

# ABSTRACT 582-T: STRUCTURAL PHENOTYPE PREDICTS HIV-1 PROTEASE INHIBITOR RESISTANCE

M Shenderovich<sup>1</sup>, RM Kagan<sup>2</sup>, K Ramnarayan<sup>1</sup>, P Hess<sup>2</sup> and PNR Heseltine<sup>2</sup>

<sup>1</sup> Structural Bioinformatics Inc. San Diego, CA 92127

<sup>2</sup> Dept. of Infectious Diseases, Quest Diagnostics Incorporated. San Juan Capistrano, CA 92690

## ABSTRACT

**Background:** Mutations in HIV-1 drug targets lead to resistance and consequent therapeutic failure. Rules based interpretations may not accurately predict the effects of multiple mutations, while phenotypic assays are time consuming and costly. To improve resistance prediction, we developed a structural phenotyping procedure that models HIV-1 protease inhibitor complexes and rapidly evaluates binding affinities of protease inhibitors (PI) to mutant variants.

**Methods:** Models of wild type Pr-PI complexes were produced by energy optimization of published structures. Amino acid substitutions were introduced and the models were refined by energy minimization. Calculated changes in the PI binding energy ( $\Delta E_{bind}$ ) of mutant vs. wild type complexes were correlated with published estimates of the changes in  $\Delta E_{bind}$  or the ratios of phenotypic  $IC_{50}$ s. Complexes of Pr variants from clinical isolates were modeled with six PIs (amprenavir, indinavir, lopinavir, nelfinavir, ritonavir, saquinavir) and  $\Delta E_{bind}$  cutoffs corresponding to a 4x increase in  $IC_{50}$  (Virco/PhenoSense™) used to define a structural phenotype of susceptible, resistant or equivocal.

**Results:** Significant correlations ( $R^2 = 0.7 - 0.8$ ,  $P < 0.01$ ) between calculated binding energies and empiric estimates obtained from published  $K_S$  were found for saquinavir, indinavir, ritonavir and amprenavir. Resistance-associated mutations markedly increased ( $\approx 1.5$  kcal/mol)  $\Delta E_{bind}$ . For 48 Pr variants, calculated  $\Delta E_{bind}$  for 5/6 PIs showed significant correlations ( $R^2 = 0.6-0.8$ ,  $P < 0.001$ ) with fold changes in  $IC_{50}$ s (Virco Antivogram™). For a second group of 46 variants,  $\Delta E_{bind}$  for all six PIs showed good correlations ( $R^2 = 0.75-0.85$ ,  $P < 0.001$ ) with fold changes in  $IC_{50}$ s (ViroLogic PhenoSense™). The overall concordance of structurally predicted resistance with phenotype exceeded 80% for both data sets.

**Conclusions:** A novel computational method for structure-based phenotyping of HIV-1 protease was developed that correlates the binding energies of six PIs to HIV-1 Pr and changes in inhibitor  $IC_{50}$ s as measured by phenotyping. This method rapidly predicts drug resistance of clinical HIV-1 Pr variants, with an accuracy approaching that of cell-based phenotypic assays.

## INTRODUCTION

Antiretroviral drugs targeting the reverse transcriptase (RT) and protease (Pr) enzymes of HIV-1 have become a standard part of many treatment regimens for HIV-1 infection (Carpenter et al., 2000). Resistance arises through the accumulation of mutations in the RT or Pr genes of incompletely suppressed virus in the presence of antiretroviral drugs (Shafer et al., 2000 and references therein).

Genotypic testing for resistance is a relatively rapid and inexpensive method to identify known Pr and RT resistance mutations and is recommended for use in clinical practice in defined situations (Hirsch et al., 2000). However, genotypic assays may incorrectly predict resistance for complex mutational patterns or when a new antiretroviral drug is first introduced, for which the relevant mutations are not known. Furthermore, the proliferation of rules-based systems for genotypic interpretation (Boume et al., 2001; Schapiro, 2001) is often confusing for the clinician seeking to determine the clinical significance of a particular mutation pattern.

Phenotypic assays may afford a more reliable measure of resistance for complex mutational patterns or when the genotypic resistance pattern has not been determined (Hertogs et al., 1998; Petropoulos et al., 2000). However, these assays are more time consuming and expensive, and the clinically significant cutoff for the fold change in  $IC_{50}$  for many drugs is not yet known.

A hybrid approach using the genotypic determination of the viral sequence matched against a database of banded viral genotypes and phenotypes results in a *VirtualPhenotype™*, a computational prediction of viral resistance (Virco NV, Mechelen, Belgium; Larder et al., 2000). This method also provides a quantitative estimate of drug resistance. However it may fail to give a prediction when few phenotypic matches for a particular variant are identified in the database as is the case for rare mutational patterns or for newly introduced drugs.

We describe here a structural approach to HIV-1 resistance testing, whereby molecular modeling of the viral protease sequence is used to calculate the change in binding energy to the mutant Pr relative to the wild type reference sequence for each of six FDA-approved PIs. We find that the change in binding energy can be directly correlated to the change in fold resistance in two commercially available phenotypic assays and present a model for a structural phenotyping resistance assay based on these correlations.

**Table 1: Regression analysis of theoretical  $\Delta E_{bind}$  and experimental changes in binding energy<sup>1</sup>**

A: ViroLogic PhenoSense™				
Pr	N (outliers)	R <sup>2</sup>	S.E. (kcal/mol)	S.E. (fold) <sup>2</sup>
amprenavir	46 (0)	0.83	0.36	1.8
indinavir	46 (0)	0.80	0.43	2.0
lopinavir	35 (0)	0.81	0.46	2.1
nelfinavir	46 (0)	0.88	0.50	2.3
ritonavir	46 (0)	0.78	0.57	2.5
saquinavir	46 (0)	0.83	0.45	2.1

B: Virco Antivogram™				
Pr	N (outliers)	R <sup>2</sup> (+outliers)	S.E. (+outliers)	S.E. (fold) <sup>2</sup> (+outliers)
amprenavir	63 (6)	0.53 (0.34)	0.42 (0.55)	2.0 (2.4)
indinavir	63 (8)	0.72 (0.51)	0.43 (0.57)	2.0 (2.5)
lopinavir	63 (3)	0.81 (0.70)	0.35 (0.44)	1.8 (2.0)
nelfinavir	63 (2)	0.70 (0.57)	0.47 (0.58)	2.1 (2.6)
ritonavir	63 (3)	0.78 (0.68)	0.54 (0.68)	2.4 (3.0)
saquinavir	61 (5)	0.66 (0.37)	0.56 (0.77)	2.5 (3.5)

**Table 2: PI  $\Delta E_{bind}$  cutoff values and concordance with phenotypic results**

A: Virologic PhenoSense™							
Pr	$\Delta E_{bind}$ cutoffs <sup>1</sup>	phenotypic concordance			equivocal predictions <sup>2</sup>		
		$c_1$	$c_2$	sensitivity	specificity	number	genotypic concordance
APV	0.7-1.4	86.7%	100%	0.907	<0.0001	5/65	4/5
IDV	0.6-1.5	94.1%	100%	0.958	<0.0001	5/65	5/5
NFV	0.7-1.0	60.6%	96.8%	0.567	<0.0001	1/65	0/1
RTV	0.7-1.4	100%	84.1%	0.754	<0.0001	4/65	3/4
SQV	0.6-1.1	68.4%	100%	0.752	<0.0001	2/65	1/2
LPV	0.3-0.7	100%	83%	0.755	<0.0001	1/35	1/1

B: Virco Antivogram™							
Pr	$\Delta E_{bind}$ cutoffs <sup>1</sup>	phenotypic concordance			equivocal predictions <sup>2</sup>		
		$c_1$	$c_2$	sensitivity	specificity	number	genotypic concordance
APV	0.7-1.4	50.0%	86.7%	0.367	0.003	6/65	4/6
IDV	0.6-1.5	87.8%	84.8%	0.709	<0.0001	7/63	3/7
NFV	0.7-1.0	62.8%	90.7%	0.633	<0.0001	3/63	2/3
RTV	0.7-1.4	78.6%	92.6%	0.71	<0.0001	8/63	5/8
SQV	0.6-1.1	62.5%	91.4%	0.561	<0.0001	4/63	1/4
LPV	0.3-0.7	83.3%	90.6%	0.743	<0.0001	6/62	4/6

<sup>1</sup>  $\Delta E_{bind}$  cutoffs in kcal/mol established as described in Methods (Fig. 2).

<sup>2</sup> 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

<sup>3</sup> Structural prediction gave equivocal value  $c_1 \leq \Delta E_{bind} \leq c_2$ . The number of such predictions reflexed to genotypic rules determination and the concordance of the genotypic calls with the phenotypic predictions are shown.

**Fig. 1: Correlation between binding energy and phenotypic resistance**

Binding energies of the Pr-inhibitor complexes were estimated using the equation:

$$E_{bind} = E_p + E_{complex} - E_{protein} - E_{PI}$$

$E_{complex}$  is the energy of the complex.

$E_{protein}$  &  $E_{PI}$  are the energies of the isolated ligand and protein

$E_p$  is an adjustable constant.

Binding energy function:  $E = E_p + E_{complex} + E_p$

$E_p$  - exact-boundary electrostatic using  $\epsilon_p = 8.0$ ;

$E_{complex}$  &  $E_{PI}$  - ECEPP/3 van der Waals and hydrogen-bonding terms;

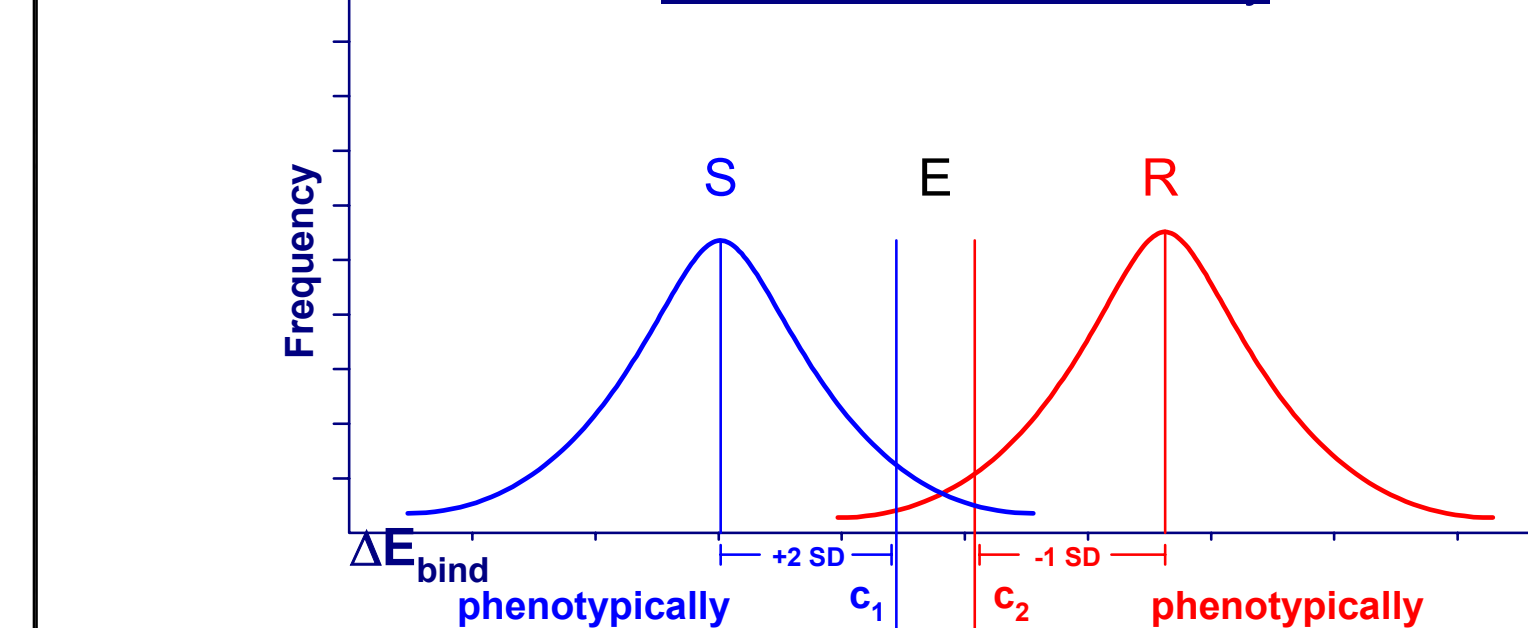
$E_p$  - side-chain entropy term.

Changes of binding energies upon Pr mutation:

$$\Delta E_{bind} (calculated) = E_{bind} (WT) - E_{bind} (mut)$$

$$\Delta E_{bind} (experimental) = RT \ln(IC_{50}mut/IC_{50}WT)$$

**Fig. 2: A semi-quantitative model for a structure-based PI resistance assay**

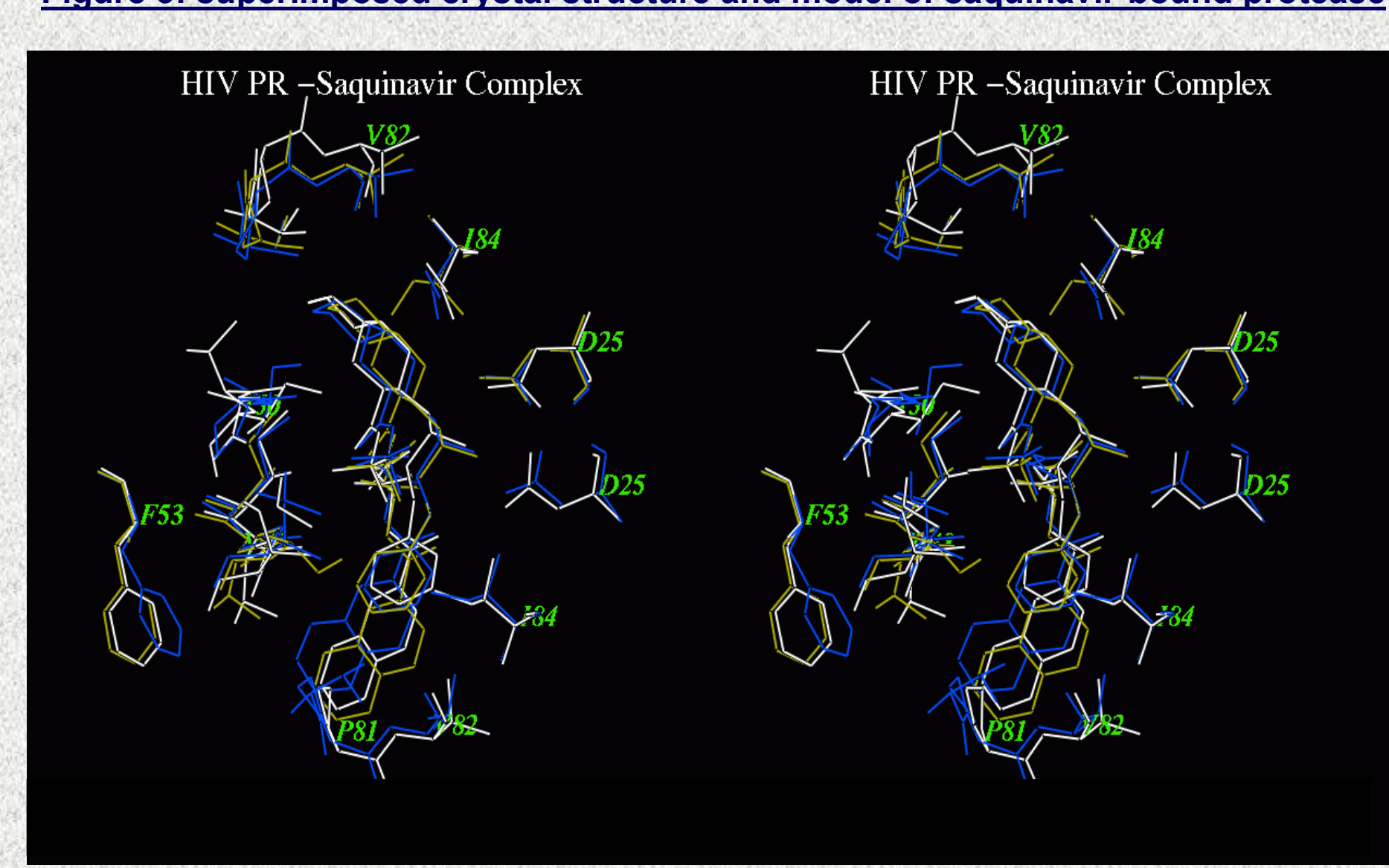


• Susceptible:  $\Delta E_{bind} < c_1$

• Resistant:  $\Delta E_{bind} > c_2$

• Equivocal:  $c_1 \leq \Delta E_{bind} \leq c_2$  (use genotypic rules)

**Figure 3: superimposed crystal structure and model of saquinavir-bound protease**



Superimposed stereo views of the X-ray crystal structure (yellow; Hong et al., 2000) and the SBI model (white) of saquinavir complex with the G48V HIV-1 protease variant. The crystal structure of G48V SOV complex is given in blue for comparison. The model correctly predicts the position and the conformation of the 148 side chain, the displacement of the F53 side chain, and movement of the SOV quinoline ring caused by the G48V substitution.

## METHODS

**Sequence determination:** HIV-1 RNA was extracted from patient plasma samples submitted for RT and Pr genotype determination by Quest Diagnostics or from 50 samples from CCGT 570 genotyped at Quest Diagnostics (Hauhrich et al., 2001). The Pr and the RT genes were amplified by RT PCR and sequenced on an ABI 3700 capillary sequencer.

**Phenotypic resistance determination:** Phenotypic assays were carried out by ViroLogic Inc. (PhenoSense™ assay, <http://www.virologic.com>) or Virco Lab Inc. (Antivogram™ assay, <http://www.vircolab.com>).

**Molecular modeling:** Models of the wild-type HIV-1 Pr-inhibitor complexes were built starting from their crystal structures (Shenderovich et al., 2001). Molecular modeling and simulations were performed using the ICM program, version 2.7 (Abagyan et al., 1994; ICM Manual, 1999). Modeling of mutant HIV protease - inhibitor complexes was performed in two steps: (1) search for an optimal conformation of each mutated side chain and in its protein surrounding, and (2) optimization of the positions of ligand and water molecules, and conformations of the protein residues located near the binding site. Binding energies of mutant Pr-inhibitor complexes were estimated as shown in Figure 1. Structural phenotype determination: the selection of semi-quantitative  $\Delta E_{bind}$  cutoffs defining susceptible, resistant or equivocal phenotypic predictions for Pr sequences is illustrated schematically in Figure 2. The cutoff ( $c_1$ ) for determining susceptibility to each PI was set at two standard deviations above the mean for the group of samples that were less than fourfold resistant to the respective PIs (PhenoSense™). The cutoff ( $c_2$ ) for defining resistance to each PI was set at one standard deviation below the mean (1.5 SD for lopinavir) for the group of samples that were greater than fourfold resistant to the respective PIs (PhenoSense™).

**Table 3: Overall concordance of structural predictions with experimental phenotype**

Structural Prediction	PhenoSense™ 1		Antivogram™ 2		
	S	R	S	R	
S	218	29 <sup>d</sup>	178	33 <sup>d</sup>	
R	5	91	25	107	
E <sup>3</sup>	4	13	E <sup>3</sup>	14	15

<sup>1</sup> 325 comparisons for 65 samples (5 PIs) and 35 comparisons for 38/65 samples with available LPV phenotypic data. The overall agreement between structural prediction and PhenoSense™ was 90%. S = Susceptible, R = Resistant, E = Equivocal.

<sup>2</sup> 315 comparisons for 63 samples (5 PIs) and 62 comparisons for 62/63 samples with available LPV phenotypic data. The overall agreement between structural prediction and Antivogram™ was 82.2%.

<sup>3</sup> Equivocal structural assignments where the change in binding energy was between the lower and the upper cutoffs. For these assignments, phenotypic resistance calls agreed with the phenotypic predictions in 14/17 cases (PhenoSense™) and 15/29 cases (Antivogram™).

<sup>4</sup> The average fold resistance for false negatives was 3.8 +/- 1.4. If biological cutoffs (Harrigan et al., 2001) are used, then only 16 false negative calls would have been reported.

<sup>5</sup> The average fold resistance for false negatives was 5.7 +/- 2.7.

**Table 4: Comparison of Structural false negatives to PhenoSense™ results<sup>1</sup>**

Pr	drug	genotype	phenotype <sup>2</sup>	SP	fold	cutoff <sup>3</sup>	$\Delta E_{bind}$ (kcal/mol)	SP fold	Protease mutations detected <sup>4</sup>
20	NFV	S	R	9.5	S	0.340	1.9	1.0	L101V15V V32I E35D M36V I47V L63 P V82I I93L
67	NFV	R	R	3.4	S	0.890	3.5	1.0	L101L R57K R L63P V77I L90M I93L I
95	NFV	S	R	2.6	S	-0.050	0.9	1.0	E35D N37H R41K I62V L63A A71T V77 I I93L
100	SQV	R	R	4.2	S	0.560	3.6	1.0	I13V L33P M36L N37D I54V L63P K70 R A71V V82A N93D L90M
114	NFV	R	R	4.2	S	0.450	2.3	1.0	L101L62V L63P A71T V77I L90M I93L
114	RTV	S	R	2.8	S	0.300	1.4	1.0	L101L62V L63P A71T V77I L90M I93L
114	RTV	S	R	2.9	S	0.350	2.4	1.0	L101L62V L63P A71T V77I L90M I93L
179	RTV	S	R	3	S	0.140	1.3	1.0	L101V K14R L33V L63Q I64V E65D
179	RTV	S	R	2.6	S	-0.197	0.6	1.0	Q76Q L101I13V E35D M36M V N37E K 43T M46I D60E I62V L63P I64V A71V A71V V77I I85V L90M I93L
183	APV	S	R	4	S	-0.190	0.7	1.0	L101I13V G16E R41K L63H I69Q A71T I72M V82I L89V L
183	SQV	R	R	5.1	S	0.090	1.4	1.0	I62V L63P I64L L90M I93L C95F
196	NFV	R	R	5.6	S	0.290	1.7	1.0	E35D M36L N37D L63P K70R A71TA
196	RTV	S	R	3.2	S	0.690	2.6	1.0	L101I13V K20I M36I K40T I54V D96E I I62V L63P A71V V82A L90M C92K
201	NFV	S	R	4.1	S	-0.450	0.4	1.0	K20I I62V L63P A71T I72T N89N S2K
228	SQV	R	R	5.7	S	-0.060	1.0	1.0	L101I13V G16E R41K L63H I69Q A71T I72M V82I L89V L
231	APV	S	R	3	S	0.440	1.6	1.0	L101 G48V I54V R57K I62V L63P I74S V82A
235	NFV	R	R	2.9	S	0.620	3.1	1.0	I15V E35D M36I I62V L63P L90M
235	RTV	S	R	3.7	S	0.660	3.3	1.0	N37T R41K M46I M I64V L90M
235	RTV	S	R	3.3	S	0.390	1.6	1.0	I15V E35D M36I I62V L63P L90M
302	APV	S	R	3.2	S	0.078	1.1	1.0	I15V I67E G L63P
302	RTV	S	R	3.9	S	0.307	1.4	1.0	I15V I67E G L63P
485	NFV	R	R	4.6	S	0.361	1.9	1.0	L101I13V N37D L63P A71T V77I V L90M
489	SQV	S	R	3.6	S	0.554	3.6	1.0	L101V E35D M36I M46I I64V L63A A71V V82A I91T I93L
524	NFV	S	R	2.6	S	-0.004	1.0	1.0	L101I13V L19M L63P I69Q

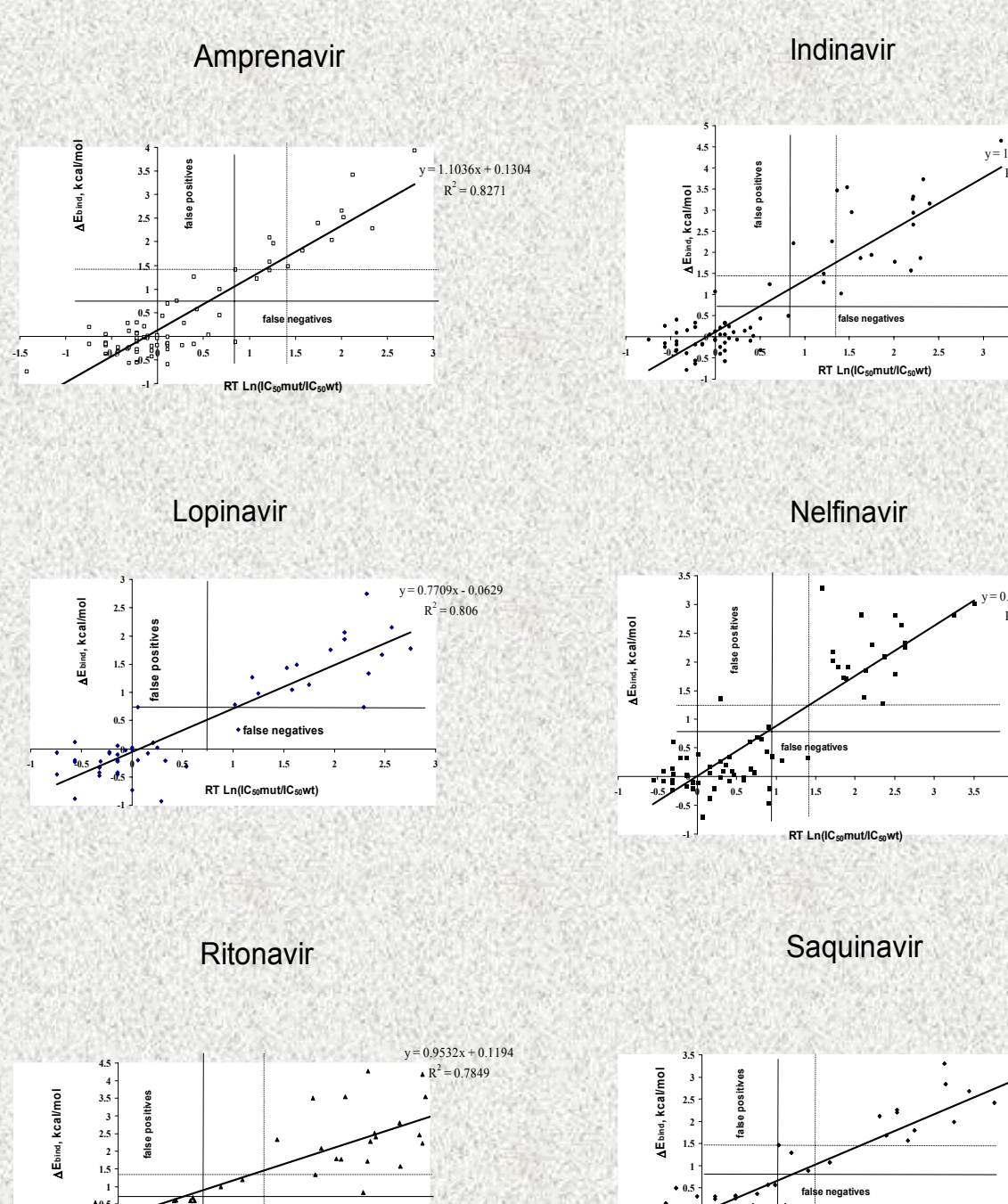
<sup>1</sup> 29 false negative predictions as shown in Table 3

<sup>2</sup> based on PhenoSense™ cutoff of > 2.5-fold increase in resistance

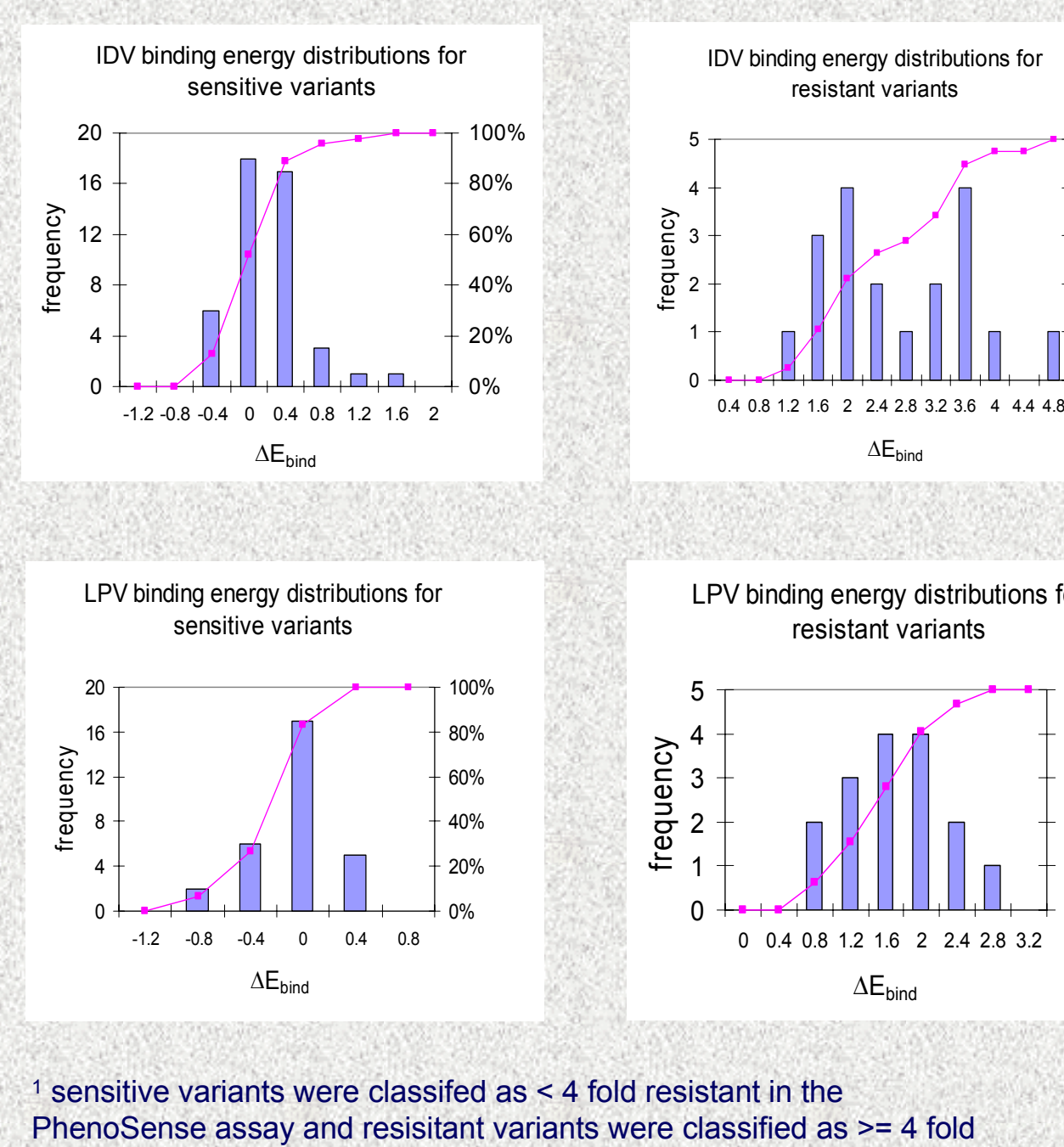
<sup>3</sup> cutoffs shown in Table 2 were used. The fold resistance predicted from the change in free energy was calculated from the linear regressions (Fig. 4). The average difference between the predicted and the experimental fold resistance was 2.1 +/- 1.7 fold

<sup>4</sup> primary resistance mutations are in red, secondary mutations are bolded and atypical amino acid substitutions at resistance-associated codons are underlined.

**Fig. 4: Correlation between binding energies fold resistance (PhenoSense™) for PI complexes**



**Figure 5: binding energy distributions for indinavir and lopinavir for sensitive and resistant variants<sup>1</sup>**



<sup>1</sup> sensitive variants were classified as < 4 fold resistant in the PhenoSense assay and resistant variants were classified as >= 4 fold resistant.

**Table 5: Comparison of false positive structural phenotype assignments to genotypic and VirtualPhenotype™ predictions<sup>1</sup>**

id	Pr	Antivogram™ phenotype	fold	SP	Structural Prediction $\Delta E_{bind}$ (kcal/mol)	genotype	Quest-Stanford	VirtualPhenotype™ <sup>2</sup>	Protease mutations detected <sup>3</sup>
A437	NFV	S	1.1	R	1.85	R (rules)	R	R (rules)	L101I13V P39Q M46I M K55K L63P A71T I I84V N88D
A531	LPV	S	1.3	R	0.78	S	R (rules)	R (rules)	L101I13V L33L F I54V L63P A71V V82AV L90ML
A531	NFV	S	1.7	R	1.48	R (rules)	R	R (rules)	N37S M46M I62V L63P G73S V77I I93L
A531	RTV	S	1.1	R	1.42	R (rules)	R	R (rules)	L101I13V E35D M36I I62V L63P G73C I64V L90M
Q228	APV	S	1.8	R	1.42	R (rules)	R	R (rules)	L101L63P A71V I72E G73S G V77I I84V L90M I93L
Q288	NFV	S	1.7	R	1.02	R (rules)	R	R (rules)	N37S M46M I62V L63P G73S V77I I90ML I93L
Q274	IDV	S	1.8	R	2.76	S	R (rules)	R (rules)	L101N37S G48V I62V L63P G73S V77I I93L
Q372	APV	S	2.2	R	1.63	R (rules)	R	R (rules)	L101I13V E35D M36I I62V L63P G73C I64V L90M
Q389	APV	S	1.7	R	2.33	R (rules)	R	R (rules)	M46M L63P I64V I74TS V82T I64V
Q389	SQV	S	1.6	R	1.63	R (rules)	R	R (rules)	L101I13V E35D M36I I62V L63P G73C I64V L90M
Q441	IDV	S	2.4	R	1.87	R (rules)	R	R (rules)	L101I244I R41K M46L G48V D106E I63P L I64V K70K V82A I84V
Q575	APV	S	0.9	R	1.45	S	R (rules)	R (rules)	L101I13V M46L I64V L63P A71V G73S V82A L90M I93L
Q581	APV	S	0.7	R	1.60	S	R (rules)	R (rules)	L101I13V I12N I15V K20K M24I M36I R41K M4 6I I54V I62V L63P I84V V82A
Q589	APV	S	1.5	R	2.34	R (rules)	R	R (rules)	M46I R57K L63A I64V A71V V82T I84V
Q589	IDV	S	2.6	R	2.8				