Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

SUPPORTING INFORMATION

Supplementary Figures S1-S5

Supplementary Table S1

Supplementary Experimental Procedures

Fine-Tuning of Microsolvation and Hydrogen Bond Interaction Regulates Substrate Channeling in the Course of Flavonoid Biosynthesis.

Julien Diharce, Jérôme Golebiowski, Sébastien Fiorucci*, Serge Antonczak*

Institut de Chimie de Nice, UMR-CNRS 7272, Faculté des Sciences, Université de Nice-Sophia Antipolis, 06108 Nice Cedex 2, France

* correspondence :
Serge Antonczak, Prof.
Institut de Chimie de Nice, UMR-CNRS 7272
Faculté des Sciences,
Université de Nice-Sophia Antipolis,
06108 Nice Cedex 2, France
Phone : (33) 492076122
e-mail : Serge.Antonczak@unice.fr

Supplementary Figure S1 : Structural analysis of DFR-LAR complexes.

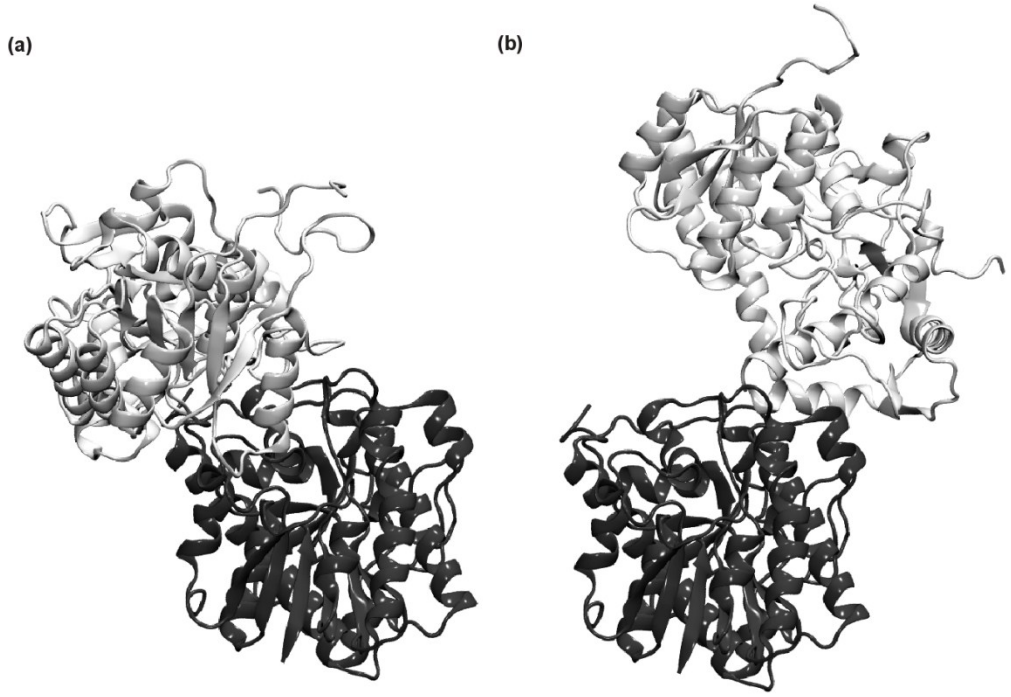

Figure S1: Complex structure of the enzyme DFR (in grey) and LAR (in black) in both PP1 (a) and PP2 (b) docking solutions. To predict the structure of the DFR-LAR complex, a virtual screening of all putative protein-protein interfaces was performed using the ATTRACT docking program.[1] 110000 protein-protein interfaces were energy minimized. The top 10 poses exhibit an ATTRACT interaction energy ranging from -9.7 to -8.4 kcal.mol^{-1}. The enzyme active sites should be closed enough to enable the substrate transfer directly from an enzyme to the following one, thus, among docking solutions, the orientation of the proteins was checked measuring the distance between the active sites. The residues I171 of LAR and T159 of DFR have been identified as key residues at the entrance of enzyme active site (cf. Supplementary Figure S3). The distance values between their C_{alpha} atoms range from 14 to 70 Å. Docking poses featuring the shortest distance (< 30 Å) were kept for MD refinement. Two poses with a I171-T159 distance of 20 and 16 Å, satisfy the structural requirements. They are ranked 4th and 7th according to the attract score with -8.7 and -8.5 kcal.mol^{-1} respectively. In the following, these complexes are subjected to molecular dynamics simulations and are respectively called PP1 and PP2 (Figure 1).

Supplementary Table S1. Structural analysis of DFR-LAR docking solutions. Description of the composition of the interface for the two selected solutions of the ATTRACT Protocol.

	PP1 Complex	PP2 Complex
Amino-acids involved at the interface	63	42
H-Bonds	12	7
Amino-acids involved in the core	23	30
Non-polar residues (%)	33	48
SASA (Å^2)	1900	2070
Ionic interaction	0	1

As defined by Lo Conte *et al.*[2], a protein-protein interaction is composed of two zones with distinct physico-chemical properties: the core and the rim of the interface. For transient protein-protein complexes, the core is mainly constituted of hydrophobic residues with few water molecules playing a structural role. The rim is partially accessible to solvent and the fraction of polar residues is much higher than in the core. The analysis performed with Ligplot+ [3] and Naccess[4] shows that the core of the interface is more compact for PP2 than for complex PP1. The core of the PP1 interface involves less residues (30 *vs* 23) and its surface is slightly higher (2070 *vs* 1900 Å^2). If we extend the analysis to the rim, the conclusions are still the same with a total number of residues of 63 and 42 for PP1 and PP2 respectively. The fraction of non polar residues is clearly in favor of cluster PP2 (33 *vs* 48 %). However, the number of hydrogen bonds at the interface is higher for PP1 than for PP2 (12 *vs* 7 H bonds). Despite some structural discrepancies, DFR-LAR complex structures exhibit transient properties, in agreement with the biological function of the complex.

Supplementary Figure S2 : Experimental free energy of binding *vs* ATTRACT score.

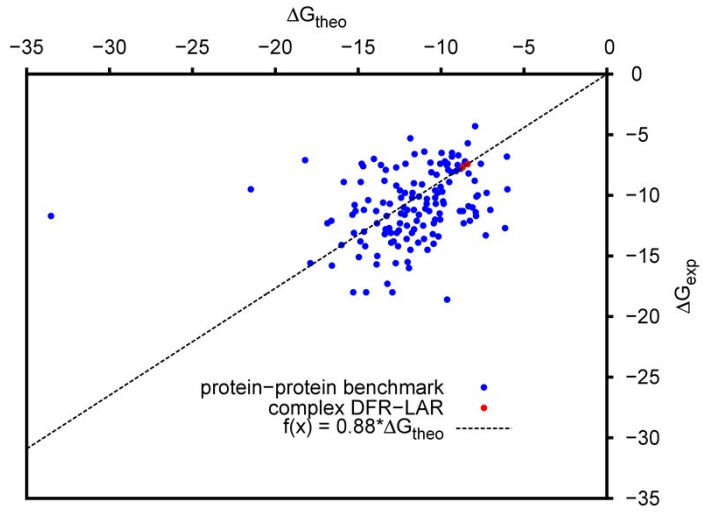

Figure S2. Experimental free energy of binding [5, 6] *vs* ATTRACT score. The two red dot indicate PP1 and PP2 DFR-LAR complexes as predicted by the ATTRACT protocol. The ATTRACT protocol is based on a knowledge based scoring function. The parameter set was optimized to predict near native complex structure in top ranked solutions[7] and not a binding affinity. However, the ATTRACT scoring function was assessed to reproduce experimental binding affinity of transient protein-protein complexes. As mentioned in Kastritis *et al.*[6], docking protocols are still not able to predict experimental binding affinities but they show relatively good performance in classifying the complexes into low, medium and high affinity groups. Comparing experimental binding affinities with the ATTRACT score, the DFR-LAR complex belongs to the weak protein-protein interaction group with an estimated K_D in the range of micromolar (~ 10^{-6} M). Analyzing top ranked solutions, the first ten complexes exhibit a high diversity of putative structures. The broad spectrum of protein-protein complex structures in conjunction with a narrow range of weak interaction energies is a typical feature of a flat energy landscape. It reflects the transient properties of the DFR-LAR complex

Supplementary Figure S3 : Egress routes from DFR and LAR active sites.

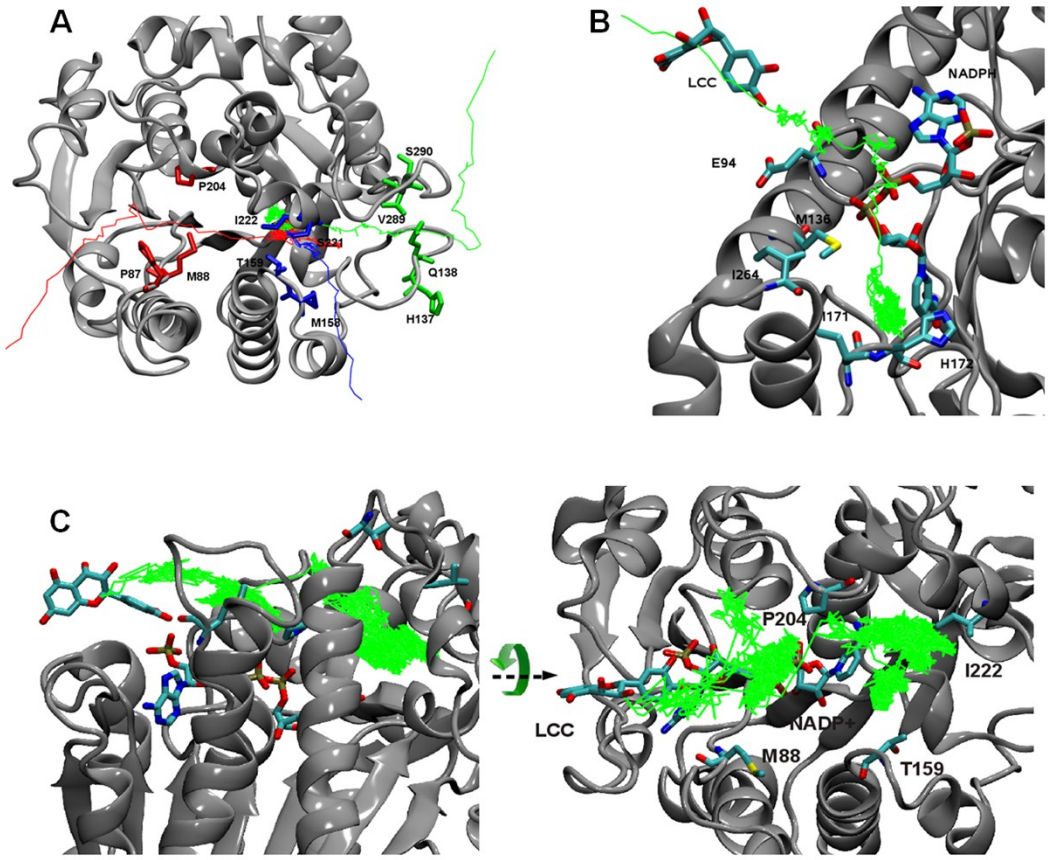

Figure S3. Egress routes from DFR and LAR active sites. (A) Ways of access to the DFR active site detected by RAMD simulations performed on DFR/NADPH/DHQ system. Each different pathway is shown in red, blue and green together with the major amino acids involved in each egress process. (B) Ways of access to the LAR active site detected by RAMD simulations performed on LAR/NADPH/LCC system. Only one route is sampled, represented in green and the amino acids involved during the diffusion are shown in *Licorice*. (C) Side view (left) and top view (right) of DFR/NADP+/LCC system. The trajectory followed by LCC during the unbinding process is shown in green and has been sampled by unconstrained Molecular Dynamics. The cofactor, the LCC ligand and the residues defining the entrance of the DFR catalytic site are shown in *Licorice*. The trajectory followed by LCC during the dissociation process corresponds to the red one previously sampled during RAMD simulations.

Supplementary Figure S4. Description of DFR and LAR active sites

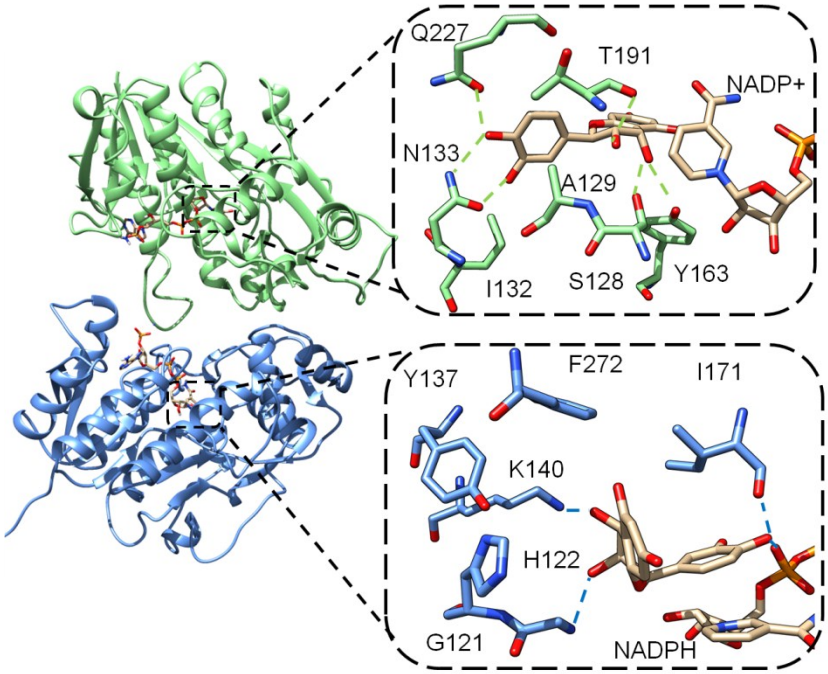

Figure S4. Description of DFR and LAR active sites. Top: LCC-DFR mode of interaction within the enzyme cavity at the beginning of the simulation. The ligand-enzyme interactions have been compared to the ones found in the crystallographic structure (PDB identifier: 2C29). The substrate is stabilized within the active site by residues Q227, N133, T191, S128 and Y163 through electrostatic interactions and by residues A129, I132 through hydrophobic interactions. The reduced cofactor NADP+ also interacts with LCC. Down: LCC-LAR mode of interactions within the enzyme cavity as found at the end of the diffusion process. As in the crystal structure (PDB identifier: 3I52), the substrate interacts with residues K140, F272, G121, H171 and the cofactor NADPH, but the metabolite conformation does not fully match to its position in the crystallographic structure. The RMSd of the LCC position between the crystal structure and the simulation is about 5 Å.

Supplementary Figure S5. Electrostatic pattern during diffusion process.

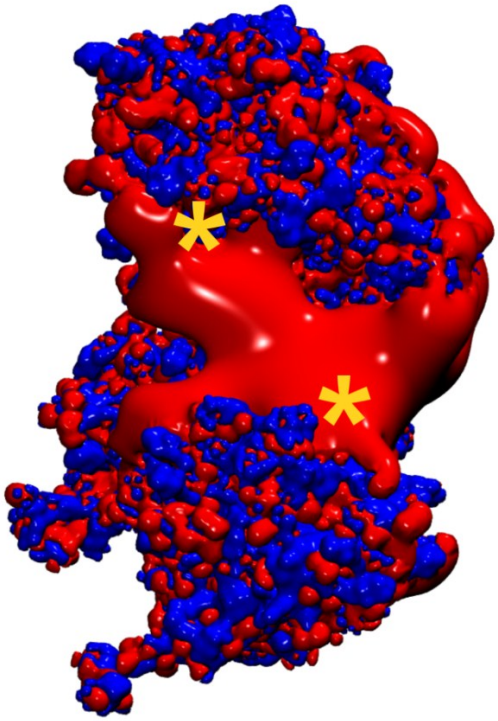

Figure S5. Electrostatic pattern during diffusion process. +6 kT/e (blue) and -6 kT/e (red) isosurfaces of the electrostatic potential of DFR-LAR complex obtained by solving the linear finite-difference Poisson-Boltzmann equations implemented in APBS. Yellow stars indicate the enzyme active sites. The protein-protein interface creates an electrostatic pattern connecting both active sites. It is mostly composed of negative charged residues bearing H-bond capacities. The driven force of the diffusion along this electrostatic pathway is thus the interaction of these residues with the different H-bond donor groups of flavonoid molecules.

Supplementary Experimental Procedures

RAMD simulations.

The paths followed by the substrate or the product between the bulk and the active site have been characterized for each enzyme. This will allow a rational choice of the protein-protein complex with two criteria: energetic and structural.

Considering DFR/NADPH/DHQ and LAR/NADPH/LCC systems independently, 30 Randomly Accelerated Molecular Dynamics (RAMD) trajectories based on different initial velocities have been produced to characterize the active site accesses. For the first case, 3 possible ways to access the DFR active site detected, and none of them are discriminated. For the second one, only one egress route is sampled (Supporting Information Figure 2). This protocol can be used to identify egress routes from a buried protein binding site to the bulk. Random acceleration was set to $0.1 kcal.mol^{-1}.Å^{-1}.g^{-1}$ and applied on the substrate every 40 steps of simulation. The calculation stops automatically when the distance between the centers of mass of the ligand and of the enzyme reaches 40 Å.

The characterization of these accesses permits a better choice of the protein-protein models provided by ATTRACT. Considering the best ranked ATTRACT structures, none of them shows proximity between DFR L2 access and the entrance in LAR active site. Thus, to define the proximity between the active sites, the distance between C_{alpha} carbon atoms of $I171_{LAR}$, located at LAR cavity entry and $T159_{DFR}$, the best candidate amongst all residues constituting DFR red egress route (see Supporting Information Figure 2A), is selected.

References

1 S. Schneider, A. Saladin, S. Fiorucci, C. Prévost and M. Zacharias, in *Computational Drug Discovery and Design*, 2012, vol. 819, pp. 221-232.

2 L. Lo Conte, C. Chothia and J. Janin, *J. Mol. Biol.*, 1999, **285**, 2177-2198.

3 R. A. Laskowski and M. B. Swindells, *J. Chem. Inf. model.*, 2011, **51**, 2778-2786.

4 S. Hubbard and J. Thornton, NACCESS, Computer program, University College London.

5 P. L. Kastritis, I. H. Moal, H. Hwang, Z. Weng, P. A. Bates, A. M. J. J. Bonvin and J. Janin, *Protein Sci.*, 2011, **20**, 482-491.

6 P. L. Kastritis and A. M. J. J. Bonvin, *J. Prot. Res.*, 2010, **9**, 2216-2225.

7 S. Fiorucci and M. Zacharias, *Biophys. J.*, 2010, **98**, 1921-1930.

8 Q. Xing, P. Huang, J. Yang, J.-Q. Sun, Z. Gong, X. Dong, D.-C. Guo, S.-M. Chen, Y.-H. Yang, Y. Wang, M.-H. Yang, M. Yi, Y.-M. Ding, M.-L. Liu, W.-P. Zhang and C. Tang, *Angew. Chem. Int. Ed. Eng.*, 2014, **53**, 11501-11505.