JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

On the Contribution of Linear Correlations to Quasi-harmonic Conformational Entropy in Proteins

Anton A. Polyansky, Antonija Kuzmanic, Mario Hlevnjak, and Bojan Zagrovic*

Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Campus Vienna Biocenter 5, Vienna, AT-1030, Austria

Supporting Information

ABSTRACT: We study the contribution of linear, pairwise atom-positional correlations (covariances) to absolute and relative conformational entropy as calculated by quasi-harmonic analysis of molecular dynamics (MD) trajectories (S_{QH} and ΔS_{QH}). By analyzing a total of 25 μs of MD simulations of ubiquitin and six of its binding partners in bound and unbound states, and 2.4 μs of simulations of eight different proteins in phosphorylated and unphosphorylated states, we show that ΔS_{QH} represents a remarkably constant fraction of a quasi-harmonic entropy change obtained if one ignores the contribution of covariance terms and uses mass-weighted atom-positional variances only (ΔS_{VAR}). In other words, the relative contribution of linear correlations to conformational entropy change for different proteins and in different biomolecular processes appears to be largely constant. Based on this, we establish an empirical relationship between relative quasi-harmonic conformational entropy and changes in crystallographic B-factors induced by different processes, and we use it to estimate conformational-entropic contribution to the free energy of binding for a large set of protein complexes based on their X-ray structures. Our results suggest a simple way for relating other types of dynamical observables with conformational entropy in the absence of information on correlated motions, such as in the case of NMR order parameters.

1. INTRODUCTION

Protein dynamics is one of the most important determinants of biological function *in vivo*. Changes in atomic motions of a protein as a consequence of folding,[1] post-translational modifications,[2] allosteric regulation,[3,4] or interactions with binding partners[5] typically have a strong impact on the overall thermodynamics of the molecule, which can be observed as a change in its conformational entropy. A number of different experimental and computational techniques provide a connection between protein dynamics and conformational entropy (see ref 6 for a recent review). Among the most commonly used such experimental methods are NMR relaxation measurements, which are able to capture ps-ns protein dynamics and estimate the absolute conformational entropy from the generalized order parameters (S^2), albeit without accounting for correlated motions directly.[7−10] Since order parameters report on the behavior of isolated bond vectors or methyl symmetry axes, they provide little information on the potentially significant, entropy-reducing correlated motions and can only give an upper bound to the actual conformational entropy. These approximations notwithstanding, using such approaches, Wand and co-workers have recently suggested that conformational entropy may account for even up to one-half of the total entropy change as measured calorimetrically in the case of calmodulin binding to different binding partners.[10] However, one should note that the total entropy change cannot be estimated without careful consideration of solvent entropy,[11] which has significant impact on the free energy change and is comparable in magnitude to conformational entropy.[12]

When it comes to computational approaches, atomic motions are typically connected with conformational entropy through the principal-component analysis of the variance−covariance matrix of protein internal (bonds-angles-torsions) or Cartesian atom-positional coordinates.[13−17] Here, eigenvalues of the mass-weighted variance−covariance matrix define angular frequencies of harmonic oscillators, which are subsequently used in the calculation of quasi-harmonic entropy (S_{QH}) via the equation for the entropy of a quantum-mechanical harmonic oscillator.[16] Although conventional calculations of S_{QH} do not correct for anharmonicity and supralinear correlations between eigenvectors of the variance−covariance matrix, and therefore only give an upper-bound estimate of conformational entropy,[15] they still represent a basis for developing more precise and computationally more efficient methods for entropy calculation.[6] Needless to say, accurate estimation of conformational entropy in realistic biomolecular systems is still a difficult open challenge for the existing experimental and computational methods alike. Moreover, development of novel, robust, and generally applicable approaches, which give an approximate estimate of the contribution of conformational entropy to the overall free energy change, is an important goal in and of itself. Despite their sometimes limited accuracy[18,19] quasi-harmonic approaches provide a useful starting point for such attempts.

Here, we study the contribution of linear, pairwise atom-positional correlations (covariances) to the Cartesian-coordinate quasi-harmonic conformational entropy. The approach we take is to evaluate the quasi-harmonic entropy of different simulated systems by first ignoring the contribution of covariance elements in the variance−covariance matrix, and

Special Issue: Wilfred F. van Gunsteren Festschrift

Received: February 1, 2012
Published: June 18, 2012

then compare the thus obtained entropy with the quasi-harmonic entropy obtained if both variance and covariance elements are taken into account. Using atomistic molecular dynamics (MD) methodology to generate a total of 15 μs of simulated time for the complex between ubiquitin (UBQ) and ubiquitin binding motif (UBM2) of human polymerase iota in bound and unbound states, 10 μs of simulated time for 5 other ubiquitin binding partners from different structural classes in bound and unbound states (UBQ binders set) and 2.4 μs of simulated time for 8 proteins undergoing phosphorylation (Phospho set),[20] we demonstrate a simple linear relationship between quasi-harmonic conformational entropy change (ΔS_{QH}) and change in quasi-harmonic entropy computed by using mass-weighted variances of atomic positions only (ΔS_{VAR}).

Based on this, we establish a connection between the change in quasi-harmonic conformational entropy in different biomolecular processes and the respective changes in crystallographic Debye–Waller factors (B-factors), which are directly related to the variance terms in the atom-positional variance–covariance matrix. We have shown previously that analysis of B-factors allows for a quantification of global conformational heterogeneity of macromolecules in crystals,[21,22] which opens up the possibility to connect this property with conformational entropy. For example, it has been shown for ferredoxin interacting with ferredoxin-NADP$^+$ reductase that an increase in B-factors of protein residues upon binding might be related to their favorable contribution to relative conformational entropy obtained by NMR.[23] Here, based on the above approach, we derive a relationship linking changes in B-factors and relative conformational entropy, and use it to estimate conformational entropy contributions in the formation of binary protein complexes for a model set of X-ray crystallographic structures.

2. METHODS

2.1. Quasi-harmonic Entropy Derived from Mass-Weighted Variance–Covariance Matrix of Atomic Positions with (S_{QH}) and without (S_{VAR}) Covariance Terms. MD-based evolution of atomic positions of N protein atoms can be used to build a mass-weighted variance–covariance matrix C with $3N \times 3N$ elements after roto-translational superposition of individual simulated structures:

$$C_{ij} = \langle M_i^{1/2}(r_i - \langle r_i \rangle) M_j^{1/2}(r_j - \langle r_j \rangle) \rangle \quad (1)$$

where $1 \le i \le 3N$, $1 \le j \le 3N$ and M is a mass matrix. Here, r_{3k-2}, r_{3k-1}, r_{3k} are Cartesian coordinates of protein atom k with mass $M_{3k-2} = M_{3k-1} = M_{3k}$ where $1 \le k \le N$, and the average is over all of the structures in a trajectory. This variance–covariance matrix is then employed in the calculation of quasi-harmonic conformational entropy using the equation for the entropy of a quantum-mechanical harmonic oscillator (see ref 6 for more details):

$$S_{QH} = k_B \sum_i \left(\frac{\hbar\omega_i/k_BT}{e^{\hbar\omega_i/k_BT} - 1} - \ln(1 - e^{-\hbar\omega_i/k_BT}) \right) \quad (2)$$

Here, $\omega_i = (k_BT/\lambda_i)^{1/2}$ are the angular frequencies derived from eigenvalues λ_i obtained by diagonalizing the mass-weighted variance–covariance matrix, and T is the temperature of the simulations, while k_B and $\hbar$ are Boltzmann's and Planck's constants, respectively.

The diagonal elements of the variance–covariance matrix C_{ii} report on atom-positional variances of individual atoms in the molecule. If one assumes that all of the atoms in the system are moving independently from each other and their covariances are all equal to zero, one obtains a new diagonal variance–covariance matrix whose mass-weighted variances are equal to its eigenvalues:

$$\lambda_i^{var} = M_i(r_i - \langle r_i \rangle)^2 = M_i \mathrm{RMSF}_i^2 \quad (3)$$

where RMSF_i is root-mean-square fluctuation along r_i coordinate of a given atom, and M is atom's mass.

In analogy with the quasi-harmonic entropy formalism,[16] we can then use these eigenvalues to calculate angular frequencies:

$$\omega_i^{var} = \sqrt{k_BT/\lambda_i^{var}} = \sqrt{M_i^{-1}k_BT}\,\mathrm{RMSF}_i^{-1} \quad (4)$$

and insert them into eq 2 to obtain:

$$S_{VAR} = k_B \sum_i \left(\frac{\hbar/\sqrt{M_i k_B T}\,\mathrm{RMSF}_i}{e^{\hbar/\sqrt{M_i k_B T}\,\mathrm{RMSF}_i} - 1} - \ln(1 - e^{-\hbar/\sqrt{M_i k_B T}\,\mathrm{RMSF}_i}) \right) \quad (5)$$

2.2. Calculation of S_{VAR} from Crystallographic B-Factors. The connection between RMSFs and S_{VAR} allows one to derive a formalism for estimating S_{VAR} from crystallographic B-factors. The connection between RMSF and B-factor for atom k is given as follows:[24]

$$\mathrm{RMSF}_k^2 = \frac{3B_k}{8\pi^2} \quad (6)$$

Positional variance of atom k is equal to the sum of positional variances for each coordinate:

$$\mathrm{RMSF}_k^2 = \mathrm{RMSF}_{xk}^2 + \mathrm{RMSF}_{yk}^2 + \mathrm{RMSF}_{zk}^2 \quad (7)$$

By assuming that the variances for each coordinate are the same, one can combine eq 6 and 7 to get:

$$\mathrm{RMSF}_{xk}^2 = \frac{B_k}{8\pi^2} \quad (8)$$

Equation 8 gives variances, which are then used in eq 5 to calculate S_{VAR} after mass-weighting.

2.3. Protein Data Sets. Conformational entropy upon binding was studied for complexes between ubiquitin (UBQ) and 6 other proteins from different structural classes (Figure 3). Because of the varying levels of sampling employed, we divide this set into two groups: (1) simulations of the complex between UBQ and the ubiquitin-binding motif (UBM2) of human polymerase iota[25] where we have simulated the two molecules for a total of 5 independent 1 μs-long trajectories in the bound and unbound state in each case, and (2) 5 other proteins from different structural classes (UBQ binders set) where we have simulated the entire complex or UBQ's binding partner alone using 1 μs-long trajectories in each case. For UBQ/UBM2 simulations, we have used structures of unbound ubiquitin (PDB code: 1UBQ),[26] unbound UBM2 (2L0G),[25] and their complex (2KTF).[25] The set named "UBQ binders" included 4 additional complexes between ubiquitin and other proteins where X-ray structures were available for both the bound and the unbound form: TSG101 domain (with PDB code for bound/unbound 1S1Q/1KPP),[27,28] UEV domain of

Table 1. Summary of MD Simulations for the Ubq/Ubm2 Set[a]

simulation[b]	PDB code[c]	no. res/atoms[d]	S_{QH}[e] (kJ mol^{-1} K^{-1})	C_F[f]	$\langle rmsd_{INIT}\rangle$[g] (Å)	$\langle rmsd_B\rangle$[h] (Å)
UBQ_1	1UBQ	76/760	23.3	0.44 ± 0.01	2.9 ± 0.9	n/a
UBQ_2	1UBQ	76/760	23.9	0.44 ± 0.01	3.1 ± 1.4	n/a
UBQ_3	1UBQ	76/760	22.9	0.44 ± 0.02	3.1 ± 1.1	n/a
UBQ_4	1UBQ	76/760	22.6	0.44 ± 0.01	2.2 ± 0.6	n/a
UBQ_5	1UBQ	76/760	24.5	0.44 ± 0.02	4.1 ± 1.3	n/a
UBQ_B1	2KTF	76/760	21.0	0.46 ± 0.01	1.7 ± 0.5	3.4 ± 0.6/3.9 ± 0.7/3.1 ± 0.7/3.4 ± 0.7/3.4 ± 0.6[a]
UBQ_B2	2KTF	76/760	22.0	0.45 ± 0.01	2.4 ± 0.9	5.6 ± 0.4/6.0 ± 0.4/5.5 ± 0.5/5.4 ± 0.5/5.7 ± 0.4[a]
UBQ_B3	2KTF	76/760	21.0	0.45 ± 0.01	2.1 ± 0.7	4.2 ± 0.5/3.2 ± 0.7/4.0 ± 0.5/4.0 ± 0.6/4.2 ± 0.5[a]
UBQ_B4	2KTF	76/760	22.1	0.45 ± 0.01	2.3 ± 0.7	2.9 ± 0.4/3.8 ± 0.5/2.7 ± 0.6/2.9 ± 0.6/2.9 ± 0.4[a]
UBQ_B5	2KTF	76/760	20.8	0.46 ± 0.01	1.7 ± 0.5	4.6 ± 0.7/4.6 ± 0.7/4.5 ± 0.7/4.5 ± 0.7/4.7 ± 0.7[a]
UBM2_1	2L0G	32/334	13.8	0.60 ± 0.03	6.1 ± 2.5	n/a
UBM2_2	2L0G	32/334	12.7	0.59 ± 0.03	4.0 ± 2.5	n/a
UBM2_3	2L0G	32/334	13.6	0.61 ± 0.03	5.8 ± 2.2	n/a
UBM2_4	2L0G	32/334	13.5	0.60 ± 0.03	5.6 ± 2.3	n/a
UBM2_5	2L0G	32/334	13.1	0.61 ± 0.02	6.5 ± 2.5	n/a
UBM2_B1	2KTF	32/334	10.9	0.66 ± 0.05	3.3 ± 1.2	8.0 ± 0.7/8.7 ± 0.7/8.7 ± 1.1/9.4 ± 0.7/7.2 ± 0.8[a]
UBM2_B2	2KTF	32/334	10.7	0.61 ± 0.03	3.7 ± 1.7	7.6 ± 0.4/8.5 ± 0.8/7.7 ± 0.8/8.1 ± 0.8/8.6 ± 0.8[a]
UBM2_B3	2KTF	32/334	11.5	0.64 ± 0.04	3.7 ± 2.1	8.3 ± 0.6/7.4 ± 0.8/7.1 ± 0.9/8.6 ± 0.9/7.4 ± 0.8[a]
UBM2_B4	2KTF	32/334	11.3	0.65 ± 0.03	3.8 ± 1.4	6.9 ± 0.9/7.8 ± 1.0/7.9 ± 1.2/7.7 ± 1.0/6.1 ± 1.1[a]
UBM2_B5	2KTF	32/334	11.2	0.62 ± 0.03	3.5 ± 1.4	8.4 ± 0.6/8.1 ± 1.3/8.4 ± 1.0/8.4 ± 0.8/7.9 ± 0.8[a]

[a]Note: 5 independent MD trajectories of the two unbound proteins were used for rmsd calculations. [b]Simulation label: UBQ_X, unbound ubiquitin simulations; UBQ_BX, bound ubiquitin simulations; UBM2_X, unbound UBM2 simulations; UBM2_BX, bound UBM2 simulations. [c]PDB code, PDB ID of initial structure. [d]No. res/atoms, number of protein residues/atoms. [e]S_{QH}, quasi-harmonic entropy. [f]C_F, compactness factor. [g]$\langle rmsd_{INIT}\rangle$, average backbone rmsd from the initial structure. [h]$\langle rmsd_B\rangle$, average backbone rmsd between bound and unbound structures.

Table 2. Summary of MD Simulations for the UBQ Binder Set

simulation[a]	PDB code[b]	no. res/ atoms[c]	S_{QH}[d] (kJ mol^{-1} K^{-1})	C_F[e]	$\langle rmsd_{INIT}\rangle$[f](Å)	$\langle rmsd_B\rangle$[g] (Å)
TSG101	1KPP	144/1480	39.6	0.41 ± 0.01	5.4 ± 0.8	n/a
TSG101_B	1S1Q	220/2240	37.1	0.42 ± 0.01	2.9 ± 0.4	4.3 ± 0.7
UBQ_B(TSG101)	1S1Q	76/760	20.1	0.44 ± 0.01	2.6 ± 0.4	3.2 ± 0.7/5.6 ± 0.5/4.0 ± 0.6/2.8 ± 0.6/4.6 ± 0.8[h]
VPS23	3R3Q	151/1493	41.2	0.41 ± 0.01	7.5 ± 1.1	n/a
VPS23_B	1UZX	227/2253	39.0	0.41 ± 0.01	4.2 ± 0.5	6.6 ± 1.1
UBQ_B (VPS23)	1UZX	76/760	20.4	0.4 ± 0.01	2.9 ± 0.6	3.4 ± 0.8/5.5 ± 0.5/3.8 ± 0.5/3.0 ± 0.7/4.4 ± 0.7[h]
UCH-L3	1UCH	229/2305	60.0	0.37 ± 0.01	5.2 ± 0.6	n/a
UCH-L3_B	1XD3	305/3065	58.1	0.40 ± 0.01	5.4 ± 1.2	9.4 ± 1.6
UBQ_B(UCH-L3)	1XD3	76/760	18.9	0.47 ± 0.01	1.7 ± 0.2	3.6 ± 0.5/5.9 ± 0.4/4.2 ± 0.5/3.2 ± 0.3/4.7 ± 0.6[h]
GGA3	1YD8[i]	92/949	26.1	0.49 ± 0.02	6.1 ± 0.5	n/a
GGA3_B	1YD8	168/1709	26.1	0.50 ± 0.01	4.5 ± 0.3	10.0 ± 0.8
UBQ_B(GGA3)	1YD8	76/760	19.2	0.46 ± 0.01	2.7 ± 0.3	3.5 ± 0.5/5.7 ± 0.4/4.5 ± 0.5/3.1 ± 0.3/4.8 ± 0.7[h]
CBL-B	2OOA	45/457	12.6	0.50 ± 0.02	4.5 ± 0.4	n/a
CBL-B_B	2OOB	121/1217	12.2	0.50 ± 0.02	4.7 ± 0.4	2.0 ± 0.6
UBQ_B(CBL-B)	2OOB	76/760	20.6	0.45 ± 0.01	3.2 ± 0.3	2.8 ± 0.7/5.4 ± 0.5/4.1 ± 0.5/2.4 ± 0.5/4.6 ± 0.7[h]

[a]Simulation label: *_B notation refers to bound proteins, UBQ_B() notation refers to ubiquitin in the bound form with the protein indicated in the parentheses. [b]PDB code, PDB ID of initial structure. [c]No. res/atoms, number of protein residues/atoms. [d]S_{QH}, quasi-harmonic entropy. [e]C_F, compactness factor. [f]$\langle rmsd_{INIT}\rangle$, average backbone rmsd from the initial structure. [g]$\langle rmsd_B\rangle$, average backbone rmsd between bound and unbound structures. [h]Five independent MD trajectories of unbound UBQ were used for rmsd calculations. [i]For GGA3 the initial bound structure of the protein was detached from ubiquitin and pre-equilibrated over 0.5 μs of MD before performing an unbound 1 μs MD production run.

Vps23 (1UZX/3R3Q),[29,30] ubiquitin hydrolase UCH-L3 (1XD3/1UCH),[31,32] and UBA domain of Cbl-b (2OOB/2OOA).[33] In addition, the UBQ binders set included one additional complex where only the bound form's X-ray structure was available: ADP-ribosylation factor binding protein GGA3 (1YD8).[34] For GGA3, lacking an experimentally determined unbound structure, the initial bound structure of the protein (1YD8) was pre-equilibrated during 0.5 μs of MD and further used as an unbound structure for the subsequent 1 μs production MD run. To ensure that each structure had the same number of atoms in the bound and unbound state, the coordinates of the missing atoms in some structures were added using those in the cognate pair as a template for structural alignment, followed by energy minimization.

The effect of phosphorylation on conformational entropy was studied for a model set of 8 proteins in native and phosphorylated forms (Phospho set,[20] Figure 4) obtained as follows. The initial set of protein structures was extracted from the RCSB Protein Data Bank using an advanced search procedure (with a query: *"Modified Res=SEP/TPO/PTR + homology <90%"*) and further filtered according to the following requirements: (1) phosphorylated residues are explicitly

present in the 3D structure; (2) both forms have the same length and sequence; (3) the molecule represents a complete protein or its biologically relevant fragment (e.g., an isolated catalytic domain), (4) the protein does not contain any missing parts (in case of X-ray structures), (5) the protein is not a membrane protein, and (6) the protein is not in complex with any binding partners. For 8 native and phosphorylated proteins, a 150 ns-long MD run was performed in each case.

The set of protein complexes and isolated binding partners that was used for the estimation of relative conformational entropy of binding using crystallographic B-factors (Prot-Binders set) was extracted from the PDB (November 2011 release) by using the following criteria: (1) X-ray structures contain only proteins (no nucleic acids); (2) their resolution is better than 2.5 Å; and (3) they were solved close to or at 100 K ($\langle T \rangle$ = 100.4 K, σ = 3.9 K). All of the complexes in our set are exclusively dimeric (number of protein entities fixed at 2, number of chains in the biological assembly fixed at 2), while unbound structures are exclusively monomeric (number of protein entities fixed at 1, number of chains in the biological assembly fixed at 1). Both structures were allowed to have as many chains per protein as were present in the PDB file (number of chains in the asymmetric unit set at 1 or more for unbound, and 2 or more for bound proteins). Connecting particular proteins with their structures was performed by using an ID-mapping tool at UniProtKB (November 2011 release). In addition, out of several possible unbound structures for a given protein, only those with a maximal sequence overlap with the bound structures were manually selected (the overlap was >90% on average). Finally, the filtering resulted in 104 different unbound proteins and 60 binary complexes with some proteins present in multiple complexes. The conformational entropy based on these structures was calculated using eq 5. Note that in cases where the bound and unbound forms of a given protein differ in the number of atoms, we rescale the entropy of the structure with the lower number of atoms by multiplying it with the ratio of the numbers of atoms in the two structures (N_{higher}/N_{lower}, please see the Supporting Information for details).

2.4. MD Simulation Details. MD simulations were performed using Gromacs 4.0.7 package[35] and Gromos 96 force fields. Ubiquitin, UBM2, and UBQ binders were modeled using Gromos 45A3 force field,[36] while proteins from the Phospho set were modeled using Gromos 43A1P force field, which includes parameters for phosphorylated residues and was previously successfully applied in MD simulations of phosphorylated peptides.[37] The SPC model was used to model water.[38] For all systems (Tables 1 and 2), a standard protocol was used: proteins were placed in water boxes, together with the necessary amount of sodium or chloride counterions to reach neutrality, and subjected to energy minimization, followed by heating to 300 K for 100 ps and subsequently unconstrained MD production simulations. All MD simulations were carried out with a time step of 2 fs using 3D periodic boundary conditions, in the isothermal−isobaric (NPT) ensemble with an isotropic pressure of 1 bar and a constant temperature of 300 K, and system coordinates were output every 1 ps. The pressure and the temperature were controlled using the Berendsen thermostat and barostat[39] with 1.0 and 0.1 ps relaxation parameters, respectively, and a compressibility of 4.5×10^{-5} bar^{-1} for the barostat. Bond lengths were constrained using LINCS.[40] Nonbonded interactions were treated using two different schemes due to different versions of Gromos 96 force field for UBQ/UBM2, UBQ binders set, and the Phospho set. In simulations of ubiquitin and its binders, van der Waals interactions were treated using a cutoff of 14 Å. Electrostatic interactions were evaluated using the reaction-field method,[41] with a direct sum cutoff of 14 Å and relative permittivity of 61. In simulations of proteins from the Phospho set, van der Waals and electrostatic interactions were truncated using a twin-range 10/12 Å spherical cutoff to match the setup used with this force field previously.

2.5. Conformational Entropy Calculations. All mass-weighted covariance matrices were constructed from the MD data using *g_covar* utility from the Gromacs[35] package. For this purpose, proteins were taken in their Gromos united-atom representation with explicit heavy atoms and polar hydrogens. Further calculations were performed in MATLAB (R2009a) using utilities specially written for this study. Eigenvalues of the covariance matrix were computed using MATLAB singular value decomposition procedures (*svd* and *pcaconv*[42]). To test convergence of S_{QH} for different separations between MD snapshots (0.05, 0.5, 1, 5, 10, and 20 ps), a 50-ns segment from the first part of a 1 μs ubiquitin trajectory was used for the calculation.

2.6. Estimation of Protein Compactness. Protein compactness factor (C_F) was calculated by using the ratio of solvent accessible surface area (SASA) of a given protein in the folded and the fully extended conformations ($SASA^{fold}/SASA^{ext}$). SASA calculations from MD snapshots separated by 100 ps were performed using PLATINUM software.[43]

2.7. Temperature-Dependent Scaling of B-factors. Crystallographic B-factors are known to exhibit temperature dependence. Considering that all the selected bound and unbound protein structures were solved at 100 K, in order to estimate conformational entropies at 300 K, we first determined a scaling factor to estimate B-factors at 300 K from those at 100 K. For this, we used a well-characterized set of monomeric eukaryotic lysozyme X-ray structures (resolution better than 2.5 Å and R_{free} < 0.25) solved at 100 and 300 K. Average B-factors were calculated for each of the subsets of lysozymes solved at these two temperatures (Figure S4, Supporting Information), and the ratio of the averages at 300 and 100 K was taken as the scaling factor (f = 1.23). However, we should emphasize that this scaling is empirical and can only be used as a rough approximation.

3. RESULTS AND DISCUSSION

3.1. Conformational Entropy Convergence. Conformational entropy obtained from MD simulations depends strongly on the level of sampling.[6,17] There are several different, mutually inter-related aspects to this fact. First, the correct determination of 3*N*-6 principal components (eigenvectors) requires that at least 3*N*+1 MD snapshots be used to build the covariance matrix. However, in real MD simulations, the main concern is the adequate coverage of the phase space, that is, the length of the simulation and the frequency at which it is sampled. To choose the optimal sampling frequency for our study, we have calculated different values of S_{QH} for a 50 ns trajectory of unbound ubiquitin sampled either every 50 fs, 500 fs, 1 ps, 5 ps, 10 ps, or 20 ps. As can be seen in Figure 1, choosing a separation between frames that is larger than 1 ps leads to an underestimation of S_{QH}. Consequently, we performed all further calculations of S_{QH} or S_{VAR} with a 1 ps separation between snapshots.

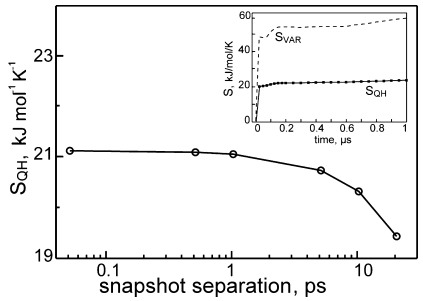

Figure 1. Quasi-harmonic entropy (S_{QH}) convergence. For a fixed trajectory length, S_{QH} depends on the separation between individual MD snapshots. The data is shown for a 50 ns segment from the equilibrated part of a ubiquitin 1 μs trajectory. Inset: Cumulative curves for S_{QH} (solid lines) and quasi-harmonic-like entropy derived from mass-weighted variances of atomic positions (S_{VAR}, dashed lines). Results are obtained using segments from 1 μs ubiquitin trajectories of the indicated length.

Second, the accuracy of S_{QH} depends on whether one adequately samples the conformational heterogeneity of a given system. In simulations of unbound ubiquitin, all protein conformations deviate from the initial structure by less than 6 Å when it comes to backbone atom-positional root-mean-square deviation (rmsd, Table 1), and most trajectories reach a 4 Å mark already within 100 ns (Figure S1, Supporting Information). Note that this relatively high rmsd for a molecule as stable as ubiquitin is due to the inclusion in the calculation of the highly flexible C-terminal tail (residues 71−76). In 4 out of 5 MD runs, protein conformations stay in the 4 Å vicinity of the initial structure when it comes to rmsd for the rest of the 1 μs trajectories (Figure S1, Supporting Information). Similarly, cumulative curves for S_{QH} and its diagonal part S_{VAR} (see the Methods section for definition) for a representative ubiquitin trajectory reach a plateau in about 150 ns (Figure 1 inset). For example, S_{QH} calculated over the first 150 ns of this trajectory reaches a value of 22 kJ mol^{-1} K^{-1}, while over the remaining 850 ns of the trajectory this value increases by 7% only. While this is in no way a measure of entropy convergence in the absolute sense, it is informative about its convergence on the scale of 1 μs studied here. Baron et al. observed a similar evolution of S_{QH} in 1 μs simulations of a small α-helical peptide.[17] Analogously to this, S_{VAR} reaches the value of 53.8 kJ mol^{-1} K^{-1} after 150 ns, which increases by 9% only at the end of the 1 μs simulation. Motivated by these observations as well as by computational limitations, we restricted the length of the simulations of proteins from the Phospho set to 150 ns. On the other hand, to exclude the contribution of the initial structural relaxation (Figure S1−S3, Supporting Information) to conformational entropy, all subsequent analysis was performed on the last 80% of each MD trajectory.

3.2. Linear Dependence of Quasi-harmonic Conformational Entropy on Protein Size. The S_{QH} calculations based on MD trajectories of UBQ, UBM2, UBQ binders, and 8 proteins from the Phospho set allow us to investigate size-dependence of quasi-harmonic conformational entropy for proteins from different structural classes. As seen in Figure 2A, we observe a strong positive linear correlation (R^2 = 0.98) between the number of protein atoms and the calculated value of S_{QH} with an average contribution per atom of approximately 26.6 J mol^{-1} K^{-1} (Figure 2A). Given that our set includes proteins ranging from 32 to 479 residues, having different secondary structure profiles and tertiary folds and exhibiting

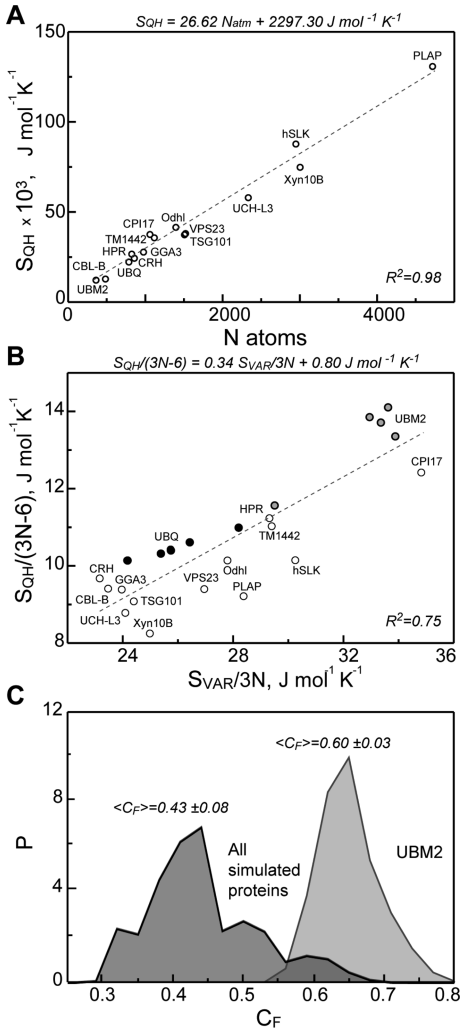

Figure 2. Absolute conformational entropy. (A) Dependence of quasi-harmonic entropy (S_{QH}) on the number of atoms in a protein. The points used for ubiquitin and UBM2 simulations are those whose average rmsd from the starting structures corresponds to the median over all five trajectories of either system. (B) Relation between S_{QH}/(3N−6) and S_{VAR}/3N. We show the results for five 1 μs MD trajectories of unbound ubiquitin (black circles) and five 1 μs trajectories of UBM2 (gray circles), five 1 μs MD trajectories of unbound ubiquitin binders (open circles), together with 8 trajectories of native proteins from the Phospho set (open circles). (C) Normalized distribution of compactness factor (ratio between SASA of folded and extended protein structures, C_F) obtained from the simulations of all the studied proteins in the native states (dark gray) and unbound UBM2 (light gray).

different levels of structuredness, the strength of the observed correlation is remarkable. Of course, S_{QH} is a function of all of the degrees of freedom in a system and is expected to exhibit size dependence, but this does not necessarily imply linearity. One way to understand its origin is as follows. In the classical limit where $\hbar\omega \ll k_BT$, the quasi-harmonic entropy equation can be linearized to

$$S_{QH} \approx k_B \sum_i (1 - \ln(\hbar\omega_i/k_BT)) = 3k_BN - 6k_B - \sum_i \ln(\hbar\omega_i/k_BT) \quad (9)$$

dx.doi.org/10.1021/ct300082q | J. Chem. Theory Comput. 2012, 8, 3820−3829

Table 3. Summary of MD Simulations for the Phospho Set[a]

simulation	PDB code	no. res/atoms	S_{QH} (kJ mol^{-1} K^{-1})	C_F	$\langle rmsd_{INIT} \rangle$ (Å)	$\langle rmsd_B \rangle$ (Å)
HPR	1PFT	87/797	26.1	0.42 ± 0.02	5.79 ± 1.86	n/a
HPR$^{(phospho)}$	1FU0	87/800	27.7	0.45 ± 0.03	6.66 ± 1.89	9.3 ± 1.3
Xyn10B	1GKL	283/2975	74.3	0.31 ± 0.01	4.87 ± 1.04	n/a
Xyn10B$^{(phospho)}$	1GKK	283/2980	76.0	0.32 ± 0.01	4.96 ± 0.75	6.0 ± 0.5
TM1442	1SBO	110/1089	35.2	0.43 ± 0.02	6.17 ± 1.17	n/a
TM1442$^{(phospho)}$	1T6R	110/1092	34.6	0.42 ± 0.01	4.68 ± 0.89	6.6 ± 0.6
CRH	1MU4	86/831	23.8	0.42 ± 0.02	9.95 ± 1.48	n/a
CRH$^{(phospho)}$	2AK7	86/834	27.3	0.43 ± 0.02	8.23 ± 1.01	7.1 ± 0.5
PLAP	1ZED	479/4689	130.2	0.33 ± 0.01	6.58 ± 1.21	n/a
PLAP$^{(phospho)}$	2GLQ	479/4681	128.2	0.32 ± 0.01	5.83 ± 0.77	8.6 ± 0.5
hSLK	2J51	288/2921	87.2	0.37 ± 0.01	9.06 ± 1.05	n/a
hSLK$^{(phospho)}$	2JFL	288/2927	86.4	0.39 ± 0.01	8.93 ± 1.54	10.3 ± 0.6
OdhI	2KB4	143/1367	41.0	0.38 ± 0.02	16.34 ± 1.57	n/a
OdhI$^{(phospho)}$	2KB3	143/1370	38.7	0.38 ± 0.02	5.46 ± 0.35	14.3 ± 0.6
CPI17	1J2M	99/1033	37.0	0.53 ± 0.04	12.32 ± 2.66	n/a
CPI17$^{(phospho)}$	2RTL	99/1036	32.8	0.45 ± 0.01	8.39 ± 1.24	14.0 ± 1.7

[a]Simulation - simulation label; PDB code - PDB ID of initial structure; No. res/atoms - number of protein residues / atoms; S_{QH} - quasi-harmonic entropy; CF - compactness factor; $\langle rmsd_{INIT} \rangle$ - average backbone rmsd from the initial structure; $\langle rmsd_p \rangle$ - average backbone rmsd between phosphorylated and native structures.

Interestingly, the prefactor of $3k_B$ in this equation equals 24.9 J mol^{-1} K^{-1}, which is close to the slope of the fit in Figure 2A of 26.6 J mol^{-1} K^{-1}. The difference between the two arguably comes from the last term in the above equation, which is also size dependent, but overall it is likely that the bulk of the linear relationship between protein size and absolute value of quasi-harmonic entropy comes from the first term and its linear dependence on N in the classical limit.

Importantly, such an effect is observed only when comparing different proteins over a large range of absolute entropies, while for each individual molecule the quasi-harmonic conformational entropy strongly depends on its internal dynamics. Finally, here, we only report on an upper bound of conformational entropy as provided by the standard quasi-harmonic analysis without including corrections for anharmonicity and supralinear correlations between eigenmodes.[6] The latter corrections would likely introduce highly nonlinear dependence on the number of degrees of freedom into the quasi-harmonic formalism, which might consequently affect the linear trend seen here.

3.3. Contribution of Variances of Atomic Positions to Conformational Entropy. We have calculated the diagonal part of the absolute quasi-harmonic conformational entropy (S_{VAR}) using mass-weighted variances of atomic positions (diagonal of the covariance matrix) from each MD trajectory using eq 5. In Figure 2B, we plot these S_{VAR} values obtained for 15 independent 1 μs runs of ubiquitin, UBM2, and UBQ binders in the unbound state, together with eight 150 ns MD simulations of native proteins from the Phospho set against the respective S_{QH} values. To avoid the artificial correlation between these two quantities due to their direct dependence on protein size (see above), we normalize both quantities by the number of degrees of freedom, that is, $3N-6$ for S_{QH} and $3N$ for S_{VAR}. As seen in Figure 2B, the linear correlation between S_{QH} and S_{VAR} is moderate ($R^2 = 0.75$), suggesting that the S_{QH} to S_{VAR} ratio is not constant and depends on the protein in question. Expectedly, the value of S_{QH} is considerably smaller than S_{VAR} in all cases since the inclusion of linear, pairwise correlations between atomic displacements (i.e., atom-positional covariances) invariably decreases the total value of conformational entropy. Overall, the off-diagonal terms decrease the value of size-normalized S_{VAR} by approximately 60−70%, but the exact magnitude of this effect is protein dependent.

Importantly, a smaller than typical contribution of covariance terms to S_{QH} ($S_{QH}/S_{VAR} \sim 0.4$) is observed in the simulations of unbound UBM2. In the unbound state, this 32-residue polypeptide displays high flexibility in the course of the simulation (rmsd from the initial conformation reaches approximately 8 Å on average after 1 μs, Figure S1, Supporting Information, with highly diverse ensembles); that is, it behaves like an intrinsically disordered protein (IDP). Given this, one might assume that the lack of a regular structure accounts for a higher S_{QH}/S_{VAR} ratio than in the case of fully structured proteins. However, for the two genuine IDPs from the Phospho set (CPI17, OdhI), we found that their S_{QH}/S_{VAR} ratios (0.34 and 0.35, respectively) are more similar to those of globular proteins than to that of UBM2. We speculate that the contribution of off-diagonal terms to S_{QH} depends more strongly on packing density than just on the presence of regular secondary or tertiary structure in a protein. Namely, high packing density (i.e., compactness in general) is expected to induce a higher level of correlated motions in proteins and vice versa. To test this idea, we used a simple estimate of protein compactness, that is, ratio of protein solvent accessible surface area in folded and fully extended conformations (C_F), to categorize proteins in our study. The distribution of C_F values obtained from simulations of all proteins in the study (see Tables 1, 2, and 3) is given in Figure 2C. A typical simulated protein exhibits C_F of 0.4 (median of the distribution). However, in accordance with our expectation, for unbound UBM2, this ratio exhibits a value of 0.6 on average (Figure 2C), making it the most loosely packed protein in our study. Interestingly, the most compact protein here is the thermostable Xyn10B (C_F = 0.3). We expect that inclusion of supralinear correlations between eigenmodes in S_{QH} calculations might even potentiate the difference between compact and more loosely packed proteins.[44]

3.4. Linear Relationship between ΔS_{QH} and ΔS_{VAR} upon Binding and Phosphorylation. Using the set of five

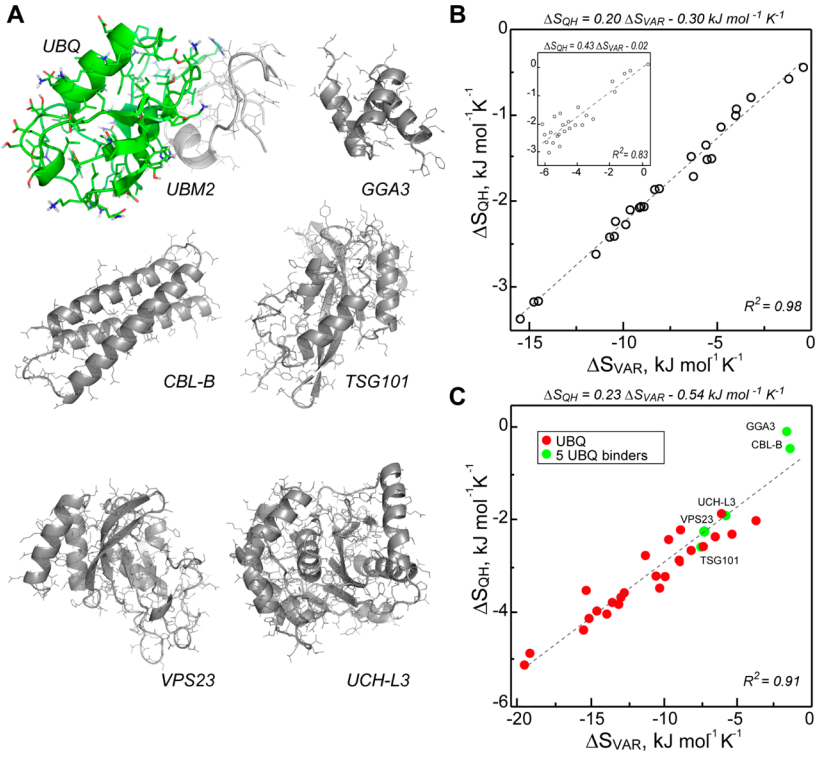

Figure 3. Binding-induced conformational entropy change. (A) Experimental structures of ubiquitin (green), UBM2 (light gray), and 5 other binding partners (dark gray). (B) ΔS_{QH} and ΔS_{VAR} upon binding for ubiquitin and UBM2. We show 25 points obtained from all the combinations of 5 trajectories of bound and 5 trajectories of unbound ubiquitin and the same for UBM2 (inset). (C) ΔS_{QH} and ΔS_{VAR} for ubiquitin (red circles) and its 5 binding partners other than UBM2 (green circles).

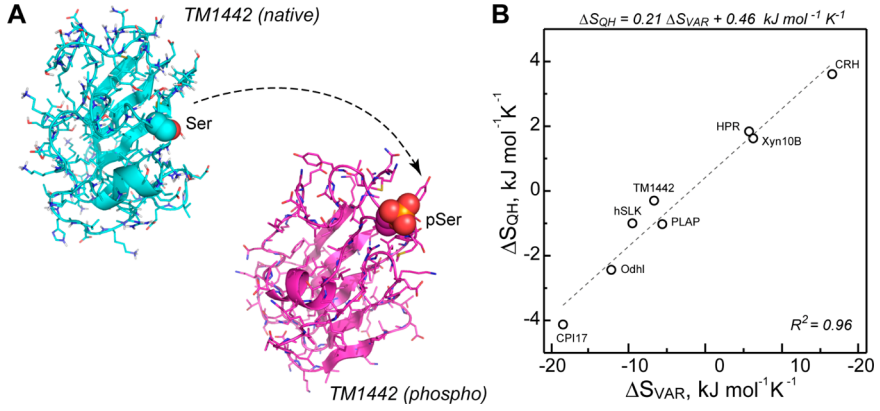

Figure 4. Phosphorylation-induced conformational entropy change. (A) 3D structures for the final MD conformations of TM1442 in native (cyan) and phosphorylated (magenta) states. Proteins are shown in cartoon representations, while the phosphorylation site is shown with spheres. (B) ΔS_{QH} and ΔS_{VAR} upon phosphorylation for 8 different proteins.

1 μs-long MD trajectories of ubiquitin in bound and unbound states (Figure 3), we estimated the change in its quasi-harmonic conformational entropy upon binding to UBM2 and other UBQ binders. All possible combinations of 5 trajectories of unbound UBQ and 5 trajectories of UBQ bound to UBM2 give 25 pairs of ΔS_{QH} and ΔS_{VAR} values, which are given in Figure 3B. Ubiquitin conformations when bound to UBM2 deviate from the unbound ones by 2.7−5.7 Å (average backbone rmsd between two trajectories, $\langle rmsd_B \rangle$, Table 1). These values are calculated over all residues in UBQ, with the highest contribution coming from the flexible tail encompassing residues 71−76. Similarly, ubiquitin bound to 5 proteins other than UBM2 (Figure 3C) adopts conformations that deviate from unbound conformers by 2.4−5.9 Å (Table 2). On the other hand, UBQ binders change their conformation upon binding anywhere from approximately 2.0 (CBL-B) to 10.0 Å (GGA3) on average. The thus obtained ΔS_{QH} values are distributed between −0.07 and −2.5 kJ mol^{-1} K^{-1} with an average of −1.4 kJ mol^{-1} K^{-1}, indicating a significant entropic penalty to binding, which is comparable in magnitude to the one reported for a 144-residues-long protein TSG101 interacting with a small peptide of −0.84 kJ mol^{-1} K^{-1}.[45] However, in the present case, entropic penalty for TSG101 bound to UBQ is almost three times higher at 2.5 kJ mol^{-1} K^{-1}.

Importantly, ΔS_{QH} values for ubiquitin binding display a strong linear correlation with ΔS_{VAR} over the whole range of values of ΔS_{QH}. The relationship between the two can be fit using an equation:

$$\Delta S_{QH} = 0.20\Delta S_{VAR} - 0.30 \text{ kJ mol}^{-1} \text{ K}^{-1} \quad (10)$$

with $R^2 = 0.98$, showing that the off-diagonal, covariance terms in the variance−covariance matrix lower the overall quasi-harmonic entropy change by approximately 80%, and that this fraction is largely invariant for very different simulated ensembles of ubiquitin. The nonzero value of the intercept is likely a consequence of the finite number of points used in the fit (Figure 3B). Similarly, for the UBQ binders set all obtained ΔS_{QH} values correlate linearly with the respective ΔS_{VAR} with $R^2 = 0.91$ (Figure 3C) and can be fit using the following linear regression equation:

$$\Delta S_{QH} = 0.23\Delta S_{VAR} - 0.54 \text{ kJ mol}^{-1} \text{ K}^{-1} \quad (11)$$

In contrast, ΔS_{QH} of UBM2 upon binding to ubiquitin does not display such a strong linear correlation with ΔS_{VAR} (Figure 3B, inset, $R^2 = 0.83$). We speculate that this difference is due to the prominently lower packing density of unbound UBM2 compared to the rest of the proteins in our study, resulting in a more varied impact of the off-diagonal terms in the covariance matrix on conformational entropy in bound and unbound states, and their less efficient cancellation upon subtraction. What is more, in this case, the off-diagonal terms lower the overall conformational entropy change by approximately 60% only.

On the other hand, we again observe a strong linear ΔS_{QH} dependence on ΔS_{VAR} for the entire Phospho set (Figure 4). Here, the attachment of the phosphate group leads either to a decrease or an increase in ΔS_{QH} (ranging between approximately −4 and 4 kJ mol^{-1} K^{-1}, Figure 4B), reflecting on the whole the level of conformational similarity between the two states as embodied in the average rmsd between phosphorylated and native trajectories (Table 3). Importantly, ΔS_{QH} exhibits very similar functional dependence on ΔS_{VAR} as in the case of ubiquitin binding, which can be fit using an equation:

$$\Delta S_{QH} = 0.21\Delta S_{VAR} + 0.46 \text{ kJ mol}^{-1} \text{ K}^{-1} \quad (12)$$

with $R^2 = 0.96$ (Figure 4B). This linear trend is maintained for 8 different proteins, whose sizes vary from 86 to 479 residues (see Table 3 for details).

Since ΔS_{VAR} is directly related to changes in protein flexibility (RMSF, see eq 5), one might expect that the dynamics in a particular region of a given molecule, such as the vicinity of the binding interface or phosphorylation site, would contribute most to the conformational entropy change. However, structural mapping of ΔRMSF values upon binding of ubiquitin to UBM2 in 5 independent MD runs (Figure S5, Supporting Information) or upon addition of the phosphate group to different proteins (Figure S6, Supporting Information) indicates significant changes of flexibility in the entire molecule with very little correlation with the distance from the binding or the phosphorylation site. Therefore, one can assume that the observed entropy changes result from the global modulation of protein dynamics.

Finally, if one combines all data sets (excluding only one prominent outlier, UBM2) and looks at changes in conformational entropy per degree of freedom (i.e., $\Delta S_{VAR}/3N$ and $\Delta S_{QH}/(3N-6)$), one still obtains a strong linear correlation ($R^2 = 0.93$) with an almost vanishing intercept (Figure 5):

$$\Delta(S_{QH}/(3N-6)) = 0.24\Delta(S_{VAR}/3N) - 0.05 \text{ J mol}^{-1} \text{ K}^{-1} \quad (13)$$

In the case of a typical individual protein (where $10^2 \leq N \leq 10^4$), this equation can be simplified to

$$\Delta S_{QH} = 0.24\Delta S_{VAR} \quad (14)$$

In other words, the contribution of linear, pairwise correlations in atomic positions to the change in quasi-harmonic conformational entropy per degree of freedom in two very different biological processes (binding and phosphorylation) and for 14 structurally and dynamically very different proteins appears to be remarkably constant and predictable.

Several comments should be provided here. First, it should be emphasized that eqs 13 and 14 are empirical fits whose success in the present context is arguably a consequence of certain general, invariant properties of protein dynamics rather than of any generic properties of variance−covariance matrices or conformational entropy itself. Second, the fact that the above correlations are markedly stronger in the case of entropy differences compared to absolute values implies a more complex connection, which is not resolved in the present study. Third, eqs 13 and 14 can only be applied to proteins with a similar level of compactness as in our model set (but not to extreme cases such as UBM2), which, however, does seem to be typical of most protein structures obtained by X-ray crystallography.

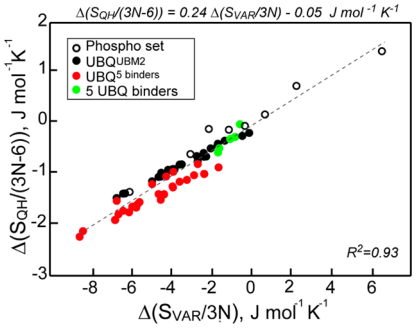

Figure 5. Linear dependence of $\Delta(S_{QH}/3N-6)$ on $\Delta(S_{VAR}/3N)$. The points are shown for 25 values from the simulations of ubiquitin in complex with UBM2 (black circle), 25 values from the simulations of ubiquitin in complex with other binders (red circles), 5 values from the simulations of 5 ubiquitin binders other than UBM2 (green circles) and 8 simulations of phosphorylated and native proteins (open circles).

3.5. Relative Conformational Entropy of Binding Obtained from Crystallographic B-Factors. We have used eq 14 together with eq 5 to estimate changes in conformational entropy upon binding for a large set of representative X-ray structures of bound and unbound proteins (see Table S1, Supporting Information, for details). The distribution of conformational entropy changes (obtained from B-factors scaled to 300 K, see the Methods section) for each individual binding partner is given in Figure 6A. Except for several extreme cases, most of $\Delta S_{binding}^{single}$ are distributed between −2 and 2 kJ mol^{-1} K^{-1} and exhibit an average of 0.87 kJ mol^{-1} K^{-1}. The median of the distribution (0.71 kJ mol^{-1} K^{-1}) is also positive, suggesting a favorable contribution of the conformational entropy to the binding free energy, which at 300 K accounts for −213 kJ mol^{-1}. The median value for our

dx.doi.org/10.1021/ct300082q | J. Chem. Theory Comput. 2012, 8, 3820−3829

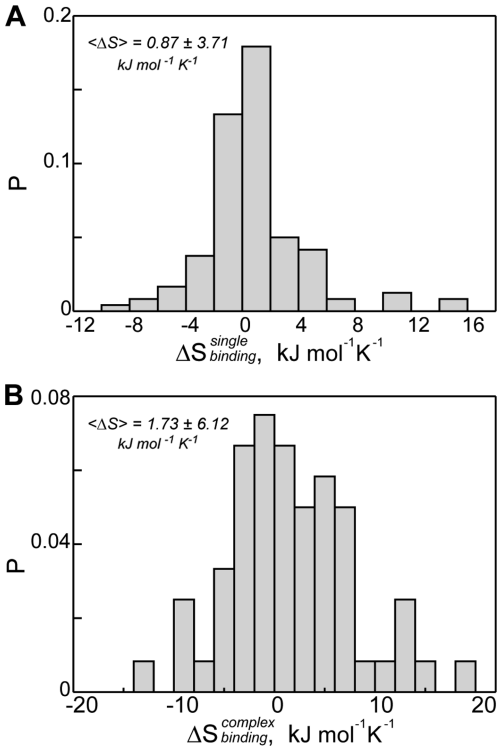

Figure 6. Relative conformational entropy of binding calculated from crystallographic B-factors. Normalized distributions for (A) 120 individual proteins and (B) their 60 complexes. In panel A, $\Delta S^{single}_{binding} = 0.24 \times (S^{bound}_{VAR} - S^{unbound}_{VAR})$, while in panel B, $\Delta S^{complex}_{binding} = 0.24 \times (S^{proteinsA+B,bound}_{VAR} - S^{proteinA,unbound}_{VAR} - S^{proteinB,unbound}_{VAR})$. All B-factors have been scaled to 300 K (see the Methods section).

distribution is approximately two times greater than the one obtained by NMR for ferredoxin-NADP$^+$ reductase (0.41 kJ mol^{-1} K^{-1})[23] and is in a similar range with other NMR-derived relative conformational entropies of binding for calmodulin (0.06−0.3 kJ mol^{-1} K^{-1}).[4,5] The mean and the median of the $\Delta S^{single}_{binding}$ distribution normalized by the number of residues are 3.0 and 4.1 J mol^{-1} K^{-1}, respectively, which is almost an order of magnitude lower than the conformational entropy penalty per residue estimated for protein folding.[46] Furthermore, the distribution of $\Delta S^{complex}_{binding}$ for protein complexes (Figure 6B) is also skewed in the direction of favorable contributions to the free energy of binding. Its mean and median are 1.73 and 1.23 kJ mol^{-1} K^{-1}, respectively, which is comparable to, but still more pronounced than, the magnitude of the conformational entropy penalty obtained in the experimental studies of antibody interactions with epidermal growth factor receptor 2 (−0.3 kJ mol^{-1} K^{-1})[47] or calculated for the UEV domain of the protein TSG101 and an HIV-derived nonapeptide (−0.8 kJ mol^{-1} K^{-1}) using the mutual information expansion formalism.[45] Importantly, favorable contribution of conformational entropy to the free energy for more than one-half of the proteins in our data set goes against the conventional wisdom assuming that binding typically incurs a conformational entropy penalty, and should be explored in future research.

4. CONCLUSION

Our results, based on extensive MD simulations in the microsecond range, have enabled us to demonstrate a simple linear connection between relative quasi-harmonic conformational entropy and its approximation obtained using mass-weighted atomic fluctuations only. Using this relation, we have estimated conformational entropy changes upon binding from the crystallographic B-factors of a large set of X-ray structures. Application of the formalism to the model protein set gives on average a reasonable estimate of favorable conformational entropy contribution in the binding free energy, which displays correspondence with the published experimental data. Importantly, our results also indicate that, in many cases, conformational entropy can have a favorable contribution to the overall binding free energy. On the other hand, a shortcoming of the described formalism is that it can be applied only to estimate the relative values of conformational entropy, with the absolute values still being out of reach. Also, reported linearity of the relationship between relative conformational entropy and its S_{VAR} approximation appear to be reliable and reproducible only for proteins above a certain level of packing density. The latter issue might be resolved by extending the formalism to explicitly include an additional contribution of protein compactness, which is a subject of our future studies. Finally, our results open up a possibility that conformational entropy estimates based on other experimental observables, which also suffer from the inability to provide information on correlated motion in proteins, such as NMR S^2 order parameters, might also be easily corrected by a linear expression of the kind presented here. This would not only help to reassess a large body of previous work in this area but also open up new possibilities when it comes to conformational entropy measurements.

■ ASSOCIATED CONTENT

Supporting Information

MD evolutions of RMSDs in UBQ/UBM2, Phospho, and 5 UBQ binders sets, B-factor distributions for eukaryotic lysozyme at 100K and 300K, RMSF changes along ubiquitin upon binding to UBM2 and proteins undergoing phosphorylation, details about PDB set used for entropy change estimations from B-factors. This information is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bojan.zagrovic@univie.ac.at.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by the Austrian Science Fund FWF (START Grant No. Y 514-B11 to B.Z.) and European Research Council (ERC Starting Independent grant to B.Z.). We thank Ruben Zubac for his extensive contributions to this work and the members of the Laboratory of Computational Biophysics at MFPL for useful advice and assistance. This work is dedicated to Wilfred F. van Gunsteren to whom the authors wish fields of force, maxima of energy, and landscapes of potential in the next 65 as well!

■ REFERENCES

(1) Mittermaier, A.; Kay, L. E. *Science* **2006**, *312*, 224.
(2) Kast, D.; Espinoza-Fonseca, L. M.; Yi, C.; Thomas, D. D. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8207.
(3) Kern, D.; Zuiderweg, E. R. *Curr. Opin. Struc. Biol.* **2003**, *13*, 748.
(4) Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831.

(5) Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. *Nature* **2007**, *448*, 325.
(6) Polyansky, A. A.; Zubac, R.; Zagrovic, B. *Methods Mol. Biol.* **2012**, *819*, 327.
(7) Akke, M.; Bruschweiler, R.; Palmer, A. G. *J. Am. Chem. Soc.* **1993**, *115*, 9832.
(8) Li, Z.; Raychaudhuri, S.; Wand, A. J. *Protein Sci.* **1996**, *5*, 2647.
(9) Yang, D.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 369.
(10) Sapienza, P. J.; Lee, A. L. *Curr. Opin. Pharmacol.* **2010**, *10*, 1.
(11) Gilson, M. K.; Zhou, H. X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21.
(12) Marlow, M. S.; Dogan, J.; Frederick, K. K.; Valentine, K. G.; Wand, A. J. *Nat. Chem. Biol.* **2010**, *6*, 352.
(13) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325.
(14) Edholm, O.; Berendsen, H. J. C. *Mol. Phys.* **1984**, *51*, 1011.
(15) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617.
(16) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289.
(17) Baron, R.; Hunenberger, P. H.; McCammon, J. A. *J. Chem. Theory Comput.* **2009**, *5*, 3150.
(18) Chang, C.-E.; Chen, W.; Gilson, M. K. *J. Chem. Theory Comput.* **2005**, *1*, 1017.
(19) Carlsson, J.; Aqvist, J. *J. Phys. Chem. B* **2005**, *109*, 6448.
(20) Polyansky, A. A.; Zagrovic, B. Protein Electrostatic Properties Predefining the Level of Surface Hydrophobicity Change upon Phosphorylation. *J. Phys. Chem. Lett.* **2012**, *3*, 973−976.
(21) Kuzmanic, A.; Zagrovic, B. *Biophys. J.* **2010**, *98*, 861.
(22) Kuzmanic, A.; Kruschel, D.; van Gunsteren, W. F.; Pannu, N. S.; Zagrovic, B. *J. Mol. Biol.* **2011**, *411*, 286.
(23) Lee, Y. H.; Ikegami, T.; Standley, D. M.; Sakurai, K.; Hase, T.; Goto, Y. *ChemBioChem* **2011**, *12*, 2062.
(24) Willis, B. T. M.; Pryor, A. W. *Thermal Vibrations in Crystallography*; Cambridge University Press: London New York, 1975.
(25) Cui, G.; Benirschke, R. C.; Tuan, H.-F.; Juranić, N.; Macura, S.; Botuyan, M. V.; Mer, G. *Biochemistry* **2010**, *49*, 10198.
(26) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531.
(27) Sundquist, W. I.; Schubert, H. L.; Kelly, B. N.; Hill, G. C.; Holton, J. M.; Hill, C. P. *Mol. Cell* **2004**, *13*, 783.
(28) Pornillos, O.; Alam, S. L.; Rich, R. L.; Myszka, D. G.; Davis, D. R.; Sundquist, W. I. *EMBO J.* **2002**, *21*, 2397.
(29) Teo, H.; Veprintsev, D. B.; Williams, R. L. *J. Biol. Chem.* **2004**, *279*, 28689.
(30) Ren, X.; Hurley, J. H. *EMBO J.* **2011**, *30*, 2130.
(31) Misaghi, S.; Galardy, P. J.; Meester, W. J.; Ovaa, H.; Ploegh, H. L.; Gaudet, R. *J. Biol. Chem.* **2005**, *280*, 1512.
(32) Johnston, S. C.; Larsen, C. N.; Cook, W. J.; Wilkinson, K. D.; Hill, C. P. *EMBO J.* **1997**, *16*, 3787.
(33) Peschard, P.; Kozlov, G.; Lin, T.; Mirza, I. A.; Berghuis, A. M.; Lipkowitz, S.; Park, M.; Gehring, K. *Mol. Cell* **2007**, *27*, 474.
(34) Prag, G.; Lee, S.; Mattera, R.; Arighi, C. N.; Beach, B. M.; Bonifacino, J. S.; Hurley, J. H. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2334.
(35) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.
(36) Schuler, L.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205.
(37) Bjarnadottir, U.; Nielsen, J. E. *Biopolymers* **2012**, *97* (1), 65−72.
(38) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Interaction Models for Water in Relation to Protein Hydration*; Reidel: Dordrecht, 1981.
(39) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(40) Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. *J. Comput. Chem.* **1997**, *18*, 1463.
(41) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451.
(42) Jolliffe, I. T. *Principal Component Analysis*, 2nd ed.; Springer: New York, 2002.
(43) Pyrkov, T. V.; Chugunov, A. O.; Krylov, N. A.; Nolde, D. E.; Efremov, R. G. *Bioinformatics* **2009**, *25*, 1201.
(44) Baron, R.; van Gunsteren, W. F.; Hünenberger, P. H. *Trends Phys. Chem.* **2006**, *11*, 87.
(45) Killian, B. J.; Kravitz, J. Y.; Somani, S.; Dasgupta, P.; Pang, Y.-P.; Gilson, M. K. *J. Mol. Biol.* **2009**, *389*, 315.
(46) Pace, C. N.; Fu, H.; Fryar, K. L.; Landua, J.; Trevino, S. R.; Shirley, B. A.; Hendricks, M. M.; Iimura, S.; Gajiwala, K.; Scholtz, J. M.; Grimsley, G. R. *J. Mol. Biol.* **2011**, *408*, 514.
(47) Bostrom, J.; Haber, L.; Koenig, P.; Kelley, R. F.; Fuh, G. *PLoS One* **2011**, *6*, e17887.