Supporting Information

Roca *et al.* 10.1073/pnas.0803405105

SI Text

The Simplified Landscape. Our current simplified model has not produced the best catalytic configurations, because the simplified enzyme substrate model has not been refined sufficiently in terms of protein–substrate interactions. This, however, does not pose a fundamental problem in our approach because the simplified model is used only as reference potential for calculations of the explicit landscape. Thus any variation of the simplified potential is allowed, with the proviso that the free energy for transfer from the simplified to the explicit model converges better when the two surfaces are similar. Thus we added the U′ term of Eq. **3** to the simplified potential. In principle, we could have evaluated the free-energy landscape from the simplified model in terms of Rg, CO, and U′, and then calculated the free energy of moving from the simplified model to the explicit model. However, at this stage we use U′ mainly to explore different ranges in the overall landscape. Fig. S1 represents the landscape in the presence of U′, whereas Fig. 3 represents the landscape without the effect of U′.

The Arrangements of the Chemical Profiles. We sorted the chemical profiles according to the rmsd distance between the ground-state structure used to start a given EVB simulation and the native structure. The results presented in Fig. 6 are based on arbitrary equal spacing between the different configurations, and the actual rmsds are given in Table S1. Subsequent studies will be needed to determine the distance between the different configurations and more importantly the activation barrier for transition between these configurations.

The Relationship Between the Chemical Fluctuations and the Folding Coordinate. In view of the interest in the relationship between the modes involved in unfolding and the modes responsible for the chemical step (see discussion in ref. 1). We evaluated the vectors along the chemical coordinate (calculated as the difference between product state and reactant state) and the vectors in the direction of the conformational coordinate (the difference between a partially unfolded structure and the folded structure). This was done starting from a partially unfolded structure Fig. S3*A* and from a structure close to the folded conformation Fig. S3*B*. As seen in Fig. S3 the two motions are almost orthogonal when we consider the large amplitude conformational motion and still significantly orthogonal when we consider the conformational change in the folded structure. Of course, the key issue is whether the fluctuations from conformational motion can be transferred to chemical motions (see ref. 1) in a non-Boltzmann way. Although we consider this to be quite unlikely, the effect of motions along the conformational axis of Fig. 6 on the overall rate constant can be explored with different assumed diffusion constants up to the inertial limit.

Reorganization Energies. This work considered the relationship between the reorganization energy and activation energies in different regions of the landscape. The meaning of the reorganization free energy is shown in Fig. S2, where we see that when the reactant and product free-energy surfaces follow a parabolic relationship (as they do in most cases), the activation barrier is correlated with the reorganization energy. In such cases the catalytic effect is fully correlated with the solvent (or protein) contribution to the reorganization energy. However, in the case of CM we have a large compensation of the charge–charge interaction in the substrate by the solvation effect from the protein. Thus, we have to consider the total solute + solvent reorganization energy, which is larger than the solvent reorganization energy. The error range of the calculated reorganization energy is ≈5 kcal/mol when averaged over several sets of initial conditions. However, in the case of Fig. 5 we have probably a larger error because we do not take the average reorganization energy but rather sort the reorganization energies according to the corresponding activation barriers.

The Effect of the Binding Free Energies. An important issue that requires more extensive discussion is the issue of the dependence of the binding free energy on the regions in the folding landscape. Although this article focuses on the $\Delta g^{\ddagger}$ of Fig. 6, it is important to consider, at least qualitatively, the effective rate constant, which is related to k_{cat}/K_M. The activation barrier of this rate constant is the sum of the binding free energy and the $\Delta g^{\ddagger}$ of the chemical step (2), so that

$$\Delta g^{\ddagger}_{p} = \Delta G_{\text{bind}} + \Delta g^{\ddagger}_{\text{cat}}. \qquad [1]$$

where $\Delta g^{\ddagger}_{cat}$ is the above chemical barrier $\Delta g^{\ddagger}$ and $\Delta g^{\ddagger}_{p}$ is the total barrier. The evaluation of $\Delta g^{\ddagger}_{p}$ is challenging because ΔG_{bind} depends on the specific region in the landscape and this dependence should be estimated.

In principle we can use our perturbation treatment where,

$$\Delta\Delta G_{\text{bind}}(R_I \rightarrow R_{II}) = \Delta\Delta G_{simp} + \Delta G_{simp\rightarrow exp}(R_{II}^{simp} \rightarrow R_{II}^{exp}) - \Delta G_{simp\rightarrow exp}(R_{II}^{simp} \rightarrow R_{I}^{exp}) \qquad [2]$$

where $\Delta\Delta G$ refers to the enzyme + substrate state, and R_{II} and R_I are points in region II and region I, respectively. Here, $\Delta G_{simp\rightarrow exp}$ is the free energy of moving from the simplified potential (that uses constraint in the case of region II) to the explicit potential that does not use constraint. The free-energy perturbation treatment that provides $\Delta G_{simp\rightarrow exp}$ is described in ref. 3. In the present case we can perform the perturbation treatment directly from R_{II} in the simplified model to both R_I and R_{II} in the explicit native region.

Now the trend in $\Delta g^{\ddagger}_{p}$ can be obtained by:

$$(\Delta g^{\ddagger}_{p})_{II} = (\Delta g^{\ddagger}_{\text{cat}})_{II} + (\Delta G_{bind})_{I} + \Delta\Delta G_{bind}(R_I \rightarrow R_{II}) \qquad [3]$$

In this way we can compare the $\Delta g^{\ddagger}_{p}$ of regions I and II.

Here, we describe the perturbation formulation mainly to clarify the direction of our future strategy (which is now in a validation stage) and present an alternative approach that has been validated by us in previous studies. That is, the binding free energy needed for the evaluation of $\Delta g^{\ddagger}_{p}$ can be estimated by the same linear response approximation (LRA) used in our studies of protein conformational changes in F_1-ATPase (4), Pol β (5), and bacteriorhodopsin (6). More specifically, we can use

$$\Delta\Delta G_{bind}(R_I \rightarrow R_{II}) = \Delta G(U_a(R_I) \rightarrow U_a(R_{II})) = \Delta G(U_a(R_I) \rightarrow U_b(R_{II})) + \Delta G(U_b((R_{II}) \rightarrow U_a(R_{II})) \qquad [4]$$

$$\Delta G(U_a(R_I) \rightarrow U_b(R_{II})) \cong \frac{1}{2}[\langle U_b - U_a\rangle_{a,R_I} + \langle U_b - U_a\rangle_{b,R_{II}}] \qquad [5]$$

where $\boldsymbol{U_a}$ and U_b are the protein-substrate explicit interaction with a charged and uncharged ligand, respectively, and where

$\langle\ \rangle_{a,R_I}$ designates trajectories on state a with a weak constraint that holds the system near the protein configuration R_I (see refs. 4 and 6). The nature of the analysis of Eq. **4** is illustrated in Fig. S4. Now because obtaining stable microscopic result for Eq. **5** requires very long simulations, we approximated the above treatment by obtaining the average $\langle U_b - U_a \rangle$ from the PDLD/S-LRA treatment (see ref. 7 for a description of this treatment) where the compensating effects that are not obtained in short simulations are represented by a dielectric effect. The results of our simulations are summarized in Table S2. As seen in Table S2, the average $\Delta\Delta G_{\text{bind}}(R_{\text{I}} \rightarrow R_{\text{II}})$ is small and is slightly negative for the monomer and slightly positive for the dimer.

Although the above analysis is quite qualitative (because the primary focus of our calculations is the dependence of $\Delta g^{\ddagger}_{cat}$ on the landscape), our calculations indicate that $\Delta\Delta G_{\text{bind}}(R_{\text{I}} \rightarrow R_{\text{II}})$ does not change the trend obtained in Fig. 6 and that $\Delta G_{\text{bind}} + \Delta g^{\ddagger}$ follows the trend of $\Delta g^{\ddagger}_{cat}$. A potentially interesting finding is the observation (from Table S2) that $\Delta\Delta G_{\text{bind}}(R_{\text{I}} \rightarrow R_{\text{II}})$ might be slightly negative in the monomer despite the fact that the binding energy in region I is larger (more negative value) than in region II. If this result is confirmed in more extensive calculations, then it may mean that the entropic effects of having a more disordered nonnative structure can help in increasing $k_{\text{cat}}/K_{\text{M}}$ in nonnative regions. However, this effect is not expected to be large.

1. Roca M, Liu H, Messer B, Warshel A (2007) On the relationship between thermal stability and catalytic power of enzymes. *Biochemistry* 46:15076–15088.
2. Warshel A, *et al.* (2006) Electrostatic basis for enzyme catalysis. *Chem Rev* 106:3210–3235.
3. Fan ZZ, Hwang JK, Warshel A (1999) Using simplified protein representation as a reference potential for all-atom calculations of folding free energy. *Theor Chem Acc* 103:77–80.
4. Strajbl M, Shurki A, Warshel A (2003) Converting conformational changes to electrostatic energy in molecular motors: The energetics of ATP synthase. *Proc Natl Acad Sci USA* 100:14834–14839.
5. Xiang Y, Goodman MF, Beard WA, Wilson SH, Warshel A (2008) Exploring the role of large conformational changes in the fidelity of DNA polymerase β. *Proteins* 70:231–247.
6. Braun-Sand S, Sharma PK, Chu ZT, Pisliakov AV, Warshel A (2008) The energetics of the primary proton transfer in bacteriorhodopsin revisited: It is a sequential light-induced charge separation after all. *Biochim Biophys Acta Bioenergetics* 1777:441–452.
7. Sham YY, Chu ZT, Tao H, Warshel A (2000) 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* 39:393–407.
8. Warshel A (1991) *Computer Modeling of Chemical Reactions in Enzymes and Solutions* (Wiley, New York).

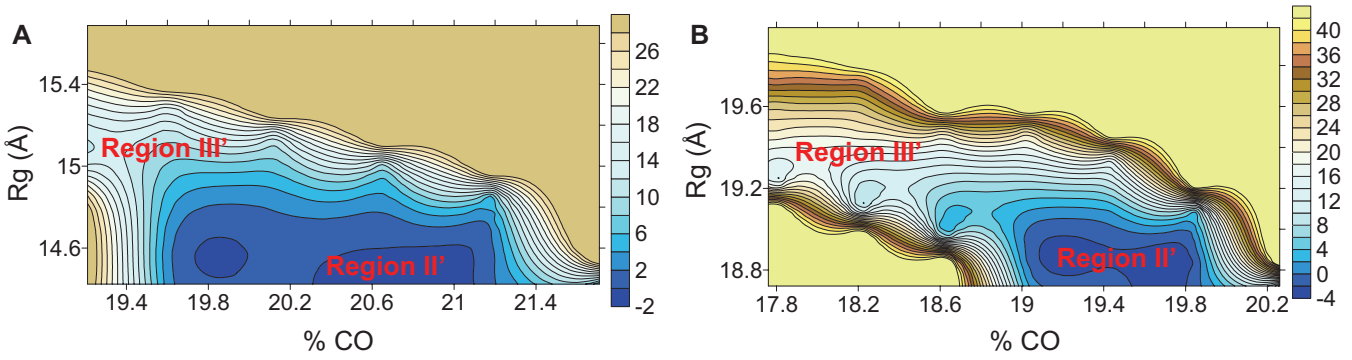

Fig. S1. Free-energy landscapes for the monomer (*A*) and dimer (*B*) structures with TSA applying an extra simplified potential (U′) that reflects the constraint on the distances between the catalytic residues and the substrate. The force constant applied was $K' = 5$ kcal/mol·Å^2. The free-energy surface is represented in terms of the radius of gyration (Rg) and the percent contact order (%CO). Region II′ designates the minimum region (that was used to generate region II in the explicit model), and region III′ is the region far from the minimum (that was used to generate region III in the explicit model). From both regions several structures were taken to calculate the activation free energies. Energies are expressed in kcal/mol and distances are in Å.

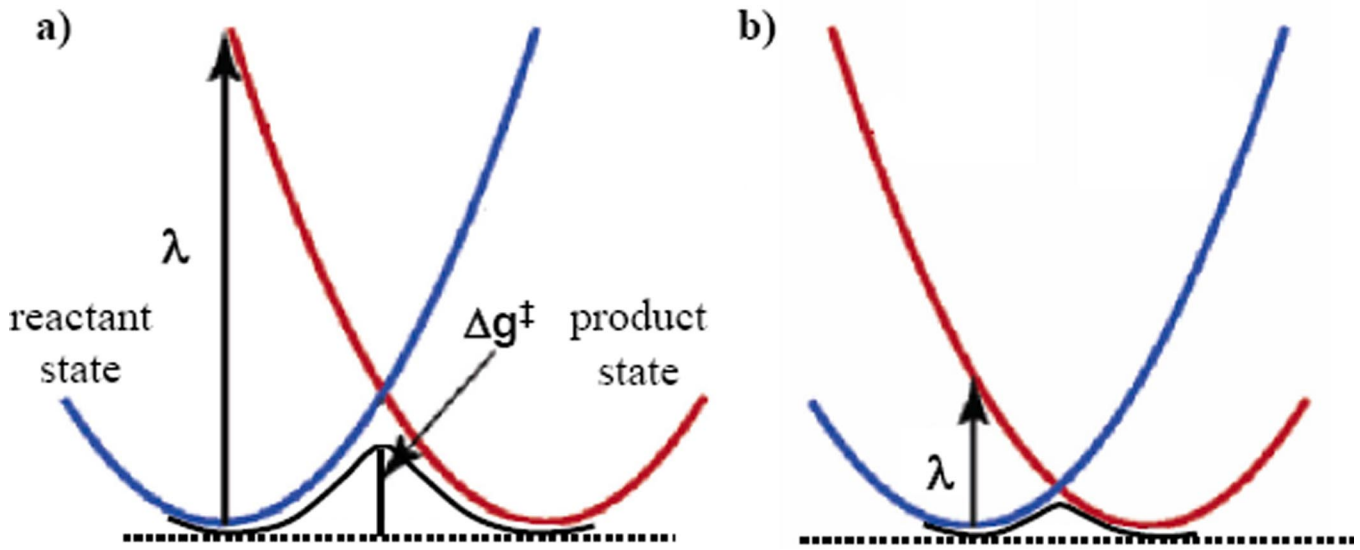

Fig. S2. Showing the correlation between the reorganization energy (λ) and the activation free energy ($\Delta g^{\ddagger}$). In cases where λ is large (e.g. *a*) the activation energy is large, whereas in cases where λ is small (e.g. *b*) the activation energy is small. However, the mixing term (H_{12}) between the states should also be considered (for more details see refs. 1 and 2).

1. Warshel A, *et al.* (2006) Electrostatic basis for enzyme catalysis. *Chem Rev* 106:3210–3235.
2. Warshel A (1991) *Computer Modeling of Chemical Reactions in Enzymes and Solutions* (Wiley, New York).

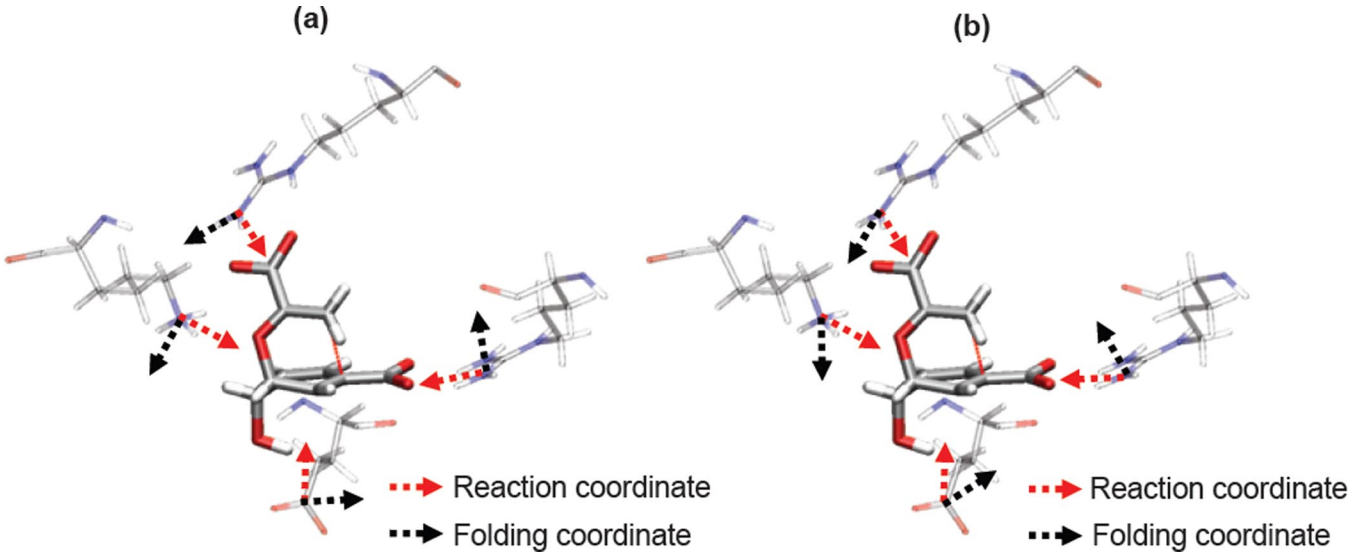

Fig. S3. View of the folding motions and the reactive modes in the reactant state of CM. Red arrows represent the vector of the motion along the reaction coordinate for the catalytic residues that interact with the substrate, and the black arrows represent the direction of the folding coordinate. This figure considers the direction of the folding coordinate starting from a partially unfolded structure (*a*) and a structure close to the folded conformation (*b*). The angles between the two vectors are 94° and 70° in *a* and *b*, respectively.

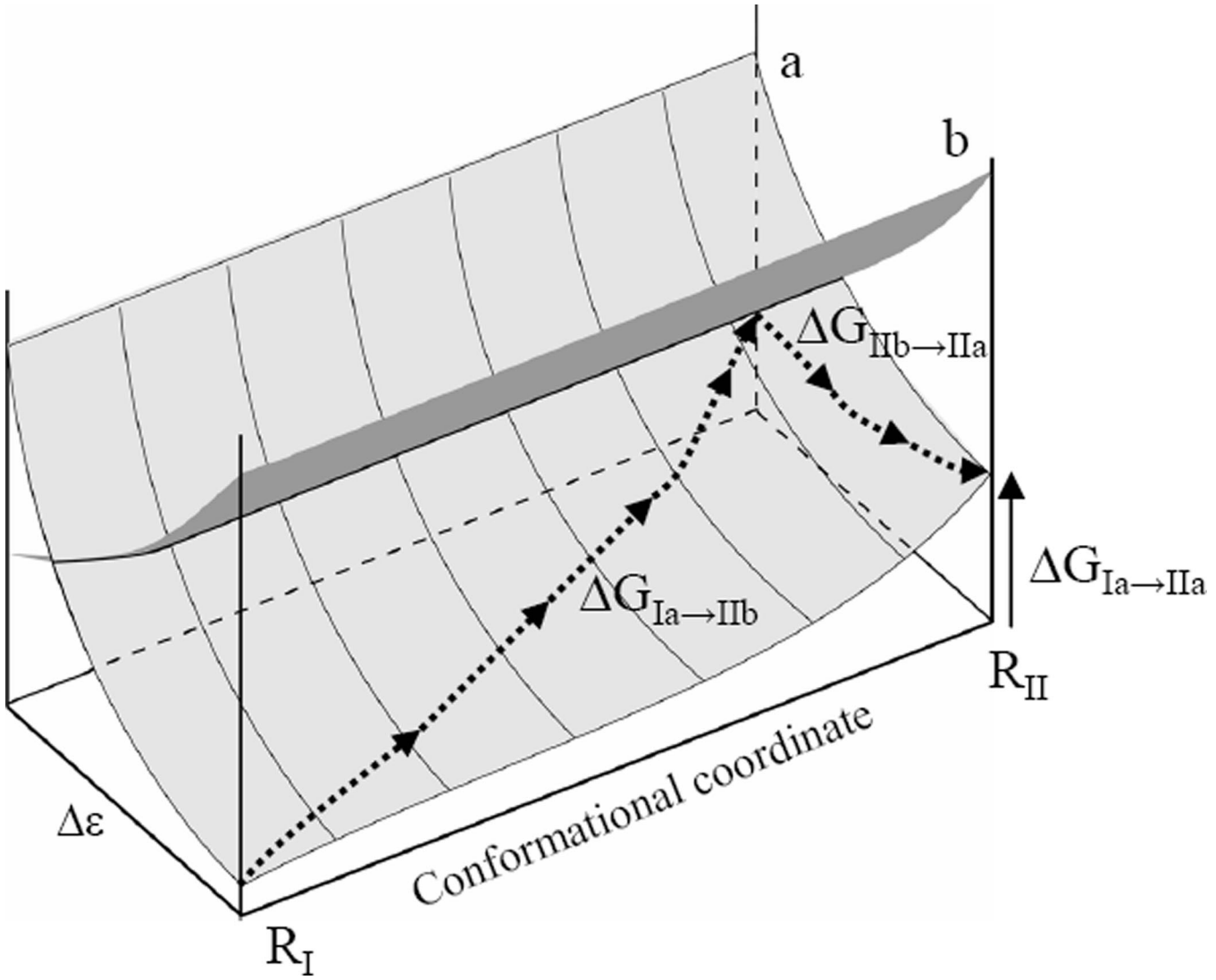

Fig. S4. Illustrating the considerations that lead to Eq. **4**, by schematically depicting two states with charged and uncharged ligand (state a and state b, respectively) in two protein configurations R_I and R_{II}. The use of the LRA approach allows us to estimate the free energy of $\Delta G(U_a(R_I) \rightarrow U_b(R_{II}))$ expressed as $\Delta G_{Ia \rightarrow IIb}$ (which includes the energetics of the protein structural change). The terms $\Delta G(U_b(R_{II}) \rightarrow U_a(R_{II}))$ and $\Delta G(U_a(R_I) \rightarrow U_a(R_{II}))$ are denoted by $(\Delta G_{IIb \rightarrow IIa})$ and $\Delta G_{Ia \rightarrow IIa}$, respectively.

Table S1. Activation barriers ($\Delta g^{\ddagger}$) and rmsd of different configurations generated in the three different regions for the monomer and the dimer

Region I		Region II		Region III	
$\Delta g^{\ddagger}$, kcal/mol	rmsd, Å	$\Delta g^{\ddagger}$, kcal/mol	rmsd, Å	$\Delta g^{\ddagger}$, kcal/mol	rmsd, Å
Monomer					
13.2	0.8	16.6	3.4	26.6	5.2
14.3	0.8	20.3	3.5	36.2	5.3
16.9	0.8	15.3	3.5	29.3	5.4
16.4	0.8	25.0	3.6	27.3	5.4
19.6	0.8	17.1	3.6	28.8	5.4
13.9	0.8	17.2	3.6	23.0	5.4
14.9	0.8	19.7	3.6	32.4	5.4
16.7	0.8	23.5	3.6	24.1	5.4
17.4	0.8	18.4	3.6	22.2	5.5
16.2	0.8	18.3	3.6	28.9	5.5
12.8	0.8	19.4	3.7	26.8	5.5
15.5	0.8	19.1	3.7	35.5	5.5
15.9	0.8	23.5	3.7	25.4	5.5
16.8	0.8	23.8	3.7	26.3	5.5
19.0	0.8	19.1	3.7	31.1	5.5
13.4	0.8	19.2	3.7	27.1	5.5
13.8	0.8	28.6	3.7	24.4	5.5
16.6	0.8	22.2	3.7	29.6	5.5
14.7	0.9	19.4	3.7	30.6	5.6
18.4	0.9	25.4	3.7	31.6	5.6
18.0	0.9	19.2	3.7	26.7	5.6
14.9	0.9	20.5	3.8	33.1	5.6
15.7	0.9	23.2	3.8	28.9	5.6
18.4	0.9	17.3	3.8	28.8	5.6
15.8	0.9	31.6	3.8	24.2	5.6
17.4	0.9	24.2	3.8	20.2	5.6
17.5	0.9	19.4	3.8	28.4	5.6
16.9	0.9	23.9	3.8	20.7	5.7
13.3	0.9	21.8	3.8	35.1	5.7
18.4	0.9	25.0	3.8	22.5	5.7
14.8	0.9	18.1	3.8	19.7	5.7
14.9	0.9	26.9	3.8	24.5	5.7
17.7	0.9	23.4	3.8	28.0	5.7
19.4	0.9	26.2	3.9	22.0	5.7
15.7	0.9	25.9	3.9	23.3	5.7
16.7	0.9	21.0	3.9	26.0	5.7
16.6	0.9	28.8	3.9	21.4	5.7
13.9	0.9	23.1	3.9	26.5	5.8
16.0	1.0	23.9	3.9	32.6	5.8
19.4	1.0	27.6	3.9	30.2	5.8
Dimer					
14.6	0.8	24.3	3.4	30.7	4.9
13.1	0.8	23.6	3.4	21.8	4.9
13.0	0.8	25.9	3.5	26.9	5.0
14.7	0.8	23.6	3.5	26.1	5.0
13.5	0.8	25.5	3.6	31.2	5.0
15.0	0.8	26.8	3.6	30.1	5.0
15.1	0.8	23.4	3.6	25.6	5.1
13.6	0.9	29.2	3.7	21.6	5.1
12.8	0.9	25.5	3.6	19.6	5.1
17.5	0.9	20.5	3.7	40.1	5.1
13.3	0.9	32.1	3.7	29.0	5.1
12.9	0.9	18.4	3.7	27.1	5.1
15.9	0.9	17.4	3.7	24.1	5.1
18.5	0.9	21.4	3.7	24.7	5.1
17.0	0.9	32.6	3.7	21.9	5.2
16.2	0.9	28.3	3.7	37.0	5.2
18.3	0.9	20.4	3.7	24.4	5.2
15.4	0.9	23.6	3.7	22.0	5.2
15.1	0.9	26.9	3.7	29.2	5.2
17.5	0.9	25.0	3.7	27.7	5.3
15.1	0.9	26.8	3.8	25.9	5.3
13.4	0.9	16.1	3.8	29.7	5.3
15.7	0.9	38.7	3.8	24.1	5.3
16.9	0.9	23.9	3.8	30.0	5.3
14.0	0.9	23.2	3.8	23.4	5.3
13.2	0.9	37.6	3.8	32.0	5.3
18.5	0.9	30.1	3.8	31.0	5.3
14.4	0.9	28.9	3.8	32.1	5.3
16.7	0.9	26.2	3.8	25.9	5.4
17.9	0.9	29.3	3.8	24.1	5.4
14.4	0.9	29.2	3.8	28.4	5.4
15.8	0.9	21.4	3.8	26.5	5.4
13.6	1.0	28.8	3.8	25.6	5.4
14.0	1.0	28.1	3.8	28.6	5.4
12.7	1.0	31.6	3.9	26.6	5.5
19.1	1.0	24.5	3.9	33.2	5.5
15.3	1.0	30.7	3.9	25.0	5.5
15.4	1.0	21.9	3.9	34.0	5.5
16.3	1.0	39.5	3.9	25.4	5.6
15.6	1.0	27.9	3.9	25.9	5.6

Table S2. Using Eqs. 3–5 in calculations of $\Delta\Delta G_{bind}(R_I \rightarrow R_{II})$ and $(\Delta g_p^{\ddagger})_{II}$ for different structures in Region I and Region II for both the monomer and dimer

			Region I			
Monomer	$\langle U_b - U_a \rangle_a$	$\langle U_b - U_a \rangle_b$	ΔG_{bind} (R_I)	$(\Delta g_{cat}^{\ddagger})_I$	$(\Delta g_p^{\ddagger})_I$	
St1	82.26	53.21	−18.90	13.20	−5.70	
St2	82.65	52.14	−21.31	14.29	−7.01	
St3	81.48	56.85	−20.83	16.91	−3.93	
St4	78.71	52.54	−17.97	16.42	−1.57	
Average	81.27	55.30	−19.75	15.20	−4.55	
			Region II			
Monomer	$\langle U_b - U_a \rangle_a$	$\langle U_b - U_a \rangle_b$	ΔG_{bind} (R_{II})	$\Delta\Delta G_{bind}$ ($R_I \rightarrow R_{II}$)	$(\Delta g_{cat}^{\ddagger})_{II}$	$(\Delta g^{\dagger})_{II}$
St1	86.62	42.98	−15.18	−2.18	20.29	18.11
St2	78.23	57.27	−19.86	2.21	16.64	18.85
St3	80.54	50.81	−16.54	0.47	25.02	25.49
St4	81.21	50.39	−18.08	−1.25	17.10	15.85
Average	81.65	51.47	−18.84	−0.19	19.75	19.56
			Region I			
Dimer	$\langle U_b - U_a \rangle_a$	$\langle U_b - U_a \rangle_b$	ΔG_{bind} (R_I)	$(\Delta g_{cat}^{\ddagger})_I$	$(\Delta g_p^{\ddagger})_I$	
St1	85.63	46.33	−19.64	14.59	−5.04	
St2	83.11	50.50	−17.58	13.10	−4.48	
St3	88.14	51.15	−21.39	13.03	−8.39	
St4	79.15	51.40	−20.41	14.68	−5.71	
Average	84.01	49.58	−19.76	13.85	−5.91	
			Region II			
Dimer	$\langle U_b - U_a \rangle_a$	$\langle U_b - U_a \rangle_b$	ΔG_{bind} (R_{II})	$\Delta\Delta G_{bind}$ ($R_I \rightarrow R_{II}$)	$(\Delta g_{cat}^{\ddagger})_{II}$	$(\Delta g^{\dagger})_{II}$
St1	86.28	49.71	−19.70	−0.33	22.36	22.03
St2	84.77	42.28	−17.59	−0.83	20.52	19.69
St3	80.56	46.96	−15.08	3.79	25.08	28.87
St4	81.01	48.00	−17.14	−0.93	19.21	18.28
Average	83.16	46.74	−17.38	0.43	21.79	22.22

The energies are given in kcal/mol. Note that $\Delta G_{bind}(R_{II})$ does not consider the energetics of the conformational changes. Furthermore, the PDLD/S-LRA value of ΔG_{bind} for both regions reflects an overestimate because we deal with strong charge–charge interactions and need a somewhat larger dielectric constant. Thus, instead of reporting $(\Delta g_p^{\ddagger})_{II}$ we report $(\Delta g^{\ddagger})_{II} = (\Delta g_{cat}^{\ddagger})_I + \Delta\Delta G_{bind}(R_I \rightarrow R_{II})$ as the "corrected" $(\Delta g_{cat}^{\ddagger})_{II}$ which reflects the change in the ground sate energy upon moving to region II. The corresponding $(\Delta g_p^{\ddagger})_{II}$ can be obtained easily by adding the average $(\Delta G_{bind})_I$ (or just $(\Delta G_{bind})_{I\ observed}$) to all the $(\Delta g_{cat}^{\ddagger})_{II}$ (see Eq. **3**).