Dynamics of Water Molecules in the Active-Site Cavity of Human Cytochromes P450

Patrik Rydberg,[†] Thomas H. Rod,[†,‡] Lars Olsen,[§] and Ulf Ryde*[,†]

Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124, S-221 00 Lund, Sweden, Atomistix A/S, Juliane Maries Vej 30, DK-2100 Copenhagen Ø, Denmark, and Biostructural Research Group, The Danish University of Pharmaceutical Sciences, Universitetsparken 2, DK-2200 Copenhagen Ø, Denmark

Received: January 17, 2007; In Final Form: March 6, 2007

We have studied the dynamics of water molecules in six crystal structures of four human cytochromes P450, 2A6, 2C8, 2C9, and 3A4, with molecular dynamics simulations. In the crystal structures, only a few water molecules are seen and the reported sizes of the active-site cavity vary a lot. In the simulations, the cavities are completely filled with water molecules, although with ∼20% lower density than in bulk water. The 2A6 protein differs from the other three in that it has a very small cavity with only two water molecules and no exchange with the surroundings. The other three proteins have quite big cavities, with 41 water molecules on average in 2C8 and 54−58 in 2C9 and 3A4, giving a water volume of 1500−2100 Å^3. The two crystal structures of 2C9 differ quite appreciably, whereas those of 3A4 are quite similar. The active-site cavity is connected to the surroundings by three to six channels, through which there is a quite frequent exchange of water molecules (one molecule is exchanged every 30−200 ps), except in 2A6. Most of the channels are observed also in the crystal structures, but two to three channels in each protein open only during the simulations. There are no water molecules close to the heme iron ion in these simulations of the high-spin ferric state (the average distance to the closest water molecule is 3.3−5 Å), and there are few ordered water molecules in the active sites, none of which is conserved in all proteins.

Introduction

The cytochromes P450 form a ubiquitous protein family with functions including the synthesis and degradation of many physiologically important compounds and the degradation of xenobiotics, for example, most drugs.[1,2] In the human genome, there are ∼60 isoforms of this protein, but five of them, 1A2, 2C9, 2C19, 2D6, and 3A4, account for the oxidation of ∼90% of the drugs on the market.[3]

Crystal structures of bacterial P450s have been known for 20 years,[4] but crystal structures of human P450s have only been known for 4 years.[5−11] The active site consists of a heme group with a central iron ion that is coordinated to a cysteine (Cys) ligand from the protein. In some crystals, there is also a sixth ligand to iron on the opposite side, typically a water molecule or an oxygen-derived ligand. According to the consensus reaction mechanism of the P450s,[2] the resting Fe(III) state of the enzyme binds the substrate. It is then reduced to Fe(II), allowing it to bind O_2. After another reduction, the O−O bond is heterolytically cleaved and a high-valent heme state (compound I) is formed, which may oxidize the substrate in various ways.

On the side of the heme group that is opposite to the Cys ligand, there is a cavity in the P450s, where the substrate binds. The size of this cavity varies strongly between the different P450s. For example, the volume of the cavity in the crystal structure of the 2C8 isoform has been estimated to 1438 Å^3,[10] whereas it is only 470 Å^3 in the 2C9 isoform.[5] In the bacterial enzymes, the cavity is even smaller, down to 79 Å^3.[12] Naturally, the size of the cavity determines the maximum size of the substrates and it also determines the dynamics of a putative substrate in the active site (e.g., if the substrate may rotate in the active site or if it must enter the cavity in the right orientation for a reaction to occur). Therefore, it is important to estimate and compare the volume of various cytochromes P450.

In crystal structures of the P450s, some water molecules are normally identified in the active-site cavity, but the number varies strongly in the different structures, ranging between 0 and ∼20 in the human structures. This means that there are extensive empty voids in most crystal structures. The hydration of cavities in proteins has been much studied,[13−19] and the consensus is that cavities of this size are normally more or less filled with water molecules (the motion of which are so large that they are not always observed in the crystal structures). In particular, it has been shown by both experiments and theoretical calculations that the active-site cavity of the bacterial P450cam contains six water molecules in the substrate-free form, although there is room for ∼10 waters.[20−23] However, the number and the dynamics of the water molecules in the human P450s is currently not known. This is a great problem, not the least in computational studies of these proteins, for which it makes a great difference if there are few or many water molecules in the cavity, both for substrate dynamics and for the reactivity of the active site. In fact, it has frequently been suggested that ordered water molecules play an essential role in the reaction mechanism of the P450s by relaying the proton transfers needed for the heterolytic cleavage of dioxygen.[24−26]

The active-site cavity of many bacterial and some human P450s is essentially closed, without any obvious channels

* To whom correspondence should be addressed. E-mail: Ulf.Ryde@teokem.lu.se. Phone: +46-46 2224502. Fax: +46-46 2224543.
[†] Lund University.
[‡] Atomistix A/S.
[§] The Danish University of Pharmaceutical Sciences.

10.1021/jp070390c CCC: $37.00
Published on Web 04/19/2007

TABLE 1: Data for the Studied Crystal Structures

	2A6	2C8	2C9o	2C9r	3A4t	3A4w
PDB code	1Z10	1PQ2	1OG2	1R9O	1TQN	1W0F
reference	11	10	5	6	7	8
resolution (Å)	1.90	2.7	2.6	2.0	2.05	2.65
number of independent subunits	4	2	2	1	1	1
reported cavity volume (Å^3)		1438[a]	470; 600[b]		1386[a]	520[b]
solvent-accessible volume[c] (Å^3)	83	441	484	819[d]	512	554
water molecules in active-site cavity	0	2	5	3	20	10
shortest Fe−water distance[e] (Å)	8.3−8.5	7.5−9.0	2.9−4.3	2.43	2.07	8.6
shortest Fe−substrate distance (Å)	3.1[f]			4.9[g]		

[a] Volume calculated by VOIDOO.[50] [b] Probe-accessible volume calculated by LIGSITE.[56] [c] Calculated by the Pocket[55] software. [d] The difference from the other structures is entirely caused by the parameters of the calculation (how much of the channel is included in the cavity). If a probe radius of 1.8 Å is used (instead of the default 1.4 Å), the solvent-accessible volumes of 2C9r and 2C9o are exactly the same. [e] A range illustrates the variation among the crystallographically independent subunits. [f] Coumarin substrate. [g] Flurbiprofen substrate.

through which substrates and products may enter and leave, but also several P450 structures exist, in which the active site is more open to the surroundings.[10,27−29] Several studies have been devoted to the identification of channels into the active-site cavity, especially for the bacterial enzymes,[30−35] but recently also for the human enzymes.[29,35,36] These studies have described 13 channels, although the extent to which they are used and the detailed dynamics of the transport seem to vary extensively between the different structures. It has also been pointed out that water molecules in the active site have to be removed when the substrate binds.[37] They can use the same channels as the substrate and the product, but they may also use other (smaller) channels.[30]

In this paper, we use molecular dynamics simulations to study the structure and dynamics of water molecules in six crystal structures of human cytochrome P450s (2A6, 2C8, 2C9, and 3A4). Molecular dynamics simulations have been performed on a few of these structures before[36,38] but with a different focus. Our simulations allow us to discuss the average number of water molecules in the active-site cavity, various definitions of the volume of the cavity, the dynamics of the water molecules, their transport in and out of the cavity, and whether there are any ordered water molecules close to heme that may be of catalytic significance.

Methods

Structures. Six structures of four different human cytochromes P450 (2A6, 2C8, 2C9, and 3A4) were studied. The coordinates were obtained from the protein data bank,[39] access codes 1Z10 (2A6),[11] 1PQ2 (2C8),[10] 1OG2,[5] 1R9O[6] (both 2C9), 1TQN,[7] and 1W0F[8] (both 3A4). In the following, the two 2C9 structures are called 2C9o (1OG2) and 2C9r (1R9O) and the two 3A4 structures are called 3A4t (1TQN) and 3A4w (1W0F). Amino acids that are not visible in the electron-density maps (amino acids 282−285 in 3A4t, 262−268 and 281−288 in 3A4w, and 38−42 and 214−220 in 2C9r) were built using the SWISS-MODEL server[41] without any template. Any ligands in the structures were removed, and only one monomer was used for the multimeric structures. Hydrogen atoms and other missing atoms were added with the CHARMM software,[42] assuming that all Asp and Glu residues are negatively charged and that the Lys and Arg residues are positively charged. The protonation status of the His residues was determined by inspection of the solvent accessibility and hydrogen-bond network around each residue and is shown in Supporting Information Table S1. The Cys iron ligand was assumed to be negatively charged, whereas the other amino-acid side chains were assumed to be neutral. The iron ion was studied in the high-spin Fe(III) state because four of the crystal structures have no sixth ligand of the heme iron (cf. Table 1) and we wanted to study all structures in the same state.

Parametrization of the Heme Group. We used the standard CHARMM parameters for the heme group (toppar_all22_prot_heme.str), including an explicit Fe−S bond. However, we recalculated the charges of the heme group and the Cys ligand with the restrained electrostatic potential (RESP)[43] as implemented in Amber 8.[44] These calculations were based on electrostatic potential points obtained with the Merz−Kollman scheme[45] at the B3LYP/6-31G* level of theory using the Gaussian 98 software package.[46] We used a model consisting of the heme group with CH_3S^- coordinated to the high-spin Fe(III) ion (Supporting Information Figure S1) with the coordinates taken from the 2C9 structure. Hydrogen atoms were added with the CHARMM software. The resulting charges are listed in Supporting Information Table S2.

To test the parameters and find the intrinsic affinity of the iron ion for a water molecule, we made a molecular mechanics minimization of the heme group, a water molecule, the Cys ligand, and its two adjacent amino acids (Asn and Ile) from a snapshot of the equilibrated 3A4w structure. This gave an Fe−O distance of 3.42 Å, which shows that the heme potential works properly, giving a five-coordinate iron ion, as in the quantum mechanical calculations.

Molecular Dynamics Simulations. Molecular dynamics (MD) simulations were performed with the CHARMM software, version 30b1.[42] Water molecules were described by the TIP3P explicit solvent model.[47] The protein was solvated in a box of water molecules, extending at least 12 Å outside the protein. The initial box size and number of water molecules are shown in Supporting Information Table S3. The solvent and hydrogen atoms were relaxed by a short molecular mechanics minimization (1100 steps) with all atoms except hydrogen atoms and water molecules harmonically restrained to their crystal positions using a mass-weighted force constant of 10 kcal/mol/amu/Å^2. Then, a 20 ps MD simulation was run with only the protein backbone restrained to the crystal positions using a mass-weighted force constant of 5 kcal/mol/amu/Å^2. Finally, the whole system was equilibrated for 1 ns and data were then collected every fifth picosecond for 4 ns. A time step of 2 fs was used throughout, and all bonds involving hydrogen atoms were constrained to the equilibrium distances by the SHAKE algorithm.[48] Electrostatic interactions were treated by the particle mesh Ewald method.[49] All simulations were performed at a constant temperature of 298 K and a constant pressure of 1 atm.

The root-mean-squared deviation (RMSD) of the various structures during the simulations are shown in Supporting Information Figure S2a and Table S4. It shows that all structures have a similar RMSD with respect to the average equilibrated

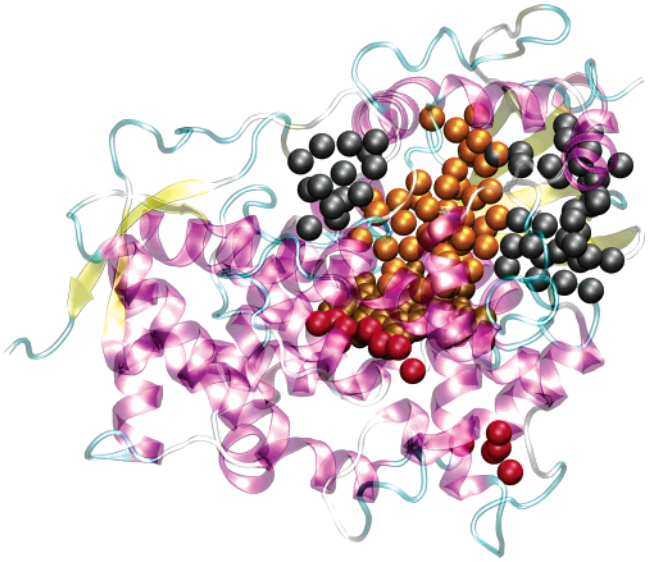

Figure 1. Illustration of the problem of defining the active-site cavity, using a snapshot from the 3A4t simulation. Water molecules in the active-site cavity are orange, those in side cavities are red, and those in channels connecting to the bulk are black. The atoms of the heme group are brown.

structure (0.8−1.1 Å, least for 2A6 and largest for 3A4). The 2C9o and 2C9r structures have similar RMSDs, which indicates that the proteins are related and that the MD simulations are proper. The same applies to the 3A4t and 3A4w structures. The RMSDs with respect to the crystal structures are somewhat larger (1.4−2.0 Å) and vary more among the various structures. The radius of gyration is also shown in Supporting Information Figure S2b and Table S4.

Finding Cavity Water Molecules. The aim of the present investigation is to study the number and dynamics of water molecules in the active-site cavity of the various cytochromes P450. Unfortunately, it is not trivial to decide which water molecules belong to this cavity,[50,51] as is illustrated in Figure 1. First, the active-site cavity is connected to the surrounding solvent by water-filled channels. Therefore, it is not obvious where the channels end and the cavity starts, especially as the shapes of the channels change during the simulations. Second, there are other small cavities in the protein, which are normally not connected to the active-site cavity but sometimes merge with it during the MD simulations. Therefore, it is hard to construct methods that always provide the correct number of water molecules in the active-site cavity (even by visual inspection, it is normally not possible to unambiguously decide the number of cavity water molecules, cf. Figure 1).

There are many algorithms and programs that identify cavity water molecules in proteins.[13,50,52−58] After having tried several of these, we decided to implement and use a simple and well-defined algorithm, called cavity_wat below. This algorithm defines bulk water molecules as those that are at least 3 Å from any protein atom and in contact with (i.e., within 2.5 Å of) other bulk water molecules (to ensure that water molecules in the middle of cavities inside the protein are not classified as bulk water molecules). All atoms (including hydrogens) are considered in these distances. Sometimes, this is not enough to avoid that water molecules in the active-site cavity are classified as bulk water molecules. Therefore, the algorithm also allows the user to discard bulk water molecules close to an arbitrary position in the simulated system (we used a distance of 13 Å from the heme iron in this investigation). Three subsequent layers of water molecules (quite arbitrarily called surface, deep surface, and channel water) are then defined as water molecules in contact with the previous layer and not belonging to an already defined layer. The remaining molecules are defined as cavity water molecules (cf. Supporting Information Figure S3).

Of course, this includes any water molecules in voids and pockets of the protein and not only the active-site cavity. Therefore, we next run a Python script intended to restrict the list to water molecules in the active-site cavity: First, all cavity water molecules below the heme plane (i.e., on the same side as the axial Cys ligand) are excluded. Then, the script selects the cavity water molecules that are in direct or indirect (i.e., by other cavity water molecules) contact with the water molecule closest to the heme iron.

These algorithms are not foolproof, and sometimes they fail. We have found four main causes for these failures. First, it may happen that the distal part of the cavity or the channels widen so much that the algorithm finds bulk water inside them. Second, the active-site cavity may occasionally merge with other cavities. Third, the mouth of a channel may become so narrow that its surroundings and all of the channel become categorized as a cavity. Fourth, some water molecules are not always directly connected to the other water molecules in the active-site cavity. To avoid the results being biased by these problems, we made a histogram of the number of cavity water molecules for each protein and manually inspected all snapshots that were found outside the 95% significance error margin. Snapshots for which the algorithm had failed were discarded. In our simulations, most problems were encountered for the 2C9 structures, because they have many wide channels to the solvent. Some problematic structures were discarded for 2C8 and 3A4, whereas for 2A6 the problems were minimal. All data were analyzed using the statistics software R,[59] version 2.1.1.

The solvent-accessible volumes of the active-site cavities were determined directly from the crystal structures (without equilibration) with the ProShape Pocket software,[55,60] using a probe radius of 1.4 Å. Since Pocket does not recognize the heme group, 200 $Å^3$ (the approximate volume of the heme group) was subtracted from the raw results. We also calculated the volume of the cavity during the simulations on the basis of the number of water molecules multiplied with the average volume of a water molecule in bulk water (30 $Å^3$ [13]).

The exchange of water molecules between the cavity and bulk was identified from the list provided by the cavity_wat and Python search script. Potential exchanges were visually inspected by the software visual molecular dynamics (VMD).[61]

Results and Discussion

Cavity Volume and Water Molecules in the Crystal Structures. According to the original publications, the volumes of active-site cavities vary vastly in the various human cytochromes P450, 470−1438 $Å^3$.[7,8,10] However, this variation is mainly caused by the fact that different methods and parameters have been used to calculate the volume (as has already been noted for the two 3A4 structures[62]). In fact, there are several definitions of volumes in proteins, as is illustrated in Figure 2:[52,63] The union of the van der Waals volumes of all atoms gives the *van der Waals volume*. If the van der Waals radii are increased by a solvent probe radius (typically the radius of water, 1.4 Å), the *solvent-accessible volume* is obtained. It is delimited by the surface obtained by rolling a sphere with the probe radius over the van der Waals surface. Finally, the *molecular volume* is similar to the van der Waals volume, but it includes also the space not accessible by the solvent probe. Consequently, the molecular volume is slightly larger than the van der Waals

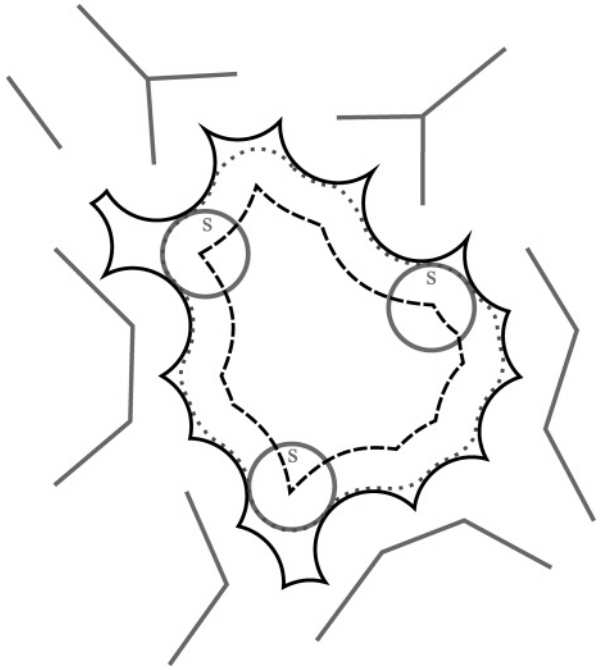

Figure 2. Schematic illustration of the definition of various cavity volumes. The van der Waals volume is shown as a solid line, the molecular volume by a dotted line, and the solvent-accessible volume by a dashed line. The circles marked with an "S" are examples of the solvent probe, and the straight gray lines are schematic protein residues.

volume, whereas the solvent-accessible volume is appreciably larger. Conversely, the volume of a cavity, delineated by a protein surface calculated by these three definitions should be largest for the van der Waals volume and smallest for the solvent-accessible volume (Figure 2).

Moreover, different programs use different methods to find cavities and their volumes, and they can give slightly different results.[13,57] Even worse, most programs require that a cavity is not accessible to bulk solvent, which for P450 structures with wide channels means that the probe radius must be increased until the channels are closed, but at the same time, this increase in probe radius will decrease the solvent-accessible volume of the cavity. Thus, depending on the probe radius, various amounts of the channels will be included in the active-site cavity.

This problem is illustrated in Supporting Information Table S5, in which the volume of the active-site cavity in 3A4t has been calculated with the VOIDOO software[50] with the three different volume definitions and two different probe radii. It can be seen that the three different volumes differ strongly (from 37 to 561 Å^3), whereas the dependence on the probe radius is smaller for this protein with small channels. However, neither of the six volumes are close to the volume calculated in the original article (1386 Å^3), although the same program was used, indicating that nondefault parameters were used. Unfortunately, the VOIDOO program fails to identify the cavity for three of the structures with default settings.

We have been able to estimate solvent-accessible volumes for all six crystal structures in this investigation with the Pocket software[55] and default settings. The results in Table 1 show that if the same definition of the volume is used, the variation in cavity size is small for four of the structures (2C8, 2C9o, 3A4t, and 3A4w), 441−554 Å^3. For the other 2C9 structure (2C9r), the program estimates a much larger cavity volume, 819 A^3. However, this difference is entirely caused by the parameters of the program: If a probe radius of 1.8 Å is used (instead of the default, 1.4 Å), the solvent-accessible cavity volumes of 2C9r and 2C9o become exactly the same. The reason for this is that the channel to the cavity is wider in 2C9r than in 2C9o and therefore becomes included in the cavity if a small probe radius is used. Therefore, we can conclude that the cavities of the 2C8, 2C9, and 3A4 structures are approximately of the same size. On the other hand, the cavity in the 2A6 structure is much smaller (79 Å^3) and, in this case, the difference is real. Finally, it is notable that the cavity sizes of the two 3A4 structures differ by 42 Å^3, which gives a rough idea of the accuracy of the volume estimates.

TABLE 2: Number of Water Molecules Found in the Active-Site Cavity during the MD Simulations

	2A6	2C8	2C9o	2C9r[a]	3A4t	3A4w
average	2	41	58	54	57	54
minimum	2	24	35	44	25	31
maximum	2	55	70	68	77	76
standard deviation	0	6	6	5	9	8
solvent-accessible volume[b] (Å^3)	83	441	484	819	512	554
solvent-accessible volume[c] (Å^3)	197	711	520	797	717	419
water volume[d] (Å^3)	60	1236	1740	1620	1703	1608

[a] Only structures for which the distal part of the active-site cavity (yellow in Figure 4) is connected to the rest of the cavity (orange in Figure 4) are included in this analysis. If all structures are included, the average, minimum, and standard deviation become 41, 20, and 11, respectively. [b] Calculated by the Pocket[55] software from the crystal structure. [c] Calculated by the Pocket[55] software at the end of the simulation. [d] The average number of water molecules multiplied by the average volume of a water molecule in bulk water, 30 Å^3.

In Table 1, we also list the number of water molecules in the active-site cavity observed in the crystal structure. Apparently, there is little correlation between the volume of the active-site cavity and the number of crystal water molecules in it. Instead, the number of reported water molecules seems to correlate with the resolution of the crystal structure, because 3A4t, which has an appreciably better resolution than the other structures (except 2A6, and 2C9r, the cavity of which are filled with a substrate), has the highest number of water molecules. However, this can also be because this structure has a charged residue, Arg-212, that points into the active-site cavity and thereby may stabilize a hydrogen-bond network between the water molecules.

With the exception of the 3A4t structure, all of the other structures give the impression that the active-site cavity is at least partly empty (the average volume of a water molecule in bulk water is ∼30 Å^3). We will see that this only reflects the fact that the water molecules are not discerned in the crystal structures, because they are too mobile and do not bind at the same position all the time.

Number of Water Molecules in the Simulations. We have analyzed the number of water molecules in the active-site cavity during the simulations as described in the Methods section. The results are collected in Table 2 and show that the 2C9 and 3A4 proteins have the largest average number of cavity water molecules, 54−58. The 2C8 protein has a somewhat lower number of water molecules (41), whereas the 2A6 protein is completely different from the other proteins, showing a minimal cavity with only 2 water molecules.

It should be noted that, in 2C9r, the active site is divided into two parts by Phe-476. In ∼60% of the snapshots, the two parts are not connected, giving a rather low number of water molecules in the active-site cavity (∼35). However, we decided to include also the other part of the cavity in the calculations, to make the results more comparable with those of the other structures. In the 2C9o structure, Phe-476 has another confor-

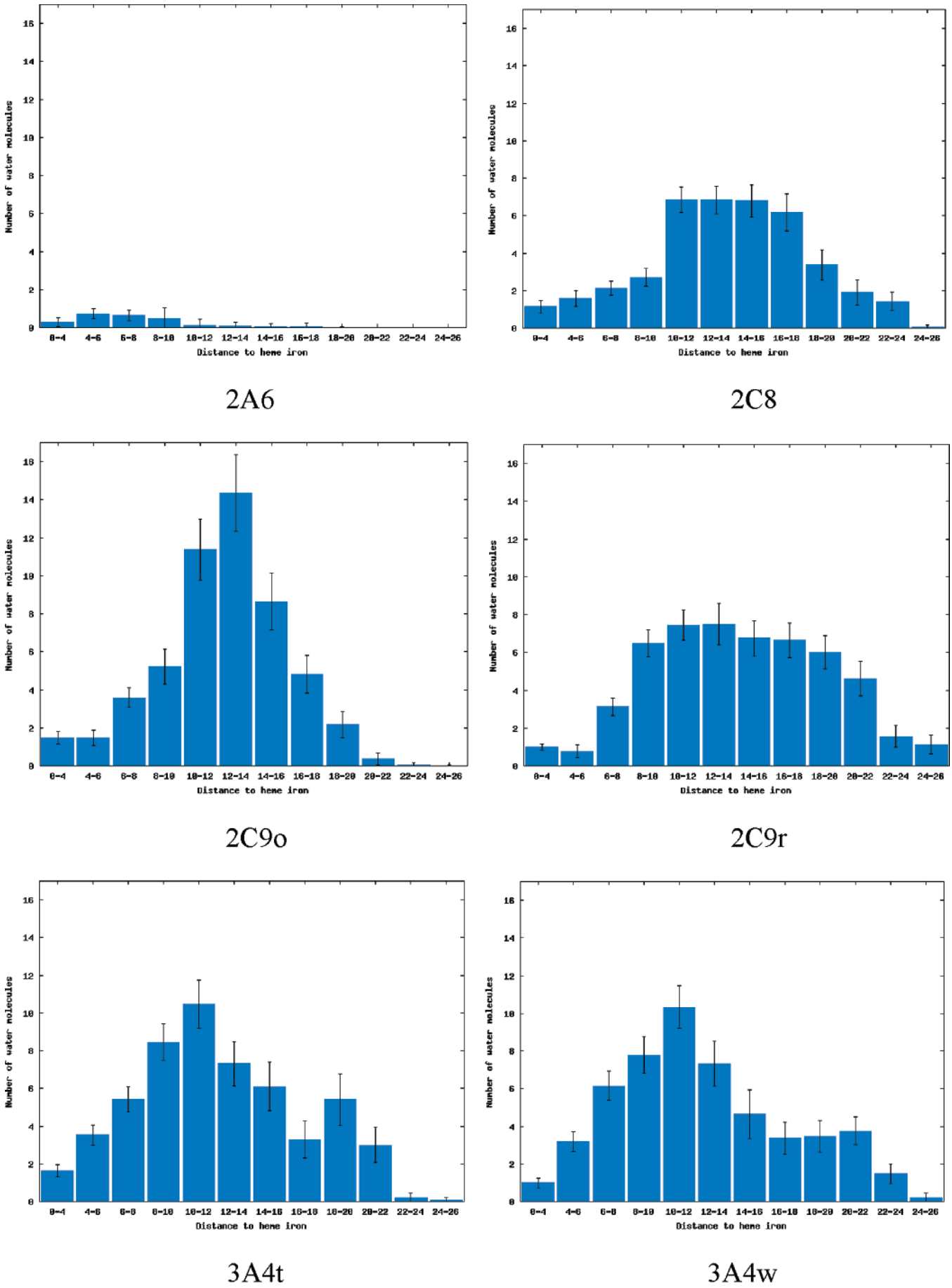

Figure 3. Distance distributions of the cavity water molecules. The number of water molecules has been plotted against the distance from the iron ion (Fe−O distance). The bars show the standard deviation in the snapshots from the MD simulations.

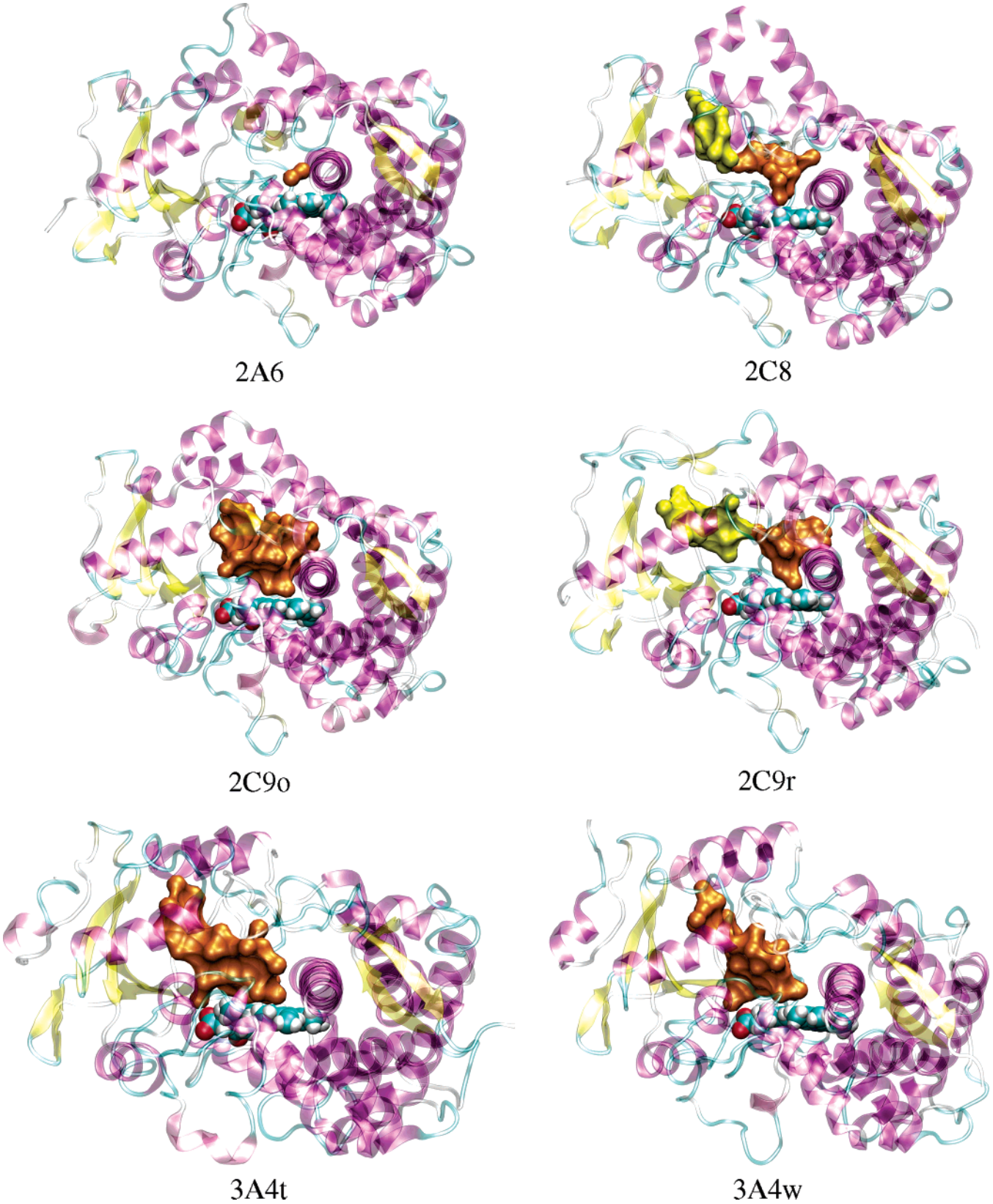

Figure 4. Picture of the active-site cavities (orange) in a typical snapshot of each protein. In 2C8 and 2C9r, the active-site cavity is split into two parts and the distal one is shown in yellow. The figures were created using VMD[61] and PovRay.[69] The molecular surface of the cavities was obtained from the cavity water molecules using the SURF algorithm[70] (with a probe radius of 2.0 Å) as implemented in VMD.[61]

mation and does not split the cavity. This change in conformation is caused by the interaction with Phe-100, which also has different conformations in the two structures. Both groups are close to amino acids 220−223, which were mutated in 2C9o, indicating that those mutations may cause the change in conformation and cavity shape. The conformation of 2C9o does not change when a substrate is bound.[5] In fact, the active-site cavity is split into two parts also in the 2C8 protein (by Ile-476), but in this protein, the two parts are always connected.

We have estimated the density of the water molecules in the

active-site cavity by calculating their surface-accessible volume and comparing it with that of the same number of water molecules in bulk water. The results indicate that the density is 15−25% lower in the active-site cavities of 2C8, 2C9, and 3A4 than in bulk water, whereas the cavity in 2A6 is large enough to room twice as many water molecules.

These results give quite a different picture of the active-site cavities than the crystal structures. First, there are many water molecules in the active-site cavity of all proteins except 2A6. In principle, the cavities are filled with water, with a density slightly lower than that in bulk water. This is of course important to know and model in theoretical simulations of the protein. It is also important for the interpretation of binding data of the proteins.

Second, the number of water molecules varies significantly among the various snapshots, showing that waters move in and out of the cavity in a rapid time scale. For example, the solvent-accessible volume estimated by the Pocket software at the end of the simulation (after removal of the water molecules) has changed by 22−270 Å^3 compared to the estimate in the crystal structure (Table 2) and similar differences are also observed for the various snapshots from the MD simulations, although the algorithm fails to give a reliable estimate of the volume for many structures, owing to the problems discussed in the Methods section (this is why we did not use this software to study the cavity in the simulation).

Third, the 2A6 protein shows only a minimal active-site cavity and in the 2C8 protein the active-site cavity is 25% smaller than that in the other two proteins.

Fourth, we see that the solvent-accessible volume is a poor estimate of the volume available to molecules like water and substrates. If we multiply the average number of water molecules in the cavity for all proteins with the average volume of a water molecule in bulk water (∼30 Å^3), the resulting volumes are 2−3 times larger than the solvent-accessible volume estimated by the Pocket[55] software. The reason for this is that part of the volumes of the water molecules are assigned to the protein, as can be seen in Figure 2. Thus, the molecular volume gives a much better estimate of the available volume.

In conclusion, cavity volumes are hard to reliably estimate, there are different definitions, and various methods (software) give different results. Undoubtedly, our estimates, based on the average number of water molecules during long MD simulations, provide a reliable estimate of the volume available for waters or a substrate. However, it must always be recognized that the number varies, owing to dynamic effects (the standard deviation is 5−9, except in 2A6).

We have tried to correlate the number of water molecules in the active site with many different properties (charge, hydrophobicity, charge per volume, etc.) of the amino acids aligning the active-site cavities but without finding any clear correlation.

The Shape of the Active-Site Cavity. The volume of the active-site cavity is interesting when trying to understand the binding of various substrates and inhibitors to the protein, but the shape of the cavity is also important. Therefore, we have studied the distribution of the cavity water molecules at various distances from the heme iron ion (Fe−O distance) in the MD simulations. The results in Figure 3 show that the cavities have quite different shapes. All cavities have a maximum width around 10−14 Å, except for the small cavity in 2A6. However, in the 2C9 proteins, the size falls smoothly off toward both smaller and larger distances, whereas the distribution is more uneven and flat in 2C8. The 3A4 cavity shows a waist at a distance of 16−18 Å, because the cavity extends out under the F′ helix and becomes larger again once it is beyond the Phe cluster, which forms a roof over most of the cavity.

TABLE 3: Distance of the Water Molecule Closest to the Heme Iron Ion in the MD Simulations (Fe−O Distance in Å)

	2A6	2C8	2C9o	2C9r	3A4t	3A4w
average	5.0	3.6	3.3	3.4	3.4	3.6
minimum	2.5	2.5	2.6	2.4	2.5	2.5
maximum	9.1	6.0	5.4	5.9	4.6	5.2
standard deviation	1.5	0.5	0.3	0.5	0.4	0.4

An even clearer picture of the cavities is given in Figure 4, which shows a typical snapshot of each of the six proteins, with the cavity water molecules displayed as a surface (of course, the details in these pictures will change if a different snapshot is used, but the snapshots were selected to contain the average number of water molecules). These figures show that the shape of the cavity is similar in all proteins (except 2A6) close to the heme group, whereas the distal part of the cavity and its extension into channels differ extensively.

Interestingly, the shape of the cavity in the two 2C9 proteins is quite different, as has been noted before:[6] In 2C9o, the cavity falls off sharply at both long and short distances from the heme group, whereas the distribution is more even in 2C9r. The reason for this is mostly in the region of the cavity furthest away from the I helix. In this region, the 2C9o structure was heavily mutated and some amino acids were not discernible in the 2C9r structure. The higher number of water molecules at large distances in the 2C9r simulation comes from the part of the active site beyond Phe-476 (yellow in Figure 4), which is situated between two channels close to the β_1 sheet (pw2b and pw2d; the naming of the helices and loops is shown in Supporting Information Figure S4a) and which is sometimes connected to the main cavity, as mentioned above.

Finally, the error bars in Figure 3 show the standard deviation in the number of water molecules at various distances. They illustrate the dynamics of the waters at various distances. It can be seen that the 2A6 and 2C8 simulations have rather low and uniform standard deviations. This indicates that the cavity is well-defined at all distances, with small channels to the bulk solvent. On the other hand, the 2C9o simulations show a large variation, especially at intermediate distances (10−18 Å) from Fe, where the access channels are found.

Water Molecule close to Heme. In the resting state of cytochrome P450, the iron ion is in the low-spin Fe(III) state with a water molecule coordinated as a sixth ligand at a distance of ∼2.3 Å. However, only in two of the crystal structures (2C9r and 3A4t) is a sixth ligand of iron observed, a solvent molecule 2.44 and 2.07 Å from Fe, respectively (cf. Table 1).[6,7] Therefore, we modeled the iron ion in the five-coordinate high-spin Fe(III) state instead. As mentioned in the Methods section, our heme parameters give an optimum Fe−O distance of 3.42 Å in the absence of the surrounding protein.

From Figure 3, it can be seen that less than 1.5 water molecules on average are found closer than 4.0 Å from the heme iron in all simulations. In Table 3, we elaborate further on this measure by calculating the distance between Fe and the closest water molecule in the various snapshots of the MD simulations. It can be seen that a sixth iron ligand is not observed in any of the proteins, although the closest water molecule occasionally approaches the iron ion (down to a distance of 2.4 Å). In five of the proteins, the average Fe−O distance is ∼3.5 Å, whereas, in 2A6, it is as much as 5.0 Å. Thus, it is clear that there are no water molecules close to the iron ion in the high-spin Fe(III) state.

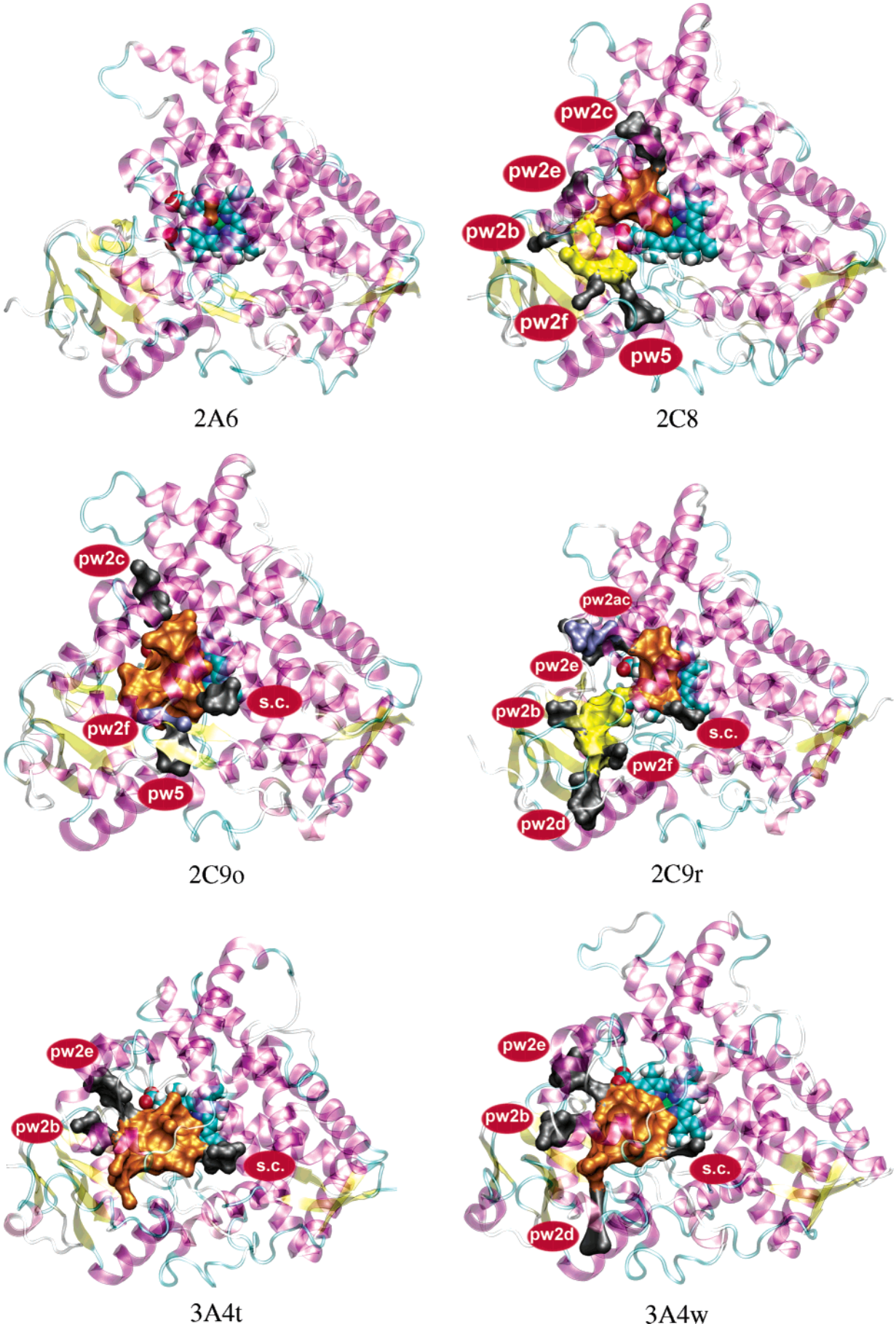

Figure 5. Channels (with names) observed in the various simulations. The channels are colored black (blue for pw2f in 2C9o and pw2ac in 2C9r). The active-site cavity is orange (and yellow for the distal part in 2C8 and 2C9r), and the heme group is atom-coded in blue, red, and green.

TABLE 4: Number of Water Molecules Exchanging through Each Channel during the MD Simulations (A "−" Means the Channel Does Not Exist in That Protein)

	2A6	2C8	2C9o	2C9r	3A4t	3A4w
pw2ac	−	−	−	10	−	−
pw2b	−	2	−	29	19	11
pw2c	−	4	14	−	−	−
pw2d	−	−	−	29	−	1
pw2e	−	5	−	18	26	19
pw2f	−	2	1	0	−	−
pw4	−	−	−	−	1	−
pw5	−	8	4	−	−	−
solvent channel	−	−	29	39	11	0
sum	0	21	48	125	57	31

Solvent-Access Channels. Many of the water molecules show sizable dynamics, and several of the cavity water molecules exchange with bulk. This exchange takes place through a number of channels, which show different extensions in the four proteins. It is highly likely that at least some of these channels are used also for the transport of substrates into and products out of the active site. Therefore, it is of great interest to characterize these channels. Wade et al.[29] have identified channels in all available (March 2006) crystal structures of cytochromes P450 using the CAVER[64] software. To facilitate the comparison with their results, we have adopted their naming of the channels. The identified channels are shown in Figure 5, and statistics of the usage of the various channels for exchange with the bulk are collected in Table 4.

The 2A6 protein is quite different from the others in that it is completely closed and no channels are visible in the crystal structure. During the simulation, no channel from the active-site cavity to the bulk opens. The pw2e channel, which runs through the BC loop, exists, but it connects to a side cavity above the loop from the K helix to the β_2 sheet (the naming of the helices and sheets is shown in Supporting Information Figure S4b). The active-site cavity is separated from the pw2e channel by the Phe-118 residue. This means that it is likely that the water molecules in the active-site cavity are not in equilibrium with bulk water for this protein (i.e., that the estimated number of water molecules in the cavity is not completely reliable). However, it is certain that the cavity of this protein is much smaller than that for the other proteins and the exchange of water molecules with the bulk is slower.

In 2C8, we observe the pw2b channel, which passes between the BC loop and the β_1 sheet, the pw2c channel, which goes between the B′−C loop and the I and G helices, and the pw2e channel. In addition, we also observe two channels, not seen in the crystal structure:[29] pw2f and pw5. pw2f runs between the F′ helix, the A helix, and the loop in the β_4 sheet, whereas pw5 runs between the A helix, the loop in the β_4 sheet, and the loop from the K helix to the β_2 sheet (note that pw5 differs from the one described by Wade[29] in that it exits above the loop from helix K to the β_2 sheet instead of below it). The pw2c channel is in the same region as the cluster of water molecules observed close to the carboxylate group of the substrate diclofenac in the structure of rabbit P450 2C5.[65] We also see a large cleft on the surface of the protein where the solvent channel should be, but this crevice is not connected to the active-site cavity at any time during the simulation. As mentioned above, the active-site cavity is almost cut off before it reaches the pw2b channel and then extends (yellow in Figure 5) from the pw2b channel to the pw4 channel. Only the pw2c and pw2e channels are directly connected to the part of the active-site cavity that is close to heme. The pw2c and pw2e channels are used equally frequently for water transport (Table 4). The pw2f channel opens by a conformational change of the side chain of Val-47 only during the last third of the simulation.

In the 2C9o protein, there are four channels, the solvent channel, which runs on top of the I helix, between the loops from the F′ to F helices and the loop in the β_4 sheet, pw2c, pw2f, and pw5 (also in this protein, pw5 exits above the loop from helix K to the β_2 sheet). The solvent channel is used twice as frequently as the pw2c channel, whereas the pw5 channel is used only four times during the 4 ns simulation. We also observe a single water molecule passing through the pw2f channel. In the 2C9r protein, on the other hand, we observe the pw2b, pw2d, pw2e, pw2f, solvent, and pw2ac channels but not the pw5 channel. The pw2d channel runs between the N-terminus and helix A, whereas the pw2ac channel is almost on top of the pw2e channel, between the BC loop and the G helix. Many water molecules exchange through the pw2b and pw2d channels, but most of them end up in the distal part of the active-site cavity (yellow in Figure 5). Most of the exchange between the bulk and the active-site cavity close to the heme group occurs through the solvent channel and the pw2e and pw2ac channels. The pw2f channel is never used, because a strongly hydrogen-bonded water molecule blocks it. The pw2b, pw2d, and pw5 channels are not observed in the crystal structures.[29]

In the 3A4 proteins, we observe three channels in both simulations, namely, the pw2b, pw2e, and solvent channels. However, the solvent channel is only used for exchange in 3A4t. In 3A4w, the solvent channel is disconnected from the active-site cavity most of the time, and when it is connected, it is so narrow that it does not allow for any exchange of water. On the other hand, we also find the pw2d channel in 3A4w, which opens for transport once. In 3A4t, the pw4 channel that runs between the F′ and G′ helices opens at a single occasion. Neither of these two channels are observed in the crystal structures.[29] In both 3A4 proteins, the pw2e channel is used almost twice as often as the pw2b channel for exchanging cavity water molecules with bulk solvent. In the simulation of 3A4t, the solvent channel is used half as often as the pw2b channel.

In general, we find that the pw2f and pw5 channels are not used much for exchange with bulk. This is probably because they are longer and narrower than the other channels. Oprea et al.[30] have suggested that there may exist another channel between an Arg residue and one of the heme propionate groups on the side of the heme group opposite to the active-site cavity (the water channel). In our simulations, we do not observe this channel at any time, not even for 2A6, which is most similar to the structures studied by Oprea et al.

There is little correlation between the polarity of the various channels and their use for water exchange. However, channels in which the water molecules have many hydrogen bonds to the protein are often not used for exchange, at least not in the time scale of our simulations.

Mobility of Water Molecules. To study the mobility of the water molecules, we determined the root-mean-squared deviation of all water molecules found in the active-site cavity at any time during the simulations with respect to the position in the first snapshot after equilibration (RMSD). The results are shown in Figure 6, in which the water molecules are color-coded according to their RMSD: Water molecules with a RMSD greater than 15 Å (black) are in the bulk at some point during the simulation, those with a RMSD of 7−15 Å (blue) may be in the bulk at some part of the simulation or they may move around extensively in the active site or the access channels, those with a RMSD of 3−7 Å (green) do not move much and remain inside the active-site cavity or the access channel

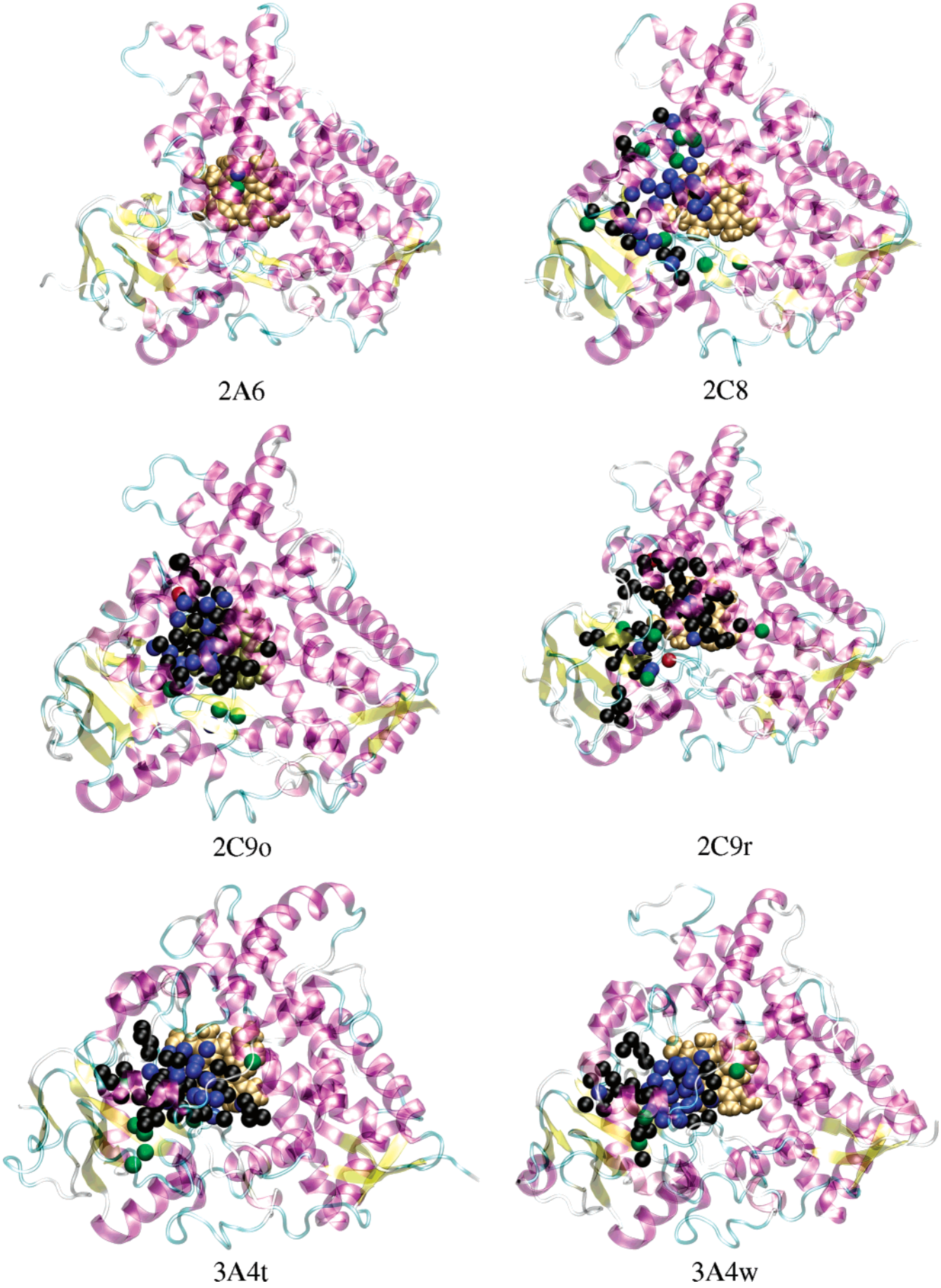

Figure 6. The first equilibrated snapshot from each protein with water molecules color-coded according to their RMSD during the simulations: <3 Å, red; 3−7 Å, green; 7−15 Å, blue; >15 Å, black.

during the entire simulation, whereas those with a RMSD of less than 3 Å (red) remain in the same position during the whole simulation.

In the 2A6 protein, only two water molecules are found in the active-site cavity. None of them exchange with the bulk during the simulation time, but they have a relatively high mobility (RMSDs of 5.3 and 7.3 Å).

In 2C8, most active-site water molecules have a quite high mobility (RMSD 7−15 Å). However, there are nine water molecules that show a lower mobility. These water molecules are confined by steric hindrance or hydrogen bonds. Water molecules in the distal part of the cavity and in the channels naturally have a higher mobility.

Water molecules in the two 2C9 simulations show a higher mobility than for the other proteins (a larger proportion of water molecules with a RMSD larger than 15 Å). Only four and six molecules have a RMSD below 7 Å in 2C9o and 2C9r, respectively. The least mobile one in 2C9o forms hydrogen bonds to the side chain of Asp-264 and the backbone nitrogen atoms of Phe-114 and Gly-111. One of the green ones in 2C9o is confined by steric hindrance during most of the simulation, whereas the other two green ones are confined into a small side cavity. In 2C9r, the least mobile molecules (the two red ones) are confined into small cages of amino acids with multiple hydrogen bonds. One of the green molecules is blocking the pw2f channel. Two other molecules are confined into small pockets at the side of the pw2b and solvent channels. The two proteins differ in that there are more mobile water molecules in 2C9r, because there is a larger exchange with the bulk.

The mobility of water molecules in the 3A4 simulations is quite similar to that of 2C8. Most water molecules have a quite high mobility, but there are also some with a lower mobility. The green one to the right in Figure 6 of 3A4w and 3A4t (RMSD of 3.0 and 3.1 Å, respectively) forms a kink in the I helix by interrupting the hydrogen bond between Phe-304 and Glu-308. Such a kink in helix I, caused by an inserted water molecule, has been observed in all high-resolution structures of bacterial P450s, although at slightly varying positions.[34] In our simulations, such kinks are observed for 2C8 and 3A4 but not for the other two proteins. There are also clusters of water molecules with a low mobility at the bottom left of the 3A4 figures. They are simply confined in the far end of the active-site cavity, close to the pw2d channel, and have nowhere to go, although a few molecules in 3A4w managed to escape.

In bacterial P450s, four well-ordered water molecules in the active site have been suggested to be important for the catalytic activity.[66] From Figure 6, it can be seen that such a network is not conserved in any of the human enzymes. Of course, the reason for this can be that we do not study a substrate-bound compound I state (the supposedly enzymatically active state), but the results clearly show that the water structure is not ordered in the high-spin ferric state to any extent, not even close to the heme group.

Conclusions

In this paper, we have studied the dynamics of water molecules in the active-site cavity of six crystal structures of four human cytochromes P450 (2A6, 2C8, 2C9, and 3A4) using MD simulations. These simulations have provided a large amount of useful information.

First, we have seen that it is hard to unambiguously delimit the active site in these proteins, because the cavity is dynamic, there are many channels to the surroundings with different sizes, and there are other cavities in the protein that occasionally may merge with the active-site cavity. Different definitions of the cavity may lead to widely different estimates of the active-site cavity volume, for example, 520 and 1386 Å^3 for the two 3A4 structures.[7,8] We do not claim that our definition of the active-site cavity is better than any other; on the contrary, we have noted several problems with it, but these can be identified by statistical methods in our extensive material of sampled snapshots. Our investigation has the advantage of studying explicit water molecules, of including the dynamics of the protein, and of using the same well-defined algorithm for all of the crystal structures of human P450s.

By this approach, we have estimated the number of water molecules in the active-site cavity in the six proteins during a 4 ns MD simulation. The results (in Table 2) show that the cavity is completely filled with water molecules, in contrast to what is observed in most crystal structures.[5,6,7,8,10] However, although the density is ∼20% lower than that in bulk water. The 2A6 protein differs strongly from the three others in that the active-site cavity is very small, containing only two water molecules, which do not exchange with bulk water during the simulation time. The 2C8, 2C9, and 3A4 proteins have much larger cavities with 41 (2C8) or 54−58 (2C9 and 3A4) water molecules on average. If we assume that the volume of each water molecule is 1.25 times the average volume of a water molecule in bulk water, this means that the volume of the active-site cavity of these proteins is 1500−2100 Å^3.

The active-site cavity is connected to the water solution by three to six channels in the various proteins (none for 2A6). All of them have been observed before for human or bacterial P450s.[29−36] However, in all structures (except 2A6), we observe channels that are not apparent in the crystal structures[29] but open only during the simulations (pw2f and pw5 in 2C8, pw2b, pw2e, and pw5 in 2C9, and pw2d and pw4 in 3A4). This shows that dynamic effects are important for the channels.

In all proteins, except 2A6, these channels are used quite extensively for the exchange of water molecules with the bulk: On average, a water molecule passes one of the channels every 30−200 ps (more seldom in 2C8 than in the other two proteins; it is possible that these values are somewhat too low, because the TIP3P water model gives a self-diffusion coefficient that is almost twice as large as the experimental one[67]). 2C8 uses mostly the pw2c and pw2e channels, and 3A4 uses mostly the pw2e and pw2b channels; however, the 2C9 proteins use the solvent channels extensively (and also the pw2c channel for 2C9o and the pw2e channel for 2C9r). It is most likely that some of these channels are also used for the transport of the substrate into and the product out of the proteins. Thus, we have seen that simple MD simulations constitute a effective method to identify channels, at least in these proteins.

Finally, the simulations have also enabled us to study the mobility of the water molecules in the human P450s. It turned out (Figure 6) that most water molecules in the active-site cavity are quite mobile. Only three water molecules in the two 2C9 structures remain in a single position (RMSD of less than 3 Å). Thus, there is no conserved water structure in the active site of these studied high-spin ferric states. In particular, there are no water molecules close to the catalytic iron ion (the average distance to the water molecule closest to the iron ion is 3.3−5 Å).

In conclusion, our MD simulations show that the 2A6 protein differs considerably from the other three studied enzymes with a very small and closed active-site cavity. The active-site cavity in 2C8 is ∼25% smaller than that in the other two proteins, whereas the 2C9 and 3A4 proteins are quite similar and the

differences between the two proteins are not larger than between the two different structures of the 2C9 and 3A4 enzymes. In particular, the two crystal structures of 2C9 seem to be quite dissimilar with quite different shapes of the cavity and with extensive differences in the channels to the active sites. It is likely that these differences are caused by the mutations employed to enhance the crystallization (especially for the 2C9o structure).[5] The two 3A4 structures have similar shapes of the cavity, although they differ slightly in the structure of the channels. However, it should be noted that two very recent structures of 3A4 with bulky ligands show an increased volume of the active-site cavity.[68] This indicates that binding of a ligand or the change of the buffer polarity may induce quite large conformational changes in the cytochromes P450. However, this does not change the general conclusion that the active-site cavity is filled with water, and we can also visually identify the same channels in these two new structures as in the 3A4t and 3A4w structures.

Acknowledgment. This investigation has been supported by funding from the research school in pharmaceutical science (FLÄK), the Swedish Research Council, the Carlsberg Foundation, and by computer resources of the High Performance Computing Center North (HPC2N) at the University of Umeå.

Supporting Information Available: Protonation status of the histidine residues in the various structures, charges of the atoms in the heme group and the Cys ligand used in the simulations, number of atoms and periodic box size for the various proteins, RMSD and radius of gyration in the various simulations, VOIDOO volumes for the 3A4t crystal structure, the active-site model for charge calculations, time evolution of RMSD and radius of gyration, definition of various water molecules by cavity_wat, and nomenclature of helices and sheets. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Graham-Lorence, S.; Peterson, J. A. *FASEB J.* **1996**, *10*, 206−214.

(2) Li, H. In *Handbook of Metalloproteins*; Messerschmidt, A., Huber, R., Poulos, T., Wieghart, K., Eds.; J. Wiley & Sons: Chichester, U.K., 2001; pp 486−502.

(3) Hodgson, J. *Nat. Biotechnol*. **2001**, *19*, 722−726.

(4) Poulos, T. L.; Finzel, B. C.; Gunsalus, I. C.; Wagner, G. C.; Kraut, J. *J. Biol. Chem.* **1985**, *260*, 16122−16130.

(5) Williams, P. A.; Cosme, J.; Ward, A.; Angove, H. C.; Vinkovic, D. M.; Johti, H. *Nature* **2003**, *424*, 464−468.

(6) Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. *J. Biol. Chem.* **2004**, *279*, 35630−35637.

(7) Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. *J. Biol. Chem* **2004**, *279*, 38091−38094.

(8) Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. *Science* **2004**, *305*, 683−686.

(9) Deleted in proof.

(10) Schoch, G. A.; Yano, J. K.; Wester, M. R.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. *J. Biol. Chem.* **2004**, *279*, 9497−9503.

(11) Yano, J. K.; Hsu, M.-H.; Griffin, K. J.; Stout C. D.; Johnson, E. F. *Nat. Struct. Mol. Biol.* **2005**, *12*, 822−823.

(12) McLean, M. A.; Maves, S. A.; Weiss, K. E.; Krepich, S.; Sligar, S. G. *Biochem. Biophys. Res. Commun.* **1998**, *252*, 166−172.

(13) Rashnin, A. A.; Iofin, M.; Honig, B. *Biochem.* **1986**, *25*, 3619−3625.

(14) Ernst, J. A.; Clubb, R. T.; Zhou, H. X.; Groenenborn, A. M.; Clore, G. M. *Science* **1995**, *267*, 1813−1817.

(15) Zhang, L.; Hermans, J. *Proteins: Struct., Funct., Genet.* **1996**, *24*, 433−438.

(16) Roux, B.; Nina M.; Pomès, R.; Smith, J. C. *Biophys. J.* **1996**, *71*, 670−681.

(17) Denisov, V. P.; Halle, B. *Faraday Discuss.* **1996**, *103*, 227−244.

(18) Hubbard, S. J.; Gross, K.-H.; Argos, P. *Protein Eng.* **1994**, *7*, 613−626.

(19) Williams, M. A.; Goodfellow, J. M.; Thornton, J. M. *Protein Sci.* **1994**, *3*, 1224−1235.

(20) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *Biochemistry* **1986**, *25*, 5314−5322.

(21) DiPrimo, C.; Hoa, G. H. B.; Douzou, P.; Sligar, S. *Eur. J. Biochem.* **1990**, *193*, 383−386.

(22) Helms, V.; Wade, R. C. *Proteins: Struct., Funct., Genet.* **1998**, *32*, 381−396.

(23) Helms, V.; Wade, R. C. *J. Am. Chem. Soc.* **1998**, *120*, 2710−2713.

(24) Poulos, T. L. *Biochem. Biophys. Res. Commun*. **2005**, *338*, 337−345.

(25) Guallar, V.; Baik, M.-H.; Lippard, S. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6998−7002.

(26) Gullar, V.; Friesner, R. A. *J. Am. Chem. Soc.* **2004**, *126*, 8501−8508.

(27) Scott, E. E.; He, Y. A.; Wester, M. R.; White, M. A.; Chin, C. C.; Halpert, J. R.; Johnson, E. F.; Stout, C. D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13196−13201.

(28) Podust, L. M.; Kim, Y.; Arase, M.; Neely, B. A.; Beck, B. J.; Bach, H.; Sherman, D. H.; Lamb, D. C.; Kelly, S. L.; Waterman, M. R. *J. Biol. Chem.* **2003**, *278*, 12214−12221.

(29) Cojocaru, V.; Winn, P. J.; Wade, R. C. *Biochim. Biophys. Acta* **2007**, *1770*, 390−401.

(30) Oprea, T. I.; Hummer, G.; García, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2133−2138.

(31) Lüdemann, S. K.; Lounnas, V.; Wade, R. C. *J. Mol. Biol.* **2000**, *303*, 797−811.

(32) Lüdemann, S. K.; Lounnas, V.; Wade, R. C. *J. Mol. Biol.* **2000**, *303*, 813−830.

(33) Winn, P. J.; Lüdemann, S. K.; Gauges, R.; Lounnas, V.; Wade, R. C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5361−5366.

(34) Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A. *Biochemistry* **2001**, *40*, 13456−13465.

(35) Wade, R. C.; Winn, P. J.; Schlichting, I.; Sudarko. *J. Inorg. Biochem.* **2004**, *98*, 1175−1182.

(36) Schleinkofer, K.; Sudarko, J.; Winn, P. J.; Lüdemann, S. K.; Wade, R. C. *EMBO Rep.* **2005**, *6*, 584−589.

(37) Loida, P. J.; Sligar, S. G. *Biochemistry* **1993**, *32*, 11530−11538.

(38) Park, H.; Lee, S.; Suh, J. *J. Am. Chem. Soc.* **2005**, *127*, 13634−13642.

(39) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(40) Deleted in proof.

(41) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. *Nucleic Acids Res.* **2003**, *31*, 3381−3385.

(42) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.

(43) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(44) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kolman, P. A. *AMBER 8*; University of California, San Francisco, 2004.

(45) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431−439.

(46) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Knudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.9; Gaussian, Inc.: Pittsburgh PA, 1998.

(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(48) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(49) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(50) Kleywegt, G. J.; Jones, T. A. *Acta Crystallogr.* **1994**, *D50*, 178−185.

(51) Hubbard, S. J.; Argos, P. *Protein Eng.* **1995**, *8*, 1011−1015.

(52) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379−400.

(53) Connolly, M. L. *Science* **1983**, *221*, 709−713.
(54) Liang, J.; Edelsbrunner, H.; Fu, P.; Sudhakar, P. V.; Subramaniam, S. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 18−29.
(55) http://www.cs.ucdavis.edu/∼koehl/ProShape/.
(56) Hendlich, M.; Ripmann, F.; Barnickel, G. *J. Mol. Graphics* **1997**, *15*, 359−363.
(57) Chakravarty, S.; Bhinge, A.; Varadarajan, R. *J. Biol. Chem*. **2002**, *277*, 31345−31353.
(58) Laurie, A. T. R.; Jackson, R. M. *Bioinformatics* **2005**, *21*, 1908−1916.
(59) R: A Language and Environment for Statistical Computing, R development core team, ISBN 3-900051-07-0.
(60) Liang, J.; Edelsbrunner, H.; Woodward, C. *Protein Sci.* **1998**, *7*, 1884−1897.
(61) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(62) Scott, E. E.; Halpert, J. R. *Trends Biochem. Sci.* **2005**, *30*, 5−7.
(63) Liang, J.; Edelsbrunner, H.; Fu, P.; Sudhakar, P. V.; Subramaniam, S. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 1−17.
(64) Petrek, M.; Otyepka, M.; Banas, P.; Kosinva, P.; Koca, J.; Damborsky, J. *BMC Bioinformatics* **2006**, *7*, 316−324.
(65) Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dijols, S.; Dansette, P. M.; Mansuy, D.; Stout, C. D. *Biochemistry* **2003**, *42*, 9335−9345.
(66) Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science* **2000**, *287*, 1615−1622.
(67) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2001**, *114*, 363−366.
(68) Ekroos, M.; Sjögren, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13682−13687.
(69) Persistence of Vision Pty. Ltd. Persistence of Vision Raytracer, 2004 (version 3.6), http://www.povray.org/download.
(70) Varshney, A.; Brooks, F. P.; Wright, W. V. *IEEE Comput. Graphics* **1994**, *14*, 19−25.