

A Molecular Dynamics Study of the Ligand Release Path in Yeast Cytosine Deaminase

Lishan Yao,^{*†} Honggao Yan,^{†‡} and Robert I. Cukier^{*‡}

^{*}Department of Chemistry, [†]Department of Biochemistry and Molecular Biology, and [‡]Michigan State University Center for Biological Modeling, Michigan State University, East Lansing, Michigan 48824

ABSTRACT Yeast cytosine deaminase, a zinc metalloenzyme, catalyzes the deamination of cytosine to uracil. Experimental and computational evidence indicates that the rate-limiting step is product release, instead of the chemical reaction step. In this work, we use molecular dynamics to suggest ligand exit paths. Simulation at 300 K shows that the active site is well protected by the C-terminal helix (residues 150–158) and F-114 loop (residues 111–117) and that on the molecular dynamics timescale water does not flow in or out of the active site. In contrast, simulation at 320 K shows a significant increase in flexibility of the C-terminal helix and F-114 loop. The motions of these two regions at 320 K open the active site and permit water molecules to diffuse into and out of the active site through two paths with one much more favored than the other. Cytosine is pushed out of the active site by a restraint method in two directions specified by these two paths. In path 1 the required motion of the protein is local—involving only the C-terminal helix and F-114 loop—and two residues, F-114 and I-156, are identified that have to be moved away to let cytosine out; whereas in path 2, the protein has to rearrange itself much more extensively, and the changes are also much larger compared to the path 1 simulation.

INTRODUCTION

Yeast cytosine deaminase (yCD) catalyzes the deamination of cytosine to uracil. Cytosine deaminase is present in bacteria and fungi as part of the pyrimidine salvage pathway but is absent in mammals (1). yCD can also catalyze the deamination of the prodrug 5-fluorocytosine (5-FC) to the anticancer drug 5-fluorouracil (5-FU). Thus, a potential system for gene-directed enzyme prodrug therapy combines yCD and the prodrug 5-FC, since the product, 5-FU, inhibits DNA synthesis and is thus a potent chemotherapeutic agent (2).

X-ray structures of yCD in the apo form (3) and in complex with the potent inhibitor 2-pyrimidinone (3,4) are available. yCD is a homodimer with one active site in each protomer. The yCD protomer contains a center β sheet ($\beta 1 \sim \beta 5$, parallel to each other) with two α helices ($\alpha 1$ and $\alpha 5$) on one side and four α helices ($\alpha 2$, $\alpha 3$, $\alpha 4$, and $\alpha 6$) on the other side. Each active center contains a single catalytic zinc ion that is tetrahedrally coordinated with a histidine (H-62), two cysteines (C-91 and C-94), and a water molecule in the substrate-free enzyme or the inhibitor in the complex. Surprisingly, the structure of apo yCD is essentially the same as that of the transition state analog complex with the active site completely buried. Previous molecular dynamics (MD) simulations for the apo form and reactant various intermediate and product complexes (5) all showed a buried active site during the ~ 2.0 -ns simulations. The protein overall is quite rigid in those simulations, consistent with NMR backbone nuclear Over-

hauser effect data and order parameters (L. Yao and H. Yan, unpublished data), which describe motion in the picosecond-nanosecond timescale. Quench-flow experiments performed to study the reaction mechanism indicate that product release is the rate-limiting step during the activation of 5-FC (6). A slow product release process was also demonstrated by a ^{19}F -NMR study, with the product release rate determined to be $\sim 13 \text{ s}^{-1}$ (6). To enhance the product release rate, and therefore improve the efficiency of this enzyme, which is the essential goal of our study, the product release process has to be understood.

The slow product release is most likely due to subtle conformational transitions in the protein. To identify potential ligand release paths, we have performed MD simulations with the use of a restraint method that is derived from umbrella sampling methods. The aim of this study is to explore the dynamic process of product release rather than constructing a potential of mean force (PMF) since, first, the product release path is unknown and, second, the product release rate is in the seconds timescale, which is much longer than an MD simulation can reach. Even the use of umbrella sampling-based methods (7) in this case where there must be large barriers to product release would make the calculation of a PMF extremely difficult. It would also be less meaningful for our purposes. Instead, we believe that identifying the product release path is more important and can guide experimental mutagenesis studies, which in turn could confirm our computational work. For these reasons we adopt a methodology that is quite similar to steered molecular dynamics (SMD), which has been used extensively in studying protein ligand binding and unbinding process in various systems (8,9). The only difference is that in steered MD the ligand is usually pushed continuously, whereas in our case the ligand is pushed out to be centered around a series of distances for a given

Submitted October 10, 2006, and accepted for publication December 11, 2006.

Address reprint requests to Professor Honggao Yan, Dept. of Biochemistry and Molecular Biology, Michigan State University, E. Lansing, MI 48824-1322. Tel.: 517-353-5282; Fax 517-353-9334; E-mail: yanh@msu.edu; or Professor R. I. Cukier, Dept. of Chemistry, Michigan State University, E. Lansing, MI 48824-1322. Tel.: 517-355-9715 x 263; Fax: 517-353-1793; E-mail: cukier@cem.msu.edu.

© 2007 by the Biophysical Society

0006-3495/07/04/2301/10 \$2.00

doi: 10.1529/biophysj.106.098921

number of MD steps, and then the rest of the system is allowed to relax to the given distance restraint.

But, before pushing the ligand out of the active site, potential paths must be chosen. The small root mean square deviation (RMSD) between the apo and transition state analog complex x-ray structures, including the residues that surround the active site, suggests that for yCD typical nanosecond simulations at 300 K can hardly give us valid information about the ligand release path and the residues that are involved in the release. So, the strategy we adopt is to perform one regular simulation for the apo enzyme at 300 K and another at 320 K. By comparing these two trajectories, we might be able to identify the “soft spot” of the protein. The higher temperature provides larger thermal fluctuations and therefore a better chance for the protein to overcome certain barriers to reach some conformations that are less populated at 300 K. In this case, these conformations would correspond to the opening of the active site while still maintaining the protein’s secondary and tertiary structure. The simulation results show that the flexibility of certain residues does increase significantly more than others. In particular, the F-114 loop (residues 111–117) and C-terminal helix (residues 150–158) that cover the active site do fluctuate more when comparing the 300-K and 320-K simulations. The motions of these two regions at 320 K open the active site and permit water molecules to diffuse into and out of the active site through two paths. Based on the identification of these paths, cytosine was pushed out of the active site and the residues that respond to the release monitored. In this manner, motions of yCD regions required to release the ligand could be suggested.

METHODS

MD simulation of the yCD apo form

The parameterization of the force field for Zn complexes was described in detail in a previous article (5). The protonation states of the ionizable residues were set to their normal ionization states at pH 7. The protein was surrounded by a periodic box of approximate dimensions $90 \times 88 \times 70 \text{ \AA}$ leading to solvation by $\sim 17,000$ TIP3P water molecules. Six Na^+ ions were placed by the LEAP program (10) to neutralize the -6 charges of the model system. The parm94 version of the all-atom AMBER force field was used to model this system.

MD simulations were carried out using the SANDER module in AMBER 7.0 (11) at constant pressure and temperature. The time step was 2 fs, and the SHAKE algorithm was used to constrain all bonds involving hydrogen atoms. A nonbond pair list cutoff of 8.0 \AA was used and the pair list updated every 25 steps. The pressure and the temperature (300 K and 320 K) of the system were controlled during the MD simulations by Berendsen’s method (12). To include the contributions of long-range electrostatic interactions, the particle mesh Ewald method was used (13). A 2-ns simulation was performed at 300 K, with 500 ps of equilibration time. The 320-K simulation protocol is slightly different from the 300 K one. After a 1-ns simulation, 3 ns of simulation were continued. Meanwhile, four other 1-ns independent simulations were started by reassigning the atom velocities to the last snapshot of the first nanosecond simulation based on the Maxwell-Boltzmann distribution at 320 K. So effectively, an 8-ns simulation was done at 320 K. The first 500 ps of the regular simulation and the first 200 ps of the velocity-reassigned simulations were treated as equilibration periods and therefore not included in the data analysis.

The coordinates were saved every 2 ps and were analyzed with the Moil-View program and the PTRAJ module of AMBER 7.0. Principal component

analysis (PCA) was performed by using the gcovar and ganeig modules in GROMACS to filter out the fast motions (14).

Pushing ligand out of the active site

A 2-ns MD simulation was previously performed for the yCD uracil (product) complex at constant temperature (300 K) and volume (5). The simulation showed that the active sites in both protomers are buried and the protein overall is quite rigid, similar to the 300-K 2-ns simulation of the apo form. But, the O4 of uracil is covalently linked to the Zn atom of the active site, which raises a difficulty for the product release path study. On the other hand, the reactant cytosine has a very similar structure to the product uracil and similar active site interactions. Furthermore, the yCD cytosine complex 300-K MD simulation also shows that the protein is rigid and that the active site remains buried (5). Thus, cytosine is used as the ligand to be pushed out of the active site.

To explore how cytosine gets out of the active site, a restraint was introduced to push it along the two paths that we identified by examining the data from the 320-K apo form simulation. The first path is in between the F-114 loop (residues 111–117) and the C-terminal helix (residues 150–158), and the second path is in between loop 1 (residues 53–61) and the C-terminal helix. A harmonic potential was added between the center of mass (COM) of the cytosine ring heavy atoms and the COM of backbone heavy atoms of H-50 (S-89) in path 1 (2). The reasons to choose H-50 are that it is from the core region ($\beta 2$) and therefore quite rigid, and it is along the path 1 direction of ligand release. Similar reasoning was used in choosing S-89 in path 2. The starting structure was taken from one snapshot of the cytosine complex MD simulation. The equilibrium distance was increased 0.1 \AA once every 120 ps to give an effective force to push cytosine out of the active site. The 120-ps simulation in between two neighboring pushes allows the system to relax and permits a better statistical evaluation of the force needed to push the ligand out. Cytosines were pushed 13 \AA along path 1 in protomer 1 and path 2 in protomer 2 in a total of 15.6 ns of simulation time, giving the effective pushing rate of $\sim 0.82 \text{ \AA/ns}$.

The pushing velocity is an important parameter in the simulation. Considering that the experimental rate of product release is 13 s^{-1} , which is impossible to simulate with current computer resources, we use a pushing speed as slow as is practical to mimic the unbinding process. The speed we use is considerably slower than the rate typically used (15–17). As noticed in an unbinding study of an acetylcholinesterase ligand complex (16), the slower the pushing rate, the smaller the irreversible work needed to push the ligand out. The slower pushing rate would allow the ligand to explore more configuration space and be more likely to find the real path(s). On the other hand, our test simulations with a faster pushing velocity (7 \AA/ns), starting from different initial structures, showed no sign of distortion of the overall structure of the protein. Thus, the velocity we adopted (0.82 \AA/ns) is a reasonable compromise between computational cost-based considerations and the genuine property of the system. As discussed above, the aim of this study is to discover potential ligand release paths and consequent protein dynamics rather than an evaluation of the PMF for ligand release.

The force constant used is $20 \text{ kcal/(mol}\cdot\text{\AA}^2)$, which corresponds to a thermal fluctuation of distance of $\sim 0.25 \text{ \AA}$. This pushing rate provides a continuous sampling of the cytosine release path. Test simulations with the 7-\AA/ns pushing speed showed that this force constant is large enough to push cytosine out efficiently—similar to the pushing simulations performed on another system we studied (18) and consistent with what other researchers used in ligand-unbinding studies (8,15,19). The restraint force was calculated and averaged over each window. The first 20 ps of simulation in a given window was treated as the equilibration period and therefore not included in the force calculation. But, in the RMSD calculation between the instantaneous MD trajectory and the crystal structure, all the snapshots were used since no apparent difference can be seen between the equilibration and data collection period in terms of structure.

Since the two ligands were pushed out simultaneously along two different paths, it might cause some synchronized artificial effect. To make sure this is not the case, a simulation where a single ligand was pushed along path 1 (at the rate of 7 \AA/ns) was carried out. No changes in the dynamics of

the active site in the adjacent protomer were found, probably because the two active sites are ~ 18 Å apart, and the pushing simulation results in the ligand farther away from the adjacent active site.

RESULTS AND DISCUSSION

Apo form simulations at 300 K and 320 K

yCD is a homodimer with 158 residues in each subunit. The data analysis was based on each subunit to improve the statistics because the crystal structure shows that the two subunits are the same, except for several N-terminal residues. Because previous MD simulations showed that the N-terminus is far more flexible than the rest of the protein (5) and is also far away from the protein active site, the first 14 residues were not included in the fitting for the calculation of the RMSD from the crystal structure or the root mean-square fluctuation (RMSF) from the average structure.

The time evolutions of the α -carbon (CA)-based RMSD of MD snapshots from the crystal structure stabilize after 500 ps, indicating that the two systems are in equilibrium during the analyzed trajectory (data not shown). Fig. 1 shows RMSF plots at 300 and 320 K for the two different subunits, as well as the RMSF calculated from the crystal B-factors. As can be seen, they are consistent with the crystal B-factors. The protein overall is quite rigid with average RMSF ~ 0.41 Å and ~ 0.44 Å for protomer 1 and 2, respectively, at 300 K. The increase of temperature from 300 K to 320 K causes a significant increase of protein flexibility (Fig. 1). The average RMSFs are ~ 0.62 Å for protomer 1 and ~ 0.67 Å for protomer 2 at 320 K, $\sim 50\%$ larger than those at 300 K. This flexibility increase mainly comes from the relatively flexible regions at 300 K, among which the C-terminal helix shows the largest RMSF increase in both protomers (Fig. 1). Since the C-terminal helix covers the active site, as shown in the crystal structure (3,4), the significant increase of RMSF in this region might be important to open up the active site and let the reactant enter and exit it.

To explore more details about the protein fluctuation changes due to the increase of temperature, a PCA (20,21) was used to analyze the MD trajectories. PCA attempts to decompose the overall atom fluctuations into a small set of modes that encompass a large percentage of the overall RMSF^2 and a much larger set of modes that contribute a relatively small amount. In this way, we can focus on the small set of modes that represent slow, large-scale motions because large portions of the flexibility change usually come from these modes. By comparing these significant motion modes at the two temperatures, we can obtain instructive information about large-scale motions that might be responsible for ligand binding or release.

As shown in Table 1, the total variance (RMSF^2) of the CA atoms (residues 15–158) is ~ 25 Å² and ~ 29 Å² for protomer 1 and 2, respectively, at 300 K. The 320-K simulation gives much bigger variances with ~ 65 Å² for protomer 1 and ~ 76 Å² for protomer 2. It is surprising that the total fluctuations approximately double even though the temperature only

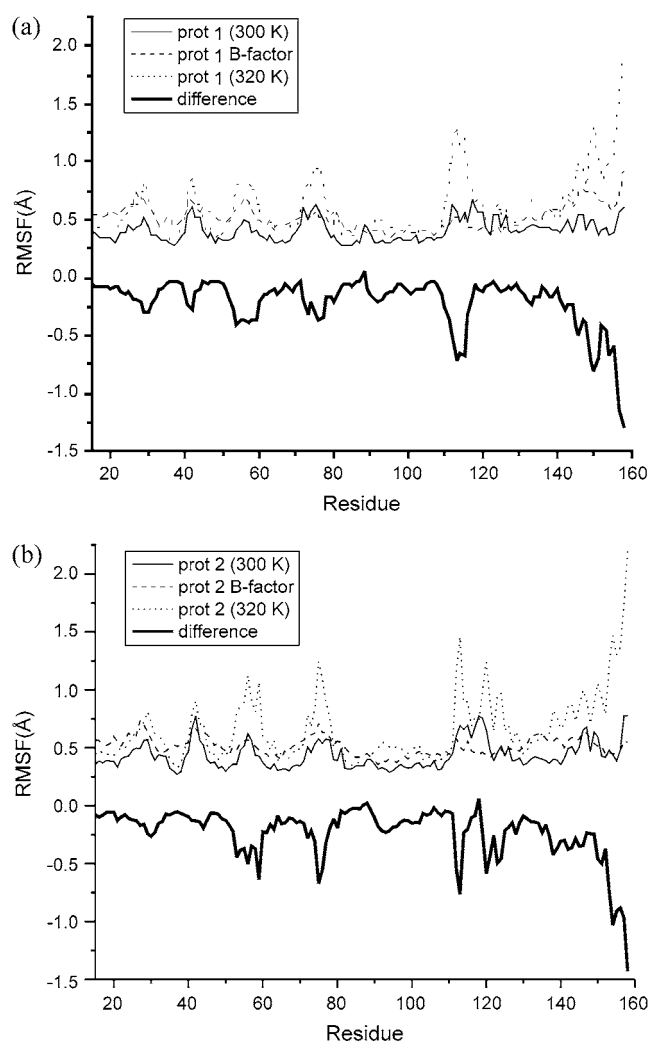


FIGURE 1 CA RMSFs of protomer 1 (a) and protomer 2 (b) at 300 K (solid line) and 320 K (dotted line). The crystal B-factors were converted (23) to RMSFs (dashed line). The heavy black lines show the RMSF differences between the 300 K and 320 K data.

increases $\sim 7\%$. Assuming harmonic vibrational motion for the CA atoms, the increase of total variance should scale with the temperature increase. The large deviation from this expectation might be due to two main reasons. First, the increase of temperature gives a better chance to overcome certain barriers and therefore explore configuration space more effectively. Second, the longer simulation time at 320 K (8 ns) can also cover more configuration space than the 2-ns simulation at

TABLE 1 Total CA fluctuations (RMSF^2) and percentage contributions from the first five PCA modes at 300 K and 320 K

	300 K		320 K	
	Protomer 1	Protomer 2	Protomer 1	Protomer 2
CA fluctuation (Å ²)	25.4	29.4	65.1	76.1
First five modes	24%	29%	48%	47%

300 K. Nevertheless, it is significant that a small increase of temperature triggers such a dramatic fluctuation change without denaturing the protein and indicates that new, slower functional motion modes might be captured in the 320-K simulation.

To see whether these suppositions are true, the PCA was used to filter out the fast motions and keep the slow ones of primary interest. The first five modes contribute 24% and 29% at 300 K compared with 48% and 47% at 320 K to the total RMSF² for protomers 1 and 2, respectively. The fluctuation difference in the first five modes between these two temperatures is $\sim 25 \text{ \AA}^2$ ($\sim 27 \text{ \AA}^2$) for protomer 1 (2), contributing 63% (73%) of the total fluctuation difference. Thus the major motion differences between the 300 K and 320 K simulations were captured in the first five modes.

The RMSF plots of individual residues at these two temperatures can provide information about which regions have large flexibility changes. Fig. 2 shows the CA RMSFs corresponding to the first five modes. The main difference between the 300 K and 320 K results is from the flexible regions. Though the trajectories of these two protomers show some differences of the flexibility increase in certain regions, such as residues 120–125 (part of helix 4) and 144–149 (part of helix 5), which is probably due to incomplete sampling of conformation space, both protomers show consistently large RMSF differences in the C-terminal helix and F-114 loop regions. To get a better view of the flexibility changes, a schematic graph was generated by superimposing the 50 MD CA snapshots projected onto the first five modes for individual protomers at these two simulation temperatures (Fig. 3). Clearly, the C-terminal helix and F-114 loop show much larger fluctuations at 320 K than those at 300 K. Furthermore, it reveals that the motion of these two regions can expose the active site to the solvent, whereby water molecules can move freely in and out of the active site, unlike the 300-K simulation in which the active site is completely buried (Fig. 3). We identified a total of 27 water molecules that diffuse into the active site and 11 water molecules that diffuse out through the path in between the C-terminal helix and F-114 loop in both protomers (Fig. 4). The 300-K MD simulation shows that the side chains of F-114, W-152, and I-156 cover the active site, whereas F-114 and I-156 block this path. The increase of flexibility of the C-terminal helix and F-114 loop at 320 K moves these two side chains slightly away from their original positions; as a consequence the active site is opened up. Three residues, C-91, F-114, and I-156, were used to define qualitatively the entrance of this water path (Fig. 4). Labeling the C-91 CA-F-114 CZ distance as d1, the C-91 CA-I-156 CD distance as d2, and the F-114 CZ-I-156 CD distance as d3 and parametrically plotting these distances for the 300-K and 320-K simulations in three-dimensions (3D) reveals that at 300 K the distances are restrained in the lower left corner, whereas at 320 K they are spread from the lower left toward the upper right in the orientation of the figure. The average distances and their fluctuations for d1, d2, and d3 are $5.25 \pm 0.72 \text{ \AA}$, $8.04 \pm$

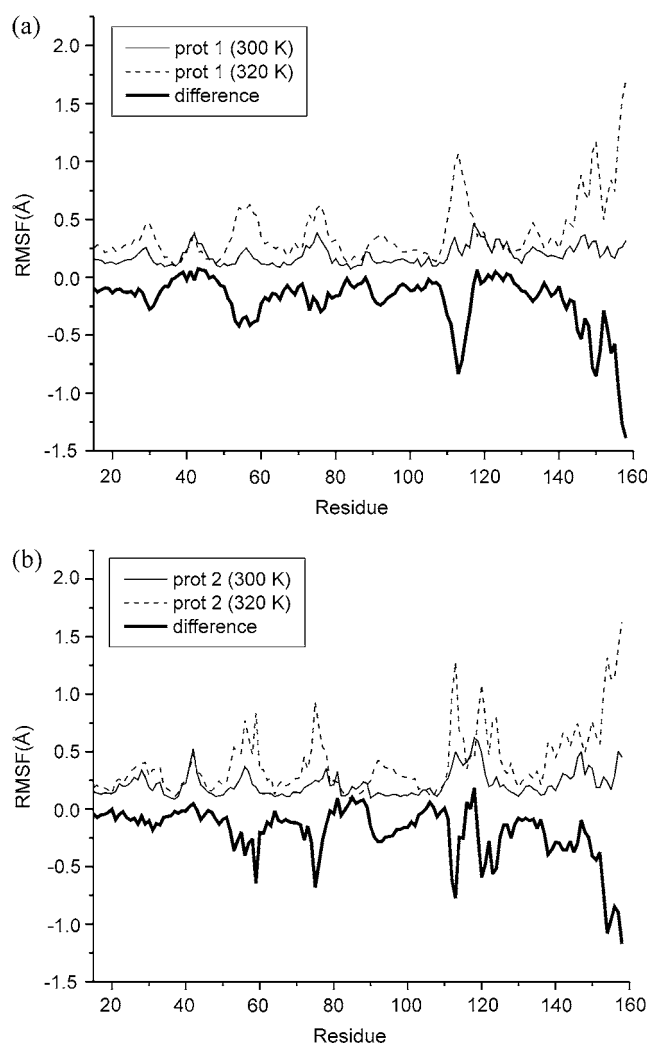


FIGURE 2 CA RMSF plots corresponding to the first five PCA modes at 300 K and 320 K and their difference: (a) protomer 1, and (b) protomer 2.

1.04 \AA , and $4.65 \pm 0.64 \text{ \AA}$ at 300 K compared with $7.73 \pm 1.90 \text{ \AA}$, $10.20 \pm 1.28 \text{ \AA}$, and $6.12 \pm 1.20 \text{ \AA}$ at 320 K in protomer 1. All the distances and their fluctuations are increased at 320 K, indicating that the mouth of the active site has opened up and the breathing motion is larger at this temperature. The 320 K data, projected onto the three two-dimensional (2D) planes formed from these distances in Fig. 5 *b*, show that the two-state behavior evident in Fig. 5 *a* mainly arises from the C-91 CA-F-114 CZ distance coordinate. Thus, there are barriers that are overcome during the simulation at the higher temperature.

It is interesting that one water molecule moves out of the active site in protomer 2 through the path in between the C-terminal helix and the loop between $\alpha 2$ and $\beta 2$ (residues 51–61), which we refer to as loop 1 (Fig. 4). As shown in Fig. 2, loop 1 is also relatively more flexible at 320 K. Thus, two water diffusion paths were identified. But, it seems that the path in between the F-114 loop and C-terminal helix (path 1) is much more favorable than the path in between the

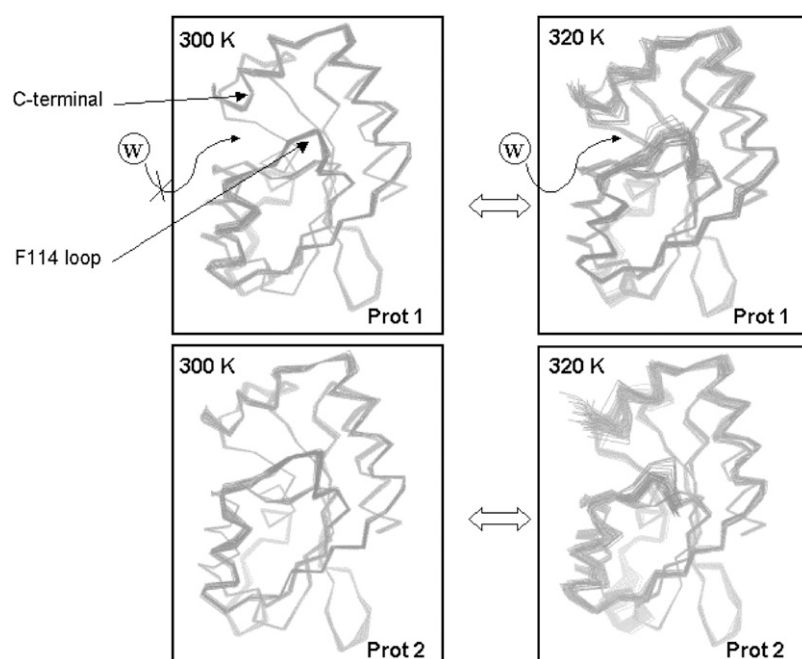


FIGURE 3 Schematic graph of 50 trajectory snapshots of CA atoms projected out of the first five PCA motion modes at 300 K and 320 K. The F-114 loop and C-terminal helix were labeled to give a better view. Water molecules can diffuse into the active site at 320 K through the path in between F-114 loop and C-terminal helix but not at 300 K.

C-terminal helix and loop 1 (path 2), particularly considering that the natural reactant cytosine and the product uracil are much larger than a water molecule. There are 28 events observed for water passing through path 1 but only one event for water passing through path 2. Further investigations of these two paths were carried out by pushing cytosine out of the active site along these two paths as described below.

To summarize the apo simulation results, the 320-K MD simulation increases the flexibility of the protein, and most of the increase comes from some large-scale (slow) motion modes that have relatively large barriers to overcome. The use of high temperature and a longer simulation time provides a greater chance to overcome these barriers and obtain a better description of the protein dynamics. Two pathways were identified for water diffusion into and out of the active site, which are potential

paths for ligand binding or release due to the increased motion of loop 1, the F-114 loop, and the C-terminal helix.

Pushing cytosine out of the active site

To determine whether the two water paths identified in the apo protein simulation are sensible paths for product release, the reactant cytosine was pushed 13 Å away from the active site through these two paths in a 15.6-ns MD simulation. We chose to push cytosine out of the active site because in the Michaelis complex cytosine is not coordinated with the Zn (22). Along the pushing paths, the structure and flexibility changes for the overall protein and individual regions can be investigated and the cytosine-protein restraint force calculated to understand better these two paths and their influence on the overall protein and its various regions. Several residues were identified that may potentially contribute to the slow product release process.

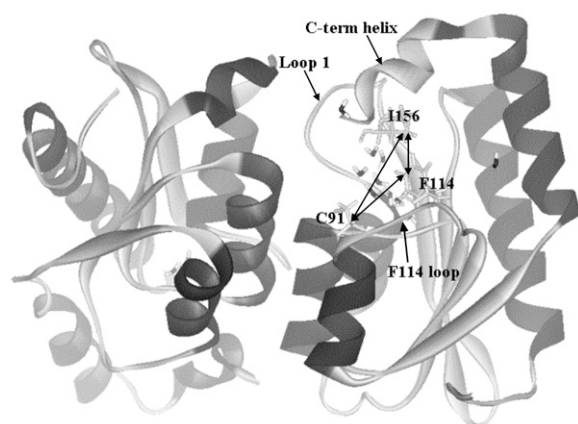


FIGURE 4 One snapshot of the apo MD simulation at 320 K. Water molecules diffuse into the active site through the triangle mouth defined by C-91, F-114, and I-156.

Path 1

In path 1, a harmonic restraint was added between the COM of the cytosine heavy atoms and the COM of the backbone heavy atoms of H-50. This residue is from the β_2 region, which is quite rigid and aligned along the path. So, this residue was selected as the anchor to implement the restraint force, with the parameters given in Section II.

Fig. 6 *a* shows the RMSD (relative to the cytosine yCD complex crystal structure) of all the CAs versus time, with zero time corresponding to a snapshot from a normal MD simulation of the complex. Over the first 7 ns the RMSD is stable at ~ 0.9 Å, which is comparable to the normal MD complex simulation, and then increases slightly to ~ 1.1 Å. The whole protein is quite rigid and the perturbation of the

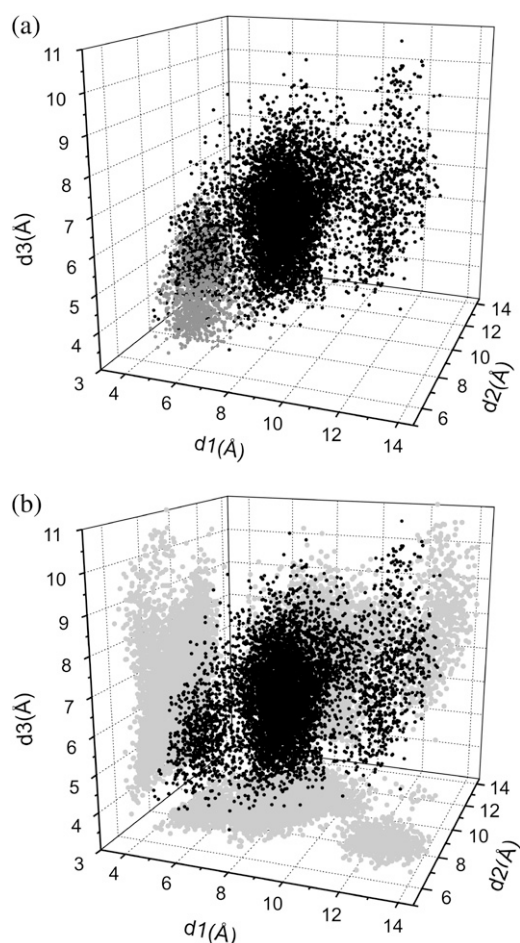


FIGURE 5 (a) 3D plots of distances d1 (C-91 CA-F-114 CZ), d2 (C-91 CA-I-156 CD), and d3 (F-114 CZ-I-156) at 300 K (black) and 320 K (gray). (b) The 320-K data plot (black) with the three 2D projections (gray).

overall structure is rather small during the ligand release. Fig. 6 *b* shows the average pushing force in each window. From the RMSD plot and the force plot (Fig. 6, *a* and *b*) it appears that there are two periods for the reactant exit; a first period from the beginning to around the seventh nanosecond and a second period from the seventh nanosecond to the end. In the first period, the overall structure changes are quite small. The main barrier comes from the direct hydrogen-bonding interactions lost between cytosine and the residues in the active site, which can be confirmed by a hydrogen-bond analysis. Fig. 6 *c* shows the lifetime of the hydrogen bonds during the pushing simulation. There are five hydrogen bonds between cytosine and the protein, including N-51, G-63, E-64, and D-155, when cytosine is well bound in the active site. All these hydrogen bonds are eventually lost during the first period of the pushing simulation. The hydrogen bond with G-63 breaks first, at ~ 3 ns, primarily because G-63 sits at the bottom of the active site and the hydrogen-bond donor, the backbone amide, is quite rigid. Then E-64 and N-51 lose their three hydrogen-bond interactions with cytosine at ~ 5 ns. Lastly, the hydrogen bond between D-155 and cytosine is broken at ~ 7 ns. During

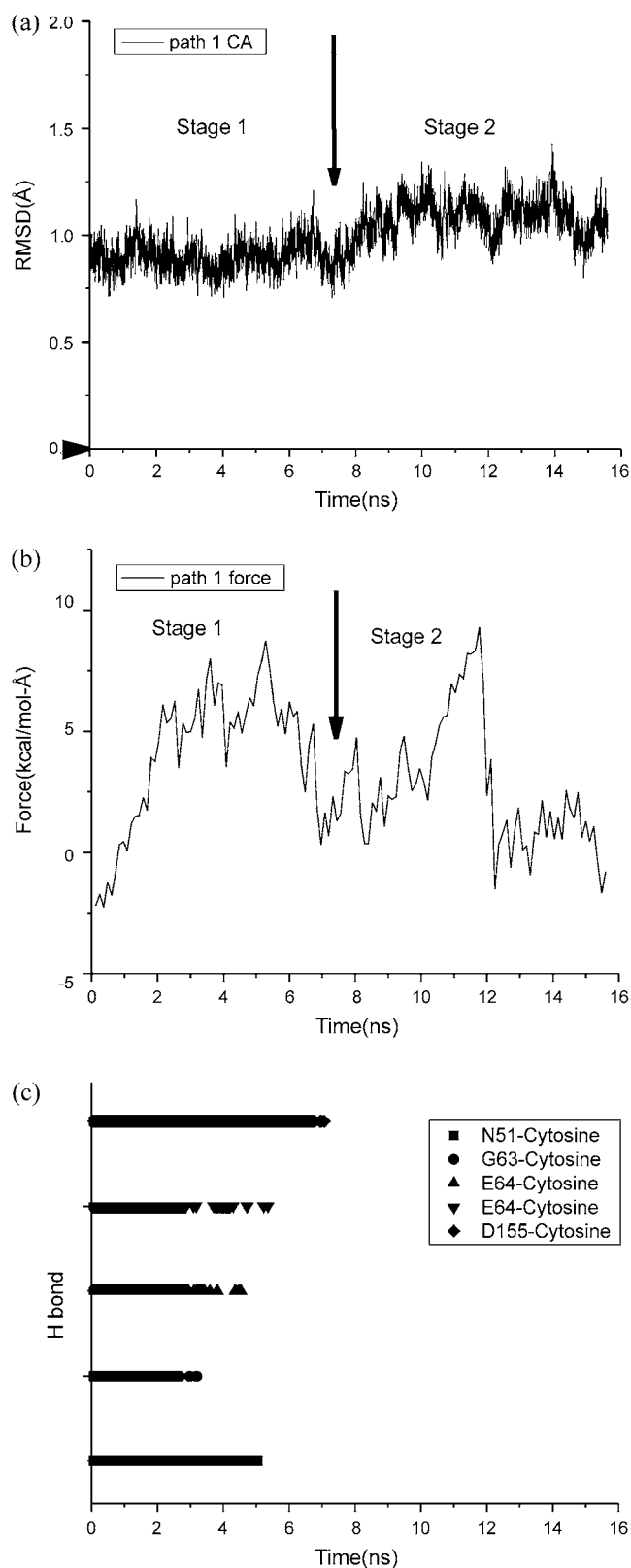


FIGURE 6 (a) Total CA RMSD versus time obtained by comparing the MD snapshots with the apo yCD crystal structure. (b) The window-averaged restraint force between cytosine and the protein in each window. (c) Hydrogen-bond lifetimes between cytosine and the protein during the cytosine pushing.

the second period, since all the original hydrogen bonds with cytosine have broken, the force needed to push the ligand out probably mainly comes from steric hindrance and conformational rearrangements of the residues along the path.

To understand how the cytosine-pushing procedure affects the dynamics of individual regions or residues, CA RMSFs are plotted in Fig. 7. Fig. 7*a* shows the CA RMSFs of the regular MD simulation. They are quite similar to those of the regular simulation of the apo enzyme at 300 K. Fig. 7, *b* and *c*, indicates that, during the pushing of cytosine, the protein overall becomes more flexible and the RMSF increases when the first and second stages of the pushing are compared with the regular simulation. The total fluctuations of the CA atoms are 35.8 \AA^2 , 50.6 \AA^2 , and 59.3 \AA^2 for the regular, stage 1, and stage 2 simulations, respectively. Since the active site remains buried during the regular simulation, the protein has to move certain regions or residues to open up the active site to let cytosine out, which causes the observed increase in fluctuation. Fig. 7, *b* and *c*, shows that these RMSF increases are mainly concentrated in the C-terminal helix and F-114 loop, which is consistent with the results of the 320 K apo form simulation (Fig. 1). The increases in the RMSFs of these two regions suggest that motions of these two parts are important for ligand release. Therefore, the motion of the protein required for ligand release is local, and the structural perturbation is small. It appears that the motion of the C-terminal helix is larger in stage 2 than that in stage 1. The MD trajectory shows that at the end of stage 2 cytosine moves out of the active site completely. The tip of the C-terminus moves close to the active site, partly to occupy the space vacated by cytosine and, consequently, blocks the active site, which effectively imparts to the C-terminal helix a relatively large motion. The unbinding process of cytosine can be illustrated by a schematic graph of the CA fluctuations. In a manner similar to the apo form simulation analysis, PCA was used to filter out the high frequency motions and only the first five modes were used to construct the schematic plot that is composed of the superposition of 50 MD snapshots. As shown in Fig. 8*a*, the whole protein is rather rigid, including the F-114 loop and C-terminal helix in the regular 2-ns MD simulation of the cytosine complex. Fig. 8*b* shows the increase of the fluctuation in these two regions in the first stage of pushing, and Fig. 8*c* describes a similar increase. Note that some of the snapshots show that the C-terminal helix moves closer to the active site in the second stage. A movie of the trajectory reveals that, in stage 2, the C-terminal helix and F-114 loop undergo a concerted flapping motion, which effectively opens up the active site and subsequently closes it after the release of cytosine (see discussion below). Besides the changes of these two regions, a small helix (residues ~ 75 – 80) also shows enhanced fluctuations during the pushing simulation (Fig. 7, *b* and *c*). Further inspection shows that this region is in close contact with the C-terminal helix of the adjacent protomer and suggests that this enhanced fluctuation of the small helix might be caused by the motion of that

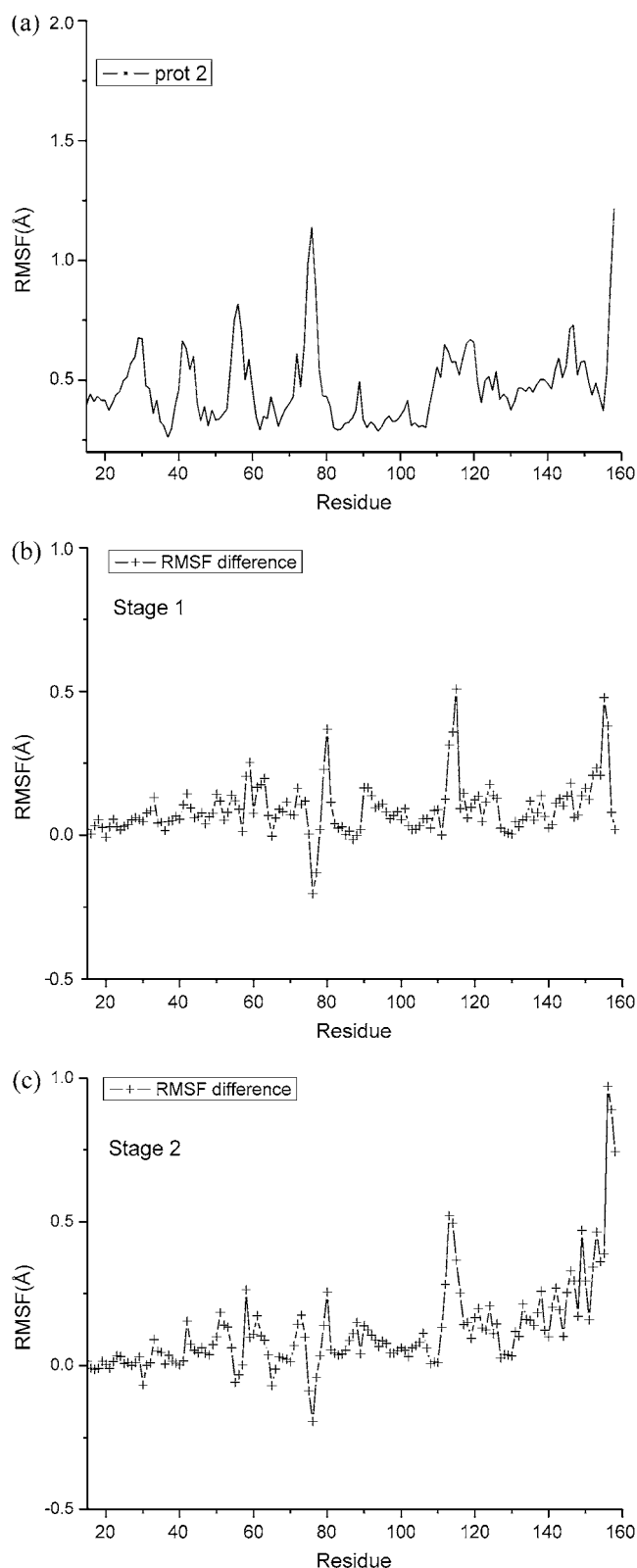


FIGURE 7 (a) CA RMSFs in 2-ns regular MD simulation. (b) RMSF difference obtained by comparing the first stage of cytosine pushing with the regular simulation. (c) RMSF difference obtained by comparing the second stage with the regular simulation.

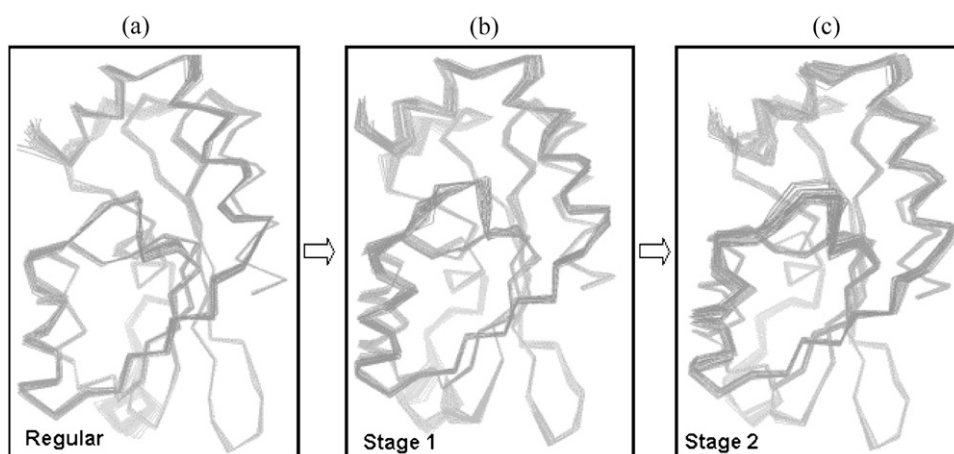


FIGURE 8 Schematic graphs of 50 trajectory snapshots of CA atoms projected out of the first five PCA motion modes: (a) Regular simulation. (b) Pushing simulation stage 1. (c) Pushing simulation stage 2.

C-terminal helix where cytosine was pushed out along path 2 simultaneously.

The MD trajectory shows that the cytosine release path is slightly different from the water diffusion path that was found in the apo form simulation. Unlike water diffusing through the triangle mouth defined by C-91, F-114, and I-156, the cytosine release path drifts slightly away to lie in between F-114 and I-156 (Fig. 9). The mass-weighted RMSDs of F-114 and I-156 are plotted in Fig. 10. The RMSD of F-114 changes dramatically during the pushing simulation but eventually returns to be close to its initial position. Compared with F-114, the motion of I-156 is smaller (Fig. 10) but has a similar pattern to the whole C-terminal helix, and at the end it also goes back to ~ 1 Å RMSD, indicating that this residue returns to its original conformation. Apparently, F-114 undergoes a large flapping motion during stage 2, which opens up the active site and assists the release of cytosine, whereas the C-terminal helix moves slightly away to make the entrance larger. Then both come back to cover the active site after the release of cytosine. It appears that the C-terminal helix moves as a whole block, whereas F-114 moves in addition to the loop movement, because there is no secondary structure restraint in the loop region. As a result, the motion scale of F-114 is much larger than that of I-156.

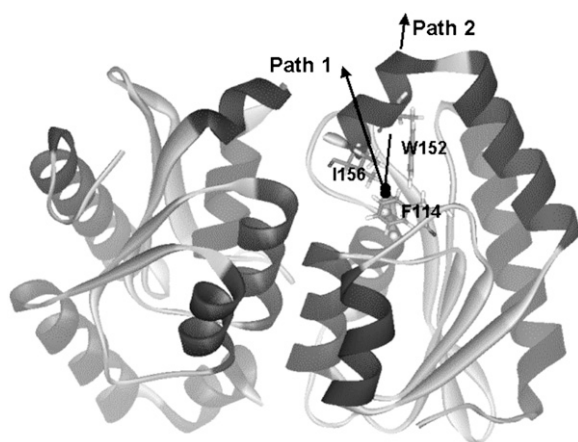


FIGURE 9 The cytosine release paths found from the pushing simulation.

Path 2

Cytosine in protomer 1 was pushed out along path 2, which is defined as in between the C-terminal helix and loop 1 (Fig. 9). According to the 320-K simulation of the apo enzyme, it seems that this path is much less popular for water to diffuse through than path 1. Fig. 11 *a* shows the RMSD along path 2 of all the CA atoms based on superposition with the crystal structure. It increases from ~ 0.8 Å to ~ 1.4 Å, which is larger than that of path 1, which increases from 0.9 Å to 1.1 Å, suggesting that the overall protein has to move more when cytosine releases through this path. This could make the release harder. The total CA fluctuation is 92.5 Å² (for protomer 2), much larger than that in the regular simulation (25 Å² (29 Å²) for protomer 1 (2)). The CA RMSF differences are presented in Fig. 11 *c*. It appears that many regions in the protein have larger fluctuations, especially those flexible regions identified in the regular 320 K apo form simulation (Fig. 1). Therefore, it seems that the protein has to rearrange many regions in a concerted manner when cytosine releases through this path, in contrast to

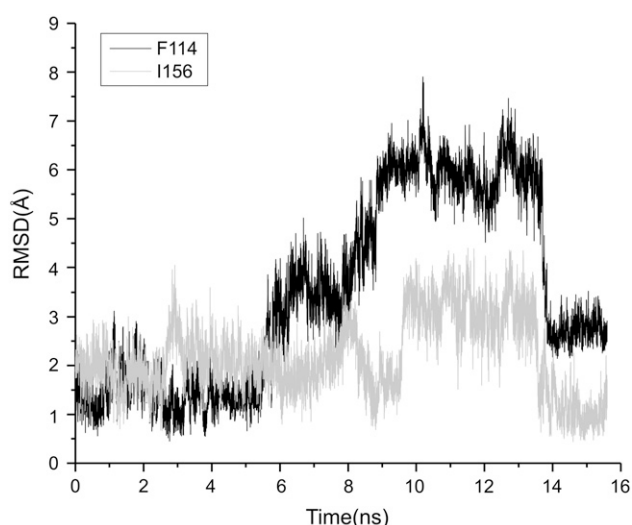


FIGURE 10 Mass-weighted RMSD plots of F-114 and I-156 along pushing path 1 of cytosine.

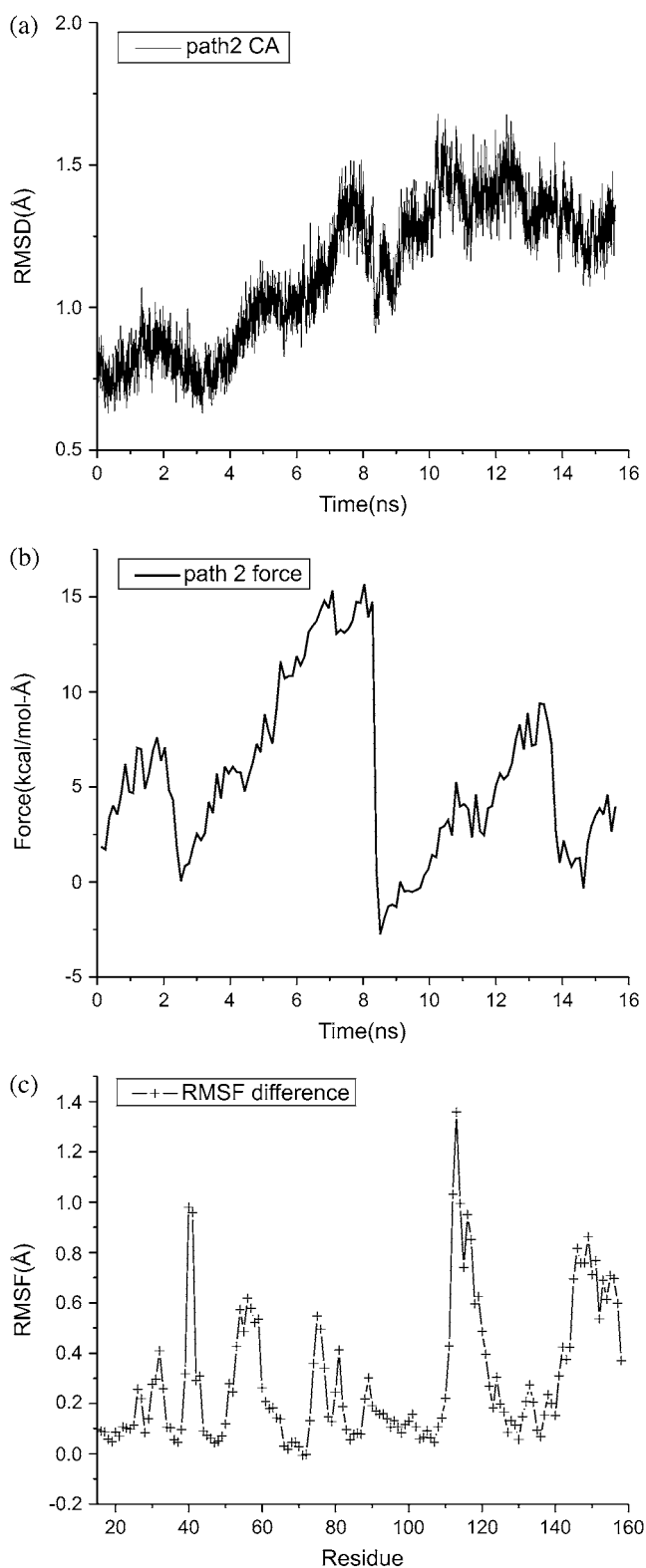


FIGURE 11 (a) RMSD of all CA atoms compared with the cytosine yCD crystal structure. (b) The window-averaged restraint force calculated for cytosine pushing along path 2. (c) The CA fluctuation differences between the regular and cytosine pushing simulations.

path 1 where only the F-114 loop and C-terminal helix regions need to move away and the perturbation is rather local and small. An analysis of the MD trajectory shows that along this path cytosine first pushes aside V-31, N-51, and D-155 and breaks the hydrogen bonds with the latter two residues and then moves away from D-151 (which is salt bridged with R-148). After that, it rearranges the side chain of F-54 to gain enough space to exit. It appears rather hard for a ligand to move through this path. The force used to push cytosine out is presented in Fig. 11 *b*. The average force is ~ 5.6 kcal/mol-Å, which is about two times larger than that in path 1 (~ 3.2 kcal/mol-Å), implying that more work has to be done in path 2. The root mean-square force fluctuation is also larger in path 2 (4.5 kcal/mol-Å) than path 1 (2.8 kcal/mol-Å), so the energy surface may be rougher along path 2.

Thus, from the energetic and structural points of view, path 2 is not favorable compared with path 1 in the current simulations. To release the ligand along path 2, many regions have to move, including the C-terminal helix and F-114 loop, which are the only parts required to move significantly when cytosine is released along path 1. In fact, the F-114 loop has the largest fluctuation changes in the path 2 simulation because cytosine pushes the C-terminal helix close to the F-114 loop and therefore moves this loop away.

It appears that in both pathways the C-terminal helix is very important, acting like a lid covering the active site. In both unbinding simulations (paths 1 and 2) the C-terminal helix moves as a whole block. The x-ray structure shows that this helix is negatively charged (-4 , including D-151, E-154, D-155, and E158), and there are four salt bridges with residues from other regions, including D-151-R-148, E-154-R-53, E-154-R-73* (from the adjacent protomer) and D-155-R-53, which restrain this helix. During the pushing simulation, all these salt bridges are well maintained.

CONCLUSIONS

In this study, we used MD to explore possible ligand release paths from yCD. The 300-K apo enzyme simulation maintains a closed active site, consistent with the crystal structure. Overall, the protein is quite rigid, a feature that has been confirmed by NMR backbone relaxation experimental data (L. Yao and H. Yan, unpublished). Finding exit pathways for ligands when a protein has a deep cleft leading to the active site can often be inferred by examination of crystal structures. For yCD, the active site is completely buried and the choice of potential paths by simple inspection of a crystal structure could be misleading. Here, the use of a higher temperature simulation was essential for identifying potential exit paths. The 8-ns 320-K simulation shows a strikingly large increase of flexibility in some regions of the protein while maintaining the secondary and tertiary structure of the protein. The increases of flexibility in the C-terminal helix and F-114 loop regions, which open up the closed active site, are most significant. Water molecules are found to diffuse into the active site

through two paths. One path is in between the F-114 loop and C-terminal helix, and the other is in between loop 1 and the C-terminal helix. The former path appears to be much more popular, with the mouth of the entrance defined by a triangular area formed by the residues C-91, F-114, and I-156. This area and its fluctuations are much larger at 320 K than at 300 K (Fig. 5), which allows water molecules to frequently move in and out of the active site.

Cytosine was pushed out of the active site along these two paths by 13 Å in a 15.6-ns simulation. The results show that in path 1 the required motion of the protein is quite local, only involving the C-terminal helix and F-114 loop. Two residues, F-114 and I-156, are identified that have to be moved away to let cytosine out; whereas in path 2, the protein has to rearrange itself quite globally, and the changes are also much larger compared to the path 1 simulation. The average force along the path and its fluctuation are larger in path 2 than in path 1, suggesting that the barrier is higher and the energy surface along the release path is rougher in path 2. Several residues have to be moved away including V-31, N-51, and D-155 first, and then D-151, and finally F-54. This path is used much less frequently for water molecules relative to path 1, and may be difficult for cytosine exit because of its larger size in comparison with water. However, path 2 cannot be excluded on the basis of the simulations alone, in view of the limited timescale available to MD simulations.

The simulations show that, for both paths, the C-terminal helix is critical for ligand release. Note that the C-terminal helix is well restrained by four salt bridges and several hydrogen bonds with residues in other regions. By weakening these salt bridge interactions, the flexibility of the C-terminal helix might be increased, and that could make cytosine release faster. Or, on the other hand, by mutating F-114 or I-156 (residues along the path) to smaller residues, ligand release also could be assisted. As discussed above, opening the active site limits the ligand escape process. From the experimental point of view, the protein modifications described above might be able to improve the product release rate and thereby enhance the protein's activity. To this end, experimental mutagenesis studies are under way. It will be interesting to see whether they will confirm our computational study.

This work was supported by the Quantitative Biology Modeling Initiative of Michigan State University, the Michigan Center for Biological Information of the Michigan Life Sciences Corridor, and National Institutes of Health grants (GM58221 to H.Y. and GM47274 to R.I.C.).

REFERENCES

- Nishiyama, T., Y. Kawamura, K. Kawamoto, H. Matsumura, N. Yamamoto, T. Ito, A. Ohya, T. Katsuragi, and T. Sakai. 1985. Antineoplastic effects in rats of 5-fluorocytosine in combination with cytosine deaminase capsules. *Cancer Res.* 45:1753–1761.
- Morris, S. M. 1993. The genetic toxicology of 5-fluoropyrimidines and 5-chlorouracil. *Mutat. Res.* 297:39–51.
- Ireton, G. C., M. E. Black, and B. L. Stoddard. 2003. The 1.14 Å crystal structure of yeast cytosine deaminase: evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. *Structure.* 11:961–972.
- Ko, T. P., J. J. Lin, C. Y. Hu, Y. H. Hsu, A. H. J. Wang, and S. H. Liaw. 2003. Crystal structure of yeast cytosine deaminase. Insights into enzyme mechanism and evolution. *J. Biol. Chem.* 278:19111–19117.
- Yao, L. S., S. Sklenak, H. G. Yan, and R. I. Cukier. 2005. A molecular dynamics exploration of the catalytic mechanism of yeast cytosine deaminase. *J. Phys. Chem. B.* 109:7500–7510.
- Yao, L. S., Y. Li, Y. Wu, A. Z. Liu, and H. G. Yan. 2005. Product release is rate-limiting in the activation of the prodrug 5-fluorocytosine by yeast cytosine deaminase. *Biochemistry.* 44:5940–5947.
- Frenkel, D., and B. Smit. 1996. *Understanding Molecular Simulation: From Algorithms to Applications.* Academic Press, San Diego, CA.
- Shen, L. L., J. H. Shen, X. M. Luo, F. Cheng, Y. C. Xu, K. X. Chen, E. Arnold, J. P. Ding, and H. L. Jiang. 2003. Steered molecular dynamics simulation on the binding of NNRTI to HIV-1 RT. *Biophys. J.* 84:3547–3563.
- Isralewitz, B., M. Gao, and K. Schulten. 2001. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* 11:224–230.
- Pearlman, D. A., D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, and P. Kollman. 1995. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91:1–41.
- Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsue, H. Gohlke, R. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, C. Singh, P. Weiner, and P. A. Kollman. 2002. AMBER7. University of California, San Francisco.
- Berendsen, H. H. C., J. P. M. Postma, W. F. Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
- Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.
- Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model. (Online).* 7:306–317.
- Xu, Y. C., J. H. Shen, X. M. Luo, I. Silman, J. L. Sussman, K. X. Chen, and H. L. Jiang. 2003. How does huperzine A enter and leave the binding gorge of acetylcholinesterase? Steered molecular dynamics simulations. *J. Am. Chem. Soc.* 125:11340–11349.
- Niu, C. Y., Y. C. Xu, Y. Xu, X. M. Luo, W. H. Duan, I. Silman, J. L. Sussman, W. L. Zhu, K. X. Chen, J. H. Shen, and H. L. Jiang. 2005. Dynamic mechanism of E2020 binding to acetylcholinesterase: a steered molecular dynamics simulation. *J. Phys. Chem. B.* 109:23730–23738.
- Zhang, D. Q., J. Gullingsrud, and J. A. McCammon. 2006. Potentials of mean force for acetylcholine unbinding from the alpha7 nicotinic acetylcholine receptor ligand-binding domain. *J. Am. Chem. Soc.* 128:3019–3026.
- Yao, L. S., H. G. Yan, and R. I. Cukier. 2006. Mechanism of dihydroneopterin aldolase: a molecular dynamics study of the apo enzyme and its product complex. *J. Phys. Chem. B.* 110:1443–1456.
- Gerini, M. F., D. Roccatano, E. Baciocchi, and A. Di Nola. 2003. Molecular dynamics simulations of lignin peroxidase in solution. *Biophys. J.* 84:3883–3893.
- Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins. Structure, Function, and Genetics.* 17:412–425.
- García, A. E. 1992. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68:2696–2699.
- Sklenak, S., L. Yao, R. I. Cukier, and H. Yan. 2004. Catalytic mechanism of yeast cytosine deaminase. An ONIOM computational study. *J. Am. Chem. Soc.* 126:14879–14889.
- McCammon, A., and S. C. Harvey. 1987. *Dynamics of Proteins and Nucleic Acids.* Cambridge University Press, Cambridge.