Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 43 (2008) 2518–2534
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

Prediction of HIV-1 protease inhibitor resistance by Molecular Modeling Protocols (MMPs) using GenMol™ software

G. Pèpe [a,*], J. Courcambeck [b], R. Perbost [b], P. Jouanna [a,c], P. Halfon [b]

[a] GCOM2, UMR-CNRS 6114, Faculté des Sciences de Luminy, Case 901, 13288 Marseille Cedex 9, France
[b] Genoscience, 23 Rue de Friedland, 13006 Marseille, France
[c] Géosciences Montpellier, Université Montpellier 2, CNRS, CC 60, 34095 Montpellier, France

Received 15 May 2007; received in revised form 28 January 2008; accepted 29 February 2008
Available online 14 March 2008

Abstract

This paper investigates the contribution of Molecular Modeling to (i) predict and (ii) understand more fundamentally HIV drug resistance. Based on a new automated GenMol™ module, these goals are approached by Molecular Modeling Protocols (MMPs), respectively, (i) the Molecular Modeling Phenotype Protocol (MMPP) and (ii) the Molecular Modeling Phenotype–Genotype Protocol (MMGPP). Section 2 recalls clinical practice with a reference case study and Section 3 presents atomistic simulation tools. Section 4 is the heart of the paper. In Section 4.1, MMPP drug resistance prediction is based on correlations between fold resistances versus binding energies on 2959 HIV-1 complexes with 6 protease inhibitors. Based on a drug sensitivity twofold criterion, modeling prediction is able to replace long and costly phenotype tests. In Section 4.2, MMGPP enlightens drug resistance by investigating steric and energetic residues/inhibitor interaction. Section 5 gives a synthesis on modeling contribution to drug resistance prediction. In conclusion, the most promising trend consists of MMP automats that are able to suggest a real time diagnosis taking into account the history of each patient, to enrich databases and to develop therapy strategy and new drugs.

Keywords: HIV resistance; HIV mutations; Molecular modeling; GenMol™

1. Introduction

For stopping HIV-1 virus proliferation, the main targets of antiretroviral treatments are virus reverse transcriptase and protease enzymes. However, virus can become drug resistant due to target genetic mutations. Therapy is guided in clinical practice by fold resistance phenotype test and mutation genotype X-ray test, leading to a genotype–phenotype cross-analysis. However, a great help would be brought to therapy by solving two problems:

- PB1 consists of reducing the cost of phenotype tests and their delay in order to fight the virus in its early development.
- PB2 consists of understanding more fundamentally the mutation/drug resistance mechanism to elaborate new treatment strategies or new drug.

For helping such a research, two Molecular Modeling Protocols (MMPs) are proposed using the GenMol™ software [1]:

(i) The "Molecular Modeling Phenotype Protocol" (MMPP) is based on inhibitor/virus protease binding energy, a fundamental information escaping to clinical phenotype. This new contribution to PB1 is coming after other studies [2–7] and especially studies in Refs. [8,9].
(ii) The "Molecular Modeling Genotype–Phenotype Protocol" (MMGPP) is a tentative of focusing on PB2 by a steric and energetic investigation of interconnection between a protease inhibitor (PI) and a residue (R).

* Corresponding author. Tel.: +334 91 82 94 04; fax: +334 91 82 94 07.
E-mail address: pepe@luminy.univ-mrs.fr (G. Pèpe).

0223-5234/$ - see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.02.046

Both MMPs are validated by comparison with clinical phenotype and/or genotype tests recalled in Section 2, where a reference case study is extracted from the ''Stanford HIV Drug resistance Data Base''. Section 3 recalls usual molecular modeling tools and validates GenMol docking–optimization process leading to a protease/drug complex. Section 4 is the heart of the paper. In Section 4.1, MMPP drug resistance prediction is investigated on different correlations between the fold resistance FR and the difference ΔE between binding energies of complexes (PI + wild protease WP) and (PI + mutated protease MP). Correlation over all PIs and correlations relative to each PI are finally split into sensitive and resistant sub-correlations. Based on a twofold criterion, a numerical prediction of virus sensitivity is obtained, replacing long and costly phenotype tests. In Section 4.2, MMGPP enters more deeply into the drug resistance mechanism, correlating phenotype–genotype clinical data with steric or energetic differences between (PI + wild residue WR) and (PI + mutated residue MR). After a synthesis in Section 5 on modeling contributions to resistance prediction, conclusion sketches some trends, the most promising consisting of MMP automats able to suggest in real time antiretroviral treatment of each patient, to enrich databases and to elaborate therapy strategies or new drugs.

2. Virus/drug approach in clinical practice

When multiplication of HIV wild virus (WV) is stopped by an antiretroviral drug (ARV), such as a protease inhibitor PI, mutations can occur. If the drug remains active after mutation the mutant is said to be sensitive. On the contrary the mutant is said to be resistant. Drug resistance study is based on phenotype and genotype tests briefly recalled hereunder to present the case study in view of MMP validation.

2.1. Phenotype

A phenotype test consists of growing in vitro an HIV sample coming from a patient, simultaneously with a wild virus culture. A dose of ARV is added to both samples and the inhibition replication of the virus is measured. The drug resistance to a given PI of a mutated virus (MV) compared to the resistance of the wild virus (WV) is quantified by the ''fold resistance'' (FR_{PI}):

$$\begin{aligned} &FR_{PI} = (IC_{50}MV)/(IC_{50}WV);\\ &IC_{50}MV = \text{PI concentration for obtaining 50\% inhibition of MV},\\ &IC_{50}WV = \text{PI concentration for obtaining 50\% inhibition of the WV} \end{aligned} \quad (1a)$$

The fold resistance can be rewritten as function of constants relative to the associated protease, called ''association constants'':

$$\begin{aligned} &FR_{PI} = (IC_{50}MV)/(IC_{50}\,WV) = (KWTP)/(KMP);\\ &KWTP = \text{association constant relative to the wild protease (WP)},\\ &KMP = \text{association constant relative to the mutated protease (MP)} \end{aligned} \quad (1b)$$

Phenotype data, extracted from the ''Stanford HIV Drug Resistance Data Base – Feb 2004'', are used in Section 4 for validating drug prediction by the Molecular Modeling Phenotype Protocol (MMPP). Phenotype tests are costly (1200 €) and the delay of virus culture (10 days) is too long for a rapid therapy decision. These major drawbacks push towards replacing experimental tests by modeling prediction.

2.2. Genotype

2.2.1. Genotype information

The purpose of a genotype test, today available at low cost in some minutes by automats, is to detect protease mutations. Among the 99 residues belonging to the wild protease (WP), essentially 21 wild residues (WR) are important for PIs' virus sensibility. A mutated protease (MP) includes one or several mutated residues (MR) characterized by a symbol, for instance K57N. Letter K identifies the amino-acid in the WP, 57 the position in the code chain and N the code of the mutated amino-acid.

Studying mutations requires a statistical approach on a great number of genotype tests. In such an ensemble, each sample S corresponds in general to several mutated residues (MRs) and several protease inhibitors (PIs). Moreover, several parameters are generally missing, such as the patient identification (is he always the same or not for different complexes?), the sequence of PIs treatments, the delay of treatment, etc.

2.2.2. Mutation frequency: a new statistical index

In view of a better understanding of mutation sequences, we propose to introduce the notion of ''mutation frequency'' (MF). The statistical index $MF(R)_{PI}$, relative to a residue R and an inhibitor PI, is defined as follows. Suppose for instance that, within reported cases, 595 samples are treated by the inhibitor PI = NFV. Within this NFV class, the total number of residues R = L10 (mutated or not) is of course equal to 595. If only 327 of them appear in the MR genotype list, the corresponding mutation frequency is $MF(L10)_{NFV} = (327/595) \times 100 = 55\%$.

A mutation frequency analysis is illustrated below using the genotype list extracted from the ''Stanford HIV Drug Resistance Data Base'' (Virologic set) at the end of February 2004. The 2959 HIV-1 samples are distributed in PI classes: 595 for nelfinavir (NFV), 547 for indinavir (IDV), 476 for ritonavir (RTV), 540 for saquinavir (SQV), 482 for amprenavir (APV) and 319 for lopinavir (LPV). Atazanavir (ATV) was not taken into account because its MP/PI X-ray structures are not available and not possible in our modeling approach. When one mutation is not identified, signaled by x or z code in the data, the most probable mutation occurrence was deduced from the associated mutations. In Fig. 1, the mutation

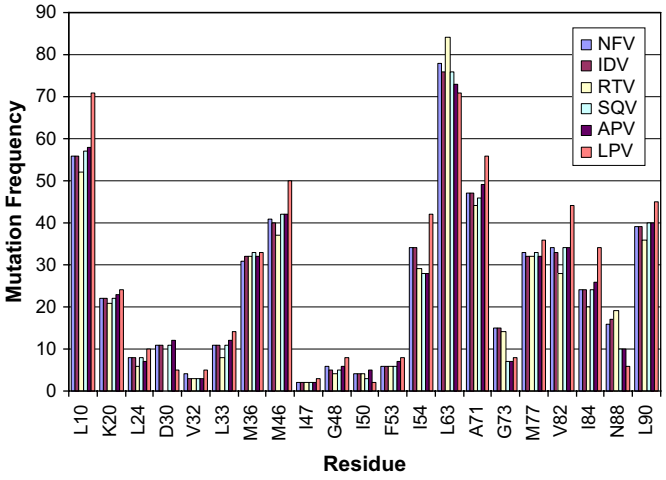

Fig. 1. Mutation Frequency $MF(R)_{PI}$ of the 21 HIV-1 protease residues R versus PIs: NFV, IDV, RTV, SQV, APV, LPV.

frequency $MF(R)_{PI}$ is reported for all the 21 residues important for PIs' virus sensibility, in the presence of the 6 PIs.

2.3. Genotype–phenotype approach

Coupling both genotype and phenotype information leads to a description of fold resistance (FR) of mutated residues MRs versus inhibitor PIs. A 3-D graph of the fold resistance function FR (MR,PI) would be confusing, due to the number of points (2959), the number of MRs (21) and the number of PIs (6). The graph of this function can be converted into a 2-D table, where columns are relative to the 6 PIs and lines are relative to the 21 Rs important for PIs' virus sensibility whatever be their post-mutation structure, except for M50 with two mutated forms M50V and M50L. The fold resistance FR (MR,PI) is coded by a color convention as shown in Fig. 2 extracted from the "Stanford HIV Drug resistance Data Base (October, 2004)". (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.) Six levels of drug resistance of the MRs are distinguished, between high-level resistance (dark blue) and hyper sensibility (white with a star).

Drug resistance of a given PI can be estimated using Fig. 2 as soon as a residue R is declared mutated in the genotype list. For instance if residue M46 is declared mutated, it will be resistant to all PIs except NFV and SQV, the latter having a fold resistance $FR = 0$.

2.4. Discussion and trends

Taking into account the mutation frequency, introduced in Section 2.2.2, would lead to an improved genotype–phenotype analysis. Indeed, a given residue can mute very often in the presence of a drug but have no incidence on the patient's health if its drug resistance is low. On the contrary, a residue of low mutation frequency can have a strong incidence on the patient if its drug resistance is high.

However, before exploiting this new information to weight statistically the drug resistance, attention must be drawn to

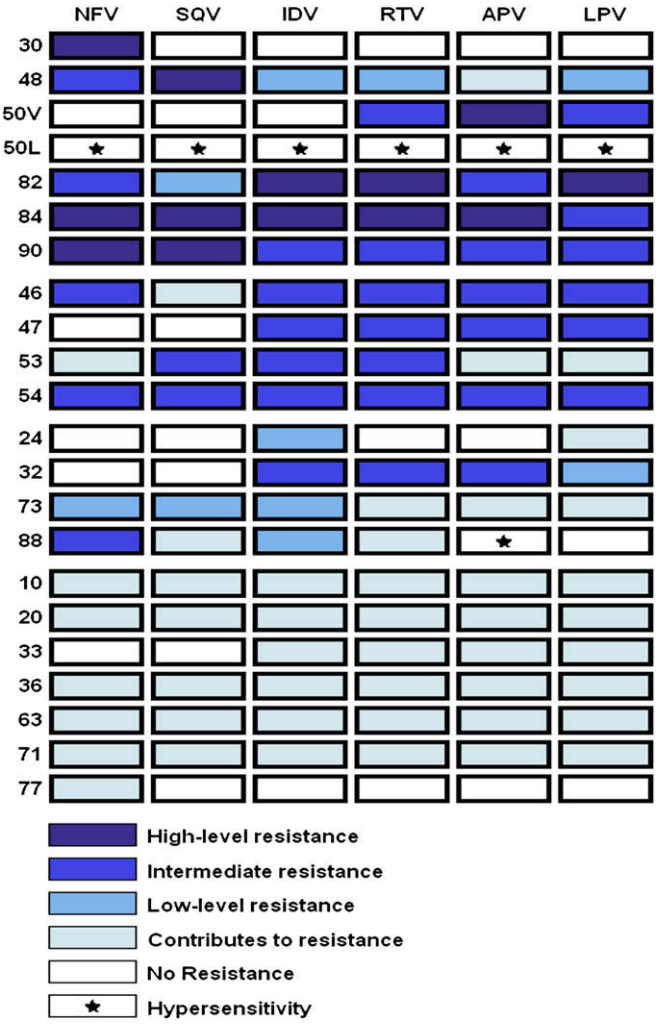

Fig. 2. Degrees of drug resistance FR versus possible mutated residues MR treated by different PIs [after "Stanford HIV Drug resistance Data Base (October, 2004)"].

possible bias due to incomplete information in data. Such is the case in Fig. 1, where the mutation frequency $MF(R)_{PI}$ appears to be quasi-identical for the 6 PIs whatever be the residue R. This unexpected coincidence suggests some correlation between hidden parameters: samples coming from the same or different patients, mutations after a simple or multi-therapy, chronology of drug administration, etc. Consequently data bank should be first refined to escape possible bias described in Section 2.2.2. Such database improvement appears essential not only for usual therapy but also for linking clinical data with modeling drug resistance prediction.

3. Virus/drug approach in molecular modeling

3.1. Drug resistance mechanism and molecular modeling

Stopping HIV development by inhibiting its protease is a mechanism observed at the microscopic level. The function of the protease is to cut long protein chains to fabricate new virus components. The drug appears as an allurement for the

protease which can remain clamped to it and thus neutralized. This mechanism suggests that ''binding energy'' between the protease and the drug is a fundamental criterion for quantifying virus sensitivity or resistance to an antiviral treatment. This quantity, not accessible by experiment, explains the specific contribution of molecular modeling in antiretroviral investigation. Fig. 3 shows a wild HIV-1 protease and its active site, with indication of the residues important for PIs' virus sensitivity.

3.2. Molecular modeling tools

3.2.1. Steric tools

Inter-atomic distances between atoms is a first geometrical criterion used in steric MMGPP (Section 4.2.1). The relative distance between similar structures is measured by the Root Mean Square Distance (RMSD) for example between WP and MP (Section 3.4).

3.2.2. Energetic tools

Access to energy is specific to modeling, contrary to geometry tools applicable also to direct X-ray observation. Theoretically, under any temperature and pressure (*T*,*P*) conditions, the adequate potential is not energy but free enthalpy or Gibbs energy. In fact, energy is here assimilated to free enthalpy, inter-atomic potentials being calibrated in clinical (*T*,*P*) conditions. Energies are defined with an arbitrary origin. Thus only energy differences are intrinsic, explaining their importance in the following.

For complex structures, energy is accessible by Molecular Mechanics with the help of ab initio. GenMol™ possesses its own Molecular Mechanics covalent and non-bond potentials. Coulomb charges are obtained using the ab initio modified Del Re method [10,11]. Both Molecular Mechanics and ab initio computations are treated by GenMol [12–15].

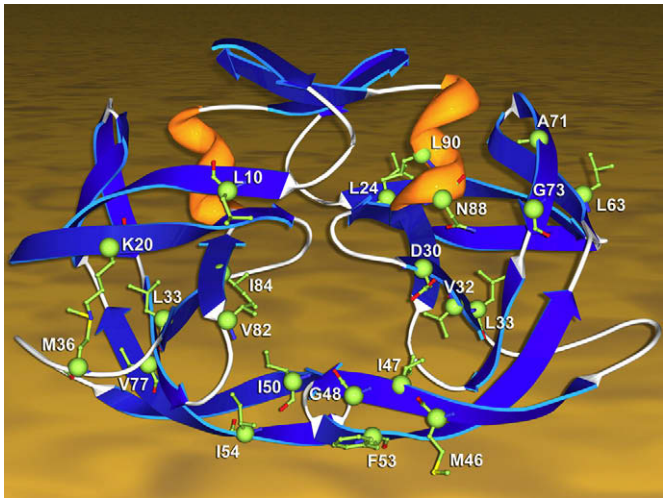

Fig. 3. Wild HIV-1 protease image displaying the residues important for PIs' virus sensitivity, the central hole corresponding to the protease active site. The two symmetrical chains A and B are shown. However, for clarity PIs are indicated only on one chain.

3.3. Building MP/PI complexes

Powerful X-ray sources lead to a complete X-ray description of structures including proton positions when good quality crystals are available. In that case, molecular modeling leads directly to the binding energy of WP/PI or MP/PI complexes. However, such complete structures are not usually available and binding energy computation requires the following operations.

3.3.1. Step 1: importing WP and PI structures from X-ray data bank

Modeling directly atomic clusters is possible for simple structures as PIs but remains difficult today due to the lack of necessary geometry accuracy for binding energy calculation on complex structures such as proteases associated to PIs. Consequently structures used for calculations are imported from the Protein Data Bank (PDB) [16].

The resolution of X-ray must be excellent for allowing precise computation. In case of low resolution, an artifice can be used. For instance, for overcoming the low resolution (2.8 Å) of the LPV X-ray structure (1MUI) and considering the similarity between LPV and RTV, 1HXW structure was used for computing binding LPV energies. A better correlation was obtained in that way with phenotype data. Similar attempts made with other PIs were abandoned, because they lead to poor correlations.

3.3.2. Step 2: building a mutated protease MP

Starting from a wild protease structure given in a PDB data bank, GenMol simulates mutations as follows. The former lateral chains corresponding to mutated residue are first withdrawn and then replaced by the new lateral chains (together with their protons) added in some geometrical position. These lateral chains and the neighborhood ones are repositioned by rotation around σ bonds in order to minimize the local non-bond interactions (van der Waals, coulombic and hydrogen bonds). Finally a complete optimization of all atomic interactions by the conjugate gradient method is applied to the total structure for obtaining the mutated MP protease structure at the lowest strain energy level.

For this global optimization, the first decision consists of choosing protons' refinement convergence limit CL1 and then all atoms' refinement convergence limit CL2. For instance, in Section 4.1.4, different (CL1, CL2) couples were tested for obtaining the best possible separation between $\Pi_{\mathrm{PI\ sen}}$ and $\Pi_{\mathrm{PI\ res}}$ sub-ensembles, leading in that case to adopt CL1 = 0.05 kcal/Å/mol and CL2 = 1.0 kcal/Å/mol. If lower values were chosen, no separation would be found. This means that optimized MP/PI structures must stay as close as possible to the reference X-ray structure.

A second decision consists of choosing between explicit or implicit hydration models. In explicit water models, water molecules are added to water molecules belonging to the X-ray structure in order to complete the water shell. Implicit water model described in Ref. [17] was also tested. However, hydration effects on the molecular complex were not considered in

the following, because they have no incidence on correlations derived in Section 4.1.

3.3.3. Step 3: docking mutated protease MP with protease inhibitor PI

Associating MP and PI is obtained by docking. The PI structure, with water molecules linked to it, is introduced into the protease active site at the X-ray position given in the PDB data bank. PDB codes are as follows: 1HXW for WP/RTV or MP/RTV (1.8 Å resolution), 1HPV for WP/APV or MP/APV (1.9 Å resolution), 1OHR for WP/NFV or MP/NFV (2.1 Å resolution), 2BPX for WP/IDV or MP/IDV (2.8 Å resolution), 1HXB and 1FB7 for WP/SQV or MP/SQV (with, respectively, a 2.3 and 2.6 Å resolution).

3.3.4. Step 4: optimizing MP/PI complex and obtaining binding energy

The docked complex is optimized in two steps. A first optimization is done by slightly translating and rotating the PI rigid molecule in its environment assumed to be rigid, followed by rotations around σ bonds of the lateral chains contacting the PI in order to minimize the binding energy.

Then a complete energy refinement by the conjugate gradient method is applied to the complete complex. The binding energy $E_{\mathrm{MP/PI}}$ is equal to the total energy $E_{(\mathrm{MP+PI})\text{-opt}}$ of the final optimized complex minus $E_{\mathrm{MP\text{-}opt}}$ energy of MP and minus $E_{\mathrm{PI\text{-}opt}}$ energy of the PI (values extracted from $E_{(\mathrm{MP+PI})\text{-opt}}$), plus an arbitrary constant:

$$\text{Binding energy } E_{\mathrm{MP/PI}} = E_{(\mathrm{MP+PI})\text{-opt}} - \left(E_{\mathrm{MP\text{-}opt}} + E_{\mathrm{PI\text{-}opt}}\right) + \text{arbitrary constant} \quad (2)$$

3.4. Discussion

Before any further development, the physical meaning of simulated MP/PI complexes must be validated by comparison with X-ray observed MP/PI structures. Table 1 gives a guide for correlating simulated and observed complexes. Of course preceding steps 1 and 2 have no equivalent for an imported X-ray complex. On the other hand, a numerical optimization in step 4 appears strange for the naturally optimized MP/PI. In fact comparing step 3 and step 4 for the imported structure offers a first level of validation. A second level of validation is based on comparing imported and simulated structures' energies in step 4.

Table 1
Comparing study steps of simulated and observed MP/PI complexes

	Simulated MP/PI complex	Observed MP/PI complex
Step 1	Importing WP and PI from X-ray	–
Step 2	Mutation of WP into MP	–
Step 3	Docking MP with PI ⇒ $(\mathrm{MP/PI})_{\mathrm{simu}}$	Importing MP/PI from X-ray ⇒ $(\mathrm{MP/PI})_{\mathrm{obs}}$
Step 4	Optimizing $(\mathrm{MP/PI})_{\mathrm{simu}}$ ⇒ $(\mathrm{MP/PI})_{\mathrm{simu\_opt}}$	Optimizing $(\mathrm{MP/PI})_{\mathrm{obs}}$ ⇒ $(\mathrm{MP/PI})_{\mathrm{obs\_opt}}$

3.4.1. First level of validation

Comparison between $(\mathrm{MP/PI})_{\mathrm{obs}}$ structure in step 3 and its numerically optimized image $(\mathrm{MP/PI})_{\mathrm{obs\_opt}}$ in step 4 can be done uniquely on a steric criterion, because energy of the natural structure is ignored. An excellent RMSD between both structures is a first modeling requirement, indicating that ab initio charges and atomic bond or non-bond potentials are adequate. An illustration is given (Fig. 4) for MP = V82A and PI = RTV. Distinction between $(\mathrm{MP/RTV})_{\mathrm{obs}}$ and $(\mathrm{MP/RTV})_{\mathrm{obs\_opt}}$ structures is unappreciable to the eye and the RMSD = 0.65 Å between the PI and the main chain atoms.

3.4.2. Second level of validation

The second level of validation consists of comparing $(\mathrm{MP/PI})_{\mathrm{obs}}$ in step 3 with $(\mathrm{MP/PI})_{\mathrm{simu\_opt}}$ in step 4. In fact, an excellent fit in the first level of validation allows replacing $(\mathrm{MP/PI})_{\mathrm{obs}}$ by its optimized image $(\mathrm{MP/PI})_{\mathrm{obs\_opt}}$. Thus comparing $(\mathrm{MP/PI})_{\mathrm{obs\_opt}}$ with $(\mathrm{MP/PI})_{\mathrm{simu\_opt}}$ can be done in step 4 on both steric and energetic criteria. A validation based on the energetic criterion is essential in Section 4, where the key for linking clinical observations to simulation predictions is energy.

The fit between $(\mathrm{V82A/RTV})_{\mathrm{obs\_opt}}$ and $(\mathrm{V82A/RTV})_{\mathrm{simu\text{-}opt}}$ structures is excellent (Fig. 5), with a RMSD = 0.5 Å between the PI and the main chain atoms.

Comparison between binding energies of both structures is also excellent:

$$E_{\mathrm{MP/PI\ obs\_opt}} = -88.5\ \mathrm{kcal/mol}$$

$$E_{\mathrm{MP/PI\ simu\_opt}} = -88.7\ \mathrm{kcal/mol}$$

$$\Delta E_{\mathrm{obs\text{-}simu}} = \left(E_{\mathrm{MP/PI\ obs\_opt}}\right) - \left(E_{\mathrm{MP/PI\ simu\_opt}}\right) = 0.2\ \mathrm{kcal/mol}$$

$$\Delta E_{obs-simu}/E_{\mathrm{MP/PI\ obs\_opt}} = 0.2\ (\%)$$

In short, molecular simulation for building protease/PI complexes (see Section 4.1 for MMPP) is validated by the above illustrations which show an excellent fit between the observed and simulated structures, using either steric or energetic comparison criteria. This validation extends to residue/PI complexes (see Section 4.2 for MMGPP).

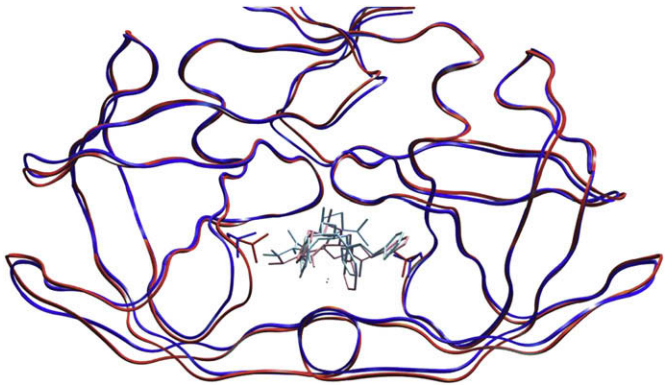

Fig. 4. Best molecular fit between $(\mathrm{V82A/RTV})_{\mathrm{obs}}$ and $(\mathrm{V82A/RTV})_{\mathrm{obs\text{-}opt}}$ structures.

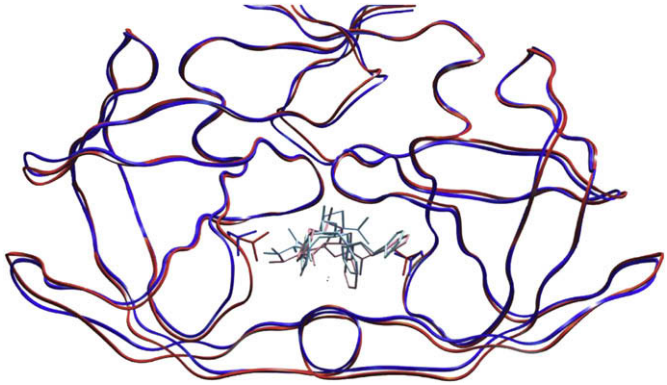

Fig. 5. Best molecular fit between $(V82A/RTV)_{simu\text{-}opt}$ and $(V82A/RTV)_{obs\text{-}opt}$ structures.

4. Bridging clinical and modeling practices for drug resistance prediction

4.1. The Molecular Modeling Phenotype Protocol (MMPP)

4.1.1. Drug resistance prediction by MMPP

MMPP aims at replacing long and costly phenotype tests by modeling. The drug resistance is clinically measured by the virus growth rate in the presence of an inhibitor PI. Bridging this growth rate directly with a computational simulation would suppose to mimic the virus duplication, an unrealistic target today. However, it has been demonstrated [18,19] that phenotypic tests, characterizing the virus growth rate by the fold resistance FR_{PI} defined by Eq. (1) or Eq. (2), can be related to the difference ΔE_{PI} between WP/PI and MP/PI binding energies, accessible by modeling:

$$\Delta E_{PI} = E_{WP/PI} - E_{MP/PI};$$
$$E_{WP/PI} = \text{binding energy of complex WP/PI},$$
$$E_{MP/PI} = \text{binding energy of complex MP/PI} \quad (3)$$

The theoretical bridging relation is:

$$\Delta E_{PI} = RT \ln(FR_{PI}) \quad (4)$$

$R = 8.314$ J/g mol K, universal gas constant; $T =$ absolute temperature (K); $RT = 300 \times 1.987\text{E} - 3 = 0.596$ kcal/mol in clinical conditions at $T = 300$ K.

However, both clinical and modeling quantities are obtained statistically. Consequently this relation becomes in practice a correlation:

$$\Delta E \Leftrightarrow RT \ln(FR) \quad (5)$$

The quality of FR prediction is directly related to the quality of this correlation. Three types of correlations are considered in the following:

(a) The simplest is a brute correlation between ΔE_{Π} and FR_{Π} for an ensemble Π of complexes extracted from a data bank without any distinction between the different PIs:

$$\Delta E_{\Pi} \Leftrightarrow RT \ln(FR_{\Pi}) \quad (6)$$

(b) A first refinement consists of investigating a correlation within the ensemble Π_{PI} of complexes related to the same PI:

$$\Delta E_{PI} \Leftrightarrow \ln(FR_{PI}) \quad (7)$$

(c) Finally, the fundamental question of the therapist is: "Does the virus be sensitive or resistant to a PI?". To solve this binary problem, a partition is considered in ensemble Π_{PI} based on an arbitrary fold resistance $FR = 2.5$. Sub-ensemble $\Pi_{PI\ sen}$ includes "sensitive" cases ($FR \leq 2.5$) and sub-ensemble $\Pi_{PI\ res}$ includes "resistant" cases ($FR > 2.5$). Correlation defined by Eq. (7) can now be split into sub-correlations:

$$\Delta E_{PI\ sen} \Leftrightarrow RT \ln(FR_{PI\ sen}) \quad (8a)$$

$$\Delta E_{PI\ res} \Leftrightarrow RT \ln(FR_{PI\ res}) \quad (8b)$$

4.1.2. Brute correlation $\Delta E_{\Pi} \Leftrightarrow RT \ln(FR_{\Pi})$

Brute correlation Eq. (6) was first experimented in the ensemble Π including the 2959 HIV-1 complexes of the "Stanford HIV Drug Resistance Data Base − Feb 2004". Binding energies $E_{WP/PI}$ and $E_{MP/PI}$ necessary for obtaining ΔE_{PI} by Eq. (3) are computed by the automated version of GenMol.

Correlation is reported (Fig. 6) between $(RT \ln FR_{\Pi})$ and $(-\Delta E_{\Pi})$. A regression line forced to pass by the origin, according to the theoretical relation Eq. (4), remains in the vicinity of a free regression line. However, such a brute correlation is relatively loose, with regression coefficients $r^2 = 0.54$ and $r^2 = 0.61$.

4.1.3. Correlations $\Delta E_{PI} \Leftrightarrow RT \ln(FR_{PI})$ for the 6 PIs

Dispersion in brute correlation Eq. (6) can be investigated by correlation Eq. (7) $\Delta E_{PI} \Leftrightarrow RT \ln(FR_{PI})$ specific to each sub-ensemble Π_{PI}. For these correlations (Fig. 7) regression lines are free or forced to pass through the origin. Table 2

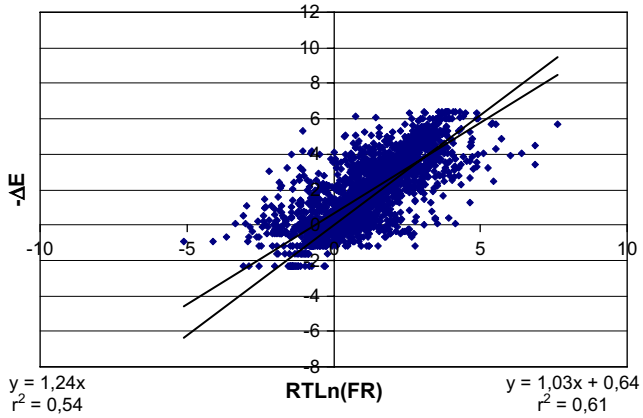

Fig. 6. Brute correlation in Π between $x = RT \ln(FR)$ and $y = -\Delta E$ with regression coefficients $r^2 = 0.54$ and $r^2 = 0.61$ on 2929 data (30 outliers were removed).

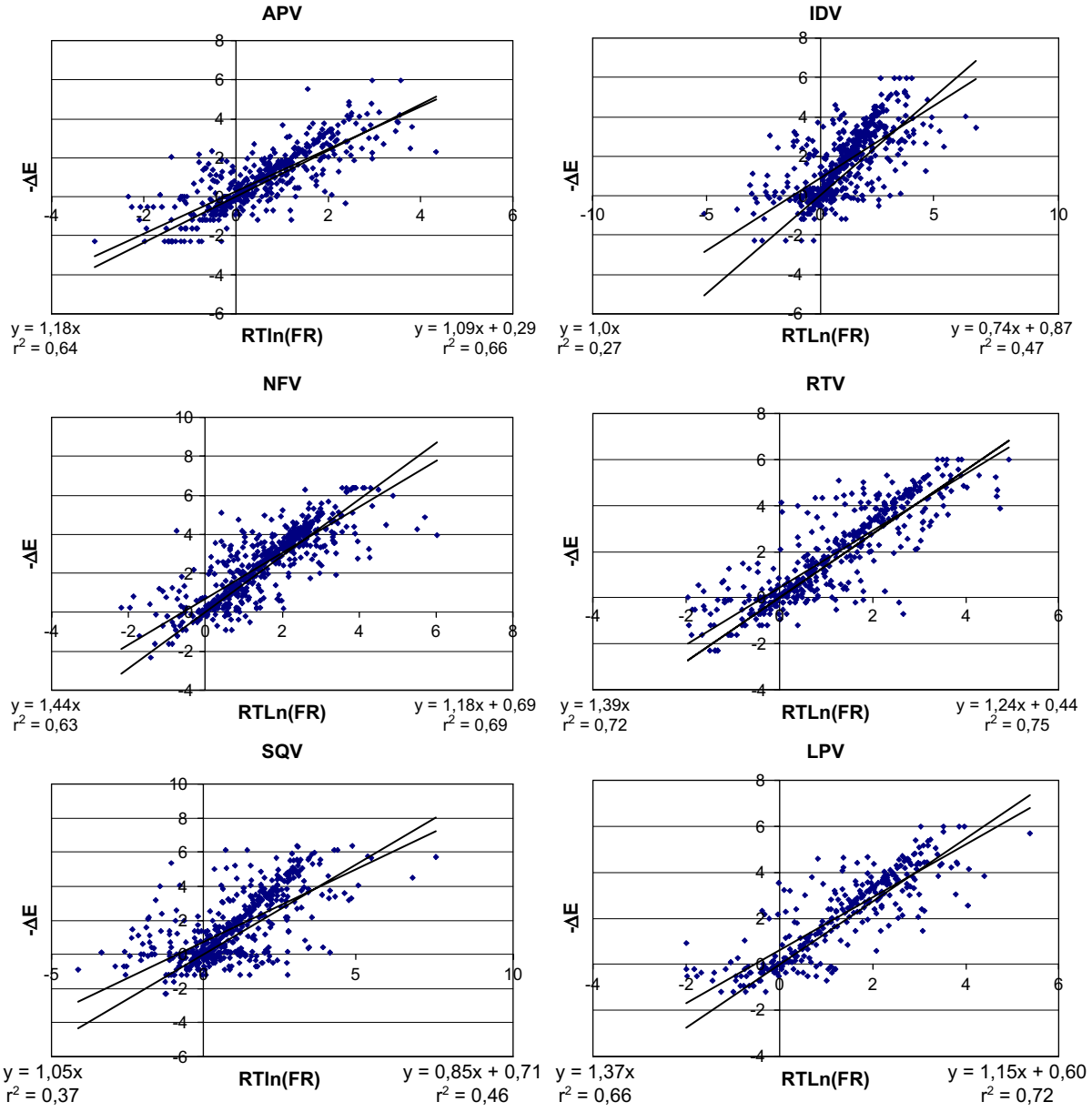

Fig. 7. Correlations between $x = RT\ln(\mathrm{FR})$ and $y = -\Delta E$ for the 6 PIs. Note: Aberrant points (“outliers”) of the distributions related to calculation convergence in false energy wells have been removed (4 for APV, 6 for IDV, 3 for NFV, 5 for RTV, 8 for SQV and 4 for LPV).

gives the resulting regression and concordance coefficients, both criteria indicating a good (bold) or average correlation.

4.1.4. Correlations $\Delta E_{\mathrm{PI\ sen}} \Leftrightarrow RT\ln(\mathrm{FR}_{\mathrm{PI\ sen}})$ and $\Delta E_{\mathrm{PI\ res}} \Leftrightarrow RT\ln(\mathrm{FR}_{\mathrm{PI\ res}})$

Finally “sensitive” $\Pi_{\mathrm{PI\ sen}}$ and “resistant” $\Pi_{\mathrm{PI\ res}}$ sub-ensembles were considered in each Π_{PI} ensemble.

Table 2
Performance analysis of correlations $\Delta E_{\mathrm{PI}} \Leftrightarrow RT\ln(\mathrm{FR}_{\mathrm{PI}})$

PI	APV	IDV	NFV	RTV	SQV	LPV
Number of cases	482	547	595	476	540	319
Free regression coefficient, r^2	**0.66**	0.47	**0.69**	**0.75**	0.46	**0.72**
	(478)	(541)	**(592)**	**(471)**	(532)	**(315)**
Concordance coefficient	**80**	74	**84**	**89**	69	**83**

4.1.4.1. Binding energies of MP/PI complexes. Fig. 8 gives the distribution of binding energies $E_{\mathrm{MP/PI}}$ for the $6\Pi_{\mathrm{PI\ sen}}$ and $6\Pi_{\mathrm{PI\ res}}$ sub-ensembles, i.e. the number of samples for each binding energy class. The width of each energy class is equal to 0.3 or 0.6 kcal/mol, in blue for “sensitive” complexes and in red for “resistant” complexes. Green arrows indicate the binding energy $E_{\mathrm{WP/PI}}$ between the wild protease and each PI.

Table 3 summarizes for the 6 PIs (column 1) binding energies with the wild protease $E_{\mathrm{WP/PI}}$ (column 3), average binding energies $<E_{\mathrm{MP/PI}}>_{\mathrm{sen}}$ with their standard deviation $\mathrm{SD}_{\mathrm{sen}}$ for complexes $\in \Pi_{\mathrm{PI\ sen}}$ (column 4), and similarly $<E_{\mathrm{MP/PI}}>_{\mathrm{res}}$ with $\mathrm{SD}_{\mathrm{res}}$ for complexes $\in \Pi_{\mathrm{PI\ res}}$ (column 5).

Simulating a phenotype test on a sample S leads to a particular value E_{simu} of the binding energy $E_{\mathrm{MP/PI}}$. To conclude on

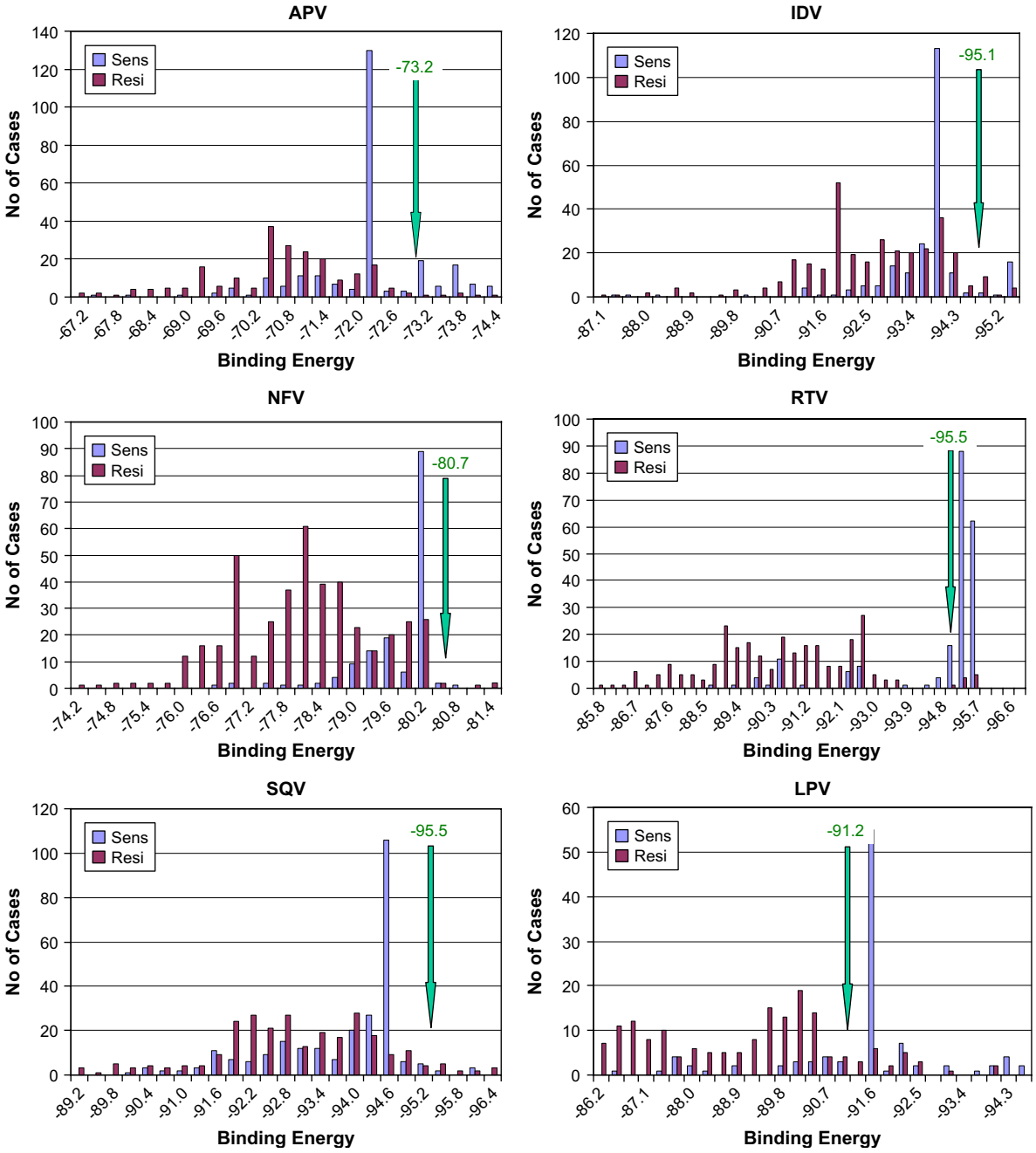

Fig. 8. Distribution of MP/PI binding energies (kcal/mol) for APV (482 structures), IDV (547 structures), NFV (595 structures), RTV (476 structures), SQV (540 structures) and LPV (319 structures). Blue refers to drug "sensitive" cases of $\Pi_{\text{PI sen}}$ and red to "resistant" cases of $\Pi_{\text{PI res}}$. Green arrows indicate PI binding energies with the wild protease. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

mutated virus sensitivity, the value E_S must simultaneously satisfy two conditions:

(i) E_{simu} must belong statistically to ensemble $\{E_{\text{MP/PI sen}}\}$ and
(ii) E_{simu} must be excluded statistically from ensemble $\{E_{\text{MP/PI res}}\}$.

Two cases may happen depending on the separation or imbrication of sensitive and resistant ensembles:

Case 1: statistical separation $\{E_{\text{MP/PI sen}}\} \cap \{E_{\text{MP/PI res}}\} = \{0\}$ at $(\text{SD}_{\text{sen}} + \text{SD}_{\text{res}})$

$$\left[< E_{\text{MP/PI}} >_{\text{sen}} + 1\ \text{SD}_{\text{sen}}\right] \text{less than} \left[< E_{\text{MP/PI}} >_{\text{res}} - 1\ \text{SD}_{\text{res}}\right] \quad (9)$$

To obtain a sensitivity diagnosis, avoiding resistance or dubious case between $\Pi_{\text{PI sen}}$ and $\Pi_{\text{PI res}}$, the necessary condition on E_{simu} is:

$$E_{\text{simu}} \leq \left[< E_{\text{MP/PI}} >_{\text{sen}} + 1\ \text{SD}_{\text{sen}} \right] \quad (10)$$

Table 3
Binding energies between PIs and wild protease (c3), mutated sensitive (c4) or mutated resistant (c5), plus upper binding energy limit insuring sensitivity (c6)

(c1)	(c2)	(c3)	(c4)	(c5)	(c6)
PI		$E_{WP/PI}$	$<E_{MP/PI}>_{sen} + 1\ SD_{sen}$	$<E_{MP/PI}>_{res} - 1\ SD_{res}$	$E_{MP/PI\ limit} = inf\ \{c4;c5\}$
APV	Energies (kcal/mol)	−73.2	−72.25 + 1.10	**−70.66 − 1.62**	−72.28
	Number of cases	482	224	258	
IDV	Energies (kcal/mol)	−95.1	−93.36 + 1.38	**−92.58 − 1.61**	−94.19
	Number of cases	547	222	325	
NFV	Energies (kcal/mol)	−80.7	−79.70 + 0.71	**−78.10 − 1.32**	−79.42
	Number of cases	595	161	434	
RTV	Energies (kcal/mol)	−95.5	**−94.45 + 1.36**	−90.30 − 2.00	−93.09
	Number of cases	476	207	269	
SQV	Energies (kcal/mol)	−95.5	−93.76 + 1.41	**−92.98 − 1.66**	−94.64
	Number of cases	540	267	273	
LPV	Energies (kcal/mol)	−91.2	−91.23 + 1.73	**−88.28 − 2.41**	−90.69
	Number of cases	319	105	214	

Inf $\{[<E_{MP/PI}>_{sen} + 1\ SD_{sen}];[<E_{MP/PI}>_{res} - 1\ SD_{res}]\}$ is bold-faced

Case 2: statistical imbrication $\{E_{MP/PI\ sen}\} \cap \{E_{MP/PI\ res}\} \neq \{0\}$ at $(SD_{sen} + SD_{res})$

$$\left[< E_{MP/PI} >_{res} - 1\ SD_{res}\right] \text{less than} \left[< E_{MP/PI} >_{sen} + 1 SD_{sen}\right] \quad (11)$$

To avoid resistance, the necessary condition on E_{simu} is:

$$E_{simu} \leq \left[< E_{MP/PI} >_{res} - 1\ SD_{res}\right] \quad (12)$$

This condition implies automatically sample $S \in \Pi_{PI\ sen}$ due to Eq. (11).

Finally, conditions Eqs. (10) and (12) can be condensed in a unique expression by defining a twofold upper sensitivity binding energy limit $E_{MP/PI\ limit}$:

$$E_{simu} \leq E_{MP/PI\ limit} = \inf\left\{\left[< E_{MP/PI} >_{sen} + 1\ SD_{sen}\right]; \left[< E_{MP/PI} >_{res} - 1\ SD_{res}\right]\right\} \quad (13)$$

Upper limits $E_{MP/PI\ limit}$ are reported for each PI in Table 3 (column 6).

4.1.4.2. Differences of binding energies between WP/PI and MP/PI complexes. To get intrinsic results, the above binding energies are reported to wild complex binding energies (Table 3, column 3). Fig. 9 gives the distribution of differences $\Delta E_{PI\ sen} = E_{WP/PI} - E_{MP/PI\ sen}$ (blue) and $\Delta E_{PI\ res} = E_{WP/PI} - E_{MP/PI\ res}$ (red) in kcal/mol. The "frequency" refers to the number of MP/PI structures relative to each ΔE class; curves are smoothed for clarity. Blue and red vertical lines indicate energy average $<\Delta E_{PI\ sen}>$ and $<\Delta E_{PI\ res}>$.

If correlations Eqs. (8a) and (8b) were perfectly distinct, blue "sensitive" $\Delta E_{PI\ sen}$ and red "resistant" $\Delta E_{PI\ res}$ distributions should be entirely disconnected but are in practice overlapping. Discriminating "sensitive" and "resistant" distributions can be analyzed by different means:

(i) Firstly, the difference between blue and red distributions can be appreciated to the eye (Fig. 9). A clear separation appears for RTV, LPV, NFV and APV but overlapping is more important for IDV and APV.
(ii) Then separating blue and red graphs can be based on the distance between average values $<\Delta E_{PI\ sen}>$ and $<\Delta E_{PI\ res}>$, mentioned in Fig. 9 by blue and red vertical lines. The larger their interval, the better the separation. Separation is good for RTV, LPV, NFV and APV (with separation, respectively, equal to 4.2, 3.0, 1.6 and 1.6 kcal/mol) but loose for IDV or SQV (separation equal to 0.8 kcal/mol).
(iii) Finally, comparison between blue and red distributions can be based on Tukey–Kramer's coefficient "p" which quantifies the statistical difference between two distributions. Correlation analysis is done according to the non-parametric Spearman's approach, ranking the two variable series and thus avoiding Gaussian assumption used in standard Pearson's approach. Values of "p" were found smaller than 0.0001 except "p" = 0.033 for IDV. All "p" values smaller than 0.05 indicate a significant separation between all $\Pi_{PI\ sen}$ and $\Pi_{PI\ res}$.

However, a strict separation between sensitive and resistant distributions is not required for applying the twofold criterion. Thus, whatever be the separation quality, sensitivity determination can be based on an intrinsic twofold criterion $\Delta E_{PI\ limit}$ deduced from Eq. (13):

$$\Delta E_{PI\ limit} = E_{WP/PI} - E_{MP/PI\ limit} = \sup\left\{\left[E_{WP/PI} - < E_{MP/PI} >_{sen} - SD_{sen}\right]; \left[E_{WP/PI} - < E_{MP/PI} >_{res} + SD_{res}\right]\right\} \quad (14)$$

Finally, whatever be the separation between $\Pi_{PI\ sen}$ and $\Pi_{PI\ res}$ sub-ensembles, sensitivity of the virus is discriminated by a unique condition when $E_{WP/PI}$ and E_{simu} are given by modeling:

$$\Delta E_{simu} = E_{WP/PI} - E_{simu} \geq \Delta E_{PI\ limit} \quad (15a)$$

or in absolute value because energy values are negative:

$$|\Delta E_{simu}| = \left|E_{WP/PI} - E_{simu}\right| \leq |\Delta E_{PI\ limit}| \quad (15b)$$

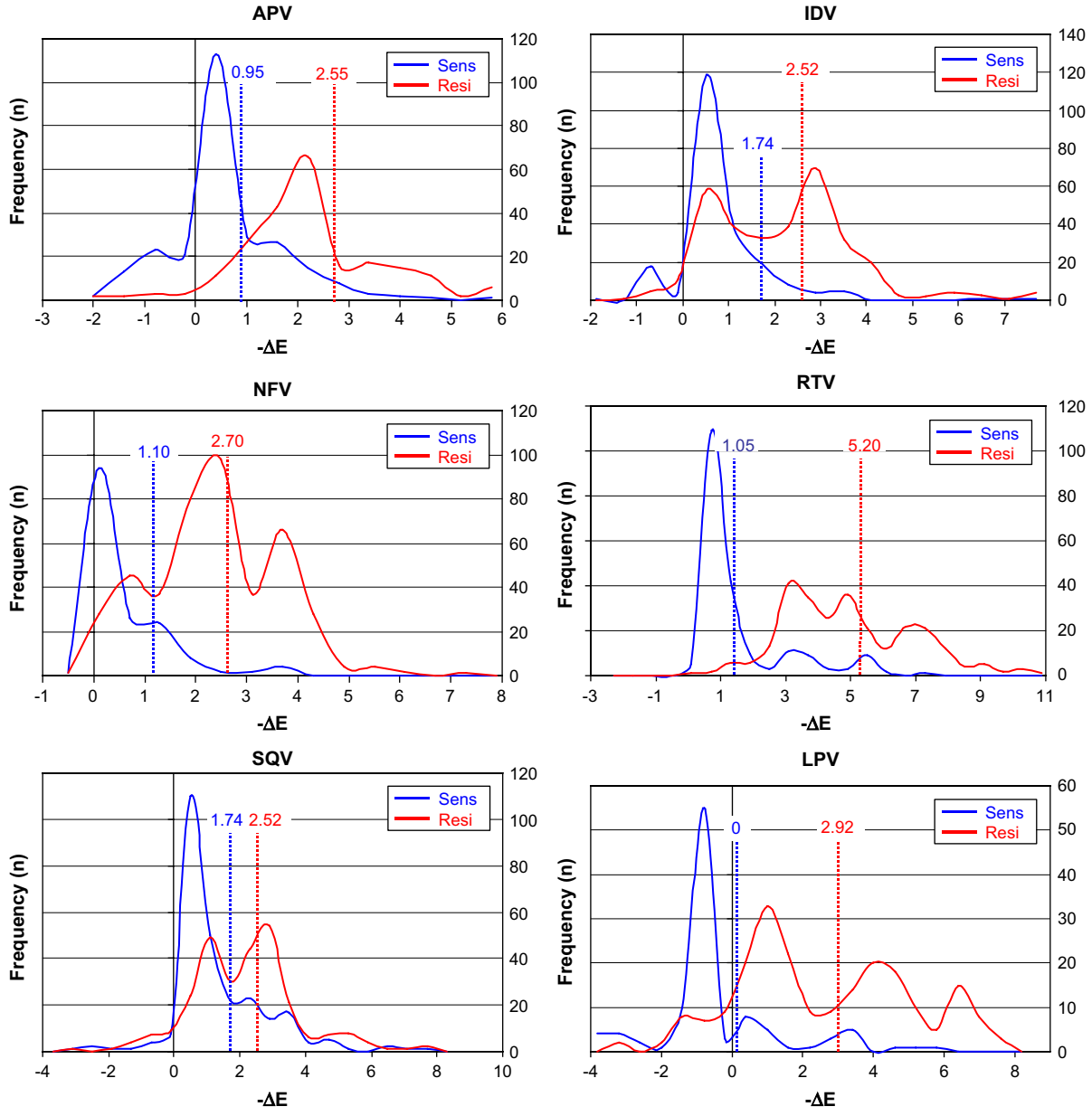

Fig. 9. Distribution of binding energy differences $\Delta E_{\text{PI sen}}$ and $\Delta E_{\text{PI res}}$ (kcal/mol) for the six "sensitive" $\Pi_{\text{PI sen}}$ (blue) and $\Pi_{\text{PI res}}$ "resistant" (red) sub-ensembles. Vertical lines refer to energy averages $<\Delta E_{\text{PI sen}}>$ (blue) and $<\Delta E_{\text{PI res}}>$ (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Application of Eq. (15a) to the different PIs leads to the following sensitivity conditions extracted from data given in Table 3:

$$\text{Sensitive to APV if } 0 \geq \Delta E_{\text{simu}} \geq \Delta E_{\text{APV limit}} = -73.2 + 72.28 = -0.9 \quad (16a)$$

$$\text{Sensitive to IDV if } 0 \geq \Delta E_{\text{simu}} \geq \Delta E_{\text{IDV limit}} = -95.1 + 94.19 = -0.9 \quad (16b)$$

$$\text{Sensitive to NFV if } 0 \geq \Delta E_{\text{simu}} \geq \Delta E_{\text{NFV limit}} = -80.7 + 79.42 = -1.3 \quad (16c)$$

$$\text{Sensitive to RTV if } 0 \geq \Delta E_{\text{simu}} \geq \Delta E_{\text{RTV limit}} = -95.5 + 93.09 = -2.4 \quad (16d)$$

$$\text{Sensitive to SQV if } 0 \geq \Delta E_{\text{simu}} \geq \Delta E_{\text{SQV limit}} = -95.5 + 94.64 = -0.9 \quad (16e)$$

$$\text{Sensitive to LPV if } 0 \geq \Delta E_{\text{simu}} \geq \Delta E_{\text{LPV limit}} = -91.2 + 90.69 = -0.5 \quad (16f)$$

4.1.5. Discussion of MMPP method and results

Replacing a phenotype test by a MMPP numerical run depends upon the fundamental correlation bridging the clinical

fold resistance FR and modeling results $\Delta E_{PI} = E_{WP/PI} - E_{MP/PI}$. This correlation depends first on the quality of the phenotype test, but above all on the quality of FR data. The quality of brute correlation and also correlation relative to each PI is limited by a number of missing parameters, such as treatments with different PIs, sequence of successive treatments and series of patients, etc. On the other hand, correlation depends of course on the quality of simulation. In this respect, our results can be compared with recent studies given in Refs. [8,9].

First of all, simulation techniques are quite different. The above cited studies are obtained by a technique of flexible docking and molecular dynamics, when GenMol™ is a molecular mechanics code, obtaining structures' optimization by an original energy minimization algorithm. Moreover an automatic and very efficient mutated protease/PI docking replaces very long molecular dynamics operations. The cpu time is divided by a coefficient of the order of 100 and a larger number of cases can be explored on any laptop. Moreover traditional approach can fail to converge for instance on cyclic chains.

Study [8] is performed on a limited number of complexes (78 protease variants). A good correlation is obtained between binding energy and fold resistance ($0.7 < r^2 < 0.8$) on the 6 PIs recorded in the Virologic data set. This kind of approach was extended by study [9] to a larger number of cases (1792 mutants), obtaining equivalent correlations with the Virologic data set. Our results lead to a similar correlation quality, at least for a global correlation as shown by comparing Fig. 5 in Ref. [8] and our Fig. 6. However, refined correlation for the 6 PIs are unfortunately not available in the above cited works to compare with our Fig. 7. It would have been interesting to confirm the disparity of correlations as function of the PIs (Fig. 7), compared to the global correlation (Fig. 6). Fortunately, these regression graphs are not directly used for modeling virus sensitivity prediction neither in cited works nor in our study.

Prediction of the drug effectiveness depends fundamentally on the sensitivity limit criterion. Criteria are slightly different in the above cited works and in our study. They remain anyway arbitrary for at least two reasons. First of all, the fold resistance limit between sensitive and resistant cases is arbitrary. To be on the safer side, we decided to take FR = 2.5 when Shenderovich et al. [8] preferred FR = 4. However, this safer choice seems to induce a more difficult separation between sensitive and resistant cases, as it appears comparing both studies in the following. This first choice being made, the second arbitrary choice is the criterion to declare sensitivity to a drug without any risk of resistance. Shenderovich et al. [8] are defining an equivocal zone for ΔE_{bind} between two cut-off $c1 = <\Delta E_{PI\ sen}> + 2\ SD_{sen}$ and $c2 = <\Delta E_{PI\ res}> - 1\ SD_{res}$, the difference c1 − c2 being found positive for all PIs ([8], Table 3). In the general case, this equivocal zone can disappear due to an imbrication between sensitive and resistant zones (c1 > c2). (Note that in Ref. [8] absolute values of binding energies are considered when we kept binding energies with their natural negative sign.) To take into account both separation or imbrication possibilities, we defined a twofold criterion $\Delta E_{PI\ limit}$ by Eq. (14). This criterion insures in any case to be in a sensitive case, avoiding dubious or resistant cases. A safe suggestion to the therapist is expressed by conditions Eq. (16a) to Eq. (16f). In our case, Table 3 shows that the second limit $[<E_{MP/PI}>_{res} - 1\ SD_{res}]$ prevails upon the first one $\{[<E_{MP/PI}>_{sen} + 1\ SD_{sen}]$ for all PIs, except for RTV where the separation between $\Pi_{PI\ sen}$ and $\Pi_{PI\ res}$ is the best. A sensitivity diagnosis based on a unique cut-off similar to c1 would be wrong for all PIs (except for RTV), because they include drug resistance possibility. Explaining the behaviour difference between RTV and other PIs escapes at first glance to a direct comparison between their structures (Fig. 10) and requires a deeper MMPP insight of binding energies as function of the PIs type.

4.2. The Molecular Modeling Genotype–Phenotype Protocol (MMGPP)

The preceding MMPP study is based on the gross binding energy between a protease, wild or mutated, with a given inhibitor. However, the complete knowledge of drug resistance given by the genotype–phenotype synthesis (Fig. 2) was not used. For exploiting these full clinical data including genotype tests, molecular modeling gives the opportunity to analyze the interaction between a drug molecule and the different residues, before or after mutation. Such is the purpose of the MMGPP approach. "Steric MMGPP" rests on a pure geometry analysis of the drug/residue couple, while "Energetic MMGPP" introduces the "interaction energy" between drug and residue.

4.2.1. Steric MMGPP

4.2.1.1. Steric analysis. Analysis of the X-ray image of a WP/PI structure allows quantifying the "distance" between a PI inhibitor molecule in the protease active site and a residue WR important for PIs' virus sensibility in the wild protease WP. This "distance" (PI–WR) is appreciated by the number of "contacts" between atoms of both entities. A "contact" is said to be existing when the distance between two atoms is lower than the sum of their van der Waals radii + 3 Å (calculation includes the residue main chain atoms). This first indication could be refined by comparing contacts in wild and mutated protease, but data will not be homogeneous between experimental WP structure and simulated MP structure. However, it was observed that generally the number of contacts is quite similar before and after mutation, even if residues and their environment are modified by mutation. Fig. 11 gives the number of "contacts" between PIs and residues WR important for PIs' virus sensibility for the 2959 HIV-1 complexes given in the "Stanford HIV Drug Resistance Data Base – Feb 2004".

4.2.1.2. Drug resistance and steric MMGPP. Comparing these results with the genotype–phenotype data (Fig. 2) leads to the following correlation between the drug resistance and the distance (PI–WR) for residues important for PIs' virus sensibility. All residues with more than one contact are manifesting a medium-level resistance when the number of contacts is

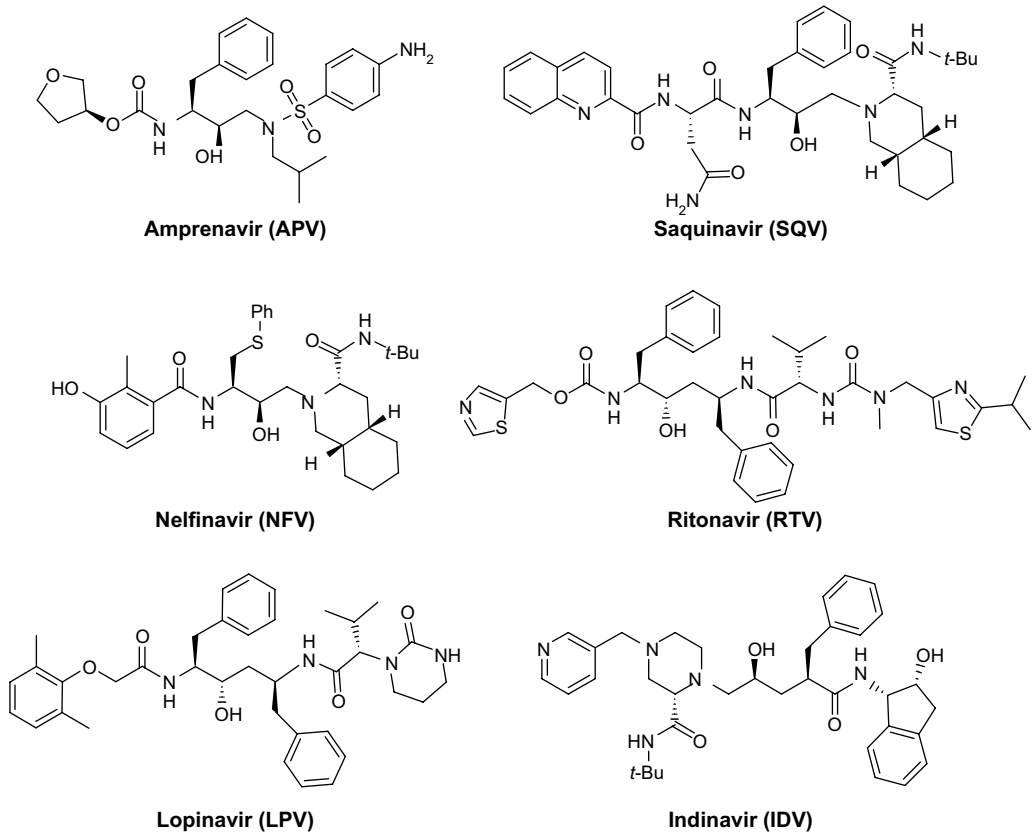

Fig. 10. Chemical structures of HIV-1 protease inhibitors.

small and high-level resistance when the number of contacts is high. A contrario, residues with no ''contact'' or 1 ''contact'' (L10, K20, D30, L33, M36, I50, L63, A71, G73, V77), far away from the active site of the wild protease, generally do not induce strong resistance when mutated. In Fig. 2 they are mentioned only as contributors to resistance when associated with other mutations. However residue L90, in this primary approach (see discussion below), appears as an exception, with a high-level resistance with all PIs and no ''contact'' before reaching a cut-off of 2 vdW radii + 6 Å.

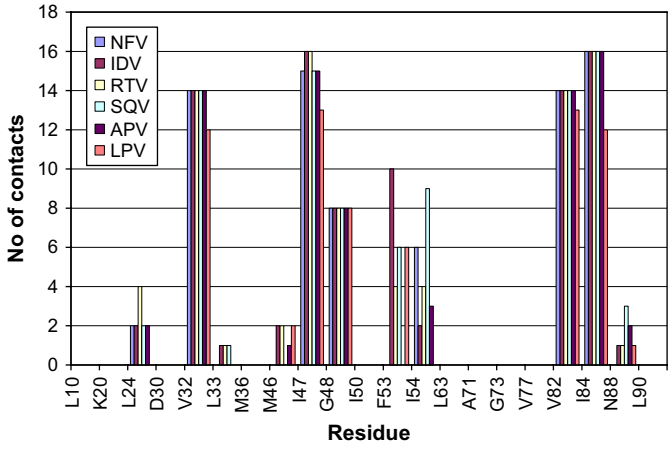

Fig. 11. Numbers of ''contacts'' between atoms of the 6 PIs and atoms of the 21 residues important for PIs' virus sensitivity, corresponding to inter-atomic distances lower than 2 vdW radii + 3 Å.

4.2.2. Energetic MMGPP

4.2.2.1. Interaction energy analysis. Another parameter that is able to take into account the relation between a PI and a residue is their interaction energy. Definition of interaction energy $IE_{WR/PI}$ or $IE_{MR/PI}$, between, respectively, a wild or mutated residue R and a PI, is similar to binding energy definition given by Eq. (2):

$$IE_{WR/PI} = E_{(WR+PI)\text{-opt}} - \left(E_{WR\text{-opt}} + E_{PI\text{-opt}}\right) + \text{arbitrary constant} \quad (17a)$$

$$IE_{MR/PI} = E_{(MR+PI)\text{-opt}} - \left(E_{MR\text{-opt}} + E_{PI\text{-opt}}\right) + \text{arbitrary constant} \quad (17b)$$

Thus, influence of a PI on a residue may be appreciated by the difference $\Delta IE_{R/PI}$

$$\Delta IE_{R/PI} = IE_{WR/PI} - IE_{MR/PI};$$
$$E_{WR/PI} = \text{interaction energy between WR and PI},$$
$$E_{MR/PI} = \text{interaction energy between MR and PI} \quad (18)$$

The parallel between MMPP protease binding energies and MMGPP residue interaction energies is illustrated in Fig. 12. These energies are related by the following relations:

$$E_{WP/PI} = \sum_{1\ to\ 99} IE_{WR/PI} = \left(IE_{W1/PI} + \cdots + IE_{WR/PI} + \cdots + IE_{W99/PI}\right) \quad (19a)$$

In case of a unique residue mutation of residue WR into MR, the total binding energy $E_{MP/PI}$ becomes:

$$E_{MP/PI} = \sum_{1\ to\ 99} IE_{MR/PI} = \left(IE'_{W1/PI} + \cdots + IE_{MR/PI} + \cdots + IE'_{W99/PI}\right) \quad (19b)$$

with $IE_{MR/PI}$ interaction energy relative to the mutated residue MR, and all other terms $IE'_{W1/PI}$ up to $IE'_{W99/PI}$ being interaction energies relative to non-mutated WR residues. However, $IE_{W1/PI} \neq IE'_{W1/PI}$, etc. because of the configuration change of the whole protease due to the unique MR mutation. This leads to:

$$\Delta E_{PI} = E_{WP/PI} - E_{MP/PI} = \sum_{1\ to\ 99} \Delta IE_{R/PI} = \left[\left(IE_{W1/PI} - IE'_{W1/PI}\right) + \cdots + \left(IE_{WR/PI} - IE_{MR/PI}\right) + \cdots + \left(IE_{W99/PI} - IE'_{W99/PI}\right)\right] \quad (20)$$

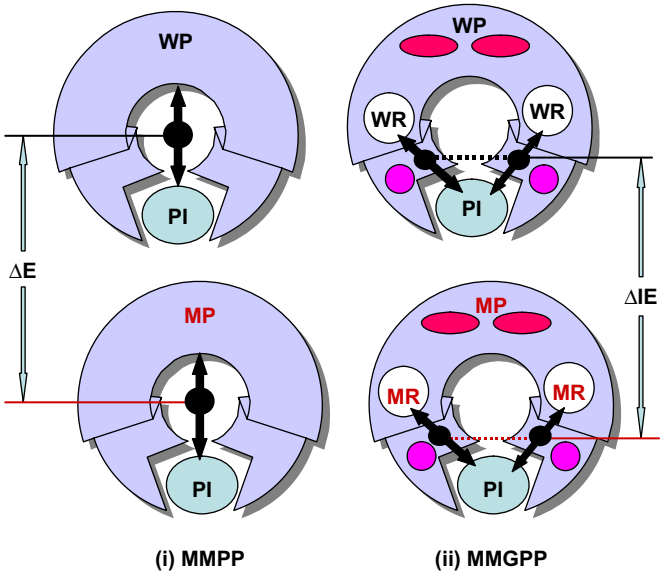

Fig. 12. Schematic binding and interaction energies computed by GenMol™: (i) MMPP (Molecular Modeling Phenotype Protocol) gives the energy difference $\Delta E_{PI} = E_{WP/PI} - E_{MP/PI}$ between binding energies of an inhibitor PI associated, respectively, with a wild protease WP and a mutated protease MP. Carriage return (ii) MMGPP (Molecular Modeling Genotype–Phenotype Protocol) gives the energy difference $\Delta IE_{R/PI} = IE_{WR/PI} - IE_{MR/PI}$ between interaction energies of an inhibitor PI associated, respectively, with a wild residue WR and a mutated residue MR.

4.2.2.2. Computing interaction energies. Interaction energies are computed starting from X–ray PI/protease complexes given in the PDB data bank [16]. The different operations are as follows:

- Start from the original complex WP/PI given in the PDB data bank, for instance 1HPV for the PI = APV. Make an energetic optimization of the complex WP/PI and extract the interaction energy $IE_{WR/PI}$ between the 21 wild residues WR (important for PIs' virus sensibility) and the 6 PIs.
- Then restarting from the wild original protease WP, generate the 21 possible monomutated protease MPs. Associate each MP with one of the 6 PIs and optimize the (21 × 6) MP/PI structures. For each optimized MP/PI, extract the interaction energy $IE_{WR/PI}$ between each of the 21 MRs and each of the 6 PIs.
- Finally deduce the interaction energy differences $\Delta IE_{R/PI} = IE_{WR/PI} - IE_{MR/PI}$ relative to the (21 × 6) couples (R, PI).

4.2.2.3. Interaction energies between PI = RTV and residues before any mutation. Fig. 13 gives the modulus of interaction energies $IE_{WR/PI}$ (WR = 1->99) according to the sequence order of the 99 wild residues of both chains contributing to the binding energy $E_{WP/PI}$. Dark values correspond to the 21 residues important for PIs' virus sensibility. Fig. 14 gives the same results but classified by increasing interaction energy values.

4.2.2.4. Impact of a residue mutation over ΔE_{PI} (MMPP) and $\Delta IE_{R/PI}$ (MMGPP). Table 4 gives, for PI = RTV, the impact on binding energies of the mutation of the 21 residues important for PIs' virus sensibility. Binding energies before mutation are given in column (c1) and after mutation in column (c2). Difference, before and after each monomutation, is given in column (c3) for binding energies relative to the total protease ($\Delta E_{PI} = E_{WP/RTV} - E_{MP/RTV}$) and in column (c4) for interaction energies of the mutated residue R ($\Delta IE_{R/RTV} = IE_{WR/RTV} - IE_{MR/RTV}$). The difference between variations of binding energy and interaction energy are given in column (c5). Column (c6) recalls intensity of drug resistance (Fig. 2), the number of stars being proportional to the resistance.

4.2.3. Discussion of steric and energetic MMGPP

Both steric and energetic MMGPPs investigate the relation between the drug and a mutated residue and its impact on the whole drug resistance phenomenon.

Steric MMGPP shows that in nearly all cases, drug resistance is directly influenced by the proximity of the mutated residue. Such is the case of I84, with a maximum number of contacts (Fig. 11) and a very high resistance with all PIs (Fig. 2). However, drug resistance created by mutation of L90 into L90M appears as a strong exception. This particularity can be explained by kinetics [20] or refined crystallographic studies of the protease [21,22]. Study [23] realized with SQV shows that the L90M mutation has a structural

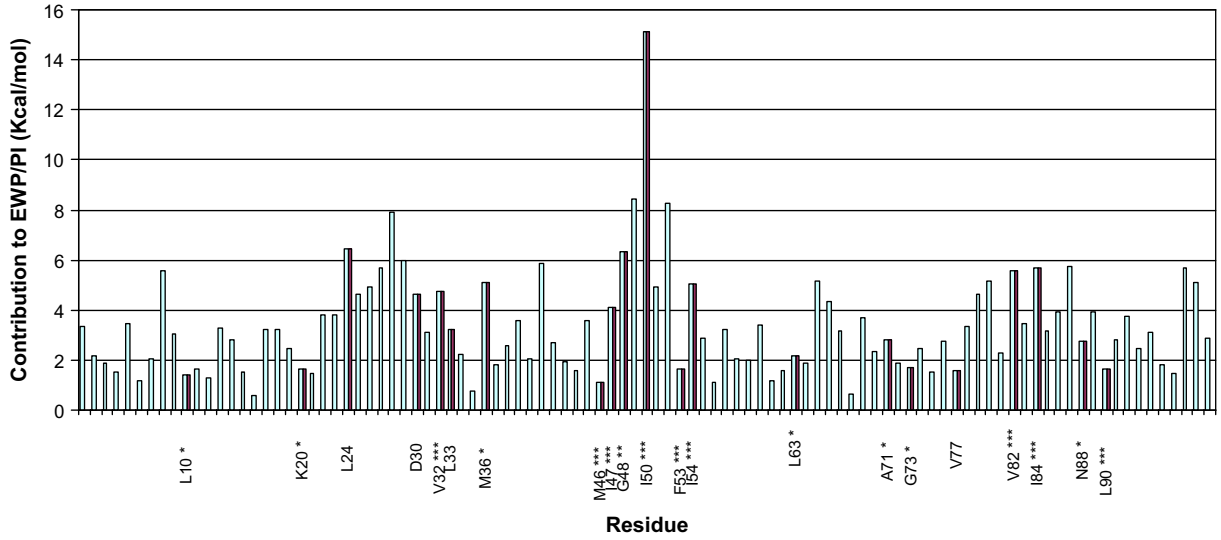

Fig. 13. Modulus of interaction energies $|E_{WR/PI}|$ of the 99 wild residues of both chains according to the sequence order contributing to the modulus of the wild protease binding energy $|E_{WP/PI}|$. Dark values correspond to the residues important for PIs' virus sensitivity.

effect reported up to L38–P44 in the opening and closing hinge region of the protease, thus indirectly acting on the escaping drug mechanism.

Energetic MMGPP approach confirms the general incidence of a mutated residue on its environment. The values in Table 4 show that the term ΔE_{PI} in Eq. (20) does not result uniquely from the term $(IE_{WR/PI} - IE_{MR/PI})$ relative to the single mutant residue because terms $(IE_{W1/PI} - IE'_{W1/PI})$ etc. are not equal to zero. All residues, important for PIs' virus sensibility or even other residues, can participate to the total binding energy difference $\Delta E_{PI} = E_{WP/PI} - E_{MP/PI}$. This explains why values in columns (c3) and (c4) are entirely different in Table 4. For instance, the mutation I47V induces a high resistance with an increase of 2.07 kcal/mol of the binding energy; however, the contribution of the interaction energy of this residue is only 0.67 kcal/mol. Contribution of interaction energies can even be opposite to binding energy variation as for F53L for instance. In other words a single mutation and "a fortiori" multi-mutations affect the whole protease energy distribution.

Finally, the complexity of a detailed study at the level of residues, as shown by the above steric or energetic approaches, appears as a justification of the statistical MMPP approach. Correlation between binding energy computation and clinical observations incorporates stochastically direct or indirect incidences of several mutations in the prediction of drug sensitivity.

5. Synthesis on MMPs contribution to drug resistance prediction

A first rapid test is the genotype X-ray test which gives the actual state of the virus, i.e. the ensemble $\{R_m\} = (R_1, R_2,$

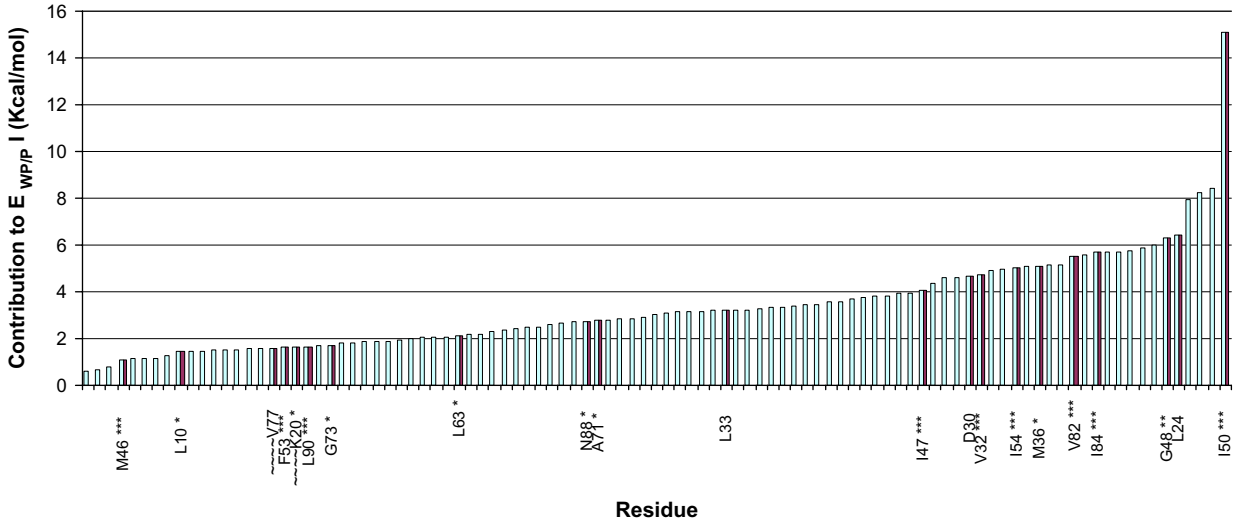

Fig. 14. Modulus of interaction energies $|E_{WR/PI}|$ of the 99 wild residues of both chains classified in increasing value order and contributing to the modulus of the wild protease binding energy $|E_{WP/PI}|$. Dark values correspond to the 21 residues important for PIs' virus sensitivity.

Table 4
Impact of a monoresidue mutation over differences $\Delta E_{RTV} = E_{WP/RTV} - E_{MP/RTV}$ between protease/RTV binding energies before and after mutation, compared to differences $\Delta IE_{R/RTV} = IE_{WR/RTV} - IE_{MR/RTV}$ between residue/RTV interaction energies before and after mutation (unit: kcal/mol)

(c1)	(c2)	(c3)	(c4)	(c5)	(c6)
Wild residue	Mutated residue	Difference of binding energy $E_{WP/RTV} - E_{MP/RTV}$	Difference of interaction energy $IE_{WR/RTV} - IE_{MR/RTV}$	Difference (c3) − (c4)	Drug resistance intensity (Fig. 2)
L10	L10I	−1.15	0.28	−1.43	*
K20	K20R	−1.79	−0.34	−1.45	*
L24	L24I	−1.03	−1.15	0.12	
D30	D30N	−0.41	−0.52	0.11	
V32	V32I	−1.99	−2.17	0.18	***
L33	L33F	0.67	0.07	0.60	*
M36	M36I	0.29	−2.69	2.98	*
M46	M46I	−0.60	0.04	−0.64	***
I47	I47V	−2.07	−0.67	−1.40	***
G48	G48V	0.90	1.81	−0.91	**
I50	I50L	−5.32	−4.40	−0.92	–
F53	F53L	−0.36	4.00	−4.36	***
I54	I54V	−0.92	−0.93	0.01	***
L63	L63P	−0.76	0.45	−1.21	*
A71	A71V	0.21	−0.47	0.68	*
G73	G73T	−0.65	2.52	−3.17	*
V77	V77i	−0.03	−0.01	−0.02	
V82	V82A	−1.00	−0.65	−0.35	****
I84	I84V	−1.53	−1.92	0.39	****
N88	N88D	0.58	2.08	−1.50	*
L90	L90M	−1.03	0.81	−1.84	***

…R_k,… R_n) of mutated residues. Based on this list, what is the information available on the resistance of a drug D_g using the data synthesis given in Fig. 2?

(a) If (R_k, D_g) ⊂ high resistance cases, for at least one $R_k \in \{R_m\}$, the drug D_g must be considered a priori as inefficient to remain on safety, even if other mutations can in some occasions attenuate this resistance. A phenotype test is generally not realized in that case.
(b) If (R_k, D_g) ⊄ high resistance cases for any $R_k \in \{R_m\}$:
  (b1) If only one R_k is mutated, and if (R_k, D_g) ⊂ low resistance or inexistent resistant cases, the drug can be used without necessitating confirmation by a phenotype test.
  (b2) But if several $R_k \in \{R_m\}$ are indicated with the mention ''intermediate or low-level resistance'' or ''contributes to resistance'', a phenotype test becomes absolutely necessary to judge the cumulative effect of mutated residues on the treatment by the drug D_g.

In this last eventuality, which is in fact the most general and most delicate case, modeling appears as an ideal tool. A great help can be brought to the therapist by suggesting, in quasi-real time and for practically no cost, a drug sensitivity diagnosis. Eventually this numerical diagnosis can be confirmed by a classical delayed and costly phenotype test.

6. Conclusion

Symbiosis between clinical genotype–phenotype tests and molecular modeling offers new potentialities for helping the therapist in the hard task of predicting HIV drug resistance.

To exploit all molecular modeling capacities, a first improvement consists of getting more detailed data bank (treatment chronology of each patient, multi-mutations and multi-drugs, etc.) and to refine consequently data analysis as proposed for instance by the mutation frequency index.

Concerning numerical tools, they should become fit for use exactly as other clinical tools. This supposes first rapid algorithms for avoiding huge cpu resources and a possible implementation on usual machines. Such is the case of GenMol™ with an optimization algorithm dividing classical atomic optimization cpu by a coefficient of the order of 100. Moreover GenMol™ possesses an automated module to fast generate 3-dim molecular structure of mutated proteins and their association with virus inhibitor. The next step consists of integrating MMPs into automatic and conversational simulators, immediately usable by any therapist on his laptop.

Combining such clinical data and efficient molecular modeling tools, a phenotype simulation is now possible in real time and at a very low cost by the Molecular Modeling Phenotype Protocol (MMPP). This issue is based on correlation between binding energy of a protease/drug complex and clinical fold resistance, plus a precise X-ray genotype. Prediction of virus sensitivity to a PI is obtained with a reasonable degree of confidence based on the twofold $\Delta E_{PI\ limit}$, insuring sensitivity and avoiding any possible dubious or resistant cases.

This MMPP statistical protocol is completed by the more detailed MMGPP study at the level of mutated residues. This refined steric or energetic study shows the complexity, even in monomutation cases, of the reorganization of a protease after a residue mutation. Finally MMGPP justifies MMPP sensitivity approach, operating on the total mutated protease configuration.

Numerical molecular modeling opens trends first in elaborating new therapy strategies, following each patient, and adapting the treatment to the mutation history of the patient's disease. Then numerical tools participate to improve database by more detailed and chronological information. Finally they appear as an eminent tool for a larger understanding of mutations and new drug development.

Acknowledgments

Authors thank G. Pénaranda for his help in statistical analysis.

Appendix A. Main abbreviations

$\langle$aaa$\rangle$ average value of "aaa" statistical variable
APV amprenavir
ARV antiretroviral drug
ATV atazanavir
CL convergence limit
CL1 convergence limit for protons' refinement
CL2 convergence limit for all atoms
2-D two-dimensional
3-D three-dimensional
E_{simu} energy value obtained by any simulation
$E_{MP/PI\ limit}$ twofold upper sensitivity binding energy limit
$E_{WP/PI}$ binding energy of wild protease complex WP/PI
$E_{MP/PI}$ binding energy of mutated protease complex MP/PI
FR fold resistance
FR_{lim} limit of FR between sensitivity and resistance ($FR \leq 2.5$)
FR_{PI} FR for any PI = $(IC_{50}MV)/(IC_{50}WV) = (KWTP)/(KMP)$
$FR_{PI\ sen}$ FR values relative to ensemble Π_{PI} sen
$FR_{PI\ res}$ FR values relative to ensemble Π_{PI} res
FR_{Π} FR relative to any sample of ensemble Π
HIV Human Immunodeficiency Virus
HIV-1 Human Immunodeficiency Virus-1
inf $\{a;b\}$ minimum value of both a,b values
$IC_{50}MV$ PI concentration for obtaining 50% inhibition of the MV
$IC_{50}WV$ PI concentration for obtaining 50% inhibition of the WV
IDV indinavir
$IE_{WR/PI}$ interacting energy of wild residue WR with PI
$IE_{MR/PI}$ interacting energy of mutated residue MR with PI
KMP association constant relative to a MP
KWTP association constant relative to a WP
LPV lopinavir
MF mutation frequency = (number of MR)/(total number of WR)
$MF(R)_{PI}$ mutation frequency of residue R in the presence of inhibitor PI
MMP Molecular Modeling Protocol
MMGPP Molecular Modeling Genotype–Phenotype Protocol
MMPP Molecular Modeling Phenotype Protocol
MP/PI Mutated Protease/Protease Inhibitor structure
$(MP/PI)_{obs}$ MP/PI observed structure by X-ray
$(MP/PI)_{simu}$ MP/PI docked structure obtained by simulation
$(MP/PI)_{obs\_opt}$ MP/PI observed structure after optimization
$(MP/PI)_{simu\ opt}$ $(MP/PI)_{simu}$ structure after optimization
MP Mutated Protease
MR Mutated Residue
MV Mutated Virus
NFV nelfinavir
p Tukey–Kramer coefficient
P pressure
PDB Protein Data Bank
PI Protease Inhibitor
(PI–WR) "distance" between PI and WR
r^2 regression coefficient
res resistant
R residue, i.e. amino-acid of a protease
R universal gas constant
RMSD Root Mean Square Distance
RTV ritonavir
sen sensitive
SQV saquinavir
S any sample
SD Standard Deviation
T temperature
vdW van der Waals
WP Wild Protease, before mutation
WR Wild Residue, before mutation
WV Wild Virus
$\Delta E_{PI\ limit}$ $E_{WP/PI} - E_{PI}$ limit
ΔE_{PI} $E_{WP/PI} - E_{MP/PI}$ for a given PI
$\langle \Delta E_{PI} \rangle$ average value of $\Delta E = \langle |E_{WP/PI} - E_{MP/PI}| \rangle$ for a given PI
$\Delta E_{PI\ sen}$ ΔE_{PI} relative to ensemble Π_{PI} sen
$\Delta E_{PI\ res}$ ΔE_{PI} relative to ensemble $\Pi_{PI\ res}$
ΔIE_{PI} $IE_{WR/PI} - IE_{MR/PI}$
ΔE_{Π} $E_{WP/PI} - E_{MP/PI}$ whatever be the PI in ensemble Π
Π ensemble of all cases found in a data bank
Π_{PI} ensemble of cases treated with a given PI
$\Pi_{PI\ sen}$ ensemble of cases sensitive to a PI treatment ($FR \leq 2.5$)
$\Pi_{PI\ res}$ ensemble of cases resistant to a PI treatment ($FR > 2.5$)

References

[1] G. Pèpe, D. Siri, GENMOL: A Fast Software for Molecular Modeling. Application to the Determination of the Psychotonic or Sedative Effects of Tricyclic Antidepressant Drugs, in: J.L. Rivail (Ed.), Modelling of Molecular Structure and Properties, Elsevier, Amsterdam, 1990, pp. 93–101.
[2] K. von der Helm, Retroviral proteases: structure, function and inhibition from a non-anticipated viral enzyme to the target of a most promising HIV therapy, Biol. Chem. 377 (12) (1996) 765–774.
[3] I.T. Weber, R.W. Harrison, Molecular mechanics analysis of drug-resistant mutants of HIV protease, Protein Eng. 12 (1999) 469–474.

[4] K. Lee, C.K. Chu, Molecular modeling approach to understanding the mode of action of L-nucleosides as antiviral agents, Antiviral Chem. Chemother. 9 (2001) 187–203.
[5] A. Altmann, N. Beerenwinkel, T. Sing, I. Savenkov, M. Däumer, R. Kaiser, S.-Y. Rhee, W.J. Fessel, R.W. Shafer, T. Lengauer, Improved prediction of response to antiretroviral combination therapy using the genetic barrier to drug resistance, Antiviral Ther. 12 (2007) 169–178.
[6] P. Volarath, R.W. Harrison, I.T. Weber, Structure based drug design for HIV protease: from molecular modeling to cheminformatics, Curr Top Med Chem 7 (10) (2007) 1030–1038.
[7] P.W. Erhardt, Medicinal chemistry in the new millennium. A glance into the future (special topic article), Pure Appl. Chem. 74 (5) (2002) 703–785.
[8] M.D. Shenderovich, R.M. Kagan, P.N.R. Heseltine, K. Ramnarayan, Structure-based phenotyping predicts HIV-1 protease inhibitor resistance, Protein Sci. 12 (2003) 1706–1718.
[9] E. Jenwitheesuk, R. Samudrala, Prediction of HIV-1 protease inhibitor resistance using a protein-inhibitor flexible docking approach, Antiviral Ther. 10 (2005) 157–166.
[10] G. Del Re, A simple MO–LCAO method for the calculation of charges distributions in saturated organic molecules, J. Chem. Soc. (1958) 4031–4040.
[11] G. Pèpe, B. Serres, D. Laporte, G. Del Re, C. Minichino, Surface electrostatic potentials on macromolecules in a monopole approximation: a computer program and an application to cytochromes, J. Theor. Biol. 115 (4) (1985) 571–593.
[12] F. Cavelier-Frontin, G. Pèpe, J. Verducci, D. Siri, R. Jacquier, Prediction of the best linear precursor in the synthesis of cyclotetrapeptides by molecular mechanics calculations, J. Am. Chem. Soc. 114 (1992) 8885–8890.
[13] G. Pèpe, M. Meyer, P. Faury, J.C. Graciet, J.C. Chermann, J.L. Kraus, A model allowing the design of modified nucleosides as HIV-RT inhibitors, Eur. J. Med. Chem. 31 (1996) 775–786.
[14] C.H. Lin, N. Gabas, J.P. Canselier, G. Pèpe, Prediction of the growth morphology of aminoacid crystals in solution. I. α-Glycine, J. Cryst. Growth 191 (1998) 791–802.
[15] J. Courcambeck, M. Bouzidi, R. Perbost, B. Jouirou, N. Amrani, P. Cacoub, G. Pèpe, J.M. Sabatier, P. Halfon, Resistance of Hepatitis C Virus to NS3–4A protease inhibitors: mechanims of drug resistance induced by R155Q, A156T, D168A and 168V mutations, Antiviral Ther. 11 (2006) 847–855.
[16] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.
[17] P.F.W. Stouten, C. Frömmel, H. Nakamura, An effective solvation term based on atomic occupancies for use in protein simulations, Mol. Simul. 10 (1993) 97–120.
[18] M. Schapira, M. Totrov, R. Abagyan, Prediction of the binding energy for small molecules, peptides and proteins, J. Mol. Recognit. 12 (1999) 177–186.
[19] M.D. Shenderovich, V. Tseitin, C.L. Fisher, K. Ramnarayan, Molecular Modeling and Binding Energy Calculations for Drug Resistant Mutations in HIV-1 Protease-inhibitor Complexes, in: M. Lebl, R.A. Houghten (Eds.), Peptides: The Wave of the Future. Proceedings of the Second International and the 17th American Peptide Symposium, American Peptide Society, San Diego, CA, 2001, pp. 418–419.
[20] B. Maschera, G. Darby, G. Palú, L.L. Wright, M. Tisdale, R. Myers, E.D. Blair, E.S. Furfine, Human immunodeficiency virus. Mutations in the viral protease that confer resistance to saquinavir increase the dissociation rate constant of the protease–saquinavir complex, J. Biol. Chem. 271 (52) (1996) 33231–33235.
[21] P. Martin, J.F. Vickrey, G. Proteasa, Y.L. Jimenez, Z. Wawrzak, A. Winters, T.C. Merigan, L.C. Kovari, "Wide-open" 1.3 Å structure of a multidrug-resistant HIV-1 protease as a drug target, Structure 13 (12) (2005) 1887–1895.
[22] B. Mahalingam, Y.F. Wang, P.I. Boross, J. Tozser, J.M. Louis, R.W. Harrison, I.T. Weber, Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site, Eur. J. Biochem. 271 (8) (2004) 1516–1524.
[23] L. Hong, X.C. Zhang, J.A. Hartsuck, J. Tang, Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: insights into the mechanisms of drug resistance, Protein Sci. 9 (10) (2000) 1898–1904.