CS279
Steven Fried
Dec. 3, 2021

Molecular Dynamics Simulation of Electrostatic Interactions between NADPH and Dihydrofolate Reductase

Introduction. Electrostatic interactions have long been known to play key roles in protein-ligand interactions, including interactions of enzymes with substrates and cofactors. One enzyme that has generated much interest in terms of the role of electrostatics in its function is dihydrofolate reductase (DHFR), which catalyzes a hydride transfer from nicotinamide adenine dinucleotide phosphate (NADPH) to dihydrofolate, producing folate as an important precursor in purine biosynthesis. In one respect, electrostatic interactions are thought to play a key role in binding of NADPH to the protein. Because NADPH is a highly anionic cofactor (with typical net charge –4), the phosphate linkages joining the nicotinamide riboside and adenosine moieties possess substantial negative charge and bind the enzyme in the vicinity of positively charged and polar residues[1-2]. Moreover, it is thought that an electric field along the direction of hydride transfer between nicotinamide donor and dihydrofolate acceptor plays a role in facilitating the charge transfer[3]. Hence, I sought here to calculate the magnitude of the electrostatic interactions of NADPH with DHFR, particularly in terms of the electric field DHFR exerts on the hydride transfer axis of NADPH, as well as on the phosphate groups that help position the NADPH ligand in its binding pocket. I approached this problem with an interest in how the protein plays a functional role in both types of interactions compared to solvent, as well as how the conformational flexibility of the protein dictates a distribution of electrostatic environments in each region.

Background. While Coulombic forces on various parts of NADPH in DHFR could be calculated directly from the crystal structure of the complex, the conformation of the enzyme imposed by crystallization conditions may not be most reflective of the native protein conformation and cofactor binding pose, and furthermore does not reflect the broader ensemble of

conformational (and electrostatic) states and substates characterizing this highly dynamic enzyme. Hence, I employed molecular dynamics (MD) simulations of NADPH in complex with *e. coli* DHFR as a means of sampling this conformation ensemble and determining the effects of protein dynamics on the electric field that NADPH experiences in its highly charged phosphate regions and on the catalytic hydride of nicotinamide. A similar MD-based approach, though also combined with quantum mechanical calculations, has been used by Liu *et al.* to determine the field along the hydride transfer axis and at various points around the cofactor binding site. One strategy for using MD simulations to measure the magnitudes of electric fields on bonds has been well characterized by the Boxer group at Stanford for small molecules in solutions[4]. However, extending this existing protocol towards a ligand within a protein should lead to more heterogeneity in terms of the ligand's environment, and a more complex workflow necessary to set up the simulations.

Methods. As a starting model of DHFR, I used the X-ray crystallographic structure 1rx9.pdb, containing a modified form of $NADP^+$ with an extra sugar attached to one of the bridging phosphates[2]. Since this ligand is slightly different from the native NADPH cofactor, I had to modify its structure slightly using the program Avogadro. I added hydrogens, removed the extra sugar group on one phosphate, and added an extra hydrogen to carbon-4 of nicotinamide, where the hydride would add from. Having modified the nicotinamide portion of the structure in this way, I then optimized only this region of the structure (fixing all other atoms in place) using a steepest-descent minimization routine in Avogadro relying upon a universal force field (UFF). The original NADPH structure I created had hydrogens on two of the phosphate groups, which may have been responsible for some early difficulties I had in keeping the ligand-protein complex stable during MD equilibration steps. Under biological conditions, it is unlikely for these phosphate groups to be protonated, so I fixed this on a subsequent attempt. I then parameterized the DHFR ligand using

Antechamber and LEaP from the AmberTools20 package, using the General Amber Force Field (GAFF)[5], because this force field has been well characterized in terms of calculating electrostatic interactions on small molecules[4] and is compatible with the Amber force field I used for the protein. The ligand parameter files were converted to GROMACS-compatible files using preexisting scripts. See the included code for the commands used in each case.

In GROMACS-2018, I converted the crystallographic structure from 1rx9 to a GROMACS structure including hydrogens and generated a topology using the AMBER99SB-ILDN force field, which has been well-optimized for proteins[6]. The commands I used to carry out these various steps are provided in the included code depository. After adding back in the ligand topology, I solvated the complex using TIP3P parameters for water. The unit cell was neutralized through the addition of 30 Na^+ and 15 Cl^- ions to bring the ionic strength to 0.1 M. The complex was minimized in energy according to steepest descent minimization until the force size was less than 0.1 kJ/mol. While restraining both the protein heavy atoms and the ligand independently, I then equilibrated the system using an *N,V,T* ensemble at 300K using a modified Berendsen thermostat for 100,000 steps of 1-fs each. This was followed by two consecutive equilibrations of similar durations and parameters in the *N,P,T* ensemble, first restraining all the protein heavy atoms and then only the alpha carbons to free up the system more gradually (Fig. 1). The simulation itself was conducted for a total of 50 ns using 1-fs timesteps, saving coordinates and forces every 10 ps. A Parinello-Rahman thermostat was used to keep the temperature at 300K, while long-range Van der Waals and Coulombic forces were handled by Verlet scheme and particle-mesh Ewald, respectively, each with a 1.2 nm cutoff. In addition to the simulation in protein, another simulation for NADPH in water was conducted using similar parameters but only one NPT equilibration step and a shorter time (1 ns), as solvent motions should be faster than protein conformation changes.

Results and Analysis. Simulations were analyzed by preexisting scripts provided by the Boxer laboratory, designed to calculate electric fields on small molecules in solvents, modified slightly to meet the specific conditions of this more complex system. The completed simulations were rerun using the same trajectory coordinates but with charge-neutral topology files. I then created new index files in GROMACS for each of the bonds along which I desired to measure an electric field, and extracted the coordinates of and forces on these atoms every 10 ps from both the original and charge-neutral simulation. The electric field projected along a bond between two atoms A and B is calculated by $F = |\vec{F}_{env}| = \frac{1}{2}(\vec{F}_{\mathrm{A}} \cdot \vec{u}_{\mathrm{AB}} + \vec{F}_{\mathrm{B}} \cdot \vec{u}_{\mathrm{AB}})$, where $\vec{F}_{\mathrm{A}}$ and $\vec{F}_{\mathrm{B}}$ are the field vectors on either atom (forces of charged trajectory minus forces of neutral trajectory, then divided by the charge of either atom from the topology) and $\vec{u}_{\mathrm{AB}}$ is the bond unit vector.

I calculated the field along the hydride transfer axis (C4 to the pro-R hydrogen, Fig. 2) and found a broad distribution of electric field values on the C-H bond induced by the protein/solvent environment. An approximately Gaussian distribution yielded a mean electric field of 14.3 MV/cm, with standard deviation 17.7 MV/cm (Fig. 3). When compared to a simulation using similar parameters for the cofactor in only water over 1 ns time, the average field calculated within the protein is slightly less positive than that value calculated in only water (27.4 MV/cm, with standard deviation 21.2 MV/cm, Fig. 4). However, my calculated electric field was still substantially different from the negative field of the protein on the hydride transfer axis reported by Liu *et al.*, who reported a field of -16.5 MV/cm resulting from just the protein and solvent on the hydride transfer axis in a complex of DHFR with NAD^+ and Folate[3]. Since the binding of folate into the substrate binding pocket could alter the protein conformation and displace many water molecules, it is not unreasonable that the negative field is not recapitulated in my result. The fact that I calculate the field on C-H in the protein to be less positive than the field in only water

indicates the protein environment does have a lowering effect on the field by about 10 MV/cm; removal of additional water by substrate binding may bring this result closer to Liu *et al.*

An additional simulation of DHFR with both NADPH and substrate (dihydrofolate) bound would help to indicate whether the field along the hydride transfer axis does become significantly negative to facilitate the catalytic redox step; hence this is a logical next step. It may also be worthwhile trying a range of different force fields to determine the robustness of the result with respect to potential energy functions used, particularly as some previous analysis has noted significant sensitivity to the parameterizations of complex molecules with uneven charge distributions, much like NADPH[4]. Performing the simulations for longer periods of time, though more computationally expensive, may also sample some other conformations of the DHFR not evaluated in this project, such as the highly dynamical movement of the M20 loop which takes place on a longer timescale[2]. Another next step could involve experimenting with quantum mechanical calculations of the field, perhaps using a density functional theory approach based on frames from the MD trajectory, which should give greater accuracy and will more closely resemble the QM/MM framework used by Liu *et al.* in their calculation of the same quantity[3].

Besides the field along the hydride transfer axis, I also investigated the electric field vectors describing interactions of DHFR with the charged phosphate regions of NADPH as a clue to the factors governing cofactor binding specificity and affinity. Specifically, what does this reveal about which phosphate is most strongly interacting with the protein, and how much does this contribute to stability of the complex? I calculated electric fields projected along each charge-carrying P-O bond expected to experience strong electrostatic interactions with its surroundings, and compared the distributions I calculated in the protein environment versus those in water. Several charged and polar groups in the vicinity of the NADPH phosphates, most notably two

arginines and a histidine, are positioned near the NADPH binding site, so I hypothesized that these phosphates would experience stronger fields as a result. In contrast to my expectations, nearly every phosphate P-O bond experienced a stronger electric field in the pure water solvent compared to the protein environment. Table 1 reports these average electric field magnitudes projected along the bonds labeled in Figure 5, where each field distribution was approximately Gaussian distributed, with the standard deviations reported in parentheses. Many phosphate bonds experience stronger electrostatic fields in water, and also a broader distribution indicative of more configurations of solvent arrangement compared to the more ordered protein. Some exceptions to this trend include the bonds P2-O1 and P-O2, whose interaction with the solvent is disrupted by the molecule bending over on itself (Fig. 6). Altogether, however, the trend in these results suggests that a charged protein environment is not necessarily more stabilizing to the charged phosphate groups of NADPH than water already is, due to water's high polarity. Rather, the charge-charge interactions of NADPH with DHFR probably help to properly position the ligand in the binding site, where other types of interactions, such as the hydrophobic effect with some of the nonpolar regions of the molecule, play a bigger role in lowering the overall free energy of the complex versus the dissociated ligand. A ligand docking algorithm may give some indication, via a scoring function, of the approximate free energy contributions that offer greatest stabilization to NADPH binding, and thus would pose a good next step to take for this aspect of the question.

In these experiments, I learned how to set up, run, and analyze a molecular dynamics simulation of a ligand bound to a protein in GROMACS. With regards to the question at hand—how DHFR uses electrostatics to promote hydride transfer and stabilize cofactor binding—my simulations show that the protein environment makes the electric field along the hydride transfer axis less positive compared to NADPH in water, potentially setting the stage for a negative field

that facilitates hydride transfer once substrate binds. Additionally, binding of the NADPH ligand is kept in position by electric fields along the phosphate bonds to nearby positive regions of DHFR, although these interactions likely do not pose significantly greater enthalpic stabilization relative to the protein in water. Overall, use of molecular dynamics to sample a range of conformational space and solvent configurations has provided new insights into how a ligand's electrostatic environment within protein compares to that in solution, with relevance to the ligand's binding specificity and catalysis of hydride transfer in enzyme function.

Figures.

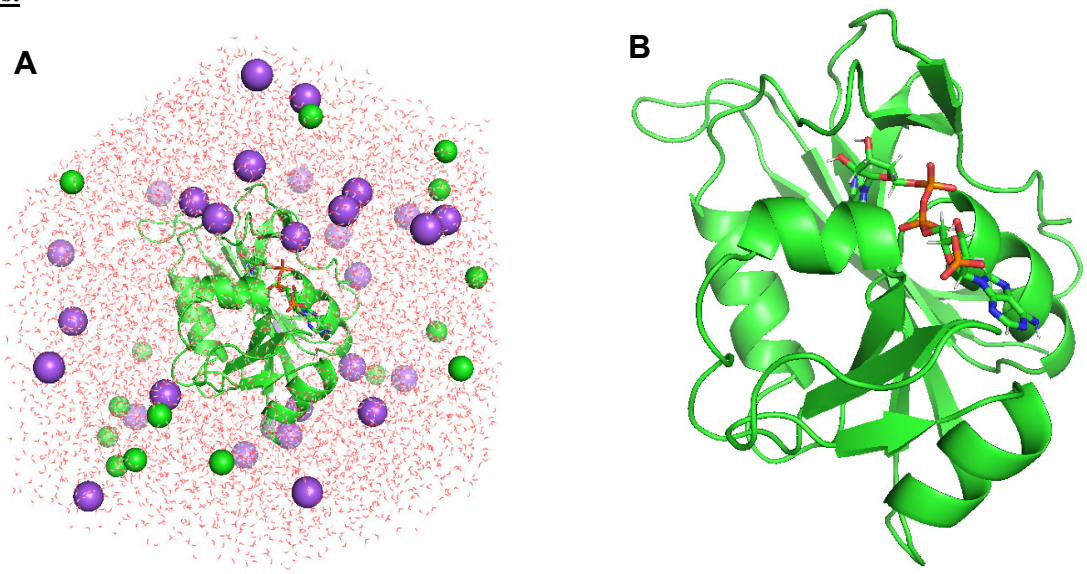

Fig. 1 (A) Solvated DHFR-NADPH complex unit cell, and (B) only the DHFR-NADPH complex after energy minimization and equilibration.

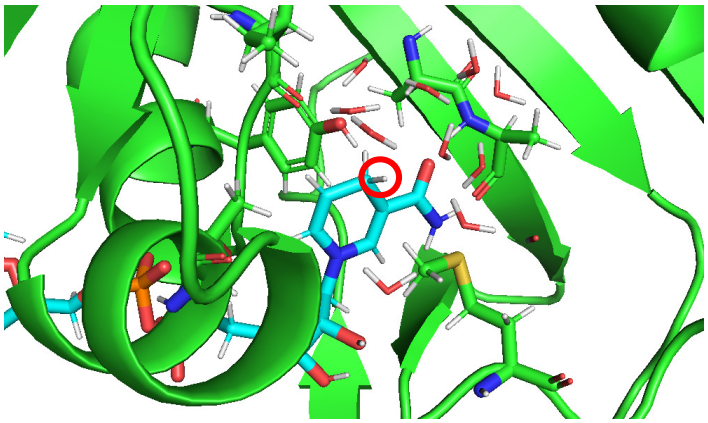

Fig. 2 Nicotinamide environment, with cofactor in blue and protein in green. Hydride is transferred from along red circled C-H axis when substrate is present (though here the substrate binding site is filled with water). Factors likely responsible for the typically positive electric field calculated along the C-H axis include the organization of the water solvent (recapitulated in the plain solvated NADPH simulation) and also the O-H proton from neighboring Tyr-100, which carries a positive charge.

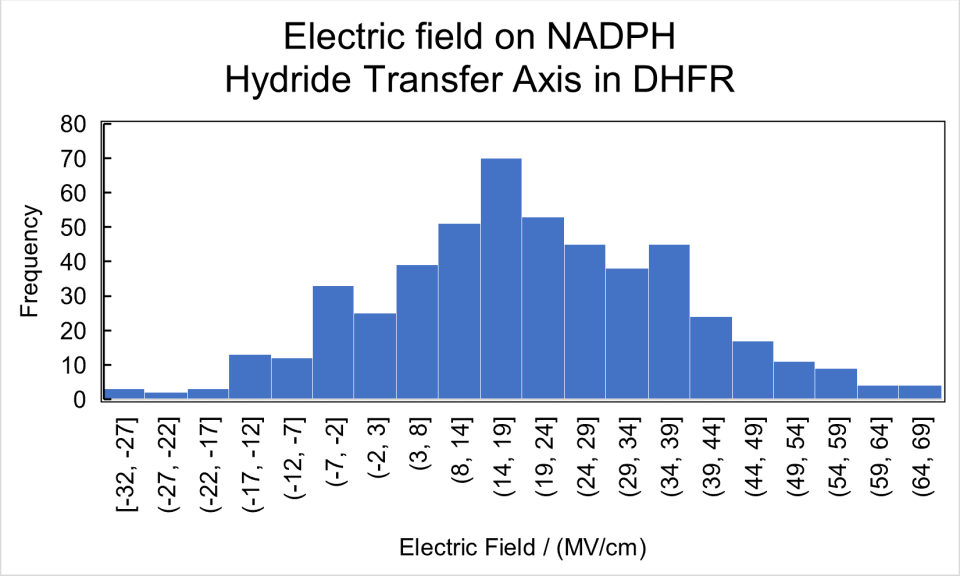

Fig. 3 Histogram showing broad distribution of electric fields experienced by the NADPH bond to its transferable hydride, within a DHFR environment. The mean field experienced is 14.3 MV/cm, with standard deviation 17.7 MV/cm.

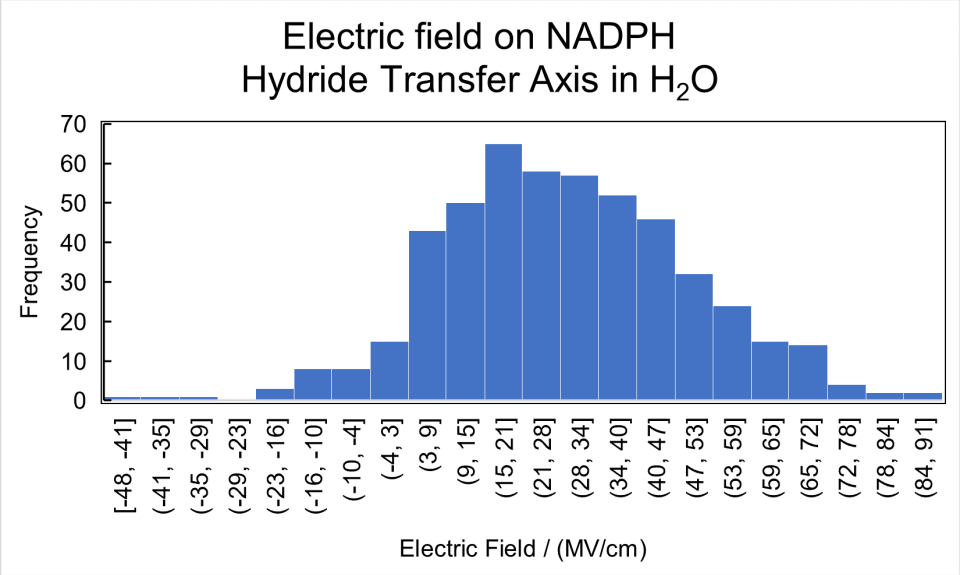

Fig. 4 Histogram showing broad distribution of electric fields experienced by the NADPH bond to its transferable hydride, within a water solvent. The mean field experienced is 27.4 MV/cm, with standard deviation 21.2 MV/cm.

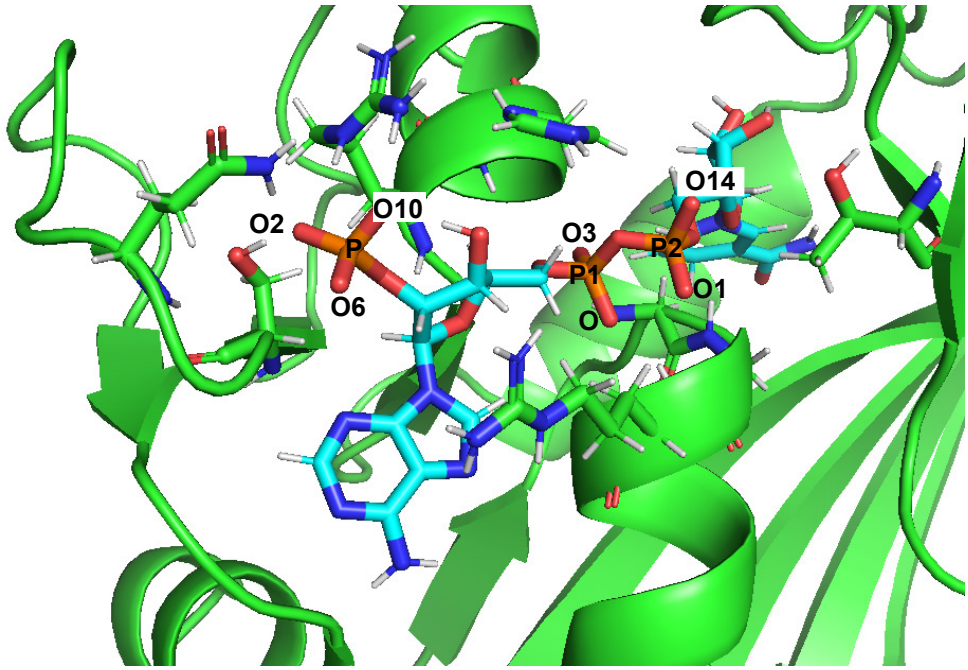

Fig. 5 Binding site of NADPH, with phosphate atoms labeled for which electric field projections were calculated.

Bond pair	**Field In DHFR in MV/cm**	**Field In Water in MV/cm**
C12-H13	14.3 (S = 17.7)	17.4 (S = 21.2)
P1-O	-10.7 (S = 15.9)	-60.7 (S = 21.5)
P1-O3	-34.3 (S = 25.9)	-48.6 (S = 33.9)
P2-O1	-68.1 (S = 25.9)	-60.7 (S = 20.0)
P2-O14	-48.7 (S = 25.9)	-63.2 (S = 24.4)
P-O2	-26.9 (S = 23.0)	-15.3 (S = 34.6)
P-O6	-25.9 (S = 28.9)	-37.7 (S = 27.5)
P-O10	-23 (S = 25.3)	-25.2 (S = 25.1)

Table 1. Average values (with standard deviations in parentheses) of the approximately Gaussian distributed electric field distributions projected along various bonds labeled in figures 2 and 5.

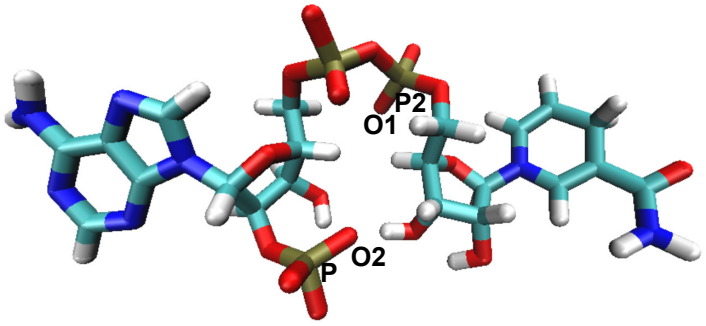

Fig. 6 Folded conformation of NADPH observed when simulating the molecule in water.

References.

1. Komatsu, K.; Nakagawa, S.; Umeyama, H.; Nakamura, H., Electrostatic Interaction Energy and Solvent Accessibility in the Methotrexate-Reduced Nicotinamide Adenine Dinucleotide Phosphate-Dihydrofolate Reductase Ternary Complex. *Chem. Pharm. Bull.* **1987,** *35* (5), 1880-1895.
2. Sawaya, M. R.; Kraut, J., Loop and Subdomain Movements in the Mechanism of Escherichia coli Dihydrofolate Reductase: Crystallographic Evidence. *Biochemistry* **1997,** *36* (3), 586-603.
3. Liu, C. T.; Layfield, J. P.; Stewart, R. J.; French, J. B.; Hanoian, P.; Asbury, J. B.; Hammes-Schiffer, S.; Benkovic, S. J., Probing the Electrostatics of Active Site Microenvironments along the Catalytic Cycle for Escherichia coli Dihydrofolate Reductase. *J. Am. Chem. Soc.* **2014,** *136* (29), 10349-10360.
4. Kozuch, J.; Schneider, S. H.; Zheng, C.; Ji, Z.; Bradshaw, R. T.; Boxer, S. G., Testing the Limitations of MD-Based Local Electric Fields Using the Vibrational Stark Effect in Solution: Penicillin G as a Test Case. *J. Phys. Chem. B* **2021,** *125* (17), 4415-4427.
5. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.
6. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010,** *78* (8), 1950-1958.