JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Funnel-Metadynamics and Solution NMR to Estimate Protein−Ligand Affinities

Laura Troussicot,[§] Florence Guillière,[§] Vittorio Limongelli,[¶,#] Olivier Walker,[§] and Jean-Marc Lancelin*[,§]

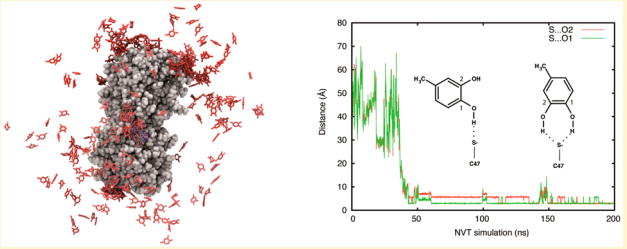
[§]Institut des Sciences Analytiques, UMR 5280, CNRS, Université de Lyon, Université Lyon 1, ENS Lyon -5, rue de la Doua, F-69100 Villeurbanne, France

[¶]Faculty of Informatics, Institute of Computational Science, Università della Svizzera Italiana (USI), via G. Buffi 13, CH-6900 Lugano, Switzerland

[#]Department of Pharmacy, University of Naples "Federico II", via D. Montesano 49, I-80131 Naples, Italy

Supporting Information

ABSTRACT: One of the intrinsic properties of proteins is their capacity to interact selectively with other molecules in their environment, inducing many chemical equilibria each differentiated by the mutual affinities of the components. A comprehensive understanding of these molecular binding processes at atomistic resolution requires formally the complete description of the system dynamics and statistics at the relevant time scales. While solution NMR observables are averaged over different time scales, from picosecond to second, recent new molecular dynamics protocols accelerated considerably the simulation time of realistic model systems. Based on known ligands recently discovered either by crystallography or NMR for the human peroxiredoxin 5, their affinities were for the first time accurately evaluated at atomistic resolution comparing absolute binding free-energy estimated by funnel-metadynamics simulations and solution NMR experiments. In particular, free-energy calculations are demonstrated to discriminate two closely related ligands as pyrocatechol and 4-methylpyrocathecol separated just by 1 kcal/mol in aqueous solution. The results provide a new experimental and theoretical basis for the estimation of ligand−protein affinities.



Fundamental properties of proteins include their capabilities to interact selectively with other molecules for particular functions as enzyme catalysis or specific receptors signaling, for instance.[1] The structural dynamics of proteins are of prime importance for their interacting properties, and the conformational changes induced upon the binding of a ligand may be the basis of their biological regulations widely referred as allostery.[2] Since the mid 20th century, crystal structures of proteins were mostly used to describe their chemical architecture both in their free and in their different bound forms to their ligands in the solid crystal state.[3] For soluble proteins, NMR introduced the possibility to study proteins in equilibrium between their different states in solution *in vitro*[4] and even *in vivo*.[5] For fast exchanges at the NMR time scales most of the NMR observables (isotropic/anisotropic chemical shifts, *J*-couplings, NOEs, etc.) are time-averaged and weighted over the ratio of the exchanging populations.[6] Currently, the most popular molecular modeling protocols widely used to interpret the results or to predict the binding include docking algorithms which are oversimplified.[7,8] These popular algorithms do not consider the dynamic properties of proteins beyond the amino acid side chain, ignore the explicit solvent and salts in solution, and clearly lead to severe limitations and possible false or elusive conclusions.

More recently, molecular dynamics (MD) simulations, using still empirical but experienced molecular force fields, conducted in an explicit solvent cell have demonstrated to describe successfully the evolution of the system under investigation in hundreds of nanoseconds (ns) up to the microsecond (μs) time scale using massively parallel calculations on equilibrated molecular systems of more than 50,000 atoms.[9,10] These simulations reach the time scales usually detected by the common NMR observables such as chemical shift, whose perturbation can be used to estimate the protein affinity to their ligands.[11] A further advance in this field is represented by the use of enhanced sampling techniques such as metadynamics and its derived form funnel-metadynamics (FM),[12−14] that have been successful to calculate the absolute protein−ligand binding free-energy with excellent convergence criteria, while keeping track of the whole binding-unbinding process. This offers a set of completely new tools for modeling the interaction equilibrium in solution that can contribute to a bright future in the area of medicinal chemistry design of protein ligands and even in the engineering of protein interactions. Here we show for the first time, with the protein human peroxiredoxin 5 (PRX5), that a computational protocol based on FM simulations correlates

Received: November 4, 2014
Published: December 30, 2014

DOI: 10.1021/ja511336z
J. Am. Chem. Soc. 2015, 137, 1273−1281

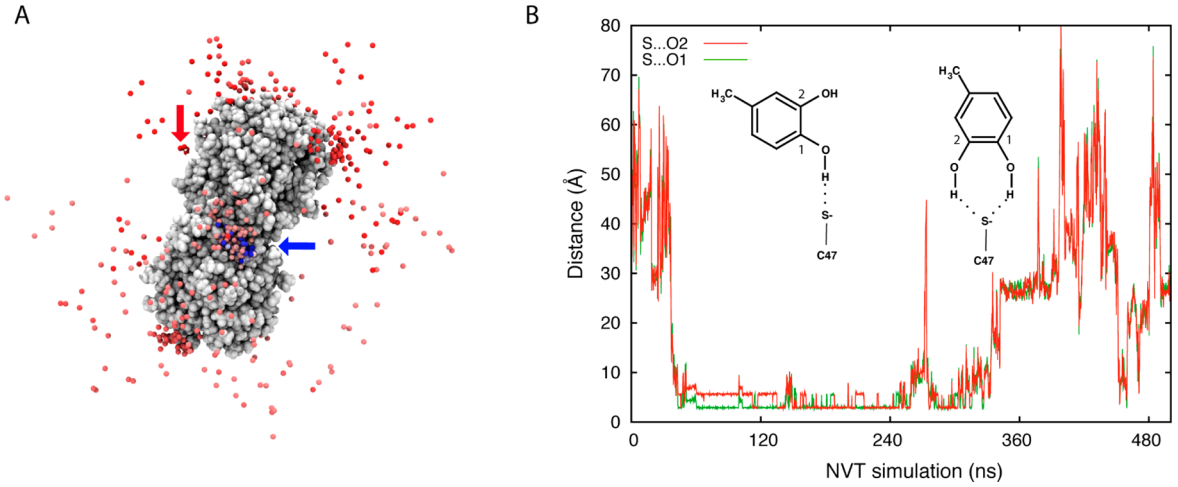


Figure 1. Unbiased NVT MD of human PRX5 and 4-methylcatechol in explicit saline (150 mM NaCl) aqueous solution where a specific binding to PRX5 is observed from a starting unbound state of 4-methylcatechol. (A) PRX5 CPK model and ball representation of O1 atom of 4-methylcatechol plotted every 100 ps steps of the first 200 ns of a 500 ns NVT production MD. The starting ligand position is indicated with a red arrow, while the final positions in the active site of the PRX chain A with a blue arrow. A color continuum from red to blue via white is applied along the MD steps for O1 atom. (B) O1 (green trace) and O2 (red trace) distances from the chain A catalytic thiolate S of C47 as the function of MD steps. 4-methyl catechol docks to the active around 40 ns with characteristic H-bonding patterns with one (different S to O1/O2 distances) or two H-bonds (similar S to O1/O2 distances).

closely to the solution NMR observables used to study the PRX5 selective affinity for simple molecules of the catechol derivatives. The catechol derivaties, pyrocatechol and 4-methylpyrocathecol, were recently found to selectively bind to the active site of PRX5 with a low-affinity. PRX5 is a member of the peroxiredoxin family involved in the peroxide homeostasis and peroxide signaling involved in crucial cell functions.[15,16] PRXs catalyze the decomposition of hydroperoxides via a complex catalytic pathway involving other proteins. The shape of the active sites of PRXs varies among the PRX family members.[17] The active site consists of central active thiolate and is thought to accept different organic peroxides along with hydrogen hydroperoxide. Our results indicate that the PRX5 affinity to their ligands can be predicted through computations with enough precision regarding the experimental NMR data in solution. The FM results confirm and complement the experimental findings, providing an extensive background information for the protein−ligand binding−unbinding process. This study provides a new way to describe the affinity equilibrium of interacting molecules in solution.

RESULTS

Unbiased MD of PRX at High Concentration of Ligands. To gain a first analysis of the protein−ligand binding and to obtain the collective variables necessary for metadynamics, two different protocols were used. For the first protocol, the homodimer was equilibrated without ligands in the active sites. Whereas the second protocol, oxidized dithiothreitol (DTT) was left in the crystal complex in one of the two active sites (chain A) of the homodimer, while the second active site was free of ligands. In both MD simulations, we positioned 6 molecules of ligands (total concentration in the water box of ∼15 mM) in the solvent near the two actives sites (chain A and chain B) at distances ranging from 15 to 25 Å (measured as the distance between the thiolate SG atom of active site Cys and one of the two oxygen atoms of ligands). A production run of 500 ns was performed to analyze the binding and unbinding events of the ligands within the protein.

In the second MD simulation, DTT unbinds in the first 10 ns from the occupied active site. As illustrated in Figure 1, one of the 6 ligands binds to the liberated active site of the chain A after ∼40 ns of free diffusion in the solvent and at the surface of the protein. The ligand binds to the active site with a RMS fluctuation of ∼1 Å for until 260 ns, then it unbinds shortly before rebinding and diffusing again in the solvent and the protein surface. The trajectories of the other ligands are given in Figure S1.

Several binding events occurred on the other active site of chain B. When the ligand is bound, two different binding modes are exchanged during the MD. The major one is strictly comparable to the positions that were described in all crystal structures of PRX for different ligands (benzoate, oxidized DTT, and catechols, see Figure S2).[18−21] This conformation shows a H-bond formed by OH at position 1 of 4-methylcatechol with an angle between the thiolate SG sulfur of Cys47, O1 and O2 atom of the ligands (S−O−O angle) of about 180° proposed to mimic the reaction transition state.[20] This conformer exchanges with another conformation where two H-bonds are formed with the two hydroxyls of the ligand, as shown in Figure 1, with a S−O−O angle of ca. 60°. In the other free active site of the homodimer, a ligand initially close to the active site binds quickly in about 5 ns, first with the double H-bonded conformation (S−O−O angle 60°), then changing into the other single H-bonded conformation (S−O−O angle 180°). These two binding conformers are also correlated with the monitoring of a protein−ligand torsional angle as Cys47SG···O1−C1−C2 (ligand). The conformers with the S−O−O angle 180° has a torsion SG···O1−C1−C2 of ±180° and those with a S−O−O angle around 60° has a torsion SG···O1−C1−C2 around 0° (see Figure S2). During the 500 ns MD run, several unbinding periods are observed ranging from a few ns to 20 ns, leaving the active site free. A direct exchange between two catechol ligands was also observed during a short period of about 2 ns before exchanging back to the first ligand.

Free-Energy Calculations Using FM Simulations. To provide more insight into the thermodynamics ruling the

DOI: 10.1021/ja511336z
J. Am. Chem. Soc. 2015, 137, 1273−1281

binding of the ligands to PRX, well-tempered metadynamics simulations were performed.[22,23] This enhanced sampling technique allows the simulation of long time scale events, such as ligand/protein binding, in a reasonable computational time duration. At the end of the calculation, the free energy landscape of the investigated process can be computed using the history-dependent bias potential added during the simulation on few degrees of freedom of the system, called collective variables (CVs). This technique has been successfully used to study complex and long time scale biological processes like DNA folding[24] and ligand−protein and ligand−DNA binding.[25−27] In the present study, in order to investigate the ligand−PRX binding process and calculate the ligand binding free energy, a recently developed metadynamics-based approach, called funnel-metadynamics (FM), was used.[14] These simulations use a funnel-restrained potential that reduces the conformational space to explore the unbound state, thus increasing the number of binding events observed during the sampling. In such a way, FM calculations lead to a quantitatively well-characterized binding free-energy surface (FES). It is important to stress that when the ligand is inside the funnel, no external potential is applied, and the system works under the standard metadynamics regime. Thus, the shape of the funnel restraint potential was purposely chosen to include the PRX binding site within the cone section (Figure 2).

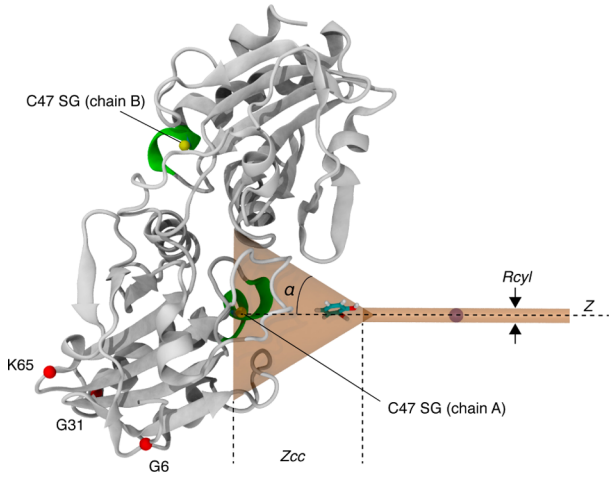


Figure 2. Funnel restraint potential applied to PRX5 enzymes with catechol ligand. The catalytic thiolate of C47 in the α-helix colored in green is represented with a yellow sphere on chains A and B of the PRX5 homodimer. Cα of G6, G31, and K65 of chain A used as positional restrains to inhibit the whole protein diffusion is represented with red spheres. The cylinder part is aligned and centered on the Z axis with a radius R_{cyl} = 1 Å. The funnel potential is limited to 35 Å from the Z-axis origin (SG atom of Cys47) by a wall potential represented by a violet sphere. The cone region is defined by a vertex height Z_{cc} of 18 Å from the origin and an angle α of 1.1 rad.

This can be done by properly setting the angle α and the distance z_{cc} parameters of the funnel restraint potential (see Materials and Methods section and ref 14 for details). In such a way the external potential does not influence the ligand exploration of the binding site, $z < z_{cc}$, while a cylindrical restraint potential is applied when the ligand is in the unbound state, $z > z_{cc}$.

Using this protocol the ligand−protein binding constant, K_b can be computed as follows:

$$K_b = \pi R_{cyl}^{\ 2} \int_{site} dz e^{-\beta[W(z)-W_{ref}]} \quad (1)$$

where $\pi R_{cyl}^{\ 2}$ is the surface of the cylinder used as restraint potential, while the potential $W(z)$ and its value in the unbound state, W_{ref}, can be derived from the potential of mean force (PMF) obtained through metadynamics calculations. β is constant, where $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant and T the temperature of the system.

The equilibrium binding constant, K_b, is directly related to the absolute protein−ligand binding free energy, $\Delta G_b^{\ 0}$, through the following formula:

$$\Delta G_b^{\ 0} = -\frac{1}{\beta} \ln(C^0 K_b) \quad (2)$$

where $C^0 = 1/1660$ Å^{-3} represents the standard concentration 1 M.

Note that the equilibrium binding constant, K_b, is the inverse of the dissociation equilibrium constant of the ligand−protein complex, K_D, usually measured by NMR experiments (see High-Precision Dissociation Constant Determination Using Solution NMR section).

Using FM simulations, Limongelli et al. have studied two different protein−ligand systems, namely trypsin/benzamidine and cyclooxygenase II (COX II) in complex with a potent inhibitor.[14] These authors have not only described with atomic details the complex ligand−protein binding mechanism but also computed the $\Delta G_b^{\ 0}$ values for both systems. The calculated $\Delta G_b^{\ 0}$ and the derived K_D fall in the range of the experimentally determined values.

The active site of PRX5 is directly accessible from the solvent, and the cone part contains all possible interacting residues of PRX5. FM simulations were then performed for more than 500 ns applying the metadynamics bias on the distance CV defined as the projection on the Z axis of the distance between Cys47 SG and the center of mass of the ligand heavy atoms. The statistics of two further CVs, specifically the distance from the funnel Z axis of the ligand center of mass and the protein−ligand torsion Cys47 SG···O1−C1 (ligand) or the SG···O1−O1 angle (see above), were collected along the FM simulations to evaluate the conformational space explored by the system (Figure S3). During the FM simulations the ligands exchange several times between the different bound conformational states diffusing in the solvent. Binding and unbinding events occurred 8−10 times over the whole 500 ns of FM. The binding events take approximately 10−30 ns, separated by unbound periods of 20−30 ns. The calculation convergence was monitored using the bias deposition as well as the potential mean force (PMF) $W(z)$ and $\Delta G_b^{\ 0}$ calculated as described before.[14] FES as a function of CVs different from that biased in the FM calculation (Figure 3) can be reconstructed using the potential added to the system and a reweighting algorithm.[28]

Absolute Binding Free-Energy. To have a quantitatively well-characterized free-energy profile, a number of recrossing events between the different states visited by the system should be observed.[14] As shown in Figure S3 both systems visit several times the bound (z < 8 Å) and the unbound states (z > 20 Å) during the simulation. Using R_{cyl} = 1 Å and $W(z)$ obtained from metadynamics (Figure 3 upper-left panel) in eqs 1 and 2, the estimate of $\Delta G_b^{\ 0}$ for catechol and 4-methylcatechol converges to −3.0 ± 0.2 and −4.2 ± 0.3 kcal/mol, respectively. To provide a picture of the convergence of the binding free-energy estimation, the free-energy difference between bound and unbound states has been computed as a function of the simulation time for either system (Figure 3 lower-left panel).

DOI: 10.1021/ja511336z
J. Am. Chem. Soc. 2015, 137, 1273−1281

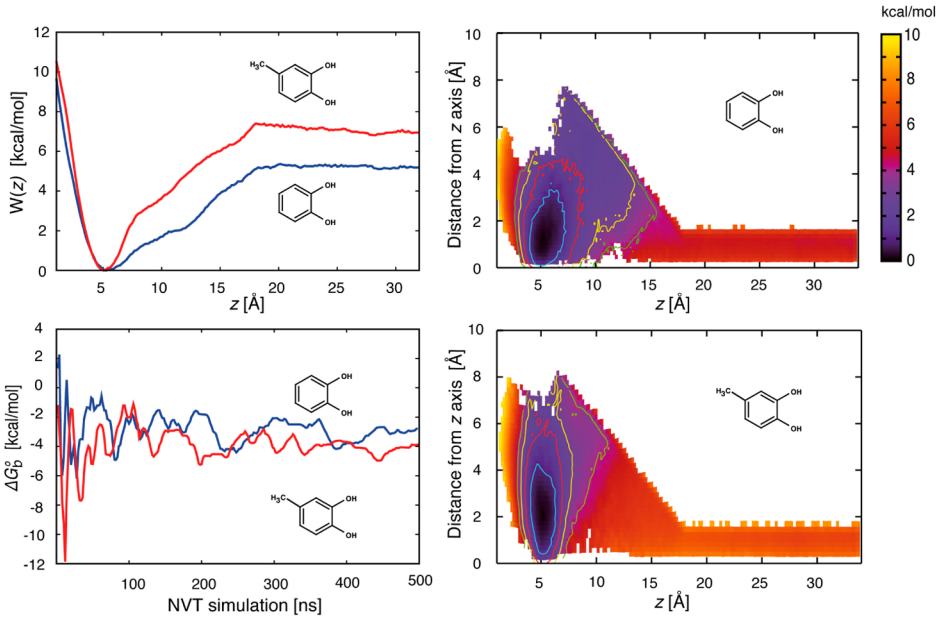


Figure 3. PMF and FES represented as a function of the different collective variables used in FM calculations and evolution of the absolute binding free energy (lower-left panel). Upper-left: compared PMF $W(z)$ of catechol (blue trace) and 4-methylcatechol (red trace) as a function of the Z-axis projection distance between Cys47 SG and the center of mass of the ligand heavy atoms. Upper/lower-right: FESs of the PRX5-ligand system for catechol (upper) and 4-methylcatechol (lower) as a function of the protein−ligand distance projection on Z axis and the protein−ligand distance from Z axis. The FESs show the lowest energy basins found by the FM calculations. Lower-left panel: evolution of absolute binding free energy during 500 ns of FM simulation for catechol (blue trace) and 4-methylcatechol (red trace). Using PMF $W(z)$ and R_{cyl} = 1 Å in eqs 1 and 2, the estimate of ΔG_b^0 for catechol and 4-methylcatechol converges to −3.0 ± 0.2 and −4.2 ± 0.3 kcal/mol, respectively. The uncertainty is calculated as the SD from the asymptotic value of the absolute protein−ligand binding free energy obtained in the last part of the simulation.

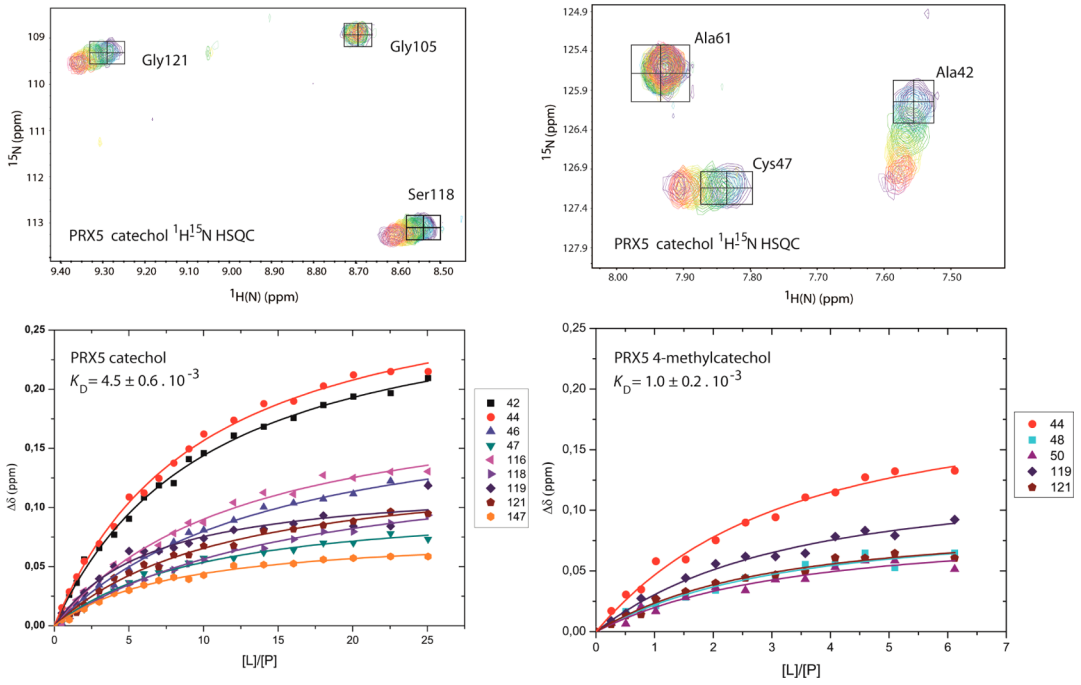


Figure 4. High-precision determination of dissociation constants K_D of human PRX5 bound to catechol and 4-methylcatechol using solution NMR. Upper-pannels: parts of the overlaid $^1H-^{15}N$ HSQC spectra of PRX5 at various concentrations of catechol in aqueous solution, 28 °C and pH 7.4. Chemical shift indicated with boxed crosses are the initial NMR chemical shift with no ligand. Lower two panels: plots of the NMR chemical shift perturbation ($\Delta\delta$) as a function of ligand-to-protein concentration ratio ($[L]/[P]$) for the amino acids sequence position indicated on the graphs. Data were nonlinearly fitted to the 1:1 protein−ligand model P + L ⇄ PL with a single dissociation K_D.

Furthermore, using eq 1 and considering that K_b is the inverse of K_D, we could estimate K_D for catechol and 4-methylcatechol, which is 6.9 ± 2.1 × 10^{-3} and 0.9 ± 0.4 × 10^{-3}, respectively.

High-Precision Dissociation Constant Determination Using Solution NMR. NMR is ideally suited for the analysis of protein interactions with dissociation constants in the μM

DOI: 10.1021/ja511336z
J. Am. Chem. Soc. 2015, 137, 1273−1281

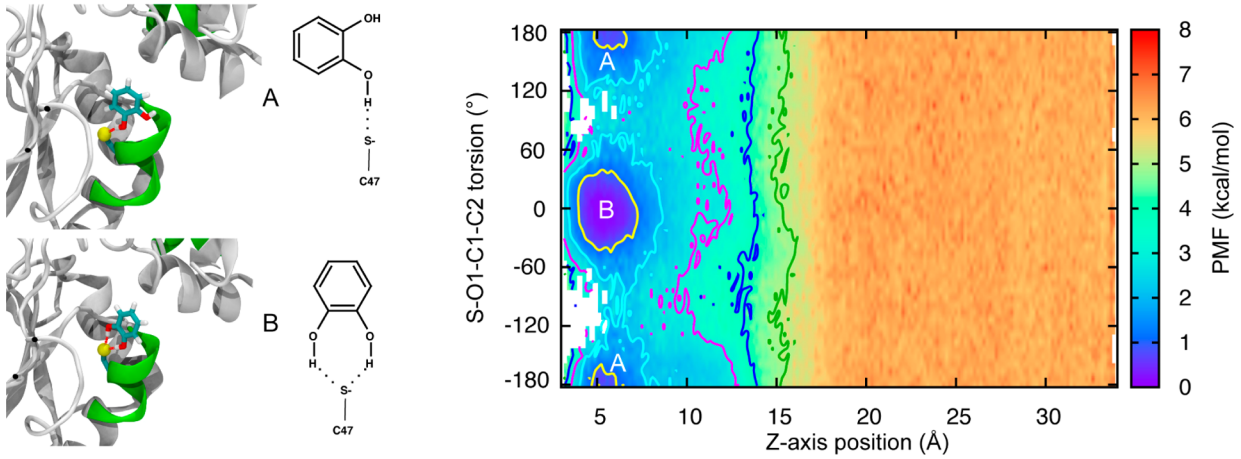


Figure 5. FES of catechol interacting basins with PRX5 evaluated using FM. Free energies are represented using iso-energetic contours spaced by 1 kcal/mol from the minima at 0 and 5 kcal/mol and a color continuum from red (highest free energy) to blue (lowest free energy) as a function of the positional projection of catechol center of mass along the *Z* axis defined according to the Material and Methods section, and the torsion defined by the catalytic thiolate SG atom of PRX5 Cys27 residue and the O1, C1, and C2 atoms of catechol. Map minimum B represents the double H-bonds binding mode, while energetic A basin represents the single aligned H-bond similar to the crystal ligand positions (see the text).

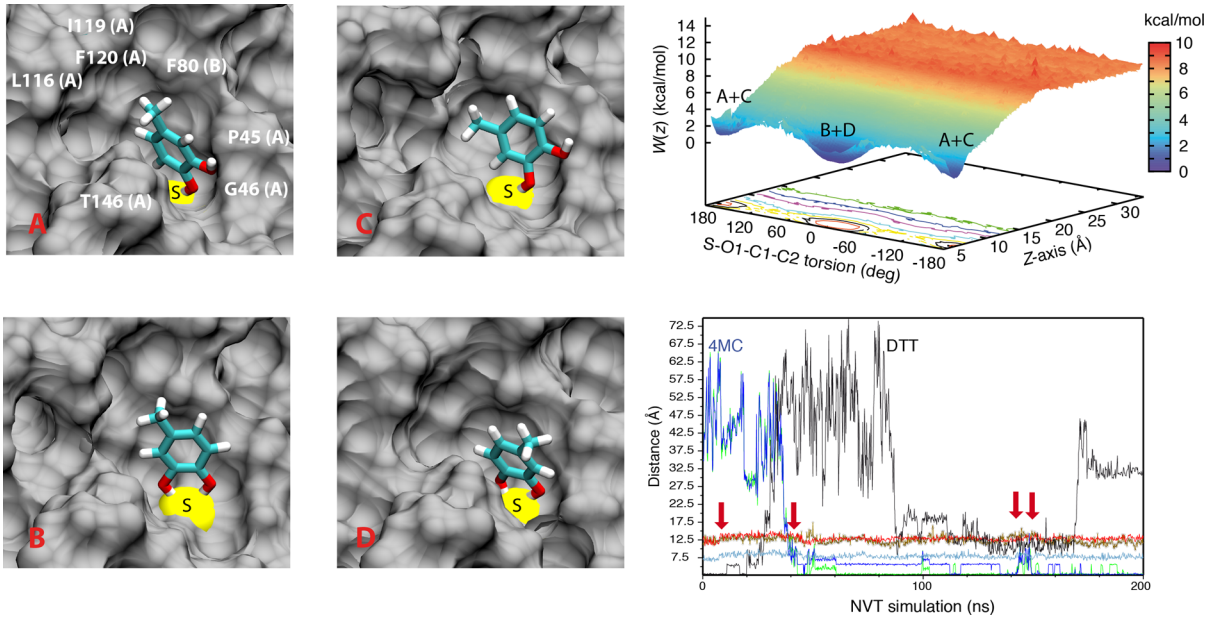


Figure 6. Binding FES of the human PRX5/4-methylcatechol complex obtained through FM (upper right), NVT unrestrained dynamics trajectories (lower right), and representation of the lowest energy ligand binding modes (left). Snapshots represent the surface accessible to the solvent of the PRX5 active site in gray and the SG atom of catalytic Cys47 in yellow. 4-methylcatechol is represented with sticks, and the surrounding amino acids are indicated either from the chain A or B of the PRX5 homodimer. Right diagrams, top: FES representation as a function of two collective variables used in funnel-metadynamics. *Z*-axis projection distance between Cys47 SG and the center of mass of the ligand heavy atoms and the protein−ligand torsion angle defined by PRX5 Cys47 SG, 4-methylcatechol O1, C1, and C2 atoms. Lower-right diagram: interatomic distances as a function of unrestrained NVT dynamics time steps during the first 200 ns of the 500 ns run. Protein(Cys47)-ligand distances measured for 4-methylcatechol (4MC) and DTT are plotted in green (SG-O1), blue (SG-O2), and black (SG to the DTT oxygen H-bonded to Cys47 in the X-ray structure), respectively, indicating the binding/unbinding events along the time steps (red arrows). Interatomic distances within the PRX5 active site including Cys47 SG with Leu116 CG (brown), Ile119 CB (red) and Phe120 (light blue) are graphed.

to mM range in the fast exchange equilibrium on the NMR time scale.[29,30] $^{1}H-^{15}N$ NMR data of PRX5 can be used to titrate the reduced form PRX5 against catechol and 4-methylcatechol.[31] Since the oxidized form dithiothreitol (DTT), a reagent commonly used to reduce disulfide bridges of proteins, was reported as a ligand of PRX5,[20] we preferred to use the unrelated reducing agent tris(2-carboxyethyl)phosphine (TCEP) hydrochloride, as not to interfere with the ligand binding. Figure 4 shows overlay of $^{1}H-^{15}N$ HSQC spectra recorded after each increasing concentrations of the ligands. Specific amino acids are affected by the interaction and present significant chemical shift perturbations (CSPs). The corresponding CSPs were then quantified carefully. The most pronounced CSPs are the $^{1}H-^{15}N$ amide group of amino acids in the active site, near the catalytic cysteine (Cys47). Residues Ala42, Thr44, Gly46, and Cys47 for catechol and Thr44, Ser48, and Thr50 for the methylated homologue are mostly affected. Some amide groups of amino acids located in the loop and the α-helix adjacent to the active site also present

DOI: 10.1021/ja511336z
J. Am. Chem. Soc. 2015, 137, 1273−1281

significant CSPs (Leu116, Ser118, Ile119, Gly121, and Thr147 for catechol and Ile119, Gly121 for 4-methylcatechol). Observed significant CSPs suggest that interaction of PRX5 with catechol and 4-methylcatechol is indeed specific despite their low-affinity.

When the most significant CSPs are plotted against the ratio ligand-to-protein concentrations $[L]/[P]$, hyperbolic isotherms are obtained as shown in Figure 4. The data can be best fitted using a nonlinear fitting protocols corresponding to a 1:1 protein−ligand binding model with a single K_D. Overall catechol induced more CSPs than 4-methylcatechol. Catechol induced to the most (Ala42 and Thr44) 0.2 ppm of CSPs, while 4-methylcatechol induced to the most CSPs between 0.1−0.15 ppm. However, the averages of the data (each titration was duplicated) shown in Figure 4 lead to K_D's of 4.5 ± 0.6 and 1.0 ± 0.2 kcal/mol at 301 K in phosphate buffer saline (PBS) solution pH 7.4 (NaPi 10 mM, NaCl 137 mM, KCl 3.3 mM) TCEP 2 mM, for catechol and 4-methylcatechol, respectively.

DISCUSSION

Details of the Ligand−Protein Interaction. When the ligand occupies the protein active site, the two binding modes (protein−ligand S−O−O angles 60° and 180°, and protein−ligand SG−O1−C1−C2 torsion 0° and 180°, respectively) exchange regularly, and the ligand samples the whole conformational space dynamically. For catechol, which owns a C_{2v} symmetry, the two SG−O1−C1−C2 torsion values correspond to two different binding conformations, as shown in Figure 5. For 4-methylcatechol with a C_h symmetry the two SG−O1−C1−C2 torsion values correspond to more possible conformations in which the methyl group could interact differently within the PRX5 active site. In the crystal state, the methyl group is oriented in the active site interacting with neighboring hydrophobic side chains such as Leu116, Ile119, Phe120, and Phe80 (Figure 6A). Due to the protein dynamics, the active site changes its shape during the ligand binding and unbinding events. This is clearly shown in Figure 6 where the distances between Cys47 SG atom, and the side chains of Leu116 and Ile119 are plotted during a standard MD calculation. In particular, one can observe that, on average, these distances elongate during the unbinding events and shorten in the binding by about 2 Å.

The different ligand binding conformations could not be discriminated just by observing the FES associated with the two protein−ligand distance CVs given in Figure 3 (right panels).

Catechol. The different ligand binding modes can be identified reconstructing the FES as a function of the distance CV on Z axis and the protein−ligand SG−O1−C1−C2 torsion (Figure 5). This operation is possible using the reweighting algorithm of Bonomi et al.[28] Looking at the FES shown in Figure 5 two free-energy minima can be found. The first, basin A, corresponds to the ligand binding mode with a single protein−ligand H-bond, including the conformations found by X-ray.[20,21] The second basin, B, is energetically equivalent to A and represents the binding mode with two protein−ligand H-bonds.

4-Methylcatechol. As for catechol, all the bound 4-methylcatechol conformations fall in the free-energy basins reported in Figure 6. Here, two most populated binding modes are found: conformer A, which is similar to the crystal conformation with a single H-bond, and conformer B, which shows the chelate H-bond interaction. In this state, the methyl group is orientated to the hydrophobic cavity formed by residues such as Phe80 (chain B), Phe120, and Leu116. Two less populated states are also present in both energy basins. Conformer C is a minor state of basin A where the ligand is slightly moved from A due to the steric hindrance of the ligand methyl group. Conformer D is a minor state of basin B where the methyl group points toward the solvent where it is not possible to form favorable hydrophobic interactions. The weaker ligand−protein interactions formed in C and D if compared with A and B lead us to consider these poses the first binding event of 4-methyl-cathecol in the active site before reaching its final position. Alternatively, they can be considered the first unbinding event of the ligand from the catalytic site of the enzyme.

Calculated Absolute Binding Free Energies and Experimental High-Precision Dissociation Constants Determined by NMR. Table 1 summarizes the calculated absolute binding free energies ΔG_b^0 using FM protocol and the dissociation constants K_D determined using NMR. For each ligand the experimental value matches very well the calculated one.

Table 1. Comparison of Calculated and Experimental Dissociation Constants K_D and Absolute Binding Free Energy ΔG_b^0 of Catechol and 4-Methylcatechol for Human PRX5

	catechol	4-methylcatechol
$\Delta G_b^{0\,a}$		
FM[b]	−3.0 ± 0.2	−4.2 ± 0.3
NMR	−3.2 ± 0.1	−4.1 ± 0.1
K_D (10^{-3})		
FM[b]	6.9 ± 2.1	0.9 ± 0.4
NMR	4.5 ± 0.6	1.0 ± 0.2

[a]kcal/mol. [b]Mean value over the last 100 ns of 500 ns FM.

Since both ligands are soluble in water, we checked that the presence of 0−8% DMSO does not induce a significant difference. DMSO alone causes some limited CSP of $^1H-^{15}N$ other amino acids including Ser48 and Gly46 that were also used for the K_D determinations affecting marginally the determined dissociation constant values. Previously published data, determined using NMR saturation transfer difference spectroscopy, for the catechol-PRX5 system reported a K_D of 3.3 ± 0.5 × 10^{-3} that is very close to our estimate.[31] The use of DTT instead of TCEP reducing agent does not affect significantly the measured values. We just noticed that DTT caused a pronounced long-term precipitation of the PRX, modifying the PRX5 concentration. However, if the NMR titration is done within a day or less, no significant differences of the data were noticed.

Structural Source of the Difference in ΔG_b^0 between the Two Ligands of PRX5. Based on the analysis of the FESs, the source of stabilizing interactions between the catechol ligands and PRX5 is clearly the possibility to form specific patterns of H-bonds between its two hydroxyl groups and the thiolate of the catalytic Cys47.

Given the estimated pKa of catalytic PRX Cys ranging from 5 to 6, more than 90% of Cys47 is a thiolate in solution at pH 7 and the source of the reactivity of PRX's toward their substrate peroxides.[20] The specific conformer's geometry of the type A in Figure 5 was proposed to be an analog of the transition state[20] and lead to a well-defined free-energy basin. This sort of stabilizing interaction is reproduced with 4-methylcatechol

DOI: 10.1021/ja511336z
J. Am. Chem. Soc. 2015, 137, 1273−1281

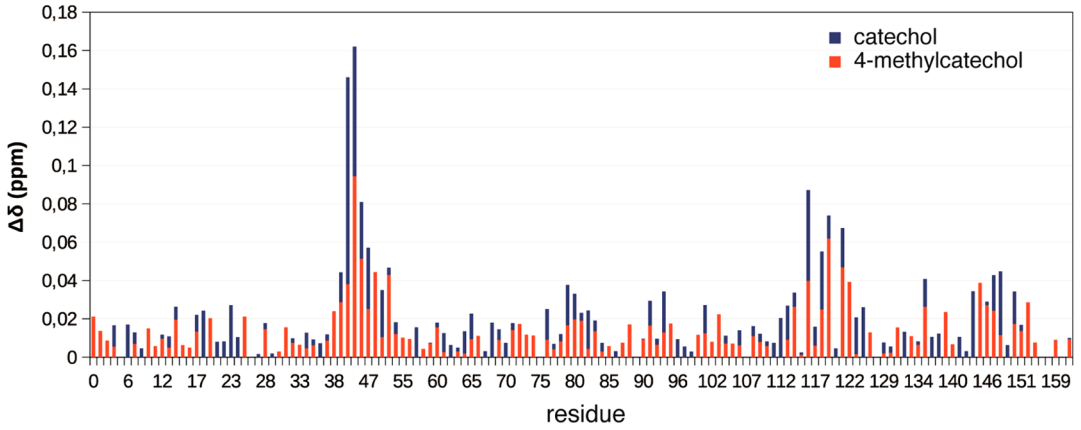


Figure 7. Superimposed histograms of the chemical shift perturbations $\Delta\delta$ caused to human PRX5 $^{1}H-^{15}N$ NMR signals by catechol (blue bars) and 4-methylcatechol (red bars) at concentrations twice their respective determined dissociation constant K_D's giving a comparable 75% of the PRX5 saturation each. Missing data due to spectral overlaps or nonobserved amino acids are not represented.

extending the stabilizing van der Waals interactions laterally with Leu116, Leu119, Phe120 side chains of the chain A and also the side chain of Phe80 of the chain B. Dynamically and on average, the active site contracts upon binding the 4-methylcatechol to optimize those additional van der Waals interactions. As a result, the lowest free-energy basin of this ligand presented in Figure 3 is more extended compared to that of catechol. This contributes, with other factors such as the higher polarity of catechol that leads to more favorable interactions with the solvent, to a gain in ΔG_b^0 of about 1 kcal/mol. We stress that the differences in ΔG_b^0 for the two ligands is accounted for both experimentally using NMR and theoretically by the metadynamics calculations.

MD Calculations and NMR. Our work demonstrated for the first time that free-energy calculations using enhanced sampling simulations correlate with NMR-determined affinities for weakly interacting protein−ligand system with dissociation constants at room temperature in the order of 10^{-3}. Recent studies with MD and FM were conducted for more affine systems with K_D in the range from 10^{-6} to 10^{-9}.[10,14] NMR observables and isotropic chemical shift in particular are essentially an average over the NMR time scales (ms to s for chemical shifts) of the conformations sampled by the system. The relevant modeling of the NMR observables should then rely on the full statistical description of the complete exchanging conformations.

The use of bias in metadynamics and external potentials as the funnel-shaped potential used in FM protocols allows faster and more complete sampling than standard simulations. The two essential states of an interacting system (the free state in the solvent and the bound forms) exchange more, and the several possible bound conformers of low energies are better sampled. At the convergence, ΔG_b^0 reaches a mean stable value, and the system motion can be considered diffusive. Thus, the use of biased MD simulation allows a quantitative description of the interactions. This description is by far more accurate than that obtained using faster but less accurate approaches, such as molecular docking, where the intrinsic dynamical character of proteins is neglected. An example of the relevance of using MD-based protocols in modeling NMR parameter can be given by the consideration of the two ligands CSPs induced at a comparable ratio between free protein and bound protein. For instance, at ligand concentrations near the determined K_D (i.e., at 50% of the saturation) as shown in Figure 7, 4-methylcatechol induced on average significantly less CSPs than catechol despite a lower K_D. The difference in CSPs can be ascribed to the larger number of ligand−protein interactions formed by 4-methylcatechol if compared with those of catechol. This aspect is explained by the FESs computed for both ligands. In fact, while catechol shows a certain free-energy basin, 4-methylcatechol has a free-energy basin extending to a more larger distance from the Z axis (vertical axis in Figure 3). This reflects a greater number of iso-energy binding conformations. All the protein motions due to these interactions are averaged in the time scale of the NMR experiment, resulting in lower CSPs for 4-methylcatechol. Such considerations, and the possibility to describe quantitatively through simulations the whole exchanging systems at a time scale relevant to NMR spectroscopy, were nearly ignored until now for interacting protein−ligand systems.

We believe that this first approach of coupling enhanced sampling calculations to NMR and other microscopic and macroscopic experimental data is the only way to rationally interpret the interacting systems in solution. The resulting knowledge may contribute significantly to the possible design and engineering of the interactions to conceive high-affinity ligand from low-affinity primary hits.

CONCLUSION

We have shown herein that MD, complemented by metadynamics, correlates not only in the energetics of a protein−ligand interacting system (accurate estimates of ΔG_b^0) but also in the whole dynamical exchanges occurring on time scale relevant to NMR spectroscopy. The consequences are important for the comprehensive understanding of the protein−ligand affinity and useful for a proper engineering and design of new interacting molecules.

MATERIALS AND METHODS

Molecular Dynamics. All MD runs were prepared starting from the conformation of the crystal structure of PRX5 complexed with oxidized DTT (3MNG.pdb).[20] PRX5 homodimers were constructed from appropriate crystallographic C2 symmetry mates. Simulations were carried out with the AMBER99SB-ILDN force field[32−34] for the protein and the TIP3P water model for the explicit solvent.[35] Crystallographic water molecules were kept to save in particular the low-diffusion water molecules buried in the protein structure. Active site Cys 47 was modeled a thiolate (residue CYM of Amber), and 48 Na^+ and 46 Cl^- were added to account for the experimental conditions

DOI: 10.1021/ja511336z
J. Am. Chem. Soc. 2015, 137, 1273−1281

used in NMR for ca. 150 mM NaCl aqueous solution. The Amber charges were applied to protein, ions, and water atoms, and the restrained electrostatic potential charges were used for the ligands using the Antechamber program suite[36] and the General Amber Force Field GAFF.[37] Using ACEMD code,[9] the system was minimized and equilibrated under constant pressure and temperature (NPT) conditions at 1 atm and 300 K using a time step of 4 fs thanks to the use of the hydrogen mass repartitioning scheme[38] implemented in ACEMD, nonbonded cutoff of 9 Å, rigid bonds and particle-mesh Ewald long-range electrostatics with a grid of 82-83-92 with spacing of 1 Å. The systems were equilibrated first using 500 steps of steepest-descent minimization, followed by running 0.1 ns of the isothermal NVT ensemble, using a Langevin thermostat set at 300 K, followed by 5 ns of the isothermal−isobaric NPT ensemble using a Langevin thermostat at the same temperature and the Berendsen barostat of ACEMD. During minimization and equilibration, the heavy protein and ligands atoms were restrained spacialy using 10 kcal mol^{-1} $Å^{-2}$ spring constant. The magnitude of the restraining force constant was then reduced to1 kcal mol^{-1} $Å^{-2}$ during 100 ps of NVT. The barostat was switched on at 1 atm for 1 ns of NPT simulation. During that period the force constant of the position constrain of all heavy atoms was gradually reduced every 100 ps steps by a factor 0.65 to a final value of about 0.1 kcal mol^{-1} $Å^{-2}$, allowing the systems to relax gently. Finally, the volume was allowed to relax for further 4 ns under NPT conditions reaching a final box size of 82.7-83.7-93.9 $Å^3$. During this run, Cα atoms of the protein and ligands were restrained with a 1 kcal mol^{-1} $Å^{-2}$ harmonic potential to prevent the system reorienting. Then, a production run of 500 ns in the NVT ensemble was performed.

Funnel-Metadynamics.[14] Unrestrained simulations were run at the same temperature and solvent model in the NVT ensemble for the indicated simulation times. The PLUMED plugin[28] was used to carry out metadynamics calculations. Since the funnel external potential is fixed in the space,[14] PRX5 dimer is diffusion restrained by three atoms chosen far from the dimer interface and the active site as shown in Figure 2. The bias was added on a distance CV (see Results section). A Gaussian width of 0.35 Å was used, and a Gaussian deposition frequency of 0.5 kcal mol^{-1} every 2 ps (1 kcal = 4.18 kJ) was initially used and gradually decreased on the basis of the adaptive bias with a ΔT of 3600 K. Trajectories were analyzed using VMD software.[39]

Protein Production and Purification. Recombinant Human PRDX5 was expressed without its mitochondrial targeting sequence as a N-terminal 6xHis-tagged protein using *Escherichia coli* M15 (pRep4) strain (plasmid coding for the protein was generously given by B. Knoops lab at University of Louvain, Belgium). To produce ^{15}N uniformly labeled PRDX5, cells were grown at 37 °C in M9 minimal medium containing $^{15}NH_4Cl$ 1g/L as sole source of nitrogen and supplemented with thiamine 20 μg/mL, metal traces solution, and the two antibiotics of resistance, kanamycin and ampicillin (50ug/mL of each). Produced PRDX5 was then purified using nickel-affinity chromatography with a Ni^{2+}-NTA column (Qiagen) as previously described[31] and dialyzed against a PBS pH 7.4 (NaH_2PO_4/Na_2HPO_4 10 mM, NaCl 137 mM, KCl 3.3 mM). Protein concentration was measured according to the UV absorbance at 280 nm based on a molar extinction coefficient at that wavelength of 5500 M^{-1} cm^{-1} calculated[40] for one chain of PRX5 (concentration expressed in molar concentration of monomer in solution).

NMR Experiments. All spectra were recorded at 28 °C with a Varian Inova 600 MHz NMR spectrometer equipped with a 5 mm standard triple resonance ($^1H/^{13}C/^{15}N$) inverse probe with a *Z*-axis field gradient. The concentration of DMSO-d_6 did not exceed 8% in the NMR samples, and DMSO's influence was tested separately. All NMR data sets were processed using NMRPipe/NMDraw package,[41] and ^{15}N chemical shift changes were assigned using NMRViewJ software.[42] Ligand interactions with U-^{15}N-PRX5 were characterized using CSPs. A series of $^1H-^{15}N$ HSQC experiments were acquired upon addition of specific ligand concentrations. NMR samples in typical series contained 400 μM of the reduced ^{15}N-labeled protein. TCEP 0.1 M was added to maintain the protein in its reduced form in PBS pH = 7.4 to a final concentration in TCEP of 2 mM. To avoid possible aggregation of the proteins, an increasing volume of a concentrated stock solution of ligands (110 mM solutions in DMS0 or H_2O) was incrementally added until saturation was reached. To derive the corresponding binding constant K_D (see below), spectral perturbations were quantified as the combined amide CSPs $\Delta\delta$ according to the following expression:[43]

$$\Delta\delta = \sqrt{\frac{\Delta\delta_H{}^2 + \frac{1}{25}\Delta\delta_N{}^2}{2}} \tag{3}$$

Dissociation Constant Determination from NMR CSPs. Titration data were analyzed assuming that the observed chemical shift perturbation $\Delta\delta$ is a weighted average between the two extreme values corresponding to the free ($\Delta\delta = 0$) and ligand-bound states ($\Delta\delta = \Delta\delta_B$). PRX5 was assumed to contain two independent binding sites corresponding to the equivalent chains A and B of the symmetrical homodimer. CSPs data were then nonlinearly fitted against the theoretical 1:1 model of protein−ligand binding of the simple binding equilibrium:

$$P + L \rightleftarrows PL \tag{4}$$

Where P stands for the protein, L for the ligand and PL for the protein−ligand complex. This equilibrium has dissociation constant K_D expression:

$$K_D = \frac{[P_0][L_0]}{[PL]} - [P_0] - [L_0] + [PL] \tag{5}$$

where $[P_0]$ and $[L_0]$ are the total concentrations in protein and ligand, respectively, and $[PL]$ the concentration of protein−ligand complex at the equilibrium. A 1:1 model with one ligand molecule bound per PRX5 gives[44]

$$\Delta\delta = \frac{[PL]\Delta\delta_B}{[P_0]} \tag{6}$$

so

$$\Delta\delta = \frac{\Delta\delta_B(([P_0] + [L_0] + K_D) - \sqrt{([P_0] + [L_0] + K_D)^2 - 4[P_0][L_0]}}{2[P_0]} \tag{7}$$

where the chemical shift perturbations ($\Delta\delta_B$) and the dissociation constant (K_D) were fit with nonlinear regression by using a in-house Matlab (The MathWorks, Inc.) based program.[45] Error margins on final K_D values were calculated using a Monte Carlo simulation assuming 20% of experimental error on protein and ligand concentrations under a cumulative error on volume and weigh-in. All titrations were achieved in duplicate and the results expressed as the averages.

■ ASSOCIATED CONTENT

Supporting Information

Figures S1−S3 providing trajectories collective variable analysis of unrestrained MD and restrained FM calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*jean-marc.lancelin@isa-lyon.fr

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Prof. Bernard Knoops and André Clippe from University of Louvain-la-Neuve, Belgium, for the generous gifts of biological materials to express the human peroxiredoxin in *Escherichia coli*. We thank Acellera Company for providing of a complementary 6 month ACEMD multi-GPU software

DOI: 10.1021/ja511336z
J. Am. Chem. Soc. 2015, 137, 1273−1281

licence and for the generous gift of a Nvidia Tesla K20 supercomputing unit. Funding of this project was provided by a grant from Région Rhône-Alpes, France, no. 13-0118962-01 to L.T. and J.-M.L., and a grant from the Swiss National Supercomputing Centre (CSCS) under project ID s557 to V.L.

REFERENCES

(1) Kuriyan, J.; Konforti, B.; Wemmer, D. *The Molecules of Life: Physical and Chemical Principles*; Garland Science: New York, 2012.
(2) Changeux, J. P. *Protein Sci.* **2011**, *20*, 1119.
(3) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *Arch. Biochem. Biophys.* **1978**, *185*, 584.
(4) Wüthrich, K. *Angew. Chem.* **2003**, *42*, 3340.
(5) Serber, Z.; Keatinge-Clay, A. T.; Ledwidge, R.; Kelly, A. E.; Miller, S. M.; Dotsch, V. *J. Am. Chem. Soc.* **2001**, *123*, 2446.
(6) Ernst, R. R. B, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Clarendon Press: Oxford, 1987; Vol. *14*.
(7) Goodsell, D. S.; Morris, G. M.; Olson, A. J. *J. Mol. Recognit.* **1996**, *9*, 1.
(8) Guo, L.; Yan, Z.; Zheng, X.; Hu, L.; Yang, Y.; Wang, J. *J. Mol. Model.* **2014**, *20*, 2251.
(9) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. *J. Chem. Theory Comput.* **2009**, *5*, 1632.
(10) Buch, I.; Giorgino, T.; De Fabritiis, G. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184.
(11) Williamson, M. P. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1.
(12) Limongelli, V.; Bonomi, M.; Marinelli, L.; Gervasio, F. L.; Cavalli, A.; Novellino, E.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5411.
(13) Barducci, A.; Bonomi, M.; Parrinello, M. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826.
(14) Limongelli, V.; Bonomi, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6358.
(15) Wood, Z. A.; Poole, L. B.; Karplus, P. A. *Science* **2003**, *300*, 650.
(16) Knoops, B.; Goemaere, J.; Van der Eecken, V.; Declercq, J. P. *Antioxid. Redox Signal.* **2011**, *15*, 817.
(17) Wood, Z. A.; Schroder, E.; Robin Harris, J.; Poole, L. B. *Trends Biochem. Sci.* **2003**, *28*, 32.
(18) Smeets, A.; Marchand, C.; Linard, D.; Knoops, B.; Declercq, J. P. *Arch. Biochem. Biophys.* **2008**, *477*, 98.
(19) Declercq, J. P.; Evrard, C.; Clippe, A.; Stricht, D. V.; Bernard, A.; Knoops, B. *J. Mol. Biol.* **2001**, *311*, 751.
(20) Hall, A.; Parsonage, D.; Poole, L. B.; Karplus, P. A. *J. Mol. Biol.* **2010**, *402*, 194.
(21) Aguirre, C.; ten Brink, T.; Guichou, J. F.; Cala, O.; Krimm, I. *PloS One* **2014**, *9*, e102300.
(22) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.
(23) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(24) Limongelli, V.; De Tito, S.; Cerofolini, L.; Fragai, M.; Pagano, B.; Trotta, R.; Cosconati, S.; Marinelli, L.; Novellino, E.; Bertini, I.; Randazzo, A.; Luchinat, C.; Parrinello, M. *Angew. Chem.* **2013**, *52*, 2269.
(25) Limongelli, V.; Marinelli, L.; Cosconati, S.; La Motta, C.; Sartini, S.; Mugnaini, L.; Da Settimo, F.; Novellino, E.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 1467.
(26) Grazioso, G.; Limongelli, V.; Branduardi, D.; Novellino, E.; De Micheli, C.; Cavalli, A.; Parrinello, M. *J. Am. Chem. Soc.* **2012**, *134*, 453.
(27) Di Leva, F. S.; Novellino, E.; Cavalli, A.; Parrinello, M.; Limongelli, V. *Nucleic Acids Res.* **2014**, *42*, 5447.
(28) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961.
(29) Fielding, L. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51*, 219.
(30) Markin, C. J.; Spyracopoulos, L. *J. Biomol. NMR* **2012**, *53*, 125.
(31) Barelier, S.; Linard, D.; Pons, J.; Clippe, A.; Knoops, B.; Lancelin, J. M.; Krimm, I. *PloS One* **2010**, *5*, e9744.
(32) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(33) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
(34) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950.
(35) Jorgensen, W. L.; Madura, J. D. *J. Am. Chem. Soc.* **1983**, *105*, 1407.
(36) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157.
(37) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *25*, 247.
(38) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *J. Comput. Chem.* **1999**, *20*, 786.
(39) Humphrey, W.; Dalke, A.; Fau - Schulten, K.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.
(40) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S. e.; Wilkins, M.; Appel, R.; Bairoch, A. In *The Proteomics Protocols Handbook*; Walker, J., Ed.; Humana Press: New York, 2005; p 571.
(41) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277.
(42) Johnson, B. A.; Blevins, R. A. *J. Biomol. NMR* **1994**, *4*, 603.
(43) Farmer, B. T.; Constantine, K. L.; Goldfarb, V.; Friedrichs, M. S.; Wittekind, M.; Yanchunas, J.; Robertson, J. G.; Mueller, L. *Nat. Struct. Mol. Biol.* **1996**, *3*, 995.
(44) Schumann, F.; Riepl, H.; Maurer, T.; Gronwald, W.; Neidig, K.-P.; Kalbitzer, H. *J. Biomol. NMR* **2007**, *39*, 275.
(45) Lange, A.; Ismail, M. B.; Riviere, G.; Hologne, M.; Lacabanne, D.; Guilliere, F.; Lancelin, J. M.; Krimm, I.; Walker, O. *FEBS Lett.* **2012**, *586*, 3379.

DOI: 10.1021/ja511336z
J. Am. Chem. Soc. 2015, 137, 1273−1281