Tackling the problem of HIV drug resistance

ABSTRACT

The virally-encoded HIV-1 protease is an effective target for antiviral drugs, however, treatment for HIV infections is limited by the prevalence of drug resistant viral mutants. In this review, we describe our three-pronged approach to analyze and combat drug resistance. Understanding the molecular basis for resistance due to protease inhibitors is a key initial step in this approach. This knowledge is being employed for the design of new, improved inhibitors with high affinity for resistant mutants as well as wild type enzyme. In parallel with experimental studies of diverse mutants and inhibitory compounds, we are developing efficient algorithms to predict drug resistance phenotype from genotype data. This approach has important practical applications in the clinic where genotyping is recommended for individuals with new infections.

INTRODUCTION

Human immunodeficiency virus (HIV) is the infectious agent for the HIV/AIDS pandemic with over 35 million people infected worldwide and an estimated 34 million deaths since 1981 [1]. Antiretroviral drugs remain the primary treatment and preventative options in the absence of an effective vaccine [2]. The standard therapeutic intervention uses combinations of drugs that inhibit different stages in the viral lifecycle: cell entry and fusion, reverse transcription, integration, and maturation [3]. Individuals with high risk of infection can benefit from pre-exposure prophylaxis or PrEP [4]. The genetic diversity of the virus, which arises from the high error rate of the reverse transcriptase and the rapid replication of the virus, is the major challenge for both therapy and PrEP [5]. The viral genomes can be classified into the more common HIV type 1 and the rarer type 2, while HIV-1 comprises four groups: M (subdivided into 9 subtypes A-J), O, N, and P with distinct geographical distributions. An infected individual may contain a mixture of evolving viral strains with a total estimated population of about $10^{10}$ HIV virions. This genetic diversity and rapid selection of variants enable the virus to escape the immune system and the existence of resistant mutants limits the effectiveness of drugs [6]. Due to this serious problem, resistance testing is recommended for newly diagnosed individuals, or those failing therapy [7]. A list of current mutations associated with resistance to the antiviral drugs is available and updated regularly [8]. In addition to the problem of drug resistance, treatment of HIV infections is hampered by the inaccessible reservoirs of latent virus due to its integration into the host cell genome.

The virally-encoded protease has proved a valuable target for antiviral drugs and a model for structure-guided design of inhibitors. The mature HIV protease is an aspartic protease comprised of two 99-amino acid subunits [3]. The protease acts during viral maturation to cleave the viral Gag and Gag-Pol polyprotein precursors into the separate enzymes and structural proteins [9,10]. This process must occur in an ordered fashion and the initial step is release of the mature protease by autocatalytic proteolysis from the Gag-Pol precursor. Hence, the protease is essential for viral replication and an excellent target for antiviral inhibitors. Inhibitors of the protease will act to decrease production of the other viral enzymes, reverse transcriptase and integrase, as well as the matrix, capsid and nucleocapsid structural proteins. Most protease inhibitors are based on transition state analogs of the peptide substrates. Historically, Alexander Wlodawer’s group was a pioneer in crystallographic analysis of HIV protease with inhibitors and reported the first crystal structure of HIV protease in complex with a peptidic inhibitor [11]. Over the next years, a number of other co-crystal structures were described during the search for antiviral inhibitors [12]. Drugs that inhibit the HIV protease were first introduced in the clinic in 1995, and improved the survival of HIV-infected individuals over therapy with only reverse transcriptase inhibi-
Protease inhibitors are potent antiretroviral agents with effects on multiple stages of the lifecycle, including viral entry as well as maturation [13].

Due to the small size of the enzyme and relative ease of structural analysis, the majority of the clinical protease inhibitors have been designed with the aid of X-ray crystal structures of the protease-inhibitor complexes, as reviewed in [14]. Currently, more than a thousand crystal structures have been reported for HIV protease or its mutants in the apo form and in the presence of inhibitors. This plethora of structural information has revealed how the protease binds substrates or inhibitors and guided the design of antiviral inhibitors. In addition, the structural data have improved our understanding of the reaction mechanism of the enzyme. Reaction intermediates have been trapped in the crystals for several steps in the reaction pathway, including the quasi-stable tetrahedral intermediate [15]. More recently, our neutron crystal structures of per-deuterated HIV protease with clinical inhibitors have given important insights into the detailed geometry of protons, hydrogen bonds and water molecules in the active site [16,17].

To date, nine protease inhibitors have been approved for clinical use. The inhibitors and their clinical effects and resistance mutations are reviewed in [18]. The majority of these drugs were designed to inhibit the wild type enzyme, however, the newest drugs, tipranavir and darunavir, were designed to target drug resistant protease variants. As measured by isothermal titration calorimetry, darunavir and tipranavir show high binding affinity of 5–10 and 20 picomolar, respectively, for wild type protease, while the earlier drugs, saquinavir, indinavir, nelfinavir and amprenavir exhibit poorer binding affinities of 0.2–0.4 nM [19,20]. The most recently approved drug, darunavir, has a broader clinical application due to fewer adverse side effects compared to treatment with tipranavir [18]. Moreover, several studies have demonstrated that physiological concentrations of darunavir and saquinavir also inhibit the autocatalytic processing of the protease from its precursor [21-23]. Although alternate binding sites in the protease have been proposed as drug targets, most design efforts have focused on competitive inhibitors that bind in the active site cavity of the protease. Current strategies tackle the challenges of drug resistance and poorly accessible reservoirs of virus.

**Table 1.** Highly resistant protease variants and their binding affinity for darunavir (DRV).

<table>
<thead>
<tr>
<th>Protease</th>
<th>$K_d$ (nM)</th>
<th>Relative $K_d$</th>
<th>Amino Acid Substitutions</th>
</tr>
</thead>
</table>

Underlined bold indicates major resistance mutations from http://hivdb.stanford.edu/DR/PIResiNote.html. *Mutant selected in laboratory [33]; other mutants were identified in clinical isolates. Binding affinity measurement from [34], [35], [36].

**Molecular Basis for Resistance to Protease Inhibitors**

HIV resistance to drugs poses a critical challenge for effective therapy. The virus evolves resistant variants to protease inhibitors by balancing the need for the protease to recognize diverse cleavage sites while rejecting the inhibitors. Since the mutations are independent results of a stochastic process, coordinated mutation of both the cleavage sites and protease can only arise slowly. Analysis of the structures and activities of protease mutants has proved valuable for understanding the molecular basis for drug resistance. This knowledge can also be applied to optimize the design strategy for new inhibitors targeting resistant variants. Clinical resistance to protease inhibitors is associated with mutations in the protease gene and different drugs elicit distinct mutations [8]. The structural location of major non-polymorphic mutations associated with drug resistance is illustrated in figure 1A. The scale of the sequence variation is demonstrated by a recent comprehensive analysis of >100,000 HIV-infected individuals in 143 countries [24]. In the protease sequences, 47% of positions had one or more types of amino acid residues occurring with at least 1% prevalence. Clusters of variation were observed between residues in the surface loops of the protease and in the short helix (Fig. 1A).

Many of the initial “major” mutations alter the protease binding site for inhibitors and substrates, and often produce defects in protease activity and viral replication. Consequently, the virus evolves additional “minor” or “accessory” mutations that may act to increase the protease activity and viral fitness [25]. These mutations occur in regions distal to the inhibitor binding site. Short insertions of amino acids have been seen and may act to increase viral fitness [26]. As the virus evolves towards higher resistance, multiple mutations accumulate in the protease and also in its
cleavage sites in the precursor polyproteins [25]. The most highly resistant proteases show changes in 20 or more amino acids in different regions of the structure, as reviewed in [27]. A variety of different sets of mutations occur in protease variants from drug resistant clinical isolates. Additional mutations in HIV Gag and Pol proteins co-evolve to compensate for loss of fitness due to protease mutations [28-31]. The massive number of potential combinations of mutations poses a severe problem for predictions of resistance from sequence, as discussed in a later section.

Structural and biochemical analyses have demonstrated the changes due to individual mutations. The structural effects of the major mutations in the inhibitor binding site have been described, however, the roles of many of the accessory mutations are poorly understood. Three general effects have been observed for protease bearing a single mutation, as reviewed in [32]. 1) Mutation of residues in the binding site (D30, V32, I47, I50, V82, I84) can directly alter the protease interactions with inhibitors. 2) Mutation of residues at the dimer interface (L24I, I50V, F53L) or in the hydrophobic core (L76V) can decrease the protease stability. 3) Mutation of residues in the flexible flaps can influence the dynamics of flap opening and closing. Recent experimental studies from several groups have examined the structural and enzymatic properties of a handful of highly resistant protease variants with up to 22 mutations, as reviewed in [27]. These proteases demonstrate several orders of magnitude lower affinity for the drugs. Selected examples of protease variants with poor affinity for the clinical inhibitor, darunavir, are listed in table 1. These highly resistant proteases from clinical isolates or a laboratory selected strain (P51) [33] exhibit 35-50 nM binding affinity for darunavir or ~10,000-fold worse than the value of 5 pM for wild type enzyme [34-36]. The protease sequences contain 14 to 20 mutations, including 5 to 8 major mutations. Only mutations of L10I/F and I54V/L/M occur in all four examples, suggesting multiple evolutionary pathways lead to high level resistance.

Our studies have focused on the extremely resistant variant PR20 with 20 substitutions relative to the wild type sequence. In contrast to wild type enzyme, autoprocessing of the precursor bearing the PR20 mutations is not significantly inhibited by darunavir and saquinavir [34]. Analysis of the crystal structures of PR20 and other multiply mutated variants showed two general changes compared to the wild type enzyme. 1) The highly resistant mutants often lose interactions with inhibitors due to direct and indirect changes in the binding site [35,37]. 2) The extreme mutants exhibit highly variable flap conformations in the absence of inhibitor [37,38]. In the absence of inhibitor, dimers of PR20 have been observed in diverse symmetric and asymmetric conformations as illustrated in figure 1B [37,39]. The mutations alter the dynamics of the transition between open and closed flap conformations, shifting toward the open conformation in the absence of inhibitor [39-42]. These structural and dynamic changes can be targeted in the designs of improved inhibitors.

STRATEGY FOR DEVELOPMENT OF NEW INHIBITORS TARGETING RESISTANT MUTANTS

A comprehensive description of diverse compounds, including natural products, which inhibit HIV-1 protease, is presented in [43]. The general design strategy pursued by our colleague, Arun Ghosh, is to introduce new interactions of the inhibitors with conserved regions of the protease dimer, in particular the main chain [44]. The potent antiviral inhibitor, darunavir, was designed based on this strategy with a bis-tetrahydrofuran (bis-THF) group at P2, and confirmed to form additional hydrogen bond interactions with the protease main chain [45,46]. Following this success, evaluation of a variety of new chemical substituents in the darunavir scaffold resulted in a series of antiviral inhibitors, described in [44]. The most potent of these inhibitors have been assessed against protease variants bearing single and multiple substitutions, as summarized in [27]. Crystal structures of inhibitor-protease complexes were solved and analyzed to understand the enzyme inhibition data and improve the design strategy.

Knowledge of the structural changes in the highly resistant variants such as PR20 has led to new insights for inhibitor designs [37]. Compounds were designed with 1) large groups at P2 to better fit in the enlarged S2 subsite, and 2) substituents providing interactions with the flexible flaps. Examples of inhibitors showing improved
efficacy on resistant virus are shown in figure 2. These compounds were designed with larger substituents at P2 compared to bis-THF in darunavir and introduce new interactions with the wild type protease. We have employed the highly resistant PR20 variant to assess the binding of inhibitors [47,48]. Structural and calorimetric studies have identified antiviral inhibitors with higher binding affinity than darunavir for PR20 (Tab. 2). GRL0519 with tris-THF at P2 instead of the bis-THF of darunavir, has similar ~40 nM binding affinity for PR20. However, GRL04410 with an oxymethyl substituent on the bis-THF moiety and GRL5010 with gem-difluoro modification exhibit binding affinity for PR20 of, respectively, 10 and 20-fold better than darunavir [48]. Crystallographic analysis demonstrated that the hydrogen bond interactions of these inhibitors with protease are conserved in wild type enzyme and the PR20 structure. Inhibitors introduce new interactions with the flexible flaps, such as the unusual halide interactions with the carbonyl oxygen of Gly48 seen for GRL5010 (Fig. 2E). These novel antiviral inhibitors are promising candidates for future pharmacological development targeting highly resistant viral strains. Moreover, incorporation of fluorine in GRL5010 increases the lipid solubility and shows improved penetration of the blood-brain barrier [49]. Development of a drug based on this compound would help to eradicate the viral reservoirs in the central nervous system.

PREDICTION OF RESISTANCE FROM SEQUENCE

Currently, genotyping is recommended for new HIV infections or for individuals failing therapy in order to identify the presence of resistant mutations and guide the choice of drugs. Genotype assays for drug resistance are preferred over phenotype assays due to the advantages in terms of speed and cost [50]. The ability to rapidly and affordably sequence HIV from infected individuals opens the door to both the practical question of predicting drug resistance prior to therapy and the theoretical problem of understanding sequence and structural evolution of the virus under drug selection. Machine learning is a good method for a computational approach to these problems. Drug resistance can be predicted from genotype data by two general techniques: rule-based genotype interpretation systems and machine learning algorithms [51].

<table>
<thead>
<tr>
<th>Compound</th>
<th>Ki</th>
<th>IC50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Darunavir</td>
<td>16pM</td>
<td>4nM</td>
</tr>
<tr>
<td>GRL0519</td>
<td>5.9pM</td>
<td>1.8nM</td>
</tr>
<tr>
<td>GRL4410</td>
<td>2.9pM</td>
<td>2.4nM</td>
</tr>
</tbody>
</table>

Figure 2. Chemical structure, inhibition value ($K_i$) and antiviral effect (IC$_{50}$) of darunavir and selected new inhibitors. A–D) darunavir, GRL0519, GRL4410 and GRL5010. E) Halide interactions of P2 group of GRL5010 (yellow sticks) with Gly48 in protease flap (gray sticks).

Table 2. Binding affinity ($K_d$, nM) of selected antiviral inhibitors to PR20 and wild type protease.

<table>
<thead>
<tr>
<th></th>
<th>DRV</th>
<th>GRL0519</th>
<th>GRL04410</th>
<th>GRL5010</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR</td>
<td>0.005</td>
<td>0.005</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td>PR20</td>
<td>41</td>
<td>39.5</td>
<td>4.3</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Information in table 2 was taken from [34,47,48].
Supervised machine learning takes a labeled set of features, such as a sequence and activity data, and builds a computational model that reproduces the correlation between the features and labels [52]. The art of machine learning is in developing approaches that select meaningful features while excluding spurious features. With drug resistance due to HIV protease variants, the obvious set of features is the sequence of the protease coupled either to the relative drug resistance or a thresholded resistant/non-resistant label.

In the absence of structural data, machine learning and rule based approaches typically have predictive accuracies in the range of 60-70% [51,53]. While the overall accuracy is less than ideal, the individual tools can be remarkably consistent, for example classifying the same set of mutations as resistant with a 95% reproducibility [51]. This shows that the problem of predicting is well posed for machine learning, but that the set of features is insufficient when only sequence data are used. When structural data are included as features in the training set, the accuracy jumps into the 90-95% range independent of the type of machine learning algorithm used [53].

The limitation of using sequence data on its own is immediately obvious to a structural biologist. Proteins are not a linear set of letters, but are folded into complex and beautiful three-dimensional structures. Residues that are far apart in the linear sequence may actually be close together in space. The three dimensional context of the mutations is lost if only the sequence is used as a representation. A single mutation might contribute to drug resistance when its spatial neighbors are of one kind, but could be neutral or increase sensitivity with other kinds of neighbors. This introduces an apparent non-linearity when only sequence data are used.

Bose et al. [54] studied representations that could efficiently encode structure and sequence for machine learning. The encoding of the structure must be translationally and rotationally invariant because there is no privileged reference frame in the biological system. This immediately leads to using either distance measures or graph representations for the structure. Distances as features are problematic because it is difficult to define an automatic way to select which distances to use. Additionally, experimental or modeling errors in the distances mean that the data are inherently fuzzy. Graphs, which simply state that two atoms or residues are in contact with each other by some criterion, are less sensitive to errors, and can be rigorously defined. Bose et al. [54] examined several types of graphs and algorithms to reduce the graphs to a constant-sized data point. They found that Delaunay triangulation was the best graph for the problems they studied. They also found the graph could be reduced to an upper triangular matrix by summing over the kinds of amino acids on each end of an arc.

Yu et al. [53,55] applied these encodings to genotype/phenotype data for drug resistance of HIV protease and reverse transcriptase mutants. These results showed significantly higher classification accuracy than the purely sequence based approaches. In addition, the unified encoding of sequence and structure can be used with regression analysis to predict the magnitude of resistance with high accuracy [55,56].

In the next stage of this study, mutants representing common features of high level resistance were selected from the data [56]. This selection was designed to give a tractable number of mutants for further analysis by biochemical and biophysical experiments. Combining mutants that represented high resistance to more than a single inhibitor resulted in a single sequence with high resistance to 6 drugs. The mutant with this sequence was designated PR$_{S17}$ due to the presence of 17 substitutions relative to a standard wild type protease. PR$_{S17}$ was verified to exhibit poor binding to six tested clinical inhibitors, 50 to >10,000-fold worse than wild type enzyme [36]. Therefore, PR$_{S17}$ will be added to our list of extremely resistant mutants for evaluation of new antiviral inhibitors.

**CONCLUDING REMARKS AND FUTURE PERSPECTIVES**

HIV infections and drug resistance are likely to continue as a problem in the absence of an effective vaccine, due to the high genetic variation, rapid turnover and existence of poorly accessible reservoirs of virus. This emphasizes the critical need for the development of targeted treatment based on genotype data and new antiretroviral drugs for both therapy and pre-exposure prophylaxis. Our studies of HIV protease tackle the challenge of drug resistance on several fronts by: 1) developing new algorithms to predict resistance from genotype data, 2) elucidating the molecular basis for resistance, and 3) incorporating this knowledge in the design strategy for novel antiviral inhibitors. Based on our studies of HIV protease and those of other groups, we propose a new paradigm for drug resistance. High level resistance to drugs does not rely on a handful of major mutations, but rather requires the coordinated effects of multiple substitutions to remodel the protease and its substrates. This new paradigm must be addressed in the interpretation of genotype data and in the design of antiviral inhibitors.

**REFERENCES**


34. Huang L, Li Y, Chen C (2011) Flexible catalytic site conformations implied in modulation of HIV-1 protease autoprocessing reactions. Retrovirology 8: 79


Pokonać problem lekooporności wirusa HIV

Irene T. Weber¹,², Robert W. Harrison²

¹Department of Biology, Georgia State University, Atlanta, GA, USA
²Department of Computer Science, Georgia State University, Atlanta, GA, USA

¹e-mail: iweber@gsu.edu

Słowa kluczowe: HIV/AIDS, krystalografia rentgenowska, enzymy, bioinformatyczne systemy uczące się

STRESZCZENIE

Kodowana przez wirusowy materiał genetyczny proteaza HIV-1 jest skutecznym celem dla leków antywirusowych. Leczenie zakażeń HIV jest jednak znacznie utrudnione przez powszechne występowanie lekoopornych mutantów wirusa. W niniejszym artykule przeglądowym autorzy opisują stosowane przez nich kompleksowe podejście mające na celu analizę i zwalczenie lekooporności. Zrozumienie molekularnego podłoża oporności na inhibitory proteaz to kluczowy, pierwszy krok tego procesu. Następnie, zdobyta wiedza jest wykorzystywana podczas projektowania nowych, udoskonalonych inhibitorów o wysokim powinowactwie do enzymu występującego u opornych mutantów, jak i u dzikiego typu wirusa. Równolegle do badań nad różnorodnymi mutantami i inhibitormi trwa opracowywanie skutecznych algorytmów umożliwiających przewidywanie fenotypu oporności wirusa na podstawie informacji o jego genotypie. Podejście to ma ważne zastosowanie praktyczne w klinycystyce, gdzie zaleca się przeprowadzenie genotypowania wirusa u pacjentów z nowymi zakażeniami.