## Supporting Information

# A Novel Strategy to Study Electrostatic effects in Chemical Reactions. Differences Between the Role of Solvent and the Active Site of Chalcone Isomerase in a Michael Addition. 

J. Javier Ruiz-Pernía ${ }^{\S}$, Sergio Martı̂ ${ }^{\S}$, Vicent Moliner**, Iñaki Tuñón * ${ }^{*}$<br>Departament de Química Física, Universitat de València, 46100 Burjassot, (Spain);<br>Departament de Química Física i Analítica; Universitat Jaume I, 12071 Castellón (Spain)<br>$\ddagger$ Universitat de València<br>§ Universitat Jaume I<br>*to whom correspondence should be addressed

ignacio.tunon@uv.es
moliner@uji.es

## Computational Details

Hybrid quantum mechanics/molecular mechanics ( $\mathrm{QM} / \mathrm{MM}$ ) potentials, within the electrostatic embedding approach, are employed in this work to study the transformation of 6'-deoxychalcone (4,2', 4'-trihydroxychalcone) into (2S)-5deoxyflavanone ( $7,4^{\prime}$ '-dihydroxyflavanone), as depicted in Scheme I of the paper. The substrate is partially deprotonated at the 2' position in physiological conditions in aqueous solution, a requisite to produce an intramolecular Michael addition to the $\alpha, \beta$ double bound. ${ }^{1,2}$ Briefly, in our simulation models, the reactant molecule ( 30 atoms) constitutes the QM subsystem, which is described by means the AM1 Hamiltonian. ${ }^{3}$ The rest of the system, protein or water molecules, is described by MM force fields, OPLS-AA ${ }^{4,5}$ and TIP3P, ${ }^{6}$ respectively. In order to study the reaction in aqueous solvent the substrate was placed in a cubic box (of side $55.5 \AA$ ) of TIP3P water molecules centered on the $\mathrm{O}^{\prime}$, atom of the QM subsystem. The enzymatic reaction was studied placing the substrate in the active site of a Chalcone Isomerase whose initial coordinates were taken from the X-ray crystal structure 1EYQ. ${ }^{7}$ The whole system (enzyme plus substrate) was placed inside a cubic box ( $79.5 \AA$ of side) of water molecules centred on the QM subsystem. During the QM/MM simulations we employed NVT ensemble with a reference temperature of 300 K and a switched cut-off radius of $13.5 \AA$ for all kind of interactions. Periodic Boundary Conditions were employed for the simulation of the aqueous system, while for the enzymatic reaction after equilibration all atoms beyond $24 \AA$ of O2' atom of the substrate were kept frozen.

Transition State Theory offers a very reasonable framework to study the catalytic activity of systems like enzymes through the analysis of the effects on the activation free energy in comparison with the counterpart process in aqueous solution. ${ }^{8}$ In general terms, the activation free energy is estimated by means the Potential of Mean Force (PMF or $\mathrm{W}(\xi)$ ) obtained as a function of a distinguished reaction coordinate $\xi$. $\mathrm{W}(\xi)$ is related to the normalized probability of finding the system at a particular value of the chosen coordinate, as shown in equation S :
$W(\xi)=C-k T \ln \int \rho\left(x^{N}\right) \delta\left(\xi\left(x^{N}\right)-\xi\right) d x^{N}$
where $\rho\left(\mathrm{x}^{\mathrm{N}}\right)$ is the probability density of finding the system at a particular configuration $\left(\mathrm{x}^{\mathrm{N}}\right)$, the Dirac delta functions allow this to be evaluated only at the desired values of the selected coordinates and $C^{\prime}$ is an integration constant. The activation free energy of the reaction can be then obtained as: ${ }^{9}$

$$
\begin{equation*}
\Delta G^{\ddagger}(\xi)=W\left(\xi^{\ddagger}\right)-\left[W\left(\xi^{R}\right)+G_{\xi}^{R}\right] \tag{S2}
\end{equation*}
$$

where the $\ddagger$ and R superscripts indicate the value of the reaction coordinate at the TS and reactants, respectively, and $G_{\xi}^{R}$ is the free energy cost associated to setting the reaction coordinate to a particular value in the reactants state.

In this work we trace two-dimensional PMFs or Free Energy Surface (FES) using, two different coordinates: a solute coordinate (r) and a solvent coordinate (s). This FES can be expressed as:
$W(r, s)=C^{\prime}-k T \ln \int \rho\left(x^{N}\right) \delta\left(r\left(x^{N}\right)-r\right) \delta\left(s\left(x^{N}\right)-s\right) d x^{N}$

A one-dimensional PMF may be recovered tracing the minimum free energy path on $\mathrm{W}(\mathrm{r}, \mathrm{s})$ and integrating over the direction perpendicular to the path, while the activation free energy can be estimated integrating also along the direction of the reaction path in the reactants state. In our case we use as solute coordinate the distance between the nucleophile and the $\mathrm{C}^{\beta}$ atom:
$r=d\left(O 2^{\prime}-C^{\beta}\right)$
As solvent coordinate we use the electrostatic potential created by the MM environment on the nucleophile:
$s=\sum_{j=1}^{j=M} \frac{q_{M M}^{j}}{\left|r_{j}-r_{o 2^{\prime}}\right|}$

The FESs corresponding to the reactions in aqueous solution and the enzyme were obtained using the weighted histogram analysis method (WHAM) combined with the umbrella sampling approach. ${ }^{10,11}$ Harmonic restraints were applied to reference values of the $r$ and $s$ coordinates:
$V_{r}=\frac{1}{2} K_{r}\left(r-r_{o}\right)^{2} ; \quad V_{s}=\frac{1}{2} K_{s}\left(s-s_{o}\right)^{2}$
Molecular Dynamics then will preferentially explore the most probable configurations of the system around the reference values $r_{0}$ and $s_{0}$. Simulations were performed at 61 different reference values of $r$ in a range from 3.2 to $1.4 \AA$ with an interval of $0.03 \AA$, using an umbrella with a force constant of $2500 \mathrm{~kJ} \cdot \mathrm{~mol}^{-1} \cdot \AA^{-2}$. For each of these values, 33 different simulations were carried out changing the value of $s$ in a range from 50 to $170 \mathrm{~kJ} \cdot \mathrm{~mol}^{-1} \cdot \mathrm{e}^{-1}$ using an umbrella with a force constant of $0.2 \mathrm{~kJ}^{-1} \cdot \mathrm{~mol} \cdot \mathrm{e}^{2}$. The behaviour of the electrostatic potential with the value of the force constant is illustrated in Figure S4. The selected value offers a good compromise between the number of simulation windows needed and the simulation time per window. Each simulation window consisted in 10 ps of production with a time step of 1 fs . As shown in Figure S5 these conditions lead to an efficient overlap of the values of the electrostatic potential corresponding to a selected value of the solute coordinate. For all the calculations, we employed a modified version of fDYNAMO. ${ }^{12}$

Finally, because the AM1 semiempirical Hamiltonian provides a systematically too high potential energy barrier, single-point corrections at a higher level of theory (MP2/6$31+\mathrm{G}^{* *}$ ) were performed using the coordinates of a reactant and a transition structure located in the enzymatic active site. The structures and the correction to the potential energy barrier are provided in Tables S1 and S2.

## References

(1) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S. J. Phys. Chem. B 2006, 110, 17663-17670.
(2) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I. J. Am. Chem. Soc. 2007, 129, 91179124.
(3) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. J. Am. Chem. Soc. 1985, 107, 3902-3909.
(4) Jorgensen, W. L.; Tiradorives, J. J. Am. Chem. Soc. 1988, 110, 16571666.
(5) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. J. Phys. Chem. B 2001, 105, 6474-6487.
(6) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J. Chem. Phys. 1983, 79, 926-935.
(7) Jez, J. M.; Bowman, M. E.; Dixon, R. A.; Noel, J. P. Nat. Struct. Biol. 2000, 7, 786-791.
(8) Marti, S.; Roca, M.; Andres, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. Chem. Soc. Rev. 2004, 33, 98-107.
(9) Schenter, G. K.; Garrett, B. C.; Truhlar, D. G. J. Chem. Phys. 2003, 119, 5828-5833.
(10) Torrie, G. M.; Valleau, J. P. J. Comput. Phys. 1977, 23, 187-199.
(11) Roux, B. Comput. Phys. Commun. 1995, 91, 275-282.
(12) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. J. Comput. Chem. 2000, 21, 1088-1100.


Figure S1. Radial Distribution Functions (RDF) of water atoms around O2' for the reactants state in aqueous solution. Up: RDF obtained at high values of the electrostatic potential; low: RDF obtained at low values of the electrostatic potential. Blue line represents the RDF of the oxygen atom of water molecules around the O2' atom while the red line represents the RDF of the hydrogen atoms of water molecules around the $\mathrm{O} 2^{\prime}$ atom.


$$
\begin{aligned}
& \mathrm{r}=3.2 \AA \\
& \mathrm{~s}=70 \mathrm{kcal} \cdot \mathrm{~mol}^{-1} \cdot \mathrm{e}^{-1}
\end{aligned}
$$



$$
\begin{aligned}
& r=3.2 \AA \\
& \mathrm{~s}=150 \mathrm{kcal} \cdot \mathrm{~mol}^{-1} \cdot \mathrm{e}^{-1}
\end{aligned}
$$

Figure S2. Representative snapshots of chalcone reactants in aqueous solution corresponding to low (left) and high (right) values of the electrostatic potential on the nucleophile.


Figure S3. PMFs for the Michael addition of 6'-deoxychalcone in aqueous solution (blue) and in the active site of CHI (red) obtained as a function of the solute coordinate $\mathrm{r}=\mathrm{O} 2^{\prime}-\mathrm{C}^{\beta}$.


Figure S4. Fluctuations of the electrostatic potential obtained for different values of the force constant $\left(0.2 \mathrm{~kJ}^{-1} \cdot \mathrm{~mol} \cdot \mathrm{e}^{2}\right.$ red line, and of $0.02 \mathrm{~kJ}^{-1} \cdot \mathrm{~mol} \cdot \mathrm{e}^{2}$ blue line) for a simulations centred at $\mathrm{s}=85 \mathrm{kcal} \cdot \mathrm{mol}^{-1} \cdot \mathrm{e}^{-1}$ and $r=1.85 \AA$.


Figure S5. Distribution of electrostatic potential values obtained for the simulation windows corresponding to the enzyme at $\mathrm{r}=1.85 \AA$ and using a force constant of $0.2 \mathrm{~kJ}{ }^{-1} \cdot \mathrm{~mol} \cdot \mathrm{e}^{2}$ for the solvent coordinate.

Table S1. Cartesian coordinates (in $\AA$ ) of a transition and a reactant structure corresponding to the Michael Addition of 6'-deoxychalcone in the active site of CHI.

## Transition Structure

| 8 | -.3933912147 | 2.9627023602 | -6.8479153530 |
| :--- | ---: | ---: | ---: |
| 6 | -7.8945669452 | 4.0085241131 | -6.6143724444 |
| 6 | -8.2152203944 | 4.4354135591 | -7.9856206477 |
| 6 | -8.2202136614 | 3.4790794169 | -9.0150093701 |
| 6 | -8.5529692910 | 3.8217215816 | -10.3193265577 |
| 6 | -8.9060782704 | 5.1529005751 | -10.6010518249 |
| 6 | -8.9222499514 | 6.1183179550 | -9.5838784468 |
| 6 | -8.5741591644 | 5.7549771538 | -8.2859996262 |
| 8 | -9.2250741249 | 5.4552826875 | -11.9045728802 |
| 1 | -9.5876860928 | 6.3631763833 | -11.9274640890 |
| 1 | -8.5506828811 | 3.0682321006 | -11.1211827575 |
| 1 | -7.9679520812 | 2.4381037704 | -8.7603779581 |
| 1 | -9.2142991208 | 7.1546688209 | -9.8067237871 |
| 1 | -8.6057043613 | 6.5124120065 | -7.4871918120 |
| 1 | -8.4438067474 | 3.0907384960 | -6.2965292299 |
| 6 | -7.5161970050 | 4.9278315419 | -5.6323094558 |
| 1 | -7.4565716458 | 6.0011969550 | -5.8319875385 |
| 6 | -6.9730163016 | 4.4424359594 | -4.4076318822 |
| 8 | -6.9070489141 | 5.1512219403 | -3.3578320217 |
| 6 | -6.3954291776 | 3.0909920671 | -4.4307448302 |
| 1 | -6.4496852279 | 2.8839092210 | -2.2872740966 |
| 6 | -6.1081174882 | 2.4297599758 | -3.2298155023 |
| 6 | -5.4276068344 | 1.2277986412 | -3.1959496093 |
| 1 | -5.2432089547 | 0.6894689299 | -2.2583750273 |
| 6 | -4.9792257240 | 0.6755330160 | -4.4160697417 |
| 8 | -4.2964953723 | -0.5009699823 | -4.3165109703 |
| 1 | -3.9612510465 | -0.7586516687 | -5.1998532972 |
| 1 | -4.9088558030 | 0.8431280552 | -6.5746048189 |
| 6 | -5.2488019700 | 1.2967627203 | -5.6356839233 |
| 6 | -6.0318476246 | 2.4779114690 | -5.6670043737 |

## Reactants structure

| 8 | -5.3394197872 | 3.0864493798 | -7.0914157732 |
| :--- | ---: | ---: | ---: |
| 6 | -8.0067662122 | 3.7911052496 | -7.6653349987 |
| 6 | -8.5159952701 | 4.3247384024 | -8.9202399131 |
| 6 | -8.3267651564 | 3.6131890148 | -10.1165466177 |
| 6 | -8.8186466337 | 4.1020191529 | -11.3213506766 |
| 6 | -9.5332762024 | 5.3109375038 | -11.3268433575 |
| 6 | -9.7613983336 | 6.0176887464 | -10.1380769442 |
| 6 | -9.2475498145 | 5.5221993408 | -8.9454130159 |
| 8 | -10.0169309066 | 5.7448370739 | -12.5407419802 |
| 1 | -10.2998189799 | 6.6735462368 | -12.4365290919 |
| 1 | -8.6641335021 | 3.5459771237 | -12.2583691336 |
| 1 | -7.7845042687 | 2.6544424364 | -10.1019815983 |
| 1 | -10.3327420968 | 6.9574585619 | -10.1474101691 |
| 1 | -9.4372661929 | 6.0690364348 | -8.0080682798 |
| 1 | -7.8507316803 | 2.6964442775 | -7.6287032335 |
| 6 | -7.7461360831 | 4.5583969862 | -6.5982021487 |
| 1 | -7.8586138659 | 5.6563828871 | -6.6398809328 |
| 6 | -7.3805452158 | 4.0577727660 | -5.2608285239 |
| 8 | -7.8369054109 | 4.7132507845 | -4.2801188960 |
| 6 | -6.5650561392 | 2.8867606957 | -5.0574286567 |
| 1 | -7.5135685284 | 2.4809498187 | -3.1607325865 |
| 6 | -6.7337395825 | 2.1492034245 | -3.8660693917 |
| 6 | -5.9976122125 | 1.0301110759 | -3.5688671576 |
| 1 | -6.1761047897 | 0.4349923956 | -2.6636773035 |
| 6 | -5.0248008236 | 0.5920753901 | -4.5145520688 |
| 8 | -4.3366231286 | -0.5357531842 | -4.1729728355 |
| 1 | -3.7676681575 | -0.7962000801 | -4.9257785058 |
| 1 | -4.0396891713 | 0.9653862346 | -6.4032077291 |
| 6 | -4.7954689153 | 1.2929234546 | -5.6786583493 |
| 6 | -5.5734219245 | 2.4577886742 | -6.0041815070 |

Table S2. Gas phase energy barriers (in $\mathrm{kcal} \cdot \mathrm{mol}^{-1}$ ) obtained from single point calculation on the structures given in Table S1.

|  | AM1 | AM1//MP2/6-31+(d,p) |
| :---: | :---: | :---: |
| $\Delta \mathrm{E}^{\ddagger}$ | 22.6 | 15.0 |

