Supporting Information

Exploring the Mechanism of Covalent Inhibition: Simulating the Binding Free Energy of $\alpha$-Ketoamide Inhibitors of the Main Protease of SARS-CoV-2

Dibyendu Mondal, Arieh Warshel*

Department of Chemistry, University of Southern California, 3620 McClintock Avenue, Los Angeles, California 90089, United States

*warshel@usc.edu

S1. The Empirical Valence Bond (EVB) Method

The theoretical background of the EVB method has been extensively discussed in many of our previous works.\textsuperscript{1-2} Here we are mentioning only a few aspects of the method to explain the functional forms of the EVB potentials and its parameters.

The EVB method is used to study the energetics of a chemical process and the reacting system is generally represented as a superposition of two (or more) diabatic states. Each state is expressed as a combination of two different force fields. Those are a) empirical quantum mechanical and b) molecular mechanics (MM) force fields.

In the case of two states representation, the Hamiltonian of the reacting system can be defined as

$$ H_{EVB} = \begin{bmatrix} H_{ii} & H_{ij} \\ H_{ij} & H_{jj} \end{bmatrix} $$

(S1)

where the diagonal terms represent the potential energy of the diabatic states and the off-diagonal term ($H_{ij}$) denotes the extent of coupling between these states. $H_{ii}$ can be approximated as a function of the reacting bonds which are being broken either in the $i^{th}$ or $j^{th}$ state. The solution of the corresponding secular equation gives the ground state energy surface of the system.

The potential energy function ($H_{ii}$) of a diabatic state $i$ can be expressed as:

$$ H_{ii} = \alpha_{gas}^i + U_{intra}(R,Q) + U_{ss}(R,Q,r,s) + U_{ss}(r,q) $$

(S2)

In eq. S2, $R$ and $r$ denote, the atomic coordinates of a fragment of the reaction center (solute) in the $i^{th}$ diabatic state and the coordinates of the atoms in the rest of the system, respectively. The
partial charges of the solute and rest of the atoms in the system are denoted with $Q$ and $q$ in eq. S2. The first term in eq. S2 is the gas phase energy of the $i^{th}$ diabatic state. This energy is defined as the energy of the reacting system wherein all fragments are assumed to be infinitely separated. The value of this energy is parameterized based on energetics of a reference reaction. If the activation free energy and the reaction free energy of the reference reaction are obtained from quantum mechanics (QM) based calculations or experiments, then our calculated energetics of the reference reaction can be matched with experimental or QM results by parameterizing the gas phase energy and the coupling constant ($H_{ij}$) simultaneously. In the EVB formalism, it is assumed that the $\alpha_{gas}^i$ and $H_{ij}$ values are constant between a reference and a protein reaction.

The term $U_{intra}$ in eq. S2 represents the intramolecular potential of the solute. In this work, the bonds that were formed or broken during a chemical reaction were represented using a Morse potential of the following form:

$$V_b = D_M [1 - e^{-\gamma (b - b_0)}]^2$$  \hspace{1cm} (S3)

The variable $b$ in eq. S3 denotes the bonding distance. All the other bonds in the solute were expressed by harmonic potentials. The other bonding parameters such as bond-angle, proper dihedral and improper dihedral angles were represented with MM force field and the values were taken from the ENZIMIX force field. The non-bonding interactions among the solute atoms are taken in two separate ways whether the pair of atoms: (a) never form bonds in any diabatic states, and (b) form bonds only in one of the diabatic states. While in first case 12-6 Lennard Jones potentials were used, for the second case, an exponential function as in eq. S4 was used.

$$V_{nb} = C e^{-\rho r_{ij}}$$  \hspace{1cm} (S4)

$r_{ij}$ in eq. S4 denotes the distance between two solute atoms. All the parameters, $D_M$, $\gamma$, $b_0$, $C$ and $\rho$ in eq. S3 and S4 are reported in Table S2.1-S2.4.

The last two terms of eq. S2, $U_{Ss}$ and $U_{ss}$ denote the intermolecular interaction potential between the solute ($S$) and the surrounding ($s$) atoms; and between all surrounding atoms (surrounding-surrounding ($s-s$)), respectively.

To calculate the complete reaction profile (see eq. S1), the system should move from one diabatic state to the other. This is achieved by running a free energy perturbation (FEP) based molecular
dynamic simulation on a mapping potential, $\epsilon_m$. The mapping potential is represented as a linear combination of the diabatic potentials of the starting state (state 1), and the final state (state 2) of the reaction. Thus, for a two-states representation of the reaction, the mapping potential takes the form:

$$\epsilon_m = \lambda_m \epsilon_1 + (1 - \lambda_m) \epsilon_2 \text{ where } 0 \leq \lambda_m \leq 1 \quad (S5)$$

The mapping parameter, $\lambda_m$ varies between 0 and 1 while the system is changing from the initial state to the final state.

The associated change in the free energy $\Delta G(\lambda_m)$ can be expressed as:

$$\Delta G(\lambda_m) = \Delta G(\lambda_0 \rightarrow \lambda_m) = \sum_{i=0}^{m-1} \delta G(\lambda_i \rightarrow \lambda_{i+1}) \quad (S6)$$

where

$$\delta G(\lambda_i \rightarrow \lambda_{i+1}) = -\left(\frac{1}{\beta}\right) \ln \left[ \langle e^{(-\epsilon_i - \epsilon_{i+1})/\beta} \rangle_i \right] \quad (S7)$$

In S7, $\beta = 1/k_B T$, $k_B$ is the Boltzmann constant, and $T$ is the temperature, kept constant throughout the simulations. The angular bracket ($\langle \rangle_i$) operator averages the quantity placed within the brackets, with respect to the mapping potential $\epsilon_i$.

Finally, the activation free energy of the reaction, $\Delta g^\neq$ is calculated, using the following free energy perturbation/umbrella sampling (FEP/US) equation:

$$\exp[-\Delta g(X^n)\beta] = \exp[-\Delta G(\lambda_m)\beta] \langle \exp[-\left(E_g(X^n) - \epsilon_m(X^n)\right)\beta] \rangle_m \quad (S8)$$

where $X^m$ is the reaction coordinate, taken in terms of a given energy gap $\epsilon_2 - \epsilon_1$. 


Figure S1. Depiction of all the chemical groups included in region I of different EVB simulations. The Arabic numbers designate all the atoms in region I of the EVB simulations. Note that not all the depicted atoms were included in region I of every EVB simulations (see main text).

S2. EVB parameters

Table S2.1. The Partial Charges of Region I Atoms and All the EVB Parameters Used for the First Proton Transfer (PT1) Step

| Residue | Atom No. | Partial Charges | Atom Type |
|---------|----------|-----------------|-----------|
|         |          | Reactant State  | Product State | Reactant State  | Product State |
| cysteine| 1        | -0.138          | 0.243       | C0              | C0           |
|         | 2 and 3  | 0.155           | -0.009      | H0              | H0           |
|         | 4        | -0.412          | -1.225      | S0              | S-           |
|         | 5        | 0.241           | 0.423       | H0              | H0           |
|         | 6        | -0.445          | -0.382      | C0              | C0           |
|         | 7 and 8  | 0.149           | 0.151       | H0              | H0           |
|         | 9        | 0.264           | 0.383       | C+              | C+           |
| histidine| 10      | -0.439          | -0.273      | N0              | N0           |
|         | 11       | 0.402           | 0.434       | H0              | H0           |
|         | 12       | 0.288           | -0.046      | C+              | C+           |
|    |   |   |    |    |
|----|---|---|----|----|
| 13 | 0.122 | 0.258 | H0 | H0 |
| 14 | -0.672 | -0.161 | N- | N0 |
| 15 | 0.037 | -0.219 | C+ | C+ |
| 16 | 0.145 | 0.281 | H0 | H0 |
| 17 | 0.407 | 0.407 | C+ | C+ |
| 18 | -0.407 | -0.407 | O- | O- |
| 19 | 0.592 | 0.592 | C+ | C+ |
| 20 | -0.592 | -0.592 | O- | O- |

Inhibitor

non-bonding exponential

| Atom Type | C (kcal/mol) | \( \rho(\text{Å}^{-1}) \) |
|-----------|--------------|--------------------------|
| C0        | 91.0         | 2.5                      |
| H0        | 5.0          | 2.5                      |
| S0        | 53.0         | 2.5                      |
| S-        | 90.0         | 2.5                      |
| C+        | 91.0         | 2.5                      |
| N0        | 60.0         | 2.5                      |
| N-        | 60.0         | 2.5                      |
| O-        | 90.0         | 2.5                      |

Morse potential

| Atom pair | D (kcal/mol) | \( \gamma(\text{Å}^{-1}) \) | \( b_0 (\text{Å}) \) |
|-----------|--------------|-----------------------------|----------------------|
| S0-H0     | 94.0         | 1.4                         | 1.345                |
| N0- H0    | 100.0        | 2.0                         | 0.988                |

EVB coupling parameters

| off-diagonal (kcal/mol) | gas phase shift (kcal/mol) |
|-------------------------|-----------------------------|
|                         |                             |
| Residue     | Atom No. | Partial Charges |             | Atom Type       |             |
|-------------|----------|-----------------|-------------|-----------------|-------------|
|             |          | Reactant State  | Product State| Reactant State  | Product State|
| PT1 (no inhibitor) | -       | -               | -           | -               | -           |
| PT1 (with inhibitor) | -       | -               | -           | -               | -           |
| cysteine    | 1        | 0.243           | -0.010      | C0              | C0          |
|             | 2 and 3  | -0.009          | 0.060       | H0              | H0          |
|             | 4        | -1.225          | -0.410      | S-              | S0          |
|             | 5        | 0.423           | 0.423       | H0              | H0          |
| histidine   | 6        | -0.382          | -0.382      | C0              | C0          |
|             | 7 and 8  | 0.151           | 0.151       | H0              | H0          |
|             | 9        | 0.383           | 0.383       | C+              | C+          |
|             | 10       | -0.273          | -0.273      | N0              | N0          |
|             | 11       | 0.434           | 0.434       | H0              | H0          |
|             | 12       | -0.046          | -0.046      | C+              | C+          |
|             | 13       | 0.258           | 0.258       | H0              | H0          |
|             | 14       | -0.161          | -0.161      | N0              | N0          |
|             | 15       | -0.219          | -0.219      | C+              | C+          |
|             | 16       | 0.281           | 0.281       | H0              | H0          |
| Inhibitor   | 17       | 0.407           | 0.11        | C+              | C0          |
|             | 18       | -0.407          | -1.00       | O-              | O-          |
|             | 19       | 0.592           | 0.950       | C+              | C+          |
|             | 20       | -0.592          | -0.760      | O-              | O-          |

Table S2.2. The Partial Charges of Region I Atoms and All the EVB parameters Used for the Nucleophilic Attack (NA) Step
non-bonding exponential

| Atom Type | C (kcal/mol) | ρ (Å⁻¹) |
|-----------|--------------|---------|
| C0        | 91.0         | 2.5     |
| H0        | 5.0          | 2.5     |
| S0        | 53.0         | 2.5     |
| S⁻        | 90.0         | 2.5     |
| C⁺        | 91.0         | 2.5     |
| N0        | 60.0         | 2.5     |
| N⁻        | 60.0         | 2.5     |
| O⁻        | 90.0         | 2.5     |

Morse potential

| Atom pair | D_M (kcal/mol) | γ (Å⁻¹) | b0 (Å) |
|-----------|---------------|---------|--------|
| S0-C0     | 90.0          | 1.4     | 1.86   |

EVB coupling parameters

\[ H_{ij} = \begin{cases} \text{off-diagonal} & \text{gas phase shift (kcal/mol)} \\ Ae^{-\mu(r-r_0)} \end{cases} \]

|                      | A= 78.5       | -118.0 |
|----------------------|---------------|--------|
| NA (HIS41 not in region I) | µ=0.1; r₀=1.86 Å | |
|                      | A= 85.5       | -157.0 |
| NA (HIS41 in region I)  | µ=0.1; r₀=1.86 Å | |

Table S2.3. The Partial Charges of Region I Atoms and All the EVB Parameters Used for the Concerted (PT1 and NA) Reaction
| Atom Type | C (kcal/mol) | \( \rho (\text{Å}^{-1}) \) |
|-----------|--------------|-----------------|
| C0        | 91.0         | 2.5             |
| H0        | 5.0          | 2.5             |
| S0        | 53.0         | 2.5             |
| C+        | 91.0         | 2.5             |
| N0        | 60.0         | 2.5             |

non-bonding exponential

Inhibitor
### Morse potential

| Atom pair | \( D_M \) (kcal/mol) | \( \gamma \) (\( \text{Å}^{-1} \)) | \( b_0 \) (\( \text{Å} \)) |
|-----------|-----------------------|-------------------------------|-----------------|
| S0-H0     | 94.0                  | 1.4                           | 1.345           |
| N0- H0    | 100.0                 | 2.0                           | 0.988           |
| S0-C0     | 90.0                  | 1.4                           | 1.86            |

### EVB coupling parameters

\[
H_{ij} = A e^{-\mu(r-r_0)}
\]

- gas phase shift (kcal/mol)

\[
A = 82.5
\]

\[
\mu = 0.1; \quad r_0 = 1.86 \text{ Å}
\]

---

**Table S2.4. The Partial Charges of Region I Atoms and All the EVB Parameters Used for the Second Proton Transfer (PT2) Step**

| Residue | Atom No. | Partial Charges | Atom Type |
|---------|----------|-----------------|-----------|
|         |          | Reactant State  | Product State | Reactant State | Product State |
| cysteine| 1        | -0.010          | -0.052     | C0            | C0            |
|         | 2 3      | 0.060           | 0.084      | H0            | H0            |
|         | 4        | -0.410          | -0.218     | S0            | S0            |
|         | 5        | 0.423           | 0.392      | H0            | H0            |
|         | 6        | -0.382          | -0.445     | C0            | C0            |
|         | 7 8      | 0.151           | 0.149      | H0            | H0            |
| histidine| 9        | 0.383           | 0.264      | C+            | C+            |
|         | 10       | -0.273          | -0.439     | N0            | N0            |
|         | 11       | 0.434           | 0.402      | H0            | H0            |
|         | 12       | -0.046          | 0.288      | C+            | C+            |
| Inhibitor | C (kcal/mol) | \( \rho \) (Å) |
|----------|--------------|-----------------|
| C0       | 91.0         | 2.5             |
| H0       | 5.0          | 2.5             |
| S0       | 53.0         | 2.5             |
| C+       | 91.0         | 2.5             |
| N0       | 60.0         | 2.5             |
| N-       | 60.0         | 2.5             |
| O-       | 90.0         | 2.5             |
| O0       | 53.0         | 2.5             |

**Morse potential**

| Atom pair | \( D_M \) (kcal/mol) | \( \gamma \) (Å\(^{-1}\)) | \( b_0 \) (Å) |
|-----------|-----------------------|-----------------|--------------|
| N0- H0    | 100.0                 | 2.0             | 0.988        |
| O0-H0     | 102.0                 | 2.0             | 0.980        |

**EVB coupling parameters**

| off-diagonal (kcal/mol) | gas phase shift (kcal/mol) |
|-------------------------|----------------------------|
| -                       | -16.0                      |
S3. Semi-microscopic Version of the Protein Dipole Langevin Dipole (PDLD) Method for Noncovalent Binding Free Energy Calculations

We have used the semi-microscopic version of the Protein Dipole Langevin Dipole method in the linear response approximation, with a scaled non electrostatic term (PDLD/S-LRA/β) to calculate the non-covalent binding free energies of the Mpro-13b complexes. The thermodynamic cycle shown in Figure 4 of ref. 5 is used to calculate the binding free energies. The effective PDLD/S potential for a single protein-inhibitor configuration in bound (B) and unbound (UB) states are calculated using the following equations:

\[ U_{elec, l}^P = \left( \Delta G_{sol}^{l+p} - \Delta G_{sol}^{l+p} \right) \left( \frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w} \right) + \Delta G_{sol}^l \left( 1 - \frac{1}{\varepsilon_p} \right) + \frac{U_{q\mu}^l}{\varepsilon_p} + \frac{U_{intra}^l}{\varepsilon_p} \]  

\[ U_{elec, w}^w = \left( \Delta G_{sol}^l \left( \frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w} \right) + \Delta G_{sol}^l \left( 1 - \frac{1}{\varepsilon_p} \right) + \frac{U_{intra}^l}{\varepsilon_p} \right)_{UB} \]  

where, \( \Delta G_{sol}^l \) denotes the solvation free energy of any group of atoms of the protein(p) (or inhibitor (l)) “l” in water. \( \varepsilon_w \) and \( \varepsilon_p \) denote the dielectric constant of water and protein respectively. The significance of \( \varepsilon_p \) can be found in ref 6. In this work \( \varepsilon_p=4 \) has been used. \( U_{q\mu}^l \) is the electrostatic interaction between the inhibitor charges (q) and the protein dipoles (\( \mu \)) in vacuum (a standard PDLD notation). \( U_{intra}^l \) is the intramolecular electrostatic interaction for l. The uncharged state (nonpolar) of the inhibitor is denoted with \( l' \).

The electrostatic interaction energy \( U_{elec, l} \) in eqs. S9 and S10 are obtained from a single configuration of protein-inhibitor complex. As a result, that energy does not properly represent the protein reorganization. Thus, the linear response approximation (LRA) is used to capture the protein reorganization. In this approximation, we calculate the average of the effective potential over the trajectories of the protein-inhibitor complex in their polar form (l) as well as nonpolar form (\( l' \)). Thus, in PDLD/S-LRA/β the electrostatic part of the binding free energy (\( \Delta G_{bind}^{elec} \)) can be expressed as,

\[ \Delta G_{bind}^{elec} = \frac{1}{2} \left[ \langle U_{elec, l}^P \rangle_l + \langle U_{elec, l}^P \rangle_{l'} - \langle U_{elec, l}^w \rangle_l - \langle U_{elec, l}^w \rangle_{l'} \right] \]
where the term $\langle U_{\text{elec}, l}^P \rangle_t$ designates an average of the effective potential $U_{\text{elec}, l}^P$ over a protein configuration generated with respect to the protein force field, which includes zero partial charges of the inhibitor.

Additionally, the non-electrostatic energy part of the binding free energy is calculated by scaling the van der waals (vdw) interaction energy of the polar form of the inhibitor with $\beta=0.25$. Thus, the following equation is used to calculate the binding free energy using the PDLD/S-LRA/$\beta$ method:

$$
\Delta G_{\text{bind}}^{\text{PDLD/S-LRA}/\beta} = \Delta G_{\text{bind}}^{\text{elec}} + \beta \left[ \langle U_{\text{vdw}, l}^P \rangle_t - \langle U_{\text{vdw}, l}^w \rangle_t \right]
$$

(S12)

References

1. Warshel, A.; Weiss, R. M., An Empirical Valence Bond Approach for Comparing Reactions in Solutions and in Enzymes. *J Am Chem Soc* 1980, 102 (20), 6218-6226.
2. Kamerlin, S. C. L.; Warshel, A., The empirical valence bond model: theory and applications. *Wires Comput Mol Sci* 2011, 1 (1), 30-45.
3. Lee, F. S.; Chu, Z. T.; Warshel, A., Microscopic and Semimicroscopic Calculations of Electrostatic Energies in Proteins by the Polaris and Enzymix Programs. *Journal of Computational Chemistry* 1993, 14 (2), 161-185.
4. Singh, N.; Warshel, A., Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. *Proteins* 2010, 78 (7), 1705-1723.
5. Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A., Examining methods for calculations of binding free energies: LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease. *Proteins-Structure Function and Genetics* 2000, 39 (4), 393-407.
6. Schutz, C. N.; Warshel, A., What are the dielectric "constants" of proteins and how to validate electrostatic models? *Abstr Pap Am Chem S* 2002, 223, C75-C75.
7. Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A., Calculations of Antibody Antigen Interactions - Microscopic and Semimicroscopic Evaluation of the Free-Energies of Binding of Phosphorylcholine Analogs to Mcpc603. *Protein Eng* 1992, 5 (3), 215-228.