Supporting Information

Exploring the Drug Resistance of HCV Protease

Garima Jindal,[a‡] Dibyendu Mondal[a‡] and Arieh Warshel[a]*

[‡]Equal Contribution

[a]Department of Chemistry, University of Southern California, 3620 McClintock Avenue, Los Angeles, California 90089, USA

Email (Corresponding Author): warshel@usc.edu

EVB Simulations

The EVB method is widely used by our group as well as others and provides a fast and effective approach to study enzymatic catalysis.[1-2] Here we only provide a concise description of the method and complete details of the method can be found elsewhere.

The potential energy surface is represented as a combination of two force fields. The classical ENZYMIX force field is used to simulate the protein that does not take part in the reaction, whereas a quantum empirical valence bond force field is used to treat the reaction centre. The quantum force field is used to represent the changes in the electronic charges as one moves from one diabatic state to another. The ground state potential energy surface is constructed by mixing the two diabatic states that represent the reactant and product.

In the EVB force field, the bonded atoms are defined using a Morse potential. It also includes repulsive functions for atoms that are not bonded along with functions for angles, torsions etc. For a two diabatic state system, the actual ground state potential is represented as:

$$Eg = c_1^2\varepsilon_1 + c_2^2\varepsilon_2 + 2(c_1^2c_2^2)^{1/2}H_{12} \quad \text{(S1)}$$

ε_1 and ε_2 represent the potential of the two diabatic states.

The coefficients are determined by diagonalizing the matrix,

$$H = \begin{pmatrix} \varepsilon_1 & H_{12} \\ H_{12} & \varepsilon_2 \end{pmatrix} \quad \text{(S2)}$$

The mixing term H_{12} is represented as:

$$H_{12} = Aexp\{-\mu R\} \quad \text{(S3)}$$

where, A and u are empirical constants that are calibrated using experimental free energy profile. The H12 parameters are assumed to be the same in the protein and solution.

To simulate the bond formation/breaking between two EVB states, MD simulations are first carried out on a mapping potential which is represented as:

$$\varepsilon_m = (1-\lambda_m)\varepsilon_1 + \lambda_m\varepsilon_2 \qquad (0 \le \lambda_m \le 1) \quad \text{(S4)}$$

where, λ_m is a parameter that is changed from 0 to 1 in N+1 windows as the initial state is changed to the final state. The Free Energy perturbation (FEP) is used to calculate the free energy change between two consecutive steps:

$$\Delta G_{m\to m+1} = -\beta^{-1}ln\left\langle e^{\{-\beta[\varepsilon_m(\lambda_{m+1})-\varepsilon_m(\lambda_m)]\}}\right\rangle_m \quad \text{(S5)}$$

$< >_m$ represents an average over different configurations when system moves on the ε_m potential.

To calculate the activation free energy, the free energy functional that represents the adiabatic ground state surface is used:

$$\Delta g(x') = \sum_{m=0}^{i-1} \Delta G_{m\to m+1} - \beta^{-1} ln\left\langle \delta(x - x') e^{\{-\beta[E_g - \varepsilon_m(\lambda_i)]\}} \right\rangle_{\varepsilon_m} \quad \text{(S6)}$$

where, $\sum_{m=0}^{i-1} \Delta G_{m\to m+1}$ = Free energy difference between the first and *i*th mapping potential, E_g = Energy of the ground state and δ = Dirac delta function and the inner broken brackets. The diabatic free energy profiles of the reactant and product represent microscopic equivalent[3] of the Marcus parabolas.[4]

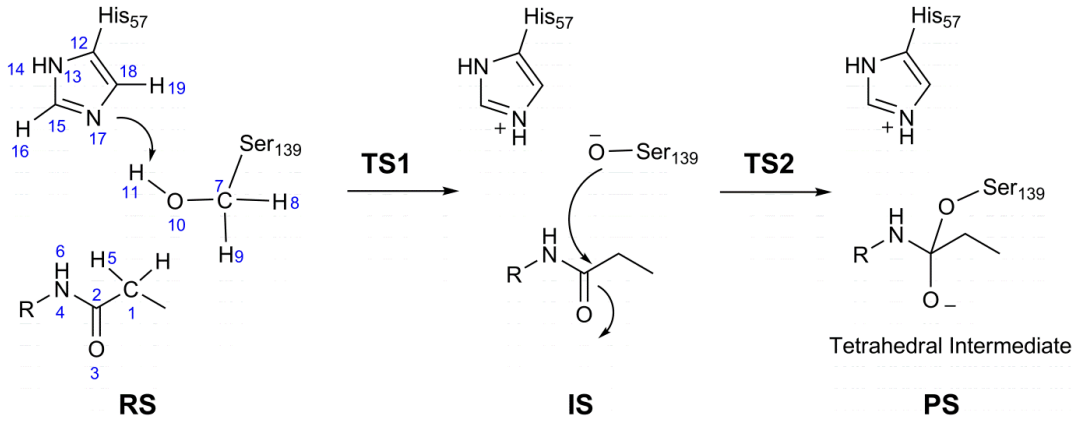
Practical Simulations

The EVB parameters were first calibrated by carrying out the reaction in solution. The calibration was done against values obtained from previous ab initio studies (ref. 38 of main manuscript). This study is one of the most elaborate and careful investigation of amide hydrolysis in the presence of bases such as water, histidine and ammonia. For our calculations, we took the barrier with histidine as the base. The reaction free energy for the first proton transfer step is taken to be 10.0 kcal/mol and the overall activation energy is 26.0 kcal/mol. The free energies of all systems are provided in Table S1. The charges for the EVB states are calculated using the B3LYP functional, which has been used previously in several of our earlier studies and the results have been in good agreement to experiments. A large basis set; 6311+G** is used to obtain correct geometries.

Table S1. Reaction Free Energies and Activation Free Energies in kcal/mol for the Proton Transfer and Nucleophilic Step

System	ΔG^{o}	$\Delta G^{\ddagger}$ (nucleophilic step)	$\Delta G^{\ddagger}$ (calc)	$\Delta G^{\ddagger}$ (obs)
Water	9.2	15.7	24.9	26.0
wt	9.1	9.2	18.3	18.0
R155Q	8.5	9.3	17.8	18.9
R155K	9.8	8.5	18.3	-
D168A	9.4	8.8	18.2	18.3
D168V	9.6	8.9	18.5	17.9
A156T	10.2	8.8	19.0	17.8
S138T	9.9	8.4	18.3	-
Q41R	9.7	8.0	17.7	-
F43V	9.2	9.0	18.2	-

It should be noted that the formation of the tetrahedral formation is considered to take place in a two step process as shown in Scheme S1. Both steps were calibrated separately using our EVB calculations. The charges used for all diabatic states are provided in Table S2a and other parameters are given in Tables S2b-g.



Scheme S1. Atom numbering scheme for different diabatic states involved in the formation of the tetrahedral intermediate.

Table S2. EVB parameters

a- Atomic Charges for the Reactant, Transition and Product States

Atom	Atom Name	RS	IS	PS
1	C	0.170985	0.170985	0.296533
2	C	0.477748	0.477748	0.807678
3	O	-0.661652	-0.661652	-0.938605
4	N	-0.166045	-0.166045	-0.858822
5	H	0.076321	0.076321	0.100735
6	H	0.199652	0.199652	0.306903
7	C	0.272519	0.749443	-0.098104
8	H	0.010004	-0.177708	0.068716
9	H	0.01004	-0.178760	0.061468
10	O	-0.736564	-1.213810	-0.47286
11	H	0.445078	0.375120	0.375120
12	C	0.251007	0.331709	0.331709
13	N	-0.350725	-0.198957	-0.198957
14	H	0.359219	0.383266	0.383266
15	C	0.233853	-0.017313	-0.017313
16	H	0.111936	0.242060	0.242060
17	N	-0.650642	-0.095596	-0.095596
18	C	-0.023670	-0.283507	-0.283507
19	H	0.133826	0.251631	0.251631

b- Morse Bond Parameters ;$\delta M(b) = D_M(1 - e^{-\mu(b-b_0)})^2$

Bond Type	D_M	b_0	μ
C-H	98.3	1.10	2.0
C-N	94.0	1.40	2.0
C-C	96.0	1.54	0.8
N-H	98.3	1.10	2.0
C7-O10	93.0	1.50	0.8
C2-O3 (RS)	93.0	1.25	2.0
C2-O3 (PS)	93.0	1.40	2.0
C15-H16/C18-H19	100.4	1.10	2.0

c- Angle Parameters; $V_\theta(\theta) = 1/2\, k_\theta(\theta - \theta_0)^2$

Angle type	**½ k_θ**	**θ_0**
C-N-C	50.0	120.0
C-N-H	50.0	120.0
C-C-N	50.0	120.0
H-C-N	50.0	120.0
C-C-O	50.0	120.0
C-C-O (PS)	50.0	109.4
N-C-O	50.0	120.0
N-C-O (PS)	50.0	109.4
C-O-C (PS)	80.0	109.4

d-. Dihedral parameters; $V_\varphi(\varphi) = k_\varphi(1 + \cos(n\varphi - \varphi_0))$

Dihedral Type	**k_φ**	**n**	**φ_0**
H-C-C-O	2.0	3.0	0.0
H-C-C-O (PS)	1.0	3.0	0.0
A-B-C-D	15.0	2.0	180.0

e-. Improper torsion parameters; $V_\varphi(\varphi) = k_\varphi(1 + \cos(n\varphi - \varphi_0))$

Dihedral Type	**k_φ**	**n**	**φ_0**
H-N-C-C	30.0	2	180.0
C-N-N-H	15.0	2	180.0
N-C-C-H	30.0	2	180.0
C-N-C-H	15.0	2	180.0
C-O-C-N	15.0	2	180.0
C-O-C-N (PS)	30.0	2	180.0

f- Nonbonded Parameters (EVB atom wise parameters for atoms bonded in one of the EVB states)

$$V_{\text{nb}} = \sqrt{C_A * C_B} e^{-\sqrt{\beta_A \beta_b} * r}$$

Atom Type	**C**	**β**
H0	5.0	2.5
C+	91.0	2.5
N+	60.0	2.5
O0	53.0	2.5
O-	90.0	2.5
C0	91.0	2.5

g- Nonbonded Parameters (EVB atom wise parameters for atoms never bonded)

$$V_{\text{nb}} = \frac{\text{vdwa}_A * \text{vdwa}_B}{r^{12}} - \frac{\text{vdwb}_A * \text{vdwb}_B}{r^6}$$

Atom Type	**vdwa**	**vbwb**
H0	7.0	0.0
C+	632.0	24.0
N+	774.0	24.0
O0	774.0	24.0
O-	1400.0	24.0
C0	632.0	24.0

Other EVB parameters

The off diagonal elements are represented using an exponential function:

$$H_r = \text{Ae}^{-\mu r}$$

For the proton transfer step:

A = 35.5, μ = 2.5

gas shift 2, α = 69.0 (Gas phase shift 2 corresponds to shifting the parabola corresponding to the EVB state 2)

For the nucleophilic attack:

A = 50.0, μ = 2.5

gas shift 2, α = 8.0

Table S3. RMSD Values of Mutant Proteins with the Natural Substrate with respect to the Wild Type

System	RMSD/Å
R155K	0.333
R155Q	0.320
A156T	0.373
D168A	0.352
D168V	0.330

PDLD Binding Free Energy Calculations:

We have used PDLD/S-LRA[5-6] and PDLD/S-LRA/β[7] methods to calculate the binding free energies for all drug-protein as well as substrate-protein complexes. In both methods, explicitly represented water molecules (used during MD simulation) are replaced by langevin dipoles (LD) to calculate the electrostatic contribution to the total free energy. The cycle shown in figure 4 of ref. 6 is used to calculate the binding free energies. The effective PDLD/S potential for a single protein-ligand configuration is calculated using the following equations:

$$U^{p}_{elec,l} = \left\{ \left(\Delta G^{l+p}_{sol} - \Delta G^{l'+p}_{sol} \right) \left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w} \right) + \Delta G^{l}_{sol} \left(1 - \frac{1}{\varepsilon_p} \right) + \frac{U^{l}_{q\mu}}{\varepsilon_p} + \frac{U^{l}_{intra}}{\varepsilon_p} \right\}_{B} \qquad \text{(S7)}$$

$$U^{w}_{elec,l} = \left\{ \Delta G^{l}_{sol} \left(\frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w} \right) + \Delta G^{l}_{sol} \left(1 - \frac{1}{\varepsilon_p} \right) + \frac{U^{l}_{intra}}{\varepsilon_p} \right\}_{UB} \qquad \text{(S8)}$$

where, ΔG^{i}_{sol} denotes the solvation free energy of the "*i*" group in water. Ideally, ΔG_{sol} should be scaled by $1/(1-1/\varepsilon_w)$, but due to high dielectric constant of water (ε_w) this correction is neglected here. $U^{l}_{q\mu}$ is the electrostatic interaction between the ligand and the protein dipoles in vaccum (standard PDLD notation). U^{l}_{intra} is the intramolecular electrostatic interaction for the ligand. The interaction energy, U is obtained from a single protein-ligand complex configuration, and therefore it does not properly represent the protein reorganization. The linear response approximation (LRA) is therefore used to capture the protein reorganization.[8-9] In this approximation, we calculate the average of the effective potential U over the

trajectories of the protein-ligand complex in their polar form as well as nonpolar form. Thus, in PDLD/S-LRA the electrostatic free energy can be expressed as,

$$\Delta G_{bind}^{elec/PDLD/S-LRA} = \frac{1}{2}\left[\left\langle U_{elec,l}^{p}\right\rangle_{l} + \left\langle U_{elec,l}^{p}\right\rangle_{l'} - \left\langle U_{elec,l}^{w}\right\rangle_{l} - \left\langle U_{elec,l}^{w}\right\rangle_{l'}\right] \quad \text{(S9)}$$

where the $<>_i$ denotes the MD average at the *i*th state.

The non-electrostatic energy calculations in PDLD/S-LRA consider the hydrophobic contribution (using field dependent hydrophobic term),[5] effect of water penetration and van der Waals effect explicitly. In the case of PDLD/S-LRA/β method, the non-electrostatic energy is calculated by scaling the van der waals interaction energy of the polar form of the ligand with β=0.25. Thus, the following equation is used to calculate the binding free energy using PDLD/S-LRA/β:

$$\Delta G_{bind}^{PDLD/S-LRA/\beta} = \Delta G_{bind}^{elec/PDLD/S-LRA} + \beta\left[\left\langle U_{vdw,l}^{p}\right\rangle_{l} - \left\langle U_{vdw,l}^{w}\right\rangle_{l}\right] \quad \text{(S10)}$$

Note that configurational entropic contribution has not been accounted in our current study.

Vitality Calculations

The vitality value[10-12] can be represented as:

$$\gamma = K_i\left(\frac{k_{cat}}{K_M}\right) \quad \text{(S11)}$$

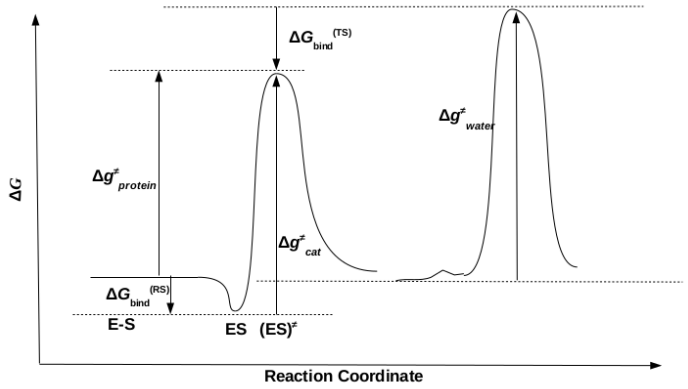


Figure S1. Key parameters that define the catalytic effect. E = enzyme, S = Substrate E + S = Enzyme + Substrate isolated in water, ES = Enzyme substrate complex and $(ES)^{‡}$ = enzyme substrate complex at the transition state.

Using Figure S1 the following quantities can be defined:

$$K_M = K_{bind}(RS) \tag{S12}$$

$$K_{bind}(RS) = \exp\left(-\Delta G_{bind}(RS)/RT\right. \tag{S13}$$

$$k_{cat} = \left(\frac{k_B T}{h}\right) \exp\left(-\frac{\Delta G^{\ddagger}_{cat}}{RT}\right) \tag{S14}$$

$$k_{water} = \left(\frac{k_B T}{h}\right) \exp\left(-\frac{\Delta G^{\ddagger}_{water}}{RT}\right), where\ \Delta G^{\ddagger}_{water} = -RTln\ k_{water} + RTln\left(\frac{k_B T}{h}\right) \tag{S15}$$

$$K_i = \exp\left(-\frac{\Delta G_{bind}(drug)}{RT}\right) \tag{S16}$$

where, ΔG_{bind}(RS) and ΔG_{bind}(TS) are the binding free energies of the reactant and transition states respectively, K_i is the inhibition constant of the drug, $\Delta g^{\ddagger}$ represent the activation barriers for different systems, T is temperature, k_B is the Boltzmann constant, h is the Planck constant and R is the gas constant.

$$\ln\left(\frac{\mathrm{k_{cat}}}{K_{bind}(RS)}\right) = \ln\left(\frac{k_B T}{h}\right) - \Delta g^{\ddagger}_{protein}/RT \text{ S17} \tag{S17}$$

where, $\Delta g^{\ddagger}_{protein} = \Delta g^{\ddagger}_{cat} - \Delta G_{bind}(RS)$

$$\Delta g^{\ddagger}_{protein} = -RTln\left\{\frac{k_{cat}}{K_{bind}(RS)}\right\} + RTln\left(\frac{k_B T}{h}\right) \tag{S18}$$

$$\Delta G_{bind}(TS) = \Delta g^{\ddagger}_{protein} - \Delta g^{\ddagger}_{water} \tag{S19}$$

Using equations (S15) and (S18),

$$\Delta G_{bind}(TS) = -RTln\left\{\frac{k_{cat}}{K_{bind}(RS)}\right\} + RTlnk_{water} \tag{S20}$$

Using equations (S17) and (S20),

$$ln\left\{\frac{k_{cat}}{K_{bind}(RS)}\right\} = lnk_{water} - \left\{\frac{\Delta G_{bind}(TS)}{RT}\right\} \quad \text{(S21)}$$

Using equation (S11),

$$ln\gamma = lnK_i + \ln\left(\frac{k_{cat}}{K_M}\right) \quad \text{(S22)}$$

$$ln\gamma = lnK_i + \ln\left(k_{cat} + K_{bind}(RS)\right) \quad \text{(S23)}$$

Using equations (S16) and (S21),

$$ln\gamma = \left(\frac{1}{RT}\right)\{\Delta G_{bind}(drug) - \Delta G_{bind}(TS)\} + lnk_{water} \quad \text{(S24)}$$

The vitality value of a mutant with respect to the native protein can finally be written as:

where m = mutant and n = native protein.

References

(1) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*. New York, 1991.
(2) Kamerlin, S. C. L.; Warshel, A. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 30-45.
(3) King, G.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682-8692.
(4) Marcus, R. A. *J. Chem. Phys.* **1965**, *43*, 679-701.
(5) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161-185.
(6) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. *Proteins: Struct. Funct. Bioinform.* **2000**, *39*, 393-407.
(7) Singh, N.; Warshel, A. *Proteins Proteins: Struct. Funct. Bioinform.* **2010**, *78*, 1705-1723.
(8) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. *Protein Eng.* **1992**, *5*, 215-228.
(9) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764*, 1647-1676.
(10) Gulnik, S. V.; Suvorov, L. I.; Liu, B.; Yu, B.; Anderson, B.; Mitsuya, H.; Erickson, J. W. *Biochemistry* **1995**, *34*, 9282-9287.
(11) Singh, N.; Frushicheva, M. P.; Warshel, A. *Proteins: Struct. Funct. Bioinform.* **2012**, *80*, 1110-1122.
(12) Ishikita, H.; Warshel, A. *Angew. Chem. Int. Ed.* **2008**, *47*, 697-700.