DDT Vol. 7, No. 24 December 2002

Structural pharmacogenomics, drug resistance and the design of anti-infective super-drugs

Edward T. Maggio, Mark Shenderovich, Ron Kagan, Dean Goddette and Kal Ramnarayan

Large-scale comparative analysis of drug-target polymorphism structures enables the rational design of next generation 'super drugs' – drugs that are less prone to development of drug resistance or that work for the largest possible fraction of the patient population. Furthermore, knowledge of the drug-target-shape repertoire that exists within the patient population enables predictions of likely clinical trial outcomes and response rates for drug efficacy. This gives information on the optimal drug candidates before the initiation of clinical trials. The economic impact of incorporating pharmacogenomics insights early on in the drug discovery process will be substantial and will afford significant competitive advantages to companies that successfully incorporate this technology.

Edward T. Maggio*
Mark Shenderovich
Dean Goddette
Kal Ramnarayan
Structural Bioinformatics
10929 Technology Place
San Diego, CA 92127, USA
*e-mail:
emaggio@strubix.com
Ron Kagan
Dept of Infectious Diseases
Quest Diagnostics
San Juan Capistrano
CA 92690, USA

▼ Structural variants (polymorphisms) exist for all human, as well as for all infectious disease, protein drug-targets. The distribution of drug-target polymorphisms within the worldwide human population is, practically speaking, more or less static, and most drugs typically work for 40–70% of patients. In the case of infectious diseases, however, development of drug resistance is a dynamic process arising in response to high mutation rates and drug selection pressures [1]. Highly active antiretroviral therapy for HIV-1, although resulting in dramatic suppression of viral replication, has also furnished a strong selective force for the emergence of drug-resistant variants [1,2]. Here, the distribution of polymorphisms can be extreme. For example, examination of the HIV-1 reverse transcriptase (RT) and HIV-1 protease structural variant databases (Variome™ modules) established by Structural Bioinformatics (SBI; http://www.strubix.com) and Quest Diagnostics (http://www.questdaignostics.com), each containing >50,000 patient specific structures, has revealed that no two patients have exactly the same sequence – each individual patient exhibits a unique sequence or structural variant for these drug targets. Because a principal component determining drug efficacy is the distribution of drug-target structural variants within the patient population [3–5], the ability to systematically analyze comparative structural information from databases of polymorphism structure offers a means to rationally design a new generation of 'super drugs' – drugs that are less prone, or not subject, to the development of drug resistance. In addition, these drugs should work with the largest fraction of the patient population, which is characteristic property of many 'best in class' drugs.

So-called 'best in class' drugs often command a 50% or more market share, while the remaining market might by divided among as many as half a dozen or more pharmaceutical company competitors. Therefore, the economic impact of incorporating pharmacogenomics insights early on in the drug discovery and optimization processes can afford a tremendous competitive advantage to companies that recognize this potential.

At present, there is no way of predicting the response rate for drug efficacy before clinical trials or to influence the outcome of such trials in a positive manner during the drug discovery and development processes. However, as mentioned previously, by incorporating protein structural knowledge at the drug discovery and optimization stages, synthesis of drug candidates can be focused toward compounds that will 'satisfy' the binding sites or

PII: S1359-6446(02)02514-X

consensus binding elements of the largest fraction of structural variants, and will thus be likely to exert efficacy in the largest possible fraction of the patient population. Additionally, foreknowledge of the drug-target-shape repertoire distribution that pre-exists in the patient population can enable certain predictions to be made concerning likely clinical trial outcomes. This gives information on the selection of optimal drug candidates from the available candidates before clinical trials commence, thus further maximizing the commercial impact of drug discovery efforts.

HIV-RT and HIV protease

In June 1999, SBI entered into a strategic alliance with Quest Diagnostics for the application of structural pharmacogenomics to clinical testing and pharmaceutical research. Quest Diagnostics generates proprietary clinical sequences for HIV-RT and HIV protease that SBI uses to create structural variant databases. The sequence data are current, clinically relevant, and are associated with temporal and anonymized demographic and clinical information. SBI applies computational proteomics expertise to accurately model each of the structural variants, compute their interactions with anti-HIV drugs, and organize the structural and anonymized patient data in a relational database. The large sample size provides a representative spectrum of HIV-1 mutations that can provide meaningful statistics. The database currently contains >56,000 HIV-1 protease and >70,000 HIV-1 RT sequences and structures.

Sampling is current and on-going and ~30,000 patient sequences are added annually for each protein. The sequence data deposited in the Variome™ database modules is generated in Quest Diagnostics' Clinical Laboratory Improvement Act (CLIA™) and ISO9001 certified laboratory, and as such, is uniquely of high quality, uniform and reproducible. The goal of creating these HIV-1 structural variant databases is to provide novel insights through comparative analysis of drug-target polymorphisms on a large scale.

Variome™ modules can be easily mined for information to answer key questions for the pharmaceutical industry (see Box 1 for some examples). The Variome™ database provides us with quantitative answers to questions that we might have thought impractical or even impossible to ask. Some examples of the kind of information available through large-scale structural proteomics analysis are presented below.

Data on the rise and fall of HIV-1 resistant mutants for any selected time period from early 1998 to the present can be tracked in the database. Figure 1 shows the percentage of resistant HIV-1 strains detected over the 51-month period beginning in early 1998. The most salient feature during this timeframe is the gradual decline in nucleoside RT inhibitor (NRTI) resistance, protease inhibitor (PtdIns) resistance, and multi-drug resistance (MDR) after they peaked in early 1999. By contrast, the level of non-nucleoside RT inhibitor (NNRTI) resistance has not shown the same post-1999 decline and remains at 40%. Figure 2 shows the trends in resistance to nevirapine, a non-nucleoside RT inhibitor, over the same 51-month period, determined from the Variome™ RT database module. This data can be further differentiated by regional, national or international segmentation as required.

Box 1. Some important questions that are answerable through structural pharmacogenomics

1. What are the structural features of our disease target that don't mutate?
2. What are the common structural features of the binding pocket for the greatest fraction of the patient population?
3. How do our current candidate drug molecules relate to the invariant portions of the drug target structures?
4. Which of our drug candidates binds to drug target in the largest fraction of the patient population?
5. Will the use of our drug vary by geographical location, age or gender?
6. What are the structural exclusions for our drug? Which mutations are likely to block our drug and against which mutations is our drug most effective?
7. Which drug candidate should we advance to clinical trials? Which is the best backup compound?
8. Which specific target structural variant (polymorphism) from the many possibilities should we select for use in HTS?
9. How can we vary compound structure to increase activity?
10. What modifications can we make to our drug to reduce or eliminate drug resistance in the largest fraction of the polymorphism universe?
11. What trends in target structure are emerging for a particular infectious disease target?
12. Against which mutations is our drug effective, where our competitor's drug is not?
13. In what percent of patients will our drug and our competitor's drugs be effective today?
14. Is our drug or our competitor's drug encountering increasing or decreasing resistance?

Protein structural insights

Variome™ modules provide a statistical distribution of the drug-target-shape repertoire within the patient population. The database of structures enables researchers to identify

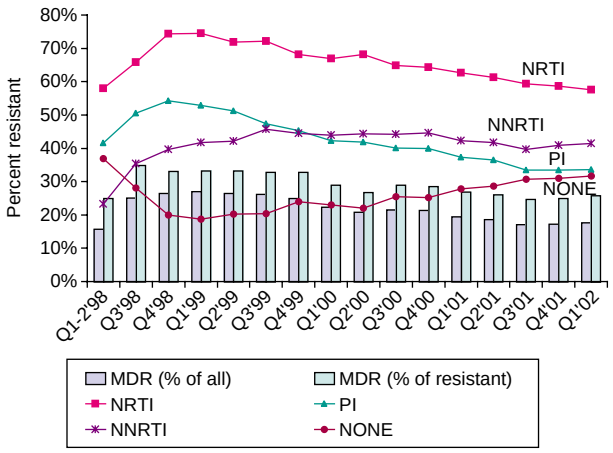

Figure 1. Percentage of HIV-1 strains detected over a 51-month period resistant to protease inhibitors (PI), non-nucleoside reverse transcriptase inhibitors (NNRTI), or nucleoside reverse transcriptase inhibitors (NRTI), along with the percent showing no resistance (NONE). Multi-drug resistance (MDR), for all patients or for resistant patients only, is represented in bar graph format.

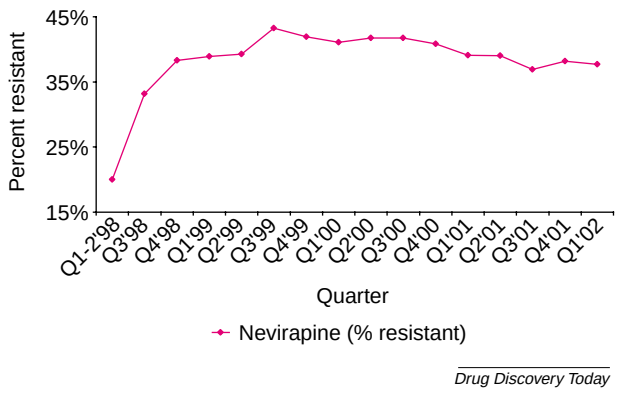

Figure 2. Trend for the development of Nevirapine® genotype-associated resistance in 47,000 patient samples present in the Variome® RT database module over a period of approximately four years.

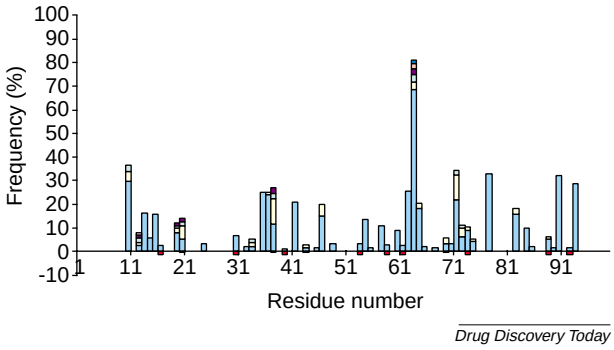

Figure 3. A plot of the mutation frequency against the amino acid position in HIV protease (10,591 patient samples). It is clearly seen why drug resistance is such a severe clinical problem. From the sequence variation in ~10,000 HIV infected individuals, within the particular timeframe sampled here, we can see that, many amino acids mutated with frequencies of 30, 20 or 10%. However, there are some amino acids that are seen to have just 0–1% mutation frequency. (50/99 residues with mutation frequencies >1%; 18/99 residues mutate to more than one amino acid identity.)

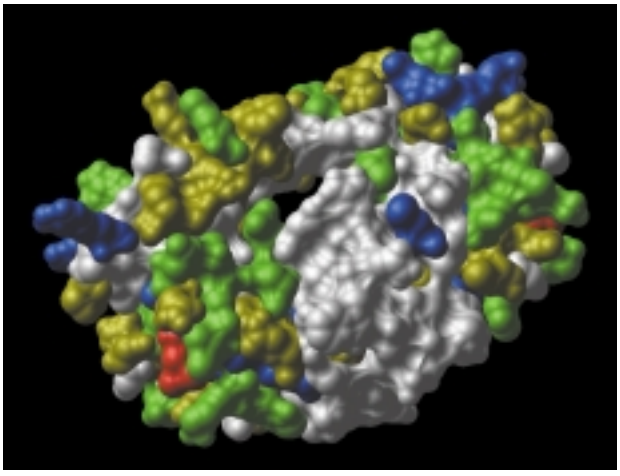

Figure 4. Point mutation frequency for HIV-1 protease mapped on to the surface. Areas shown in white exhibit 0–1% mutation frequency. Areas in red exhibit >50% mutation frequency. Yellow, green and blue areas display residues with mutation frequencies less than 5, 20 and 50%, respectively.

the structurally conserved and variable regions in infectious disease targets. Figures 3 and 4 illustrate the power of comparative analysis and data extraction from large numbers of protein structures simultaneously. For example, if we look at the mutation frequency at each position within the amino acid sequence of HIV protease (Fig. 3), we see why drug resistance is such a severe clinical problem. Looking at the protease sequences from >10,000 HIV infected individuals, within the particular timeframe sampled here, we can see that many amino acids mutated with frequencies of 10–30%.

At the same time, some amino acids are seen to have 0–1% mutation frequency. These conserved positions appear spread across the entire sequence. However, when we look at mutation frequency in three dimensions mapped back on to the protein surface (Fig. 4), the conserved residues are not spread out at all. In fact, they form patches, or islands, of conserved 3D structure. Conserved areas, shown in white, have 0–1% mutation frequency and highly mutating areas are shown in other colors with increasing mutation frequency represented from yellow to green to

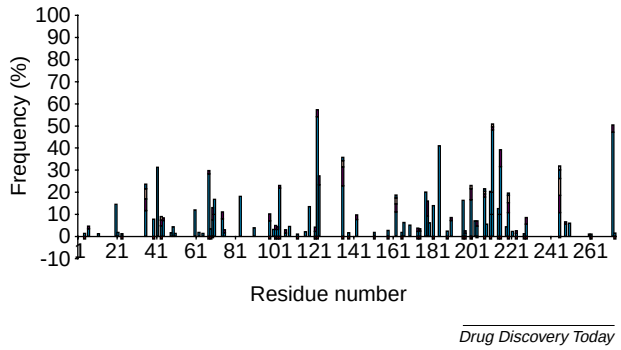

Figure 5. Point mutation frequency for HIV-1 reverse transcriptase (11,145 patient samples). 102 of 273 residues exhibit mutation frequencies >1%; 35 of 273 residues mutate to more than one amino acid identity.

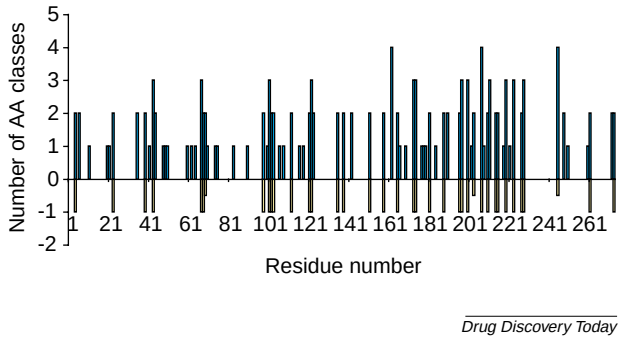

Figure 7. The number of amino acid classes allowed at each residue position for HIV-1 reverse transcriptase. 38 positions (flagged below the X-axis) frequently exhibit mutation to an amino acid class different from the original class.

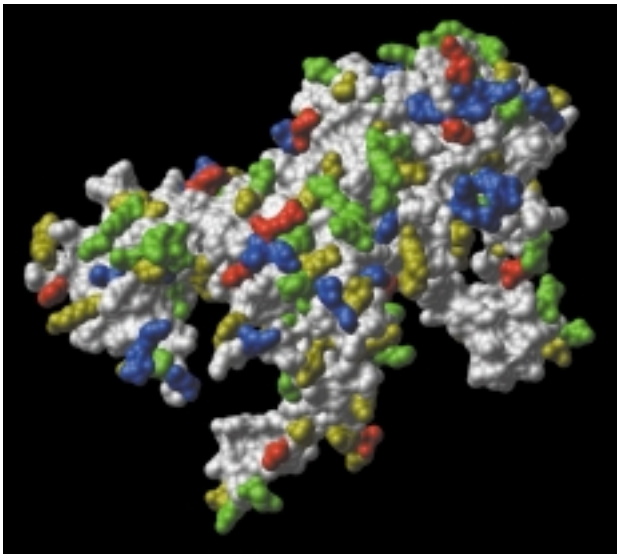

Figure 6. Point mutation frequency for HIV-1 reverse transcriptase mapped on to the surface. Areas shown in white exhibit 0–1% mutation frequency. Areas in red exhibit >50% mutation frequency. Yellow, green and blue areas correspond to frequencies less than 3, 10 and 25%, respectively.

red. Conservation of these patches in >10,000 individuals is a dramatic finding – this is the Achilles' heel of HIV-1 protease. The picture can be further clarified when temporal changes in the mutating residues are analyzed. Introduction of new drugs into the patient population following regulatory approval causes some shifts in the distribution of mutations. However, certain residues remain unchanged even under the selection pressure of multiple, structurally distinct, drugs. In the case of HIV-1 protease, there are three conserved (white) patches (two more, not shown, occur on the back of the 3D image) that HIV can not afford to mutate without losing viability. Thus, these areas present the locations and orientations of the structural element with which the next generation of anti-retroviral drugs should be designed to interact and bind, to avoid or suppress development of drug resistance. Obviously, this concept has broad applicability to many, perhaps most, infectious disease drug targets. Figures 5 and 6 show a similar analysis for HIV-RT.

Drilling down further into the types and nature of changes that occur as drug-target mutations accumulate, either naturally or as a result of drug selection pressures, provides further insight into what is likely to be productive or unproductive in modifying the chemical structure of drug leads. For example, in the case of RT, at certain positions within the sequence, mutations outside of the original chemical class of aminoacyl side-chain (i.e. hydrophobic, hydrophillic or charged) are not permitted. Other residues permit one, two, three or even four different classes of side chains upon examination of observable mutations across the database (Fig. 7). An even finer degree of structural discrimination can be observed in that certain positions within the amino acid sequence permit mutations that flip between a small number of specific amino aminoacyl sidechains. The identity of the allowed aminoacyl side chains defines the available steric volume for chemical modification of a drug lead and could also define other permitted or excluded chemical or structural characteristics. Because this information has been gathered from a statistically large sampling of naturally occurring mutations (>50,000 for each target), accumulated over an extended timeframe and under evolving drug selection pressures as each new drug was introduced, the 'freedom-space' of individual side chains to mutate has been extensively sampled. As a result, the structural conclusions should remain valid as the virus continues to mutate. This information can provide useful

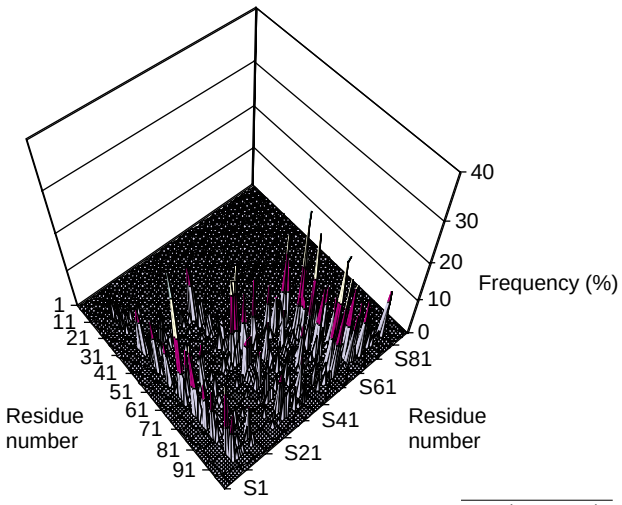

Figure 8. HIVPR pair mutation. Point mutations induce selection pressures of their own. The paired mutation frequency derived from analysis of a 10,591 unique-structure subset of SBI's Variome™ HIV protease structural variant database shows that residues co-mutate with defined statistical frequencies.

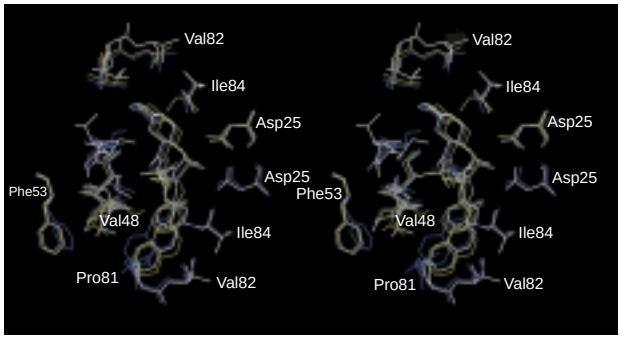

Figure 9. Superimposed stereo views of the X-ray crystal structure (yellow; from Ref. [14]) and the SBI model (white) of saquinavir complexed to the G48V HIV-1 protease variant. The crystal structure of G48 SQV complex is given in blue for comparison. The model correctly predicts the position and the conformation of the V48 side chain, the displacement of the F53 side chain, and movement of the SQV quinoline ring caused by the G48V substitution [9].

insights in classical medicinal chemistry optimization. It could also be used in guiding or focusing the development of combinatorial libraries for further optimization, as well as for expansion, or even circumvention, of structure-related composition of matter intellectual property.

Because each single mutation exerts a greater or lesser effect on protein structure and function, a first mutation could provide a selection pressure for a second mutation that acts in a compensatory fashion to re-establish a structural or physical feature diminished as a result of the first mutation [3]. Indeed, when the paired mutation frequencies for ~10,000 structural variants of HIV-1 protease are examined, this is exactly what is observed [6]. Numerous paired mutations become evident as represented by the peaks in Figure 8.

Prediction of clinical outcomes

Another application of structural pharmacogenomics is the rapid quantitative prediction of drug binding, and thus, drug resistance phenotyping. Although clinicians still largely rely on retrospective tables that correlate historical drug resistance with previously observed mutations [7,8], the computational pharmacoproteomics method described by Shendrovich *et al.* [9,10] is truly prospective in nature. This means that newly arising mutations could be immediately characterized without the need to develop 6–12 months of historical laboratory drug resistance phenotyping observations [9]. Table-based algorithm methods must also be calibrated to the relevant clinical population because mutations vary geographically [5]. Laboratory methods, effectively the 'gold standard' in drug resistance phenotyping, can pick up newly appearing mutations; however, these are slow and typically take 3–6 weeks to complete [11–13].

In addition to rapid patient-specific drug-resistance phenotyping for the management of antiretroviral therapy, structural pharmacogenomics can be used for the prediction of clinical trial outcomes. By computationally analyzing the interaction of a putative drug with a sampling of the polymorphism structural repertoire that a new drug will encounter in the clinic, it is possible to predict the binding effectiveness of a new drug before the initiation of expensive clinical evaluation in patients. This is a straightforward application and extension of the recently reported successful computational prediction of drug resistance phenotypes for HIV-1 protease on a patient-by-patient basis [9]. The interaction and movements of side-chains in individual polymorphism structures in response to drug binding can be computed. For example, comparison of the computed structural changes upon drug binding with the observed changes in the corresponding X-ray crystallographic structure [14] of the saquinavir–HIV protease complex is shown in Fig. 9.

The sensitivity and specificity for predicting drug resistance phenotypes on a patient-by-patient basis for various Food and Drug Administration (FDA; http://www.fda.gov) approved drugs, based upon the computed relative energy (ΔE_{bind}) of interaction of each drug with each patient's unique drug-target polymorphism structure, are summarized in Table 1. The numbers appear to be quite good in comparison to what is achievable in the laboratory at present.

Table 1. Comparison of computational drug resistance phenotype with laboratory phenotyping*

	ΔE_{bind} Cutoff values (kcal/mol)					
Protease Inhibitor	**Sensitive below**	**Resistant above**	**Sensitivity**	**Specificity**	**Kappa2**	**p**
Amprenivir®	0.7	1.4	86.7%	100%	0.907	<0.0001
Indinavir®	0.6	1.5	94.1%	100%	0.958	<0.0001
Nelfinavir®	0.7	1.0	60.6%	96.8%	0.567	<0.0001
Ritonavir®	0.7	1.4	100%	84.1%	0.754	<0.0001
Saquinovir®	0.6	1.1	68.4%	100%	0.752	<0.0001
Lopinavir®	0.3	0.7	100%	83%	0.755	<0.0001

*Virologic PhenoSense® (Virologic, http://www.virologic.com).

Kappa is a measure of inter-assay agreement: kappa >0.75: excellent agreement; 0.4< kappa <0.75: good agreement; kappa <0.4: poor agreement [9].

Nevirapine© and Tipranavir© (Boehringer Ingelheim, http://www.boehringer-ingelheim.com), Saquinivair® (Roche, http://www.roche.com), Amprenivir® (GlaxoWellcome, http://www.gsk.com), Indinavir® (Merck, http://www.merck.com), Nelfinavir® (Pfizer, http://www.pfizer.com), LPV (Lopinavir®) and Ritonavir® (Abbot, http://www.abbot.com).

However, perhaps what is most important, with respect to increased efficiency and speed in drug design, is that answers can be generated overnight rather than in 3–6 weeks (the typical timeframe for laboratory drug-resistance phenotyping).

Another way of looking at the interaction of putative drugs or drug leads with a set of drug target polymorphisms is to dock the molecules to the individual polymorphisms and measure the distance of the drug to each residue within the protein. Such an analysis for five different protease inhibitors is shown in Fig. 10. The mutation frequency profile is superimposed at the bottom of the graph and this analysis clearly identifies those aminoacyl side chains that mutate with a high frequency and are in close proximity to the bound drug, those that are proximal and stable, those that are distal and mutate, and those that are distal and are stable.

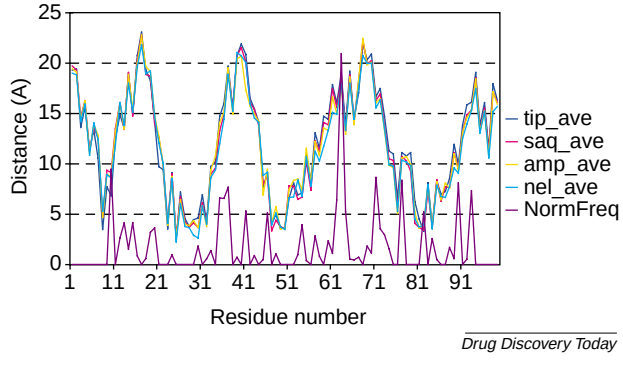

Figure 10. A plot of average ligand–protein distances versus normalized mutation frequency reveals regions in close contact with drugs that are also likely to mutate.

The Variome™ structural pharmacogenomics technology has broad applications in the rational design of highly effective infectious disease therapies – bacteria as well as viruses, drugs as well as vaccines – that offer the prospect of stable efficacy in the face of drug selection pressure. In addition to naturally occurring infectious agents, it is clear that there is significant value in biodefense related applications in rational drug and vaccine design, and in threat assessment and prediction. Box 2 gives a partial list of useful questions and potential insights obtainable from structural pharmacogenomics analyses.

Box 2. Useful questions in analyzing essential drug target: drug (ligand) interactions – potential insights from structural pharmacogenomics

1. Which mutations affect drug interactions for my drug-lead?
2. What are the common structural features of the binding pocket across the largest segment of the patient population?
3. What are the structural exclusions for my drug lead or lead series? Should certain scaffolds be excluded because of anticipated steric conflicts with current or demonstrably allowable future mutations?
4. How many drug–main-chain interactions are there between my drug and the receptor? Are they independent of sequence?
5. What residues are the known drugs contacting?
6. What other essential (conserved) side chain or backbone features, capable of providing additional selectivity or affinity, are close to the drug-binding site?
7. What other space in the binding pocket is available for new interactions? Can I use this information to focus design of combinatorial libraries?
8. How do specific mutations affect the specificity of my drug or drug lead? What structural changes can be made to prevent recurring resistance?
9. What structural changes are suggested to existing active molecules that might circumvent potential patent infringement or permit the establishment of new and patentable intellectual property – both offensively and defensively?

Summary

Structural pharmacogenomics is an exciting new field that is just now beginning to exert impact on the pharmaceutical industry including fostering the rational design of 'best-in-class' drugs, preclinical selection or ranking of drug candidates, clinical trial analysis, epidemiology, patient outcomes assessment and even 'evidence-based marketing'.

As we have tried to show, many novel and powerful technical insights are becoming available for the first time through steadily advancing computational proteomics means. Access to protein structure on a large scale enables prediction of the Achilles' heel of infectious disease targets, paving the way for rational design of 'super drugs' that circumvent or slow down the development of drug resistance. Structural pharmacogenomics will begin to provide a practical way to address some important implications of the distribution of polymorphisms by race, gender and genetic predisposition to disease, insuring that new drugs will work with the largest fraction of the world's patient population. It will also enable drug companies to select the best drug candidates to advance to the clinic, thus improving clinical trial outcomes and significantly cutting the cost of developing and commercializing new drugs, which is beneficial for drug companies and patients alike. Prediction of drug-resistance phenotypes on a patient-by-patient basis will not only greatly impact the treatment of infectious diseases, such as HIV, but will also be increasingly important in the treatment of cancer.

In many ways, earlier concepts about the application of pharmacogenomics to the pharmaceutical field were limited. They consisted primarily of retrospective analyses – usually involving identification of responders versus non-responders through individual patient testing. This approach is clearly valuable for certain life and death situations; for example, in the treatment of certain cancers with expensive or potentially toxic drugs that are effective with only a subset of patient-specific cancer phenotypes: Genentech's (http://www.gene.com) Herceptin® cancer therapeutic is one example. Such applications will continue to offer great benefit to patients, particularly in crucial situations where the added cost of pre-testing all patients before electing to begin therapy can be justified. However, practical economic considerations dictate that, whenever information becomes available to enable scientists to develop drugs that work in the majority of patients, pharmaceutical companies will incorporate this information to produce increasingly better drugs that work for the largest possible fraction of the patient population, more quickly and at a reduced cost. As pharmaceutical companies increasingly begin to adopt structural pharmacogenomics analyses as a fundamental insight in drug discovery, everyone with a stake in their continued success – patients, physicians, healthcare providers, insurers and drug company employees and shareholders alike – stand to benefit immensely.

References

1 Shafer, R.W. *et al.* (2000) The genetic basis of HIV-1 reverse transcriptase and protease inhibitors. *AIDS Rev.* 2, 211–228
2 Richman, D.D. (2001) HIV chemotherapy. *Nature* 410, 995–1001
3 Perno, C.F. *et al.* (2001) Secondary mutations in the protease region of human immunodeficiency virus and virologic failure in drug-naive patients treated with protease inhibitor-based therapy. *J. Infect. Dis.* 184, 983–991
4 Wang, W. and Kollman, P.A. (2001) Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. *Proc. Natl. Acad. Sci. U. S. A.* 98, 14937–14942
5 Harrigan, P.R. *et al.* (2001) World-wide variation in HIV-1 phenotypic susceptibility in untreated individuals: biologically relevant values for resistance testing. *AIDS* 15, 1671–1677
6 Maggio, E.T. and Ramnarayan, K. (2001) Recent developments in computational proteomics. *Drug Discov. Today* 6, 996–1004
7 Boulme, R. *et al.* (2001) Comparative evaluation between five automated resistance interpretation algorithms. *Antivir. Ther.* 6 (Suppl. 1), 121
8 Larder, B.A. *et al.* (2000) Quantitative prediction of HIV-1 phenotypic drug resistance from genotypes: the virtual phenotype (VirtualPhenotype). *Antivir. Ther.* 5 (Suppl. 3), 49
9 Shenderovich, M. *et al.* (2002) Structural Phenotype Predicts HIV-1 Protease Inhibitor Resistance. Presented at the *9th Conference on Retroviruses and Opportunistic Infections*, 24–28 February 2002
10 Shenderovich, M. *et al.* (2001) Structural pharmacogenomic approach to the evaluation of drug resistant mutations in HIV-1 protease. *J. Clin. Ligand Assay* 24, 140–144
11 Qari, S.H. *et al.* (2002) Comparative analysis of two commercial phenotypic assays for drug susceptibility testing of human immunodeficiency virus type 1. *J. Clin. Microbiol.* 40, 31–35
12 Petropoulos, C.J. *et al.* (2000) A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* 44, 920–928
13 Dam, E. *et al.* (2001) Comparison of HIV-1 resistance phenotypes obtained by two different assay systems. *Antiviral Ther.* 6 (Suppl. 1), 122
14 Hong, L. *et al.* (2000) Crystal structure of an *in vivo* HIV-1 protease mutant in complex with saquinavir: insights into the mechanisms of drug resistance. *Protein Sci.* 9, 1898–1904