Determination of protein structural ensembles by hybrid-resolution SAXS restrained Molecular Dynamics.

Cristina Paissoni[1,], Alexander Jussupow[2], and Carlo Camilloni[1,*]*

[1]Dipartimento di Bioscienze, Università degli Studi di Milano, via Celoria 26, 20133 Milano, Italy [2]Department of Chemistry and Institute of Advanced Study, Technical University of Munich, Garching 85747, Germany.

SAXS, Molecular Dynamics, Conformational Ensembles, Protein Dynamics, Ubiquitin, Metainference, Metadynamics.

ABSTRACT. SAXS experiments provide low-resolution but valuable information about the dynamics of biomolecular systems, which could be ideally integrated in MD simulations to accurately determine conformational ensembles of flexible proteins. The applicability of this strategy is hampered by the high computational cost required to calculate scattering intensities from three-dimensional structures. We previously presented a metainference-based hybrid resolution method that makes atomistic SAXS-restrained MD simulation feasible by adopting a coarse-grained approach to efficiently back-calculate scattering intensities; here, we extend this technique, applying it in the framework of multiple-replica simulations with the aim to investigate the dynamical behavior of flexible biomolecules. The efficacy of the method is assessed on the

K63-diubiquitin multi-domain protein, showing that inclusion of SAXS-restraints is effective in generating reliable and heterogenous conformational ensemble, also improving the agreement with independent experimental data.

1. INTRODUCTION

Biomolecules in solution can be characterized by a different extent of conformational dynamics depending on the specific system and experimental conditions[1–3]. While the dynamics of single domain proteins in native condition is generally limited to fluctuations around a well-defined structure, fully disordered proteins can only be described as statistical ensembles of conformations. In between these cases multi domain proteins connected by linker region can populate multiple states generally characterized by a different size[4].

Experimentally the characterization of conformational heterogeneity can be achieved by employing multiple solution techniques like nuclear magnetic resonance (NMR), Förster resonance energy transfer (FRET) and small angle X-ray scattering (SAXS)[1,2]. The latter has the advantage to be label free, to work with systems of any size and in essentially all experimental conditions[5]. An atomistic interpretation of scattering data could benefit from its combination with computational techniques, as Molecular Dynamics (MD) simulations, which could provide an accurate physical model to generate reliable conformational ensembles in agreement with SAXS data[6]. Common approaches employ SAXS to reweight conformational ensembles *a posteriori*, making use of statistically founded theoretical frameworks[7–12]. Recently, few methods in which SAXS experimental data are integrated in MD to drive conformational sampling have been proposed, nevertheless their application is hindered by the high computational cost required to

calculate scattering intensities[13–15]. As this is a major issue in the field, multiple strategies have been suggested to alleviate this problem[16–19].

In a previous work[20], we developed a MD-based multi-resolution strategy to efficiently refine protein-DNA and protein-RNA complexes integrating SAXS experimental data with metainference[21]. According to this strategy, MD is run with full atomistic details, using standard atomistic force-field, while the back-calculation of SAXS intensities is performed in a coarse-grain fashion, based on the Martini force field[22]. In the refinement protocol conformational averaging was not considered, under the assumption that a single structure, representing the most populated state of the system, could reliably reproduce all the measured experimental data used as restraints. In this work, we aim to further extend this approach to investigate the conformational space of biomolecules that can adopt multiple conformations in solution.

Here we applied our multi-resolution strategy to investigate the conformational ensemble of K63-linked diubiquitin (K63-Ub_2). Diubiquitins represent an ideal test system as they are known to populate multiple conformational states due to the presence of a highly flexible linker connecting the C-terminal of the distal ubiquitin with either a lysine or the N-terminus methionine of the proximal domain (Figure 1A)[23–29]. In particular, the heterogeneity of K63-Ub_2 conformational space is supported by the presence of numerous crystallographic structures of this protein, free or in complex with diverse targets, displaying different degrees of opening and arrangements of the two subunits[30–36]. Furthermore, studies based on different biophysical techniques, including SAXS, NMR, cross-linking and FRET, support the idea that K63-Ub_2 in solution populates a dynamic ensemble, with a preference for compact states[28,29,37]. This equilibrium between multiple states is considered critical in modulating the affinity of diubiquitin towards its biological partners[28].

In the following we present our SAXS-restrained all-atom metadynamics metainference (M&M)[1,38,39] simulation of K63-Ub_2, performed with the hybrid resolution approach, in comparison with an unrestrained reference simulation, in which the same setting was used but for the inclusion of experimental data. The conformational ensembles obtained from the two simulations (metainference and unrestrained), once assessed using independent experimental data available in literature from NMR paramagnetic relaxation enhancement (PRE) experiments[29], allow us to suggest that M&M SAXS restrained simulations can be efficiently employed to characterize the conformational ensemble of dynamic systems. All the methods described in this paper are freely available in the PLUMED-ISDB module[40] of the PLUMED library[41], further all the input files used are available on the PLUMED-NEST repository[42], as plumID:19.057.

2. THEORY AND METHODS

2.1 **Metainference**

Metainference allows integrating experimental data with prior information, generally represented by a molecular mechanic force field[21], taking into account the effect of conformational averaging and other sources of errors. In the case of Gaussian noise, the metainference energy can be written as[43]:

$$E_{MI} = E_{FF} + \frac{k_B T}{2} \sum_{i=1}^{N_d} \sum_{r=1}^{N_r} \frac{[d_i - \lambda f_i(\boldsymbol{X})]^2}{(\sigma_{r,i}^B)^2 + (\sigma_{r,i}^{SEM})^2} + E_\sigma,$$

where E_{FF} is the energy of the force field, k_B the Boltzmann constant, T the temperature, d_i the set of N_d experimental data, $f_i(\boldsymbol{X}) = \frac{1}{N_r} \sum_{r=1}^{N_r} f_i(X_{\boldsymbol{r}})$ is averaged over the N_r replica, $f_i(X_{\boldsymbol{r}})$ is the forward model used to predict observable i from conformation $X_{\boldsymbol{r}}$, $\sigma_{r,i}^B$ is an uncertainty parameter that describes random and systematic errors, $\sigma_{r,i}^{SEM}$ is the standard error of the mean

related to the conformational averaging and E_σ is an energy term that accounts for normalization of the data likelihood and error priors. Monte Carlo sampling is used to sample the uncertainty $\sigma_{r,i}^B$ and optionally a scaling parameter λ that relates experimental and back-calculated data (as in the case of SAXS experiment): these parameters are inferred during the simulation along with the model of the system. Importantly, if only one replica is considered, metainference becomes equivalent to the Inferential Structure Determination approach[44]; conversely, if $\sigma_{r,i}^B = 0$ (i.e. in absence of data and forward model errors) it is equivalent to the replica-averaged MaxEnt modelling[45].

Metainference can be combined with metadynamics (M&M) to accelerate the exploration of the conformational space[38,46]. In particular, it was proposed to apply it in combination with parallel bias metadynamics[47] (PBMetaD), which allows to use a larger number of collective variables (CVs) applying multiple low-dimensional bias potentials and therefore reducing the risk of missing slow degrees of freedom. In M&M multiple copies of simulation are run in parallel, where all the replica use the same conditions and force field and share the bias potential as in the case of multiple-walkers method[48]. The coupling of metainference and metadynamics is given by the calculation of the average forward model $f_i(\boldsymbol{X})$, where each replica contributes differently to the average with a weight $w(X_r)$ depending on the bias potential V_{PB} according to: $w(X_r) = e^{\frac{V_{PB}(CV(X_r),t)}{k_B T}}$.

2.2 **Coarse-grain approach to calculate scattering intensities**

Given a molecule of *N* atoms, the average scattering intensity for multiple copies randomly oriented can be computed via the Debye equation as:

$$I(q) = \sum_{i=1}^{N} \sum_{j=1}^{N} f_i(q)\, f_j(q) \frac{\sin(q r_{ij})}{q r_{ij}}. \quad (1)$$

Here r_{ij} indicates the distance between the atoms ij, $q = 4\pi \sin\vartheta\, /\lambda$ is the magnitude of the scattering vector, 2ϑ is the scattering angle and λ is the X-ray wave length. The form factor $f_i(q)$ of the i-th atom can be computed using the Cromer-Mann analytic function[49] and can be corrected using the Fraser approach[50] to account for the effect of the displaced solvent.

The calculation of scattering intensities with the Debye equation is extremely expensive from a computational perspective as it requires the evaluation of pairwise distances between all the atoms in the biomolecules, resulting in a problem of complexity $O(N^2)$. Among the strategies adopted to overcome this problem, it has been proposed to coarse-grain the scattering calculation[17,18]. According to this method, which is well justified by the low resolution of SAXS data, the molecule is represented as a collection of M beads, each comprising a variable number of atoms; herein the number of atoms per bead can be tuned to optimize the balance between accuracy and computational efficiency. If the form factors $F(q)$ of the beads are known the scattering intensities can then be computed as:

$$I(q) = \sum_{i=1}^{M} \sum_{j=1}^{M} F_i(q)\, F_j(q) \frac{\sin(q R_{ij})}{q R_{ij}}, \quad (2)$$

where R_{ij} indicates the distance between the center of mass of beads ij and with the sum running over the number of beads. The complexity is therefore reduced from $O(N^2)$ to $O(M^2)$. The form factors $F(q)$ for custom beads can be computed adopting the Single Bead Approximation averaging over multiple structures[17]. Alternatively, form factors for beads based on Martini force field[22] are available[18] and were previously implemented in PLUMED-ISDB module[20,40].

2.3 **Hybrid-resolution SAXS-driven metainference simulations**

SAXS experimental data can be used to improve MD simulations using the metainference approach. Importantly metainference is able to consider the ensemble-averaged nature of SAXS measurements; additionally, it can deal with both random and systematic errors in experimental data, as well as with the inaccuracies of the forward model (i.e. the calculation of SAXS intensities from the 3D-structures). This is particularly relevant in SAXS back-calculation when coarse-grain forward models are used and when the excess of electron density in the hydration shell is neglected. Recently, we implemented in PLUMED-ISDB a hybrid multi-resolution strategy to perform full atomistic MD simulations in which SAXS intensities, computed at a coarse-grain level based on Martini force field, are used as restraints within the metainference framework[20] (see Figure 1A). To achieve this goal the virtual positions of the Martini beads are computed on-the-fly and are used in combination with Martini form factors[18] for SAXS calculations. The computational efficiency of this strategy can be further improved using a multiple time-step protocol, where the metainference bias is applied only every few time steps[51]. In our previous work we demonstrated the reliability of the hybrid resolution approach for single-replica simulations in which two protein-nucleic acids complexes were refined against SAXS data.

Here we extended the described approach to multi-replica M&M simulations, with the aim to exhaustively explore the conformational space of flexible biomolecules, able to populate multiple conformational states.

2.4 **Computational details of the simulations**

K63-Ub_2, for which both SAXS and Paramagnetic Resonance Enhancement (PRE) experimental data are available[29,37], was used as test-system. As starting model for the simulations, we used

the chains B and C of PDB 2ZNV[34]: the K63R and D77 mutations in distal and proximal ubiquitin, respectively, were maintained to be coherent with SAXS measurements. MD simulations were performed with GROMACS 2018[52], PLUMED 2[41] and the PLUMED-ISDB[40] module, using the Amber ff03w force field[53] with TIP4P/2005 water model[54] and scaled protein-water Lennard Jones parameters[55]. The choice of this force field, that was specifically designed to increase molecules solvation, avoiding collapsed states and nonspecific protein-protein interactions, was guided by the fact that we expect an equilibrium between open and compact states of K63-Ub_2 with only transient inter-domain contacts. The system was solvated in a periodic dodecahedron box, initially 1.2 nm larger than the protein in each direction, and neutralized. After an initial energy minimization to a maximum force of 100 kJ/mol/nm, the solute was equilibrated under NVT condition at the temperature of 300 K for 50 ps using the Berendsen thermostat[56]; then Berendsen barostat was used to equilibrate the system in the NPT ensemble to the target pressure of 1 atm for 200 ps. The equilibration phase was followed by an initial plain-MD simulation of 100 ns, from which a pool of well-equilibrated conformations was extracted to be used as staring models for the subsequent runs. During the production runs in the NPT ensemble, the md integrator was employed with a time step of 2 fs; the temperature was maintained at 300 K using the Bussi thermostat[57] and the pressure was controlled with Parrinello-Rahman barostat[58]. Bonds were constrained with the LINCS algorithm[59], using a matrix expansion of the order of 6 and 2 iterations per step. Electrostatic was treated by using the particle mesh Ewald scheme[60] with a short-range cut-off of 0.9 nm and a Fourier grid spacing of 0.12 nm. The van der Waals interaction cut-off was set to 0.9 nm.

Two metadynamics multi-replica simulations were performed: 1) the metainference one, consisting of 18 replica, in which metainference was used to enforce the agreement with SAXS

data and 2) the reference one, consisting of 8 replica, in which the same settings of simulation (1) were used but without the inclusion of experimental restraints. PBMetaD was performed in combination with well-tempered metadynamics[61] and the multiple-walker scheme[48], where Gaussians with an initial height of 0.6 kJ/mol were deposited every 1 ps using a bias factor of 22. Four CVs were biased: two of them (hydContacts and polContacts) counts the number of the hydrophobic and polar contacts between the two ubiquitin domains, the other two (TICAcv1 and TICAcv2) are the results of the linear combination of numerous angles as determined by a Time-lagged Independent Component Analysis[62] (TICA) performed on the initial 100 ns plain MD simulation (see Supplementary Information for more details). The width of the Gaussians was determined with the dynamically-adapted gaussian approach[63], using a time window of 2 ps to estimate CVs fluctuations and setting as minimum values for the width 0.01, 0.05, 0.01 and 0.01 for hydContacts, polContacts, TICAcv1 and TICAcv2, respectively.

Experimental SAXS intensities for K63-Ub_2 are available in the SASDCG7[37] entry of the SASDB database[64]. For the metainference simulation, a set of 15 representative SAXS intensities at different scattering vectors, ranging between 0.06 $Å^{-1}$ and 0.20 $Å^{-1}$ and equally spaced, were included as restraints. These representative intensities were extracted from the experimental data, where a 21-point running average was performed to reduce the influence of experimental noise. Metainference was applied every 10 steps, using a single Gaussian noise per data-point and sampling a scaling factor between experimental and calculated SAXS intensities with a flat prior between 0.5 and 1.5.

For both simulations, each replica was evolved for 350 ns, resulting in a total simulation time of 6.3 μs for the metainference simulation and of 2.8 μs for the reference one. Convergence was checked using the block analysis procedure, in which free-energy profiles are computed over

different blocks of simulations and lastly, the weighted average error along the free energy profile is computed as a function of the block length (Figure S1).

3. RESULTS AND DISCUSSION

The hybrid AA-CG approach was applied to characterize the conformational ensemble of K63-Ub_2. Specifically, we performed: (1) a multi-replica M&M simulation, integrating the available experimental SAXS data with the coarse-grain approach every 10 steps; (2) an unrestrained reference simulation, adopting the same setting of the metainference run but without the inclusion of experimental data to drive the sampling. This second simulation is used for comparison and allows a deeper understanding of the role played by SAXS restraints.

3.1 The hybrid AA-CG approach is computationally efficient.

In **Figure 1B** we compared the performances of a plain atomistic MD simulation (yellow) and of all-atom metainference simulations where SAXS restraints were included every step with atomistic forward model (green) (that is Eq. 1 in theory and methods), every step (purple) or every 10 steps with CG forward model based on Martini force field (blue) (that is Eq. 2 in theory and methods). Of note, the use of the hybrid AA-CG approach can significantly improve the performances of SAXS-driven MD simulations if compared to the ones adopting atomistic scattering evaluation. This gain can be further increased using a multiple time-step protocol (**Figure 1B**, blue line), in which the metainference bias is applied every few time steps. This strategy is well justified in the case of SAXS data, which are characterized by slow temporal fluctuations, and allows to approach the performances of plain MD simulations.

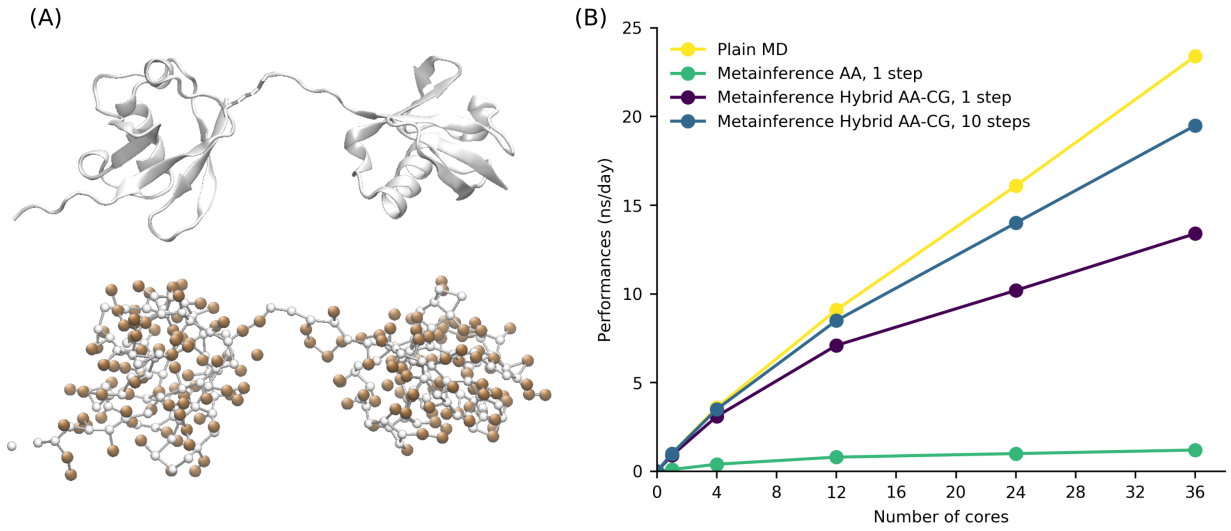

Figure 1. (A) K63-Ub_2 (2473 atoms) represented with cartoon (top) or highlighting the centers of the 328 Martini beads (bottom), colored in white and orange for backbone and sidechain, respectively. (B) Performances, as a function of number of cores, for plain MD and for metainference simulations (with all-atom or coarse-grain forward model, computed every 1 or 10 steps). Performances are estimated on Intel Xeon E5-2697 2.30 GHz for a single replica of K63-Ub_2 in water.

3.2 **Monitoring Metainference simulation.**

To evaluate on-the-fly the effectiveness of SAXS restraints in the metainference simulation, we monitored the correlation between back-calculated and experimental data as a function of the simulation time (**Figure 2 and S2**). Both the simulations showed a high correlation between experiments and forward model, as a consequence of the intrinsic low resolution of SAXS data. However, the comparison revealed a better agreement in the metainference than in the unrestrained simulation (**Figure 2**), confirming the efficacy of our restraints. This is supported

by other statistical properties, including the sum of square deviation and the slope/intercept of the linear fit (**Figure S2**).

We also monitored the intensity of experimental restraints, which depends on the square sum of the uncertainty parameters $\sigma_{r,i}^{B}$ and $\sigma_{r,i}^{SEM}$. To this aim we computed the distribution across the metainference ensemble of both $\sigma_{r,i}^{B}$, which is associated with experimental and forward model inaccuracies, and $\sigma_{r,i}^{SEM}$, i.e. the standard error of the mean over the replicas. We observed a broader distribution of the sampled parameter $\sigma_{r,i}^{B}$ with respect to $\sigma_{r,i}^{SEM}$(**Figure S3**), with greater uncertainties associated to smaller scattering angles (where indeed the global conformation mostly influences SAXS profiles). The values of $\sigma_{r,i}^{SEM}$are always within the range sampled by $\sigma_{r,i}^{B}$, indicating that the two sources of error comparably contribute to the restraint weight and suggesting that the number of replicas (which concurs in determining the magnitude of $\sigma_{r,i}^{SEM}$) is sufficient.

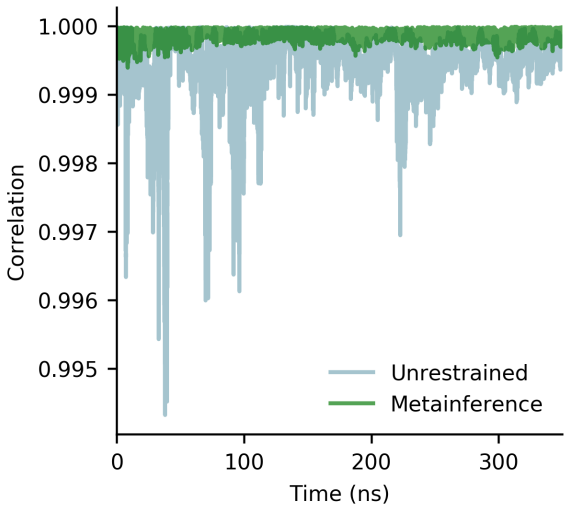

Figure 2. Correlation, as a function of the simulation time, between experimental and back-calculated SAXS intensities. The intensities considered are the ones used as restrained in the metainference simulation. The back-calculated data are averaged over the replica and computed with the non-approximated atomistic Debye equation (1).

We finally computed the χ^2 value between the experimental and back-calculated SAXS data, considering the whole conformational ensembles sampled within the unrestrained or the metainference simulations. To this aim, we used as scaling factor λ, relating experimental and calculated data, the one sampled via Monte Carlo in our metainference run, which converged to a value of 0.984 (**Figure S4**). The computed χ^2 values, 1.25 for the metainference and 1.82 for the unrestrained run, respectively, confirm that the metainference ensemble provides a better match with SAXS experimental data, even if the difference appears to be not that striking.

3.3 **Comparison of the conformational ensemble obtained from SAXS-driven and unrestrained simulations.**

Despite the relatively small improvement in the agreement with SAXS, the introduction of SAXS restraints has a relevant influence on the resulting ensembles. While the dynamic of the single ubiquitin units is only negligibly affected by the restraints, as testified by the comparison of the RMSD distribution of the single domains in metainference and unrestrained simulations (**Figure S5**), vice-versa, the relative positioning of the two ubiquitin units is strongly influenced by the introduction of SAXS data. Indeed, we observed that the probability density function of the gyration radius shows a wider distribution in the metainference simulation, according to which K63-Ub_2 exists in a dynamic ensemble comprising both extended and compact conformations (**Figure 3 and S6)**. Here, it is worth to observe that introducing experimental

restraints in MD does not result in a reduction of the conformational space sampled, as it could be naively guessed; on the opposite, it could help in broadening the conformational space visited, encouraging the exploration of regions that could be, otherwise, under-sampled.

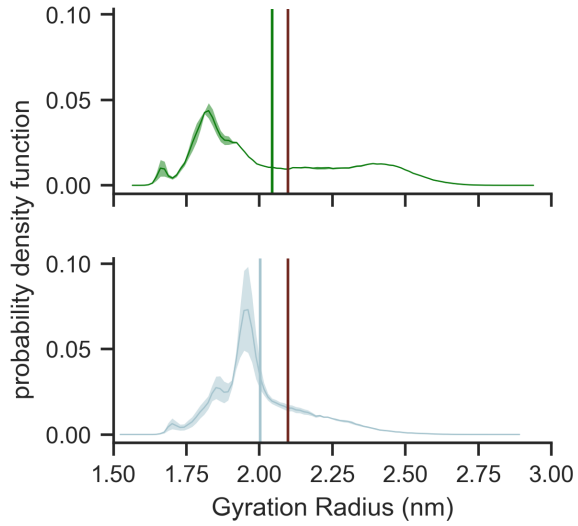

Figure 3. Distribution of the gyration radius in the metainference (green) and in the unrestrained (light blue) conformational ensembles. The vertical bars indicate the average back-calculated (green or light blue) and the experimental (red) gyration radius. The shade indicates the standard deviation, computed comparing the first and second halves of simulations.

To have a better description of the generated conformational ensembles, we reconstructed the 2d-free energy landscapes in a space defined by the Cα-gyration radius and a global dihedral angle θ (used also as metadynamics CVs, see **Figure S7**), that describes the relative orientation of the two ubiquitin domains. Interestingly, the coordinates in this space of the available K63-Ub_2 PDB structures mostly fall in regions characterized by low free-energy according to metainference simulation, supporting the reliability of the sampled conformation (**Figure 4**). Conversely, regions compatible with the extended structures, adopted by all the crystallized free K63-Ub_2,

are insufficiently sampled within the unrestrained simulation. A comparison with Förster Resonance Energy Transfer data available in literature[28,37] confirms that the unrestrained simulation over-sampled compact conformations, missing the extended ones (Supplementary Text **S2** and Figure **S8**).

The inspection of the 2d-fes (**Figure 4**) also revealed that the relative positioning of the two domains is different in the two ensembles, especially when focusing on compact conformations. This suggests that, in the two simulations, different Ub-Ub interfaces are preferred. The contact map analysis (**Figure S9**) confirmed that diverse groups of residues are involved in the inter-subunit interactions in the two cases, where the major differences relies in the residues of distal ubiquitin (residues 42-49 according to metainference, residues 8-12 according to the unrestrained simulation). Based on this observation, it would be tempting to hypothesize that SAXS restraints could help in identifying reliable protein-protein interfaces. To confirm this hypothesis, we proceeded by validating our conformational ensemble against Paramagnetic Relaxation Enhancement (PRE) data.

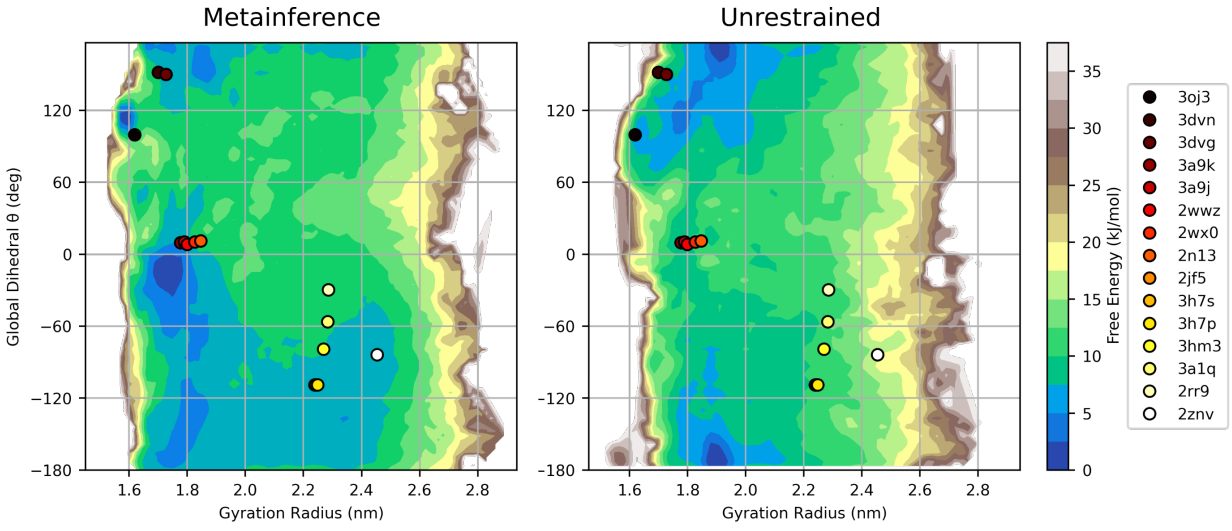

Figure 4. 2D-free energy surface for K63-Ub_2, derived by the metainference (left panel) and the unrestrained (right panel) ensembles, as a function of the Cα-gyration radius and the global dihedral angle θ (see Supplementary **Text S1 and Figure S7**). The coordinates of the available PDB structures in this space are plotted with points (colored from red to white, from more compact to extended conformations). To make the Cα-gyration radius comparable with the one back-calculated from the PDB structures, only residues 1-72 of the two ubiquitin domains were considered.

3.4 **Validation and analysis of the Ub-Ub interfaces.**

PRE experiments from NMR are particularly suited to provide information about inter subunit distances in multi domain proteins. In these experiments, after conjugation of a specific residue with a paramagnetic probe, PRE can be measured for the other domain, where PRE values are proportional to the inverse sixth power of the distance between the paramagnetic centre and the nuclei. Due to this functional form, PRE data are extremely sensible to closed states even if poorly populated[65]. Therefore, comparison of conformational ensemble against PRE is particularly indicated to validate the Ub-Ub interfaces of the compact states and their relative population.

Liu and co-workers previously acquired inter-subunit PRE data for K63-Ub_2, conjugating the paramagnetic probe on residues N25 or K48 of the distal ubiquitin, after N25C/K48C mutations, and detecting many large PRE for some residues of the proximal unit[29,66]. We back-calculated the same PRE values from our metainference and unrestrained conformational ensembles, approximating the paramagnetic centre–nuclei distances with the distances between the Cβ atom of N25 or K48 and all the amide hydrogens of the proximal ubiquitin. As this approximation is

reasonably quite strong, we evaluated an error of ± 3Å on the estimation of these distances, which finally gave us an estimation of the minimum/maximum PRE values. We observed that experimental N25-PRE are in good agreement with the ones calculated from the metainference ensemble (**Figure 5, upper-left panel**), suggesting that the compact interfaces are correctly sampled in our metainference run. Conversely, the unrestrained ensemble fails to reproduce N25-PRE for the proximal unit residues 8-14 and misses the high PRE observed for residues 46-47 (**Figure 5, lower-left panel**). On the opposite, both the metainference and the unrestrained ensembles totally failed in reproducing K48-PRE (**Figure 5, right panels**), showing high PRE in regions not detected by experiments (e.g. proximal residues 18-26) and vice-versa (e.g. residues 45-49). As the comparison with N25-PRE gave good indication about the reliability of our metainference ensemble in sampling correct Ub-Ub interfaces, we investigated the reasons underlying the bad agreement with K48-PRE by analysing the energetic contributions of each residue to the interface formation. We found that, according to metainference ensemble, K48 of proximal ubiquitin is important in stabilizing electrostatic interactions at the interface (**Fig 6A**). Based on this, we hypothesize that the introduction of the paramagnetic probe at this site, along with the K48C mutation, could strongly destabilize relevant inter-domains contacts, leading to a rearrangement of the positioning of the two subunits and finally giving rise to PRE values that could not be comparable with the one of wt-K63Ub2. Importantly, we verified that this is not the case for N25C, where neither Coulomb nor Lennard-Jones interactions seem to play a major role in stabilizing the Ub-Ub interfaces (**Figure 6A and S10**).

In order to have a deeper insight into the sampled Ub-Ub interfaces, we analyzed the conformational minima identified by our metainference run. Based on the 2d free-energy of **Figure 4**, we classified the closed states (defined as the conformers with Cα-gyration radius<2.1

nm) in 4 different regions: region A, with -100º<θ<=60º, region B, with 60º<θ<=140º, region C, with θ<=-100º or θ>140º and gyration radius <=1.82 nm, and region D, with θ<=-100º or θ>140º and gyration radius >1.82 nm (see **Figure S11**). Region A, B, C and D have a population of ~33%, ~8%, ~12% and ~14%, together accounting for the 67% of the conformational space, whereas the remaining 33% is populated by extended conformations (gyration radius >2.1 nm). The two minima associated with regions A and B present well-defined Ub-Ub interfaces (**Figure 6 B-C and Figure S12);** on the opposite, the broad regions C and D contain quite heterogenous conformational states (**Figure S12**), suggesting that additional variables could be useful to subdivide the compact conformational space in even smaller minima and supporting the idea that many different possible interfaces can be transiently populated by K63-Ub_2. Inspection of both the structures of minima A and B and of the related energy matrices (**Figure S12**) allowed us to characterize more in detail these interfaces. In minimum A we observed many charged residues involved (**Figure 6 B**): next to the K63-linker region, the positively charged R72-R42 of distal ubiquitin contact E16-E18 of the proximal domain; further distal K48 engages both polar interactions with proximal charged residues D21-E24-E52, as well as stabilizing hydrophobic contacts with T22 and/or T55. In minimum B (**Figure 6 C**) R72-R42 of distal ubiquitin play again an important role by contacting the negatively charged E64 of the proximal unit; further interactions involve the aliphatic side-chains of distal residues E24-N25 and the hydrophobic proximal residues F45-A46.

Overall, our analysis revealed the involvement of many charged residues in the Ub-Ub interface and suggests that K63-Ub_2 prefer electrostatic interfacial contacts, being hindered by steric constraints to interact via the common I44/I36 hydrophobic patches, in line with previous reports[67]. Our results are in agreement with previous mutagenesis experiments concerning the E64 residue

of the proximal unit, which play a major role in minimum B interface. Indeed, it was reported that E64 is important for the stabilization of closed conformations, where an E64R mutation was shown to decrease the binding affinity toward ligands, known to bind the K63-Ub_2 closed states, via an entropically-driven mechanism. Herein, our results support the conformational selection mechanism proposed by Liu and coworkers[29] for K63-Ub_2 ligand recognition.

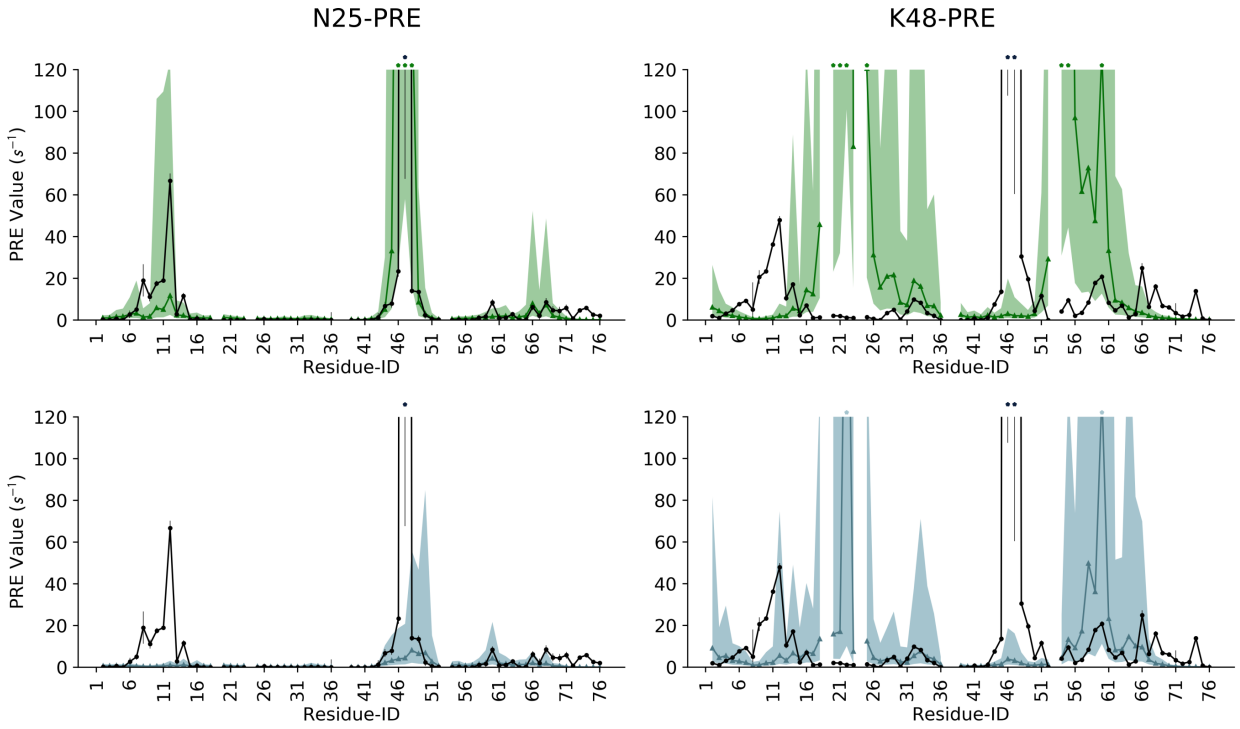

Figure 5. Comparison of experimental (black line) and back-calculated inter-subunit PRE for the residues of K63-Ub_2 proximal ubiquitin, with the paramagnetic probe conjugated at N25C (left panels) or K48C (right panels) of the distal ubiquitin. The area between the minimum/maximum back-calculated PRE values, considering a $\pm$3Å error on the estimation of probe-nuclei distances, is coloured with green or light-blue shades, for metainference and unrestrained ensembles, respectively. The respective back-calculated PRE, without distance correction, are

shown with green and light blue lines. PRE values higher than 120 s^{-1} are indicated with a star on the top of the graph.

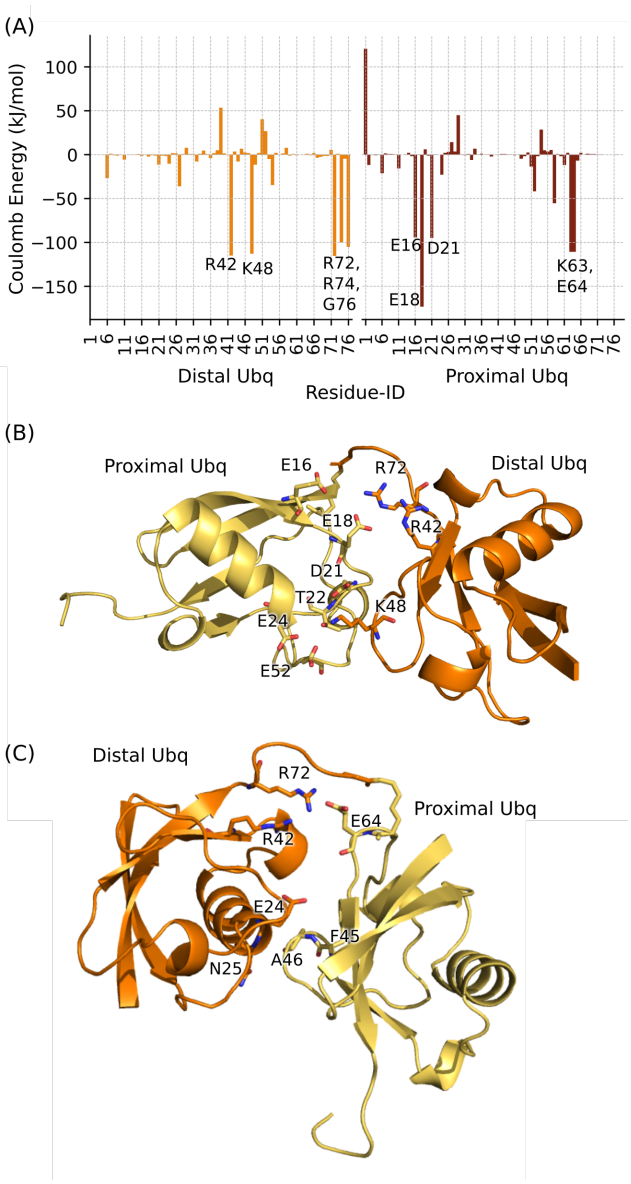

Figure 6. (A) Per-residue Coulomb energy obtained summing over the residue-residue energetic contributions for pairs of residues belonging to the two different Ub domains. Residues of distal and proximal ubiquitin are colored in orange and red, respectively; lowest energy peaks are labelled. (B,C) Representative conformations extracted from minima A and B, respectively. Relevant residues for the interface formation are highlighted in sticks.

4. CONCLUSIONS

In this work we presented a hybrid-resolution MD-based strategy, useful to determine structural ensembles providing an accurate interpretation of SAXS data. The hybrid AA-CG approach, in which SAXS intensities are computed at a coarse-grain level while simulations are run with atomistic force field, makes the SAXS-driven simulations feasible in terms of computational efficiency without losing atomistic details. Of notice, the proposed method is able to deal with highly flexible systems, aiding in the estimation of the population of the different existing conformational states.

To prove the efficacy of the method, it has been here applied to study the conformational ensemble of the multidomain protein K63-Ub_2. Our results reveal that the inclusion of SAXS restraints significantly influences the relative positioning of the different sub-units, without altering the dynamics of the single domain. The effectiveness of the approach in improving the reliability of the conformational sampling is supported by multiple indirect validations as well as by quantitative comparison with independent experimental data (e.g. PRE from NMR).

Given the efficacy of the hybrid resolution method, we propose that it could be used as a strategy to provide accurate interpretation of scattering data for multi-domains and/or disorder proteins. Importantly, the approach is implemented in the freely available PLUMED software and all the data and PLUMED input files required to reproduce the results reported in this paper are available on the PLUMED-NEST repository (www.plumed-nest.org)[42], as plumID:19.057.

ASSOCIATED CONTENT

The following files are available free of charge.

Supporting Information (PDF). Supplementary Text 1: description of metadynamics Collective Variables. Supplementary Text 2: Comparison with FRET data. Figures: 1. Statistical parameters to monitor metainference simulation; 2. Scaling factor as a function of time; 3. Probability distribution of the uncertainty parameters; 4-5. RMSD and minimum distance probability distribution of the single ubiquitin domains; 6. Graphical representation of the global dihedral angle θ; 7. FRET profiles; 8. Contact map of the metainference and unrestrained ensemble; 9. Per-residue Lennard-Jones interface contribution; 10. Region definition in the 2d-fes; 11. Energy matrices and conformational ensembles for the metainference minima.

AUTHOR INFORMATION

Corresponding Author

* cristina.paissoni@unimi.it, carlo.camilloni@unimi.it

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

Any funds used to support the research of the manuscript should be placed here (per journal style).

Notes

Any additional relevant notes should be placed here.

ACKNOWLEDGMENT

We acknowledge CINECA for an award under the ISCRA initiative, for the availability of high-performance computing resources and support.

ABBREVIATIONS

K63-Ub_2, K63-linked diubiquitin; MD, Molecular Dynamics; M&M, metadynamics metainference; PBMetaD, parallel bias metadynamics; CV, collective variable; AA, all-atom; CG, coarse-grain; SAXS, small angle X-ray scattering; PRE, paramagnetic relaxation enhancement; NMR, nuclear magnetic resonance; FRET, Förster resonance energy transfer; TICA,Time-lagged Independent Component Analysis.

REFERENCES

(1) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. Principles of Protein Structural Ensemble Determination. *Current Opinion in Structural Biology*. Elsevier Current Trends February 1, 2017, pp 106–116. https://doi.org/10.1016/j.sbi.2016.12.004.

(2) Wei, G.; Xi, W.; Nussinov, R.; Ma, B. Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? The Diverse Functional Roles of Conformational Ensembles in the Cell. *Chemical Reviews*. American Chemical Society June 8, 2016, pp 6516–6551. https://doi.org/10.1021/acs.chemrev.5b00562.

(3) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature*. December 12, 2007, pp 964–972. https://doi.org/10.1038/nature06522.

(4) Papaleo, E.; Saladino, G.; Lambrughi, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **2016**, *116* (11), 6391–6423. https://doi.org/10.1021/acs.chemrev.5b00623.

(5) Kikhney, A. G.; Svergun, D. I. A Practical Guide to Small Angle X-Ray Scattering (SAXS) of Flexible and Intrinsically Disordered Proteins. *FEBS Lett.* **2015**, *589* (19 Pt A), 2570–2577. https://doi.org/10.1016/j.febslet.2015.08.027.

(6) Hub, J. S. Interpreting Solution X-Ray Scattering Data Using Molecular Simulations. *Current Opinion in Structural Biology*. April 2018, pp 18–26. https://doi.org/10.1016/j.sbi.2017.11.002.

(7) Shevchuk, R.; Hub, J. S. Bayesian Refinement of Protein Structures and Ensembles against SAXS Data Using Molecular Dynamics. *PLOS Comput. Biol.* **2017**, *13* (10), e1005800. https://doi.org/10.1371/journal.pcbi.1005800.

(8) Antonov, L. D.; Olsson, S.; Boomsma, W.; Hamelryck, T. Bayesian Inference of Protein

Ensembles from SAXS Data. *Phys. Chem. Chem. Phys.* **2016**, *18* (8), 5832–5838. https://doi.org/10.1039/c5cp04886a.

(9) Bowerman, S.; Rana, A. S. J. B.; Rice, A.; Pham, G. H.; Strieter, E. R.; Wereszczynski, J. Determining Atomistic SAXS Models of Tri-Ubiquitin Chains from Bayesian Analysis of Accelerated Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2017**, *13* (6), 2418–2429. https://doi.org/10.1021/acs.jctc.7b00059.

(10) Cheng, P.; Peng, J.; Zhang, Z. SAXS-Oriented Ensemble Refinement of Flexible Biomolecules. *Biophys. J.* **2017**, *112* (7), 1295–1301. https://doi.org/10.1016/j.bpj.2017.02.024.

(11) Róycki, B.; Kim, Y. C.; Hummer, G. SAXS Ensemble Refinement of ESCRT-III CHMP3 Conformational Transitions. *Structure* **2011**, *19* (1), 109–116. https://doi.org/10.1016/j.str.2010.10.006.

(12) Hummer, G.; Köfinger, J. Bayesian Ensemble Refinement by Replica Simulations and Reweighting. *J. Chem. Phys.* **2015**, *143* (24). https://doi.org/10.1063/1.4937786.

(13) Björling, A.; Niebling, S.; Marcellini, M.; Van Der Spoel, D.; Westenhoff, S. Deciphering Solution Scattering Data with Experimentally Guided Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2015**, *11* (2), 780–787. https://doi.org/10.1021/ct5009735.

(14) Chen, P. C.; Hub, J. S. Interpretation of Solution X-Ray Scattering by Explicit-Solvent Molecular Dynamics. *Biophys. J.* **2015**, *108* (10), 2573–2584. https://doi.org/10.1016/j.bpj.2015.03.062.

(15) Kimanius, D.; Pettersson, I.; Schluckebier, G.; Lindahl, E.; Andersson, M. SAXS-Guided Metadynamics. *J. Chem. Theory Comput.* **2015**, *11* (7), 3491–3498. https://doi.org/10.1021/acs.jctc.5b00299.

(16) Svergun, D.; Barberato, C.; Koch, M. H. CRYSOL - A Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28* (6), 768–773. https://doi.org/10.1107/S0021889895007047.

(17) Yang, S.; Park, S.; Makowski, L.; Roux, B. A Rapid Coarse Residue-Based Computational Method for X-Ray Solution Scattering Characterization of Protein Folds and Multiple Conformational States of Large Protein Complexes. *Biophys. J.* **2009**, *96* (11), 4449–4463. https://doi.org/10.1016/j.bpj.2009.03.036.

(18) Niebling, S.; Björling, A.; Westenhoff, S. MARTINI Bead Form Factors for the Analysis of Time-Resolved X-Ray Scattering of Proteins. *J. Appl. Crystallogr.* **2014**, *47* (4), 1190–1198. https://doi.org/10.1107/S1600576714009959.

(19) Berlin, K.; Gumerov, N. A.; Fushman, D.; Duraiswami, R. Hierarchical {\it O}({\it N}) Computation of Small-Angle Scattering Profiles and Their Associated Derivatives. *J. Appl. Crystallogr.* **2014**, *47* (2), 755–761. https://doi.org/10.1107/S1600576714004671.

(20) Paissoni, C.; Jussupow, A.; Camilloni, C. Martini Bead Form Factors for Nucleic Acids and Their Application in the Refinement of Protein–Nucleic Acid Complexes against SAXS Data. *J. Appl. Crystallogr.* **2019**, *52* (2), 394–402. https://doi.org/10.1107/s1600576719002450.

(21) Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M. Metainference: A Bayesian Inference Method for Heterogeneous Systems. *Sci. Adv.* **2016**, *2* (1), e1501177. https://doi.org/10.1126/sciadv.1501177.

(22) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini Model. *Chem. Soc. Rev.* **2013**, *42* (16), 6801–6822. https://doi.org/10.1039/c3cs60093a.

(23) Komander, D.; Rape, M. The Ubiquitin Code. *Annu. Rev. Biochem.* **2012**, *81* (1), 203–

229. https://doi.org/10.1146/annurev-biochem-060310-170328.

(24) Vincendeau, M.; Hadian, K.; Messias, A. C.; Brenke, J. K.; Halander, J.; Griesbach, R.; Greczmiel, U.; Bertossi, A.; Stehle, R.; Nagel, D.; et al. Inhibition of Canonical NF-κ B Signaling by a Small Molecule Targeting NEMO-Ubiquitin Interaction. *Sci. Rep.* **2016**, *6* (1), 18934. https://doi.org/10.1038/srep18934.

(25) Castañeda, C. A.; Chaturvedi, A.; Camara, C. M.; Curtis, J. E.; Krueger, S.; Fushman, D. Linkage-Specific Conformational Ensembles of Non-Canonical Polyubiquitin Chains. *Phys. Chem. Chem. Phys.* **2016**, *18* (8), 5771–5788. https://doi.org/10.1039/c5cp04601g.

(26) Castañeda, C. A.; Kashyap, T. R.; Nakasone, M. A.; Krueger, S.; Fushman, D. Unique Structural, Dynamical, and Functional Properties of K11-Linked Polyubiquitin Chains. *Structure* **2013**, *21* (7), 1168–1181. https://doi.org/10.1016/j.str.2013.04.029.

(27) Berlin, K.; Castañeda, C. A.; Schneidman-Duhovny, D.; Sali, A.; Nava-Tudela, A.; Fushman, D. Recovering a Representative Conformational Ensemble from Underdetermined Macromolecular Structural Data. *J. Am. Chem. Soc.* **2013**, *135* (44), 16595–16609. https://doi.org/10.1021/ja4083717.

(28) Ye, Y.; Blaser, G.; Horrocks, M. H.; Ruedas-Rama, M. J.; Ibrahim, S.; Zhukov, A. A.; Orte, A.; Klenerman, D.; Jackson, S. E.; Komander, D. Ubiquitin Chain Conformation Regulates Recognition and Activity of Interacting Proteins. *Nature* **2012**, *492* (7428), 266–270. https://doi.org/10.1038/nature11722.

(29) Liu, Z.; Gong, Z.; Jiang, W.-X.; Yang, J.; Zhu, W.-K.; Guo, D.-C.; Zhang, W.-P.; Liu, M.-L.; Tang, C. Lys63-Linked Ubiquitin Chain Adopts Multiple Conformational States for Specific Target Recognition. *Elife* **2015**, *4*. https://doi.org/10.7554/elife.05767.

(30) Kulathu, Y.; Akutsu, M.; Bremm, A.; Hofmann, K.; Komander, D. Two-Sided Ubiquitin

Binding Explains Specificity of the TAB2 NZF Domain. *Nat. Struct. Mol. Biol.* **2009**, *16* (12), 1328–1330. https://doi.org/10.1038/nsmb.1731.

(31) Komander, D.; Reyes-Turcu, F.; Licchesi, J. D. F.; Odenwaelder, P.; Wilkinson, K. D.; Barford, D. Molecular Discrimination of Structurally Equivalent Lys 63-Linked and Linear Polyubiquitin Chains. *EMBO Rep.* **2009**, *10* (5), 466–473. https://doi.org/10.1038/embor.2009.55.

(32) Weeks, S. D.; Grasty, K. C.; Hernandez-Cuebas, L.; Loll, P. J. Crystal Structures of Lys-63-Linked Tri- and Di-Ubiquitin Reveal a Highly Extended Chain Architecture. *Proteins Struct. Funct. Bioinforma.* **2009**, *77* (4), 753–759. https://doi.org/10.1002/prot.22568.

(33) Sato, Y.; Yoshikawa, A.; Yamashita, M.; Yamagata, A.; Fukai, S. Structural Basis for Specific Recognition of Lys 63-Linked Polyubiquitin Chains by NZF Domains of TAB2 and TAB3. *EMBO J.* **2009**, *28* (24), 3903–3909. https://doi.org/10.1038/emboj.2009.345.

(34) Sato, Y.; Yoshikawa, A.; Yamagata, A.; Mimura, H.; Yamashita, M.; Ookata, K.; Nureki, O.; Iwai, K.; Komada, M.; Fukai, S. Structural Basis for Specific Cleavage of Lys 63-Linked Polyubiquitin Chains. *Nature* **2008**, *455* (7211), 358–362. https://doi.org/10.1038/nature07254.

(35) Sato, Y.; Yoshikawa, A.; Mimura, H.; Yamashita, M.; Yamagata, A.; Fukai, S. Structural Basis for Specific Recognition of Lys 63-Linked Polyubiquitin Chains by Tandem UIMs of RAP80. *EMBO J.* **2009**, *28* (16), 2461–2468. https://doi.org/10.1038/emboj.2009.160.

(36) Newton, K.; Matsumoto, M. L.; Wertz, I. E.; Kirkpatrick, D. S.; Lill, J. R.; Tan, J.; Dugger, D.; Gordon, N.; Sidhu, S. S.; Fellouse, F. A.; et al. Ubiquitin Chain Editing Revealed by Polyubiquitin Linkage-Specific Antibodies. *Cell* **2008**, *134* (4), 668–678. https://doi.org/10.1016/j.cell.2008.07.039.

(37) Liu, Z.; Gong, Z.; Cao, Y.; Ding, Y. H.; Dong, M. Q.; Lu, Y. B.; Zhang, W. P.; Tang, C. Characterizing Protein Dynamics with Integrative Use of Bulk and Single-Molecule Techniques. *Biochemistry* **2018**, *57* (3), 305–313. https://doi.org/10.1021/acs.biochem.7b00817.

(38) Bonomi, M.; Camilloni, C.; Vendruscolo, M. Metadynamic Metainference: Enhanced Sampling of the Metainference Ensemble Using Metadynamics. *Sci. Rep.* **2016**, *6* (1), 31232. https://doi.org/10.1038/srep31232.

(39) Camilloni, C.; Pietrucci, F. Advanced Simulation Techniques for the Thermodynamic and Kinetic Characterization of Biological Systems. *Advances in Physics: X*. Taylor & Francis January 15, 2018, pp 885–916. https://doi.org/10.1080/23746149.2018.1477531.

(40) Bonomi, M.; Camilloni, C. Integrative Structural and Dynamical Biology with PLUMED-ISDB. *Bioinformatics* **2017**, *33* (24), 3999–4000.

(41) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613. https://doi.org/https://doi.org/10.1016/j.cpc.2013.09.018.

(42) The PLUMED Consortium. A Community Effort to Promote Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat. Methods* **2019**.

(43) Löhr, T.; Jussupow, A.; Camilloni, C. Metadynamic Metainference: Convergence towards Force Field Independent Structural Ensembles of a Disordered Peptide. *J. Chem. Phys.* **2017**, *146* (16), 165102. https://doi.org/10.1063/1.4981211.

(44) Rieping, W.; Habeck, M.; Nilges, M. Inferential Structure Determination. *Science (80-.).* **2005**, *309* (5732), 303–306. https://doi.org/10.1126/science.1110428.

(45) Cavalli, A.; Camilloni, C.; Vendruscolo, M. Molecular Dynamics Simulations with

Replica-Averaged Structural Restraints Generate Structural Ensembles According to the Maximum Entropy Principle. *J. Chem. Phys.* **2013**, *138* (9), 094112. https://doi.org/10.1063/1.4793625.

(46) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562–12566. https://doi.org/10.1073/pnas.202427399.

(47) Pfaendtner, J.; Bonomi, M. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *J. Chem. Theory Comput.* **2015**, *11* (11), 5062–5067. https://doi.org/10.1021/acs.jctc.5b00846.

(48) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B* **2006**, *110* (8), 3533–3539. https://doi.org/10.1021/jp054359r.

(49) Cromer, D. T.; Waber, J. T. Scattering Factors Computed from Relativistic Dirac–Slater Wave Functions. *Acta Crystallogr.* **1965**, *18* (1), 104–109. https://doi.org/10.1107/S0365110X6500018X.

(50) Fraser, R. D. B.; MacRae, T. P.; Suzuki, E. An Improved Method for Calculating the Contribution of Solvent to the X-Ray Diffraction Pattern of Biological Molecules. *J. Appl. Crystallogr.* **1978**, *11* (6), 693–694. https://doi.org/10.1107/S0021889878014296.

(51) Ferrarotti, M. J.; Bottaro, S.; Pérez-Villa, A.; Bussi, G. Accurate Multiple Time Step in Biased Molecular Simulations. *J. Chem. Theory Comput.* **2015**, *11* (1), 139–146. https://doi.org/10.1021/ct5007086.

(52) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

https://doi.org/https://doi.org/10.1016/j.softx.2015.06.001.

(53) Best, R. B.; Mittal, J. Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *J. Phys. Chem. B* **2010**, *114* (46), 14916–14923. https://doi.org/10.1021/jp108618d.

(54) Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123* (23), 154707. https://doi.org/10.1063/1.2121687.

(55) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10* (11), 5113–5124. https://doi.org/10.1021/ct500569b.

(56) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

(57) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 14101. https://doi.org/10.1063/1.2408420.

(58) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.

(59) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1998**, *18* (12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

(60) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

https://doi.org/10.1063/1.464397.

(61) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2). https://doi.org/10.1103/PhysRevLett.100.020603.

(62) McCarty, J.; Parrinello, M. A Variational Conformational Dynamics Approach to the Selection of Collective Variables in Metadynamics. *J. Chem. Phys.* **2017**, *147* (20), 204109. https://doi.org/10.1063/1.4998598.

(63) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, *8* (7), 2247–2254. https://doi.org/10.1021/ct3002464.

(64) Valentini, E.; Kikhney, A. G.; Previtali, G.; Jeffries, C. M.; Svergun, D. I. SASBDB, a Repository for Biological Small-Angle Scattering Data. *Nucleic Acids Res.* **2015**, *43* (D1), D357–D363.

(65) Marius Clore, G.; Iwahara, J. Theory, Practice, and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chem. Rev.* **2009**, *109* (9), 4108–4139. https://doi.org/10.1021/cr900033p.

(66) Liu, Z.; Tang, C. Ensemble Structure Description of Lys63-Linked Diubiquitin. *Data Br.* **2016**, *7*, 81–88. https://doi.org/10.1016/j.dib.2016.02.003.

(67) Wagner, N. D.; Russell, D. H. Defining Noncovalent Ubiquitin Homodimer Interfacial Interactions through Comparisons with Covalently Linked Diubiquitin. *J. Am. Chem. Soc.* **2016**, *138* (51), 16588–16591. https://doi.org/10.1021/jacs.6b09829.