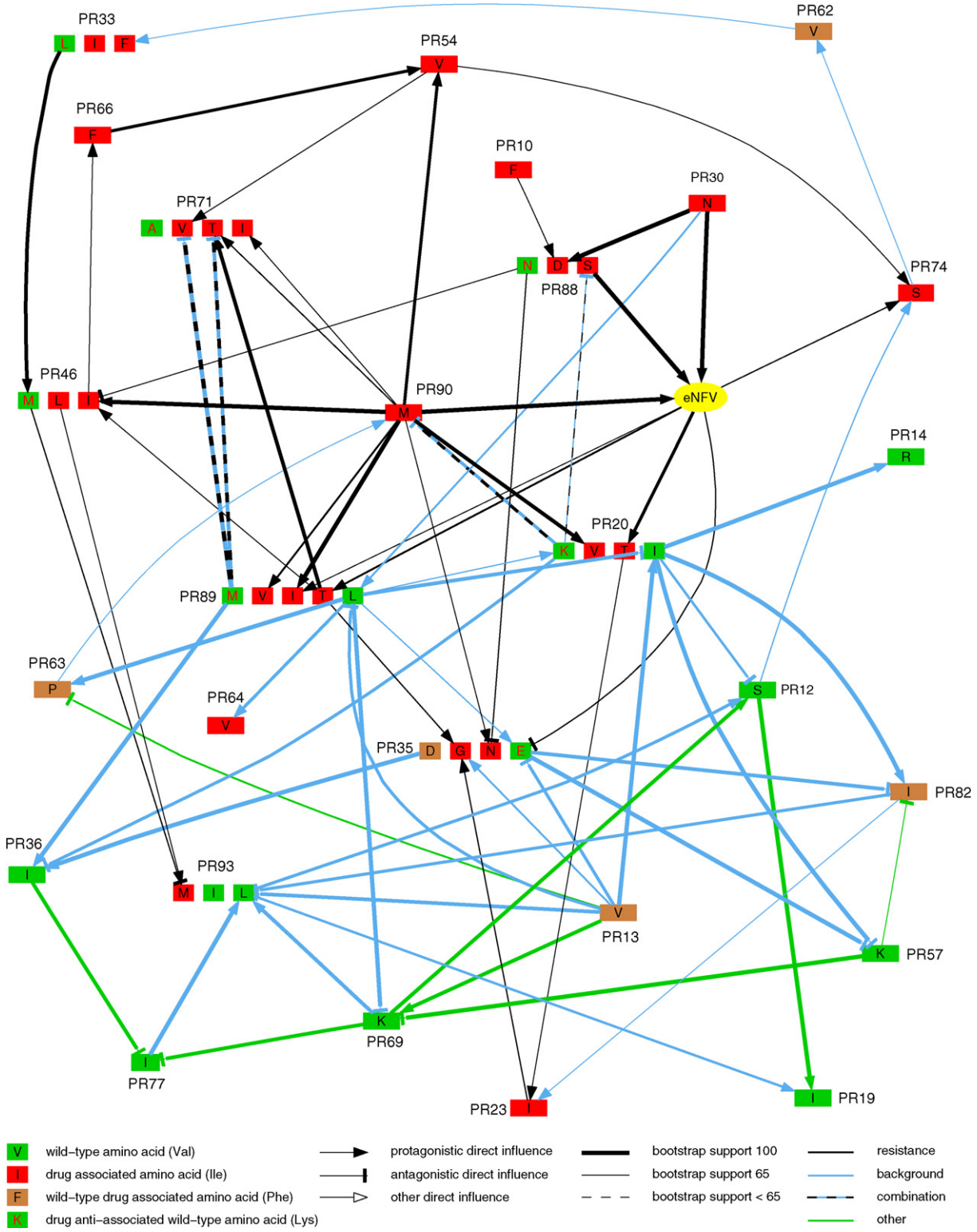


Fig. 2. Annotated NFV experience Bayesian network showing direct influences between NFV-associated mutations, polymorphisms and NFV treatment (eNFV). An arc represents a direct dependency between the corresponding variables and thickness is proportional to bootstrap support. Arc color indicates whether 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.

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%), Al (12%), D (3%), Fl (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.

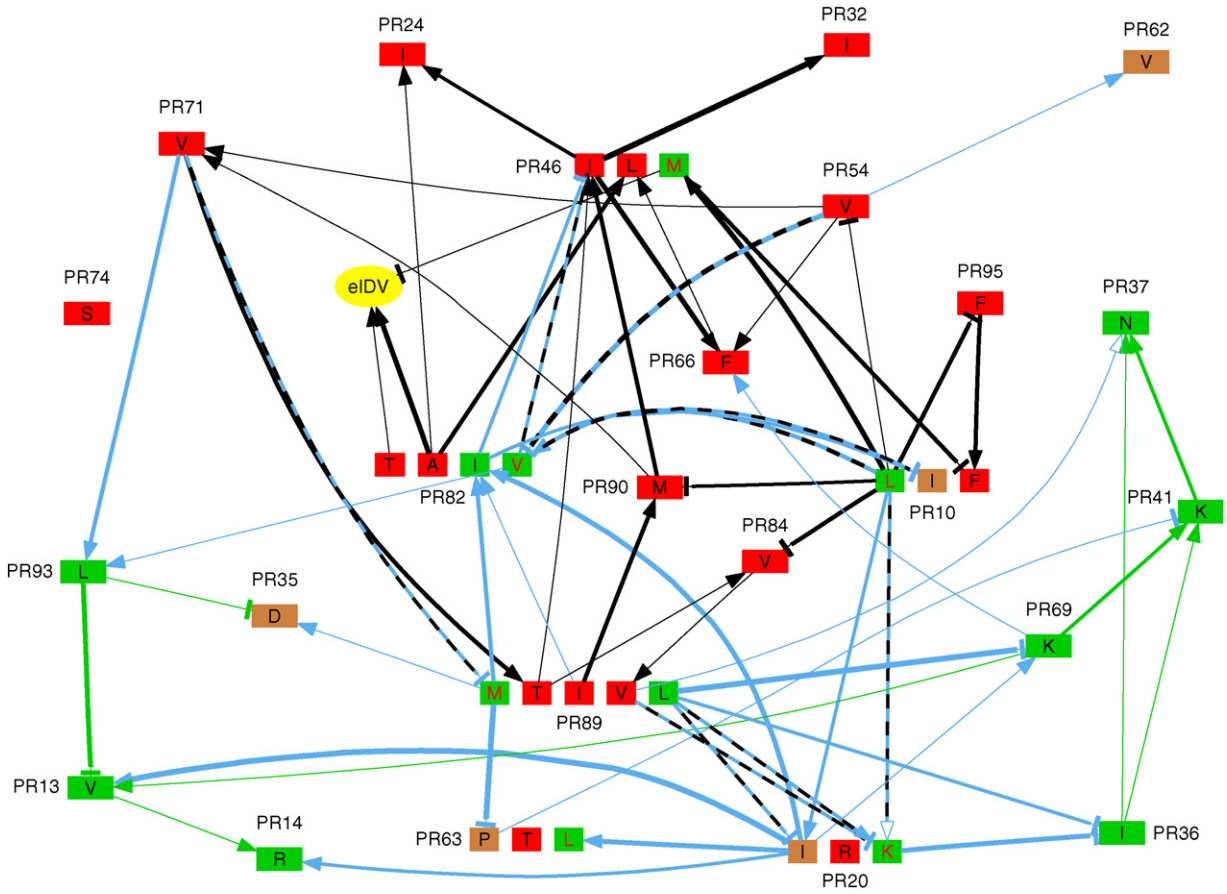


Fig. 3. Annotated IDV experience Bayesian network showing direct influences between IDV-associated mutations, polymorphisms and IDV treatment (elDV). Legend as in Fig. 2.

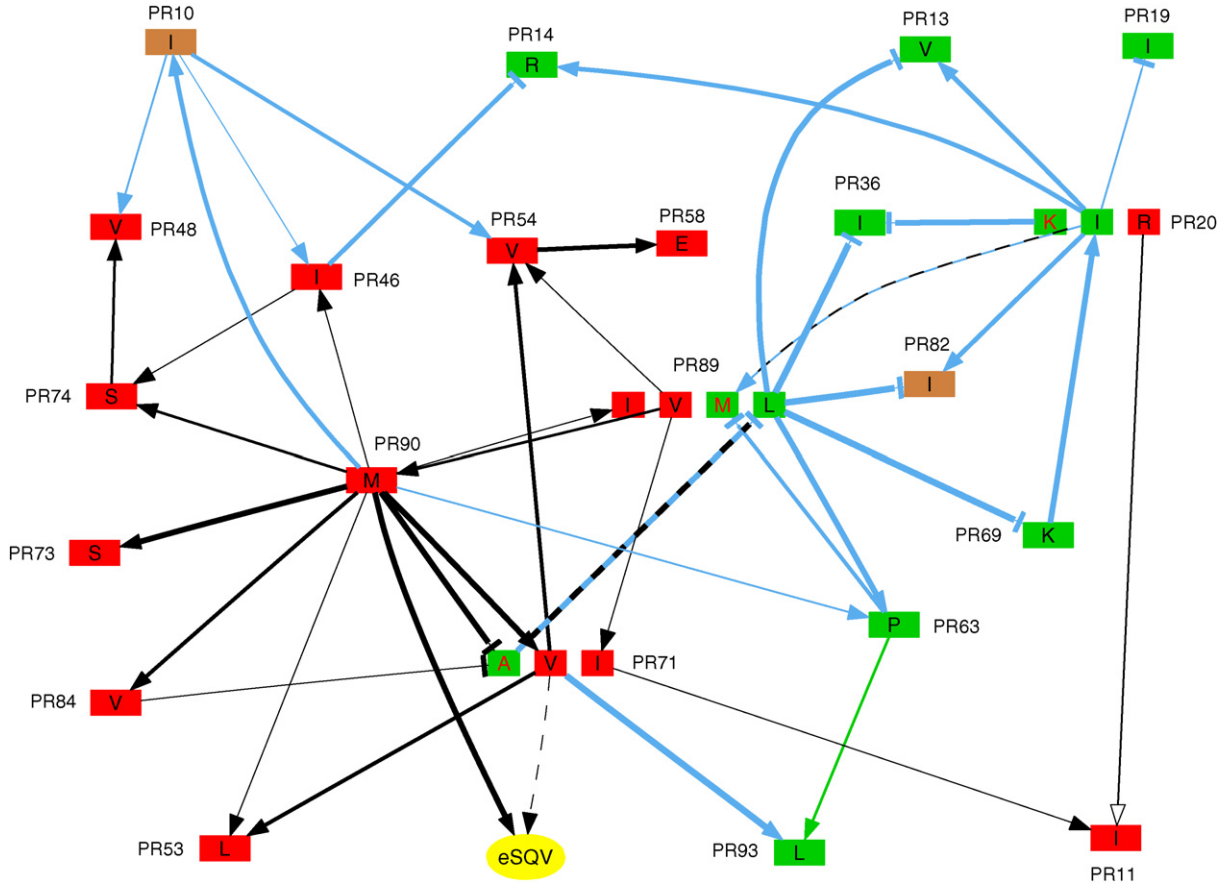


Fig. 4. Annotated SQV experience Bayesian network showing direct influences between SQV-associated mutations, polymorphisms and SQV treatment (eSQV). Legend as in Fig. 2.

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.

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 Al and Fl. 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.

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

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.