# Dynamics, Hydration, and Motional Averaging of a Loop-Gated Artificial Protein Cavity: The W191G Mutant of Cytochrome *c* Peroxidase in Water as Revealed by Molecular Dynamics Simulations†

Riccardo Baron*,‡,§ and J. Andrew McCammon‡,§,||,⊥

*Department of Chemistry and Biochemistry, Center for Theoretical Biological Physics, Department of Pharmacology, and Howard Hughes Medical Institute, University of California at San Diego, La Jolla, California 92093-0365*

*Received May 7, 2007; Revised Manuscript Received June 5, 2007*

ABSTRACT: Five molecular dynamics simulations of the W191G cavity mutant of cytochrome *c* peroxidase in explicit water reveal distinct dynamic and hydration behavior depending on the closed or open state of the flexible loop gating the cavity, the binding of ($K^+$ or small molecule) cations, and the system temperature. The conformational spaces sampled by the loop region and by the cavity significantly reduce upon binding. The largest ordering factor on water dynamics is the presence of the $K^+$ ion occupying the gated cavity. Considerable water exchange occurs for the open-gate cavity when no ligand or cation is bound. In all cases, good correspondence is found between the calculated (ensemble-averaged) location of water molecules and the water sites determined by X-ray crystallography experiments. However, our simulations suggest that these sites do not necessarily correspond to the presence of bound water molecules. In fact, individual water molecules may repeatedly exchange within the cavity volume yet occupy on average these water sites. Four major conclusions emerge. First, it seems misleading to interpret the conformation of protein loop regions in terms of single dominant structures. Second, our simulations support the general picture of Pro 190 *cis*−*trans* isomerization as a determinant of the loop-opening mechanism. Third, receptor flexibility is fundamental for ligand binding and molecular recognition, and our results suggest its importance for the docking of small compounds to the artificial cavity. Fourth, after validation against the available experimental data, molecular dynamics simulations can be used to characterize the dynamics and exchange of water molecules and ions, providing atomic level and time-dependent information otherwise inaccessible to experiments.

Dynamics and flexibility play a key (thermo)dynamic role in the function and stability of biological systems. The mobility of protein surface loops is a determinant of molecular recognition and association. Conformational gating is a fundamental mechanism used by enzymes to recruit specific substrates or reactive intermediates to their active site (*1*−*9*). A subtle stereochemical modification of surface amino acid configuration, for example, a *cis*−*trans* isomerization, may underlie very sophisticated machineries to regulate enzymatic activity (*10*) and to time biochemical pathways in living organisms (*11*).

Our current knowledge of these dynamical processes is commonly based on the comparison of time and/or ensemble-averaged structures as inferred by X-ray diffraction experiments on protein crystals or (solid and liquid phase) nuclear magnetic resonance (NMR) experiments (*12*). However, the (bio)physical interpretation of these experiments is often nontrivial even when including dynamical information (*13*−*17*).

Molecular dynamics (MD[1]) simulation with empirical force fields is a theoretical tool of steadily increasing importance to study the properties of (bio)molecular systems at atomic resolution (*17*, *18*). These calculations and their underlying models based on classical mechanics give fundamental insight into the behavior of complex macromolecules and help to understand the relationship among their structure, dynamics, and function. One of the major benefits of MD simulations is to complement experimental results by providing not only averages but also distributions and time series in principle of any physically definable observable. A remarkable example is the application of MD simulations based on explicit solvent models to follow the dynamics of water molecules and interpret ensemble-averaged experimental data concerning

† This work was supported, in part, by grants from the National Science Foundation (NSF) and the National Institutes of Health (NIH).

* Corresponding author. Phone: +1-858-534-2913. Fax: +1-858-534-4974. E-mail: rbaron@mccammon.ucsd.edu.

‡ Department of Chemistry and Biochemistry.

§ Center for Theoretical Biological Physics.

|| Department of Pharmacology.

⊥ Howard Hughes Medical Institute.

[1] Abbreviations: NMR, nuclear magnetic resonance; MD, molecular dynamics; W191G, Trp 191 to Gly cavity mutant of cytochrome-*c* peroxidase; **ref**, apo closed-gate; **bb**, complex with best binder; **open**, apo open-gate; **rst**, restrained backbone closed-gate; **rsto**, restrained-backbone open-gate; SPC, single point charge; PDB, protein data bank; RMSD, root-mean-square difference; RMSF, root-mean-square fluctuation; RGYR, radius of gyration; SASA, solvent-accessible surface area; nr. wt, total number of water molecules in the cavity; $n_{ex}$, total number of water exchanging events; nr. ex-wt, total number of water molecules exchanging; nr. nex-wt, total number of water molecules that never leave the cavity; $\tau_{max}$, maximum residence time of a specific water molecule inside the cavity.

10.1021/bi700866x CCC: $37.00 
Published on Web 08/24/2007

the hydration of protein surface and buried cavities (*19*−*27*).

The W191G cavity mutant of cytochrome *c* peroxidase from *Escerichia coli* (W191G) and the binding of heterocyclic cation ligands into its engineered cavity were characterized experimentally (*6*, *28*−*32*). Induced binding has been proposed for specific ligands of comparatively larger sizes (i.e., imidazo[1,2-*a*]pyridine, benzimidazole, and indoline), which would involve a ligand-gated, hinged loop rearrangement upon binding to the buried protein cavity (*6*, *32*). Removal of Trp 191 not only creates a ligand-binding cavity but also appears to introduce sufficient flexibility in residues 190−195 to enable the opening of a pathway for ligands to reach the buried cavity. Interestingly, X-ray crystallographic experiments elucidated the structures of several ligand−protein complexes, including those for which the loop rearrangement is more pronounced and causes a shift between closed and open-gate ensembles of structures. Particularly, benzimidazole was suggested to induce a full opening of the cavity, while two compounds (imidazo[1,2-*a*]pyridine and indoline) would induce partial opening in the surface loop (*6*, *32*). The trapping of the closed and open alternate states provides a unique view of the extreme (ensemble-averaged) stages of loop rearrangement, as revealed by superimposing the X-ray structures of the apo and W191G-benzimidazole complexes (see Figure 1). The flap movement possibly has two well-defined hinge points (Pro 190 and Asn 195) as previously observed for other protein-surface loops, both from native (*4*, *5*, *8*) and engineered (*33*) proteins. Kinetics and thermodynamic experiments suggest that the loop rearrangement is substantially dominated by the *cis*−*trans* isomerization of Pro 190 (*6*).

Most peptide bonds preferentially populate the *trans* isomer under unrestrained conditions, mainly because the amide hydrogen (*trans* isomer) offers less steric repulsion to the preceding $C^{\alpha}$ atom than does the following $C^{\alpha}$ atom (*cis* isomer). However, peptide bonds are able to populate both the *cis* and *trans* isomers (*34*), and this effect is more pronounced for bonds to Pro because the *cis* and *trans* isomers of the X-Pro peptide are nearly isosteric (*35*, *36*) (where X is the preceding amino acid). The fraction of X-Pro peptide bonds in the *cis* isomer under unstrained conditions is generally found in the range between 10 and 40% (*35*, *36*). Prolyl *cis*−*trans* isomerizations are known to be slow reactions that occur in the time range of minutes at standard conditions. Their importance as rate-determining steps in protein folding has been widely recognized on the basis of experiments (*37*, *38*) and computer simulations (*39*) on model systems. Yet, their (thermo)dynamic and kinetic role in conformational rearrangements for larger proteins is only partially understood. Particularly relevant for their biological implications are the change of thermodynamics and kinetics due to conformational restraining of peptide chains involving X-Pro bonds (*40*, *41*). Concerning protein stability, it is also unknown what determines the specific conformation of a prolyl bond in native and folded protein structures. Apparently, the coupled interactions with the three-dimensional fold are important because they move the *cis*−*trans* equilibrium toward peptidyl bonds in a *cis* conformation even when the original amino acid Pro is mutated to another amino acid (*42*). On the contrary, recent investigations of the side-chain effects on peptidyl-prolyl *cis*−*trans* isomerization for model oligopeptides suggest that the prolyl bond conformation in proteins is determined by local effects (*43*).

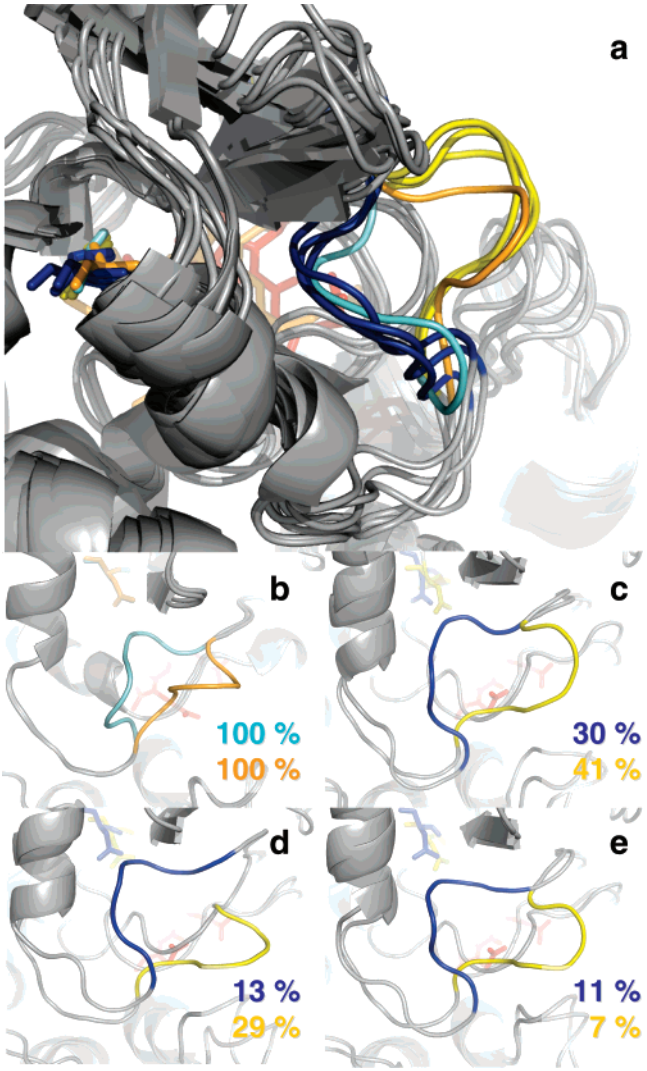


FIGURE 1: Dominant conformations found for the loop-gating region (residues 190−195) from X-ray crystallography experiments and **ref** or **open** simulations. (a) X-ray structures (*57*) in the closed (cyan tube) and open (orange tube) states are superimposed to the central members of the three most populated clusters of structures from **ref** (blue tube) and **open** (yellow tube) simulations. Enlarged images are shown separately with the same color code for the (b) X-ray structure and the central member of the (c) first, (d) second, and (e) third most populated clusters of structures. Populations are reported using the corresponding colors. An identical orientation criterion (superimposition of all backbone $C^{\alpha}$ atoms) was used for all structures in (a) or all structures in (b)−(e). See the Materials and Methods section for computational details and Table 1 for reference codes.

Similar activation free energy (44 kJ $mol^{-1}$), enthalpy (113 kJ $mol^{-1}$), and entropy (235 J $K^{-1}$ $mol^{-1}$) upon binding to the W191G cavity were measured for benzimidazole and 1,2-dimethylimidazole despite their differences in size and hydrophobicity (*6*). These two compounds may or may not (respectively) induce the opening of the cavity gate in solution. A similar activation enthalpy was attributed to *cis*−*trans* isomerization of a refolding protein (*44*). Changes upon binding of (solute plus solvent) free energy (in the range −30/−16 kJ $mol^{-1}$), enthalpy (in the range −96/−17 kJ

mol$^{-1}$), and entropy (in the range −230/17 J K$^{-1}$ mol$^{-1}$) have been reported for 13 additional ligands (*32*). However, the mechanism involved and the link between structural and thermodynamic changes is not clear because of (i) statistical averaging occurring over multiple pathways for the closed−open transition, as expected for large-scale protein motions involving many torsional degrees of freedom, (ii) the solvent contribution to changes of thermodynamic quantities, and (iii) possible physico-chemical differences between the force field experienced by a protein in solution vs a protein in a crystal. Experimental studies comparing the structural heterogeneity of protein loops in solution with corresponding X-ray crystal structures have been recently reported in the literature (*3*, *5*, *8*) and also address points (i) and (iii). Here, we characterize the shift in distributions for the key torsional angles involved in the loop rearrangement and the corresponding changes of cavity hydration, which are of fundamental importance for molecular docking.

The W191G artificial mutant protein works as a highly selective receptor for cations (*6*, *32*) comparatively to known artificial cavities (*45*). However, it is only partially known how physico-chemical factors, such as electrostatics, solvation, hydrophobicity, and hydrogen bonding, combine to make evolved natural binding sites distinguishable from more common surface clefts (*46*). X-ray experiments suggest the presence of five conserved water sites in the inner cavity of W191G. Depending on the compounds binding the cavity, some of the water molecules (two or more) are expelled, together with the $K^+$ cation occupying the cavity of the W191G apo structure (*32*). Yet, the characterization of cavity hydration in the two states may be limited by intrinsic motional averaging, resolution, and refinement constraints. Particularly, it is possible to imagine that the W191G engineered protein cavity may contain disordered water molecules (invisible to X-ray measurements), which remain favorably confined into the cavity volume because of hydrogen-bond interactions, the cooperative interaction with other disordered water molecules, and a large enthalpy−entropy compensation. This behavior has been described for hydrophobic cavities on the basis of X-ray and NMR experiments (*47*, *48*), and its thermodynamic interpretation is significantly different from that concerning water molecules bound to a protein also proposed for the W191G cavity. In the latter case, a large entropy loss characterizes the binding process of a water molecule to a protein (*49*), and large enthalpy−entropy compensation occurs upon ligand binding.

Prompted by the interest for the W195G system and its relevance as a test case for cavity complementation and ligand−protein binding, we investigated these aspects using computer simulations. The binding of small molecules to the W195G cavity has been employed as a test case for free energy calculations (*50*) and docking methods (*51*). However, no theoretical study focused on cavity dynamics, flexibility, and hydration has been reported to date.

Here, we present five MD simulations of the W191G system in explicit water at apparent pH 4.5: the apo reference system, the W191G in complex with the best binder currently known (i.e., 2-amino-5-methylthiazole), and the apo system in its open state conformation (at 300 K). Two additional simulations of the closed and open states of W191G have been performed at higher temperature (500 K) with the main-chain backbone atoms positionally restrained close to the experimental X-ray structures (see Table 1).

Table 1: Simulated Systems

| ref code | ref | bb | open | rst | rsto |
|---|---|---|---|---|---|
| PDB initial coordinates[a] | 1AA4 | 1AEN | 1RYC | 1AA4 | 1RYC |
| *T* [K] | 300 | 300 | 300 | 500 | 500 |
| restrained atoms | | | | $C^\alpha$ | $C^\alpha$ |
| ligand | $K^+$ | 2a5mt | | | |
| nr. of $K^+$ ions | 8 | 7 | 8 | 8 | 8 |
| total system charge | 0 | 0 | 0 | 0 | 0 |
| cofactor | heme | heme | heme | heme | heme |
| nr. of solute atoms | 3078 | 3089 | 3078 | 3078 | 3078 |
| nr. of solvent molecules | 16826 | 16315 | 17030 | 16826 | 17030 |
| nr. of solvent molecules initially in cavity,[b] $r_c = 0.5$ (nr. of which from crystallographic water sites)[a] | 5 (3) | 2 (2) | 5 (1) | 5 (3) | 5 (1) |
| nr. of solvent molecules initially in cavity,[b] $r_c = 0.75$ (nr. of which from crystallographic water sites)[a] | 11 (6) | 3 (3) | 9 (3) | 11 (6) | 9 (3) |
| nr. of atoms in the system | 53564 | 52042 | 54176 | 53564 | 54176 |
| equilibration period [ns] | 5.0 | 5.0 | 5.0 | 7.0 | 7.0 |
| equilibrium period [ns] | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 |

[a] From refs *6* and *32*. [b] After energy minimization to relax the solvent molecules around the protein.

The present article focuses on six major aspects of protein dynamics and hydration: (i) the stability of the MD simulations and their comparison with the available experimental data, (ii) the conformational space sampled by the loop region and its flexibility in the separate thermodynamic ensembles, (iii) the conformational sampling of the cavity and its changes upon ligand binding, (iv) the comparison of surface solvent accessibility between X-ray crystal structures and liquid-phase MD simulation ensembles, (v) the characterization of hydration, and (vi) the characterization of water dynamics and exchange inside the cavity and their changes upon ligand binding and gate opening. Additional analyses address the role of specific water molecules and their residence times, the behavior of ions, and the exchange of solvent between the cavity and bulk solvent depending on the conformation and temperature of the system.

## MATERIALS AND METHODS

*Molecular Model and Computational Procedures.* Trajectories for five different cytochrome *c* peroxidase W191G mutant systems in explicit water (**ref**, apo closed-gate; **bb**, complex with best binder; **open**, apo open-gate; **rst**, restrained backbone closed-gate; **rsto**, restrained-backbone open-gate) were generated using the GROMOS05 biomolecular software (*52*). The force field parameters and charges were chosen on the basis of the 45A4 GROMOS force field (*53*, *54*) (including improved bonded parameters for the heme cofactor) to reproduce the experimental conditions of apparent pH 4.5 (*6*, *32*). The GROMOS compatible SPC model for water (*55*) and the previously reported parameters for ions (*56*) were used in all simulations.

The initial configurations for W191G, the ligand, the ions (whenever available), and the crystallographic water sites were taken from the PDB (*6*, *57*) and solvated in (pre-equilibrated) rectangular boxes large enough to avoid any interactions between mirror images under periodic boundary conditions. Solvent molecules not identified by crystallographic water sites and in close contact with the solute were deleted if the minimum distance between non-hydrogen atoms fell below 0.23 nm. All systems were neutralized by

additional ions substituted to randomly chosen water molecules (but enforcing minimum reciprocal distances of 1.5 nm). A summary of the simulated systems is reported in Table 1.

A steepest-descent energy minimization was performed in order to relax the solvent and ion configuration, while the position of protein atoms and of the crystallographic-water oxygens were restrained by means of a harmonic potential (force constant of 2000 kJ $mol^{-1}$ $nm^{-2}$). Next a steepest-descent energy minimization was performed without restraints to eliminate any residual strain. The energy minimizations ended whenever the energy change per step became smaller than 0.5 kJ $mol^{-1}$. MD simulations were initialized from the energy-minimized configurations with atomic velocities taken from Maxwell−Boltzmann distributions at 50 K, while the position of protein atoms were restrained using a harmonic potential (force constant of 2000 kJ $mol^{-1}$ $nm^{-2}$). Each system was then gradually brought to the desired temperature (specified in Table 1) in 11 (**rst** and **rsto**, 19) consecutive 250 ps periods of simulations. During each period, the reference temperature was incremented by 25 K and the force constant decreased by 250 kJ $mol^{-1}$ $nm^{-2}$. The procedure was different for **rst** and **rsto** simulations: an initial restraining potential of 2000 kJ $mol^{-1}$ $nm^{-2}$ was enforced on all $C^{\alpha}$ backbone atoms during the minimization and heat-up phases, while no restraining was applied on the remaining protein atoms.

An initial equilibration period was performed for each system and extended for 5−7 ns (as specified in Table 1) to reach full equilibration of the energy components and of the system pressure (not shown). For system **open**, a restraining potential of 1500 kJ $mol^{-1}$ $nm^{-2}$ on all $C^{\alpha}$ backbone atoms was initially employed (and gradually removed) during the first 2 ns of the equilibration phase to keep the loop region (residues 190−195) in its open state. In all cases, only the 10 ns periods after the equilibration of each trajectory were used for analysis. For **rst** and **rsto** simulations, a restraining potential of 2500 kJ $mol^{-1}$ $nm^{-2}$ was maintained on all $C^{\alpha}$ backbone atoms during the entire MD runs.

Newton's equations of motion were integrated using the leap-frog algorithm (*58*) with a 2 fs time step. The SHAKE algorithm (*59*) was applied to constrain all bond lengths with a relative geometric tolerance of $10^{-4}$. The simulations were carried out in the N,p,T ensemble (at a pressure of 1 atm and reference temperatures as specified in Table 1) by separately coupling the temperature of solute and solvent degrees of freedom to a heat bath (*60*) (relaxation time 0.1 ps) and by coupling the pressure (estimated on the basis of an atomic virial) to a pressure bath (*60*) via isotropic coordinate scaling (relaxation time 0.5 ps; isothermal compressibility 4.574 $10^{-4}$ [kJ $mol^{-1}$ $nm^{-3}$]$^{-1}$). Non-bonded interactions were truncated at a distance of 1.4 nm and recalculated every time step in the range 0.0−0.8 nm and every five time steps in the range 0.8−1.4 nm, using a twin-range cutoff scheme (*61*). A reaction-field correction (*62*) was applied to account for the neglected interactions beyond 1.4 nm, using a relative dielectric permittivity of 61 for the SPC water model (*63*). A fast grid-based pairlist-construction algorithm (*64*) was employed (cell-mask edge of 0.4 nm; atomic-level cutoff) as implemented in the GROMOS05 MD++ module (*52*).

*Analysis Procedure.* Snapshots were extracted every 1 ps along each of the five equilibrated 10 ns MD trajectories and used for analysis.

Structural fitting was performed by (i) superimposing their centers of mass (to remove overall translation) and (ii) performing an atom-positional least-square fitting procedure (to remove overall rotation) using all $C^{\alpha}$ atoms (*65*), unless otherwise specified. The occurrence of secondary structure elements was monitored according to the definition by Kabsch and Sander (*66*).

The loop region includes six residues: Pro 190, Gly 191, Gly 192, Ala 193, Ala 194, and Asn 195. Two cavity definitions based on a molecular radius (i.e., a solvent molecule is considered to be inside the sphere whenever its center of mass fell within the cutoff distance) were employed for the analyses. The side-chain aliphatic carbon of Asp 235 was the reference center of the monitored properties in all cases because of its fundamental role in ligand binding (*29*, *32*).

First, for flexibility and properties of the residues near the cavity (i.e., torsional-angle normalized-probability distributions and radius of gyration of cavity atoms), the choice was made by including all cavity residues within a radius $r_c$ of 1.0 nm from the side-chain aliphatic carbon of Asp 235. These residues (His 175, Leu 177, Lys 179, Thr 180, Pro 190, Asn 195, Phe 202, Met 230, Met 231, Leu 232, and Asp 235) account for the properties of the cavity and of the surrounding environment. Pro 190 and Asn 195 are the hinge points for the loop-region rearrangement.

Second, in the case of water and ion properties, the cavity was defined by a sphere centered on Asp 235. Each of the five MD trajectories was filtered, selecting only the solvent molecules within 11 spheres of equally spaced radii $r_c$ in the range 0.4−0.9 nm. These 55 reduced trajectories were employed to calculate (i) nr. wt, the total number of water molecules in the cavity for each trajectory structure, (ii) $n_{ex}$, the total number of water exchanging events from or into the cavity, (iii) nr. ex-wt, the total number of water molecules exchanging, (iv) nr. nex-wt, the total number of water molecules that never leave the cavity, and (v) $\tau_{max}$, the maximum residence time of a specific water molecule inside the cavity. Quantity (i) describes the effective solvent occupancy of the cavity at a given simulation time. Quantity (ii) describes the total number of exchange events from the cavity to the external environment or vice versa. Quantity (iii) and (iv) describe the total number of (labeled) waters involved in the exchanging process between the cavity and the external environment (or vice versa) and the number of waters that reside in the cavity, respectively. The maximum residence times, $\tau_{max}$, describe the longest period a specific water molecule remains inside the cavity. Therefore our analysis of water dynamics captures all individual water-exchange events even when these do not contribute to a change in nr. wt (e.g., one water molecule leaving and another entering the cavity, simultaneously). The properties described and their dependence on $r_c$ can be used to characterize the disordered nature of water dynamics inside the cavity.

Trajectory structures were clustered into batches of similar configurations using the backbone atom-positional root-mean-square difference (RMSD) of residues 190−195 as

similarity criterion and the clustering algorithm previously described (*67*) and applied to flexible molecules (*14*, *67*−*69*). A cutoff of 0.1 nm was chosen after evaluation of the dependence of cluster populations against the total number of clusters found for each simulation. In this case, the structures were superimposed using only the $C^{\alpha}$ atoms of the residues belonging to the cavity (see above).

Solvent-accessible surface area (SASA) values were calculated using the program NACCESS (*70*) for individual protein structures after removing water and ion molecules. In the case of the **bb** ensemble, the ligand atoms were kept to include the desolvation effect due to ligand binding. Partial contributions to the total SASA value were calculated after dividing the contributions in five main classes of particles: (i) all atoms, (ii) polar side chains (all oxygens and nitrogens in the side chains), or (iii) non-polar side chains (all non-oxygens and non-nitrogens in the side chains), and (iv) side chains (all side-chain atoms) or (v) main chain (all backbone N, C, and O atoms). For this purpose, $C^{\alpha}$ atoms were classified as side-chain atoms in order to assign a side-chain solvent accessibility to Gly residues also and are therefore not included in the main chain group.

Hydrogen bonds were defined to have a maximum hydrogen-acceptor distance of 0.25 nm and a minimum donor−hydrogen acceptor angle of 135°. To identify multi-center hydrogen bonds, looser criteria for maximum distance (0.3 nm) and minimum angle (125°) were used. We note that so defined hydrogen-bond occurrences will discriminate hydrogen-bonding interactions formed by a specific donor group with distinguishable water molecules. Therefore, corresponding relative time occurrences for X-ray structures (Table S1, Supporting Information) should be interpreted as the sum over all possible hydrogen-bond relative occurrences to a given donor group by all exchanging water molecules. In the case of multi-center hydrogen bonds, the values from X-ray water sites should be interpreted as the sum of all occurrences to a given donor group.

The comparison of crystallographic water sites with the ensemble-averaged location of water in the simulations was based on the work by Schiffer and van Gunsteren (*21*). The radius of a water site was defined on the basis of the resolution of the corresponding X-ray structure (*57*). Ensemble-averaged hydration-density maps were calculated as described in the Results section using the VMD software (*71*).

Self-diffusion coefficients were calculated using the mean-square displacements of the water oxygens and the $K^+$ ions inside and outside the cavity. Corresponding averages $\langle D \rangle_{nex}$ over the non-exchanging water molecules (nex) and over the $K^+$ ions diffusing in bulk water $\langle D \rangle_{Kout}$ are also reported, together with their standard deviation from the mean values.

## RESULTS

*Stability of the Simulations.* To assess the stability of the five MD simulations, a collection of properties was monitored as a function of time. Figure S1 (Supporting Information) shows the time series of the formation of secondary structure elements according to the definition by Kabsch and Sander (*66*). The apo (**ref**), best binder−W191G complex (**bb**), and the open (**open**) simulations at 300 K are all stable along the pre-equilibrated 10 ns periods (see Table 1 for reference codes and simulation details). The different secondary structure elements remain well defined, and their relative occurrences are similar throughout the **ref**, **bb**, and **open** trajectories (3-helix, 2, 2, and 2%; 4-helix, 42, 41, and 40%; 5-helix, 0, 0, and 0%; bend, 19, 22, and 20%; beta-bridge, 2, 1, and 2%; beta-strand, 5, 5, and 5%; and turn, 7, 7, and 8%, respectively). An overall good agreement is found with corresponding values from experimental X-ray structures (*57*) (3-helix, 5, 6, and 6%; 4-helix, 45, 47, and 47%; 5-helix, 0, 0, and 0%; bend, 15, 13, and 12%; beta-bridge, 3, 2, and 3%; beta-strand, 4, 5, and 4%; and turn, 9, 8, and 8%, based on 1AA4, 1AEN, and 1RYC, respectively). However, we note comparatively higher formation of bend regions on the basis of our MD simulations. This discrepancy can be explained considering that our simulations concern proteins in the liquid phase, while X-ray measurements are carried out in the presence of crystal-packing forces. In the case of **rst** and **rsto** simulations, the same secondary structure elements are stable along the 10 ns periods, albeit the relative average formation of 4-helix elements decreases (18 and 19.2%, respectively) and the formation of bend elements increases (27 and 22%, respectively), as a consequence of the higher thermal motion (not shown).

We note that employing high-temperature simulations with positional restraining potentials on the backbone atoms can enhance the conformational sampling of the protein side chains but can result as well in arbitrary changes of the secondary structure element formation with respect to the X-ray structure at standard conditions. Therefore, the analyses of **rst** and **rsto** simulations will be limited in the following to cavity residues and water dynamics.

The backbone $C^{\alpha}$ atom-positional root-mean-square deviation (RMSD) between the trajectory structures and the corresponding X-ray references (*57*) are of small magnitude overall and remain stable along the simulation time, showing that a good agreement with experiment is present in the simulated trajectories. The backbone $C^{\alpha}$ atom-positional root-mean-square fluctuation (RMSF) in the **ref** and **open** trajectories displays peaks of comparable intensities, and the most flexible regions of the proteins sample their configurational spaces to a similar extent. The region of residues 185−195 shows a pronounced mobility, comparatively the lowest for the **bb** and the highest for the **open** simulations. The bend regions of residues 50−65 and 125−140 are also highly mobile, and a reduced flexibility is again systematically observed for the **bb** simulation (compare Figure S1a, b, and c, Supporting Information). Interestingly, a similar observation holds for all of the peaks in the region 140−205, largely mapping to the cavity residues. These results are consistent with a general good stability of the simulations and with a change of backbone flexibility upon ligand binding.

The stability of the heme cofactor was also monitored for all five simulations by calculating the distance of the Fe atom from the reference residue Asp235 and the RMSF of all heme atoms (see Figure S2, Supporting Information). In all cases, the results show that the cofactor is stably fluctuating behind the cavity region, the Fe atom is correctly coordinated, and the flexibility is reasonably distributed among the cofactor atoms and similarly along different trajectories.

Correspondence between MD simulation and X-ray experiments (*28*, *29*, *32*) is found also with respect to the behavior of the $K^+$ ions, as inferred from the time series of

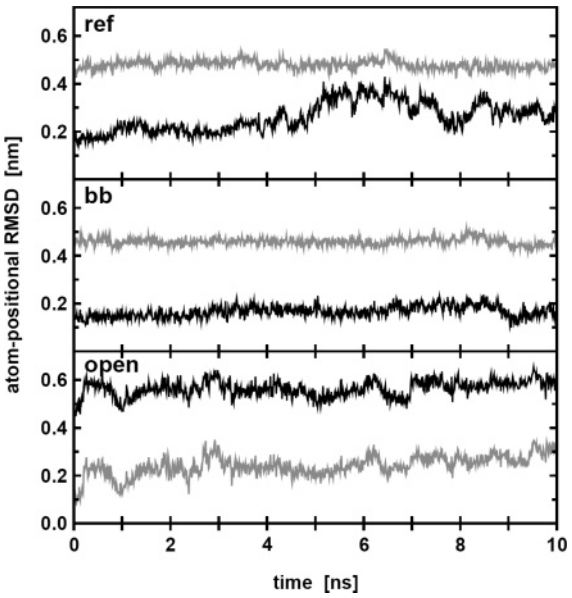


FIGURE 2: Time series of the backbone $C^{\alpha}$ atom-positional root-mean-square deviation (RMSD) of the loop region of residues 190−195 of **ref**, **bb**, and **open** trajectory structures from closed (black lines) and open (gray lines) X-ray crystallography structures (*57*). All structures were superimposed using all backbone $C^{\alpha}$ atoms of W191G. See Table 1 for reference codes.

the ion−cavity distances (not shown) and their distributions (see Figure S3, Supporting Information). The $K^+$ ion binding the apo cavity remains inside the W191G cavity during the entire **ref** simulation and shows (i) an average coordination number of 3.2 with surrounding water oxygens and (ii) a stably fluctuating distance from Asp 235 (not shown). None of the 7 $K^+$ ions additionally present in solution ever enters the cavity. Consistently, not one of the 7 or 8 $K^+$ ions reaches the buried cavity volume throughout the **bb** and **open** simulations, in agreement with the experimental observation of competitive binding for small (positively charged) molecules to the W191G cavity (*6*, *32*). Interestingly, tens of ion exchanges are observed in **rst** and **rsto** simulations showing that an increment of temperature from 300 to 500 K enhances the exchange of ions between the bulk solvent and the cavity.

*Conformational Spaces of the Gating Loop.* Figure 2 shows the loop region atom-positional RMSD time series calculated for our **ref**, **bb**, and **open** MD simulations alternatively using the corresponding closed or open X-ray crystallography structures (*57*). Significantly larger deviations are observed on the basis of the X-ray open reference structure (PDB ID: 1RYC) with respect to values calculated from the X-ray closed reference structures (PDB ID: 1AA4 and 1AEN, respectively) throughout the **ref** and **bb** trajectories. The picture inverts for the RMSD time series calculated between **open** trajectory structures and closed or open reference structures (PDB ID: 1AA4 and 1RYC, respectively). The trajectory structures of the **open** simulation are characterized by open-loop configurations during the entire (pre-equilibrated) 10 ns period, suggesting general stability for this thermodynamic state on the nanosecond time scale. This observation based on the open-gate apo ensemble of structures supports previous suggestions concerning the slow kinetics of *cis*−*trans* isomerization for Pro 190 (*6*).

Figure 1 shows the dominant conformations adopted by the loop region and the corresponding (ensemble-averaged) configurations as determined by X-ray crystallography experiments. Trajectory structures were clustered into batches of similar configurations using the RMSD of the loop region 190−195 as a similarity criterion (see the Materials and Methods section). For both **ref** and **open** simulations, the central members of the three most populated clusters of structures are displayed. They represent 30 and 41%, 13 and 29%, 11 and 7% for the first, second, and third clusters, respectively. In the case of the ligand-bound W191G, corresponding values of 92, 6, and 1% are found (**bb**, not shown). In all cases, the three most populated central member structures have low backbone $C^{\alpha}$ atom-positional RMSD values from the X-ray models (within 0.2 nm). For the **ref** and **open** simulations, there are 19 or 14 clusters populated by at least 1%, totaling 94 or 97% of the entire ensemble, respectively. For the **bb** simulation, we find a significantly smaller number of 3 clusters populated by at least 1%, totaling 99% of the entire ensemble. Interestingly, the reduced diversity of loop-region conformations for **bb** versus **ref** and **open** simulations reflects a sizable reduction of flexibility when W191G is bound to 2-amino-5-methylthiazole.

We note that such considerations would not emerge by a simple comparison of RMSD time series, although a comparatively reduced configurational sampling of the loop region for the 2-amino-5-methylthiazole−W191G complex matches the more stable RMSD values found for **bb** versus **ref** and **open** trajectories (see Figure 2b vs 2a and c). Additionally, we stress that a cluster analysis such as the one presented herein may supply useful information to characterize the flexibility of mobile regions in proteins, which is generally problematic on the basis of ensemble-averaged structures only (*72*, *73*).

Figure S4 (Supporting Information) displays the backbone torsional angle normalized probability distributions characterizing the loop-gating mechanism around residues Pro 190 and 195 ($\omega$, Pro 190: $C^{\alpha}-C-N-C^{\alpha,\ ring}$; $\phi$, Asn 195: $C-N-C^{\alpha}-C$; $\psi$, Asn 195: $N-C^{\alpha}-C-N$). For **ref, bb**, and **open** simulations at 300 K, these distributions are compared to the corresponding experimental values based on X-ray structure (Figure S4, Supporting Information, vertical dashed lines). The distributions calculated from MD simulations systematically overlap with the experimental values from X-ray crystal structures. The dihedral angle $\psi$ displays a comparatively broader distribution based on the **ref** simulation and additionally accesses the dihedral space around 120°. Interestingly, an evident shift of about 180° is observed when comparing the $\omega$ dihedral angle distributions of closed (**ref** and **bb**) and open (**open**) gate ensembles because of the Pro 190 *cis*−*trans* isomerization. The $\phi$ and $\psi$ backbone conformational sampling for Pro 190 is shifted toward the helical region with respect to unrestricted backbone chains (not shown). A similar effect upon backbone constraining was previously reported on the basis of computer simulations of model systems (*41*). The corresponding distributions for $\phi$ and $\psi$ of Asn 195 shift by about 40 and 150 upon opening the gating loop, respectively. Similar effects upon loop opening can be observed from **rst** and **rsto** simulations at 500 K (see Figure S4, Supporting Information). Remarkably, the loop can also access the $\psi$ dihedral space of Asn 195, characteristic of the open-loop configuration in the absence of a charged ligand inside the W191G cavity (Figure S4,

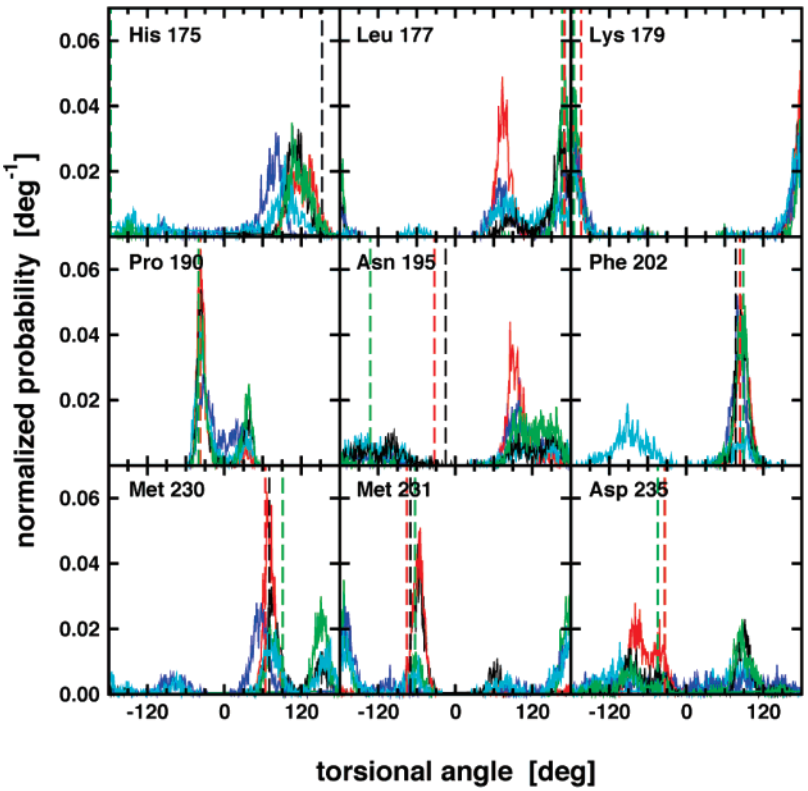


FIGURE 3: Side-chain torsional-angle normalized-probability distributions characterizing the W191G cavity. The side-chain torsional angles $\chi$ ($C^{\alpha}-C^{\beta}-X^{\gamma}-X^{\delta}$; with X being either a C, N, or O atom depending on the amino acid) of His 175, Leu 177, Lys 179, Pro 190, Asn 195, Phe 202, Met 230, Met 231, and Asp 235 are shown as calculated from **ref** (black), **bb** (red), **open** (green), **rst** (blue), and **rsto** (cyan) simulations, respectively. The vertical dashed lines indicate the dihedral-angle values based on the corresponding X-ray structures (black, 1AA4; red, 1AEN; green, 1RYC, (*57*)). The effect of a higher simulation temperature can be quantified by comparing **ref** (300 K) with **rst** (500 K) and **open** (300 K) with **rsto** (500 K) simulation ensembles, for the closed and open-gate configurations, respectively. The effect of ligand binding can be quantified by comparing **ref** with **bb**. See Table 1 for reference codes.

Supporting Information, bottom panel, **ref** vs **bb** ensembles).

These results show general agreement with the considerations inferred from ensemble-averaged X-ray structures (*6*), provide additional insight on the loop-opening mechanism, and support the general picture of Pro 190 *cis*−*trans* isomerization as a determinant of the loop-opening mechanism. The marked shifts of dihedral distributions suggest that the dihedral transitions of the hinge residues Pro 190 and Asn 195 drive the loop conformational change. The different sampling of the $\psi$ dihedral spaces in the **ref** and **bb** simulations is in line with the general observation of Pro *cis*−*trans* isomerization in folded proteins being governed also by cooperative (*42*) (long-range) as opposed to only local (*43*) (short range) interactions.

*Conformational Spaces of the Cavity*. Figure 3 shows the side-chain torsional angle normalized probability distributions for key residues inside the W191G cavity region, as calculated from our five MD simulations in the liquid phase and on the basis of the three corresponding X-ray crystal structures (*57*). Different sampling behavior characterizes the cavity residues considered. Yet, a few general trends can be defined concerning the effects due to ligand binding and temperature.

All calculated distributions systematically overlap with the corresponding experimental values, with the exception of Asn 195. However, the latter residue spans the widest range of values among the three different X-ray crystal structures. This shows that characterizing the flexible nature of an amino acid side chain may be difficult on the basis of ensemble-averaged crystal structures only. Yet, it can be revealed by the corresponding conformational heterogeneity observed among different crystals (whenever available) or from a sufficient sampling based on MD simulations. Lys 179 and Pro 190 display dihedral angle distributions, which are similar among the different simulations. The Pro 190 side chain shows two-peak distributions, which additionally sample the dihedral region around 30° in all simulations. The side chain of Leu 177, Asn 195, Met 231, and Asp 235 show comparatively narrower distributions in the presence of the ligand 2-amino-5-methylthiazole in the cavity (**bb**). His 175, Leu 177, Phe 202, Met 230, Met 231, and Asp 235 display significantly broader distributions upon increasing the temperature to 500 K (**rst** and **rsto**). Of special interest is the behavior of the negatively charged side chain of Asp 235, which defines the key electrostatic interactions and the geometry for cation ligand binding (*29*, *32*). We observe a pronounced reduction of the configurational spaces accessed upon ligand binding. Instead, the Asp 235 side chain additionally samples the dihedral angle values around 90° for the closed and open-loop ensembles of structures at both temperatures (**ref**, **open**, **rst**, and **rsto**), which are not revealed on the basis of X-ray crystal structures (see Figure 3).

The results presented in this section suggest that small compounds docking to an artificial cavity may experience significantly different interactions depending on (i) the X-ray crystal structure considered (e.g., apo, complex, or open-

Table 2: Comparison of Experimental and Calculated Ensemble Averaged Solvent-Accessible Surface Areas[a]

| | < SASA > [nm$^2$] | | | | |
|---|---|---|---|---|---|
| ref code | all | polar | non-polar | side chains | main |
| | | X-ray structure | | | |
| 1AA4 | 129 | 55 (43) | 74 (57) | 111 (86) | 18 (14) |
| 1AEN | 129 | 54 (42) | 75 (58) | 111 (86) | 18 (14) |
| 1RYC | 132 | 55 (42) | 77 (58) | 113 (86) | 19 (14) |
| | | MD simulation | | | |
| ref | 138 | 65 (47) | 73 (53) | 120 (87) | 18 (13) |
| bb | 131 | 61 (47) | 70 (53) | 114 (87) | 17 (13) |
| open | 148 | 73 (49) | 75 (51) | 129 (87) | 19 (13) |
| rst | 169 | 72 (43) | 97 (57) | 144 (85) | 25 (15) |
| rsto | 256 | 100 (39) | 156 (61) | 215 (84) | 41 (16) |

[a] The values calculated using *all*, *polar* or *non-polar*, and *side chain* or *main chain* atom subgroups are reported separately. Relative percent values are given in parentheses (with respect to *all* values). See the Materials and Methods section for computational details and the definitions of the subgroups. X-ray structures are as in ref *57*.

gate configurations), (ii) the physico-chemical nature of the ensemble of structures (e.g., one ensemble-averaged X-ray crystal structure or an ensemble of structures from MD simulations in the liquid phase), and (iii) the temperature of the system. These effects are determinants of ligand-binding thermodynamics and result in significantly different configurational entropies for the cavity residues involved in ligand binding (not shown). Yet, longer simulation periods are required to fully address entropy contributions and are currently being pursued.

*Surface Solvent Accessibility.* Figure S5a (Supporting Information) shows the time series of the solvent-accessible surface area (SASA) values during our five MD trajectories. The fluctuations are relatively small (maximum standard deviation from the mean values of 2.8 nm$^2$ observed for **rsto**) in agreement with the stability of W191G throughout the equilibrated 10 ns periods (see above).

Table 2 reports the SASA values based on single X-ray crystal structures and the ensemble average values based on the five MD simulations. Systematically smaller values characterize the reference X-ray crystal structures (Table 2, second column). A comparatively lower value is expected and found for the **bb** versus **ref** simulation (131 vs 138 nm$^2$) as a consequence of ligand binding. Instead, a larger average value is expected and found for the **open** versus **ref** simulation (148 vs 138 nm$^2$), upon opening of the loop gating the cavity. Comparatively larger increments are evident for **rst** and **rsto** ensembles of structures upon increasing the temperature (while keeping the backbone C$^\alpha$ positionally restrained close to the initial X-ray structure).

The origin of a different solvent accessibility between X-ray crystal structures and explicit solvent MD simulations can be detailed considering the partial contributions to the overall (all) SASA values separately based on the polar or non-polar and the side chains or main chain subgroups of atoms only (see Table 2 and the Materials and Methods section). Considering either the polar or the side chains subgroups, systematically larger average values are found for the **ref**, **bb**, and **open** trajectories with respect to the corresponding values based on X-ray structures (polar: 65 vs 55, 61 vs 54, 73 vs 55 nm$^2$, respectively; side chains: 120 vs 111, 114 vs 111, 129 vs 113, respectively). However, considering either the non-polar or the main chain subgroups, systematically similar values are found for the same trajectories structures with respect to the X-ray structures (non-polar: 73 vs 74, 70 vs 75, 75 vs 77, respectively; main: 18 vs 18, 17 vs 18, 19 vs 19). The major deviations seem to be dominated by the polar and side chain subgroups compared to non-polar and main chain subgroups. The side chains are the determinants of SASA value changes for W191G, which show stable backbone C$^\alpha$ atom positional RMSD time series (Figure S1, Supporting Information). These trends can be explained considering that crystal-packing forces restrict the motion in the protein, especially for those surface regions mainly contributing to the solvent-accessible surface. Instead, MD simulations in explicit solvent more closely reproduce the physical phase of proteins in solution.

In order to correlate the SASA values for the different ensembles with the plasticity of the cavity, the radius of gyration (RGYR) for the cavity residues only was calculated throughout the MD trajectories (Figure S5b, Supporting Information). These values are stable along the 10 ns periods (maximum standard deviation of 0.02 nm observed for **rsto**). Interestingly, SASA ensemble averages (Table 2) change accordingly to RGYR ensemble averages (**ref**, 0.86 nm; **bb**, 0.82 nm; **open**, 0.89 nm; **rst**, 1.01; **rsto**, 1.22), suggesting that solvent accessibility depends on the conformational sampling of the cavity residue side chains (see also Figure S5, Supporting Information).

*Ensemble Averaged Properties of Cavity Water.* In this section, we describe a number of analyses based on our five explicit solvent MD simulations to characterize the hydration in the W191G artificial cavity depending on the gating state of the cavity, ligand binding, and system temperature.

Table S1 (Supporting Information) reports the solute−solvent hydrogen bonds to specific (labeled) water molecules in the cavity region of W191G occurring for at least 5% of the time throughout the 10 ns pre-equilibrated trajectory periods. Only a reduced number of water molecules fulfill the (distance and geometry) criteria for (two or three centers) hydrogen bonding to protein cavity residues at 300 K. No hydrogen bonds are maintained by specific water molecules to a cavity residue for at least 5% of the MD trajectory period in the case of the simulations **rst** and **rsto** at 500 K.

The comparison between X-ray crystallographic water sites and the water density along the simulations can be related to the structural properties of the W191G cavity based on ensemble-averaged hydration-density maps (Figure 4). These maps describe isosurfaces of arbitrary values of average water density along the MD trajectories. The water sites detected by X-ray diffraction experiments (*57*) systematically overlap onto calculated high-density regions within the approximations of Debye−Weller factors commonly employed in crystallographic refinement, which assume fully harmonic motions of biomolecules in crystals (*73*). This demonstrates an excellent correspondence between the location of crystallographic water sites and the average presence of water in the simulations. The same correspondence was verified by considering the water density within spheres of radius corresponding to the resolution of the crystallographic model (*21*). We stress that this correspondence does not imply the presence of permanently bound waters in the cavity. On the basis of **ref** and **bb** simulations, we observe that water molecules can exchange repeatedly among the different high-density regions. This effect is more pronounced for the

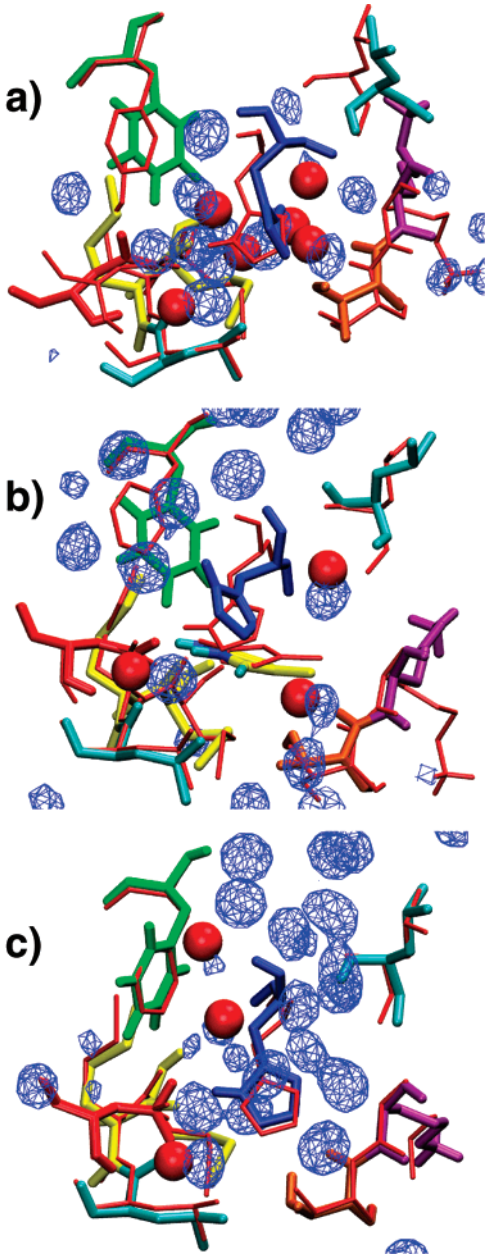


FIGURE 4: Ensemble-averaged hydration-density maps represented as the number of times the oxygen atom of any water molecule falls within a (0.05 $nm^3$) grid element throughout the (a) **ref**, (b) **bb**, or (c) **open** simulations. The blue isosurface defines volume regions with 200 hits per grid element, i.e., about 100 times the hydration density of bulk water. Corresponding crystallographic water sites are represented with red spheres with diameters corresponding to the X-ray structure resolution (57). Cavity residues from the starting structure of the equilibrated trajectories are displayed using the following color coding: His 175 (blue), Leu 177 (cyan), Lys 179 (purple), Thr 180 (orange), Phe 202 (green), Met 230 (yellow), Met 231 (yellow), Leu 232 (cyan), and Asp 235 (red). Their reference orientation based on corresponding X-ray crystal structures are shown as thin red sticks (superimposed using the $C^{\alpha}$-atoms of the cavity residues only). In the case of **bb**, the 2-amino-5-methylthiazole ligand is represented from the last MD snapshot (thick sitcks) and on the experimental structure (thin red sticks). For graphical purposes, slightly different orientations are used in (a), (b), and (c).

disordered water molecules in the **open** ensemble.

Figure S6 (Supporting Information) shows the structural properties of water based on water−oxygen radial distribution functions centered on the aliphatic carbon atom of Asp 235 and calculated from either (i) the entire period or (ii) only the last 2 ns of each MD trajectory. The corresponding integrals $\xi$ describe the average number of water molecules at a given distance from Asp 235. The curves based on (i) or (ii) overlap, showing that cavity water−oxygen positions are reasonably well equilibrated after about 5−7 ns of MD simulation (see Table 1). However, deviations are noticed for the **open** simulation, suggesting that the solvent molecules occupying an open W191G cavity are rather disordered (see below). The **ref** cavity evidences stronger ordering than the **bb** and **open** ones, concerning the first solvation shell. First peaks of similar intensity but broader shape are observed for **rst** and **rsto** simulations at 500 K, with respect to corresponding simulations **ref** and **open** at 300 K.

The radial distribution functions also capture the dependence of the solvent structure on the loop-gating conformational state considered (see Figure S6, Supporting Information, distance range 0.4−0.9 nm). The lower intensities based on **ref** and **bb** simulations (and the lower values for the $\xi$ integrals) account for the desolvation effect induced by the gating loop in its closed state. For **ref**, **bb**, and **open** simulations, $\xi$ integrals display minima at a distance of about 0.9 nm, corresponding to the average distance between Asp 235 and the gating loop. Interestingly, none of the $\xi$ curves vanishes in correspondence to the gating-loop region, demonstrating the average presence of a few water molecules at a distance corresponding to the gating loop. This observation agrees with our results concerning water dynamics and exchange between the cavity and the bulk solvent for all simulations (see below).

Figure 5a summarizes the results of hydration in the W191G cavity depending on the cavity depth. The results are reported on the basis of a variable cavity radius to capture the behavior of the water molecules in close contact with Asp 235 (small $r_c$ values) in the volume region in contact with bulk solvent (large $r_c$ values). A higher occupancy is expected and found for the apo versus holo W191G cavity, which accommodates on average about 5 more water molecules ($\langle$nr. wt$\rangle$ at $r_c = 0.75$ nm, **ref**, 10.0 vs **bb**, 5.2). Upon opening of the cavity gate, about 4 additional water molecules are located on average within the cavity volume ($\langle$nr. wt$\rangle$ at $r_c = 0.75$ nm, **ref**, 10.0 vs **open**, 13.8). Significantly larger average occupancies (and corresponding fluctuations) are found for simulations at 500 K ($\langle$nr. wt$\rangle$ at $r_c = 0.75$ nm, **rst**, 19.3 ± 3.0 and **rsto**, 28.2 ± 4.5; not shown).

The following observations emerge by comparison with cavity flexibility and surface solvent accessibility. First, the average number of water molecules in the cavity correlates well with both RGYR values for the cavity and the overall SASA values (compare Figure 5 with Figure S5a and b, Supporting Information), demonstrating that the W191G cavity accounts for most of the overall solvent-accessible surface area and that its volume determines the number of water molecules that can on average reside therein. Second, the same values do not correlate directly with the intensity of the first peak in the water−oxygen radial distribution functions (Figure S6, Supporting Information), and larger water occupancies are found for **ref** (closed gate; $K^+$ bound into the cavity) versus **open** (open gate; no $K^+$ inside the cavity) states for the cavity volume closest to Asp

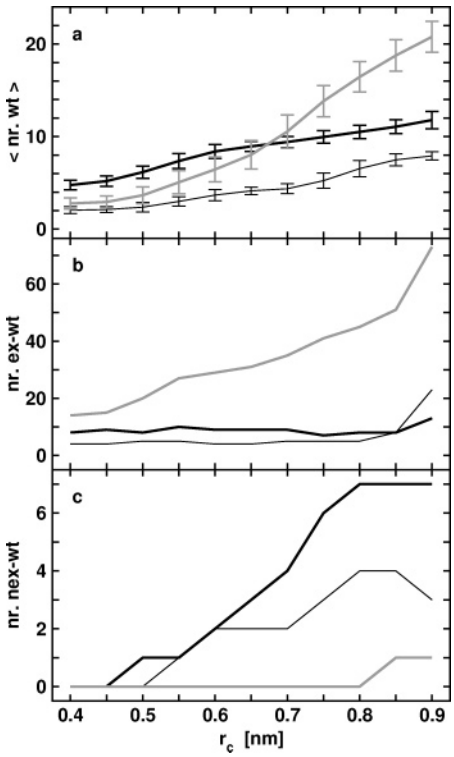


FIGURE 5: Analysis of water dynamics in the W191G cavity of cytochrome *c* peroxidase depending on the cavity radius ($r_c$) based on **ref** (black thick lines), **bb** (black thin lines), and **open** (gray thick lines) molecular dynamics simulations. (a) ⟨nr. wt⟩, ensemble average of the total number of water molecules inside the cavity (and their standard deviations as vertical bars). (b) nr. ex-wt, total number of exchanging water molecules. (c) nr. nex-wt, total number of never-exchanging water molecules. See Table 1 for reference codes and the Materials and Methods section for computational details.

235 (Figure 5a, $r_c$ < 0.65 nm). The ordering electrostatic interactions formed among cavity water molecules, polar cavity residues, and the $K^+$ ion binding (see Figure S3, Supporting Information) are in line with the experimental description of a highly ordered network of interactions formed by the $K^+$ and $Na^+$ cations and the surrounding coordinated water molecules inside the cavity of the W191G apo structure (*28*, *29*, *32*).

*Water Dynamics and Exchange.* Here, we present a group of analyses concerning water dynamics in the W191G cavity and water exchange between the cavity and the bulk solvent that supply complementary information with respect to X-ray crystallography experiments in that (i) only positionally ordered water molecules can be detected by X-ray crystallography because the observed electron density represents a superposition of all atomic positions during the experiment time, which typically last hours; (ii) the location of water oxygen on the basis of X-ray crystallography measurements may be nontrivial because of resolution and refinement constraints (*21*, *73*); (iii) the average location and the thermodynamics of conserved water sites may differ for protein crystals and proteins in solution as suggested by experimental (*47*, *48*, *59*, *73*−*75*) and theoretical (*20*, *23*−*26*, *75*) studies; and (iv) X-ray diffraction data were collected at a lower temperature (288 K) than ligand-binding thermodynamic measurements data (298 K) (*6*, *32*).

Figure 6a and b show the time series of the total number of water molecules (nr. wt) and of the total number of

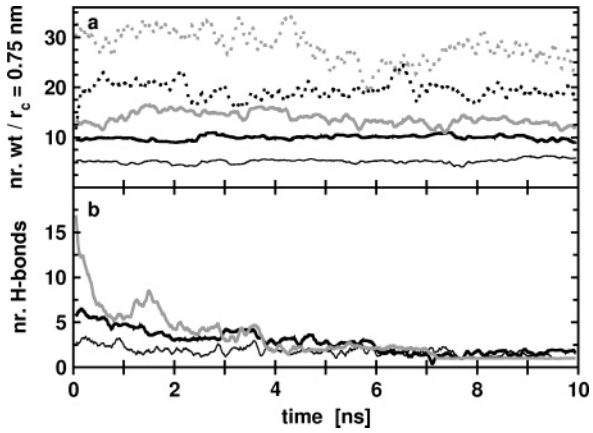


FIGURE 6: Time series of (a) the total number of water molecules inside the cavity (nr. wt) and (b) the total number of (solute−solvent and solvent−solvent) hydrogen bonds inside the cavity (nr. H-bonds). Graphs refer to **ref** (black thick lines), **bb** (black thin lines), **open** (gray thick lines), **rst** (dotted black lines), and **rsto** (dotted gray lines) simulations, respectively. The cavity was defined by a radius $r_c$ = 0.75 nm from Asp 235. For graphical purposes, running averages (over periods of 10 ps) are shown. See Table 1 for reference codes and the Materials and Methods section for computational details.

(solute−solvent and solvent−solvent) hydrogen bonds (nr. H-bonds) inside the cavity ($r_c$ = 0.75 nm), respectively. A decrease of hydrogen-bond formation during the pre-equilibrated 10 ns periods can be observed on the basis of trajectories **ref**, **bb**, and **open** at 300 K. The decrease is more evident when starting the simulation from an open-gate cavity (**open**). Yet, the $K^+$ ion occupying the cavity of the apo structure is on average coordinated to 3.2 water molecules throughout the **ref** simulation. The observation for the **open** simulation of the largest decrease of solvent−solvent hydrogen-bond relative occurrences is consistent with a disordered behavior of open-cavity water molecules. For **rst** and **rsto** simulations at 500 K, only a limited number of (solute−solvent or solvent−solvent) hydrogen bonds are formed inside the cavity (time occurrence <5%).

Figure 5b displays the number of water molecules exchanging (nr. ex-wt) for increasing cavity volumes. This quantity describes the disorder of water molecules exchanging from or toward a cavity of given radius. Similar values are found for **ref** and **bb** simulations at 300 K corresponding to a closed-cavity state (nr. ex-wt at $r_c$ = 0.75 nm; **ref**, 7 vs **bb**, 5). A large increase of exchanging waters is observed upon the opening of the cavity gate (nr. ex-wt at $r_c$ = 0.75 nm; **ref**, 7 vs **open**, 41). The exchanging behavior of cavity waters is substantially perturbed when the temperature increases to 500 K (nr. ex-wt at $r_c$ = 0.75 nm; **rst**, 5980 and **open**, 8227; not shown).

Figure 5c shows the number of water molecules never exchanging (nr. nex-wt) for increasing cavity volumes. This quantity describes the local order of water molecules confined into a cavity of a given radius. We find nex-wt values at $r_c$ = 0.75 nm of 6 and 3 for **ref** and **bb** simulations, respectively. In the case of **open**, **rst**, and **rsto** simulations, no water molecule is permanently residing within the cavity. The largest number of confined waters is observed for the apo cavity occupied by a $K^+$ ion, in agreement with the network of electrostatic interactions described in this case (see above).

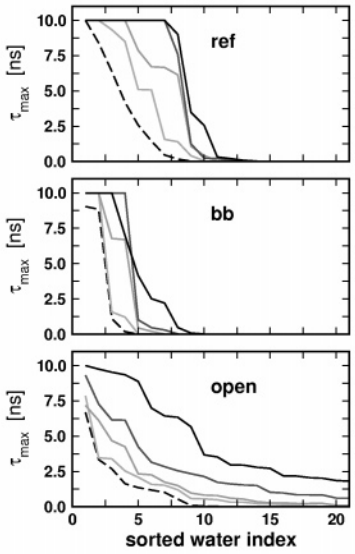


FIGURE 7: Sorted values of the largest 20 maximum residence times ($\tau_{max}$) of water molecules entering the W191G cavity for **ref**, **bb**, and **open** simulations. The results for increasing cavity volumes ($r_c$ = 0.5: dashed line, 0.6, 0.7, 0.8, and 0.9 nm) are displayed with corresponding darker lines. See Table 1 for reference codes and the Materials and Methods section for computational details.

Figure 7 displays the 20 largest sorted values of the maximum residence time ($\tau_{max}$) for individual water molecules entering the W191G cavity as calculated for **ref**, **bb**, and **open** simulations and on the basis of different cavity radii (in the range $r_c = 0.5-0.9$ nm). We find that water molecules near Asp 235 have comparatively larger $\tau_{max}$ values when the cavity is occupied by $K^+$ and the gating loop is in its closed state (**ref** and **bb** vs **open**). However, a number of molecules (**ref**, 6; **bb**, 2; **open**, 4) show maximum residence times longer than 5 ns for all simulations at 300 K. On the contrary, no water molecule shows maximum residence times larger than 0.50 and 0.25 ns for **rst** and **rsto** simulations at 500 K, respectively. The number of these (slowly or never) exchanging water molecules is systematically close to the average number of occupancy of the deeper cavity (compare with ⟨nr. wt⟩ values in Figure 5a; $r_c < 0.5$ nm). This observation holds also for larger cavity radii in the case of **ref** and **bb** simulations. Interestingly, different curves are observed for **ref** and **bb** (characterized by steep decreases) versus **open** (characterized by gradual decrease) simulations, demonstrating that $\tau_{max}$ values capture the different properties of exchanging waters depending on the state of the loop gate.

The comparison between cavity and bulk solvent dynamic diffusion properties supplies additional insight on the behavior of water. The self-diffusion coefficients $\langle D \rangle_{nex}$ averaged over the water molecules permanently inside the cavity (nr. nex-wt at $r_c = 0.75$ nm, **ref**: 6; **bb**: 3) are 3.1 $10^{-10}$ and 4.3 $10^{-10}$ $m^2$ $s^{-1}$, that is, about a factor of 8 or 10 smaller than that estimated for bulk water, on the basis of experiments (2.4 $10^{-9}$ $m^2$ $s^{-1}$ (*76*)) and the SPC water model (4.2 $10^{-9}$ $m^2$ $s^{-1}$ (*77*)), respectively. Similarly, the self-diffusion coefficient of the $K^+$ ion bound to the apo W191G cavity is about 10 times smaller than the average value $\langle D \rangle_{Kout}$ over the 7 $K^+$ ions present in solution (3.2 $10^{-10}$ and 4.0 $10^{-9}$ $m^2$ $s^{-1}$, respectively).

Some relevant trends concerning water dynamics and exchange can be summarized in the following key points:

(i) Exchange of water molecules between the cavity and bulk solvent occurs (to different extent) in all simulations suggesting that this process is not fully prevented by closing the loop gate and that sufficient flexibility of this region allows the passage of at least a few water molecules.

(ii) In the case of **ref** versus **bb**, a reduced number of water molecules is involved in exchange events when the cavity is occupied by the ligand; in the case of the apo structure occupied by one $K^+$ ion, we find that one molecule never leaves the deeper part of the cavity (Figure 5c; nr. ex-wt, $r_c < 0.6$ nm); in the **ref** and **bb** cases 6 and 3 water molecules never leave the whole cavity volume, respectively (Figure 5c; nr. nex-wt, $r_c = 0.75$ nm), in agreement with the presence of 6 and 3 water sites determined by X-ray crystallography experiments (Table 1). The corresponding crystallographic water sites are in qualitative agreement with calculated ensemble-averaged hydration-density maps (Figure 4a and b); comparatively higher ⟨nr. wt⟩ and nr. nex-wt values and lower $n_{ex}$ values for **ref** are found concerning the water dynamics of the whole cavity (Figure 5a and c; $r_c = 0.75$ nm).

(iii) In the case of **ref** versus **open**, a reduced number of water molecules occupies on average the deeper volume of the cavity near Asp 235 (Figure 5a; ⟨nr. wt⟩, $r_c < 0.6$ nm), and a larger number occupies on average the whole cavity (Figure 5a; ⟨nr. wt⟩, $r_c = 0.75$ nm; also cf. Figure 4a vs 4c) when the gating loop is in its open state. Water exchange is significantly enhanced, and the number of exchanging water molecules significantly increased upon opening the cavity gate, concerning both the deeper and the whole cavity volumes (Figure 5b; nr. ex-wt, $r_c = 0.5$ and 0.75 nm). No water molecule resides permanently in the cavity when the gating loop is in its open state (Figure 5c; nr. nex-wt, $r_c < 0.8$ nm). The 3 crystallographic water sites are in qualitative agreement with calculated ensemble-averaged hydration-density maps (Figure 4c).

(iv) In the case of **rst** and **rsto** versus **ref** and **open**, that is, a comparison between simulations at 300 and 500 K (in the second case keeping the backbone $C^\alpha$ positionally restrained close to the initial X-ray structure), the temperature increase results in significantly larger values of ⟨nr. wt⟩ (and of their standard deviations); about a factor of 10 larger $n_{ex}$ values; and more than a factor of 100 larger nr. ex-wt values. No water molecule resides permanently in the cavity at 500 K, even when the simulation starts from a closed-gate cavity.

## DISCUSSION

The artificial cavity of the W191G mutant of cytochrome *c* peroxidase shows distinct dynamics and hydration properties depending on the closed or open state of the flexible gating loop, the presence of ($K^+$ or small molecule) binding cations, and the temperature of the system. The molecular dynamics simulations presented herein show good general stability and agreement with the experimental data available.

The following conclusions emerge on the dynamics of the W191G protein. First, the conformational space sampled by the loop region shows a sizable reduction when the cavity hosts a small positively charged compound. Second, the conformational sampling of the cavity shows sizable reduc-

tion upon binding as quantified by analyzing backbone flexibility and the distributions of the dihedral spaces accessed by cavity residues. Third, the comparison of the surface solvent accessibility between single X-ray crystal structures and liquid-phase MD simulation ensembles indicates possible differences concerning solvent-exposed regions of the protein surface.

The following conclusions emerge concerning water dynamics and exchange in the W191G cavity. First, water disorder within the cavity and exchange between the cavity and bulk solvent are dominated by the presence of one charged molecule or ion, by the state of the cavity gate, and by the system temperature. The largest ordering factor is the presence of the $K^+$ ion occupying the gated W191G cavity, and cooperative interaction among several water molecules and polar residues define a network of interactions in qualitative agreement with that determined experimentally (*28*−*32*). Second, the opening of the cavity toward the bulk solvent introduces a large perturbation on the dynamic behavior of cavity waters. Considerable water exchange between the cavity and the bulk solvent occurs in the open-gate ensemble when no ligand or cation is bound. This suggests that cavity water molecules are rather disordered and that the open W191G cavity of cytochrome *c* peroxidase may experience hydration similar to that experimentally characterized for human interleukin-1$\beta$ (*48*). Yet, water diffusion is on average significantly slower inside the cavity than in the bulk solvent. Third, concerning the dependence of the number of water exchange events on the system temperature we observed effects in line with the behavior recently reported for the water molecules occupying a buried cavity of cytochrome *c* on the basis of residence time and coordination number calculations (*23*).

The simulations presented supply detailed information concerning the structural role of specific water molecules and allow more general conclusions on the interpretation of X-ray crystallography experiments elucidating time and ensemble averaging effects. Water sites determined by X-ray crystallography experiments systematically correspond to highly favorable average location of water molecules from simulation, within the approximations commonly employed in crystallographic refinement (*22*, *73*). However, our simulations suggest that the water sites located on the basis of X-ray crystallography data have to be interpreted as spheres defining the favorable (ensemble-averaged) presence of water molecules. In fact, individual water molecules may repeatedly exchange among water sites within the cavity volume yet occupy these sites on average. This suggests that X-ray crystallographic water sites correctly placed after model refinement are only a necessary (not sufficient) condition to imply the presence of permanently bound water molecules. This is of fundamental importance to design effective drug design methodology for ligand docking and will be further investigated.

Four major conclusions are apparent from our study. First, it seems misleading to interpret experimentally obtained conformations of protein loop regions in terms of single dominant structures. In fact, a variety of conformations (clusters of structures) is found. This observation is in line with the conformational heterogeneity of flexible surface loops described in the recent literature. We have shown that the dominant loop configurations in solution can be successfully captured using a clustering method previously described (*67*) and applied in the context of flexible peptides (*14*, *67*−*69*). We suggest that this procedure is a useful tool to characterize the flexibility of mobile regions in biomolecules. Second, our simulations support the general picture of Pro 190 *cis*−*trans* isomerization as a determinant of the loop-opening mechanism and are in line with the general observation of Pro *cis*−*trans* isomerization in folded proteins being governed also by cooperative (*42*) (long-range) as opposed to only local (*43*) (short-range) interactions. Third, receptor flexibility appears to be a key ingredient of ligand binding and molecular recognition (*78*). Therefore, docking studies aimed at understanding molecular association processes in the liquid phase should involve the extensive sampling of the receptor conformational space. Fourth, molecular dynamics simulations can be used to characterize the effects regulating the dynamics and exchange of water molecules and ions, providing atomic level and time-dependent information otherwise invisible to experiments. Future NMR experiments and free energy and entropy calculations will elucidate the different thermodynamic roles of bound and unbound cavity waters.

## ACKNOWLEDGMENT

R.B. warmly thanks Cristina S. Pereira, Donald Hamelberg, and Cesar A. F. de Oliveira for useful discussions. The Center for Theoretical Biological Physics of the University of California at San Diego is acknowledged for the computing resources employed.

## SUPPORTING INFORMATION AVAILABLE

Solute−solvent hydrogen-bond occurrences in the W191G cavity (Table S1); time series of secondary structure formation (Figure S1); analysis of heme cofactor stability (Figure S2); distributions of Asn 295−ion distances (Figure S3); backbone torsional angle distributions characterizing the loop-gating mechanism around Pro 190 and Asn 195 hinge points (Figure S4); time series of SASA and W191G cavity RGYR values (Figure S5); and radial distribution functions for water oxygens in the cavity (Figure S6). This material is available free of charge via the Internet at http://pubs.acs.org.

## REFERENCES

1. Herschlag, D. (1988) The role of induced fit and conformational changes of enzymes in specificity and catalysis, *Bioinorg. Chem. 16*, 62−96.
2. Young, R. D., Frauenfelder, H., Johnson, J. B., Lamb, D. C., Nienhaus, G. U., Philipp, R., and Scholl, R. (1991) Time- and temperature dependence of large-scale conformational transitions in myoglobin, *Chem. Phys. 158*, 315−327.
3. Taylor, J. C., and Markham, G. D. (1991) Conformational dynamics of the active site loop of S-adenosyl-methionine synthetase illuminated by site-directed spin labeling, *Arch. Biochem. Biophys. 415*, 164−171.
4. Falzone, C. J., Wright, P. E., and Benkovic, S. J. (1994) Dynamics of a flexible loop in dihydrofolate reductase from *Escherichia coli* and its implication for catalysis, *Biochemistry 33*, 439−442.
5. Williams, J. C., and McDermott, A. E. (1995) Dynamics of the flexible loop of triophosphate isomerase: the loop motion is not ligand gated, *Biochemistry 34*, 8309−9319.
6. Fitzgerald, M. M., Musah, R. A., McRee, D. E., and Goodin, D. B. (1996) A ligand-gated, hinged loop rearrangement opens a channel to a buried artificial protein cavity, *Nat. Struct. Biol. 3*, 626−631.

7. Zhou, H-X., Wlodek, S. T., and McCammon, J. A. (1998) Conformation gating as a mechanism for enzyme specificity, *Proc. Natl. Acad. Sci. U.S.A. 95*, 9280−9283.
8. Aparicio, R., Ferreira, S. T., and Polikarpov, I. (2003) Closed conformation of the active site loop of rabbit muscle triosephosphate isomerase in the absence of substrate: evidence of conformational heterogenity, *J. Mol. Biol. 334*, 1023−1041.
9. McNulty, B. C., Triphaty, A., Young, G. B., Charlton, L. M., Orans, J., and Pielak, G. J. (2006) Temperature-induced reversible conformational change in the first 100 residues of α-synuclein, *Protein Sci. 15*, 602−608.
10. Stein, R. L. (1993) Mechanism of enzymatic and nonenzymatic prolyl *cis-trans* isomerization, *Adv. Protein Chem. 44*, 1−24.
11. Eckert, B., Martin, A., Balbach, J., and Schmid, F. X. (2005) Prolyl isomerization as a molecular timer in phage infection, *Nat. Struct. Biol. 12*, 619−623.
12. Creighton, T. E. (1993) *Proteins: Structure and Molecular Properties*, W. H. Freeman, New York.
13. Daura, X., Antes, I., van Gunsteren, W. F., Thiel, W., and Mark, A. E. (1999) The effect of motional averaging on the calculation of NMR-derived structural properties, *Proteins: Struct., Funct., Genet. 36*, 542−555.
14. Baron, R., Bakowies, D., and van Gunsteren, W. F. (2004) Carbopeptoid folding: effects of stereochemistry, chain length, and solvent, *Angew. Chem., Int. Ed. 43*, 4055−4059.
15. Zagrovic, B., and Pande, V. S. (2004) How does averaging affect protein structure comparison on the ensemble level? *Biophys. J. 87*, 2240−2246.
16. Zagrovic, B., and van Gunsteren, W. F. (2006) Comparing atomistic simulation data with the NMR experiment: how much can NOEs actually tell us?, *Proteins: Struct., Funct., Bioinf. 63*, 210−218.
17. van Gunsteren, W. F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D. P., Glättli, A., Hünenberger, P. H., Kastenholz, M. A., Oostenbrink, C., Schenk, M., Trzesniak, D., van der Vegt, N. F. A., and Yu, H. (2006) Biomolecular modeling: goals, problems, perspectives, *Angew. Chem., Int. Ed. 45*, 4064−4092.
18. Adcock, S. A., and McCammon, J. A. (2006) Molecular dynamics: survey of methods for simulating the activity of proteins, *Chem. Rev. 106*, 1589−1615.
19. Wade, R. C., Mazor, M. H., McCammon, J. A., and Quiocho, F. A. (1991) A molecular dynamics study of thermodynamic and structural aspects of the hydration of cavities in proteins, *Biopolymers 1*, 919−931.
20. Brunne, R. M., Liepinsh, E., Otting, G., Wüthrich, K., and van Gunsteren, W. F. (1993) A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations, *J. Mol. Biol. 231*, 1040−1048.
21. Schiffer, C. A., and van Gunsteren, W. F. (1999) Accessibility and order of water sites around proteins: a crystallographic time-averaging study, *Proteins: Struct., Funct., Genet. 36*, 501−511.
22. Luise, A., Falconi, M., and Desideri, A. (2000) Molecular dynamics simulation of solvated azurin: correlation between surface solvent accessibility and water residence times, *Proteins: Struct., Funct., Genet. 39*, 56−67.
23. Garcia, A. E., and Hummer, G. (2000) Water penetration and escape in proteins, *Proteins: Struct., Funct., Genet. 38*, 261−272.
24. Marchi, M., Sterpone, F., and Ceccarelli, M. (2002) Water rotational relaxation and diffusion in hydrated lysozyme, *J. Am. Chem. Soc. 124*, 6787−6791.
25. Henchman, R. H., Tai, K., Shen, T., and McCammon, J. A. (2002) Properties of water molecules in the active site gorge of acetylcholinesterase from computer simulation, *Biophys. J. 82*, 2671−2682.
26. Hamelberg, D., and McCammon, J. A. (2004) Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method, *J. Am. Chem. Soc. 24*, 7683−7689.
27. Gorfe, A. A., and Caflisch, A. (2005) Functional plasticity in the substrate binding site of β-secretase, *Structure 13*, 1487−1498.
28. Fitzgerald, M. M., Churchill, M. J., McRee, D. E., and Goodin, D. B. (1994) Small molecule binding to an artificially created cavity at the active site of cytochrome c peroxidase, *Biochemistry 33*, 3807−3818.
29. Fitzgerald, M. M., Trester, M. L., Jensen, G. M., McRee, D. E., and Goodin, D. B. (1995) The role of aspartate-235 in the binding of cations to an artificial cavity at the radical site of cytochrome c peroxidase, *Protein Sci. 4*, 1844−1850.
30. Musah, R. A., Jensen, G. M., Rosenfeld, R. J., McRee, D. E., and Goodin, D. B. (1997) Variation in strength of an unconventional C-H to O hydrogen bond in an engineered protein cavity, *J. Am. Chem. Soc. 119*, 9083−9084.
31. Musah, R. A., and Goodin, D. B. (1997) Introduction of novel substrate oxidation into cytochrome c peroxidase by cavity complementation: oxydation of 2-aminothiazole and covalent modification of the enzyme, *Biochemistry 36*, 11665−11674.
32. Musah, R. A., Jensen, G. M., Bunte, S. W., Rosenfeld, R. J., and Goodin, D. B. (2002) Artificial protein cavities as specific ligand-binding templates: characterization of an engineered heterocyclic cation-binding site that preserves the evolved specificity of the parent protein, *J. Mol. Biol. 315*, 845−857.
33. Grochulski, P., Li, Y., Schrag, J. D., and Cygler, M. W. (1994) Two conformational states of *Candida rugosa* lipase, *Protein Sci. 3*, 82−91.
34. Eyles, S. J. (2001) Proline not the only culprit? *Nat. Struct. Biol. 8*, 380−381.
35. Schmid, F. X. (1993) Prolyl isomerase: enzymatic catalysis of slow protein-folding, *Annu. Rev. Biophys. Biomol. Struct. 22*, 123−143.
36. Vanhoof, G., Goossens, F., Meester, I., Hendriks, D., and Scharpé, S. (1995) Proline motifs in peptides and their biological processing, *FASEB J. 9*, 736−744.
37. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues, *Biochemistry 14*, 4953−4963.
38. Grathwol, C., and Wütrich, K. (1981) NMR studies of the rates of proline *cis-trans* isomerization in oligopeptides, *Biopolymers 20*, 2623−2633.
39. Fischer, S., Bunbrack, R. L., Jr., and Karplus, M. (1994) *Cis-trans* imide isomerization of the proline dipeptide, *J. Am. Chem. Soc. 116*, 11931−11937.
40. Shi, T., Spain, S. M., and Rabenstein, D. L. (2004) Unexpectedly fast *cis/trans* isomerization of Xaa-Pro peptide bonds in disulfide-constrained cyclic peptides, *J. Am. Chem. Soc. 126*, 790−796.
41. Hamelberg, D., Shen, T., and McCammon, J. A. (2005) Phosphorilation effects on *cis/trans* isomerization and the backbone conformation of serine-proline motifs: accelerated molecular dynamics analysis, *J. Am. Chem. Soc. 127*, 1969−1974.
42. Tweedy, N. B., Nair, S. K., Paterno, S. A., Fierke, C. A., and Christianson, D. W. (1993) Structure and energetics of a non-proline *cis*-peptidyl linkage in a proline-202 → alanine carbonic anhydrase II variant, *Biochemistry 32*, 10944−10949.
43. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M., and Fischer, G. (1998) Side-chain effects of peptidyl-prolyl *cis/trans* isomerization, *J. Mol. Biol. 279*, 449−460.
44. Koide, S., Dyson, H. J., and Wright, P. E. (1993) Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization, *Biochemistry 32*, 12299−12310.
45. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., and Matthews, B. W. (1992) Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect, *Science 255*, 178−183.
46. Ringe, D. (1995) What makes a binding site a binding site? *Curr. Opin. Struct. Biol. 5*, 828−829.
47. Buckle, A. M., Cramer, P., and Fersht, A. R. (1996) Structural and energetic responses to cavity-creating mutations in hydrophobic cores: observation of a buried water molecule and the hydrophilic nature of such hydrophobic cavities, *Biochemistry 35*, 4298−4305.
48. Ernst, J. A., Clubb, R. T., Zhou, H. X., Gronenborn, A. M., and Clore, G. M. (1995) Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR, *Science 267*, 1813−1817.
49. Dunitz, J. D. (1994) The entropic cost of bound water in crystals and biomolecules, *Science 264*, 670.
50. Banba, S., and Brooks, C. L. (2000) Free energy screening of small ligands binding to an artificial protein cavity, *J. Chem. Phys. 113*, 3423−3433.
51. Brenk, R., Vetter, S. W., Boyce, S. E., Goodin, D. B., and Shoichet, B. K. (2006) Probing molecular docking in a charged model binding site, *J. Mol. Biol. 357*, 1449−1470.
52. Christen, M., Hünenberger, P. H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D. P., Heinz, T. N., Kastenholz, M. A., Kräutler, V., Oostenbrink, C., Peter, C., Trzesniak, D., and van Gunsteren, W.

F. (2005) The GROMOS software for biomolecular simulation: GROMOS05, *J. Comput. Chem. 26*, 1719−1751.
53. van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P., and Tironi, I. G. (1996) *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Hochschulverlag an der ETH Zürich/Biomos, Zürich/Groningen, Germany.
54. Soares, T. A., Hünenberger, P. H., Kastenholz, M. A., Kräutler, V., Lenz, T., Lins, R. D., Oostenbrink, C., and van Gunsteren, W. F. (2005) An improved nucleic acid parameter set for the GROMOS force field, *J. Comput. Chem. 26*, 725−737.
55. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981) Interaction Models for Water in Relation to Protein Hydration, in *Intermolecular Forces* (Pullman, B. E., Ed.), pp 331−342, D. Reidel Publishing Company, Dordrecht, The Netherlands.
56. Åqvist, J. (1990) Ion-water interaction potentials derived from free energy perturbation simulations, *J. Phys. Chem. 94*, 8021−8024.
57. The PDB structures used in this study are available from the RCSB Protein Data Bank, http://www.rcsb.org. The PDB ID reference codes of the structures used in this study are as follows: 1AA4 (**ref**; **rst**), 1AEN (**bb**), and 1RYC (**open**; **rsto**). The corresponding models were refined at resolutions of 0.21, 0.21, and 0.18 nm, respectively.
58. Hockney, R. W. (1970) The potential calculation and some applications, *Methods Comput. Phys. 9*, 136−211.
59. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-integration of Cartesian equations of motion of a system with constraints: molecular-dynamics of n-alkanes, *J. Comput. Phys. 23*, 327−341.
60. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular-dynamics with coupling to an external bath, *J. Chem. Phys. 81*, 3684−3690.
61. van Gunsteren, W. F., and Berendsen, H. J. C. (1990) Computer-simulation of molecular-dynamics: methodology, applications, and perspectives in chemistry, *Angew. Chem., Int. Ed. Engl. 29*, 992−1023.
62. Tironi, I. G., Sperb, R., Smith, P. E., and van Gunsteren, W. F. (1995) A generalized reaction field method for molecular-dynamics simulations, *J. Chem. Phys. 102*, 5451−5459.
63. Heinz, T. N., van Gunsteren, W. F., and Hünenberger, P. H. (2001) Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations, *J. Chem. Phys. 115*, 1125−1136.
64. Heinz, T. N., and Hünenberger, P. H. (2004) A fast-construction algorithm for molecular simulations under periodic boundary conditions, *J. Comput. Chem. 25*, 1474−1486.
65. McLachlan, A. D. (1979) Gene duplications in the structural evolution of chymotrypsin, *J. Mol. Biol. 128*, 49−79.
66. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features, *Biopolymers 22*, 2577−2637.
67. Daura, X., van Gunsteren, W. F., and Mark, A. E. (1999) Folding-unfolding thermodynamics of a $\beta$-heptapetide from equilibrium simulations, *Proteins: Struct., Funct., Genet. 34*, 269−280.
68. Baron, R., Bakowies, D., van Gunsteren, W. F., and Daura, X. (2002) $\beta$-peptides with different secondary-structure preferences: how different are their conformational spaces? *Helv. Chim. Acta 85*, 3872−3882.
69. Baron, R., Bakowies, D., and van Gunsteren, W. F. (2005) Principles of carbopeptoid folding: a molecular dynamics simulation study, *J. Pept. Sci. 11*, 74−84.
70. Hubbard, S. J., and Thornton, J. M. (1993) *NACCESS*, Computer program, Department of Biochemistry and Molecular Biology, University College London, UK.
71. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics, *J. Mol. Graphics 14*, 33−38. .
72. Hünenberger, P. H., Mark, A. E., and van Gunsteren, W. F. (1995) Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations, *J. Mol. Biol. 252*, 492−503.
73. Garcia, A. E., Krumhansi, J. A., and Frauenfelder, H. (1997) Variation on a theme by Debye and Waller: from simple crystal to proteins, *Proteins 29*, 153−160.
74. Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S., and Zaccai, G. (1998) Protein hydration in solution: experimental observation by X-ray and neutron scattering, *Proc. Natl. Acad. Sci. U.S.A. 95*, 2267−2272.
75. Leitner, D. M., Havenith, M., and Gruebele, M. (2006) Biomolecule large-amplitude motion and solvation dynamics: modeling and probes from THz to X-rays, *Int. Rev. Phys. Chem 25*, 553−582.
76. Krynicki, K., Green, C. D., and Sawyer, D. W. (1978) Pressure and temperature-dependence of self-diffusion in water, *Faraday Discuss. Chem. Soc. 66*, 199−208.
77. Baron, R., Trzesniak, D., de Vries, A. H., Elsener, A., Marrink, S. J., and van Gunsteren, W. F. (2007) Comparison of thermodynamic properties of coarse-grained and atomic-level simulation models, *ChemPhysChem 8*, 452−461.
78. McCammon, J. A. (2005) Target flexibility in molecular recognition, *Biochim. Biophys. Acta 1754*, 221−224.

BI700866X