bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Metastable contacts and structural disorder in the estrogen receptor transactivation domain

Yi Peng[1*], Shufen Cao[2*], Janna Kiselar[1], Xiangzhu Xiao[2], An Hsien[1], Yinghua Chen[2], Wuxian Shi[1], Wei Jiang[3], Lin Yang[4], Mark R. Chance[1], Witold K. Surewicz[2], Matthias Buck[2], and Sichun Yang[1,2†]

[1]Center for Proteomics and Department of Nutrition, School of Medicine, Case Western Reserve University, Cleveland, OH

[2]Department of Physiology and Biophysics, School of Medicine, Case Western Reserve University, Cleveland, OH

[3]Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, IL

[4]Photon Sciences Directorate, Brookhaven National Laboratory, Upton, NY

*Equal contribution

[†]Correspondence: sichun.yang@case.edu or (216) 368 5793

Keywords: intrinsically disordered protein, SAXS, multi-technique modeling

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The N-terminal transactivation domain (NTD) of estrogen receptor alpha, a well-known member of the family of intrinsically disordered proteins (IDPs), mediates the receptor's transactivation function to regulate gene expression. However, an accurate molecular dissection of NTD's structure-function relationships remains elusive. Here, using small-angle X-ray scattering (SAXS), nuclear magnetic resonance, circular dichroism, and hydrogen exchange mass spectrometry, we show that NTD adopts an unexpectedly compact, mostly disordered conformation that undergoes structural expansion upon chemical denaturation. By combining SAXS and hydroxyl radical footprinting measurements, we derive ensemble-structures that represent the natively compact and disordered NTD. The resulting contact map for the ensemble reveals that the NTD features metastable regional and long-range contacts, including specific interactions between residues I33 and S118 that pervade the ensemble. Mutation at S118, a known regulatory site via phosphorylation, promotes conformational changes and increases coactivator binding, confirming its important structural contributions. These findings extend our understanding of IDPs' structure and function, and how specific metastable and/or transient structural interactions within an IDP can mediate critical regulatory functions of disordered proteins.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

The N-terminal transactivation domain (NTD) of estrogen receptor alpha is critical for the receptor's constitutive activation via its engagement with coactivator proteins to regulate transcriptional initiation (1-6). The NTD is an intrinsically disordered protein (IDP), a widely observed structural feature for transactivation domains in various members of the nuclear receptor superfamily (7-9) as well as many other transcription factors such as p53 (10, 11). Due to the functional significance of the NTDs, they are of interest for drug targeting. In particular, drug candidates that bind an intrinsically disordered region in the NTD of the androgen receptor (AR) have been developed (12-14). However, understanding a drug's mechanism of action or the rational of targeting the NTD usually requires in-depth insights into their molecular features. Such knowledge has been hindered by the difficulties in characterizing the structural ensembles of NTDs in particular and IDPs in general. In this paper, we specifically attempt to understand the structure-function relationships of the NTD of the estrogen receptor alpha (ERα) with regard to the structural features that determine its transactivation function.

Current knowledge about the ERα-NTD is limited due to the technical challenges of characterizing IDPs in general (15, 16). Only limited structural information is available for ERα-NTD from circular dichroism (CD) spectra and chemical shifts from nuclear magnetic resonance (NMR) spectroscopy (5, 6). These have demonstrated that that the isolated NTD is largely unstructured with poorly resolved chemical shifts, and with a high content of random coils and minimal helicity. Recently, biophysical techniques such as small-angle X-ray scattering (SAXS), fluorescence resonance energy transfer (FRET), and hydrogen exchange mass spectrometry (HX-MS) have revealed interesting structural features of IDPs (17-19), in particular RXRα-NTD (20), AR-NTD (21), and PR-NTD (8). The application of such approaches to ERα-NTD, however, has not been previously reported to date.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To investigate ERα-NTD structural features under native and denaturing conditions, we conducted multiple, in-solution biophysical studies using a recombinant ERα-NTD protein. We first acquired data from SAXS and HX-MS measurements for the native NTD. HX-MS confirms its overall structural disorder, consistent with the observation by CD and NMR, while SAXS data demonstrate that the NTD is more compact than expected. To examine the influence of chemical denaturants on the NTD, we acquired additional SAXS, CD, and ^{1}H-^{15}N NMR data for chemically-denatured states and showed that the NTD undergoes structural expansion upon chemical denaturation, indicating a denaturant-dependent loss of structural features that drive the native compactness. To gain insights into the structural basis of this compact disorder, we performed all-atom explicit-solvent simulations to create a structural library of NTD conformations, and supplemented macroscopic SAXS data with hydroxyl radical footprinting coupled with mass spectrometry, which reports solvent accessibility for residue side chains (22-24). The combination of these complementary techniques was used to refine and assess ERα-NTD ensemble-structure models. We identified long-range interactions involving residues S118 and I33 driving compactness within the ensemble and showed that S118 mutation mediates conformational changes and alters coactivator binding. Overall, our multi-technique biophysical approach allowed us to elucidate specific structural features of the ERα-NTD that mediate its transactivation function, and our findings raise important questions as to the potential for metastable contacts to mediate functions for other nuclear receptors and/or IDPs in general.

Results

Compact disorder of ERα-NTD

We purified the ERα-NTD segment (SNA-M1-Y184; Fig. 1A and Fig. 1B) via a refolding process from inclusion bodies of bacterial expression and conducted multiple biophysical measurements to confirm the structural disorder and structural features of this construct under

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

native conditions. First, analysis of circular dichroism (CD) and 2D-NMR spectra reaffirms the disordered nature of the native state as observed by Warnwark and colleagues (5). Negative ellipticity at 200 nm indicates that the NTD has a high content of random coils with minimal helicity (Fig. 1C). Moreover, the ^{1}H-^{15}N TROSY spectra exhibit a characteristic spectral crowding in a random-coil region around the 8.0 ppm (Fig. 1D) due to the NTD's limited chemical shift dispersion (25). We confirm the overall lack of ordered secondary structure by HX-MS measurements (26-30). The HX-MS experiments show nearly 100% deuterium incorporation for the native NTD already after 1 min of exchange (Fig. S1). The only region showing less than complete exchange maps to N- and C-terminal peptides 3-12 and 166-175, respectively. Even though this local protection is very small and potentially within the margin of a systematic error of the HX-MS experiments used, it could also be explained by the presence of residual secondary structure under native conditions as indicated by CD data (see below). Overall, these HX-MS data across the protein sequence suggest that this NTD construct is highly dynamic, consistent with a high level of structural disorder and very little ordered secondary structure as observed by 2D NMR and CD spectroscopy.

We next acquired SAXS data for the NTD in aqueous solution. The scattering intensity I(q) (Fig. 1E) yields a radius of gyration value of R_g^{expt} = 31.0 ± 0.2 Å, using the Sosnick approach (31). A Kratky plot of $K(q) = (qR_g)^2 * I(q)$ shows the characteristics of disordered or unfolded proteins (Fig. 1F), compared to the bell-shaped K(q) distribution typically observed in globular well-folded proteins. It is worth noting that many other IDPs approximately followed a power law of $R_g^{pred} = 2.54 * N^{0.522}$ (32); using this power law, the predicted value is R_g^{pred} = 39.0 Å for the 187-residue NTD, considerably greater than the measured value, even when we consider experimental uncertainties. The apparent difference between R_g^{pred} and R_g^{expt} indicates that ERα-NTD adopts a more compact conformation in its natively disordered state than would be generally expected for IDPs.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To further characterize the protein's biophysical properties, we calculated the mean hydrophobicity of the NTD at <*H*> = 0.446 (Fig. S2) using the Kyte-Doolittle scale (33, 34), and determine its mean net charge <*q*> = 0.003 per residue based on the sequence (Fig. 1B). Figure S3 shows that ERα-NTD deviates from the expected combination of <*H*> and <*q*> for IDPs in the two-dimensional H/q space, as it is seen to be to the "right" of Uversky's proposed boundary line that separates most IDPs and globular folds (35, 36). Of note, similar biophysical characteristics are observed for the 130-residue disordered RXRα-NTD which has values of <*H*> = 0.466 and <*q*> = 0.006 (20), with placement at a similar location in the two-dimensional H/q space. Moreover, we find a balanced charge distribution for ERα-NTD at f_+ = 0.080 and f_- = 0.086, the fraction of negatively charged (Asp+Glu) and positively charged (Arg+Lys) residues, respectively, consistent with the "weak polyampholyte" regime of IDPs (37). Overall, these biophysical characteristics of ERα-NTD, at the boundary of native globular and disordered proteins, are consistent with the compact-disordered conformation that the NTD adopts.

Structural expansion by chemical denaturation

To examine the effect of chemical denaturants on the NTD, we acquired CD data for the NTD in the presence of guanidine hydrochloride (GdnHCl), a chemical denaturant that can disrupt interactions driving the compactness. The minimal ellipticity at 222 nm observed for the native protein is reduced in the presence of GdnHCl (Fig. 2A), indicating the existence of residual secondary structure in the native state. The ^{1}H-^{15}N TROSY NMR spectra also show increases in structural disorder upon addition of chemical denaturants (Fig. 2B). In the latter experiment, GdnHCl was replaced by urea, a denaturing agent that is more compatible with NMR experiments. By overlapping the TROSY spectra for the NTD in the urea buffer, we observe that NMR peaks are gradually shifted to the downfield region and that the proton's chemical shift dispersion is narrowed (Fig. 2B top). These NMR peak changes in the presence of a denaturant, together with reduced CD signal at 222 nm, indicate an increase in the level

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of structural disorder as well as a disappearance of residual secondary structure. We next acquired SAXS data for the NTD in the presence of 3 M and 6 M GdnHCl. Compared with the native R_g^{expt} value of 31.0 Å, the protein exhibits R_g^{expt} values of 38.0 Å and 43.4 Å in the presence of 3 M and 6 M GdnHCl, respectively (Fig. 2C). The latter R_g^{expt} value for the denatured NTD is consistent with those reported for chemically-denatured proteins via a power (38), by which we estimate a R_g^{pred} value of 44.0 Å for the denatured NTD. This denaturant-dependent size change for the NTD is also consistent with other IDP polypeptide chains that expand with increasing denaturant concentration (39). In this case, however, the native NTD protein is more compact than expected for a typical IDP, and the structural changes seen in NMR and CD upon denaturation apparently disrupt interactions driving the compact state.

To further characterize this variation of the NTD conformation, we used SAXS data to estimate the pair-wise distance distribution p(*r*), a histogram of all pair-wise distances (*r)* between protein atoms. By using an indirect Fourier transformation (40), we calculated the p(*r*) function. For native NTD, this function shows an asymmetric peak below r = 50 Å (Fig. 2D), whereas in the presence of 3 M GdnHCl, it is considerably broadened (with a flat plateau), and more so in the presence of 6 M GdnHCl, in which it shows a broad distribution with a long tail up to r = 150 Å. The p(*r*) analysis shows that the NTD undergoes a considerable expansion from its native compactness with a large-scale stretching out of its polypeptide chain as a result of chemical denaturation.

Ensemble structures constructed by combining SAXS and FP-PF data

As our studies reveal distinctive features for the native NTD, we attempt to determine the structural basis for the unexpectedly compact state. Given the low-resolution nature of SAXS, however, it is recognized that SAXS data alone are not sufficient to provide details of ensemble-structures. To gain additional information, sequence- or position-specific data from other

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

techniques such as NMR and FRET have been used in conjunction with SAXS for characterizing IDPs (39, 41-44). Here, we supplement SAXS data by hydroxyl radical protein footprinting (FP) coupled with mass spectrometry, where transforming the data to protection factors (PFs) can report solvent accessibility of residue side chains. The value of this approach in modeling protein structures and protein-protein interactions has been demonstrated in our previous studies (45). To measure FP-PF data, purified NTD samples were exposed to a focused synchrotron X-ray white beam, where radiolysis-generated hydroxyl radicals react with solvent accessible side chains, and the sites and rates of oxidation are monitored by tandem mass spectrometry MS (22). The extent of modification is quantified as a function of X-ray exposure time (see examples in Fig. S4), yielding a measured rate of footprinting (k_{fp}) for each site. Dividing the measured rates by those expected for a full accessible site, the intrinsic reactivity with hydroxyl radical (k_R), provides a residue-specific protection factor (PF = k_R/k_{fp}) (24). The natural log of the PF values provides a quantitative residue-specific map on solvent accessibility; high logPF values reflect more protection from solvent (e.g., residues I33 and H124 with logPF = 2.69 and 3.03, respectively), while lower values (e.g., F97 with logPF = 0.68) reflect greater solvent accessibility. Table S1 shows the log PF values for the set of 16 residues that were quantitatively analyzed after Lys-C/Asp-N and/or pepsin digestions. These footprinting protection factor (FP-FP) data provide information about solvent accessibility at the various sites that can be integrated with SAXS data to facilitate ensemble-structure construction for the natively compact-disordered NTD.

We utilized a maximum-likelihood method for ensemble-structure construction that makes use of a structural library of NTD conformations in conjunction with SAXS and FP-PF data, following the general approaches available in the literature (39, 46-50). To generate the conformational library, various computational methods ranging from coarse-grained to simplified all-atom models have been developed (44, 51-54). Due to the prior knowledge about

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the NTD compactness, here we used an accumulative set of 35-μs all-atom explicit-solvent molecular dynamics simulation trajectories (Methods). The resulting pool of NTD conformations is combined with and compared against experimental SAXS and FP-PF data simultaneously to construct the best-fit ensemble-structures. The goodness of fit between the ensemble-calculated and experiment data is assessed by two unit-less functions χ^2 (Eq. 1) and φ^2 (Eq. 2), each measuring the difference between calculated and experimental SAXS and FP-PF data, respectively. To avoid potential over-fitting, we applied a maximum-likelihood method (Eq. 3, Eq. 4, and Fig. S5). As such, a broad probability distribution $\{P_i\}$ is determined for the ensemble-structures. This multi-technique modeling for the NTD resulted in an ensemble of top 10 clusters of structures (Fig. 3A) that contributes to a total of 16% of the entire population and an ensemble of top 100 structures that accounts for about 80% of the population (Fig. 3B).

Comparison of the ensemble-calculated and experimental SAXS profiles shows a reasonably good agreement (χ^2 = 1.7; Fig. 3C). In addition, a weighted average for side-chain solvent-accessible surface area (SA) is estimated for each residue and plotted against its corresponding logPF value (Fig. 3D). The SA vs. logPF plot has a correlation coefficient of -0.76 (P-value < 10^{-6}), illustrating the overall consistency of the ensemble-structures with experimental data. Previous correlation analyses of this kind have been performed on globular folded proteins, which typically exhibit a wide range of logPF values. However, most of the observations here have logPF values from 1 to 2 (SA values of 80-110 Å^2), and have few values at the extremes of being either highly solvent-protected or highly solvent-accessible (e.g., upper left or lower right quadrants). In order to remove any potential bias due to residue size, we normalized the SA data into the fractional SA (fSA) values, using the maximum possible SA for the residue (55). In a re-plotted fSA vs. logPF in Fig. S6, 12 out of the 16 points (circled) are clustered in the “middle” of the plot in a restricted range. This contrasts with globular folded proteins that exhibit as many residues in the “middle” range as in the high/low fSA regions (56,

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

57). This is presumably because the existence of a balanced distribution, including both high (e.g., fSA > 0.8) and low (e.g., fSA < 0.4) solvent-accessible residues, correlates with structural stability and maintains the overall fold for globular folded proteins. By comparison, the majority of NTD residues are in an intermediate range of fSA (and logPF) values, consistent with its dynamic but compact nature. This relatively narrow fSA distribution is also consistent with the modest chemical shift dispersion observed in NMR spectra and may reflect a structural signature of this NTD. Retrospectively, this broad distribution for the ensemble is also consistent with the typical behavior of disordered proteins, where an ensemble of structures is often required to characterize the structural properties of IDPs.

We next performed an ensemble contact-map analysis for the NTD ensemble-structures (Methods). The resulting contact-map shows that the NTD has multiple medium and long-range contacts formed (within residues 11-75), pointing to a 'collapsed' segment (for a portion of the distribution) at its N-terminus, compared to its C-terminus (Fig. 4A). This regional collapse is consistent with the higher hydrophobicity of its N-terminal sequence than the C-terminal (Fig. S2). Moreover, long-range interactions are formed (e.g., frequently observed in the ensemble) between two 5-residue peptides, each centered at I33 and S118 (marked by an arrow in Fig. 4A and illustrated in Fig. 4B). Residue I33 is predicted to have a SA value of 52 $Å^2$ from the ensemble-structures (Table S1), indicating that I33 is more protected from solvent exposure than the typical or average member of the ensemble. To gain insight into the nature of these conformations and their distributions, we examined the explicit structural distribution of I33, compared to L44, which has a similar/compatible side-chain size but with a higher SA value indicating much greater solvent exposure. It can be seen that based on the top 100 structures (Fig. 4C), I33 has far more ensemble members with SA values below 50 $Å^2$, compared to L44 that has a greater number of conformations above 100 $Å^2$. I33 is on average solvent-protected to a greater degree, presumably due to its “metastable” interaction with S118 (see below) that

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

tilts the conformational distribution towards greater protection, compared to L44. While the interaction itself is presumably relatively weak compared to tertiary contacts for a globular well-folded protein, the metastable nature of this long-range interaction could provide a critical structural role driving the function of the NTD, as has been widely hypothesized (58).

To examine the functional role of S118 and to test its structural interaction predicted by our ensemble-structures, we mutated the non-charged S118 residue to a charged residue Asp (D) (i.e., S118D), as a phosphomimetic mutation. The ellipticity at 222 nm shows that S118D increases the content of ordered secondary structure (Fig. 4D). To gain insight into the impact on the overall protein, we further acquired SAXS data for the S118D mutant, yielding the R_g^{expt} value of 38.7 Å (Fig. 4E and Fig. S7), which is greater than that of the wild-type (R_g^{expt} = 31.0 Å) and smaller than the denatured state (R_g^{expt} = 43.4 Å for the NTD in the presence of 6 M GdnHCl). Mutation-dependent conformational changes are also confirmed by peak shifts and appearance/disappearance in the ^{1}H-^{15}N TROSY spectra of S118D (Fig. 4F; arrows). While detailed NMR assignments to identify specific residues involved will be the subject of future studies, these spectrum changes indicate that S118D mutation leads to modest conformational changes that spread out beyond the mutation site itself, in accord with the sizeable R_g change as observed by SAXS. Overall, these data collectively show that the mutation propagates its structural impact and physically extends the polypeptide chain, presumably by destabilizing S118-mediated long-range contacts, while counterintuitively resulting in a gain of local secondary structure.

We also examined the impact of the S118D mutation on the NTD binding to TBP, a general transcription factor TATA box-binding protein, known as an ER coactivator in the transcriptional machinery (3, 59). We estimated a weak-to-modest binding affinity of 53 ± 12 μM between the wild-type NTD and TBP, using surface plasmon resonance (SPR) spectrometry (Fig. S8), consistent with the previously reported binding affinity in the micromolar range (5).

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The NTD-TBP binding is further demonstrated by NMR experiments, which show multiple peak shifts and appearance/disappearance in the ^{1}H-^{15}N spectra of the wild-type (WT) upon addition of TBP (Fig. S9A). Compared to the WT, however, the addition of TBP leads to a greater extent of peak shifts/changes in the ^{1}H-^{15}N spectra of the S118D mutant (Fig. S9B), indicating that S118D increases coactivator binding as a result of the conformational changes. This increased coactivator binding is further supported by the SPR measurements, where S118D increases the TBP-binding affinity more than ten-fold to K_d = 3.9 ± 0.4 μM (Fig. S10), compared to the WT's K_d = 53 ± 12 μM. Together, these data show that the S118D mutation induces sizable conformational changes within the NTD and further increases the coactivator binding.

Discussion

The NTD accounts for a large portion of the total transcriptional activities of most nuclear receptors in terms of transactivation function. While it is recognized early on that ERα-NTD is structurally disordered, little is known about its molecular details, compared to extensive knowledge available for other regions of the receptor such as the DNA-binding domain and the hormone-binding domain (Fig. 1A). This study provides a comprehensive characterization of the NTD's rather distinctive structural features for both native and chemically-denatured states. Our biophysical measurements show that the NTD is mostly disordered as revealed by CD and NMR, although SAXS data point to considerable compactness, in accord with the modest hydrophobicity and nearly-zero net charge that is more common for globular proteins. The ensemble structures were observed to undergo a loss of residual secondary structure and a structural expansion from a narrow single-peak distance distribution in the native state to a broad distribution in chemically-denatured states, indicative of melting of pre-existing structural features associated with the compactness and/or helicity. Of note, this phenomenon of a large expansion with >40% increase in R_g from 31 to 43 Å was observed for ERα-NTD

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

only; a generalization to and comparison with other IDPs such as RXRα-NTD and AR-NTD has not yet been attempted.

The ensemble-structures constructed by multi-technique modeling, which combines SAXS and FP-PF data together with a library of simulation-generated conformations, provide a rare opportunity to observe the molecular details of ERα-NTD. We acknowledge that this modeling process cannot provide a unique solution to understanding the ERα-NTD structures in part due to the ensemble-average and/or conformer-coexistence nature of both SAXS and FP-PF data. We further note that several residues had extreme logPF values not explained by the modeling. In particular, H124 was observed to be the most protected, but our ensemble-structure models could not explain its high degree of solvent protection (Fig. 3D), which points to next-steps for future ensemble refinements. Nonetheless, given the strong evidence that validates our constructed ensemble-structures for ERα-NTD, our multi-technique approach could be applied for the structure determination of other IDP systems such as RXRα-NTD.

In a broader context, IDPs are increasingly appreciated as functionally important for coactivator recruitment of protein-protein interactions involved in cellular signaling and transcriptional regulation. Typical IDPs are viewed as a diverse population of interchangeable structures or conformers that co-exist in a concerted equilibrium. Here, the ensemble-structure studies for ERα-NTD add to our understanding of the range of IDPs' structure-function capability. A large portion of the ensemble population for ERα-NTD favors a regional collapse at the N-terminus contributing to the overall compactness. Moreover, a subset of the population, roughly three out of the top 10 structures in the ensemble, adopts long-range contacts near I33 and S118 that presumably tilt the dynamic equilibrium in favor of more compact conformers. The extended NTD ensemble, accompanied by the increase of ordered secondary structure, possesses structural elements or flexibility that permits the IDP to interact more productively with the folded TBP protein, resulting in stronger coactivator binding. Functionally, S118 is a

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

known phosphorylation site that modulates the receptor's transcriptional activity (60), and an increase in the level of S118 phosphorylation has been associated with endocrine resistance via direct recruitment of coactivators (61). As a phosphomimetic mutation, S118D directly links a phosphorylation-driven shift in the NTD conformational equilibrium in favor of coactivator recruitment, providing a molecular basis for understanding S118-associated endocrine resistance in breast cancer. S118 phosphorylation is also critical for association with a cis/trans isomerase pin1 as observed in a truncated NTD segment, while mutation of S118 to alanine prevents the association (6). Such mutation-induced modulation is also observed in other IDPs such as the transcription factor HIF-1α's transactivation domain (19) and a disordered PAGE4 protein (62). In conclusion, as we note above that the hydrophobicity and mean net charge of RXRα-NTD is similar to that of ERα-NTD, our findings suggest that a comprehensive characterization of RXRα-NTD and/or other NTDs by these approaches may reveal additional examples of metastable long-range contacts that fine-tune the conformational equilibrium to execute particular molecular functions as a general mechanism of mediating nuclear receptor structure and function.

Methods. The ERα-NTD segment (residues M1-Y184) was inserted into a pMCSG7 vector and expressed in *E. coli* cells. Multiple in solution biophysical measurements (CD, NMR, SAXS, HX-MS, FP-PF, and SPR) were performed. Subsequent multi-technique modeling was conducted by combining MD simulation results with experimental SAXS and FP-PF data for the construction of the NTD ensemble-structures. See details in Supporting Information.

Acknowledgments. We thank Sayan Gupta for assistance with footprinting experiments, Geof Greene for inspiration of this project, and Marc Parisien and Wenwei Zhang for comments. The work of S.Y. was supported by a NIH grant (GM114056). Use of the synchrotron sources was supported by the U.S. Department of Energy (DE-AC02-98CH10886) and by NIH (EB009998).

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Tora L, *et al.* (1989) The human estrogen receptor has two independent nonacidic transcriptional activation functions. *Cell* 59(3):477-487.

2. Metivier R, Petit FG, Valotaire Y, & Pakdel F (2000) Function of N-terminal transactivation domain of the estrogen receptor requires a potential alpha-helical structure and is negatively regulated by the A domain. *Mol Endocrinol* 14(11):1849-1871.

3. Sadovsky Y, *et al.* (1995) Transcriptional activators differ in their responses to overexpression of TATA-box-binding protein. *Mol Cell Biol* 15(3):1554-1563.

4. Lavery DN & McEwan IJ (2005) Structure and function of steroid receptor AF1 transactivation domains: induction of active conformations. *Biochem J* 391(Pt 3):449-464.

5. Warnmark A, Wikstrom A, Wright APH, Gustafsson JA, & Hard T (2001) The N-terminal regions of estrogen receptor alpha and beta are unstructured in vitro and show different TBP binding properties. *Journal of Biological Chemistry* 276(49):45939-45944.

6. Rajbhandari P, *et al.* (2012) Regulation of estrogen receptor alpha N-terminus conformation and function by peptidyl prolyl isomerase Pin1. *Mol Cell Biol* 32(2):445-457.

7. Kumar R & Thompson EB (2012) Folding of the glucocorticoid receptor N-terminal transactivation function: dynamics and regulation. *Mol Cell Endocrinol* 348(2):450-456.

8. Kumar R, *et al.* (2013) Regulation of the structurally dynamic N-terminal domain of progesterone receptor by protein-induced folding. *J Biol Chem* 288(42):30285-30299.

9. Li J, Motlagh HN, Chakuroff C, Thompson EB, & Hilser VJ (2012) Thermodynamic dissection of the intrinsically disordered N-terminal domain of human glucocorticoid receptor. *J Biol Chem* 287(32):26777-26787.

10. Lee CW, Martinez-Yamout MA, Dyson HJ, & Wright PE (2010) Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein. *Biochemistry* 49(46):9964-9971.

11. Dawson R, *et al.* (2003) The N-terminal domain of p53 is natively unfolded. *J Mol Biol* 332(5):1131-1141.

12. Myung JK, *et al.* (2013) An androgen receptor N-terminal domain antagonist for treating prostate cancer. *J Clin Invest* 123(7):2948-2960.

13. De Mol E, *et al.* (2016) EPI-001, A Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor. *ACS Chem Biol* 11(9):2499-2505.

14. De Mol E, *et al.* (2018) Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators. *Structure* 26(1):145-152 e143.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

15. van der Lee R*, et al.* (2014) Classification of intrinsically disordered regions and proteins. *Chem Rev* 114(13):6589-6631.

16. Oldfield CJ & Dunker AK (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* 83:553-584.

17. Lindorff-Larsen K*, et al.* (2004) Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein. *J Am Chem Soc* 126(10):3291-3299.

18. Choy WY & Forman-Kay JD (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J Mol Biol* 308(5):1011-1032.

19. Berlow RB, Dyson HJ, & Wright PE (2017) Hypersensitive termination of the hypoxic response by a disordered protein switch. *Nature* 543(7645):447-451.

20. Belorusova A*, et al.* (2016) Solution Behavior of the Intrinsically Disordered N-Terminal Domain of Retinoid X Receptor alpha in the Context of the Full-Length Protein. *Biochemistry* 55(12):1741-1748.

21. Tien AH & Sadar MD (2018) Order within a Disordered Structure. *Structure* 26(1):4-6.

22. Takamoto K & Chance MR (2006) Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annu Rev Biophys Biomol Struct* 35:251-276.

23. Wang L & Chance MR (2017) Protein Footprinting Comes of Age: Mass Spectrometry for Biophysical Structure Assessment. *Mol Cell Proteomics* 16(5):706-716.

24. Huang W, Ravikumar KM, Chance MR, & Yang S (2015) Quantitative mapping of protein structure by hydroxyl radical footprinting-mediated structural mass spectrometry: a protection factor analysis. *Biophys J* 108(1):107-115.

25. Sahu D, Bastidas M, & Showalter SA (2014) Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods. *Anal Biochem* 449:17-25.

26. Englander SW & Kallenbach NR (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys* 16(4):521-655.

27. Craig PO*, et al.* (2011) Prediction of native-state hydrogen exchange from perfectly funneled energy landscapes. *J Am Chem Soc* 133(43):17463-17472.

28. Smirnovas V*, et al.* (2011) Structural organization of brain-derived mammalian prions examined by hydrogen-deuterium exchange. *Nat Struct Mol Biol* 18(4):504-506.

29. Bai Y, Milne JS, Mayne L, & Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins* 17(1):75-86.

30. Buck M, Radford SE, & Dobson CM (1994) Amide hydrogen exchange in a highly denatured state. Hen egg-white lysozyme in urea. *J Mol Biol* 237(3):247-254.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

31. Riback JA, *et al.* (2017) Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* 358(6360):238-241.

32. Bernado P & Svergun DI (2012) Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol Biosyst* 8(1):151-167.

33. Kyte J & Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157(1):105-132.

34. Wilkins MR, *et al.* (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112:531-552.

35. Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* 11(4):739-756.

36. Mao AH, Crick SL, Vitalis A, Chicoine CL, & Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci U S A* 107(18):8183-8188.

37. Das RK & Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc Natl Acad Sci U S A* 110(33):13392-13397.

38. Kohn JE, *et al.* (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci U S A* 101(34):12491-12496.

39. Borgia A, *et al.* (2016) Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. *J Am Chem Soc* 138(36):11714-11726.

40. Svergun DI (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J Appl Crystallogr* 25:495-503.

41. Ganguly D & Chen J (2009) Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states. *J Mol Biol* 390(3):467-477.

42. Zheng W, *et al.* (2016) Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment. *J Am Chem Soc* 138(36):11702-11713.

43. Fuertes G, *et al.* (2017) Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. *Proc Natl Acad Sci U S A* 114(31):E6342-E6351.

44. Krzeminski M, Marsh JA, Neale C, Choy WY, & Forman-Kay JD (2013) Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* 29(3):398-399.

45. Huang W, Ravikumar KM, Parisien M, & Yang S (2016) Theoretical modeling of multiprotein complexes by iSPOT: Integration of small-angle X-ray scattering, hydroxyl radical footprinting, and computational docking. *J Struct Biol* 196(3):340-349.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

46. Rozycki B, Kim YC, & Hummer G (2011) SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* 19(1):109-116.

47. Schwalbe M*, et al.* (2014) Predictive atomic resolution descriptions of intrinsically disordered hTau40 and alpha-synuclein in solution from NMR and small angle scattering. *Structure* 22(2):238-249.

48. Brookes DH & Head-Gordon T (2016) Experimental Inferential Structure Determination of Ensembles for Intrinsically Disordered Proteins. *J Am Chem Soc* 138(13):4530-4538.

49. Roux B & Weare J (2013) On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J Chem Phys* 138(8):084107.

50. Bernado P*, et al.* (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci U S A* 102(47):17002-17007.

51. Vitalis A & Pappu RV (2009) ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *J Comput Chem* 30(5):673-699.

52. Ghavami A, van der Giessen E, & Onck PR (2013) Coarse-Grained Potentials for Local Interactions in Unfolded Proteins. *J Chem Theory Comput* 9(1):432-440.

53. Zhang W & Chen J (2014) Replica exchange with guided annealing for accelerated sampling of disordered protein conformations. *J Comput Chem* 35(23):1682-1689.

54. Ozenne V*, et al.* (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* 28(11):1463-1470.

55. Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, & Wilke CO (2013) Maximum allowed solvent accessibilites of residues in proteins. *PLoS One* 8(11):e80635.

56. Kaur P, Kiselar J, Yang S, & Chance MR (2015) Quantitative protein topography analysis and high-resolution structure prediction using hydroxyl radical labeling and tandem-ion mass spectrometry (MS). *Mol Cell Proteomics* 14(4):1159-1168.

57. Gustavsson M*, et al.* (2017) Structural basis of ligand interaction with atypical chemokine receptor 3. *Nat Commun* 8:14135.

58. Wright PE & Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* 16(1):18-29.

59. Louder RK*, et al.* (2016) Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature* 531(7596):604-609.

60. Chen D*, et al.* (2000) Activation of estrogen receptor alpha by S118 phosphorylation involves a ligand-dependent interaction with TFIIH and participation of CDK7. *Mol Cell* 6(1):127-137.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

61. Kato S*, et al.* (1995) Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science* 270(5241):1491-1494.

62. Kulkarni P*, et al.* (2017) Phosphorylation-induced conformational dynamics in an intrinsically disordered protein and potential role in phenotypic heterogeneity. *Proc Natl Acad Sci U S A* 114(13):E2644-E2653.

63. Delaglio F*, et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277-293.

64. Smirnovas V*, et al.* (2009) Distinct structures of scrapie prion protein (PrPSc)-seeded versus spontaneous recombinant prion protein fibrils revealed by hydrogen/deuterium exchange. *J Biol Chem* 284(36):24233-24241.

65. Miao Y, Feher VA, & McCammon JA (2015) Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J Chem Theory Comput* 11(8):3584-3595.

66. Wang L, Friesner RA, & Berne BJ (2011) Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J Phys Chem B* 115(30):9431-9438.

67. Jo S & Jiang W (2015) A generic implementation of replica exchange with solute tempering (REST2) algorithm in NAMD for complex biophysical simulations. *Comput Phys Commun* 197:304-311.

68. Phillips JC*, et al.* (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781-1802.

69. Hornak V*, et al.* (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65(3):712-725.

70. Nerenberg PS & Head-Gordon T (2018) New developments in force fields for biomolecular simulations. *Curr Opin Struct Biol* 49:129-138.

71. Huang J*, et al.* (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 14(1):71-73.

72. Best RB, Zheng W, & Mittal J (2014) Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J Chem Theory Comput* 10(11):5113-5124.

73. Piana S, Donchev AG, Robustelli P, & Shaw DE (2015) Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J Phys Chem B* 119(16):5113-5123.

74. Song D, Luo R, & Chen HF (2017) The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins. *J Chem Inf Model* 57(5):1166-1178.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

75. Huang J & MacKerell AD, Jr. (2018) Force field development and simulations of intrinsically disordered proteins. *Curr Opin Struct Biol* 48:40-48.

76. Yang S, Blachowicz L, Makowski L, & Roux B (2010) Multidomain assembled states of Hck tyrosine kinase in solution. *Proc Natl Acad Sci U S A* 107(36):15757-15762.

77. Ravikumar KM, Huang W, & Yang S (2013) Fast-SAXS-pro: a unified approach to computing SAXS profiles of DNA, RNA, protein, and their complexes. *J Chem Phys* 138(2):024112.

78. Hsieh A, Lu L, Chance MR, & Yang S (2017) A Practical Guide to iSPOT Modeling: An Integrative Structural Biology Platform. *Adv Exp Med Biol* 1009:229-238.

79. Boomsma W, Ferkinghoff-Borg J, & Lindorff-Larsen K (2014) Combining experiments and simulations using the maximum entropy principle. *PLoS Comput Biol* 10(2):e1003406.

80. Conn AR, Gould N, & Toint PL (1997) A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. *Math Comput* 66(217):261-&.

81. Sobolev V, Sorokine A, Prilusky J, Abola EE, & Edelman M (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics* 15(4):327-332.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures and Captions

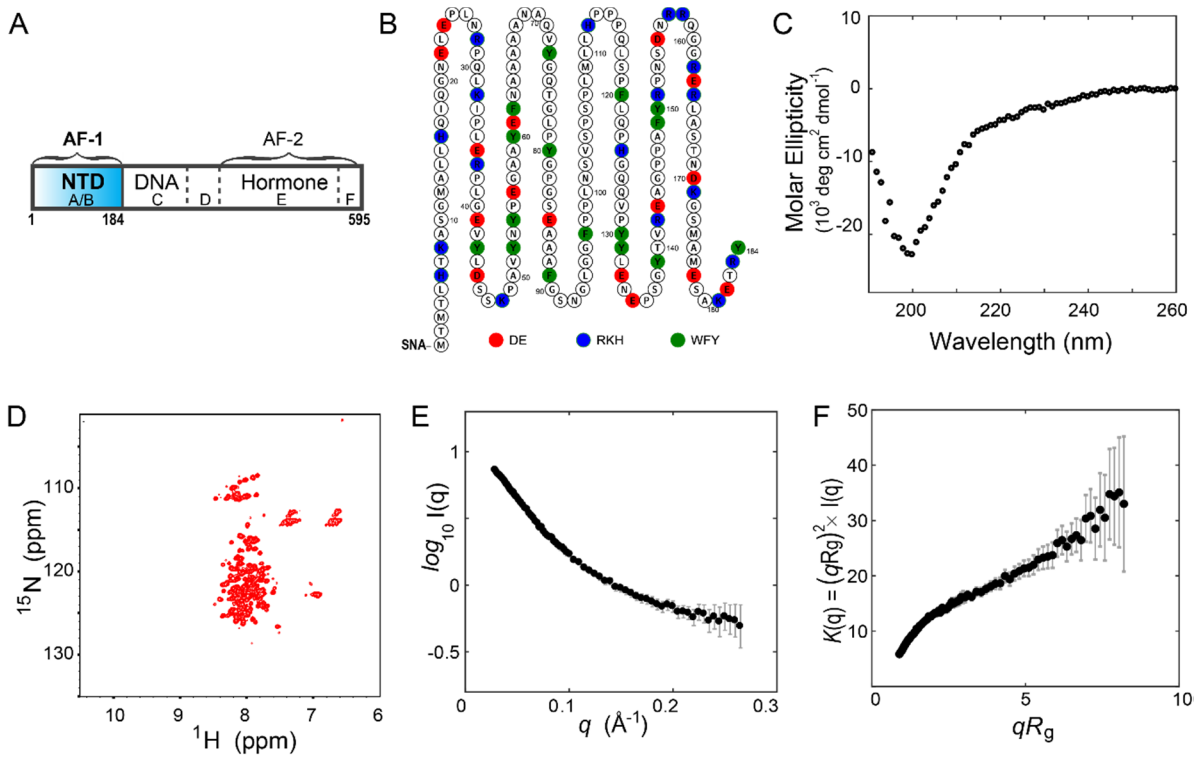

Figure 1. Structural disorder of ERα-NTD.

(A) Sub-domains of the ERα. Five domains are the NTD (A/B), a DNA-binding domain (C), a hinge region (D), a ligand-binding domain (E), and a short C-terminal domain (F).

(B) The 187-residue sequence. Charged and aromatic residues are highlighted in color.

(C) CD spectra for the NTD. (D) Two-dimensional ^{1}H-^{15}N TROSY spectrum.

(E,F) Measured SAXS data yield the R_g value of 31.0 ± 0.2 Å and its dimensionless Kratky plot shows a structurally disordered characteristic.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

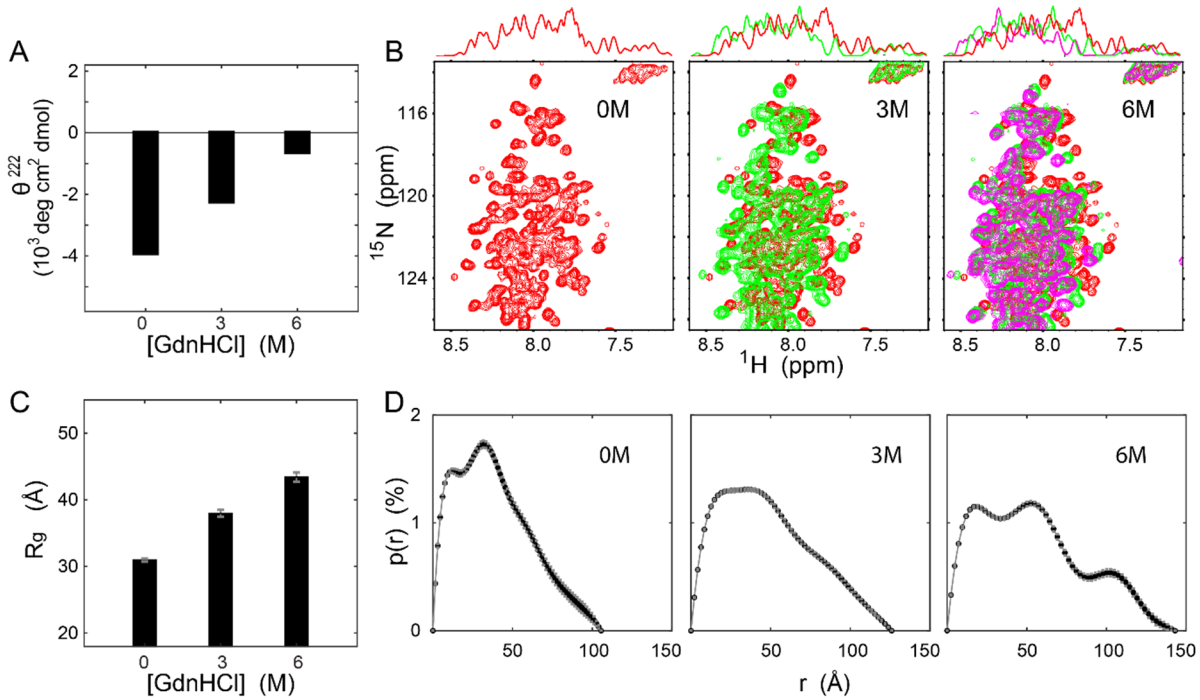

Figure 2. Structural expansion by chemical denaturation.

(A) CD signal at 222 nm for ERα-NTD in the presence of 0 M, 3 M, and 6 M GdnHCl, respectively.

(B) Two-dimensional ^{1}H-^{15}N TROSY spectra for the NTD in 0 M (red), 3 M (green), and 6 M (magenta) urea, respectively. On the top is the one-dimensional projection.

(C) SAXS-derived R_g values of 31.0 ± 0.2 Å, 38.0 ± 0.5 Å, and 43.4 ± 0.7 Å, respectively, using the Sosnick method (31).

(D) The pair-wise distance distribution p(r) derived from experimental SAXS data. Calculated using AutoGnom (40).

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

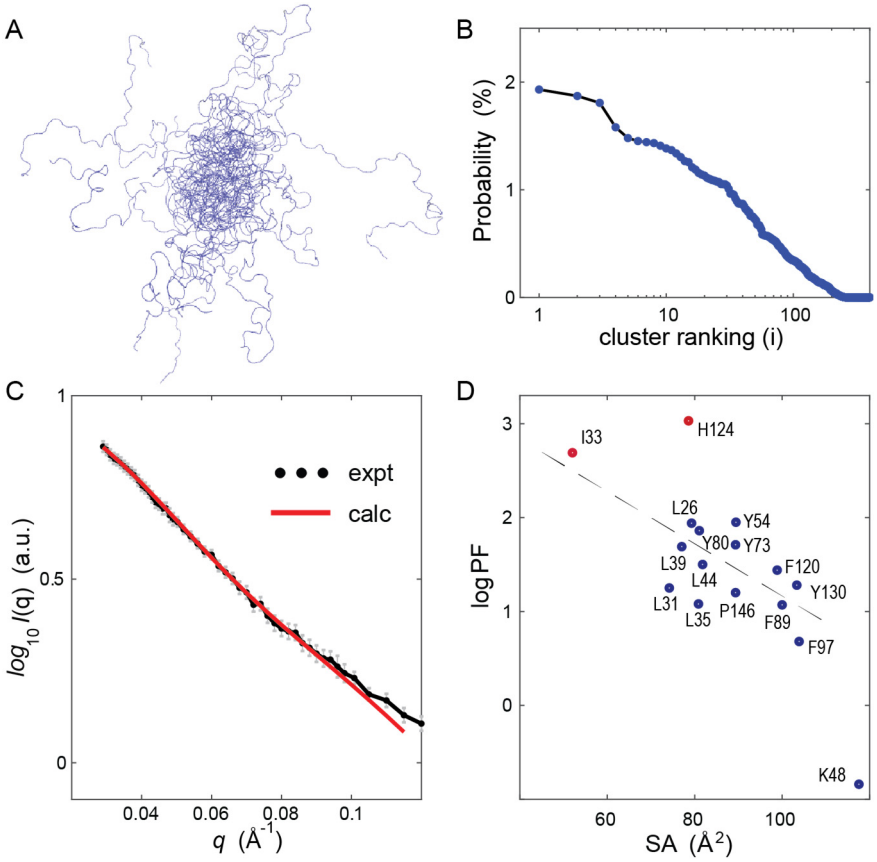

Figure 3. **Ensemble-structure construction for the native compact-disordered NTD**.

(A) Ensemble-structures constructed by integrating SAXS and FP-PF data. Shown is the ensemble of the top 10 structures.

(B) Probability distribution for the ensemble-structures.

(C) Measured and calculated SAXS profiles.

(D) Measured logPF values and ensemble-based solvent-accessible surface areas (SAs). A set of 16 residue side-chains were probed and shown.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

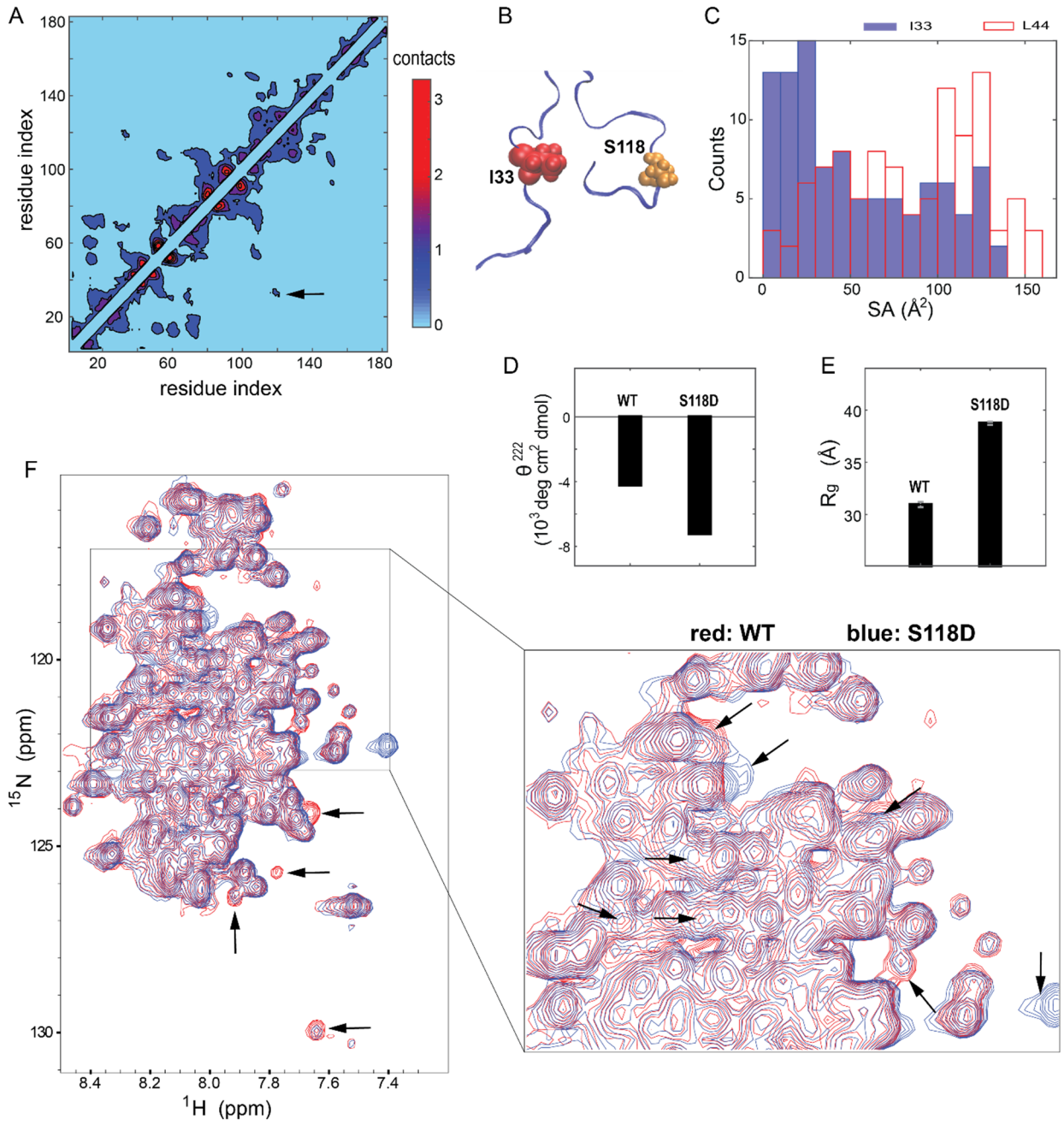

Figure 4. **Ensemble-contact map and mutation-induced conformational changes.**

(A) Contact map shows that its N-terminal region (e.g., residues 11-75) has more contacts formed than its C-terminal. The color-bar indicates the contact number, where each contact is ensemble-weighted (Methods). The metastable I33-S118 contact is marked by an arrow, where three of the top 10 structures from the ensemble make the contact.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(B) Cartoon illustration of I33-S118 interaction.

(C) Histogram for SA values for residues I33 and L44 from top 100 structures of the ensemble.

(D) CD signal at 222 nm for S118D as well as wild-type (WT).

(E) The R_g value of 38.7 ± 0.2 Å for mutation S118D, compared to the R_g value of 31.0 ± 0.2 Å for the wild-type (WT). Related SAXS data shown in Fig. S7.

(F) Overlap of ^{1}H-^{15}N TROSY spectra between WT (red) and mutant S118D (blue). Peak changes/shifts marked by arrows.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SUPPORTING INFORMATION

Metastable contacts and structural disorder in the estrogen receptor transactivation domain

Yi Peng[1*], Shufen Cao[2*], Janna Kiselar[1], Xiangzhu Xiao[2], An Hsien[1], Yinghua Chen[2], Wuxian Shi[1], Wei Jiang[3], Lin Yang[4], Mark R. Chance[1], Witold K. Surewicz[2], Matthias Buck[2], and Sichun Yang[1,2†]

[1]Center for Proteomics and Department of Nutrition, School of Medicine, Case Western Reserve University, Cleveland, OH

[2]Department of Physiology and Biophysics, School of Medicine, Case Western Reserve University, Cleveland, OH

[3]Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, IL

[4]Photon Sciences Directorate, Brookhaven National Laboratory, Upton, NY

*Equal contribution

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supporting Methods

Plasmid construct. The sequence of the N-terminal domain (NTD) of human ERα (M1-Y184) was inserted into a bacterial expression vector pMCSG7 containing an N-terminal 6-Histidine tag designed for Ni-NTA affinity purification. Verification of this expression vector was performed for DNA sequencing. His-tag was cleaved by TEV protease to generate the NTD protein with a short three-residue segment of SNA (from the expression vector) at its N-terminus.

Protein expression and purification. The NTD samples were expressed as inclusion bodies in bacterial BL21 (DE3) cells. Expression was carried out overnight at 18 °C with 0.1 mM IPTG in TB medium after the OD_{600} reached ~0.8. Cells were harvested by centrifugation at 5000 rpm, re-suspended in 50 mM Tris, pH 8.0, 500 mM NaCl with a protease inhibitor cocktail (Roche, Indianapolis, IN). The cells were disrupted using a M110Y microfluidizer (Microfluidics, Newton, MA). The pellet fraction containing the NTD inclusion bodies was solubilized in the buffer of 50 mM sodium phosphate, pH 7.4, 300 mM NaCl, 6 M guanidine hydrochloride, and 5 mM imidazole with 1 mM PMSF. Solubilized inclusion bodies were centrifuged at 18,000×g for 45 min at 4 °C. Cleared supernatant was incubated with Talon resin (Clontech) for 45 min at 4 °C, followed by a wash with a buffer of 50 mM sodium phosphate, pH 7.4, 500 mM NaCl, 5 mM imidazole, and 6 M guanidine hydrochloride. Samples were step-eluted with imidazole at a series of concentrations of 50, 250, and 500 mM. Fractions containing the His-tagged NTD protein were combined and dialyzed against a buffer of 20 mM Tris, pH 7.5, 10 mM NaCl and 0.1 mM PMSF. Refolded NTD proteins were collected and incubated with TEV protease at a TEV:NTD ratio of 1:50 overnight at 4 °C. The second step of affinity purification via Talon resin was used to remove the TEV and uncleaved His-tagged proteins. Final purification was performed using ion exchange via a HiTrap Q 5 mL column (GE Healthcare) and by gel filtration via an Enrich SEC650 column (Bio-Rad).

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Circular dichroism (CD) spectroscopy. NTD proteins in a buffer of 20 mM sodium phosphate (pH 7.4) were used for the far UV CD spectroscopy measurement at 25 °C on an AVIV 215 spectrometer with a 0.1 cm quartz cell. CD spectra were recorded in the wavelength of 190-260 nm. Each CD spectra were reported after buffer subtraction. Final spectra were averaged over three independent scans.

NMR Spectroscopy. Two-dimensional NMR spectra were recorded at 8 °C on a Bruker Avance II 800-MHz spectrometer equipped with a TXI cryoprobe. The NTD protein was prepared in the buffer of 10 mM sodium phosphate, pH 7.2, 100 mM NaCl, 0.5 mM EDTA and 0.1mM PMSF, in the presence of 0 M, 3 M, 6 M urea, respectively, with a final protein concentration of 100 μM. Here, urea was used to replace GdnHCl because of its chemical compatibility for NMR spectroscopy. Two-dimensional ^{1}H-^{15}N TROSY spectra were collected with 1024*248 (F1*F2) points and 64 scans. In addition, one-dimensional proton spectra were recorded, where the ^{1}H chemical shifts of Arginine side chains (at 6.92 ppm and 6.97 ppm) were identified and used as a reference to calibrate the ^{1}H-^{15}N TROSY spectra. Spectra were processed with NMRPipe (63) and analyzed with Sparky (Goddard TD, Kneller DG, SPARKY 3, University of California, San Francisco).

Small-angle X-ray scattering (SAXS). A flow-cell setup was used for SAXS measurements at the LiX-16-ID beamline at the National Synchrotron Light Source II of the Brookhaven National Laboratory (Upton, NY). A series of scattering images were recorded each with a 2-sec exposure for both protein and buffer samples. Final scattering intensity was reported after buffer subtraction. The X-ray energy was 13 KeV.

Amide hydrogen exchange with mass spectrometry (HX-MS). Purified NTD (4.6 ug/ul) was prepared in 20 mM sodium phosphate buffer (pH 7.4). To initiate the deuterium labeling, 1ul aliquots of purified NTD were added to 19 μl of 10 mM phosphate buffer (pH 7.3) in D_2O. After incubation at room temperature for different time periods (1 min and 5 min), samples

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

were cooled on ice in a pre-calibrated quenching 0.1 M phosphate buffer (pH 2.4) containing 7 M GdnHCl and 0.1 M TCEP. The solution was diluted 10 times with ice-cold 0.05% trifluoroacetic acid and digested for 5 min with agarose immobilized pepsin (Thermo Scientific). Peptic fragments were separated and analyzed by an LTQ Orbitrap XL mass spectrometer (ThermoElectron, San Jose, CA) as described previously (28, 64). As a result, the percent deuterium uptakes were averaged for the triplicate samples of NTD under each condition to obtain the mean percent deuterium incorporation.

Footprinting protection factor (FP-PF) analysis. NTD samples were purified into a buffer of 10 mM sodium phosphate, 0.05mM EDTA and pH 7.4. The protein concentration was adjusted to 5 μM. Samples were exposed to X-ray white beam for 0-50 ms at the 3.2.1 beamline of Advanced Light Source (Berkeley, CA) at ambient temperature, and immediately quenched with 10 mM methionine amide to prevent secondary oxidation. Subsequently, all samples were reduced with 10 mM DTT at 56 °C for 45 minutes and alkylated with 25 mM iodoacetamide at room temperature in the dark for 45 minutes. Two sets of protease digestion were performed at 37 °C using an enzyme-to-protein ratio of 1:10 (w/w). One was with Lys-C (Promega, Inc.) overnight, followed by Asp-N (Promega, Inc.) for 5 hours. The other was with pepsin (Promega, Inc.) for 2 hours. Digestion was terminated by heating samples at 95 °C for 2 min. Identification and quantification of oxidative sites were performed by liquid chromatography-mass spectrometry (LC-MS) analysis using an Orbitrap Elite mass spectrometer (Thermo Scientific, CA) interfaced with a Waters nanoAcquity UPLC system (Waters, MA). The proteolytic peptides (150 ng) were loaded on a trap column (180 μm × 20 mm packed with C18 Symmetry, 5 μm, 100 Å; Waters, MA) to desalt and concentrate peptides. Peptide mixture was eluted on a reverse phase column (75 μm x 250 mm column packed with C18 BEH130, 1.7 μm, 130 Å; Waters, MA) using a gradient of 2 to 35% mobile phase B (0.1% formic acid and acetonitrile) vs. mobile phase A (100% water/0.1 % formic

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

acid) for 60 minutes at 40 °C at a flow rate of 300 nL/min. Eluted peptides were introduced into the nano-electrospray source at a capillary voltage of 2.5 kV.

The extent of oxidation by MS was quantified as a function of X-ray exposure time. The curves were fit to an exponential decay function and the measured rate constants (k_{fp}) were divided by a measure of their intrinsic reactivity with hydroxyl radicals (k_R), thereby providing a residue-level protection factor (PF=k_R/k_{fp}) (24, 56). The PF calculation is available from the website http://www.theyanglab.org/protection.html. The log of the PF values provides an accurate surface topology map, where high logPF values reflect more protection for solvent and lower logPF values reflect greater solvent exposure.

Conformational library. A pool of 35,240 NTD structure candidates was generated from molecular dynamics simulations with an accumulative total time of 35-μs in a 1-ns recording frequency. To enhance conformational sampling, two advanced algorithms were used: Gaussian accelerated molecular dynamics (GaMD) (65) and replica exchange solute tempering (REST2) (66). First, a set of 25 GaMD trajectories (using the software AMBER16, each starting with a random configuration and lasting 1 μs) resulted in a total of 25 μs. Second, a set of 64 replicas in REST2 simulations, ranging from 300 K to 600 K were performed at the Argonne Leadership Computing Facility using the software NAMD as previously described (67, 68). Each replica lasted 160 ns, which resulted in 10 μs. In both GaMD and REST2 simulations, the molecular force field of Amber ff99SB with a TIP3P water model was used (69). Given that the force field for IDPs is still under active development (70-75), the Amber ff99SB was used here for the purpose of pose generation only, with the help from the inclusion of all trajectories from different temperatures and nonequilibrium simulations. Hence, the accuracy of the force field used is beyond the scope here, and these conformations were combined and selected by comparison with multi-technique experimental measurements. To group similar structures together and reduce the computational cost, we performed an RMSD-based clustering

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

analysis as previously described (40, 45). Using a Ca-RMSD cutoff of 5 Å, we obtained a set of N=8,491 clusters of NTD structures from the total of 35,240 simulation snapshots.

Ensemble structure construction. Based on the library of NTD conformations generated above, data from SAXS and FP-PF measurements were combined for ensemble refinement. We selected the ensemble-structures that best-fit experimental SAXS and FP-PF data simultaneously, by a weighted average over the resulting *N* clusters of structures with a probability distribution $\{\mathrm{P}_i\} = P_1, \cdots, P_N$ of each cluster (*i*), based on two unit-less functions χ^2 and φ^2. The χ^2 is defined to measure the difference between the theoretical and experimental SAXS profile by (76),

$$\chi^2(\{\mathrm{P}_i\}) = \frac{1}{N_q}\sum_q \frac{\left\{\log I^{\exp}(q) - \sum_{i=1}^{N}\log[\mathrm{P}_i \cdot I_i^{\mathrm{cal}}(q)]\right\}^2}{\sigma^2(q)}, \qquad (1)$$

where I^{cal} is calculated using f*ast-SAXS-pro* (77) and N_q is the number of scattering *q* points recorded in experimental $I^{\exp}$ (with its measurement error of $\sigma(q)$). Similarly, the φ^2 is the goodness of fit between experimental and theoretical FP-FP data by (45, 78),

$$\varphi^2(\{P_i\}) = \frac{1}{N_s}\sum_s \frac{\left\{\log PF^{\exp}(s) - \sum_{i=1}^{N}\log[P_i \cdot SA_i^{\mathrm{cal}}(s)]\right\}^2}{\delta^2(s)}, \qquad (2)$$

where SA^{cal} is the predicted side-chain SA value of each site, using the linear regression between experimental log$PF^{\exp}$ values (with an error $\delta(s)$ at each site (*s*)) and SA values of N_s sites for each structural cluster (here, N_s=16 as shown in Table S1).

To avoid potential overfitting, we performed a maximum-likelihood principle analysis by adopting a "relative entropy" contribution (46, 79),

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$$S = -\sum_{i=1}^{N} P_i \log P_i, \quad (3)$$

assuming a uniform prior, so the entire optimization of the probability distribution was achieved by minimizing an effective free energy,

$$F(\{P_i\}) = \chi^2 + \beta\varphi^2 - TS, \quad (4)$$

where β is a scaling factor for balancing two unit-less functions χ^2 and φ^2 and T is an effective temperature-like scaling factor. We chose $\beta = 0.1$ and $T = 0.05$ by minimizing $F(\beta, T)$ (see Fig. S5). To reduce the computational cost, we decreased the number of clusters to N=519 by satisfying a cutoff criteria of $\chi^2 < 5$ and a negative linear correlation between logPF and SA for each cluster. Finally, an optimal probability distribution $\{P_i\}$ was determined by numerically minimizing Eq. (4) with the constraints of $0 < P_i < 1$ and $\sum_i P_i = 1$ using a genetic algorithm (80).

Ensemble contact map. The residue-residue interaction (or contact) was calculated using the *contact of structural units* (CSU) approach (81). Each residue was the center of a window segment of 5 neighboring amino acids. Each contact was weighted over the ensemble-structures by the probability of top 100 structures that account for a total of 80% of the entire population.

Surface plasmon resonance. The binding of NTD with TBP was measured in PBSP+ (GE healthcare) buffer, pH7.4 by surface plasmon resonance (SPR) using a Biacore T100 system (GE Healthcare) at 4 °C. S series sensor chip CM4 (GE Healthcare) was preconditioned at 100uL/min with successive 20 μL pulses of 50 mM NaOH, 100 mM HCl, 0.1 mM SDS and 0.085% (v/v) H_3PO_4. NTD was immobilized at low density and TBP concentration series

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ranging from 1.5 to 25 μM were injected over the NTD surface. Data were plotted by Origin 2017 and the affinity constant was calculated by Biaevaluation3 software (GE Healthcare).

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supporting Figures and Table

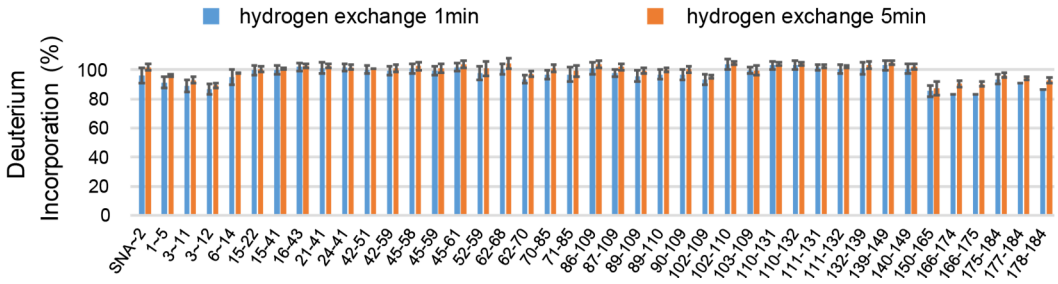

Figure S1. Deuterium incorporation of peptic fragments in HX-MS within 1 min (blue) and 5 min (red) incubation in D_2O. Error bars indicate standard deviation (3 independent experiments).

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

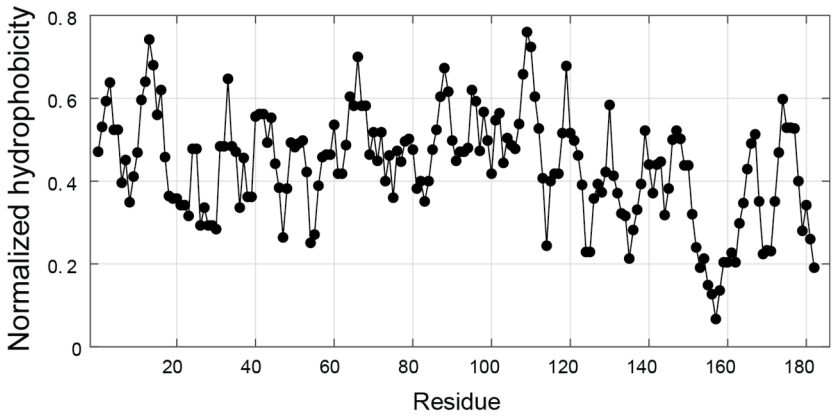

Figure S2. The hydrophobicity of amino acids in the NTD sequence. Each was calculated by the Kyte-Doolittle approximation using the ExPASy server (with a window size of 5 amino acids). The hydrophobicity of individual residues was normalized to a scale of 0 to 1. The mean hydrophobicity <*H*> is defined as the sum of the normalized hydrophobicities of all residues divided by the number of residues in the polypeptide, yielding a <*H*> value at 0.446.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

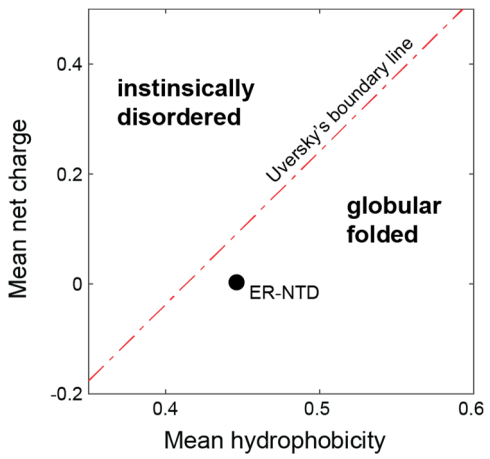

Figure S3. The two-dimensional space of mean hydrophobicity <*H*> and mean net charge <*q*> of globular and intrinsically disordered proteins. The ERα-NTD is close to the Uversky's boundary line of <*H*> = (<*q*> + 1.151)/2.785, and enters the globular-fold regime.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

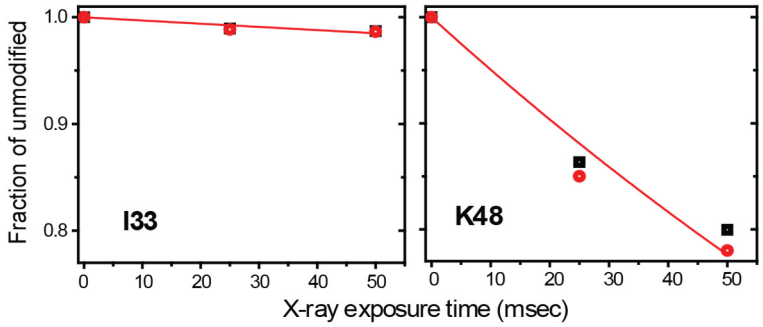

Figure S4. Hydroxyl radical footprinting dose-response curves as a function of X-ray exposure time. Red lines are the least-squares fit to an exponential decay function (i.e., a normalized fraction of unmodified residues), each yielding a rate of footprinting (k_{fp}). The division by an intrinsic reactivity with hydroxyl radicals (k_R) provides a protection factor measure (PF = k_R/k_{fp})..

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

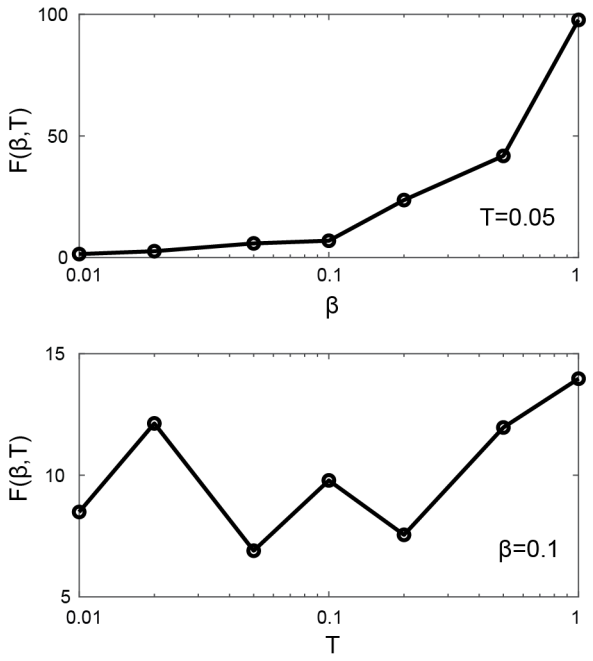

Figure S5. The effective free energy F(β,T) as function of the parameters β and T. The values of β=0.1 and T=0.05 were chosen to give the local minimum of F(β,T) (see Eq. 4).

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

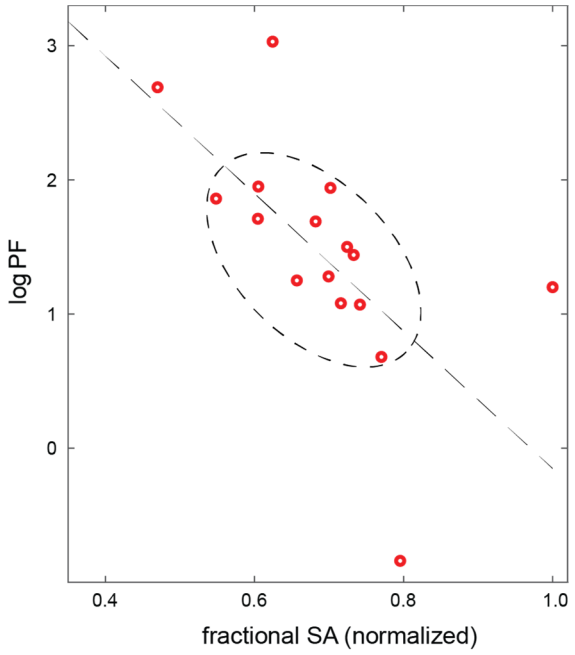

Figure S6. Measured logPF values against the fractional SA values for the set of 16 probed residues. The factional SA value was the SA weighted by each residue's maximum possible accessible area.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

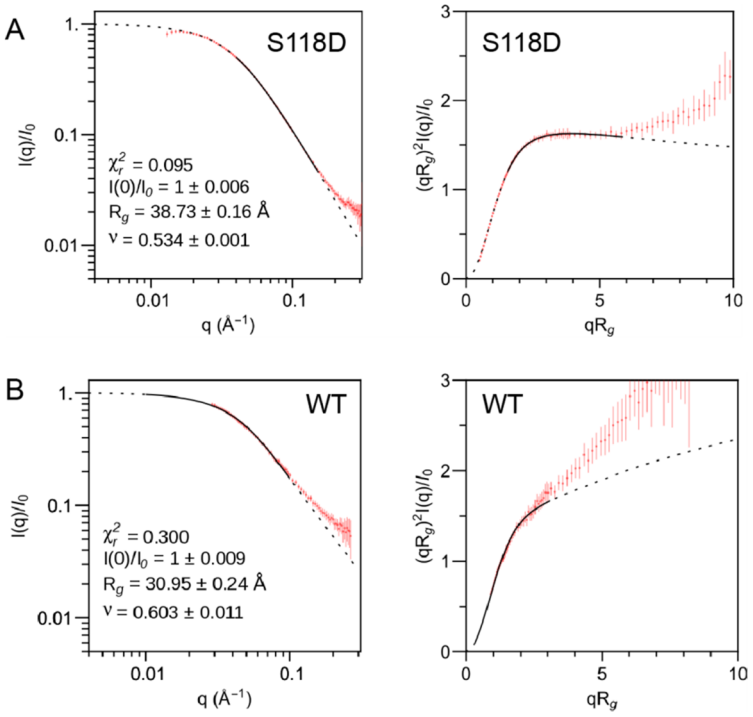

Figure S7. SAXS data and analysis for mutant S118D (A), compared to the wild-type (B). Generated using the webserver at http://sosnick.uchicago.edu/SAXSonIDPs.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

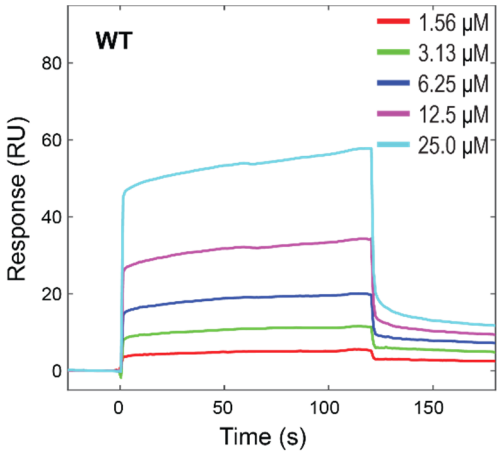

Figure S8. Binding of NTD with TBP measured by surface plasmon resonance (SPR) spectrometry. An apparent binding affinity was estimated at K_d = 53 ± 12 μM.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

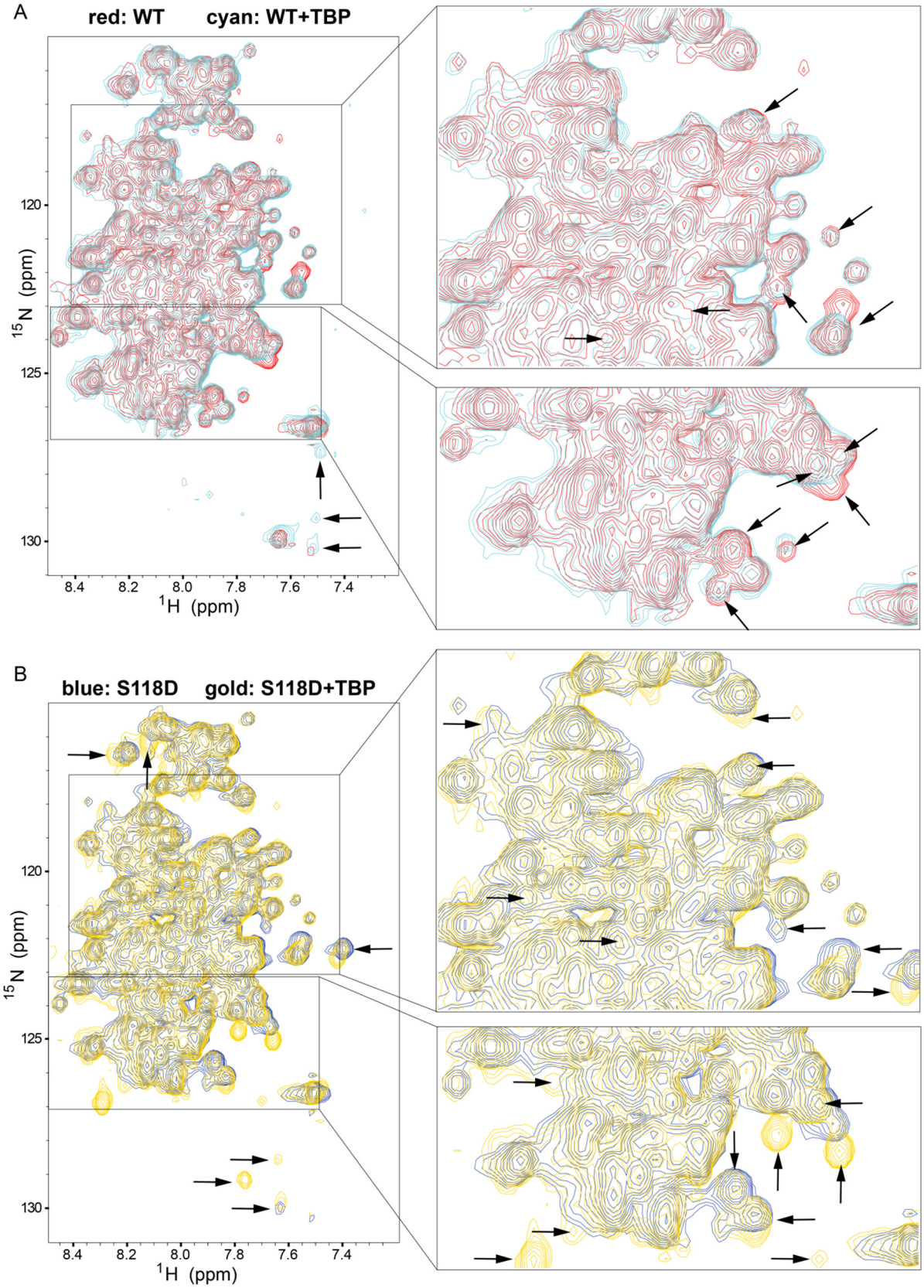

Figure S9. Mutation-induced conformational changes and alteration of coactivator TBP binding.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(A) Overlap of ^{1}H-^{15}N TROSY spectra of the WT in the absence (red) and presence (cyan) of TBP. Change/shift of peaks is marked by arrows.

(B) Overlap of ^{1}H-^{15}N TROSY spectra of S118D in the absence (blue) and presence (gold) of TBP.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

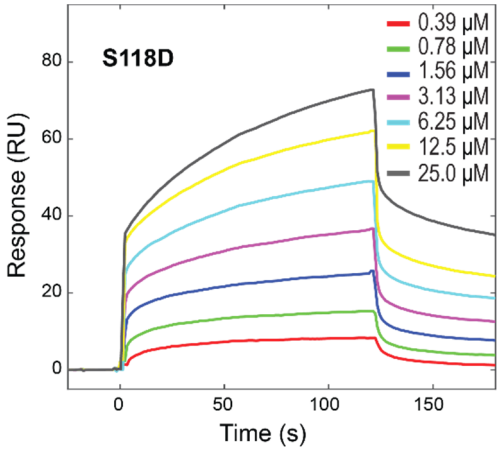

Figure S10. Measured TBP-binding for the S118D mutant using SPR spectrometry. An apparent binding affinity was estimated at K_d = 3.9 ± 0.4 μM.

bioRxiv preprint doi: https://doi.org/10.1101/332205; this version posted May 28, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S1. Hydroxyl radical footprinting protection factors (FP-PFs) for the NTD residues. Listed are the probed residues and their corresponding values for footprinting rates (k_{fp}), logPF, and SA^{pred} predicted from the ensemble-structures.

Residue	k_{fp} (s^{-1})	*log(*PF)	SA^{pred} ($Å^2$)
L26	0.63±0.01	1.94±0.02	79.3
L31	1.26±0.14	1.25±0.11	74.2
I33	0.30±0.03	2.69±0.10	52.0
L35	1.50±0.13	1.08±0.09	80.9
L39	0.81±0.04	1.69±0.05	77.0
L44	0.98±0.20	1.50±0.20	81.8
K48	5.07±0.30	-0.84±0.06	117.6
Y54	1.70±0.15	1.95±0.09	89.4
Y73	2.17±0.09	1.71±0.04	89.3
Y80	1.87±0.092	1.86±0.05	81.0
F89	3.85±0.11	1.07±0.03	100.0
F97	5.68±0.19	0.68±0.03	103.9
F120	2.64±0.11	1.44±0.04	98.9
H124	0.45±0.029	3.03±0.07	78.6
Y130	3.32±0.13	1.28±0.04	103.4
P146	0.30±0.02	1.20±0.07	89.3