

Services for prediction of drug susceptibility for HIV proteases and reverse transcriptases at the HIV drug research centre

Ola Spjuth*, Martin Eklund, Maris Lapins, Muhammad Junaid and Jarl E. S. Wikberg

Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

Associate Editor: Alfonso Valencia

ABSTRACT

Summary: The HIV Drug Research Centre (HIVDRC) has established Web services for prediction of drug susceptibility for HIV proteases and reverse transcriptases. The services are based on two proteochemometric models which accept a protease or reverse transcriptase sequence in amino acid form, and outputs the predicted drug susceptibility values. The predictions are based on a comprehensive analysis where all the relevant inhibitors are included, resulting in models with excellent predictive capabilities.

Availability and Implementation: The services are implemented as interoperable Web services (REST and XMPP), with supporting web pages to allow for individual analyses. A set of plugins were also developed which make the services available from the Bioclipse workbench for life science. Services are available at <http://www.hivdrc.org/services>.

Contact: ola.spjuth@farmbio.uu.se

Received on December 20, 2010; revised on March 2, 2011; accepted on April 6, 2011

1 INTRODUCTION

HIV protease and reverse transcriptase (RT) are two major drug targets in HIV therapy with several drugs developed to inhibit their functions. However, the HIV virus is capable to escape current antiviral therapy directed at these targets by evolving into drug-resistant variants with highly complex mutation patterns. Drugs show different efficiency for different viral strains, and it is important to pick the medication which most efficiently can block HIV on individual patient basis. Several attempts have been made to produce models to predict drug susceptibility based on the sequenced HIV (Beerenwinkel *et al.*, 2003; Liu and Shafer, 2006), but the current models are commonly trained only on data for one drug and hence do not make full use of the available information. Expert systems are also present (Liu and Shafer, 2006), which build on the experience of physicians and are hence unable to extrapolate and predict the effect of novel mutations. The services presented here are founded on *proteochemometric* models for HIV proteases and reverse transcriptases, a technology which typically produces accurate predictions based on more information (Kontijevskis *et al.*, 2008).

*To whom correspondence should be addressed.

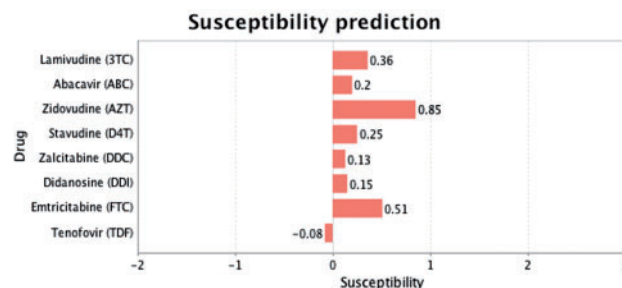


Fig. 1. The chart displays the resulting susceptibilities of an analyzed sequence (RT in this case) for various drugs. A positive value indicates a reduced susceptibility for the respective drug relative to a drug sensitive reference HIV strain, while a negative value indicates an increased susceptibility.

2 METHODS

2.1 HIV protease drug susceptibility prediction

The proteochemometric model described in Lapins *et al.* (2008) allows for prediction of drug susceptibility of seven clinically used protease inhibitors (PIs) from the HIV protease sequence. The model is based on 828 mutated HIV protease genome sequences and the experimental susceptibility data for the seven protease inhibitors measured by their ability to inhibit the replications of the respective mutated HIV variant *in vitro*; in total the model is based on 4794 protease sequence-inhibitor combinations.

2.2 HIV-RT drug susceptibility prediction

The proteochemometric model described in Junaid *et al.* (2010) allows for prediction of drug susceptibility of eight clinically used nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs). The model is based on 728 mutated HIV RT genome sequences and the experimental susceptibility data for the eight NRTIs measured by their ability to inhibit the replications of the respective mutated HIV variant *in vitro*; in total the model is based on 4495 RT sequence-inhibitor combinations.

2.3 Model validation

Both proteochemometric models were thoroughly statistically validated, among several methods using cross-validation and external predictions for new virus isolates. The protease model was validated by double loop cross-validation, with inner $Q^2 = 0.87$ and outer $Q^2 = 0.85$. RT model was validated by CV ($Q^2 = 0.89$) and by external test set comprising 30% of HIV isolates ($Q^2 = 0.86$). Q^2 is an estimate of a model's predictive performance and ranges from $-\infty$ to 1, where 1 indicates perfect prediction (see Equation 3 in Freyhult *et al.* (2005) for details on the Q^2 statistic). We also performed permutation tests to ensure that the models were not overfitted. See Lapins *et al.* (2008) and Junaid *et al.* (2010) for more details on model validation.

Table 1. Feature comparison between available servers for HIV protease and reverse transcriptase drug susceptibility predictions

| Server | URL | Algorithm | Batch | Web service | Predictions |
|--------------|--|------------------------|-------|---------------|--------------|
| HIVdb/Sierra | http://hivdb.stanford.edu/pages/webservices/ | Rules ^a | Yes | SOAP | 5 level |
| HIV-GRADE | http://www.hiv-grade.de/cms/grade/ | Rules ^b | Yes | No | 4-5 level |
| Geno2Pheno | http://www.geno2pheno.org/ | Support vector machine | Max 8 | No | Quantitative |
| PIRSpred | http://protinfo.compbio.washington.edu/pirspred/ | Linear regression | No | No | Quantitative |
| HIVDRC | http://www.hivdrc.org/services/ | Proteochemometrics | Yes | REST and XMPP | Quantitative |

^aHIVDB 6.0.7.^bGRADE 12/2008, ANRS 07/2008, HIVDB 5.1.2, Rega v8.0.1.

Batch indicates if it is possible to submit several sequences at the same time.

3 IMPLEMENTATION

The two services were implemented as Web services, available from the server *ws1.bmc.uu.se*. The services accept a mutated PI or NRTI sequence as input (in plain text or FASTA format), and outputs the predicted susceptibility values. The *HIV protease drug susceptibility prediction service*, and the *HIV-RT drug susceptibility prediction service* were implemented as XMPP services (Wagener *et al.*, 2009). A REST interface was also added to both services, to provide an additional way to invoke them. Implementation details are available from the web page <http://www.hivdrc.org/webservices>. Both REST services are registered in BioCatalogue (Bhagat *et al.*, 2010).

Web pages were established to allow for entering sequences with one or many mutations, and submit for prediction of susceptibility by the two Web services with results visualized in a chart (see Fig. 1). A summarizing feature comparison between servers for susceptibility prediction are available in Table 1. Notably, only the HIV Drug Research Centre (HIVDRC) services are available as REST and XMPP services, making them the most interoperable services available for HIV susceptibility prediction. EuResist (Zazzi *et al.*, 2011) was not included in the list as it does not predict virus susceptibility to individual drugs but the probability of success of common treatment regimens (combinations of 2–4 drugs) after 8 weeks of treatment.

A set of plugins were also constructed to integrate the HIVDRC services into the graphical platform Bioclipse (Spjuth *et al.*, 2007). The services can be accessed both from the graphical user interface and from scripts (Spjuth *et al.*, 2009).

4 CONCLUSION

We here present two Web services that predict drug susceptibility for PIs and NRTIs from the mutation patterns of different HIV strains' genome sequences. The services are based on a comprehensive analysis of HIV protease and RT inhibitors, using data from all drugs in each model, resulting in a proteochemometric model with excellent predictive ability. The implementation as Web services allows for programmatic access to the services, and for inclusion in workflows.

The binding sites for NNRTIs and NRTIs are different so these compounds cannot be incorporated into one proteochemometric model. A separate proteochemometric model for predictions of NNRTI drugs is in development.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the BMC Computing department for hosting the Web services on the server *ws1.bmc.uu.se*.

Funding: The Swedish Research Council (04X-05957); Uppsala University (KoF 07).

Conflict of Interest: none declared.

REFERENCES

- Bhagat, J. *et al.* (2010) Biocatalogue: a universal catalogue of web services for the life sciences. *Nucleic Acids Res.*, **38** (Suppl), W689–W694.
- Beerenwinkel, N. *et al.* (2003) Geno2pheno: Estimating phenotypic drug resistance from HIV-1 genotypes. *Nucleic Acids Res.*, **31**, 3850–3855.
- Freyhult, E. *et al.* (2005) Unbiased descriptor and parameter selection confirms the potential of proteochemometric modelling. *BMC Bioinformatics*, **6**, 50.
- Junaid, M. *et al.* (2010) Proteochemometric Modeling of the Susceptibility of Mutated Variants of the HIV-1 Virus to Reverse Transcriptase Inhibitors. *PLoS ONE*, **5**, e14353.
- Kontijevskis, A. *et al.* (2008) Generalized proteochemometric model of multiple cytochrome p450 enzymes and their inhibitors. *J. Chem. Inf. Model*, **48**, 1840–1850.
- Lapins, M. *et al.* (2008) Proteochemometric modeling of HIV protease susceptibility. *BMC Bioinformatics*, **9**, 181.
- Liu, T.F. and Shafer, R.W. (2006) Web resources for HIV type 1 genotypic-resistance test interpretation. *Clin. Infect. Dis.*, **42**, 1608–1618.
- Spjuth, O. *et al.* (2009) Bioclipse 2: A scriptable integration platform for the life sciences. *BMC Bioinformatics*, **10**, 397.
- Spjuth, O. *et al.* (2007) Bioclipse: an open source workbench for chemo- and bioinformatics. *BMC Bioinformatics*, **8**, 59.
- Wagener, J. *et al.* (2009) XMPP for cloud computing in bioinformatics supporting discovery and invocation of asynchronous web services. *BMC Bioinformatics*, **10**, 279.
- Zazzi, M. *et al.* (2011) Prediction of response to antiretroviral therapy by human experts and by the euresist data-driven expert system (the eve study). *HIV Med.*, **12**, 211–218.