Supporting Information

Wlodarski and Zagrovic 10.1073/pnas.0906966106

SI Text

Selection Criteria for the X-Ray Structures of Bound Ubiquitin. The structural dataset used here (Table S2) was created by filtering the set used by Lange and coworkers (1, 2) and applying more stringent selection criteria to it to make it, albeit smaller, more representative of the full diversity of ubiquitin interactions. Our main objective was to create a nonbiased dataset containing all known ubiquitin binding interfaces, but exhibiting as little as possible sampling bias for any particular interaction or ligand. Briefly, we chose only those ubiquitin X-ray structures that: (*i*) were cocrystallized with a covalently or noncovalently bound partner (including polyubiquitin chains), and (*ii*) are bound to different ligands except when the same ligand is bound to two different binding sites, or when the backbone rmsd between the two ubiquitin conformers bound to the same ligand in the same binding site exceeds 0.8 Å. Although smaller, the present dataset does fully describe all binding interfaces found in ubiquitin via crystallographic studies (3).

MDS. The procedure for generating MDS maps was as follows: (*i*) we used MDS to map all 19 X-ray structures into 2D space, (*ii*) the distances between points in this map were then kept fixed during the next part of analysis, and (*iii*) we used MDS to map each NMR structure separately to the fixed X-ray MDS map. In this way, we obtained a set of points in 2D space (Fig. 1), each representing one X-ray or NMR structure, and distributed in such a way that the distances between points are proportional to the corresponding rmsd values. However, because of the above modifications in the MDS algorithm, the position of points representing a given NMR structure is not influenced by other NMR structures, but only by the whole X-ray ensemble. The algorithm performs the above dimensionality reduction by minimizing the Sammon stress function:

$$S_{sam} = \frac{1}{\sum_{i<j} d_{ij}} \sum_{i<j} \frac{(d_{ij} - d^*_{ij})^2}{d_{ij}}. \qquad [1]$$

where d^*_{ij} and d^*_{ij} are distances (rmsd value) between structures in the original and reduced space, respectively.

KS Test. The KS test is a statistical significance test with a null hypothesis that the two distributions that are being compared are drawn from the same continuous distribution. Its statistic for a two sample test is defined by:

$$D_n = \sup_x |F_n(x) - F_{n'}(x)|,$$

where $F_n(x)$ and $F_{n'}(x)$ are empirical distribution functions for the first and second of the compared distributions, respectively:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x),$$

where X_i are elements of distribution and $I(X_i \leq$ [lt]$x)$ is the indicator function, which is equal to 1 when $X_i \leq$ [lt]x, and 0 otherwise. The null hypothesis is rejected on an α-level of significance when

$$\sqrt{\frac{nn'}{n+n'}} D_{n,n'} > K_\alpha,$$

where K_α is obtained from:

$$Pr(K \leq K_\alpha) = 1 - \alpha,$$

where K is drawn from the Kolmogorov distribution. Here, it should be noted that the KS test assumes that the samples are drawn independently from the nonparametric distribution, and by using it we neglect, to a first approximation, the possibility of local correlations.

Analysis of Additional Structural Ensembles Refined from the Lange and Coworkers (1, 2) RDC Data. The additional Backrub ensembles used in our analysis were generated by T. Kortemme and coworkers (4) using Monte Carlo Backrub moves with 3- and 12-residue segments at kT = 2.4 and 1.2, respectively. We have redone the complete local rmsd and *P* value analysis for these two ensembles to verify our findings obtained with the EROS ensemble. The results of this analysis are summarized and compared with our principal findings in Table S1. The δ-values show that local conformational deviations close to the binding site (0.5-Å range) are greater than the global deviations (25-Å range), to a degree that is in some cases greater and in some lower compared with the EROS ensemble. In general, the δ-values for the cases with the C-terminal tail included (relevant for induced fit) are significantly greater for both Backrub ensembles than for the EROS ensemble (e.g., 133% versus 78% for backbone for Backrub 1.2 and EROS, respectively), whereas without the C-terminal tail included, they are on average lower (e.g., 13% versus 30% for backbone, for Backrub 1.2 and EROS, respectively). Moreover, the *P* value analysis for the Backrub ensembles suggests that conformational changes close to the binding site are more pronounced than the global conformational changes, which is in agreement with our main findings. For example, for distances up to 5 Å from the binding site, for ≈50% of structures for both Backrub ensembles with and without the tail included, $P < 0.1$. Altogether, these results obtained on other structural ensembles, which still well reproduce the RDC data, support all of our main conclusions obtained for the EROS ensemble. Here, it should be mentioned that although the Backrub ensembles agree well with the RDC data from Lange and coworkers (1, 2), the original data were not used directly in the course of their refinement, but rather in a selection procedure (4). The structures generated in a Backrub Monte Carlo procedure were filtered for those that agree well with experimental data. This is the fundamental difference between these ensembles and the EROS ensemble, and it may be that because of it, the latter ensemble may be somewhat more physical as it was generated using time-dependent molecular dynamics.

Analysis of Correlations in the Fluctuations of the Unbound State of Ubiquitin. We have used the *g_covar* routine included in the Gromacs package (5) to obtain the mass-weighted variance-covariance matrix of structural fluctuations for the EROS ensemble of unbound ubiquitin structures after backbone-atom least-squares superposition. The calculated diagonal elements of the variance-covariance matrix, $\sigma_{i,i}$, are variances of positions of each of the atoms, whereas the off-diagonal elements, $\sigma_{i,j}$, are covariances between the *i*th and the *j*th atom, defined by: $\sigma_{i,j} = <(\mathbf{r}_i - <\mathbf{r}_i>)(\mathbf{r}_j - <\mathbf{r}_j>)>$, where $\mathbf{r}_i$ is the instantaneous position of the *i*th atom (a three-element vector containing the *x*, *y*, and *z* positions) and $<\mathbf{r}_i>$ is its average position over the entire ensemble. The variance-covariance matrix was converted into a

correlation matrix by normalizing each covariance element in the matrix by the associated variances (6):

$$c_{i,j} = \frac{\sigma_{i,j}}{\sqrt{\sigma_{i,i}\sigma_{j,j}}}.$$

In Fig. S4, we show a correlation map for the unbound EROS ensemble of ubiquitin. To relate this to binding, we have marked with crosses on the correlation map all of the atoms belonging to the residues in the binding site of ubiquitin for structure 1YD8 (see also Fig. 4). This is the structure whose induced-fit motions in the course of binding are, in terms of their magnitude, closest to the average obtained for all bound structures analyzed in this study. The variances of the binding-site atoms themselves are marked with crosses on the diagonal, together with their covariances off of the diagonal. Clearly, the atoms in the binding site exhibit significant correlated motions already in the unbound state. However, except between each other, none of these atoms exhibit significant correlated motions with any other atoms: all of the major correlated motions occur only within the binding site. One may speculate that precisely these correlated motions of the binding-site atoms contribute to conformational selection. Interestingly, there are no major long-range correlated motions between the binding site and other parts of the ubiquitin's structure, a feature that is sometimes seen in allosterically controlled proteins.

1. Lange OF, et al. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320:1471–1475.
2. Lakomek N, et al. (2008) Residual dipolar couplings as a tool to study molecular recognition of ubiquitin. *Biochem Soc Trans* 36:1433–1437.
3. Hicke L, Schubert HL, Hill CP (2005) Ubiquitin-binding domains. *Nat Rev* Mol Cell Biol 6:610–621.
4. Friedland GD, Lakomek N, Griesinger C, Meiler J, Kortemme T (2009) A correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. *PLoS Comput Biol* 5:e1000393.
5. Van Der Spoel D, et al. (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26:1701–1718.
6. Theobald DL, Wuttke DS (2008) Accurate structural correlations from maximum likelihood superpositions. *PLoS Comput Biol* 4:e43.

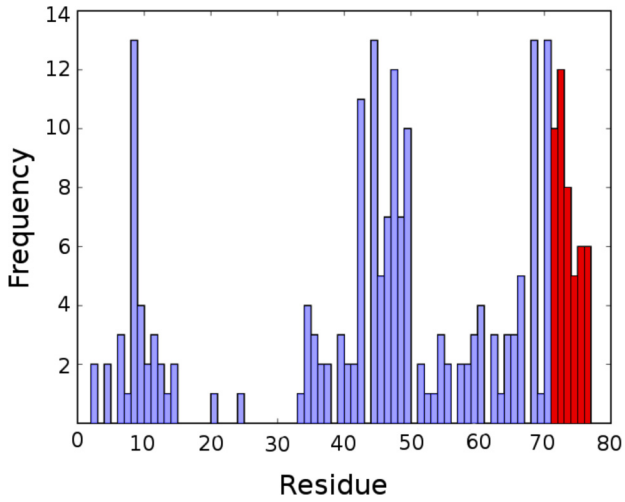

Fig. S1. The role of ubiquitin's tail in binding. Frequency of occurrence of ubiquitin's residues as a part of the binding interface, defined as all ubiquitin atoms in the 0.5-Å distance range (see *Materials and Methods*) and calculated from the X-ray ensemble; the tail residues (residues 71–76) are colored red.

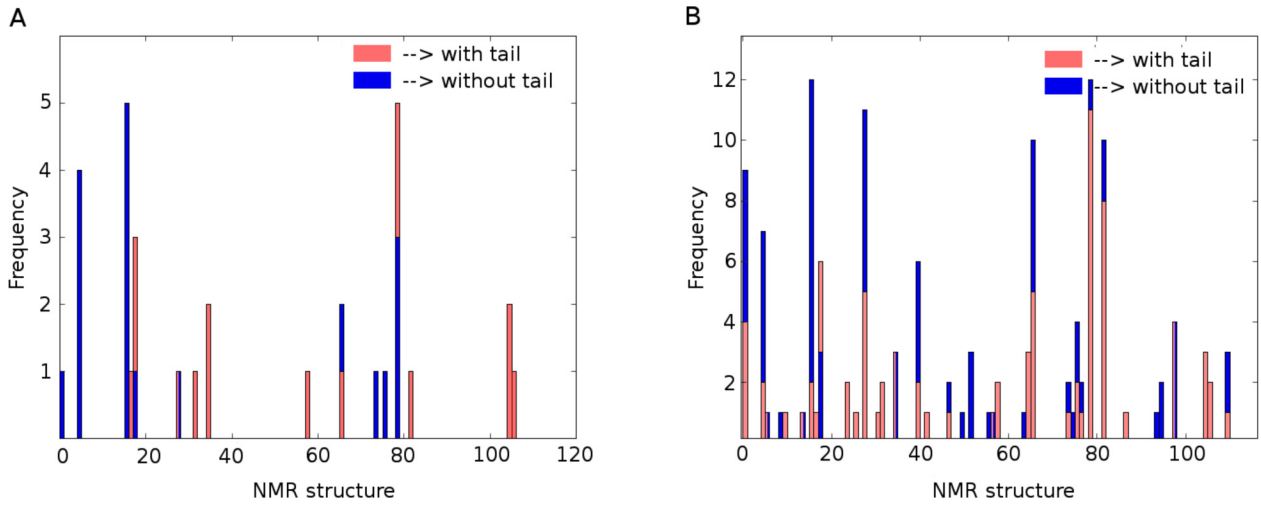

Fig. S2. Characterization of the conformationally selected structures. (*A*) The frequency of how often each unbound NMR structure of ubiquitin from the EROS ensemble is the one that is the most similar to one of the 19 bound X-ray structures in the rmsd sense (red, with tail; blue, without tail). Overall, 63% of X-ray structures are captured by the top three most similar NMR structures without the tail, whereas this number drops to 52% with the tail included. (*B*) The frequency of how often each unbound NMR structure of ubiquitin from the EROS ensemble has rmsd value within a range of 0.1 Å from the rmsd value of the most similar NMR structure to each X-ray.

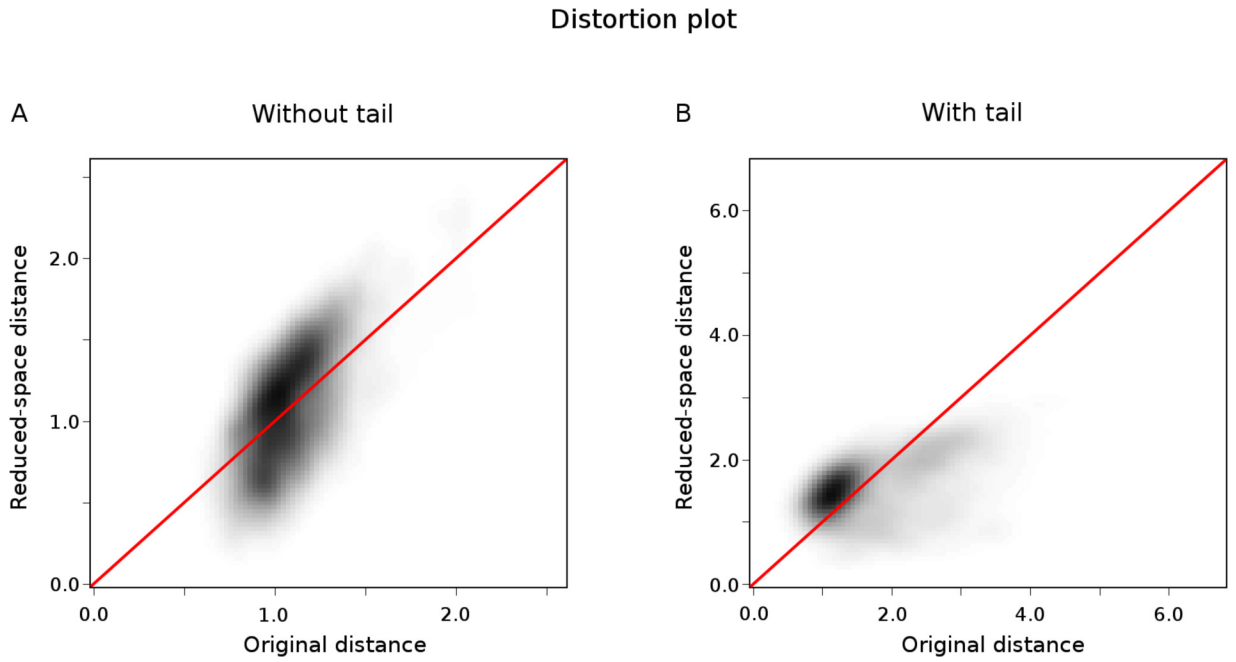

Fig. S3. MDS distortion plots. Density distortion plots describe how well the MDS method (Sammon mapping) reduces the dimensionality of the data by plotting original distances against mapped (reduced-space distance). We have used density plots to better present the most common relation between this two distances. For ideal mapping points should be spread close to red curve. An additional measure of the quality of mapping in MDS is the value of Sammon stress function, which is minimized during MDS calculations. For our dataset we obtain mean S_{sam} equal 0.056 with standard deviation: 0.016 and 0.129 with 0.055, without and with the C-terminal tail, respectively.

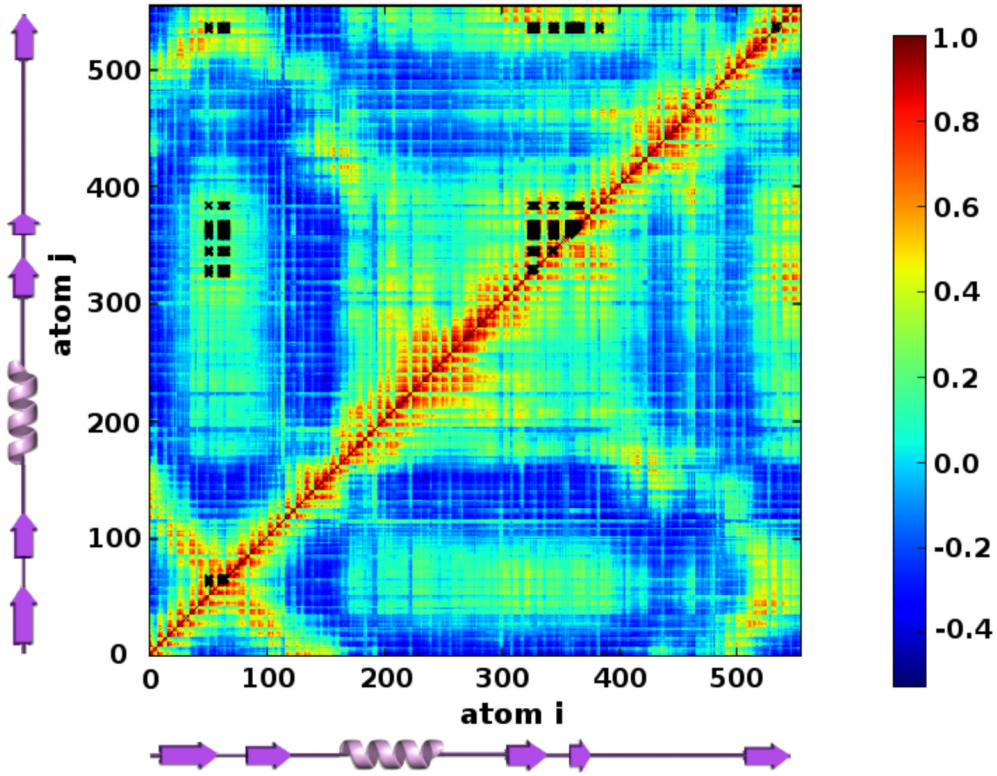

Fig. S4. Analysis of the correlations in the fluctuations of the unbound state of ubiquitin. Correlation map (normalized variance-covariance matrix) for the EROS ensemble of ubiquitin. The variances and covariances for atoms belonging to the residues in the binding site of ubiquitin for a bound complex structure 1YD8 are depicted with crosses. The secondary structure elements in ubiquitin, in the N to C direction, are depicted next to the axes.

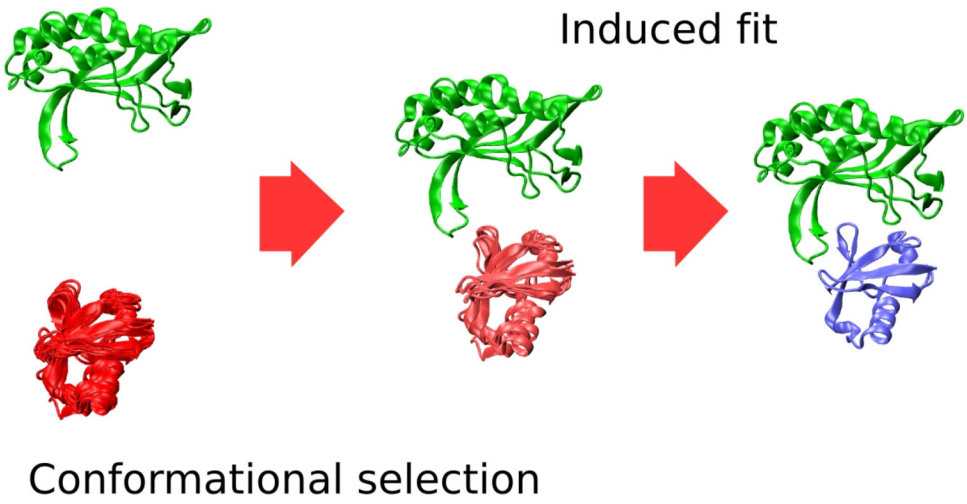

Fig. S5. Model of protein–protein interactions for ubiquitin. The first step in binding entails conformationally selecting an unbound conformation that is structurally similar to the bound state conformation from the spectrum of available conformers. After conformational selection, one observes a population shift in the direction of those conformers that are structurally optimized for binding. The last step is induced fit optimization of the interactions in the binding site, which, as shown in the present manuscript for ubiquitin, could be significant. For illustration purposes, here we use the TSG101 protein in complex with ubiquitin (PDB ID code 1S1Q).

Other Supporting Information Files

Table S1 (PDF)
Table S2 (PDF)