Letter

PIRSpred: a web server for reliable HIV-1 protein-inhibitor resistance/susceptibility prediction

Ekachai Jenwitheesuk, Kai Wang, John E. Mittler and Ram Samudrala

Department of Microbiology, University of Washington School of Medicine, Seattle, Washington, WA 98195, USA

We introduce a publicly available webserver, PIRSpred (http://protinfo.compbio.washington.edu/pirspred/), for accurate human immunodeficiency virus type 1 (HIV-1) genotypic resistance/susceptibility prediction. The server accepts mutant HIV-1 protease or reverse transcriptase (RT) enzyme sequences as input and predicts resistance/susceptibility to several FDA-approved inhibitors using three approaches: linear regression, docking with dynamics, and their consensus. The predictions made by the server outperform other publicly available HIV-1 genotypic interpretation algorithms, and can be used efficiently and economically in a clinical setting to guide treatment against HIV-1 infection.

The development of antiretroviral drug resistance severely impedes treatment of HIV-1-infected patients. The International AIDS Society-USA [1] has recommended that antiretroviral drug regimens be adjusted for patients who fail to suppress HIV-1 viral load to <200 copies per ml. Genotypic testing provides viral mutation information that might suggest new drug regimens to suppress the growth of drug-resistant variants [2]. However, interpreting genotypic results can be problematic owing to complex mutational patterns [3].

To overcome this problem, several groups have developed knowledge-based techniques, including linear regression, neural networks, decision trees, support vector machines and rule-based methods, to interpret HIV-1 genotypic data (refer to Ref. [4] and references therein). These methods generally perform well, but need large viral phenotype-genotype training sets and often make discordant phenotypic predictions [5]. The accuracies are particularly low for newly approved drugs because there are fewer genotype-phenotype results available for learning.

Physics-based approaches that evaluate the binding energy of protein–inhibitor complexes have successfully identified key mutations in parts of the enzyme that confer drug resistance and show a high correlation with the inhibitory concentration (IC_{50}) values determined by phenotypic tests (refer to Ref. [6] and references therein). This approach achieves higher accuracy than the knowledge-based approaches for newly approved drugs because it does not require a large phenotype-genotype database for learning, but is limited when considering mutations that are not located at the active site where the drug binds.

Here, we introduce the webserver PIRSpred, which uses both knowledge- and physics-based approaches, to perform accurate HIV-1 genotypic interpretations. An input mutant sequence can be submitted to the server as a complete amino acid sequence or as a list of mutations relative to the wild-type HIV-1 subtype B consensus sequence. Genotypic interpretation is performed by three modules: linear regression, docking with dynamics, and their consensus. The interpretation covers three classes of approved antiretroviral drugs: protease inhibitors, and nucleoside and non-nucleoside RT inhibitors. The PIRSpred result, returned to the user's e-mail address, provides both binary (resistant or susceptible) and quantitative (IC_{50} values and/or binding affinities) results and is easy to interpret. The PIRSpred server accuracy has been evaluated relative to other methods and has the best accuracy overall when compared with the HIV-1 genotypic interpretation servers that are publicly available [4,6,7] (Figure 1).

The linear regression algorithm used by PIRSpred is based on the assumption that drug resistance values can be quantitatively approximated by a simple linear model, in which each mutation contributes to drug resistance independently and quantitatively. Using datasets obtained from the Stanford HIV drug resistance database (http://hivdb.stanford.edu/) [8], we constructed drug resistance models by standard stepwise linear regression techniques to produce scores that were translated into quantitative IC_{50} estimates [4]. We evaluated the accuracy of our predictions relative to six other publicly available genotypic interpretation algorithms for seven protease inhibitors and ten RT inhibitors. In hold-one-out experiments, and in tests on an independent dataset, our algorithm had the best performance relative to other servers [4] (Figure 1).

The docking with dynamics protocol used by PIRSpred integrates protein flexibility (using molecular dynamics simulations) with an inhibitor flexible-docking technique to predict binding energies of protease–inhibitor complexes. The calculated binding energies made by this protocol were highly correlated (0.87) with experimental binding energies [9], enabling interpretation of HIV-1 genotypic data. The module generates a three-dimensional structure of the mutant protein using the x-ray crystallography structure of HIV-1 wild-type as a template, with mutant side chains constructed using the SCWRL v3.021 software [10]. The calculated binding energy between the mutant protein and the drug molecule is converted to an inhibitory constant (K_i). The sequence is considered resistant or susceptible on the basis of the fold change of the calculated K_i value relative to that of the wild-type.

Corresponding author: Samudrala, R. (ram@compbio.washington.edu).

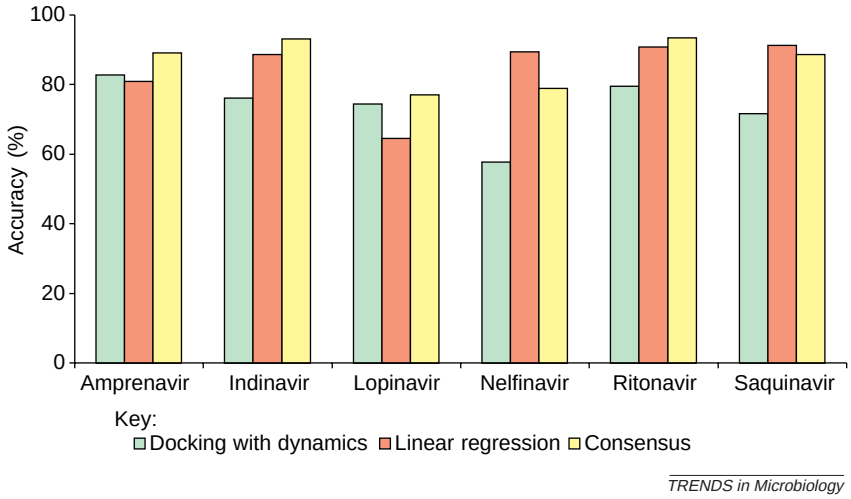

Figure 1. Comparison of the accuracy of three HIV-1 genotypic prediction methods for 1792 mutants and six FDA-approved protease inhibitors. The overall average accuracies were 73%, 84% and 86% for docking with dynamics, linear regression and their consensus, respectively. The knowledge- and physics-based approaches have different strengths and weaknesses, with the former performing well on cases where a lot of experimental genotype-phenotype data already exists, and the latter performing well on newly approved drugs. The consensus of the two produces higher overall average accuracy, indicating that drug regimen prediction based on HIV-1 genotypic data can be undertaken with higher confidence.

This physics-based approach works best on newly approved drugs (Figure 1).

The PIRSpred consensus module takes advantage of our previous finding that the consensus of complementary methods generally produces higher overall accuracies relative to the individual methods [6,7] (Figure 1). Specifically, the result is reported as resistant or susceptible when the linear regression and the docking with dynamics modules generate the same prediction. A discordant result consisting of the percent accuracy of each module is reported when the two modules generate different predictions. Consensus interpretations are especially useful for polymorphic strains with complex mutational patterns. Clinical decisions about therapeutic regimens can be undertaken with greater confidence when orthogonal algorithms make the same prediction.

Acknowledgements

This work was supported by NIH grants R21 AI52063–01, R01 HL072631 and R03 AI055394 to John E. Mittler and NSF grant DBI-0217241, NIH grant GM068152 and a Searle Scholar Award to Ram Samudrala.

References

1 Yeni, P.G. *et al*. (2004) Treatment for adult HIV infection: 2004 recommendations of the international AIDS society-U. S. A. panel. *JAMA* 292, 251–265

2 Baxter, J.D. *et al*. (2000) A randomized study of antiretroviral management based on plasma genotypic antiretroviral resistance testing in patients failing therapy. CPCRA 046 Study Team for the Terry Beirn Community Programs for Clinical Research on AIDS. *AIDS* 14, F83–F93

3 Ravela, J. *et al*. (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

4 Wang, K. *et al*. (2004) Simple linear model provides highly accurate genotypic predictions of HIV-1 drug resistance. *Antivir Ther* 9, 343–352

5 Kijak, G.H. *et al*. (2003) Discrepant results in the interpretation of HIV-1 drug resistance genotypic data among widely used algorithms. *HIV Med.* 4, 72–78

6 Jenwitheesuk, E. and Samudrala, R. (2005) Prediction of HIV-1 protease inhibitor resistance using a protein-inhibitor flexible docking approach. *Antivir Ther* 10, 157–166

7 Jenwitheesuk, E. *et al*. (2004) Improved accuracy of HIV-1 genotypic susceptibility interpretation using a consensus approach. *AIDS* 18, 1858–1859

8 Rhee, S.Y. *et al*. (2003) Human immunode.ciency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* 31, 298–303

9 Jenwitheesuk, E. and Samudrala, R. (2003) Improved prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations. *BMC Struct. Biol.* 3, 2

10 Canutescu, A.A. *et al*. (2003) A graph theory algorithm for protein side-chain prediction. *Protein Sci.* 12, 2001–2014

0966-842X/$ - see front matter
doi:10.1016/j.tim.2005.02.003