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
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