Supplementary Methods

Chemical cross-links detected by mass spectrometry have been used to generate distance restraints, which help in determining the structure of soluble proteins and protein complexes. The output of the method is a list of pairs of residue positions that are sufficiently close to be cross-linked. The approach described below attempts to overcome several challenges. First, the heterogeneity of the Δ131Δ samples. Second, the use of quantitative information about cross-linked encoded in the ion intensity ratios between Hsp90 bound and unbound states. Third, the dependence of the cross-linked fraction on the structure. Fourth, the errors in the identification and quantitation of cross-links.

Here, we build on the Inferential Structure Determination (ISD), a Bayesian framework that provides an objective way to interpret experimental data and integrate it with prior knowledge[1]. We generalized the original ISD approach allowing us to disentangle the effect of structural heterogeneity on the data from the measurement noise. As a result, we could compute multiple structural states of the Hsp90-bound and unbound Δ131Δ, as well as their populations, and the uncertainty of each cross-linked ratio.

Theory. The Bayesian approach estimates the probability of a model, given information available about the system, including both prior knowledge and newly acquired experimental data. When modeling multiple structural states of a macromolecular system, the model M includes a set of N modeled structures $X = X_i$, their population fractions in the sample w_i, and additional parameters introduced below. The posterior probability $p(M \mid D, I)$ of model M given data D and prior knowledge I is

$$p(M \mid D, I) \propto p(D \mid M, I) \cdot p(M \mid I) \quad (1)$$

where the *likelihood function* $p(D \mid M, I)$ is the probability of observing data D given M and I; and the *prior* $p(M \mid I)$ is the probability of model M given I. The likelihood function is based on the *forward model* $f(X)$ that predicts the data point that would have been observed for structure(s) X in the absence of experimental noise, and a *noise model* that specifies the distribution of the deviation between the experimentally observed and predicted data points. The *Bayesian scoring function* is defined as $S(M) = -log\left[p(D \mid M, I) \cdot p(M \mid I)\right]$ which ranks the models the same as the posterior probability. The most probable models are found by selecting the best scoring models sampled from the posterior distribution. Next, we define the components of the Bayesian scoring function specifically for the binary and quantitative cross-linking data.

System representation. The spatial proximity between molecular components could conceivably be resolved by chemical cross-linking at the residue resolution.

---

[1] Wolfgang Rieping, Michael Habeck, and Michael Nilges, "Inferential Structure Determination.," *Science (New York, N.Y.)* 309, no. 5732 (2005): 303–306.

Therefore, a residue-level coarse-grained representation of the modeled structure is appropriate, allowing us to sample the posterior distribution more efficiently than with the atomic representation. To account for correct directionality of the cross-links, ε-amino groups of cross-linked lysines are explicitly modeled as a single bead centered on the nitrogen atom.

Forward model for binary cross-linking data. The forward model predicts the presence of a given cross-linked lysine pair n after reaction time t. Here, we assume that the reaction is unimolecular, that competitive reactions are negligible, and that the rate of interconversion between states is slower than the cross-linking rate. For a given structure X_i the forward model of binary cross-link data is:

$$f_n(X_i) = \left[1-\exp\left\{-k_n(X_i)t\right\}\right] \quad (2)$$

where $k_n(X_i)$ is the conformation dependent first-order cross-linking rate discussed below. For multiple state modeling, the forward models for each individual state are mixed in the likelihood function (see below).

Forward model for quantitative cross-linking data. The forward model predicts the ratio of ion intensities $R_n^{(u,b)} = I_n^{(u)} / I_n^{(b)}$ between the Hsp90-unbound and bound Δ131Δ, for a given cross-linked lysine pair n. For two sets of structures $\{X_k^{(u)}\}$ and $\{X_i^{(b)}\}$ for the unbound and bound states, respectively, the forward model is:

$$\overline{R}_n^{(u,b)}(\{X_k^{(u)}, w_k^{(u)}, X_i^{(b)}, w_i^{(b)}\}) = \frac{\sum_{k=1} w_k^{(u)} f_n(X_k^{(u)})}{\sum_{i=1} w_i^{(b)} f_n(X_i^{(b)})} \quad (3)$$

where $w_k^{(b)}$ and $w_i^{(u)}$ are the population fractions of the structures.

Cross-linking rate. Using a coarse-grained representation, the exponent in Eq. 2 can be approximated by:

$$k_n(X_i) \cdot t \sim \alpha_n \cdot \rho(r_n)$$

where the calibration parameter α_n is the product of the reaction time t and an unknown intrinsic reaction rate k_n that depends on the local environment of lysine pair n, such as *p*K. The dependence of the rate on geometry is encoded into the cross-linking efficiency term $\rho(r_n)$, where r_n is the distance between the two cross-linked ε-amino groups. $\rho(r_n)$ approximates a complex function that in principle depends on all atomic coordinates of the system. It is computed by considering i) the uncertainty in the position of ε-amino groups, and ii) the cost of having a cross-linking geometry far from the ideal one. The uncertainty in the position of the ε-amino groups is accounted for by adding noise to their positions. Assuming that their coordinates are random variables $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_j$ distributed

around the model coordinates $\mathbf{x}_i$ and $\mathbf{x}_j$ according to a 3D normal distributions with variances τ_i^2 and τ_j^2, respectively, the efficiency can thus be expressed as:

$$\rho(r_n) = \int d\tau p(\tau) \int d\tilde{r}_n \rho(\tilde{r}_n) \cdot p(\tilde{r}_n \mid r_n, \tau) \quad (4)$$

where $\tau^2 = \tau_i^2 + \tau_j^2$. The function $p(\tilde{r}_n \mid r_n, \tau)$ is the conditional probability of having a random distance $\tilde{r}_n$ given r_n and τ, and corresponds to the Rice's distribution. We used an uniformative Jeffrey's prior for $p(\tau)$[2]. The cost of having a cross-link geometry far from the ideal one is encoded into $\rho(\tilde{r}_n)$ and it is calculated as the potential of mean force of a system consisting of the DSS cross-linker covalently attached to the two lysines, as a function of the distance of the two ε-amino groups $\tilde{r}_n$. $\rho(\tilde{r}_n)$ is obtained from the free energy $F(\tilde{r}_n)$ after the Jacobian correction[3]:

$$\rho(\tilde{r}_n) \propto (4\pi\tilde{r}_n^2)^{-1} \exp(-F(\tilde{r}_n)/k_B T) \quad (5)$$

where k_B is the Boltzmann constant and T is the temperature. $F(\tilde{r}_n)$ is estimated by atomistic molecular dynamics simulations carried on the solvated DSS molecule (Yannick Spill and Michael Nilges, personal communication).

Likelihood function for binary cross-linking data. The likelihood function for a single observed cross-link d_n and a single structure X_i equals the forward model of Eq. 2, i.e., $p(d_n \mid X_i, I) = f_n(X_i)$) and spans the interval [0,1]. The likelihood of observing a cross-link given a set $\{X_i\}$ of structures is:

$$p(d_n \mid \{X_i\}, I) = 1 - \prod_i (1 - f_n(X_i)) \quad (6)$$

which is 1 when f_n=1 for at least one conformation, and 0 when all f_n=0. The joint likelihood function $p(D \mid M, I)$ for a dataset $D = \{d_n\}$ of N_{XL} independently observed cross-links is a product of likelihood functions for each data point.

Likelihood function for quantitative cross-linking data. The likelihood function $p(D \mid M, I)$ for the dataset $D = \{R_n^{(b,u)}\}$ of N_{XL} independently measured cross-linking ion intensity ratios is a product of likelihood functions for each data point:

$$p(D \mid M, I) = \prod_{n=1}^{N_{XL}} p(R_n^{(u,b)} \mid \{X_k^{(u)}, w_k^{(u)}, X_i^{(b)}, w_i^{(b)}\}, \alpha_n, \sigma_n) \quad (7)$$

where the uncertainty σ_n shapes the likelihood function for data point d_n. To account for varying levels of noise in the data, each data point has an individual σ_n. Because the observed cross-linked ratios are non-negative numbers, we modeled the noise using a log-normal distribution:

---

[2] Jaynes and Bretthorst, *Probability Theory*.

[3] Stefan Boresch and Martin Karplus, "The Jacobian Factor in Free Energy Simulations," *J Chem Phys* 105, no. 12 (1996): 5145–5154, doi:10.1063/1.472358.

$$p(R_n^{(u,b)} \mid \{X_k^{(u)}, w_k^{(u)}, X_i^{(b)}, w_i^{(b)}\}, \alpha_n, \sigma_n) = (\sqrt{2\pi}\sigma_n R_n^{(u,b)})^{-1} \exp(-\log^2(R_n^{(u,b)} / \bar{R}_n^{(u,b)}) / 2\sigma_n^2) \quad (8)$$

Prior. The model prior $p(M \mid I)$ is defined as a product of the individual priors $p(X_i)$, $p(w_i)$, $p(\sigma_n)$, and $p(\alpha_n)$ on the state coordinates, population fractions, data point uncertainties, and calibration parameters, respectively. The priors $p(X_i)$ include terms to maintain the correct stereochemistry of the backbone and the position of ε-amino groups, to avoid clashes between components, and to incorporate foldon information from hydrogen exchange data. The prior on a set of structures is defined as $p(\{X_i\}) \propto \exp\left(-\sum_i V(X_i)\right)$ where V is a sum of spatial restraints:

$$V = V_{\text{excl.vol.}} + V_{\text{Ca-bonds}} + V_{\text{Ca-angles}} + V_{\text{Ca-dihedrals}} + V_{\text{lysine sidechains}} + V_{\text{foldons}}$$

The excluded volume restraint was implemented as a pairwise hard-sphere repulsive potential, where the volume of each Cα particle equals the average volume of the corresponding amino acid residue. The bond, angle, and dihedral terms are statistical potentials derived from crystallographic structures (see below). To improve the accuracy of the crosslinking stereochemistry, the cross-linked lysines are represented by an additional bead. This bead corresponds to the sidechain amino group position. Harmonic distance, angular, and improper restraints are applied between the sidechain bead and the Cα bead, calibrated to mimic the flexibility and the orientation of the sidechain with respect to the backbone. Foldons are encoded as distance restraints to enforce the native structure of Staphilococcal Nuclease (SN) around the corresponding backbone NH-CO hydrogen bonds.

The priors $p(w_i)$ are uniform distributions over the range from 0 to 1, with the constraint $\sum_i w_i = 1$. The priors $p(\sigma_n)$ are unimodal distributions:

$$p(\sigma_n \mid \sigma) = 2\sigma(\sqrt{\pi}\sigma_n^2)^{-1} \exp(-\sigma^2 / \sigma_n^2) \quad (9)$$

where σ corresponds to an unknown experimental uncertainty; the heavy tail of the distribution allows for outliers. A single α_n can be used for all detected cross-links, assuming that the reaction rates k_n are averaged over all lysine pairs, and given that the total reaction time t is identical for all cross-linking reactions. For quantitative modeling, the priors $p(\alpha_n)$ are uniform distributions in the range [0,1].

Secondary structure terms. The bond, angle, and dihedral terms $V_{C\alpha\ \text{bonds}}$, $V_{C\alpha\ \text{angles}}$, and $V_{C\alpha\ \text{dihedrals}}$, respectively, are statistical potentials that enforce the correct stereochemistry, as well as the correct secondary structure propensity, of the flexible backbone. The input information is the predicted secondary structure (using DSSP secondary structure symbols[4]) calculated on the Staphylococcal

[4] W Kabsch and C Sander, "Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features.," *Biopolymers* 22, no. 12 (December 1983): 2577–2637.

Nuclease crystallographic structure (PDB entry code 1STN). These terms were computed by estimating the probability that residues in a given secondary structure sequence adopts a given configuration, defined by residue-residue distances, angles, and torsion angles. The probability is derived from the MRS database of crystallographic structures with assigned secondary structure[5]. For each sequence-contiguous residue pair $(n,n+1)$, triplet $(n,n+1,n+2)$ and quintuplet $(n,n+1,n+2,n+3,n+4)$, the potentials are calculated as:

$$V_{C_\alpha-Bonds}(r_n;\ S_n,S_{n+1}) = -log\left(\frac{\sum_k\sum_i \delta(r_n - r_i^{(k)})\delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}}{\sum_k\sum_i \delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}}\right)$$

$$V_{C_\alpha-Angle}(\alpha_n;\ S_n,S_{n+1},S_{n+2}) = -log\left(\frac{\sum_k\sum_i \delta(\alpha_n - \alpha_i^{(k)})\delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}\delta_{S_{n+2},S_{i+2}^{(k)}}}{\sum_k\sum_i \delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}\delta_{S_{n+2},S_{i+2}^{(k)}}}\right)$$

$$V_{C_\alpha-Dihedral}(\tau_n,\tau_{n+1};\ S_n,S_{n+1},S_{n+2},S_{n+3},S_{n+4}) =$$
$$= -log\left(\frac{\sum_k\sum_i \delta(\tau_n - \tau_i^{(k)})\delta(\tau_{n+1} - \tau_{i+1}^{(k)})\delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}\delta_{S_{n+2},S_{i+2}^{(k)}}\delta_{S_{n+3},S_{i+3}^{(k)}}\delta_{S_{n+4},S_{i+4}^{(k)}}}{\sum_k\sum_i \delta_{S_n,S_i^{(k)}}\delta_{S_{n+1},S_{i+1}^{(k)}}\delta_{S_{n+2},S_{i+2}^{(k)}}\delta_{S_{n+3},S_{i+3}^{(k)}}\delta_{S_{n+4},S_{i+4}^{(k)}}}\right)$$

where r_n, α_n, and τ_n are respectively the distance, the angle, and the torsion angle between sequence-contiguous residue pairs, triplets, and quadruplets starting from residue n; n, i, and k are respectively indexes for the residue in the model, the residue in the database structure, and the structure number in the database; δ is the Kronecker delta function; $S_n \in \{H,E,C\}$ is the secondary structure symbol for residue n, where H,E,C correspond to helical, beta, and random coil. The denominator on the left side of each equation is the normalization term over the given secondary structure sequence.

The dihedral term corresponds to the joint probability of having the torsion angles τ_n and τ_{n+1} at given values, given that the secondary structure sequence is $S_n,S_{n+1},S_{n+2},S_{n+3},S_{n+4}$. This term enforces the secondary structure geometry on the Cα model more effectively than a term that depends on a single torsion angle τ_n.

To increase the flexibility of the backbone, every native secondary structure potential term was mixed with a random coil secondary structure potential term. The mixing factors were determined by trial and error to have reversible folding of secondary structure elements along the sampling calculation (30% native and 70% random coil).

Bayesian scoring function. The multi-state Bayesian scoring function for binary cross-linking data is:

$$S(M) = -\sum_{n=1}^{N_{XL}} \log[p(d_n \mid \{X_i\})] - \sum_{i=1} \log[p(X_i)]$$

[5] M L Hekkelman and G Vriend, "MRS: a Fast and Compact Retrieval System for Biological Data.," *Nucleic Acids Research* 33, no. Web Server issue (July 1, 2005): W766–9.

For quantitative cross-linking data, to facilitate the sampling of the posterior probability distribution, we eliminate its dependence on uncertainties σ_n. This goal was achieved by numerical integration (*i.e.*, marginalization) of the posterior distribution with respect to uncertainties σ_n.
Thus, the marginal likelihood function is:

$$p(R_n^{(u,b)} \mid \{X_k^{(u)}, w_k^{(u)}, X_i^{(b)} w_i^{(b)}\}, \alpha_n, \sigma) = \int d\sigma_n p(\sigma_n \mid \sigma) p(R_n^{(u,b)} \mid \{X_k^{(u)}, w_k^{(u)}, X_i^{(b)} w_i^{(b)}\}, \alpha_n, \sigma_n)$$

$$= \sqrt{2}\sigma\left(\pi R_n^{(u,b)}(2\sigma^2 + \log^2(R_n^{(u,b)} / \bar{R}_n^{(u,b)}))\right)^{-1}$$

Hence, the multi-state Bayesian scoring function for quantitative cross-linking data is:

$$S(M) = -\sum_{n=1}^{N_{XL}} \log[\, p(R_n^{(u,b)} \mid \{X_k^{(u)}, w_k^{(u)}, X_i^{(b)} w_i^{(b)}\}, \alpha_n, \sigma) \cdot p(\alpha_n)] +$$
$$- \sum_{i=1} \log[\, p(X_i^{(b)}) \cdot p(w_i^{(b)})] - \sum_{k=1} \log[\, p(X_k^{(u)}) \cdot p(w_k^{(u)})]$$

Sampling. Metropolis Monte Carlo enhanced by replica exchange with 16 replicas was used to generate a sample of coordinates $\{X_i\}$ from the posterior distribution defined on binary cross-linking data. The moves for $\{X_i\}$ are random translation of individual beads by a maximum of 0.15 Å for each Monte Carlo step. A Gibbs sampling scheme with Metropolis Monte Carlo was applied to sample the values of the parameters α_n, w_i, and σ for the posterior distribution defined on quantitative cross-linking data. The multi-state pairs consisting of $\{X_k^{(b)}\}$ and $\{X_i^{(u)}\}$ were enumerated from the pool of highly probable multi-state structures $\{X_i\}$ generated from the binary cross-linking data modeling.

Modeling Protocol. The modeling was organized into two steps. We initially generated multi-state models based on the crosslinking binary data using the likelihood function defined by Eq. 6, combined with the priors discussed above. Three foldon priors were defined: the first used only the red foldon, the second combined the red and the yellow foldons, the third used all four (red, yellow, green and blue) foldons. For each foldon prior, the models were generated using 1, 2, 3, and 4 states, for a total of 12 modeling runs. Models were generated by Monte Carlo sampling of coordinates. About 1.5 10^6, 8.0 10^5 , 5.0 10^5 and 4.0 10^5 models were generated for each of the 1-, 2-, 3- and 4-state modeling, respectively. Finally, from each of the 12 sampling runs, the 100 best scoring multi-state models were selected, for a total of 1200 models.

We used all possible pairs of the selected models to quantitatively assess the quantitative crosslinking likelihood (Eq. 7). Each selected model was used either as a bound or unbound state, for a total of 1.44 10^6 combinations. In the second modeling step, we fixed the coordinates of the bound and unbound structures, and determined the values of the population fractions, the uncertainties, and the calibration parameters by maximizing the posterior probability. We finally

selected the three best scoring models (which are pairs of models derived from the first modeling step, Figs. S4 and S5).

Solvent accessible surface area calculation. Atomistic models were generated from coarse grained structure by using the program PULCHRA[6]. Atomistic degrees of freedom were subsequently relaxed using CHARMM force field[7] by running 50 steps of steepest descents and 100 steps of conjugated gradient. Residue solvent accessible areas were computed using the CHARMM program.

[6] Piotr Rotkiewicz and Jeffrey Skolnick, "Fast Procedure for Reconstruction of Full-Atom Protein Models From Reduced Representations.," *J Comput Chem* 29, no. 9 (July 15, 2008): 1460–1465.
[7] B R Brooks et al., "CHARMM: the Biomolecular Simulation Program.," *J Comput Chem* 30, no. 10 (July 2009): 1545–1614.