Published in final edited form as:

Mol Phys. 2012 May 10; 110(9-10): 531–536. doi:10.1080/00268976.2012.655337.

Barrier Crossing in Dihydrofolate Reductase does not involve a rate-promoting vibration

Mariangela Dametto^a, Dimitri Antoniou^a, and Steven D. Schwartz^{a,b,c,*}

^aDept of Biophysics, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461, USA

^bDept of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461, USA

^cInstitut des Hautes Études Scientifiques, 91440 Bures-sur-Yvette, France

Abstract

We have studied atomic motions during the chemical reaction catalyzed by the enzyme dihydrofolate reductase of *Escherichia coli* (EcDHFR), an important enzyme for nucleic acid synthesis. In our earlier work on the enzymes human lactate dehydrogenase and purine nucleoside phosphorylase, we had identified fast sub-ps motions that are part of the reaction coordinate. We employed Transition Path Sampling (TPS) and our recently developed reaction coordinate identification methodology to investigate if such fast motions couple to the reaction in DHFR on the barrier-crossing timescale. While we identified some protein motions near the barrier crossing event, these motions do not constitute a compressive promoting vibration, and do not appear as a clearly identifiable protein component in reaction.

1. Introduction

Work by several groups in the last decade has identified enzymes in which dynamical motions seem to play an important role in chemical reaction [1–3]. A system that has received a great deal of attention in these studies of the interplay of catalysis and dynamics is dihydrofolate reductase (DHFR).

DHFR acts in the cell cycle and generates a precursor to nucleotides used to synthesize DNA molecules. The role of DHFR is to reduce dihydrofolate (DHF) to tetrahydrofolate (THF) with the help of the cofactor nicotinamide adenine dinucleotide phosphate (NADPH). The enzyme has a globular structure and contains four β helices (labeled B, C, E and F), eight β sheets (labeled A–H), and other amino acids making the loops between the elements of the secondary structure of DHFR. The helices flank the sheet, two helices being on each side of the sheet. The main flexible loop regions are the M20 loop (residues 14–24), the FG loop (residues 116–125) and the GH loop (residues 64–71). The action of DHFR has been studied both experimentally [4–9] and theoretically [10–13].

The chemical step of the reaction catalyzed by DHFR consists of two sequential steps: protonation of the pteridine ring (with a completely oxidized form of the folate), followed by hydride ion transfer from the cofactor (NADPH) to the DHF. The donor of the hydride is the atom C4 of the nicotinamide ring of NADPH and the acceptor is the C6 atom of the

protonated DHF. Theoretical work [11, 12] has focused on the second step mentioned above (hydride transfer), using both QM/MM and ab initio quantum descriptions of the folate substrate and amino acid residues in the active site. Important ns timescale conformational motions were identified in theoretical studies [5, 14] including the motions that close and open the M20 loop. Hammes *et al.* [15] studied a mutated enzyme and concluded that mutation results in a lower rate constant for the enzymatic reaction. Hammes-Schiffer and coworkers [11] performed mixed QM/MM molecular simulations to generate a free energy profile along chosen coordinates; from geometrical changes observed along these energy profiles they identified important motions, that they assumed are in the ns–ms timescale. They also analyzed one molecular dynamics trajectory and identified a spatially extended fs-timescale motion that is concomitant with the chemical step.

The goal of the work reported in this paper is to determine whether fast protein motions are crucial for chemistry and therefore part of the reaction coordinate, or rather a consequence of chemistry. To fulfill this goal, we examined ensembles of reactive trajectories generated using Transition Path Sampling (TPS), followed by analysis of the transition state ensemble using a method we have developed [16]. In addition to identifying protein motions that are potentially part of the reaction coordinate, we also wanted to determine if this enzyme supports an organized promoting vibration such as we have identified in other enzymes [17–19]. In other words, if there are protein motions that are part of the reaction coordinate, we wish to know if they perform obvious chemical function such as barrier compression (as in lactate dehydrogenase) or bond polarization (as in purine nucleoside phosphorylase.)

2. Methods

We performed simulations on the *E. coli* ternary complex of DHFR with the substrate DHF and the cofactor NADPH using CHARMM35 [20]. The initial coordinates of the complex were obtained from the PDB (Protein Data Bank) entry 1RX2. The folate molecule present in this crystallographic structure was substituted by the DHF molecule using the Discovery Studio package. A box of water molecules of dimensions 60×60×60 Å³ was generated around the complex, with the edges of the water box kept 10 Å away from the protein. The final complex used in the simulations has 22464 atoms: 2489 protein atoms, 69 QM atoms (NADPH and Folate atoms), 58 MM atoms (45 NADPH and 13 Folate atoms), 459 crystallographic waters and 19389 solvated water molecules. The two boundary atoms (Fig. 1) were treated with the Generalized Hybrid Orbital method [21].

For the NADPH and DHF molecules we did not use the standard CHARMM parameter files, but instead we used the topology and parameter files calculated by Garcia-Viloca *et al.* [12]. We used the AM1 semi-empirical model (excluding the water molecules) included in CHARMM27 all-atom force field for the protein. The water molecules were modeled with the TIP3 model. While the AM1 model is approximate, and known, for example, to over-estimate barrier heights, this error in quantum potential is expected to have minimal effect on the vibrational structure of the rest of the protein. As such, we expect little change were we able to implement the calculations with a more exact ab initio method.

The complex was then minimized first for 100 steps of adopted basis Newton-Raphson (ABNR) method and fixing the QM region (with respect to their positions in the crystallographic structure), followed by another 100 steps of ABNR minimization using a 100 kcal/mol-Å harmonic force applied only to the backbone of the protein. We used the protocol of Garcia-Viloca *et al.* [12]: the heating was done slowly from 0 K to 300 K for 30 ps with the harmonic restraint still being applied to the QM atoms, then, the equilibration phase was performed for 50 ps at 300 K, during which the harmonic restraint applied on the rings of the QM region was slowly reduced to 1 kcal/mol-Å.

After that, we generated reactive trajectories using the TPS method. The order parameter we used was: distance hydride-donor $< 1.5 \text{ \AA}$ and distance hydride-acceptor $> 1.5 \text{ \AA}$ for reactant basin; distance hydride-donor $> 1.5 \text{ \AA}$ and distance hydride-acceptor $< 1.5 \text{ \AA}$ for product basin. These distances were used to ensure the observation of the hydride transfer.

As is often the case in TPS simulations, initial trajectory generation often presents the greatest challenge. We generated the initial reactive trajectory by applying a harmonic force equal to $100 \text{ kcal/mol}\cdot\text{\AA}$ on the hydride-acceptor pair in a 0.2 ps MD simulation in order to hold the donor and acceptor close enough for the hydride transfer to occur in a reasonable time scale. The next iteration is generated by randomly selecting a slice along the trajectory, and then assigning a small random momentum perturbation drawn from a Boltzmann distribution to all atoms in the system. This newly generated trajectory is then propagated backward and forward in time. For this study, we generated 100 reactive trajectories with timestep 1 fs of length 500 fs . All the production runs were performed in an IBM iDataplex $\times 3650$ Linux cluster.

Once reactive trajectories have been generated, the next step is identification of the transition state ensemble. Each reactive trajectory contributes one member transition state to the ensemble: the structure that lies on the 0.5 isocommittor surface (trajectories initiated there with velocities drawn from a Boltzmann distribution have probability 0.5 to reach the reactants and 0.5 to reach the products). The complete set of these structures constitutes the transition state (TS) ensemble. Further analysis of the TS ensemble is a formidable task. Our group has suggested [16] a new method to analyze the transition state ensemble, based on a kernel Principal Component Analysis (kPCA).

This method exploits the fact that, by definition, the separatrix is the nonlinear surface spanned by variables that do not participate in the reaction coordinate, so it is a geometric object in configuration space along which progress in the reaction is fixed. Therefore, the separatrix is “thin” along directions perpendicular to it. A standard method that identifies directions of maximum variance is the Principal Component Analysis (PCA), but it has the limitation that it is a linearization procedure, while the separatrix in an enzymatic system is highly nonlinear. The kernel PCA [22] is a nonlinear generalization of PCA that is able to identify nonlinear directions along which the variance of the data is maximal. If the dominant eigenvector of kPCA dominates the variance, it represents the nonlinear direction along which the separatrix is “thickest”, and residues that do not span this direction are the ones that participate in the reaction coordinate (details can be found in [16], where we developed this method.) This method produces, with small computational effort, a good initial guess of residues that participate in the reaction. Once these residues have been identified, one can then further analyze the reactive trajectories and observe specific residue motions.

3. Results and Discussion

The output of our calculations was the ensemble of the trajectories generated by TPS, showing the progress of the reaction in full atomic detail, allowing us to examine residue motions in the timescale of the chemical event. From our TPS simulations, this timescale was found to be of the order of 10 fs and the transfer distance to be $\sim 1.25 \text{ \AA}$.

The first step in our analysis was the examination of the distances between the donor-hydride (NC4-NH42) atoms, the acceptor-hydride (C6-NH42) atoms and the donor-acceptor (NC4-C6) atoms. The donor atom (NC4) is in the NADPH molecule (Fig. 1) and the acceptor atom (C6) is in the DHF molecule. These monitored distances are shown in Fig. 2 for one of the reactive trajectories. This plot shows a compression in the donor-acceptor distance when the hydride transfer takes place. This compression of the donor-acceptor

distance is similar to what we had seen in our earlier work on the enzymes LDH and PNP, where it was also found that there are residue motions that are correlated to this compression. We have also examined overlaps of transition state structures, as identified by committor analysis, of three reactive trajectories. We saw significant similarity among these structures, indicating a common transition state structure.

The next step was the analysis of the transition state ensemble using the kernel PCA analysis, for identifying residues that are part of the reaction coordinate. Earlier works on DHFR had identified important conformational motions in ps-ns timescale, for example the motion of the FG loop [11, 23], and some of these residues (Ile14, Gly15, Phe31, Met42, Gly96, Tyr100, and Gly121) have been suggested to be crucial for catalysis. We examine motions on the timescale (fs) of the chemical event. The only other simulation that examined motions in the timescale of the chemical event was by Hammes-Schiffer [11], who found a concerted motion of several residues along the axis that connects residues Met42 and Gly121.

The results of our kPCA analysis are plotted in Fig. 3, which shows the contributions of the residues to the dominant kPCA component. As explained earlier, residues that do not contribute to this dominant component span directions perpendicular to the separatrix, which is the definition of the reaction coordinate. The residues that are found to be part of the reaction coordinate are those at the local minima in this figure (residues Ile5, Phe31, Met42, Gly43, Thr46, Ile94, Gly95, Gly96, Tyr100 and Leu110; the positions of these residues in the protein is shown in Fig. 4). Then we analyzed the motions of these residues in our set of reactive trajectories and found a pseudo-compressional motion along the curve formed by the residues Ile14, Tyr100, Gly95 and Met42 (Fig. 5). However, this motion is not along the donor-acceptor axis, as can be seen in Fig. 5. Unlike LDH and PNP, there seems to be no compressional motion of residues along the donor-acceptor axis. We should point out that the axis along the residues Ile14 and Met42 is distinct from the axis along residues Met42 and Gly121, along which a compression had been seen in a ps simulation by Hammes-Schiffer and coworkers [11]. Our results agree with the decrease in distance between the residues Met42 and Gly121 (about 18.8 Å) reported in that work, around the hydride transfer event.

We then compared the motions of NADPH's nicotinamide ring (where the donor is located) to the motions of those residues identified by kPCA that lie on the donor side (residues Ile14, Met42, Ile94, Gly95, Tyr100 and Gly121). They form a kind of pocket accommodating the NADPH molecule in the active site. We found that residue Met42 moves away from the NADPH nicotinamide ring by about 0.5 Å and that residue Gly121 gets closer (by about 0.5 Å) to the same ring. We also found that the residues Ile14 and Tyr100 are the ones that are closer to the ring, but their distance from it are kept roughly unchanged. Thus, the motion of the ring where the donor is located seem to be affected mainly by the residues Met42 and Gly121, that are each about 10 Å away. The overall movement of the residues described in this paragraph shows that even though the residues Met42 and Gly121 have a greater motion, the other residues in the pocket (Ile14, Ile94, Gly95 and Tyr100) may be involved in keeping the donor atom in a manner that it is closer to the acceptor, since these residues are the ones closer to the donor.

The above description presents a list of residues that move as the enzymatic reaction crosses the transition state. As important as this list is, it is critical to discuss the significance of these motions. If they are indeed part of the reaction coordinate, then if we started a MD trajectory from a state on the separatrix and propagate the system while keeping these residues fixed (thus, remaining on the separatrix), and finally initiating trajectories from

points along this restrained trajectory, we would get commitment probabilities close to 0.5, which we failed to find by using these residues for the restrained trajectory.

Therefore, our simulations of the DHFR catalyzed reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) indicate that while fast motions are involved in the chemical event, they are not the types of organized dynamics in which the participation of protein dynamics in the reactive event is clear. Examples of this phenomenon would include compressional motions along a donor-acceptor axis in other enzymes [17, 19]; or leaving group bond polarization caused by the compression of electronegative atoms [18]. Our results show that while there are fast motions concomitant with the chemical event, they are exactly the types of small detailed active site changes expected to accompany any charge transfer reaction. These motions do not show any obvious pattern of organization of the relative atomic displacements of transferring partners, or chemically relevant electron polarization. This result in no way speaks to the importance of any longer time networks of motions.

Acknowledgments

This work was supported by National Institutes of Health Grant GM068036.

References

- [1]. McGeagh JD, Ranaghan KE, Mulholland AJ. *Bioch. Biophys. Acta.* 2011; 1814:1077.
- [2]. Henzler-Wildman K, Thai V, Lei. M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson M, Petsko G, Karplus M, Hubner C, Kern D. *Nature.* 2007; 450:838. [PubMed: 18026086]
- [3]. Schwartz S, Schramm V. *Nature Chem. Biol.* 2009; 5:551. [PubMed: 19620996]
- [4]. Sawaya M, Kraut J. *Biochemistry.* 1997; 36:586. [PubMed: 9012674]
- [5]. Rajagopalan P, Lutz S, Benkovic S. *Biochemistry.* 2002; 41:12618. [PubMed: 12379104]
- [6]. Osborne M, Schnell J, Benkovic S, Dyson H, Wright P. *Biochemistry.* 2001; 40:9846. [PubMed: 11502178]
- [7]. Maglia G, Allemann RK. *J. Am. Chem. Soc.* 2003; 125:13372. [PubMed: 14583029]
- [8]. Boehr D, Dyson H, Wright P. *Chem. Rev.* 2006; 106:3055. [PubMed: 16895318]
- [9]. Swanwick R, Maglia G, Tey L, Allemann R. *Biochem. J.* 2006; 394:259. [PubMed: 16241906]
- [10]. Radkiewicz J, Brooks CL. *J. Am. Chem. Soc.* 2000; 122:225.
- [11]. Agarwal PK, Billeter SR, Rajagopalan PTR, Benkovic SJ, Hammes-Schiffer S. *Proc. Natl. Acad. Sci. USA.* 2002; 99:2794. [PubMed: 11867722]
- [12]. Garcia-Viloca M, Truhlar D, Gao J. *Biochemistry.* 2003; 42:13558. [PubMed: 14622003]
- [13]. Arora K, Brooks CL III. *J. Am. Chem. Soc.* 2009; 131:5642. [PubMed: 19323547]
- [14]. Thorpe I, Brooks C. *J. Phys. Chem. B.* 2003; 107:14042.
- [15]. Antikainen N, Smiley R, Benkovic S, Hammes G. *Biochemistry.* 2005; 44:16835. [PubMed: 16363797]
- [16]. Antoniou D, Schwartz SD. *J. Phys. Chem. B.* 2011; 115:2465. [PubMed: 21332236]
- [17]. Basner JE, Schwartz SD. *J. Phys. Chem. B.* 2004; 108:444.
- [18]. Saen-oon S, Quaytman S, Schramm V, Schwartz S. *Proc. Natl. Acad. Sci. USA.* 2008; 105:16543. [PubMed: 18946041]
- [19]. Quaytman S, Schwartz S. *Proc. Natl. Acad. Sci. USA.* 2007; 104:12253. [PubMed: 17640885]
- [20]. Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S, Karplus M. *J. Comput. Chem.* 1983; 4:187.
- [21]. Gao J, Amara P, Alhambra C, Field M. *J. Phys. Chem. A.* 1998; 102:4714.
- [22]. Schölkopf, B.; Smola, A. *Learning with kernels: support vector machines, regularization, optimization and beyond.* MIT Press; Cambridge, MA: 2002.
- [23]. Rod TH, Radkiewicz JL, Brooks CL. *Proc. Natl. Acad. Sci. USA.* 2003; 100:6980. [PubMed: 12756296]

Figure 1.

Schematic representation of the DHF molecule (on top) and the NADPH molecule (at the bottom) showing the QM and MM regions (divided by the straight line in the figure) used in the simulations. Atom A in the DHF molecule is the hydride acceptor atom and the atom D in the NADPH molecule is the donor atom. The D and A atoms are both in the QM region. Atoms named as B in each molecule are the boundary atoms (both carbon atoms).

Figure 2. Geometric changes, for one reactive trajectory, associated with the distance between the donor-hydride (NC4-NH42) atoms (black line), the acceptor-hydride (C6-NH42) atoms (red line) and the donor-acceptor (NC4-C6) atoms (green line).

Figure 3. Plot of the kPCA results. Projections of the residues on the principal component analysis found by kPCA (vertical axis) and the residue index (horizontal axis). The red dots at minima represent residues that are possibly part of the reaction coordinate.

Figure 4.

Representation of the most relevant residues found in the kPCA in the donor (residues Met42, Gly43, Thr46, Gly95, Gly96, and Tyr100) and acceptor (residues Ile5, Phe31 and Ile94) sides. These residues show fewer motions (described in the text) related to the donor and acceptor atoms. The cofactor molecule (NADPH - QM region) is where the donor (blue dot) is located and the substrate (DHF - QM region) is the molecule where the acceptor (cyan dot) is located.

Figure 5.

A representation of the direction of the only compressional motion seen in this work. D is the donor and A is the acceptor. The residues shown are the ones we found in the kPCA analysis having a minimum lower than 0.025, except for residue Ile14. Residue Phe31 is the only one in the acceptor side. The substrate and the NADPH molecules are shown in gray and the rest of the protein is shown as ribbons.