JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Hidden Regularity and Universal Classification of Fast Side Chain Motions in Proteins

Rajitha Rajeshwar T.,[†] Jeremy C. Smith,[‡,§] and Marimuthu Krishnan*,[†]

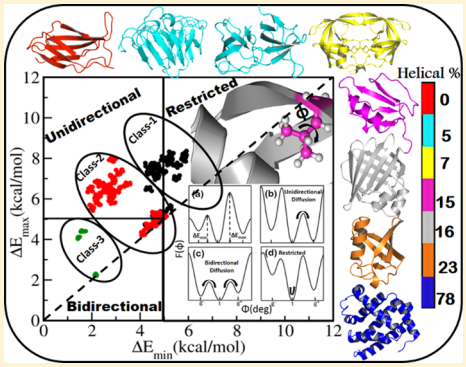
[†]Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Gachibowli, Hyderabad 500 032, India

[‡]UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, P.O. Box 2008 Oak Ridge, Tennessee 37831-6309, United States

[§]Department of Biochemistry and Molecular and Cellular Biology, University of Tennessee, M407 Walters Life Sciences, 1414 Cumberland Avenue, Knoxville, Tennessee 37996, United States

Supporting Information

ABSTRACT: Proteins display characteristic dynamical signatures that appear to be universal across all proteins regardless of topology and size. Here, we systematically characterize the universal features of fast side chain motions in proteins by examining the conformational energy surfaces of individual residues obtained using enhanced sampling molecular dynamics simulation (618 free energy surfaces obtained from 0.94 μs MD simulation). The side chain conformational free energy surfaces obtained using the adaptive biasing force (ABF) method for a set of eight proteins with different molecular weights and secondary structures are used to determine the methyl axial NMR order parameters (O^2_{axis}), populations of side chain rotamer states (ρ), conformational entropies (S_{conf}), probability fluxes, and activation energies for side chain inter-rotameric transitions. The free energy barriers separating side chain rotamer states range from 0.3 to 12 kcal/mol in all proteins and follow a trimodal distribution with an intense peak at ~5 kcal/mol and two shoulders at ~3 and ~7.5 kcal/mol, indicating that some barriers are more favored than others by proteins to maintain a balance between their conformational stability and flexibility. The origin and the influences of the trimodal barrier distribution on the distribution of O^2_{axis} and the side chain conformational entropy are discussed. A hierarchical grading of rotamer states based on the conformational free energy barriers, entropy, and probability flux reveals three distinct classes of side chains in proteins. A unique nonlinear correlation is established between O^2_{axis} and the side chain rotamer populations (ρ). The apparent universality in O^2_{axis} versus ρ correlation, trimodal barrier distribution, and distinct characteristics of three classes of side chains observed among all proteins indicates a hidden regularity (or commonality) in the dynamical heterogeneity of fast side chain motions in proteins.



INTRODUCTION

The startling diversity in the structure and dynamics of proteins illustrates the complexity and richness of protein-mediated biological processes and also demands a molecular-level understanding of the fundamental principles underpinning these processes.[1−5] The notion that structure and dynamics are governed by the underlying energy surface underscores the intricate connection between heterogeneity in functional dynamical processes and the hierarchical conformational substates of proteins.[6−10] The interactions within proteins and with surrounding species (such as other proteins, solvent molecules, and counterions) give rise to a complex energy landscape resulting in a wide spectrum of dynamics, ranging from localized atomic vibrations to large-scale collective conformational transitions.[2,7,11,12] The high-frequency harmonic motions do not alter the equilibrium structure of the protein, while significant structural changes occur during conformational transitions. These anharmonic, barrier-crossing motions enable the protein to visit different regions of the conformational space and thus play critical roles in protein function.[6,7,9−11,13−15] Conformational transitions in proteins occur at various length- (domain- to side chain-level) and time-scales (fast (ps−ns) to slow (μs−ms)).[1,8,16,17]

Many research efforts directed toward understanding the functional roles of internal motions primarily rely on experimental techniques that probed the average dynamics (i.e., dynamics averaged over all residues or probes) of proteins. For instance, fluorescence spectroscopy measures the average lifetime or relaxation rate of different fluorophores in a protein,[18−20] neutron scattering experiments determine the average mean-square displacement of the nonexchangeable protein hydrogen atoms,[2,21,22] and infrared (IR) and Raman spectroscopies probe the overall vibrational modes of different functional groups in proteins.[23,24] Although these experimental techniques are non-site-specific, they play critical roles in

Received: August 29, 2013
Published: May 20, 2014

   dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

unravelling clues about the functional importance of dynamics and about some universal features in the average dynamics of proteins. For example, it is now established that the distribution of the entropy-rich, low-frequency (below 100 cm^{-1}) collective vibrational modes of proteins follows a universal curve[25,26] and that there exists a universal temperature-dependent harmonic to anharmonic dynamical transition in proteins.[2,14,21,27]

In the light of recent progress in the experimental characterization of site-specific fast dynamics of protein side chains, it is essential to understand the heterogeneity and correlations in the side chain conformational fluctuations at different sites in a protein and to examine the commonality of these fast internal motions. In this regard, nuclear magnetic resonance (NMR) spectroscopy has emerged as a powerful technique to probe the structure and dynamics of backbones and side chains of proteins.[3,5,17,28−42] Recent advances in isotopic enrichment and biosynthetic approaches to labeling atoms at specific sites of proteins, together with new pulse sequences, have opened up new possibilities of investigating site-specific details of the structure, dynamics, and thermodynamics of high-molecular weight proteins and protein complexes in their native and less-populated excited states.[30,31,34,37−50]

Site-specific NMR relaxation studies use methyl groups in proteins as reporters of side chain dynamics.[28,31,32,37,38] The hydrogens and/or carbons of methyl groups in side chains of a protein are labeled with deuterium (^2H) and ^{13}C respectively, resulting in an isotopically enriched protein consisting of different methyl spin probes (methyl isotopomers) (CH_3, $^{13}CHD_2$, $^{13}CH_2D$) spread across different parts of the protein.[28,48] Details of side chain dynamics encoded in the nuclear Overhauser effects (NOE), longitudinal (T_1) and transverse (T_2) relaxation times of ^2H and ^{13}C spins of enriched methyl groups can be extracted using the "model-free" formalism of Lipari and Szabo, resulting in motional parameters that quantify the rates and amplitudes of motions of these methyl groups.[28,32,51−53]

Specifically, the square of an order parameter, O^2, quantifies the amplitude of methyl group motion, while τ measures the relaxation time associated with the motion. The reorientational dynamics of the symmetry axis and the rotation of the C−H bond vectors about the symmetry axis are the two major contributions that determine methyl dynamics. Assuming that these two degrees of freedom are uncoupled, O^2 can be written as a product of O^2_{rot} and O^2_{axis}, where O^2_{rot} is the order parameter for rotation about the methyl symmetry axis, and O^2_{axis} is that for motion of the symmetry axis. $O^2_{rot} = 1/9$ for a methyl group that rotates completely and possesses ideal tetrahedral geometry, and in the analysis of NMR experimental data, this value of O^2_{rot} is commonly assumed.[11,54]

Molecular dynamics simulations provide atomistic insights into molecular motions occurring in a time window accessible to site-specific NMR relaxation experiments, and the relaxation data obtained from these techniques are complementary to each other.[54−61] In MD simulation, the atomic interactions are described by an empirical potential energy function, which describes both covalent and noncovalent interactions using a suitable choice of force field parameters. The atomic forces derived from this potential function are used to solve the equations of motion, thereby tracking dynamical trajectories of individual atoms in the system.[1,54,56,57,59−61] The spatial and temporal correlations and other equilibrium properties of the system determined from 10 to 100 ns MD trajectories can be used to provide atomistic insights into NMR relaxation data.[1,54,56,57,59−61]

There are at least two approaches to compute NMR motional parameters from MD trajectories.[38,41,54,56,57,59−62] In the first approach, the reorientational time correlation functions (TCFs), $C(t)$, of the C−H bond vectors of methyl groups are computed, and the motional parameters O^2_{axis} and τ are derived using the following relationships:

$$C(t) = \langle P_2(\hat{\mu}(0)\cdot\hat{\mu}(t))\rangle \quad (1)$$

$$O^2 = \lim_{t\to\infty} C(t) \quad (2)$$

$$O^2 = O^2_{rot}O^2_{axis} = 0.111O^2_{axis} \quad (3)$$

$$\tau = \frac{1}{1-O^2}\int_0^\infty (C(t) - O^2)\,dt \quad (4)$$

where $\hat{\mu}$ is a unit vector along the C−H bond and $P_2(x) = (3x^2 - 1)/2$ is the second Legendre polynomial. The other approach is to monitor the time evolution of θ and ϕ, here $\pi - \theta$ and ϕ denote the polar and azimuthal angles determining the orientation of the symmetry axis of a given methyl group (shown in Figure 1), to determine the equilibrium orientational distribution, $P(\theta, \phi)$. O^2_{axis} can be determined from $P(\theta, \phi)$ using the following relationship.

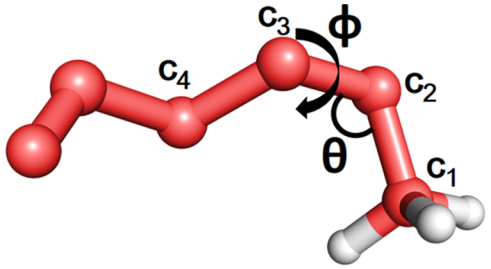


Figure 1. Schematic representation of the angular coordinates θ and ϕ determining the orientation of the symmetry axis (unit vector along C_1-C_2 bond vector) of a methyl group in a protein. C_1 is the methyl carbon and C_2, C_3, C_4 are the consecutive heavy atoms attached to C_1.

$$O^2_{axis} = \frac{3}{2}[\langle x^2\rangle^2 + \langle y^2\rangle^2 + \langle z^2\rangle^2 + 2\langle xy\rangle^2 + 2\langle xz\rangle^2 + 2\langle yz\rangle^2] - \frac{1}{2} \quad (5)$$

By expressing the Cartesian coordinates in terms of the spherical polar coordinates, the averages (denoted by the angular brackets) of different squared terms in the above equation can be calculated using the equilibrium probability distribution, $P(\theta, \phi)$.[11,63−65]

Both of the above-mentioned approaches should, in principle, give accurate estimates of O^2_{axis} and τ, provided that the force field parameters are accurate enough to capture the protein dynamics realistically and that the conformational sampling is good. In practice, it is not possible to achieve complete sampling of the allowed conformational substates of a protein (even of a moderate size) in a typical 10−100 ns MD simulation. Hence, the estimates of the NMR motional parameters obtained from simulations with inadequate sampling of phase space are subject to sampling errors.[59] In the first approach above, lack of convergence of TCFs obtained from MD trajectories is the manifestation of the sampling errors. Consequently, multiple simulations may result in different TCFs with different values of O^2_{axis} for a given methyl

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

side chain, thus giving rise to average O^2_{axis} values with significant uncertainty. In the second approach, insufficient conformational sampling results in $P(\theta, \phi)$ that deviates from the equilibrium distribution. Hence, with both of the above methods the incomplete sampling of the side chain conformational space gives rise to poor estimates of O^2_{axis} and associated relaxation times. In contrast, the protein backbone is in many simulations relatively rigid, exhibiting only restricted motion, and so the extent of conformational space to be sampled by a backbone amide bond vector is in these cases very limited, the backbone degrees of freedom do not suffer from sampling issues, and a 10−100 ns MD trajectory is sufficient to sample the backbone conformational space. Thus, the above approaches have worked well for backbone relaxation parameters.[59,66]

A computational approach that efficiently circumvents the sampling problem by using enhanced sampling methods was developed recently to determine sampling error-free side chain O^2_{axis} from molecular dynamics trajectories.[67] Using calmodulin as a model system, a nonlinear relationship was established between O^2_{axis} and the populations of side chain rotamer states which was then used to reconstruct side chain conformational free energy surfaces from O^2_{axis}.[67] In the present contribution, we extend this method to a set of eight proteins with different secondary structures and molecular weights to examine whether there is a universal nature of the correlation between methyl symmetry axis order parameters (O^2_{axis}) and populations of side chain rotamer states. The results indicate that, regardless of the molecular weight and topology of the proteins concerned, a unique correlation exists between the side chain conformational free energy surfaces and O^2_{axis} which is likely to be universal among proteins. Each methyl-containing residue type in any protein follows a characteristic nonlinear curve in the parameter space defined by O^2_{axis} and the populations of the rotamer states. The universality of this correlation involves a trimodal barrier distribution, and distinct characteristics of three classes of side chain observed in all proteins indicate a regularity (or commonality) of the dynamical heterogeneity in fast side chain motions in proteins.

SIMULATION DETAILS

Model Systems. A set of eight proteins with different secondary structures (all-alpha, all-beta, and mixed) and molecular weights (8.1−23.7 kDa) was selected. The PDB IDs of these proteins are 1EGL (70 residues, 8.1 kDa, 15% helical),[68] 1UBQ (76 residues, 8.5 kDa, 23% helical),[69] FNfn10 (93 residues, 9.93 kDa, 0% helical), 1SHF (2 × 59 residues, 13.4 kDa, 5% helical),[70] 1LIB (131 residues, 14.5 kDa, 16% helical),[71] 1J52 (154 residues, 18.6 kDa, 78% helical),[72] 1A45 (173 residues, 20.9 kDa, 5% helical),[73] 3KF1 (2 × 99 residues, 23.7 kDa, 7% helical).[74] The coordinates of FNfn10 were taken from 1FNF.[59] The choice of these proteins was based on the availability of previous experimental and molecular dynamics (MD) simulation studies.[75−77] Among the eight model proteins studied here, the experimental O^2_{axis} values are available for six of them while the remaining two were chosen from the extreme ends of the spectrum of the secondary structure composition: myoglobin (primarily an α-helical protein) and Gamma F Crystallin (primarily a β-sheet protein). Thus, the selected set of proteins reasonably covers a wide range of secondary structures and molecular weights while remaining within the limits of computational capabilities.

Molecular Dynamics Simulation. MD simulations of these proteins were carried out using NAMD2.8[78] with the CHARMM27 all-atom[79] and TIP3P[80] water force fields. For each model system, the heteroatoms were first removed from the corresponding PDB structure and the protein was solvated in a TIP3P[80] water box of suitable dimensions: 54 × 47 × 49 Å^3 (1EGL), 40 × 42 × 46 Å^3 (1UBQ), 45 × 59 × 61 Å^3 (FNfn10), 44 × 47 × 42 Å^3 (1SHF), 61 × 55 × 47 Å^3 (1LIB), 44 × 57 × 60 Å^3 (1J52), 58 × 54 × 80 Å^3 (1A45), and 54 × 48 × 58 Å^3 (3KF1). After removing the water molecules that were in hard contact with the protein, all the model systems were subjected to energy minimization using the conjugate gradient method followed by 3 ns of NPT equilibration MD runs at 1 atm pressure and at 300 K. Subsequently, MD simulations were performed in the NPT ensemble at 300 K and 1 atm pressure using a Langevin thermostat and barostat with a damping coefficient of 5 ps^{-1}. Each simulation consisted of 2 ns equilibration followed by a 10 ns production run. The equations of motion were integrated with a time step of 1 fs. The nonbonded pair-interaction potential was truncated at 12 Å and smoothed between 10 and 12 Å using a cubic switching function. Periodic boundary conditions were applied. Electrostatic interactions were computed using the particle mesh Ewald (PME) method[81] with a real space cutoff of 13 Å, and the reciprocal space interactions were computed on grids (using sixth-degree B-splines) of suitable dimensions: 58 × 50 × 54 (1EGL), 44 × 46 × 50 (1UBQ), 50 × 64 × 66 (FNfn10), 48 × 50 × 46 (1SHF), 64 × 58 × 50 (1LIB), 45 × 60 × 64 (1J52), 60 × 56 × 82 (1A45), and 56 × 50 × 60 (3KF1). Four further, independent, 10 ns production runs were also performed starting from different initial velocities.

Adaptive Biasing Force Method. The adaptive biasing force (ABF) method was used to accelerate the side chain conformational sampling, thereby providing sampling error-free estimates of the conformational free energy landscapes of the methyl-bearing side chains of the proteins. The statistical mechanical theory underlying the ABF method is briefly discussed here. Consider an N-particle system with a Hamiltonian, $H(\mathbf{R},\mathbf{P}) = U(\mathbf{R}) + T(\mathbf{P})$, where $\mathbf{R}$ and $\mathbf{P}$ denote the set of Cartesian coordinates and momenta of all particles, respectively, and $U(\mathbf{R})$ and $T(\mathbf{P})$ represent the potential and kinetic energies of the system. To determine the potential of mean force, $F(\phi)$, as a function of a chosen reaction coordinate, ϕ, the Cartesian coordinates are first transformed into a set of generalized coordinates (ϕ, $\mathbf{Q}$), where $\mathbf{Q}$ denotes the set of remaining generalized coordinates. The derivative of $F(\phi)$ with respect to ϕ is expressed as follows:

$$\frac{dF(\phi)}{d\phi} = \left\langle \frac{\partial U(\phi, Q)}{\partial \phi} - \frac{1}{k_B T}\frac{\partial \ln|J|}{\partial \phi} \right\rangle_{\phi} = -\langle f_{\phi} \rangle_{\phi} \qquad (6)$$

where J is the Jacobian associated with the transformation from the Cartesian coordinates to the generalized coordinates, k_B is the Boltzmann's constant, $\langle f_{\phi} \rangle_{\phi}$ is the average force acting along the reaction coordinate, ϕ, determined at a given value of ϕ, and the angular brackets denote the statistical averages.[82−85] In the ABF method, the reaction coordinate ϕ is divided into small windows of size $d\phi$, $\langle f_{\phi} \rangle$ is computed for each bin during the course of the simulation, and a biasing force proportional to $\langle f_{\phi} \rangle_{\phi}$ is introduced in the dynamics to generate uniform sampling along the chosen reaction coordinate.[82−85]

The Collective Variables Module[82−85] implemented in NAMD2.8 was used for performing the ABF calculations.

The torsional angle, ϕ (shown in Figure 1), of any given methyl-containing side chain was selected as the reaction coordinate. The mean force as a function of ϕ was accumulated in bins of width 1° from a total simulation time of 65 ns. The convergence of the free energy profiles was examined by monitoring the time evolution of a convergence quantity, ζ, defined as follows:

$$\zeta = \sqrt{\frac{1}{N}\sum_{\phi=-\pi}^{\pi}[F(\phi, t+\Delta t) - F(\phi, t)]^2} \quad (7)$$

where N is the number of bins used to partition the ϕ space, $F(\phi, t)$ and $F(\phi, t+\Delta t)$ are the free energy profiles of a given side chain calculated at two adjacent time windows, t and $t + \Delta t$, respectively, and Δt was set to 7.2 ns in our calculations. That is, the free energy profiles for all methyl-containing residues were recorded at regular intervals of 7.2 ns along the ABF trajectory and the value of ζ was monitored as a function of t (here $t = \Delta t, 2\Delta t, 3\Delta t, ..., n\Delta t$, and n is an integer chosen such that $n\Delta t$ corresponds to the total ABF trajectory length). ζ was observed to decrease to a value close to zero with increasing t, and the conformational sampling is considered complete when ζ is less than a set threshold value (0.05 kcal/mol) for all residues, as reported elsewhere.[67]

O^2_{axis} and Conformational Entropy from Free Energy Profiles. We have calculated O^2_{axis} of the protein side chains using two independent methods. In the first, the free energy profiles, $F(\phi)$, obtained from the ABF calculations were used to determine O^2_{axis}. The angular distribution function, $P(\theta, \phi)$ (θ and ϕ are defined in Figure 1), is calculated from $F(\phi)$ using the following relationship:

$$P(\theta, \phi) = \frac{e^{-\beta F(\phi)} e^{-\beta(\theta-\theta_0)^2/2\sigma^2}}{\int_0^{\pi}\int_{-\pi}^{\pi} e^{-\beta F(\phi)} e^{-\beta(\theta-\theta_0)^2/2\sigma^2} \sin\theta \, d\theta \, d\phi} \quad (8)$$

Here, θ_0 is the equilibrium value of θ, $\sigma/(\beta)^{1/2}$ is the width of the Gaussian distribution characterizing the fluctuations of θ around θ_0, $\beta = 1/k_BT$, k_B is the Boltzmann's constant and T is the temperature. The values of θ_0 and σ for each residue were determined by fitting the distribution of θ (Figure 2) obtained from the MD trajectories with a Gaussian function. The Cartesian coordinates of the unit vector along the symmetry axis of a given methyl group can be expressed in terms of θ and ϕ as follows: $x = \sin(\theta)\cos(\phi)$, $y = \sin(\theta)\sin(\phi)$, and $z = -\cos(\theta)$. Given $P(\theta, \phi)$ and x, y, z in terms of θ and ϕ, it is straightforward to determine $O^2_{axis,ABF}$ using eq 5 and by computing the squared averages appearing in eq 5 as follows:

$$\langle\alpha(\theta, \phi)\omega(\theta, \phi)\rangle = \int_0^{\pi}\int_{-\pi}^{\pi} \alpha(\theta, \phi)\omega(\theta, \phi)P(\theta, \phi)\sin\theta \, d\theta \, d\phi \quad (9)$$

where α and ω can be x, y, or z. In the second method, $O^2_{axis,MD}$ was computed directly from the nonaccelerated MD trajectories. To do this, translational and rotational motions of a given protein were first removed by superimposing each snapshot of the trajectory onto the reference structure. To obtain $O^2_{axis,MD}$ of any given methyl group, the time evolution of the x, y, and z components of the unit vector along the symmetry axis was used to compute time averages of the squared terms in eq 5. The order parameter values were obtained for each independent MD run performed, and $O^2_{axis,MD}$ reported here is an average over all MD runs.

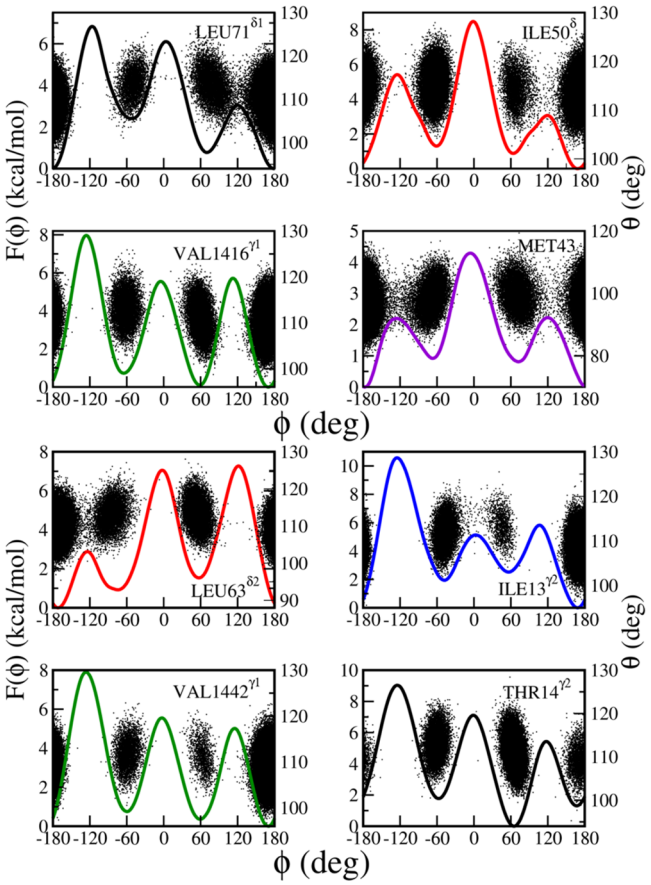


Figure 2. Scatter plots of θ/ϕ obtained from five 10 ns MD trajectories and conformational free energy profiles computed using the ABF method (solid line) shown for representative side chains of proteins: ubiquitin (black), HIV-1 protease (red), myoglobin (blue), 1A45 (violet) and FNfn10 (green).

The side chain conformational entropy was computed using the following equation:

$$S_{conf} = -k_B\int_0^{\pi}\int_{-\pi}^{\pi} P(\theta, \phi) \ln P(\theta, \phi) \sin\theta \, d\theta \, d\phi \quad (10)$$

$P(\theta, \phi)$ defined in eq 8 was used in the above equation to calculate S_{conf} for each residue.

RESULTS AND DISCUSSION

First, we emphasize that most of the results reported in the following sections were obtained from the free energy profiles determined using the ABF method.

Free Energy Profiles. Figure 2 shows scatter plots of θ/ϕ distributions (from direct MD simulations) and free energy profiles (from the ABF calculations) for a few representative side chains that do not suffer from sampling issues. The θ/ϕ scattered data were calculated from a set of five independent 10 ns trajectories obtained using nonaccelerated MD simulations. The bend angle, θ, is restricted to a narrow window around 110° for all residues while the dihedral angle, ϕ, is distributed between ϕ = 180° and ϕ = −180°. The data points are clustered around specific values of ϕ that correspond to the most preferred side chain conformations. The minima on the corresponding free energy profiles are located at ϕ values at which the clustering of data is observed in the θ/ϕ scattered plots. The spread and density of the clustered data points

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

correlate with the width and depth of the corresponding rotamer states, respectively. Also, the absence of data points between two adjacent clusters indicates the presence of a free energy barrier between these rotamer states. The similarity between the MD and ABF results (at least for well-sampled residues) together with the convergence analysis (see Simulation Details section) indicates the reliability of the computed free energy profiles. In a typical short MD simulation, the low-energy conformations dominate the Boltzmann distribution, and there is incomplete sampling of the conformations close to the barrier, preventing the determination of the exact barriers from the distributions obtained from MD simulations.

The conformational free energy profiles of all methyl groups except those in alanine residues consisted of three stable rotamer states: *gauche*$^+$ (denoted here as g^+) ($\phi \approx 60 \pm 10°$), *gauche*$^-$ (g^-) ($\phi \approx -60 \pm 10°$) and *trans* (t) ($\phi \approx \pm 180°$). The stability of each rotamer state and the magnitudes of the energy barriers separating any two adjacent rotamer states vary, depending upon the local environment and the chemical nature of the residues to which the methyl groups are attached.[86−89] Since the methyl groups in ALA residues are directly attached to the backbone, the symmetry axes of these methyl groups are expected to exhibit restricted dynamics within a rotamer state (either in g^+ or g^- state). For all ALA residues a single cluster of data observed in the θ/ϕ scatter plots with a single deep well in the corresponding free energy profiles establish this fact (Figure S1 - Supporting Information [SI]).

Figure 3 shows the free energy profiles and θ/ϕ scatter plots for a few representative side chains that inadequately sampled

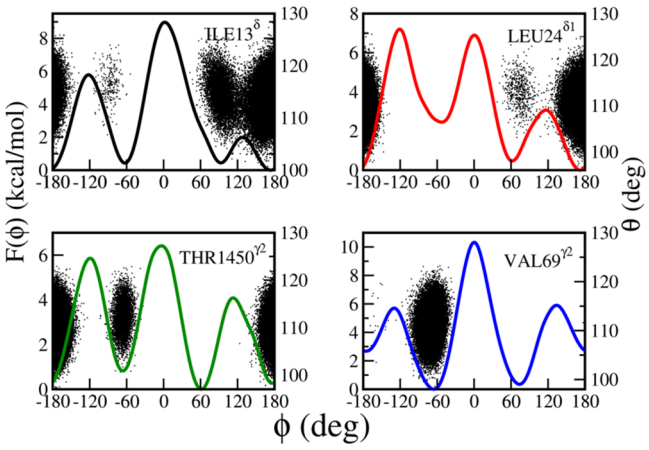


Figure 3. Scatter plots of θ/ϕ obtained from five 10 ns MD trajectories and conformational free energy profiles computed using ABF method (solid line) shown for representative side chains that exhibit poor conformational sampling during MD simulation: ubiquitin (black), HIV-1 protease (red), myoglobin (blue), 1A45 (violet), and FNfn10 (green).

the conformational space during MD simulations. The comparison of the MD and ABF results demonstrates that some rotamer states are indeed not sampled in MD simulations. $F(\phi)$ computed using the ABF method for these residues (Figure 3) indicates three minima corresponding to three stable rotamer states, with one or two of these states not being visited during the course of the MD simulations. It was observed that for each system ∼40% of methyl-containing residues were "well-behaved" (i.e., they visit all rotamer states in an ensemble of five independent 10 ns MD trajectories) while the remaining residues suffered from sampling issues to differing degrees.

Barrier Distribution. *Universal Trimodal Barrier Distribution.* We now examine the distribution of barriers (ΔF) separating side chain rotamer states on the free energy surfaces obtained from the ABF simulations. These barriers determine the kinetics of conformational transitions of side chains in proteins.[6−10,90] Each rotamer state is surrounded by two barriers (one each for forward and reverse rotamer transitions), and thus there are six activation energies for any given side chain (excluding ALA residues). Figure 4 shows the distribution

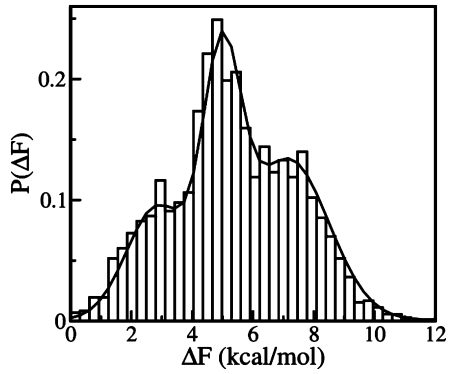


Figure 4. Distribution of free energy barriers of all methyl-containing residues of all proteins studied.

of free energy barriers, $P(\Delta F)$, obtained for all proteins investigated in this study. The corresponding distributions for individual proteins are provided in the SI (Figure S2).

The barrier distribution shows three features: an intense peak at ∼5 kcal/mol and two shoulders, at ∼3 kcal/mol and ∼7.5 kcal/mol. This trimodal distribution of barrier heights is universal among all proteins studied here (Figure S2 in the SI), indicating that some barriers are more favored than others by proteins to maintain a balance between their conformational stability and flexibility; 39% of barriers fall between 4 and 6 kcal/mol, while 25% and 36% of barriers are lower than 4 kcal/mol and greater than 6 kcal/mol, respectively. A side chain with all six barrier heights lower than 4 kcal/mol (first peak in the barrier distribution) is expected to have higher conformational flexibility, while a more rigid side chain would have all barriers higher than 6 kcal/mol. However, an examination of the barrier heights for individual residues reveals that it is rare that all six barrier heights of a given residue contribute to one single peak or shoulder in the barrier distribution; most side chains have barrier heights spread between two or three of the features.

Fourier Expansion of Free Energy Profiles. In order to examine the origin of the trimodal distribution of side chain conformational barriers in proteins, the calculated free energy profiles were decomposed into 1-, 2-, and 3-fold components by fitting them using the following truncated Fourier expansion:

$$F(\phi) = \Delta E_0 + \sum_{n=1}^{3} \frac{\Delta E_n}{2}[1 - \cos n(\phi - \delta_n)] \quad (11)$$

where δ_n is the phase angle and ΔE_0 and ΔE_n are the coefficients of the cosine expansion.[91] A few representative free energy profiles fitted with the above equation are shown in Figure S3(a) (SI). The physical interpretation of different terms

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

in the above equation is well-established: the 3-fold term is due to steric interactions arising from bond−bond repulsion, the 1-fold term is mainly due to dipolar and steric interactions, while the 2-fold term is due to charge delocalization between lone pair orbitals and the rotating bonds.[91] The absolute values of the coefficients ΔE_1, ΔE_2, and ΔE_3 are the measures of the barriers in the 1-, 2-, and 3-fold profiles, respectively. The distribution of the absolute values of ΔE_n shown in Figure S3(b) (SI) also reasonably follows a trimodal distribution similar to the overall barrier distribution shown in Figure 4 despite the fact that $P(\Delta F)$ and $P(\Delta E_n)$ represent the distributions of two different quantities (ΔF is a measure of the overall barrier while ΔE_n denotes the barrier on the n-fold [n = 1,2,3] Fourier component).

The statistics of the relative contributions of 1- to 3-fold terms to any given ΔF and to any peak in $P(\Delta F)$ are likely to have direct implications for the physical origin of the trimodal distribution of conformational barriers. Consider a barrier ΔF between a free energy minimum ($F(\phi_{min})$ at $\phi = \phi_{min}$) and a maximum ($F(\phi_{max})$ at $\phi = \phi_{max}$). The barrier $\Delta F = F(\phi_{max}) - F(\phi_{min})$ can be written as a sum of barrier contributions arising from the 1- to 3-fold Fourier components as follows: $\Delta F = \Delta F_1 + \Delta F_2 + \Delta F_3$, where $\Delta F_i = F_i(\phi_{max}) - F_i(\phi_{min})$ is the barrier contribution from the corresponding i-fold (i = 1,2,3) Fourier component $F_i(\phi)$. The ratio $\Delta F_i/\Delta F$ gives a quantitative measure of the relative contribution of the i-fold term to a given conformational barrier. The ratio $\Delta F_i/\Delta F$ can be postive or negative, depending upon whether the i-fold interaction term is involved in barrier enhancement ($\Delta F_i > 0$) or reduction ($\Delta F_i < 0$), respectively. Given these relative contributions, it becomes straightforward to examine which of the Fourier components chiefly contributes to different peaks in $P(\Delta F)$. Figure 5 shows

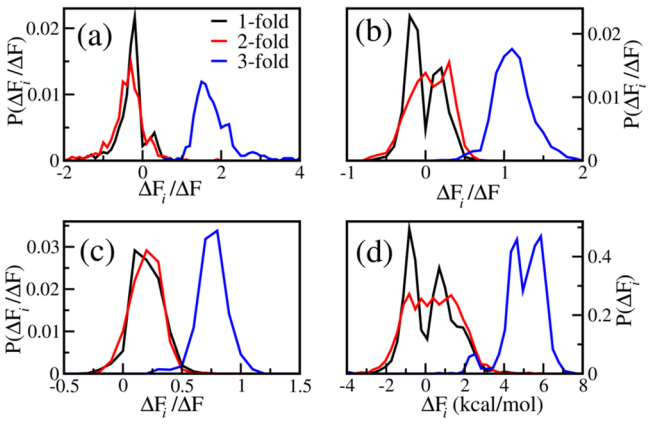


Figure 5. Distributions of $\Delta F_i/\Delta F$ for the (a) first peak, (b) second peak, (c) third peak of $P(\Delta F)$ and (d) ΔF_i are shown.

the distribution of $\Delta F_i/\Delta F$ for 1-, 2-, and 3-fold Fourier components plotted separately for ΔF belonging to different peaks in $P(\Delta F)$. It is observed that ΔF_3 takes only positive values between 2 kcal/mol and 7 kcal/mol and $P(\Delta F_3)$ exhibits two major peaks at $\Delta F_3 \approx 4.2$ kcal/mol and at $\Delta F_3 \approx 6$ kcal/mol (Figure 5d). In contrast, ΔF_1 and ΔF_2 take both positive and negative values between −3.0 and 4 kcal/mol with a broad $P(\Delta F_1)$ distribution centered around $\Delta F_1 = 0$. $P(\Delta F_2)$ is also almost centered around zero but exhibits two peaks at $\Delta F_2 = \pm 1$ kcal/mol.

It is evident from $P(\Delta F_i/\Delta F)$ distributions (Figure 5a−c) that the 3-fold term representing the steric interactions mainly contributes to the peaks in $P(\Delta F)$. $P(\Delta F_1/\Delta F)$ and $P(\Delta F_2/\Delta F)$ distributions corresponding to the first peak (0.3 kcal/mol $\leq \Delta F \leq$ 3.6 kcal/mol) of $P(\Delta F)$ are peaked at $(\Delta F_1/\Delta F) = (\Delta F_2/\Delta F) = -0.25$ and ΔF_1 and ΔF_2 are mostly negative, indicating that the 1- and 2-fold terms play a catalytic role in reducing the overall barrier, leading to low barriers on the free energy profiles. In contrast, $P(\Delta F_1/\Delta F)$ and $P(\Delta F_2/\Delta F)$ corresponding to the third peak (6.3 kcal/mol $< \Delta F \leq$ 12 kcal/mol) are shifted to the positive side with a sharp peak at $(\Delta F_1/\Delta F) = (\Delta F_2/\Delta F) = 0.15$ suggesting that all 1- to 3-fold terms provide additive positive contributions to ΔF, leading to high barriers (or rate-limiting barriers) on the free energy surface. Similarly, for the second peak (3.6 kcal/mol $< \Delta F \leq$ 6.3 kcal/mol) in $P(\Delta F)$, the 1- and 2-fold terms provide both positive and negative contributions (almost equally) to the overall ΔF.

Although the truncated Fourier expansion approach provides insights into the role of important interactions in determining the conformational barriers, the original theoretical interpretations were formulated on the basis of quantum chemical calculations on simple organic molecules in the gas phase[91] whereas $F(\phi)$ reported in the present study were obtained for protein solutions. The exact contributions of nonbonded interactions in a condensed phase to 1-, 2-, and 3-fold Fourier components and to the overall $F(\phi)$ remain unclear. A detailed further investigation of the contributions of van der Waals and electrostatic interactions to the conformational barriers is much needed to shed more light on the trimodal nature of $P(\Delta F)$, but it is beyond the scope of this paper.

Calculated Barrier Versus Experimental Activation Energy. Although NMR, neutron scattering, and theoretical studies report on barriers to methyl rotation in proteins, the trimodal distribution of barriers to methyl symmetry axis dynamics in proteins is yet to be observed experimentally.[92,93] However, caution must be exercised when comparing the calculated $P(\Delta F)$ reported in Figure 4 with the corresponding quantity obtained from experiment for the following reasons. Most experiments only determine a single activation energy per residue, from the temperature dependence of the corresponding terminal methyl relaxation time, whereas $P(\Delta F)$ reported in Figure 4 includes all six barriers in the corresponding free energy surface controlling the dynamics of the methyl symmetry axis of any given residue. Moreover, some side chain conformational transitions may occur between a few most-preferred conformers, and some rotamer states may not be visited by a side chain within the experimental time scales, as has been seen for some methyl-containing residues.[94] In such cases, the activation energy estimated from experiment does not have contributions from barriers that surround the unpopulated rotamer states.

In order to make a sensible comparison with the experimental activation energies, the average barrier, $\langle \Delta F \rangle_{ABF}$, was calculated for each methyl group from the rates of transitions across the six barriers separating different rotamer states using the following equation:[54,92,95]

$$\langle \Delta F \rangle_{ABF} = -k_B T \ln\left(\frac{1}{6}\sum_{i=1}^{6} e^{-\Delta F_i/k_B T}\right) \quad (12)$$

where k_B is the Boltzmann's constant, T is the temperature, and ΔF_i is the i^{th} (i = 1,2,...,6) barrier on the corresponding conformational free energy surface. The experimental activation energies, $\langle \Delta F \rangle_{EXP}$, were obtained from the reported exper-

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

imental methyl relaxation times (τ_e) using the following transition state theory expression:[96,97]

$$\langle \Delta F \rangle_{EXP} = k_B T \ln\left(\frac{k_B T \tau_e}{h}\right) \quad (13)$$

where h is the Planck's constant. The comparison of $\langle \Delta F \rangle_{EXP}$ and $\langle \Delta F \rangle_{ABF}$ is meaningful only for the fast residues that sample all rotamer states within the NMR experimental time scale. In the present study, by using the experimental distribution of τ_e (Figure S4 - SI) and $O^2_{axis,EXP}$, we define fast residues as those with $\tau_e \leq$ 40 ps and $O^2_{axis,EXP} \leq 0.4$; these are expected to undergo rapid rotameric transitions between all rotamer states. The correlation between $\langle \Delta F \rangle_{ABF}$ and $\langle \Delta F \rangle_{EXP}$ is shown in Figure 6 for fast methyl-bearing residues. A significant number of data points located near the diagonal line indicating a reasonable correlation between $\langle \Delta F \rangle_{ABF}$ and $\langle \Delta F \rangle_{EXP}$.

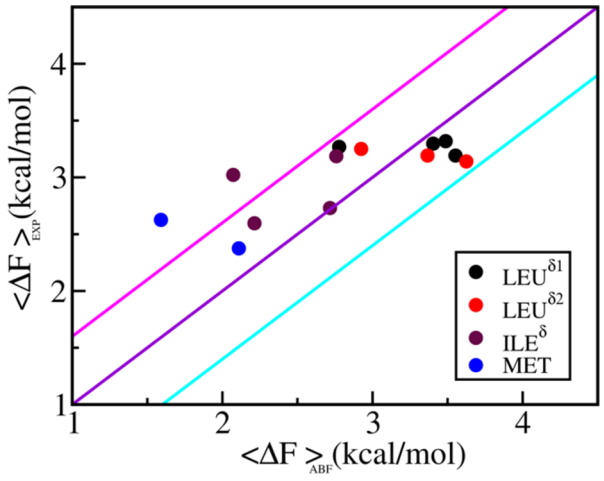


Figure 6. Correlation between $\langle \Delta F \rangle_{ABF}$ and $\langle \Delta F \rangle_{EXP}$ for fast residues is shown. The straight lines (violet, $\langle \Delta F \rangle_{EXP} = \langle \Delta F \rangle_{ABF}$; cyan, $\langle \Delta F \rangle_{EXP} = \langle \Delta F \rangle_{ABF} - 0.6$ kcal/mol; magenta, $\langle \Delta F \rangle_{EXP} = \langle \Delta F \rangle_{ABF} + 0.6$ kcal/mol) are shown as guides to eyes.

A promising new, X-ray diffraction-based experimental technique that measures the noise-filtered electron densities around side-chain dihedral angles is beginning to provide useful information on the side-chain conformational distributions of proteins in crystals.[98,99] The side-chain dihedral angle-dependent electron density obtained from this technique for different methyl-containing residues reveals information on different possible conformers and associated probabilities for these side chains in proteins: the dihedral angle at which the electron density is highest corresponds to the most populated conformer, while weak electron densities correspond to less populated conformers. The most direct way to validate the multimodal distribution of barriers observed here is to compare the conformational distributions of methyl-containing residues in proteins derived from this electron density-based approach. Since X-ray diffraction-based electron densities are available for five of the model systems studied here, we have calculated the distributions of electron densities as a function of ϕ for all methyl-containing residues using the RINGER program.[98,99] Figure S5 (SI) shows the charge density as a function of ϕ for a few representative residues together with the free energy profiles obtained from the ABF method. A reasonable correlation is evident between the electron density profiles and the free energy profiles: the maxima in the electron density profiles coincide with the minima in the corresponding free energy profiles. Although the heterogeneous distribution of electron density barriers separating different rotamer states is evident, to better understand the side-chain conformational barrier distributions in proteins a precise determination of the conformational free energy profiles from the corresponding electron density distributions on per-residue basis is much needed.

Barrier Distribution for Different Residue Types. Figure 7 shows barrier distributions for individual residue types, and Table 1 provides statistics on the contributions of barriers associated with different rotameric transitions to the peaks in $P(\Delta F)$ for all residue types. The barriers for all MET residues are less than 5 kcal/mol and thus lead to the first shoulder of the overall barrier distribution (see Figure 4). The $P(\Delta F)$ for MET consists of two peaks: the peak at ∼2 kcal/mol is due to the barriers between the *trans* and *gauche* conformers while the barriers between g^+ and g^- conformers lead to the other peak at ∼4.5 kcal/mol[100] (Table 1 and Figure 7b). Since the number of barriers between the *trans* and *gauche* states is twice that between the g^+ and g^- states, the intensity of the peak at ∼4.5 kcal/mol is almost half that of the peak at ∼2 kcal/mol. The smaller barriers for the symmetry axis dynamics of MET methyls observed here suggest that MET methyls experience less steric hindrance in proteins, leading to fast transitions between different conformers.[100]

The distributions for LEU prochiral methyls (LEU$^{\delta1}$ and LEU$^{\delta2}$) consist of three peaks at ∼3, ∼ 5, and ∼7 kcal/mol. For LEU$^{\delta2}$ methyls, the activation barriers for $t \rightarrow g^+$ and $g^- \rightarrow g^+$ rotameric transitions are relatively high ($\Delta F >$ 6 kcal/mol) and they lead to the peak at ∼7 kcal/mol while the barriers for $t \rightleftharpoons g^-$ transitions lead to the peak at ∼3 kcal/mol and the remaining barriers form the peak at ∼5 kcal/mol. Similarly, for LEU$^{\delta1}$ methyls, $t \rightarrow g^-$ and $g^+ \rightarrow g^-$ barriers lead to the peak at ∼7 kcal/mol, while the barriers for $t \rightleftharpoons g^+$ transitions lead to the peak at ∼3 kcal/mol, and the remaining barriers lead to the peak at ∼5 kcal/mol. Although $P(\Delta F)$ for LEU prochiral methyls are expected to be identical, comparison of the distributions for LEU$^{\delta1}$ and LEU$^{\delta2}$ indicates that $P(\Delta F)$ for LEU$^{\delta1}$ has sharp peaks, while the peaks are relatively broader for LEU$^{\delta2}$. Also, the fraction of barriers with $\Delta F >$ 8 kcal/mol or $\Delta F <$ 3 kcal/mol is higher in LEU$^{\delta2}$ than in LEU$^{\delta1}$. These differences can be attributed to variation in the nonbonded interactions arising from the microenvironment, and a closer examination reveals that the distributions of near-neighbor heavy atoms around the carbons of LEU$^{\delta1}$ and LEU$^{\delta2}$ methyl groups of some LEU residues can vary significantly (Figure S6 - SI). The broader distribution for LEU$^{\delta2}$ residues suggests a greater sensitivity of these conformational free energy profiles to the nonbonded interactions. Thus, LEU$^{\delta2}$ is expected to be dynamically more heterogeneous than LEU$^{\delta1}$ with a broader distribution of O^2_{axis}. A detailed discussion on the influences of $P(\Delta F)$ on O^2_{axis} distributions and $\langle O^2_{axis} \rangle$ is presented in a later section. The difference in the conformational sensitivity of LEU$^{\delta1}$ and LEU$^{\delta2}$ to microenvironmental changes observed here is in line with the experimental observation that these prochiral methyls respond differently to external magnetic fields with different relaxation parameters.[11,45,101,102]

The distribution for ILE$^{\delta}$ is also multimodal, with a broad peak below 4 kcal/mol and two more peaks at ∼5.5 and ∼8.5 kcal/mol. The barriers for $g^- \rightleftharpoons g^+$ rotameric transitions lead to the peak at ∼8.2 kcal/mol, while the barriers for $t \rightleftharpoons g^+$ transitions mainly contribute to the peaks observed below 4

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

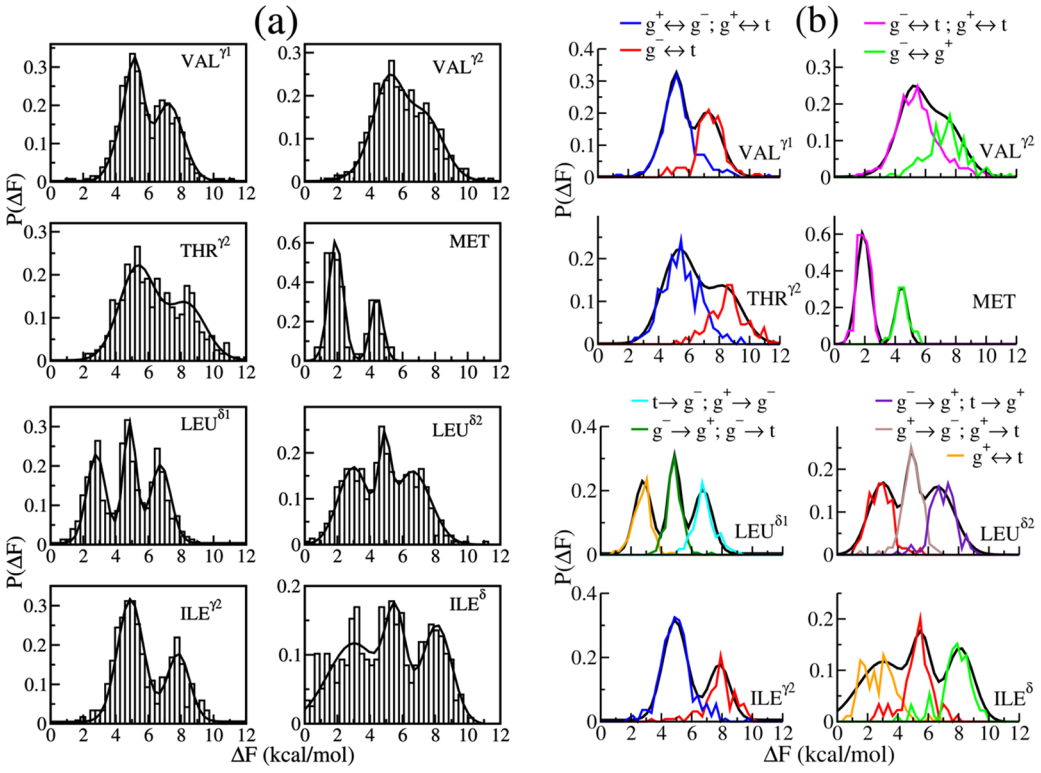


Figure 7. (a) Distribution of free energy barriers for different types of residues of all proteins studied. (b) $P(\Delta F)$ for specific side chain rotameric transitions shown for different residue types.

Table 1. Contributions of Barriers for Specific Side Chain Rotameric Transitions to Different Peaks in $P(\Delta F)$ for Different Residue Types in Proteins

class	residue type	peak I ~3 kcal/mol (~2 kcal/mol for MET)	peak II ~5 kcal/mol (~4.5 kcal/mol for MET)	peak III ~7.5 kcal/mol
Class 1	VAL$^{\gamma 1}$	–	$g^+ \rightleftharpoons g^-$ $g^+ \rightleftharpoons t$	$g^- \rightleftharpoons t$
	VAL$^{\gamma 2}$	–	$g^- \rightleftharpoons t$ $g^+ \rightleftharpoons t$	$g^+ \rightleftharpoons g^-$
	THR$^{\gamma 2}$	–	$g^+ \rightleftharpoons g^-$ $g^+ \rightleftharpoons t$	$g^- \rightleftharpoons t$
	ILE$^{\gamma 2}$	–	$g^+ \rightleftharpoons g^-$ $g^+ \rightleftharpoons$ t	$g^- \rightleftharpoons t$
Class 2	LEU$^{\delta 1}$	$g^+ \rightleftharpoons t$	$g^- \rightarrow g^+$ $g^- \rightarrow t$	$t \rightarrow g^-$ $g^+ \rightarrow g^-$
	LEU$^{\delta 2}$	$g^- \rightleftharpoons t$	$g^+ \rightarrow g^-$ $g^+ \rightarrow t$	$t \rightarrow g^+$ $g^- \rightarrow g^+$
	ILE$^{\delta}$	$g^+ \rightleftharpoons t$	$g^- \rightleftharpoons t$	$g^+ \rightleftharpoons g^-$
Class 3	MET	$g^- \rightleftharpoons t$ $g^+ \rightleftharpoons t$	$g^+ \rightleftharpoons g^-$	–

kcal/mol. Among residues with three or more peaks in $P(\Delta F)$, only ILE$^{\delta}$ exhibits a relatively broad distribution, with a peak at a relatively high value of ΔF (~8.5 kcal/mol) and with a significant number of barriers greater than 8.5 kcal/mol. A broad $P(\Delta F)$ implies that ILE$^{\delta}$ is expected to show a high degree of dynamic heterogeneity with a relatively broad distribution of O^2_{axis}.

The $P(\Delta F)$ distributions for VAL$^{\gamma 1}$, VAL$^{\gamma 2}$, THR$^{\gamma 2}$, and ILE$^{\gamma 2}$ consist of only two peaks at ~5 and ~8 kcal/mol. The barriers for $t \rightleftharpoons g^-$ transitions lead to the peak at ~8 kcal/mol for VAL$^{\gamma 1}$, THR$^{\gamma 2}$, and ILE$^{\gamma 2}$ methyl groups, while the remaining barriers lead to the peak at ~5 kcal/mol. The features of these bimodal distributions differ among these residues: $P(\Delta F)$ is relatively broad for THR$^{\gamma 2}$ and VAL$^{\gamma 2}$, while it is sharply peaked for ILE$^{\gamma 2}$. The differences between $P(\Delta F)$ of VAL prochiral methyls (VAL$^{\gamma 1}$ and VAL$^{\gamma 2}$) can be attributed to the variation of their microenvironment as shown in Figure S6 (SI). The absence of a peak at ~3 kcal/mol suggests that the average O^2_{axis} for these residues is likely to be greater than other methyl-containing residues (see later).

A common pattern that emerges from the above analyses of the distributions of barriers for individual residue types is that all δ-methyls in proteins exhibit a trimodal $P(\Delta F)$ with each of three pairs of rotameric transitions (among six possible transitions per residue) contributing to different peaks in $P(\Delta F)$. In contrast, all γ-methyls exhibit a bimodal $P(\Delta F)$ with a pair of rotameric transitions contributing to the peak at ~8 kcal/mol and the remaining transitions contributing to the other peak at ~5 kcal/mol. The peak in $P(\Delta F)$ observed at ~8 kcal/mol for all residue types (except MET) corresponds to slower rotameric transitions involving crossing of the rate-limiting barriers, while the peak at ~3 kcal/mol observed in $P(\Delta F)$ of δ-methyls corresponds to faster conformational transitions. The following empirical relationship between the peak positions in the barrier distributions and the average thermal energy per degree of freedom at room temperature (i.e., $k_BT/2$ at T = 300 K) is evident from $P(\Delta F)$: $\Delta F_n = 9n(k_BT/2)$, where ΔF_n is the n^{th} peak position with n = 1, n = 2, and n = 3 corresponding respectively to the first, second, and third peaks in $P(\Delta F)$; k_B is the Boltzmann's constant; and the factor 9 is due to the nine X−C−C−Y (Y denotes the methyl carbon in the present study) dihedral contributions to the total conformational free energy associated with the rotation about a given C−C bond.

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

Average Barrier Distribution. The residues that exhibit trimodal $P(\Delta F)$ are grouped into "Class-2", while those with a bimodal $P(\Delta F)$ (excluding MET) are grouped as "Class-1". "Class-3" contains the MET residues. Figure 8 shows the

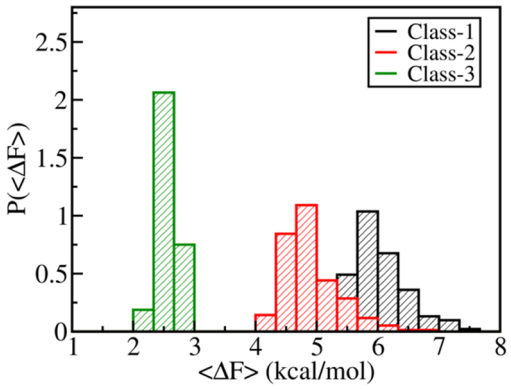


Figure 8. Distribution of average free energy barriers for three classes of side chain of all proteins.

distribution of $\langle \Delta F \rangle$, which is the average of six barriers per residue, and it was calculated separately for each residue, plotted for different classes of side chain: Class-1 ($VAL^{\gamma 1}$, $VAL^{\gamma 2}$, $THR^{\gamma 2}$, and $ILE^{\gamma 2}$), Class-2 ($LEU^{\delta 1}$, $LEU^{\delta 2}$, and ILE^{δ}), and Class-3 (MET). The $P(\langle \Delta F \rangle)$ distributions clearly distinguish these classes; the distribution for Class-2 residues is peaked around 4.8 kcal/mol, whereas Class-1 residues are peaked around 6 kcal/mol. Since Class-1 residues contribute mainly to the second and third peaks of the overall barrier distribution, their average barrier values are higher than those for other residue types belonging to other classes. The observation that $\langle \Delta F \rangle$ is smaller for Class-2 residues than the side chains of Class-1 implies that the degree of conformational sampling by Class-2 residues is higher than Class-1, and thus, for a given O^2_{axis} a Class-2 residue is expected to have a conformational entropy slightly higher than that of a Class-1 residue. This free energy-based classification of side chains is consistent with the two-parameter classification (based on O^2_{axis} and S_{conf}) of side chain dynamics in proteins.[115]

Barrier Map-Based Classification of Side Chains. The rotational diffusive motion of a side chain in a particular rotamer state is influenced by a pair of barriers surrounding that state. The i^{th} rotamer state ($i = 1$, $i = 2$, and $i = 3$ correspond to t, g^-, and g^+ rotamer states, respectively) of a given side chain has a forward barrier (ΔE^f_i) and a reverse barrier (ΔE^r_i). Let ΔE_{min} and ΔE_{max} denote the smaller and larger of ΔE^f_i and ΔE^r_i, respectively. The i^{th} rotamer state is said to be rotationally diffusive on both sides (i.e., bidirectionally diffusive as shown in Figure 9c) if both ΔE_{min} and ΔE_{max} are small (less than 5 kcal/mol), nondiffusive or restricted if both ΔE_{min} and ΔE_{max} are high (greater than 6 kcal/mol), and partially diffusive (unidirectional diffusion as shown in Figure 9b) if one of the barriers is small (less than 5 kcal/mol) and the other is large. These different types of side chain rotamer states in proteins are illustrated in Figure 9.

The ABF-based free energy profiles provide a direct access to ΔE_{min} and ΔE_{max} for all side chain rotamer states. Figure 10a shows ΔE_{min} versus ΔE_{max} (referred to here as the "barrier map") for all rotamer states of all methyl-containing residues of all proteins studied here. Each data point in this "barrier map" corresponds to a rotamer state. Since $\Delta E_{min} \leq \Delta E_{max}$, the data

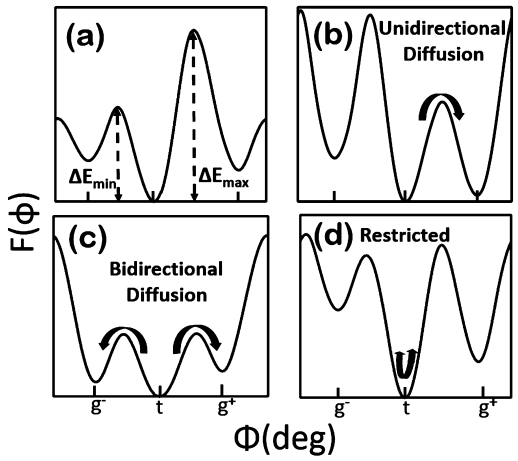


Figure 9. Schematic diagram illustrating different types of side chain rotamers based on ΔE_{min} and ΔE_{max}.

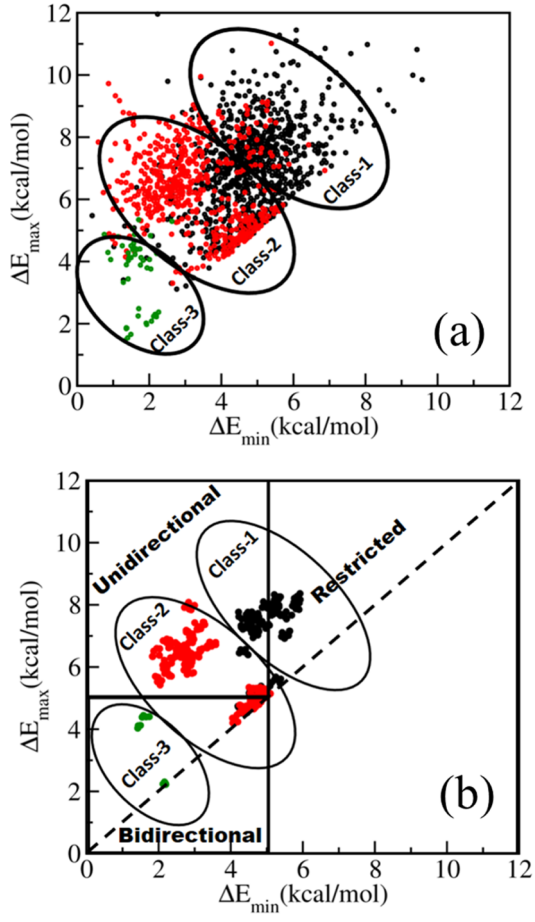


Figure 10. Barrier map representing the variation of minimum energy barrier (ΔE_{min}) with maximum energy barrier (ΔE_{max}) for (a) all rotamer states (b) high-density (greater than 66.67 $(kcal/mol)^{-2}$) regions of $\Delta E_{min}/\Delta E_{max}$ space are shown.

points are expected to be located only along the diagonal or the upper diagonal of the plot in Figure 10a. The diagonal and off-diagonal points represent the rotamer states with iso-energetic (or symmetric) and unequal (or asymmetric) barriers on either side, respectively. The degree of deviation from the diagonal and the closeness to the origin of a given off-diagonal point measure the unequalness or asymmetry in the barrier heights

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

and the degree of rotational diffusive nature of the corresponding rotamer state, respectively.

The data for Class-3 (MET) residues (Figure S7 - SI) are closer to the origin with both ΔE_{min} and ΔE_{max} less than 5 kcal/mol for almost all rotamer states. Two clusters of Class-3 residues are observed: one near the diagonal with $\Delta E_{max} = \Delta E_{min} = \sim 2.0$ kcal/mol and the second with $\Delta E_{max} = \sim 4.0$ kcal/mol and $\Delta E_{min} = \sim 2.0$ kcal/mol. The data for Class-1 ($VAL^{\gamma 1}$, $VAL^{\gamma 2}$, $THR^{\gamma 2}$, and $ILE^{\gamma 2}$) residues are widely spread with a majority of data points with $\Delta E_{min} > 5.0$ kcal/mol and $\Delta E_{max} > 7.0$ kcal/mol, while the data for Class-2 ($LEU^{\delta 1}$, $LEU^{\delta 2}$, and ILE^{δ}) residues are clustered into two groups: one near the diagonal with $\Delta E_{max} = \Delta E_{min} = \sim 5.0$ kcal/mol and the other located around $\Delta E_{max} = \sim 7.0$ kcal/mol and $\Delta E_{min} = \sim 3.0$ kcal/mol. Though the classification of side chains based on the values of ΔE_{max} and ΔE_{min} is unambiguous for many residues, some degree of overlap is observed between Class-1 and Class-2 residues.

Figure 10b shows only the high-density regions of the "barrier map" depicted in Figure 10a. These high-density regions were identified by dividing the "barrier map" into 0.3×0.3 $(kcal/mol)^2$ square grids, and only those grids with density greater than or equal to 66.7 $(kcal/mol)^{-2}$ are shown in Figure 10b. The Class-3 residues fall in the region where all rotamer states would be bidirectionally diffusive, whereas most of the Class-2 residues fall in the region corresponding to unidirectionally diffusive rotamer states. Some Class-2 residues (mainly LEU residues) falling on the diagonal have bidirectionally diffusive rotamer states. Most Class-1 residues exhibit restricted dynamics.

O^2_{axis} Distribution and Experimental Comparison. The distribution of O^2_{axis} is a good measure of the degree of dynamical heterogeneity in proteins: the broader the O^2_{axis} distribution the more is the dynamical heterogeneity. Understanding the molecular origins of the O^2_{axis} distributions and how they change with temperature and pressure is a topic of current interest, but a comprehensive understanding is yet to be achieved.[13,60,103,104] The O^2_{axis} distribution for calmodulin obtained from site-specific NMR experiments showed a trimodal O^2_{axis} distribution corresponding to three classes of residue: J-,ω-, and α-classes.[105] The J-class residues have $O^2_{axis} \approx 0.3$ due to frequent rotameric interconversion between different rotamer states, while $O^2_{axis} \approx 0.9$ for ω-class residues due to their restricted motion within a rotamer well. The α-class side chains exhibit restricted motion within a rotamer state with occasional jumps between different rotamer states, giving rise to $O^2_{axis} \approx 0.6$.[105] The O^2_{axis} distribution for calmodulin was found to be almost insensitive to the variation of temperature between 288 and 346 K.[13] The trimodal distribution of O^2_{axis} was also observed for calmodulin in a recent molecular dynamics simulation study.[67] A large number of proteins exhibited multimodal distributions that differ among proteins, indicating the nonuniversal nature of the O^2_{axis} distribution.[11,60]

The calculated $O^2_{axis,MD}$ and $O^2_{axis,ABF}$ are compared with the corresponding experimental values ($O^2_{axis,EXP}$) for six systems studied here,[75−77,106−108] and the results are shown in Figure 11. $O^2_{axis,MD}$ correlates reasonably well with $O^2_{axis,EXP}$ and the estimated correlation coeffcients (0.58−0.87) for these systems are within the range repored in earlier MD simulation studies of proteins.[41,59,60,65,109] The observation that $O^2_{axis,ABF} < O^2_{axis,EXP}$ for many residues and a fair agreement between $O^2_{axis,MD}$ and $O^2_{axis,EXP}$ point to the interplay between the kinetic and thermodynamic factors in the determination of the calculated

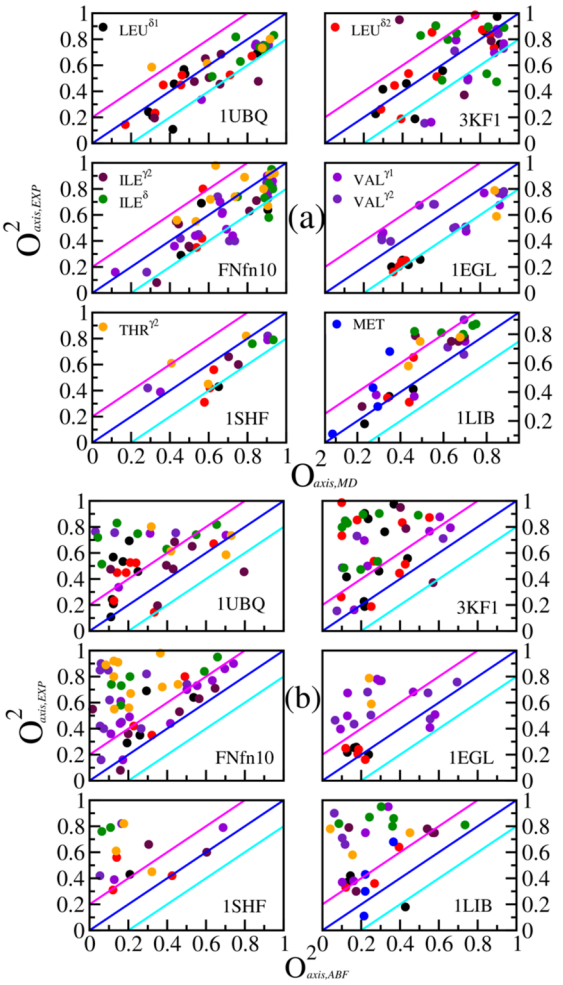


Figure 11. Comparison of the calculated (a) $O^2_{axis,MD}$ (b) $O^2_{axis,ABF}$ with experimental $O^2_{axis,EXP}$ is shown for six out of eight proteins studied here.

O^2_{axis} values. Since the ABF method is devoid of sampling problems, the conformational distributions derived from the free energy profiles obtained from this enhanced-sampling approach represent the thermodynamic equilibrium distributions, and $O^2_{axis,ABF}$ calculated from these distributions (using eq 5) can be considered as equilibrium values that are free from kinetic effects. A significant improvement in the correlation between $O^2_{axis,ABF}$ and $O^2_{axis,EXP}$ is observed (Figure S8 - SI) when limiting the comparison to only fast side chains ($\tau_e \leq 30$ ps and $O^2_{axis,EXP} \leq 0.4$) that are less susceptible to slow rotameric transitions. A systematic under-estimation of $O^2_{axis,ABF}$ (that is, $O^2_{axis,ABF} < O^2_{axis,EXP}$) observed for all $ILE^{\gamma 2}$ methyl groups (see Figure S8 - SI) was earlier attributed to the inaccuracy of the current empirical force field parameters in describing some side chain dynamics of proteins.[67] However, the $O^2_{axis,MD}$ were calculated from MD trajectories that suffer from sampling problems and some rotamer states that are not being visited during the course of these simulations (see Figure 3). Thus, kinetic factors play an important role in determining the apparent side chain conformational distributions and $O^2_{axis,MD}$ (derived from these MD trajectories), which may deviate significantly from the expected equilibrium distributions and $O^2_{axis,ABF}$, respectively. Since $O^2_{axis,EXP}$ obtained from NMR experiments are also sensitive to picosecond−nanosecond time-scale motions and some slow rotameric transitions may

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

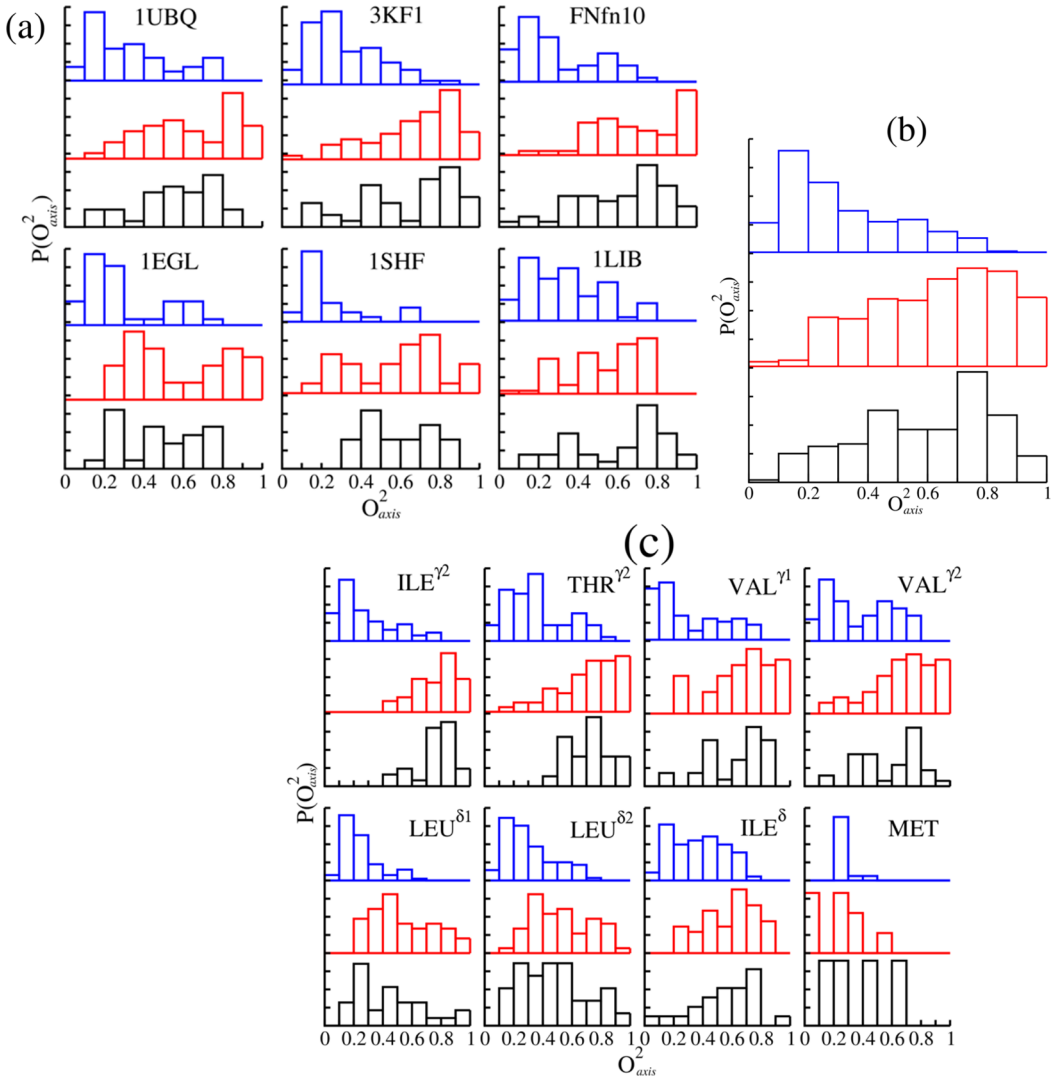


Figure 12. Distributions of O^2_{axis} for (a) individual proteins, (b) overall, (c) different residue types are shown. $P(O^2_{axis,EXP})$ (black), $P(O^2_{axis,MD})$ (red), and $P(O^2_{axis,ABF})$ (blue).

not be picked up by the experimental O^2_{axis}, a fair correlation between $O^2_{axis,MD}$ and $O^2_{axis,EXP}$ is justified. There is, of course, also scope for improvement of the empirical force field parameters to accurately describe the side chain dynamics, and this might potentially improve the correlation between the experimental and calculated order parameters.[110−112]

Figure 12a shows $P(O^2_{axis,MD})$, $P(O^2_{axis,ABF})$, and $P(O^2_{axis,EXP})$ for the different proteins under consideration in the present study. The $P(O^2_{axis,MD})$ compares reasonably well with $P(O^2_{axis,EXP})$ for all proteins, and the overall $P(O^2_{axis,MD})$ and $P(O^2_{axis,EXP})$ (shown in Figure 12b) exhibit the following similar features: a maximum at $O^2_{axis} \approx 0.75$ and a secondary peak at $O^2_{axis} \approx 0.45$. However, the individual and overall $P(O^2_{axis,ABF})$ distributions differ significantly from the corresponding experimental and MD distributions. $P(O^2_{axis,ABF})$ for all proteins shows an intense peak at $O^2_{axis,ABF} \approx 0.2-0.3$, and some exhibit secondary peaks at $O^2_{axis,ABF} \approx 0.5-0.6$ (3KF1, 1LIB and FNfn10) and $O^2_{axis,ABF} \approx 0.7$ (1UBQ). The overall $P(O^2_{axis,ABF})$ exhibits a peak at $O^2_{axis,ABF} \approx 0.2$, and the intensity decreases gradually with increasing values of $O^2_{axis,ABF}$. The observed deviation of $P(O^2_{axis,ABF})$ from $P(O^2_{axis,EXP})$ and $P(O^2_{axis,MD})$ can be attributed again to kinetic effects that play an important role in the determination of the two latter distributions, while $P(O^2_{axis,ABF})$ is solely governed by thermodynamics.

The $P(O^2_{axis})$ for different residue types (shown in Figure 12c) indicate that the O^2_{axis} distributions vary widely among residue types. Here again, $P(O^2_{axis,MD})$ qualitatively reproduces the essential features of the experimental $P(O^2_{axis,EXP})$ for many methyl-containing residue types: the positions of the most intense peaks in $P(O^2_{axis,MD})$ and $P(O^2_{axis,EXP})$ coincide reasonably well for ILE$^{\gamma 2}$, LEU$^{\delta 2}$, VAL$^{\gamma 1}$, and ILE$^{\delta}$ methyl groups. $P(O^2_{axis})$ for the prochiral methyls of LEU and VAL residues differ slightly consistently with the variation in $P(\Delta F)$ for these methyl groups. The $P(O^2_{axis})$ distributions for ILE$^{\delta}$ are relatively broader than those for ILE$^{\gamma 2}$, indicating a wide variation in the conformational flexibility of ILE$^{\delta}$ methyls in proteins in accordance with a broader $P(\Delta F)$ distribution observed for ILE$^{\delta}$ residues.

The percentage, $P_{\%}(O^2_{axis,\dagger})$, of methyl groups of a given residue type with O^2_{axis} less than a given value $(O^2_{axis,\dagger})$ can be determined from $P(O^2_{axis})$ using the following equation

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

$$P_{\%}(O^2_{axis,\dagger}) = \frac{\int_0^{O^2_{axis,\dagger}} P(O^2_{axis})\, dO^2_{axis}}{\int_0^1 P(O^2_{axis})\, dO^2_{axis}} \times 100.0 \quad (14)$$

$P_{\%}(O^2_{axis,\dagger})$ increases nonlinearly (sigmoidal-type curve) with $O^2_{axis,\dagger}$ for all classes of residue, and the calculated $P_{\%}(O^2_{axis,\dagger})$ versus $O^2_{axis,\dagger}$ curves almost reproduce the corresponding experimental curves (Figure 13a). The rate of growth of

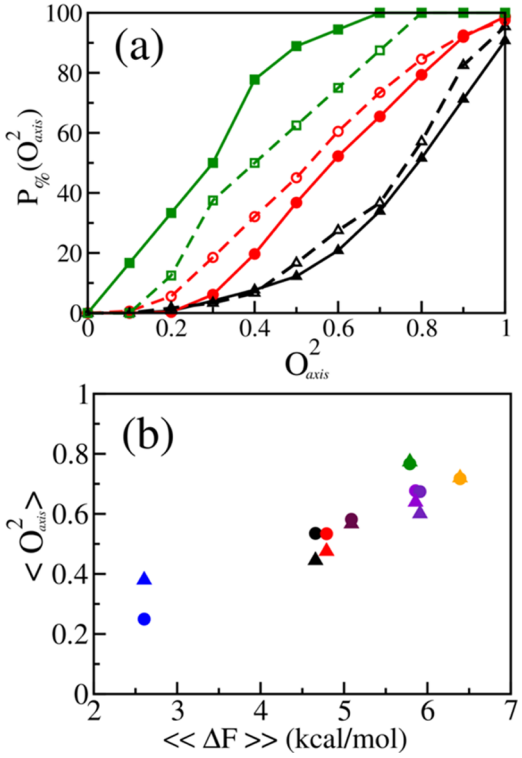


Figure 13. (a) Experimental (dashed lines) and MD-derived (solid lines) correlation between $P_{\%}(O^2_{axis})$ and O^2_{axis} shown for Class-1 (black), Class-2 (red), and Class-3 (green) residues (b) Correlation of experimental $\langle O^2_{axis,EXP}\rangle$ (filled triangle), MD-derived $\langle O^2_{axis,MD}\rangle$ (filled circle) with the average free energy barrier $\langle\langle \Delta F\rangle\rangle$ shown for different residue types: MET (blue), ILE$^{\delta}$ (maroon), LEU$^{\delta 1}$ (black), LEU$^{\delta 2}$ (red), ILE$^{\gamma 2}$ (green), VAL$^{\gamma 1}$ (indigo), VAL$^{\gamma 2}$ (violet), and THR$^{\gamma 2}$ (orange).

$P_{\%}(O^2_{axis,\dagger})$ with $O^2_{axis,\dagger}$ is different for different classes of residues; $P_{\%}(O^2_{axis,\dagger})$ for Class-1 residues is negligible for $O^2_{axis,\dagger} < 0.5$, while a significant number of Class-2 residues and ~90% of Class-3 residues have $O^2_{axis,\dagger} < 0.5$. The percentage of methyl groups with $O^2_{axis} < 0.5$ is greater for δ-methyls than that for γ-methyls, indicating a relatively higher conformational flexibility of δ-methyls than the γ-methyls in proteins.

Figure13b shows the correlation between $\langle O^2_{axis}\rangle$ and $\langle\langle \Delta F\rangle\rangle$ (these averages were calculated by averaging over all residues of a given type) for different residue types. $\langle O^2_{axis}\rangle$ increases almost linearly with $\langle\langle \Delta F\rangle\rangle$ and the calculated dependence of $\langle O^2_{axis}\rangle$ on $\langle\langle \Delta F\rangle\rangle$ is consistent with that determined from the experimental data. The $\langle O^2_{axis}\rangle$ and $\langle\langle \Delta F\rangle\rangle$ for Class-1 residues are higher than those for Class-2 residues, whereas Class-3 residues have the lowest $\langle O^2_{axis}\rangle$ and $\langle\langle \Delta F\rangle\rangle$. NMR experiments have shown a general increase in the width of $P(O^2_{axis})$ and a general decrease in the value of O^2_{axis} with increasing separation of a methyl group from the main chain, suggesting that the longer the side chain, the more variable is the dynamics.[104] These experiments also demonstrated that $P(O^2_{axis})$ for all δ-methyl-containing residues (ILE, LEU) (Class-2) are similar and are broader than those for all γ-methyl-containing residues (VAL, THR) (Class-1), indicating that it is less likely for a longer side chain to have a higher O^2_{axis} value. The higher values of $\langle O^2_{axis}\rangle$ and slower rate of growth of $P_{\%}$ with O^2_{axis} observed for Class-1 residues and the classification of residues based on free energy barriers are consistent with these experimental results.

Probability Flux and Conformational Entropy. When classifying the side chains above based on $\Delta E_{min}/\Delta E_{max}$, only the barriers neighboring individual rotamer states were considered. However, the degree of conformational sampling by a side chain is determined by whether successive rotamer states are diffusive or not. A side chain having a diffusive rotamer state with two nondiffusive adjacent rotamer states is expected to have different motional parameters and thermodynamic properties than a diffusive rotamer state surrounded by other diffusive states. This information about how successive diffusive rotamer states are correlated was absent in the classification based on barrier distribution and $\Delta E_{min}/\Delta E_{max}$.

Using the principle of detailed balance, we now examine the probability for a side chain to visit all rotamer states. There are two possible closed loops of rotamer states by which a given side chain can explore the conformational landscape: $t \rightarrow g^+ \rightarrow g^- \rightarrow t$ (referred to here as the forward loop) and $t \rightarrow g^- \rightarrow g^+ \rightarrow t$ (referred to here as the reverse loop). Each rotameric transition in a loop obeys the principle of detailed balance; for example, for the transition between t and g^+ $P(t)P(t \rightarrow g^+) = P(g^+)P(g^+ \rightarrow t)$, where $P(t)$ and $P(g^+)$ are the equilibrium probabilites of being in t and g^+ states, respectively, and $P(t \rightarrow g^+)$ and $P(g^+ \rightarrow t)$ are the transition probabilities for $t \rightarrow g^+$ and $g^+ \rightarrow t$ transitions, respectively. We define the probability flux for the forward loop as $J_F = \prod_{i=1}^{3} P(i)P(i \rightarrow i+1)$, where i = 1,2,3 denote t, g^+, and g^- states, respectively and $P(3 \rightarrow 4) = P(3 \rightarrow 1)$ due to periodicity of the conformational energy profile. Similarly, the probability flux for the reverse loop is defined as $J_R = \prod_{i=3}^{1} P(i)P(i \rightarrow i-1)$ and here $P(1 \rightarrow 0) = P(1 \rightarrow 3)$. The equilibrium probability of being in state i is given by $P(i) = e^{-\beta F_i(\phi)}/(\int_0^{2\pi} e^{-\beta F_i(\phi)}\, d\phi)$ and the transition probability $P(i \rightarrow j) \propto e^{-\beta \Delta F_{ij}}$ where ΔF_{ij} is the free energy barrier separating i and j states. The probability fluxes J_F and J_R depend on the energies, $F_i(\phi)$, of the rotamer states and the barriers separting the rotamer states.

The normalized forward probability flux, which varies from 0 to 1, is defined as follows,

$$J_F^N = \frac{\ln\left(\frac{J_F}{J_F^{min}}\right)}{\ln\left(\frac{J_F^{max}}{J_F^{min}}\right)} \quad (15)$$

where J_F^{min} and J_F^{max} are the minimum and maximum values of J_F and J_F^N denotes the normalized J_F. The condition of detailed balance, $J_F^N = J_R^N$, was verified for all residues (Figure S9 - SI). The larger the value of J_F^N the higher is the conformational sampling. Figure 14a shows J_F^N as a function of O^2_{axis} for different residues of all proteins. J_F^N increases with decreasing O^2_{axis}; residues with larger O^2_{axis} have smaller J_F^N than that of residues with smaller O^2_{axis}. The MET residues have highest $J_F^N \approx 1.0$, and Class-2 residues have slightly higher J_F^N than Class-1 residues at all values of O^2_{axis}.

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

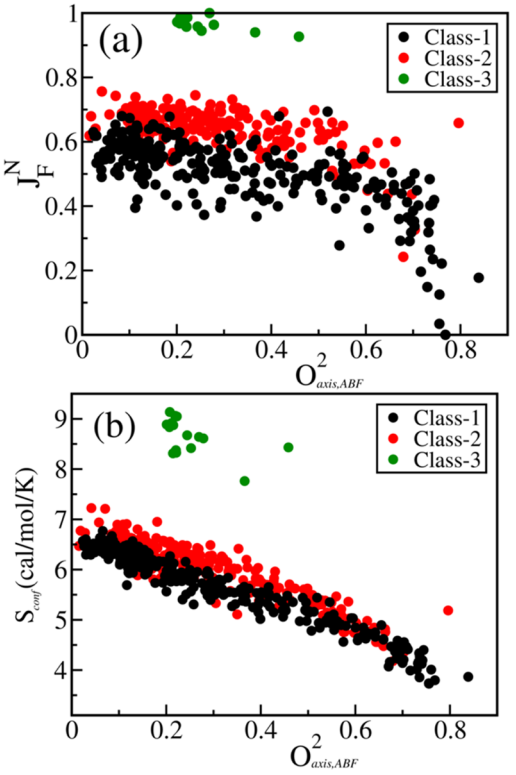


Figure 14. (a) Variation of probability flux (J_F^N) with $O^2_{axis,ABF}$ and (b) variation of conformational entropy (S_{conf}) with $O^2_{axis,ABF}$ for different classes of side chain of all proteins studied.

The conformational entropy (S_{conf}) due to the reorientational dynamics of the methyl symmetry axis also distinguishes different classes of side chain in proteins. Using the correlation between S_{conf} and O^2_{axis} and an assumption that the dynamics of a given methyl group reports on local disorder around it, an "entropy meter" was introduced recently to determine the changes in the conformational entropy and the total binding entropy of a protein upon binding to a ligand directly from the average changes in residue-weighted O^2_{axis}.[105,109,113] Since the calculated O^2_{axis} reported here for any given methyl group is also sensitive to its microenvironment, it is of interest to examine the correlation between S_{conf} and $O^2_{axis,ABF}$. Figure 14b shows S_{conf} versus $O^2_{axis,ABF}$. S_{conf} decreases with increasing $O^2_{axis,ABF}$ and their nonlinear dependence on each other is consistent with experimental and other simulation studies.[114,115] The MET residues have the largest conformational entropy, around 8.5 cal/mol/K, while S_{conf} for other residues spans between 4 and 7 cal/mol/K. The Class-2 residues have slightly higher conformational entropy than Class-1 residues, especially for $O^2_{axis,ABF}$ < 0.4. The difference in S_{conf} between Class-1 and Class-2 residues can be attributed to the fact that the conformational landscape for Class-1 residues consists of relatively higher free energy barriers than those of Class-2 residues. The observation that two side chains with the same $O^2_{axis,ABF}$ but belonging to different classes can have different S_{conf} is in accord with an earlier investigation.[115] The dependence of S_{conf} on $O^2_{axis,ABF}$ for different residue types is shown in Figure S10 (SI).

Correlation Between O^2_{axis} and Rotamer Populations. The free energy profiles, $F(\phi)$, obtained from the ABF simulations were used to determine the probability distribution of ϕ, $P(\phi)$, for individual residues. $P(\phi)$ for all residues except ALA consisted of three peaks corresponding to three rotamer states. These rotamer states were classified into major, intermediate, and minor states, depending on the intensities of the corresponding $P(\phi)$ peaks; the highest- and lowest-intensity peaks correspond, respectively, to the major and minor states. Depending upon the nature of the nonbonded interactions arising from the microenvironment surrounding a side chain, any one of the t, g^+, or g^- states can be a major rotamer state for the side chain. The statistical distribution of the major rotamer states for different residue types in all proteins investigated here is provided in Table S2 ("Statistics on Side Chain Rotamer Sequences" section in the SI). For instance, t is the major rotamer state for ~83% of ILE$^{\delta}$, ~ 67% of MET, and ~48% of LEU$^{\delta 2}$ residues.

The relative population of the three rotamer states was calculated using the following equation:

$$\rho_i = \int_{\phi_i-\pi/3}^{\phi_i+\pi/3} P(\phi)\, d\phi \qquad (16)$$

where ϕ_1, ϕ_2, and ϕ_3 denote the values of ϕ for the major, intermediate, and minor states, respectively, and ρ_1, ρ_2, and ρ_3, the corresponding populations. The values of ρ_i range from 0 to 1 and indicate the extent of conformational restriction of a given side chain. A residue with $\rho_1 \approx 1.0$ is expected to exhibit restricted motion mostly within a single rotamer state due to greater steric hindrance, while side chains with $0.3 < \rho_1 < 0.4$ experience relatively less steric hindrance and visit all rotamer states. The lowest possible value for ρ_1 is 1/3 and occurs when all three rotamer states are equally populated. The highest possible value of ρ_2 is 0.5, when both major and intermediate rotamer states are equally populated with zero population in the minor rotamer state.

Since both ρ_i and O^2_{axis} are related to the degree of conformational flexibility of side chains, it is of interest to examine any possible correlation between these two quantities. Figure 15 shows the variation of $O^2_{axis,ABF}$ with ρ_1 and ρ_2 for all

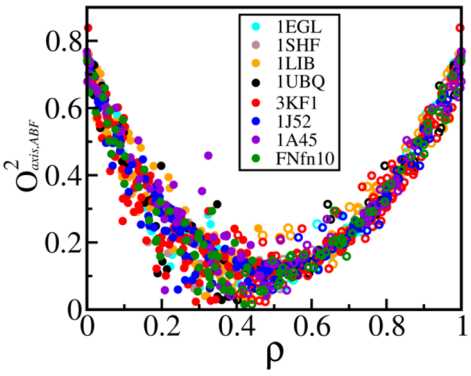


Figure 15. Dependence of $O^2_{axis,ABF}$ on the populations of major (ρ_1 - open circles) and intermediate (ρ_2 - filled circles) states color coded for individual proteins.

proteins. Regardless of the molecular weight and secondary structures of proteins, a universal nonlinear correlation is observed between $O^2_{axis,ABF}$ and the populations of the rotamer states for all proteins. A similar nonlinear correlation between O^2_{axis} and ρ_1 has been observed for protein eglin C.[65] ρ_1 increases with increasing $O^2_{axis,ABF}$, while ρ_2 decreases with $O^2_{axis,ABF}$.

There is some spread in the data for $O^2_{axis,ABF}$ < 0.3 and ρ_2 > 0.2. There are at least two reasons for this spread of data. The

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

first is that the flexible side chains with $O^2_{axis,ABF} < 0.3$ exhibit two types of rotamer transitions; in the first type a side chain visits only the major and intermediate states (two-state dynamics), while all three states are visited in the second mechanism (three-state dynamics). Since ρ_2 can take different values depending on whether it follows two- or three-state dynamics, this spreads the data in Figure 15. The other reason is that the nature of the dependence of $O^2_{axis,ABF}$ on ρ_i is different for different residue types.

The relationship between $O^2_{axis,ABF}$ and ρ_i was examined for different residue types, and the results are presented in Figure 16 for all proteins under consideration. An asymmetric,

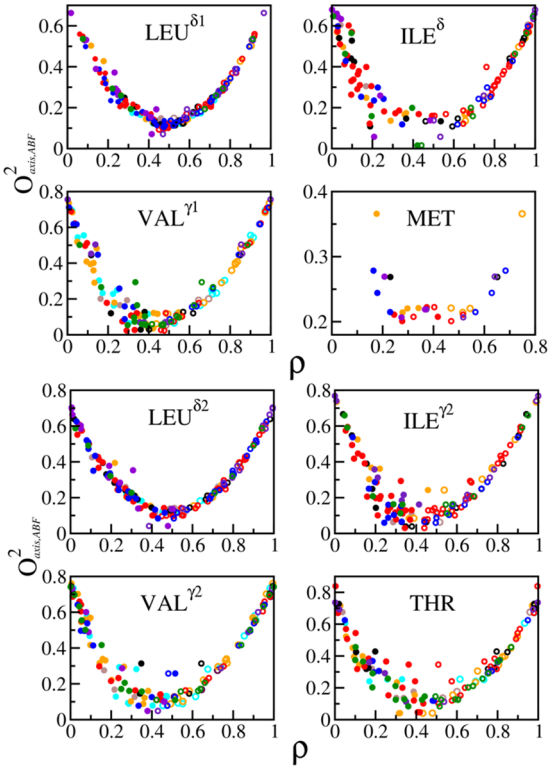


Figure 16. Dependence of $O^2_{axis,ABF}$ on the populations of major (ρ_1 - open circles) and intermediate (ρ_2 - filled circles) states for different types of residue from all proteins using the same color code.

parabolic, nonlinear dependence of $O^2_{axis,ABF}$ on ρ_1 and ρ_2 is observed for all residue types of all proteins. Though the overall parabolic pattern is seen for all residue types, the detailed geometric features, such as the curvature and well depth, are different for different residue types. The $O^2_{axis,ABF}$ versus ρ_i data were fitted using the following equation.

$$O^2_{axis,ABF} = b_0(\rho - b_1)(\rho - b_2) + 0.5(b_1 + b_2) \quad (17)$$

The fitted parameters are listed in Table S1 (SI). The range of $O^2_{axis,ABF}$, i.e., the difference between the maximum and minimum values of $O^2_{axis,ABF}$, is different for different residue types. A majority of residues have $O^2_{axis,ABF} < 0.4$ while $O^2_{axis,ABF}$ for some residues fall between 0.4 and 0.7. Only a few THR$^{\gamma2}$ and VAL$^{\gamma2}$ residues have $O^2_{axis,ABF} > 0.7$. The spread of data in the region defined by $O^2_{axis,ABF} < 0.3$ and $\rho_2 > 0.2$ is low for LEU residues but higher for ILE$^{\gamma2}$ and VAL$^{\gamma1}$. A thermodynamic consequence of the spread of data for $O^2_{axis,ABF} < 0.3$ and $\rho_2 > 0.2$ is that two residues having same $O^2_{axis,ABF}$ can have different conformational entropies depending on whether they follow two- or three-state dynamics. A residue that visits all three states is expected to have higher conformational entropy than one that visits only two states.

CONCLUSIONS

Fast side chain conformational dynamics play a role in the biological function of proteins.[11,30,33,38,116,117] High-resolution solution NMR spectroscopy uses side chain methyl spin probes to determine the amplitudes and rates of site-specific fast internal motions of proteins. In particular, the methyl axial order parameter (O^2_{axis}), which quantifies the spatial restriction of angular motion of the methyl symmetry axis, is commonly used to establish a molecular-level connection between the atomistic dynamics and thermodynamics of proteins.[5,11,13,30,38,40,44,46,47,49,50,63,65,105,113,118] Recognized as a proxy for protein conformational entropy, O^2_{axis} is used as an "entropy-meter" to measure and elucidate the essential role of the side chain entropy in many biological processes including protein−ligand binding, molecular recognition, folding and allosteric regulation.[109,113,119]

Realizing the fact that the side chain dynamics and thermodynamics are governed by the underlying conformational energy landscape, the present study employs an enhanced sampling free energy method (adaptive biasing force) and molecular dynamics simulation to characterize fast side chain dynamics by examining the conformational energy landscape of individual residues of proteins. A set of eight proteins with different molecular weights and secondary structures was chosen, and the side chain conformational free energy profiles for all methyl-bearing residues were computed using ABF. The conformational energy surface for each residue consists of three stable rotamer states separated by energy barriers of differing heights. The features of the free energy profiles (such as the barrier heights and rotamer state energies) differ significantly within and among residue types. The activation energy for inter-rotamer-state conversion in proteins ranges from ∼0.3 to ∼12.0 kcal/mol. The barriers separating the rotamer states follow a trimodal distribution with a major peak at ∼5.0 kcal/mol and two shoulders at ∼3 and ∼7.5 kcal/mol. This trimodal distribution was found to be universal among all proteins under consideration, indicating the existence of globally favored and disfavored barriers for side chain dynamics in proteins.

The relative magnitudes of the pair of barriers surrounding any given side chain rotamer state were used to determine whether the rotamer state is diffusive or restricted; a diffusive rotamer state favors interstate conversion, while restrictive states do not. A side chain with no diffusive rotamer state exhibits a higher O^2_{axis}, while a side chain with all three diffusive states has a lower O^2_{axis}. The free energy barrier-based hierarchical grading of rotamer states enables us to group protein side chains into three dynamic classes: Class-1 (VAL$^{\gamma1}$, VAL$^{\gamma2}$, THR$^{\gamma2}$, and ILE$^{\gamma2}$), Class-2 (LEU$^{\delta1}$, LEU$^{\delta2}$, and ILE$^{\delta}$), and Class-3 (MET). Two methyl-bearing side chains having the same O^2_{axis} but belonging to different classes have differing conformational entropy; the side chains of Class-2 exhibit higher entropy than Class-1 residues. These results provide conformational landscape-based evidence for the two-parameter classification (based upon conformational entropy and O^2_{axis}) of protein side chains reported earlier.[115] The O^2_{axis} correlate nonlinearly with the side chain rotamer populations and the conformational entropy, and these correlations were found to be universal among all proteins studied here.

dx.doi.org/10.1021/ja5024783 | J. Am. Chem. Soc. 2014, 136, 8590−8605

As O^2_{axis}-based high-resolution NMR studies continue to evolve in important ways to elucidate the structure, dynamics and thermodynamics of proteins in natively folded, unfolded, and 'invisible' excited states,[11,46,120,121] the present study adds to the growing knowledge about the characteristics of fast side chain motions in proteins. A major advantage of the computational method proposed here is its ability to directly relate O^2_{axis}, S_{conf}, and the rotamer populations with the conformational landscapes of proteins. The landscape-based interpretation of motional and thermodynamic parameters readily describes the heterogeneous distribution of O^2_{axis} in proteins in terms of a broader distribution of activation barriers separating the rotamer states.

Many biological processes, including ligand binding and allosteric regulation, result in a significant modulation of the protein energy landscape leading to changes in the side chain dynamics, thermodynamics, and rotamer populations.[35,86,105] The characterization of the evolution of side chain rotamer populations and flexibilities at different stages of biological processes is an active area of research, and the nonlinear dependence of O^2_{axis} on rotamer populations established here suggests that the changes in O^2_{axis} and ρ due to perturbations (including ligand binding, protein−protein association, external pressure) are governed by a universal parabolic correlation between these two quantities. The present study has illustrated the use of advanced enhanced sampling methods to gain complementary molecular-level insights into fast conformational dynamics of proteins.

■ ASSOCIATED CONTENT

Supporting Information

Additional figures and tables (Figures S1−S10, Tables S1−S2). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

m.krishnan@iiit.ac.in

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

M.K. acknowledges financial support provided by Department of Science and Technology (DST), New Delhi under SERC Fast Track scheme for Young Scientists (No. SR/FTP/PS-033/2010). R.R.T. acknowledges DST for the Inspire fellowship. J.C.S. acknowledges funding from the National Science Foundation (MCB section).

■ REFERENCES

(1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(2) Daniel, R. M.; Dunn, R. V.; Finney, J. L.; Smith, J. C. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 69−92.
(3) Palmer, A. G., III. *Curr. Opin. Biotechnol.* **1993**, *4*, 385−391.
(4) Kay, L. E.; Muhandiram, D. R.; Farrow, N. A.; Aubin, Y.; Forman-Kay, J. D. *Biochemistry* **1996**, *35*, 361−368.
(5) Kay, L. E. *Nat. Struct. Biol.* **1998**, *5*, 513−517.
(6) Frauenfelder, H.; Parak, F.; Young, R. D. *Annu. Rev. Biophys. Biol.* **1988**, *17*, 451−479.
(7) Frauenfelder, H.; Sligar, S.; Wolynes, P. *Science* **1991**, *254*, 1598−1603.
(8) Henzler-Wildman, K. A.; Lei, M.; Thai, V.; Kerns, S. J.; Karplus, M.; Kern, D. *Nature* **2007**, *450*, 913−916.
(9) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318−321.
(10) Schulz, R.; Krishnan, M.; Daidone, I.; Smith, J. C. *Biophys. J.* **2009**, *96*, 476−484.
(11) Igumenova, T. I.; Frederick, K. K.; Wand, A. J. *Chem. Rev.* **2006**, *106*, 1672−1699.
(12) Faure, P.; Micu, A.; Pérahia, D.; Doucet, J.; Smith, J. C.; Benoit, J. P. *Nat. Struct. Biol.* **1994**, *1*, 124−128.
(13) Lee, A. L.; Wand, A. J. *Nature* **2001**, *411*, 501−504.
(14) Tournier, A. L.; Smith, J. C. *Phys. Rev. Lett.* **2003**, *91*, 208106.
(15) Daniel, R. M.; Finney, J. L.; Smith, J. C. *Faraday Discuss.* **2003**, *122*, 163−169.
(16) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585−590.
(17) Sapienza, P. J.; Lee, A. L. *Curr. Opin. Pharmacol.* **2010**, *10*, 723−730.
(18) Lakowicz, J. R.; Weber, G. *Biochemistry* **1973**, *12*, 4161−4170.
(19) Lakowicz, J. R.; Weber, G. *Biochemistry* **1973**, *12*, 4171−4179.
(20) Palmer, A. G.; Hochstrasser, R. A.; Millar, D. P.; Rance, M.; Wright, P. E. *J. Am. Chem. Soc.* **1993**, *115*, 6333−6345.
(21) Doster, W.; Cusack, S.; Petry, W. *Nature* **1989**, *337*, 754−756.
(22) Becker, T.; Smith, J. C. *Phys. Rev. E* **2003**, *67*, 021904.
(23) Kim, H.; Cho, M. *Chem. Rev.* **2013**, *113*, 5817−5847.
(24) Thomas, G. J., Jr. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 1−27.
(25) Ben-Avraham, D. *Phys. Rev. B* **1993**, *47*, 14559−14560.
(26) Balog, E.; Smith, J. C.; Perahia, D. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5543−5548.
(27) Etchegoin, P. *Phys. Rev. E* **1998**, *58*, 845−848.
(28) Ruschak, A. M.; Kay, L. E. *J. Biomol. NMR* **2010**, *46*, 75−87.
(29) Yang, D. *Protein Pept. Lett.* **2011**, *18*, 380−395.
(30) Kay, L. E. *J. Magn. Reson.* **2005**, *173*, 193−207.
(31) Tugarinov, V.; Kay, L. E. *ChemBioChem.* **2005**, *6*, 1567−1577.
(32) Muhandiram, D. R.; Yamazaki, T.; Sykes, B. D.; Kay, L. E. *J. Am. Chem. Soc.* **1995**, *117*, 11536−11544.
(33) Boehr, D. D.; Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2006**, *106*, 3055−3079.
(34) Kay, L. E. *J. Magn. Reson.* **2011**, *213*, 492−494.
(35) Wand, A. J. *Nat. Struct. Biol.* **2001**, *8*, 926−931.
(36) Ishima, R.; Petkova, A. P.; Louis, J. M.; Torchia, D. A. *J. Am. Chem. Soc.* **2001**, *123*, 6164−6171.
(37) Tugarinov, V.; Ollerenshaw, J. E.; Kay, L. E. *J. Am. Chem. Soc.* **2005**, *127*, 8214−8225.
(38) Brüschweiler, R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 175−183.
(39) Lee, A. L.; Flynn, P. F.; Wand, A. J. *J. Am. Chem. Soc.* **1999**, *121*, 2891−2902.
(40) Lee, A. L.; Sharp, K. A.; Kranz, J. K.; Song, X.-J.; Wand, A. J. *Biochemistry* **2002**, *41*, 13814−13825.
(41) Showalter, S. A.; Johnson, E.; Rance, M.; Brüschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 14146−14147.
(42) Brüschweiler, R.; Wright, P. E. *J. Am. Chem. Soc.* **1994**, *116*, 8426−8427.
(43) Kay, L. E. *Nat. Struct. Biol.* **1998**, *5*, 513−517.
(44) Baldwin, A. J.; Kay, L. E. *Nat. Chem. Biol.* **2009**, *5*, 808−814.
(45) Hansen, D. F.; Neudecker, P.; Vallurupalli, P.; Mulder, F. A. A.; Kay, L. E. *J. Am. Chem. Soc.* **2009**, *132*, 42−43.
(46) Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. *Nat. Struct. Biol.* **2001**, *8*, 932−935.
(47) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224−228.
(48) Wand, A. J.; Englander, S. W. *Curr. Opin. Biotechnol.* **1996**, *7*, 403−408.
(49) Li, D.-W.; Brüschweiler, R. *J. Am. Chem. Soc.* **2009**, *131*, 7226−7227.
(50) Li, D.-W.; Showalter, S. A.; Brüschweiler, R. *J. Phys. Chem. B* **2010**, *114*, 16036−16044.
(51) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(52) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559−4570.

(53) Nicholson, L. K.; Kay, L. E.; Baldisseri, D. M.; Arango, J.; Young, P. E.; Bax, A.; Torchia, D. A. *Biochemistry* **1992**, *31*, 5253−5263.
(54) Chatfield, D. C.; Augsten, A.; D'Cunha, C.; Chatfield, D. C.; Augsten, A.; D'Cunha, C. *J. Biomol. NMR* **2004**, *29*, 377−385.
(55) Lipari, G.; Szabo, A.; Levy, R. M. *Nature* **1982**, *300*, 197−198.
(56) Chatfield, D. C.; Wong, S. E. *J. Phys. Chem. B* **2000**, *104*, 11342−11348.
(57) Chatfield, D. C.; Szabo, A.; Brooks, B. R. *J. Am. Chem. Soc.* **1998**, *120*, 5301−5311.
(58) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631−639.
(59) Best, R. B.; Clarke, J.; Karplus, M. *J. Mol. Biol.* **2005**, *349*, 185−203.
(60) Best, R. B.; Clarke, J.; Karplus, M. *J. Am. Chem. Soc.* **2004**, *126*, 7734−7735.
(61) Case, D. A. *Acc. Chem. Res.* **2001**, *35*, 325−331.
(62) Chatfield, D. C.; Augsten, A.; D'Cunha, C.; Wong, S. E. *J. Comput. Chem.* **2003**, *24*, 1052−1058.
(63) Yang, D.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 369−382.
(64) Akke, M.; Brüschweiler, R.; Palmer, A. G. *J. Am. Chem. Soc.* **1993**, *115*, 9832−9833.
(65) Hu, H.; Hermans, J.; Lee, A. L. *J. Biomol. NMR* **2005**, *32*, 151−162.
(66) Lienin, S.; Bremi, T.; Brutscher, B.; Brüschweiler, R.; Ernst, R. *J. Am. Chem. Soc.* **1998**, *120*, 9870−9879.
(67) Krishnan, M.; Smith, J. C. *J. Phys. Chem. B* **2012**, *116*, 4124−4133.
(68) Hyberts, S.; Goldberg, M.; Havel, T.; Wagner, G. *Protein Sci.* **1992**, *1*, 736−751.
(69) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531−544.
(70) Noble, M.; Musacchio, A.; Saraste, M.; Courtneidge, S.; Wierenga, R. *EMBO J.* **1993**, *12*, 2617−2624.
(71) Xu, Z.; Bernlohr, D.; Banaszak, L. *J. Biol. Chem.* **1993**, *268*, 7874−7884.
(72) Vojtěchovský, J.; Chu, K.; Berendzen, J.; Sweet, R. M.; Schlichting, I. *Biophys. J.* **1999**, *77*, 2153−2174.
(73) Norledge, B.; Hay, R.; Bateman, O.; Slingsby, C.; Driessen, H. *Exp. Eye Res.* **1997**, *65*, 609−630.
(74) Perryman, A. L.; Zhang, Q.; Soutter, H. H.; Rosenfeld, R.; McRee, D. E.; Olson, A. J.; Elder, J. E.; David Stout, C. *Chem. Biol. Drug Des.* **2010**, *75*, 257−268.
(75) Fu, Y.; Kasinath, V.; Moorman, V. R.; Nucci, N. V.; Hilser, V. J.; Wand, A. J. *J. Am. Chem. Soc.* **2012**, *134*, 8543−8550.
(76) Ishima, R.; Petkova, A. P.; Louis, J. M.; Torchia, D. A. *J. Am. Chem. Soc.* **2001**, *123*, 6164−6171.
(77) Best, R. B.; Rutherford, T. J.; Freund, S. M. V.; Clarke, J. *Biochemistry* **2004**, *43*, 1145−1155.
(78) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(79) MacKerell, A. D.; Bashford, D.; Bellott; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(80) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(81) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(82) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169−9183.
(83) Hénin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. *J. Chem. Theory Comput.* **2009**, *6*, 35−47.
(84) Hénin, J.; Chipot, C. *J. Chem. Phys.* **2004**, *121*, 2904.
(85) Chipot, C., Pohorille, A., Eds. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Springer Series in Chemical Physics, Vol. *86*; Springer GmbH: New York, 2007; pp 143−149.
(86) Krishnan, M.; Smith, J. C. *J. Am. Chem. Soc.* **2009**, *131*, 10083−10091.
(87) Hembree, W. I.; Baudry, J. *J. Phys. Chem. B* **2011**, *115*, 8575−8580.
(88) Baudry, J. *J. Am. Chem. Soc.* **2006**, *128*, 11088−11093.
(89) Batchelder, L. S.; Niu, C. H.; Torchia, D. A. *J. Am. Chem. Soc.* **1983**, *105*, 2228−2231.
(90) Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 809−813.
(91) Radom, L.; Hehre, W. J.; Pople, J. A. *J. Am. Chem. Soc.* **1972**, *94*, 2371−2381.
(92) Xue, Y.; Pavlova, M. S.; Ryabov, Y. E.; Reif, B.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2007**, *129*, 6827−6838.
(93) Roh, J.; Novikov, V.; Gregory, R.; Curtis, J.; Chowdhuri, Z.; Sokolov, A. *Phys. Rev. Lett.* **2005**, *95*, 038101.
(94) Batchelder, L. S.; Sullivan, C. E.; Jelinski, L. W.; Torchia, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 386−389.
(95) Agarwal, V.; Xue, Y.; Reif, B.; Skrynnikov, N. R. *J. Am. Chem. Soc.* **2008**, *130*, 16611−16621.
(96) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186−195.
(97) Eyring, H. *Chem. Rev.* **1935**, *17*, 65−77.
(98) Lang, P. T.; Ng, H.-L.; Fraser, J. S.; Corn, J. E.; Echols, N.; Sales, M.; Holton, J. M.; Alber, T. *Protein Sci.* **2010**, *19*, 1420−1431.
(99) Fraser, J. S.; van den Bedem, H.; Samelson, A. J.; Lang, P. T.; Holton, J. M.; Echols, N.; Alber, T. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 16247−16252.
(100) Butterfoss, G. L.; Hermans, J. *Protein Sci.* **2003**, *12*, 2719−2731.
(101) Tugarinov, V.; Kay, L. E. *J. Am. Chem. Soc.* **2004**, *126*, 9827−9836.
(102) Tugarinov, V.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 13868−13878.
(103) Fu, Y.; Kasinath, V.; Moorman, V. R.; Nucci, N. V.; Hilser, V. J.; Wand, A. J. *J. Am. Chem. Soc.* **2012**, *134*, 8543−8550.
(104) Mittermaier, A.; Kay, L.; Forman-Kay, J. *J. Biomol. NMR* **1999**, *13*, 181−185.
(105) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. *Nature* **2007**, *448*, 325−329.
(106) Constantine, K. L.; Friedrichs, M. S.; Wittekind, M.; Jamil, H.; Chu, C.-H.; Parker, R. A.; Goldfarb, V.; Mueller, L.; Farmer, B. T. *Biochemistry* **1998**, *37*, 7965−7980.
(107) Mittermaier, A.; Davidson, A. R.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 9004−9005.
(108) Clarkson, M. W.; Lee, A. L. *Biochem. J.* **2004**, *43*, 12448−12458.
(109) Kasinath, V.; Sharp, K. A.; Wand, A. J. *J. Am. Chem. Soc.* **2013**, *135*, 15092−15100.
(110) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(111) Li, D.-W.; Brüschweiler, R. *Angew. Chem.* **2010**, *122*, 6930−6932.
(112) Li, D.-W.; Brüschweiler, R. *J. Chem. Theory Comput.* **2011**, *7*, 1773−1782.
(113) Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. *Nat. Chem. Biol.* **2010**, *6*, 352−358.
(114) Li, Z.; Raychaudhuri, S.; Wand, A. J. *Protein Sci.* **1996**, *5*, 2647−2650.
(115) Glass, D. C.; Krishnan, M.; Smith, J. C.; Baudry, J. *J. Phys. Chem. B* **2013**, *117*, 3127−3134.
(116) Ishima, R.; Torchia, D. A. *Nat. Struct. Mol. Biol.* **2000**, *7*, 740−743.
(117) Palmer, A. G., III. *Chem. Rev.* **2004**, *104*, 3623−3640.
(118) Kay, L. E.; Muhandiram, D. R.; Wolf, G.; Shoelson, S. E.; Forman-Kay, J. D. *Nat. Struct. Biol.* **1998**, *5*, 156−163.
(119) Wand, A. J. *Curr. Opin. Struct. Biol.* **2013**, *23*, 75−81.
(120) Vallurupalli, P.; Bouvignies, G.; Kay, L. E. *J. Am. Chem. Soc.* **2012**, *134*, 8148−8161.
(121) Sekhar, A.; Kay, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 12867−12874.