Structural characterization of stem cell factors Oct4, Sox2, Nanog and Esrrb disordered domains, and a method to identify their phospho-dependent binding partners.

Bouguechtouli Chafiaa,[a,%] Rania Ghouil,[a] Ania Alik,[a,#] Dingli Florent,[b] Loew Damarys,[b] Theillet Francois-Xavier [a,*]

[a] *Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France.*
[b] *Institut Curie, PSL Research University, Centre de Recherche, CurieCoreTech Spectrométrie de Masse Protéomique, Paris cedex 05, France.*

* Corresponding author. E-mail address: francois-xavier.theillet@cnrs.fr
[%] Present address: Structural Motility, Institut Curie, Paris Université Sciences et Lettres, Sorbonne Université, CNRS UMR144, 75005 Paris, France.
Present address: Université de Paris, Institut Cochin, CNRS UMR8104, INSERM U1016, Paris, France

Keywords: Pluripotency Transcription Factors, Intrinsically Disordered Proteins, Post-Translational Modifications, kinases, NMR, Proteomics, Quantitative Mass-Spectrometry

Abstract

The combined expression of a handful of pluripotency transcription factors (PluriTFs) in somatic cells can generate induced pluripotent stem cells (iPSCs). Here, we report the structural characterization of disordered regions contained in four important PluriTFs, namely Oct4, Sox2, Nanog and Esrrb. Moreover, many post-translational modifications (PTMs) have been detected on PluriTFs, whose roles are not yet characterized. To help in their study, we also present a method i) to produce well-characterized phosphorylation states of PluriTFs, using NMR analysis, and ii) to use them for pull-downs in stem cell extracts analyzed by quantitative proteomics to identify of Sox2 binders.

1. Introduction

The possibility of reprogramming somatic cells to an induced pluripotency state was revealed in the 2000s, carrying great expectations in the fields of Biology and Medicine [1–3]. Induced pluripotent stem cells (iPSCs) and embryonic stem cells (ESCs) are characterized by the active state of a pluripotency network, whose core comprises the pluripotency transcription

factors (PluriTFs) Oct4, Sox2, Nanog and Esrrb (OSNE). These bind to enhancer sequences, and thus activate or repress, or even "bookmark" during mitosis, a wealth of genes related to pluripotency or cell differentiation [4–9]. Consistently, their misregulation correlates with cancer malignancy and stemness [10–14].

Comprehensive structural descriptions of OSNE are missing to our knowledge. The folded DNA-binding domains (DBDs) of OSNE have been structurally characterized, in complex with their DNA target sequences [15–19], together with the ligand-binding domain (LBD) of Esrrb [20]. Recent studies depicted even the splendid structures of Oct4 and Sox2 DBDs binding to nucleosomes, hence deciphering their "pioneer factor" abilities [21–29]. Another structure of Sox2 bound to the importin Impα3 has also been published, showing how its two Nuclear Localization Sequences (NLSs) flanking the DBD are involved in Sox2 nuclear import [30]. The other segments of OSNE have been predicted to be intrinsically disordered regions of proteins (IDRs) [31], i.e. they should have no stable tertiary fold when isolated [32–36].

These IDRs appear to have important roles in binding partners involved in epigenetic reprogramming, chromatin reorganization and in recruiting transcription or repression machineries [5,37–41]. These functions are poorly understood, and, to the best of our knowledge, no experimental characterization of the structural behavior of these regions in N- and C-terminal of DBDs has been released yet. Recent studies have shown that C-terminal regions of Oct4 and Sox2 are important for their reprogramming capacities [42,43], notably by contributing to the engagement in molecular phase-separated condensates with the Mediator complex [44]. More generally, the activating or repressive activities of IDRs of transcription factors (TFs) have been scarcely studied at the structural level: these segments are thought to contain hydrophobic patches flanked by acidic amino acids, which favors DNA-binding specificity, phase separation and low-specificity interactions, notably with the

Mediator subunit Med15 [44–53]; more specific interactions have been described in some cases [45,54,55].

Post-translational modifications (PTMs) add a layer of complexity, by being often responsible of the regulation of IDRs' interactions [32,33,56,57], and notably of TFs activity [58–61]. Post-translational modifications (PTMs) are classical carriers of cell signaling by regulating the stability and the interactions of proteins. An increasing number of PTMs have been described on OSNE IDRs in the recent years [5,37,38,62–72] [73–77], notably phosphorylation by Cyclin-dependent kinases (CDKs) [66,78–87] or by Mitogen-activated protein kinases (MAPKs) [83,88,89], or their complementary Ser/Thr O-GlcNAcylation by OGT [65,90–96]. In order to prompt future studies on this topic, we thought to establish and test an experimental strategy i) for producing well-characterized samples of post-translationally modified IDRs of PluriTFs and ii) to use these as baits in pull-down assays for identifying PTMs' related binding partners.

Hence, we characterized some of the phosphorylation reactions of Esrrb and Sox2 by p38α/β, Erk2 and Cdk1/2. Then, we showed that biotinylated chimera of Sox2 and Esrrb coupled to an AviTag peptide could be attached to streptavidin-coated beads. Finally, we loaded truncated segments of the C-terminal IDR of Sox2 (phosphorylated or not) on these beads, and exposed them to extracts of mouse Embryonic Stem Cells (mESCs) in pull-down assays, which we analyzed using quantitative mass spectrometry-based proteomics. Among the quantified (phospho)-Sox2 binders, we verified the phospho-dependent interaction between the proline cis-trans isomerase Pin1 and Sox2 using NMR spectroscopy.

2. Material and methods

2.1. Production of recombinant fragments of Oct4, Sox2, Nanog and Esrrb

We used human protein sequences, unless specified. Codon-optimized (for expressing in *Escherichia coli*) genes coding for human Oct4(aa1-145) and Oct4(aa286-380) were synthesized in the context of larger genes coding for Tev-Oct4(aa1-145)-Tev-GB1 and Tev-Oct4(aa286-380)-Tev-GB1 by Genscript and cloned into pET-41a(+) vector between SacII and HindIII restriction sites, hence permitting the expression of GST-His6-Tev1-Oct4(aa1-145)-Tev2-GB1 and GST-His6-Tev1-Oct4(aa286-380)-Tev2-GB1; Tev1 and Tev2 are the heptapeptide ENLYFQG cleavage site of the TEV protease, Tev2 is separated by GAGGAGG from GB1 (T2Q variant of the immunoglobulin binding domain B1 of the protein G from group G Streptococcus [97,98]). The C-terminal GB1 tag was added to avoid any C-terminal proteolysis of the IDR of interest during the expression and the first purification steps; we did not test constructs without this supplementary folded domain, whose necessity for the stability of the IDR is thus not proven.

The same rationale (cDNA synthesis, cloning, vectors, chimera constructs) was used for producing Nanog(aa154-305), Nanog(aa154-215), Nanog(aa154-272), Nanog(aa154-305_C185A-C227A-C243A-C251A), Nanog(aa154-272_C185A-C227A-C243A-C251A), and a very similar rationale (chimera constructs missing the C-terminal Tev2-GB1) for Sox2(aa1-42), Sox2(aa115-317_C265A), Sox2(aa115-187), Sox2(aa115-236), Sox2(aa115-282_C265A), Esrrb(aa1-102_C12A-C72A-C91A), Esrrb(aa1-102_C12A-C91A), Nanog(aa1-85) (this latter was cloned in the MfeI/HindIII restriction sites from pET-41a(+)).

The recombinant production and the purification of the protein constructs followed the procedures described previously [99], using the soluble fraction of bacterial lysates, except for the constructs containing the Sox2 C-terminal fragments. These latter constructs were recovered from the insoluble fractions of the lysates, and resolubilized in 8 M urea; these were submitted to a His-Trap purification in urea, and the last size-exclusion chromatography (SEC) had to be carried out in 2 M urea, which avoided clogging of the column and permitted

to obtain regular elution peak widths (these were otherwise extremely broad, up to 100 mL for the longest Sox2(aa115-317_C265A) construct). The samples were concentrated and stored at -20 °C, and thawed just before the NMR experiments. The Sox2 samples containing 2 M urea were submitted to 2-3 cycles of concentration/dilution in Hepes at 20 mM, NaCl at 75 mM to generate samples in urea at 0.25 or 0.125 M.

All purification steps were carefully carried out at 4 °C; protein eluates from every purification step were immediately supplemented with protease inhibitors (EDTA-free cOmplete, Roche) (together with DTT at 10 mM for cysteine-containing protein constructs), before being submitted to a concentration preparing the next purification step.

Chimera constructs of Sox2 and Esrrb IDR fragments containing a 15-mer peptide AviTag GLNDIFEAQKIEWHE were produced using procedures similar to those described earlier for OSNE constructs. The construct Sox2(aa234-317)-AviTag-His6 was soluble, and did not require to be purified in urea.

More details about the production of OSNE peptides are given in the Supplementary Material.

2.2. Production of the biotin ligase BirA and specific biotinylation of the AviTag-peptide chimera

The biotin ligase BirA was produced using recombinant production in *E. coli* BL21(DE3)Star transformed with a pET21-a(+) plasmid containing a gene coding for BirA cloned at EcoRI and HindIII restriction sites. pET21a-BirA was a gift from Alice Ting (Addgene plasmid # 20857)[100]. The expression was carried out over night at 20 °C in a Luria-Bertani culture medium. The construct was containing a His6 tag in C-terminal and was purified using a two-step purification procedure including a His-trap followed by a SEC. Details about the production of BirA are given in the Supplementary Material.

The biotinylation was executed using a rationale inspired from a published protocol [101], at room temperature during 90 minutes, in samples containing the AviTag-chimera of interest at 100 µM and BirA at 0.7 µM in a buffer containing ATP at 2 mM, biotin at 600 µM, $MgCl_2$ at 5 mM, DTT at 1 mM, HEPES at 50 mM, NaCl at 150 mM, protease inhibitors (EDTA-free cOmplete, Roche), at pH 7.0. To remove some eventual proteolyzed peptides and BirA, the biotinylated constructs were purified using a SEC in a column (Superdex 16/60 75 pg, Cytiva) preequilibrated with a buffer containing phosphate at 20 mM, NaCl at 150 mM at pH=7.4 (buffer called thereafter Phosphate Buffer Saline, PBS). The eluted fractions of interest were concentrated and stored at -20 °C.

2.3. Assignment of NMR signals from OSNE fragments, and structural propensities

The assignment strategy was the same than in previous reports from our laboratory [99]. The ^{15}N relaxation data were recorded and analyzed according to the methods described in previous reports [102]. Details are given in the Supplementary Material.

Disorder prediction were calculated using the ODINPred website (https://st-protein.chem.au.dk/odinpred) [103]. Experimental secondary structure propensities of unmodified OSNE peptides were obtained using the neighbor-corrected structural propensity calculator ncSCP [104,105] (http://www.protein-nmr.org/, https://st-protein02.chem.au.dk/ncSPC/) from the experimentally determined, DSS referenced Cα and Cβ chemical shifts as input, with a correction for Gly-Pro motifs (-0.77 ppm instead of -2.0 ppm) [106]. Some signals were too weak in 3D-spectra from Sox2(aa115-317_C265A) recorded at 950 MHz, and their chemical shifts were not defined. In these cases, chemical shifts from 3D-spectra of Sox2(115-236) or His6-AviTag-Sox2(234-317_C265A) were used to complete the lists of chemical shifts used to calculate the chemical shift propensities shown in Figure 2.

2.4. NMR monitoring of phosphorylation reactions and production of phosphorylated peptides

We performed the phosphorylation kinetics presented in Figure 4a using commercial recombinant kinases GST-p38β at 10 µg/mL (Sigma-Aldrich, ref. B4437), GST-Erk2 at 20 µg/mL (Sigma-Aldrich, ref. E1283), GST-Cdk1/CyclinA2 at 20 µg/mL (Sigma-Aldrich, stock ref. C0244) and GST-Cdk2/CyclinA2 at 20 µg/mL (Sigma-Aldrich, ref. C0495). Then, we used kinases produced in house in *E. coli*, using plasmids containing optimized genes coding for p38α(aa1-360, full-length) and Erk2(aa8-360); these were produced, activated and purified in house as described previously [99]; in-house p38α was used at 40 µg/mL for the experiments shown in Figure 4.

Phosphorylation reactions were carried out using ^{15}N-labeled IDRs at 50 µM, in Hepes 20 mM, NaCl 50 mM, DTT or TCEP at 4 mM, ATP 1.5 mM, $MgCl_2$ at 5 mM, protease inhibitors (Roche), 7.5% D_2O, pH6.8 at 25°C in 100 µL using 3 mm diameter Shigemi tubes. We monitored the phosphorylation kinetics by recording time series of ^{1}H-^{15}N SOFAST-HMQC spectra at 600 or 700 MHz, and quantifying the NMR signal intensities of the disappearing unphospho- and the appearing phospho-residues. We applied the methods that we described progressively in earlier publications [107–110]. More details are given in the Supplementary Material.

2.5. Pull-down assays

The extracts from mouse Embryonic Stem Cells (mESCs) were obtained from mESCs cultured in the conditions previously described [111]. Homogeneous extracts were obtained using DNA shearing by sonication in presence of benzonase, as described by Gingras and colleagues [112].

The pull-down assays were executed using 25 µL of streptavidin-coated magnetic beads (Magbeads streptavidine, Genscript) loaded with 1 nmol of the biotinylated bait-peptides of interest. These were incubated during one hour at room temperature with mESCs extracts, washed in PBS and eluted using a 2x Laemmli buffer. Details are given in the Supplementary Material.

2.6. Mass spectrometry-based proteomics analysis of pull-down assays

The pull-down samples were treated on-beads by trypsin/LysC (Promega). The resulting peptides were loaded and separated on a C18 columns for online liquid chromatography performed with an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific). Maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.6 Da for MS/MS peaks. The resulting files were further processed using myProMS v3.9.3 (https://github.com/bioinfo-pf-curie/myproms; Poullet et al., 2007 [2]). False-discovery rate (FDR) was calculated using Percolator [3] and was set to 1% at the peptide level for the whole study. Label-free quantification was performed using peptide extracted ion chromatograms (XICs), computed with MassChroQ [4] v.2.2.1. The complete details are given in the Supplementary Material.

2.7. Recombinant production of Pin1 and NMR analysis of its interaction with Sox2 or phospho-Sox2

The plasmid containing the gene coding for the Pin1-WW domain was a kind gift from Isabelle Landrieu. The production was executed according to the previously published protocol [113]. The NMR analysis of binding with phosphoSox2(aa115-240) were performed at 283K and pH 7.0 with the GST-Pin1-WW construct and ^{15}N-labeled Sox2(aa115-240) mixed in stoichiometric proportions, either at 50 or 10 µM for non-phospho and

phosphoSox2, respectively. The details on the NMR acquisition, processing and analysis are given in the Supplementary Information.

3. Results

3.1. Structural characterization of the N- and C-terminal regions of Oct4

We produced and purified protein constructs containing the fragments of human Oct4(aa1-145) and Oct4(aa286-360), which were both predicted to be mostly disordered (Supp. Mat 3.2.3.). The 2D ^{1}H-^{15}N HSQC NMR spectra showed crosspeaks in the spectra region where random coil peptides resonances are usually found (Figure 1). We assigned the backbone NMR signals of $^{1}H_N$, ^{15}N, $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO for both segments, which permitted to calculate experimentally derived secondary structure propensities from the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts. We did not find any sign of a stable secondary structure, the highest α-helical propensities reaching about 25 % in short stretches of about 5 consecutive amino acids. Hence, we verified experimentally that these N- and C-terminal fragments of human Oct4 are IDRs.

We can notice that Oct4 IDRs contain a high density of Prolines, which are not directly detectable in the present $^{1}H_N$-detected experiments, even though most of the $^{13}C\alpha$, $^{13}C\beta$ and resonances were characterized via HNCAB and HNCO experiments. We have shown previously that the ^{13}C-detected experiments $^{13}C\alpha^{13}CO$ permitted to observe all these Pro residues in Oct4(aa1-145) [99], whose chemical shifts were those of random coil peptides.

3.2. Structural characterization of the N- and C-terminal regions of Sox2

We produced and purified peptide fragments of human Sox2, namely Sox2(aa1-42), Sox2(aa115-187), Sox2(aa115-236), Sox2(aa115-282), Sox2(aa234-317_C265A), and Sox2(aa115-317_C265A). We also produced and purified chimera peptides His6-AviTag-

Sox2(aa115-240) and His6-AviTag-Sox2(aa234-317_C265A). These were all predicted to be disordered (Supp. Mat 3.3.3.).

We had solubility issues with all of them but Sox2(aa1-42), Sox2(aa234-317_C265A) and His6-AviTag-Sox2(aa234-317_C265A). We had to recover these troublesome peptides from the insoluble fraction of the bacterial extract, after overexpression at 37 °C. We had even to carry out our final size-exclusion chromatography (SEC) purification step in a buffer containing urea at 2 M (at 4 °C) for Sox2(aa115-236), Sox2(aa115-282), His6-AviTag-Sox2(aa115-240), and Sox2(aa115-317-C265A). The assignments of these latter constructs were achieved in 0.25-0.5 M urea, after executing 2 to 3 concentration-dilution steps. Aggregates were forming during the acquisition, which made the assignment rather painful. This behavior correlated with liquid-liquid phase separation (LLPS) propensities, which we observed a few months before such a behavior was reported by Young and collaborators [44]. The assignment of Sox2(aa115-317-C265A) was possible only at 950 MHz with the help of the previously assigned smaller fragments Sox2(aa115-236) and His6-AviTag-Sox2(aa234-317_C265A). Some stretches of amino acids were particularly difficult to observe in 3D spectra, e.g. the region aa160-185, because of an apparent fast T2 relaxation. We may investigate these phenomena in later reports.

We observed crosspeaks in the 2D ^{1}H-^{15}N HSQC NMR spectra that were all resonating in the spectral region of random coil peptides resonances (Figure 2). The spectra of the short fragments of the C-terminal region of Sox2 were exactly overlapping with those of Sox2(aa115-317_C265A) (Supp. Fig. 1). This shows that all these fragments have very similar local conformational behaviors, a phenomenon regularly observed with IDRs. We assigned the backbone NMR signals of $^{1}H_N$, ^{15}N, $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO for Sox2(aa1-42), Sox2(aa115-236), His6-AviTag-Sox2(aa234-317_C265A) and partially for Sox2(aa115-317_C265A). We aggregated the lists of chemical shifts of the C-terminal fragments, and

used them to calculate the experimental secondary structure propensities from the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts. This confirmed the absence of any stable secondary structure elements in Sox2 N- and C-terminal region. The C-terminal region is poorly soluble below 0.25 M urea; this should not affect a stable fold, so we can affirm that these regions of Sox2 are experimentally proven IDRs.

3.3. Structural characterization of the N-terminal region of Nanog

We produced and purified the N-terminal peptide fragment of human Nanog(aa1-85). All our attempts to purify C-terminal regions of Nanog failed, even after alanine-mutation of cysteines in Nanog(aa154-305), Nanog(aa154-272) and Nanog(aa154-215). We managed to resolubilize our construct GST-His6-Tev-Nanog(aa154-305) from the insoluble fraction of the bacterial extract, to partially purify it and cleave it using the TEV protease. However, the resulting Nanog(aa154-305) peptide was barely soluble in a detergent (NP-40 at 2% v/v), and not in high salt buffers, or not even in presence of urea at 4 M. The 10 Tryptophane residues are probably playing a role in this behavior, in the context of a primary structure containing not enough hydrophobic amino acids favoring stable folds.

The crosspeaks of Nanog(aa1-85) in the 2D ^{1}H-^{15}N HSQC NMR spectrum were all in the spectral region of random coil peptides resonances (Figure 3). The assignment of the backbone NMR signals of $^{1}H_N$, ^{15}N, $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO permitted to calculate experimental secondary structure propensities, which were low through the whole peptide. This Nanog N-terminal is thus confirmed to be an IDR.

3.4. Structural characterization of the N-terminal region of Esrrb

We produced and purified the N-terminal fragment of human Esrrb(aa1-102), which was predicted to be disordered (Supp Mat. 3.5.3.), in the alanine-mutated versions Esrrb(aa1-

102_C12A-C72A-C92A), Esrrb(aa1-102_C12A-C92A). This was a strategic choice to attenuate the formation of disulfide bonds; the wild-type N-terminal fragments might however be workable too. Mutating cysteines permitted to work in more comfortable conditions and to maintain our construct monomeric for longer periods of time in the next phosphorylation and biotinylation experiments. Cysteines are indeed highly solvent-accessible in IDRs and they are consequently difficult to keep in their thiol, non-disulfide forms, even in presence of fresh DTT or TCEP at neutral pH. We also produced chimera constructs Esrrb(aa1-102_C12A-C72A-C92A)-AviTag-His6 and Esrrb(aa1-102_C12A-C72A)-AviTag-His6.

All the 2D ^{1}H-^{15}N HSQC NMR spectra revealed crosspeaks in the spectral region of random coil peptides (Figure 3). These spectra are overlapping to a large extent, confirming the weak influence of the mutations of cysteines: the mutation Cys72Ala has almost no consequences on the chemical shifts, below 0.05 ppm even for the neighboring amino acids (Supp. Fig 2a,b); the mutation Cys91Ala has more impact, with chemical shifts perturbations of about 0.1 ppm for the next 5 amino acids (Supp. Fig. 2c,d), which is at least partially due to the fact that the Ala substitution favors an increase of local α-helicity (about 25 %, see Supp. Fig. 2e,f). This N-terminal fragment of human Esrrb is thus an IDR, according to the calculated secondary structure propensities (Figure 3f, Supp. Fig. 2e,f).

3.5. Phosphorylation of Esrrb and Sox2 by p38α, Erk2, Cdk1/2 as monitored by NMR spectroscopy

We reported recently the site-specific phosphorylation kinetics of Oct4 by p38α using ^{13}C-direct NMR detection [99]. Here, we used more standard ^{1}H-detected/^{15}N-filtered experiments to rapidly characterize the site preferences of MAPKs and CDKs on Esrrb and Sox2, which we thought to use as baits for performing phospho-dependent pull-down assays (see below).

To start with, we used commercial aliquots of MAPKs, namely p38β and Erk2, and CDKs, namely Cdk1/CyclinA2 and Cdk2/CyclinA2, on Esrrb(aa1-102). We observed the progressive phosphorylation of its three Ser-Pro motifs, i.e. at Ser22, Ser34 and Ser58, in agreement with the consensus motifs of these kinases [114,115]. Ser22 is the preferred target in all cases, while Ser34 is the least processed by CDKs, if at all: the commercial CDKs are poorly active in our hands, which we have verified with a number of other targets for years in the laboratory; this makes it difficult to distinguish between sites that are only mildly disfavored or those that are more stringently ignored by CDKs in NMR monitored assays.

The limited activities and high costs of commercial kinases motivated us to develop in-house capacities in kinase production. p38α was the most accessible to produce among the MAPKs and CDKs; we produced it and activated it using recombinant MKK6. We tested our home-made p38α on His6-AviTag-Sox2(aa115-240) and His6-AviTag-Sox2(aa234-317). It phosphorylated all the Ser/Thr-Pro motifs of these two peptides, and also T306 in a $PGT_{306}AI$ context, which shows a favorable Proline in position -2 [116], and a less common S212 in a $MTS_{212}SQ$ context.

Hence, we were able to generate AviTag-IDR chimera in well-defined phosphorylated states. To produce phosphorylated ^{14}N-AviTag-IDR dedicated to pull-down assays, we executed the same protocol, and monitored in parallel "identical", but ^{15}N-labeled samples by NMR. Hence, we could no the exact phosphorylation status of the ^{14}N-AviTag-IDR for the next experiments.

3.6. Structural characterization of AviTag-peptide chimera and biotinylated versions thereof

We aimed at identifying new partners of OSNE using pull-down assays. We thought to use chimera containing GST at the N-terminus, which appeared as a convenient approach: vectors

integrating a GST-coding DNA sequence for overexpression in *E. coli* are available and of common use; glutathione-coated beads are also accessible and permit efficient and specific binding of GST-containing chimera peptides. However, we were unsatisfied by the performances of the method: GST binding to glutathione-coated beads is slow at low temperature (necessary to avoid IDR proteolysis); moreover, it appeared that GST-IDRs chimera are hampered by the IDR "molecular-cloud" and are even weaker and slower to bind the beads. Our attempts to bind GST-IDRs to the beads were thus resulting in poor yields, which were not very reproducible. In the context of our aims, i.e. to establish a method allowing semi-quantitative detection of IDR binding partners, this unsatisfying lack of reproducibility was only promising supplementary variable parameters.

Thus, we decided to switch to another strategy: the use of the specifically biotinylated 15-mer peptide tag called AviTag [101]. This is efficiently and specifically biotinylated by the biotin ligase BirA (Figure 5d), which permits the high-affinity binding to streptavidin-coated beads. We designed AviTag-IDR chimera, with the AviTag in N-terminal position for Sox2-IDRs constructs, and in C-terminal for Essrb-IDR constructs. We characterized the AviTag and its impact on the IDRs of interest using NMR: the AviTag is unfolded and it does not affect the Sox2 and Esrrb fragments, according to the observed negligible chemical shift perturbations (Fig 5) – the GST-Tag provoked also only very weak chemical shift changes on Essrb(aa1-102) (Fig S3). The biotinylated AviTag peptide appears to be slightly less mobile than the common IDPs on the ps-ns timescale, according to the heteronuclear $^{1}H^{15}N$-nOes (Figure 5b). We observed that the biotinylation of the AviTag provokes weak, but distinguishable chemical shift perturbations for the close neighbors of the biotinylated lysine, but had no effect on the peptides of interest (Figure 5c). It generated also the appearance of a HN-ester NMR signal, similar to that of acetylated lysine [107,117]. Hence, we could quantify and monitor the reaction advancement using NMR, and determine the incubation time that was

necessary and sufficient to obtain a complete biotinylation of our chimera AviTag-IDRs (see 2.2.). This was one among many optimization steps permitting the production of sufficient quantities of intact IDRs for the pull-down assays.

Next, we tested the binding of the biotinylated AviTag-IDRs on streptavidin-coated beads. This produced very satisfying results, i.e. stoichiometric, specific binding in one hour with no leakage (Fig. S4). This approach was thus selected for the pull-down assays.

3.7. Identification of Sox2 binding partners in extracts from mouse Embryonic Stem Cells

We prepared the four peptides AviTag-Sox2(aa115-240) and AviTag-Sox2(aa234-317) in their non-phosphorylated and phosphorylated versions, using p38α to execute the phosphorylation reactions (Figure 6a). These peptides were also biotinylated, and later bound to streptavidin-coated beads, which we used as baits for pull-down assays in extracts from mESCs (Figure 6b). Importantly, size-exclusion chromatographies were carried out between every step to discard proteolyzed peptides, the enzymes (kinases of BirA) and their contaminants as much as possible. We performed the pull-down assays with the four samples in parallel with the same cell extract, in duplicate, and then analyzed the bound fractions using quantitative LC-MS/MS analysis (see Supp. Mat 1.6. for the full description). Hence, we could identify and evaluate the relative quantities of proteins retained by the four AviTag-peptides (Figure 6). On paper, this presents the important advantage of removing false-positive binders, which can interact unspecifically with the streptavidin-coated beads.

We present here results that should be interpreted carefully: we produced only duplicates for every condition, using one single cell extract. To deliver trustful information, the common standards in the field recommend 3 to 5 replicates, and the use of independent samples. We can still comment the results, which we consider to be interesting preliminary data. We quantified proteins with at least 2 distinct peptides across the two replicates (Figure 6C).

Among the relative quantifications with an adjusted p-value < 0.05, we observed a two-fold change or more (FC>2) of a number of transcription factors in the proteins pulled out by AviTag-Sox2(aa115-240), among which the PluriTFs Oct4 and Klf5 are significantly enriched (>4 peptides, p-value<0.02). Most of these TFs appear to bind slightly less to the phosphorylated pSox2(aa115-240). However, pSox2(aa115-240) kept apparently a capacity to bind Oct4 and Klf5 to some extent, in comparison to pSox2(aa234-317). Also, we found out that pSox2(aa115-240) was pulling out the three isoforms of CK1 (>3 peptides, p-values<10^{-5}) and the proline isomerase Pin1 (>2 peptides, p-values<2.10^{-4}), while Sox2(aa115-240) and pSox2(aa234-317) did not or much less.

3.8. NMR characterization of the interaction between Pin1 and phospho-Sox2

We decided to test the interaction between pSox2(aa115-240) and Pin1. We recorded ^{1}H-^{15}N NMR spectra of ^{15}N-labeled Sox2(aa115-240) or pSox2(aa115-240) alone or in presence of the Pin1-WW domain (natural abundance peptide, i.e. 0.6% ^{15}N, 99.4% ^{14}N, hence "NMR invisible" in ^{15}N-filtered experiments). We observed localized losses in signal intensities for the residues neighboring the three pSox2(aa115-240) phosphosites when mixed with Pin1 (Figure 7b); in contrast, no significant differences showed up in the spectra obtained with non-phosphorylated Sox2(aa115-240) in absence or presence of Pin1-WW (Figure 7a). Hence, these signal losses are the typical signs of a position-specific interaction between an IDR and a folded protein in the intermediate or slow NMR time-scale, i.e. µs-s timescale, which corresponds to submicromolar affinities for this type of molecules. This shows that our pull-down assays were capable of isolating and identifying binding partners of an IDR of Sox2 in this range of affinities.

4. Discussion

The structural biochemistry analysis reported here can be applied to the broad field of transcription factors (TFs). These are essential actors of cell signaling: they are key elements for inducing or maintaining pluripotency or differentiation, for cell proliferation or cell-cycle arrest, by activating or repressing gene transcription [61,118]. About 90% of the ~1,600 TFs contain large disordered segments (> 30 consecutive amino acids), which is particularly true for PluriTFs [31,119], and this correlation between TFs and IDRs exist in all kingdoms of life [45,120]. IDRs of TFs recruit transcription co-factors or the transcription machinery, which is still not very well characterized in detail [44–55].

Indeed, the fine understanding of TFs interactions via their IDRs appears to be hampered by the nature of these interactions: weak affinities, multivalency, possible redundancy between TFs and coacervation; post-translational modifications (PTMs), which can switch on or off IDRs' interactions, are a supplementary source of confusion when searching for binding partners. Among the difficulties, we should also mention the basic biochemistry issues: IDRs are difficult produce and manipulate *in vitro*, because they are prone to degrade or aggregate. Here, we have tried to demonstrate the feasibility and the interest of some biochemical and spectroscopic approaches to better characterize IDRs of TFs, their phosphorylation and the associated binding partners.

Like other TFs, pluripotency TFs Oct4, Sox2, Nanog and Esrrb are post-translationally modified (see the introduction), notably by CDKs and MAPKs [66,78–89]. These two classes of kinases are fundamental actors in all aspects of eukaryotic cellular life, and understanding their activity and regulation in pluripotency or differentiation is of high significance. Interesting questions are still pending: what is the phosphorylation status of OSNEs' IDRs in pluripotent cells, what is the impact on their interaction networks, and how does it affect pluripotency or differentiation? The inhibition of MAPK Erk signaling is necessary to maintain pluripotency in the standard culture conditions of ESCs and iPSCs [66,82–84], while

these cells show a high CDK activity [8,86,87]; these two kinases family have the same core consensus sites, i.e. Ser/Thr-Pro motifs, which are abundant in OSNEs' IDRs and whose phosphorylation has apparently consequences for initiating differentiation [73,78,80,83,85]. We have shown that we could produce well-defined phosphorylation status of these peptides, using recombinant kinases and NMR analysis, which makes it possible to study their interactions *in vitro*. We have also demonstrated our capacity to use these peptides as baits in pull-down assays for detecting new binding partners.

Using this approach, we have identified an interaction between phospho-Sox2(aa115-240)-pS212-pS220-pT232 and the prolyl cis-trans isomerase Pin1. Interactions between Pin1 and Oct4 or Nanog have been published earlier [81,121]. This does not represent a major surprise, because Pin1 recognize the consensus motif pSer/pThr-Pro, which is very degenerate (about one third of all the detected sites in phosphoproteomics) [122,123]; dozens of Pin1 binding partners have been reported, which leaves suspicions about the functionality of all these, and about the capacity of the intracellular pool of Pin1 molecuels to engage in so many interactions. We also detected an affinity between phospho-Sox2(aa115-240)-pS212-pS220-pT232 and three CK1 isoforms. This capacity of CK1 kinases to recognize prephosphorylated in position -3 of a potential Ser/Thr is well-known (with a hydrophobic residue in +1), and a recent structural characterization of CK1δ has shown that it was even capable of keeping some affinity for a phosphorylated product after the phosphate transfer [124].

Hence, we feel that these results are at the same time satisfying, but do reveal degenerate interactions, whose biological significance is questionable. This corresponds to one of the major drawbacks in the field of IDRs' studies: they generate interactions of weak affinities, which can be easily released during the washes of our pull-down assays. In this regard, the « proximity labeling » approaches (BioID, APEX and their derivates) appeared recently to be quite adapted to transient interactions: these methods, developed in the last ten years, use

chimera constructs containing enzymes that transfer chemical groups to their intracellular neighbors, which can be later identified by mass-spectrometry [125–128]. IDRs are very flexible, solvent-exposed and establish a lot of poorly specific transient interactions. This was raising concerns about the possible production of many false-positive if one used proximity labeling methods to detect IDRs' binding partners. This has been partially confirmed by a recent study, but this bias appears to be limited [129] . Interactomes of 109 TFs have actually been described using Bio-ID and affinity-purification MS, showing the complementarity between proximity-labeling and the pull-down approach proposed here [130]. Yeast double-hybrid, which can detect ~20 µM affinity interactions, and novel phage display approaches will also help in this task [131–134].

Another difficulty in studying IDRs of TFs is their propensity to coacervate [44–53]. Here, we have tried to use Sox2 as a bait in pull-down assays, a protein that has been later recognized to favor liquid-liquid phase separation [44,135]. We met this difficulty during the production steps, which forced us to purify most of the Sox2 constructs in urea at 2 M. We could straightforwardly observe liquid-liquid phase separation of Sox2(aa115-317) at 4 µM using DIC microscopy in presence of Ficoll-70 (data not shown), but also progressive aggregation and low solubility thresholds while working with our purified samples. These are clear limiting factors for sample production and NMR characterization, which will hamper a number of other studies on IDRs of TFs. This might also affect the results of pull-down assays: we noticed an enrichment in TFs in the samples obtained from pull-downs using Sox2(aa115-240) as a bait, which has a much higher coacervation propensity than Sox2(aa234-317). Is it possible to generate local surface liquid-liquid phase separation on the surface of streptavidin-coated beads? This might be at the same time a bless and a curse for future studies, by helping the formation of biologically significant assemblies, or by favoring unspecific, non-native macromolecules interactions.

A final bottleneck in the studies of these IDRs is the capacity to produce post-translationally modified samples. The commercial enzymes are not very well adapted to our NMR studies, because of the required quantities. Here, we have used in-house production of the kinase p38α. Since we carried out the present work, we have developed our capacities in producing activated Erk2, Cdk2/CyclinA1 and Cdk1/CyclinB1. These will be part of our future studies. TFs are indeed quite adapted to NMR investigations: they are 300 to 500 residues long and contain large IDRs (~100 amino acids) separating small folded domains (also ~100 amino acids) [31,45]. Their structural characterization would permit to understand a number of cell signaling mechanisms at the atomic scale, and possibly to identify new therapeutic targets, even though this class of proteins is notoriously difficult to inhibit [60,136,137].

5. Conclusion

We have applied NMR techniques to carry out a primary analysis of the pluripotency transcription factors Oct4, Sox2, Nanog and Esrrb, in particular of their intrinsically disordered regions. We have shown experimentally that they did not adopt a stable fold when isolated, and that we were able to conduct a residue-specific analysis. This relies on the delicate production and purification of these peptides, which are prone to proteolysis and aggregation; producing them in a well-defined post-translational modification status was an even arduous challenge. We have demonstrated the feasibility of these tasks using recombinant kinases and NMR analysis. We have also evaluated the usefulness of such protein constructs as baits in pull-down assays to identify new binding partners of IDRs. These characterizations and the associated methods provide firm basis for future investigations on transcription factors.

Declaration of Competing Interest

The authors declare that they have no competing interest.

Fundings

This work was supported by the CNRS and the CEA-Saclay, by the French Infrastructure for Integrated Structural Biology (https://frisbi.eu/, grant number ANR-10-INSB-05-01, Acronym FRISBI) and by the French National Research Agency (ANR; research grants ANR-14-ACHN-0015 and ANR-20-CE92-0013). Financial support from the IR INFRANALYTICS FR2054 for conducting the research is also gratefully acknowledged. This work was also supported by grants from the "Région Ile-de-France" and Fondation pour la Recherche Médicale (D.L. and LSMP).

Acknowledgements

We thank Thaleia Papadopoulou, Amandine Molliex and Navarro Pablo for providing mouse Embryonic Stem Cells (mESCs) extracts, and for fruitful discussions. We thank Nadia Izadi-Pruneyre for providing the bench space necessary to carry out the pull-down experiments with fresh mESCs extracts. We thank the IR-RMN (now Infranalytics), notably Nelly Morellet, François Giraud and Ewen Lescop, for their reactivity, their support, and their long-standing, efficient care of the 950 MHz spectrometer. We thank Marie Sorin, Baptiste Nguyen and Benjamin Bacri, who contributed to this study during their Master1-Master2 internships.

Supplementary material

Supporting information for this article is available online at …

References

[1] H. Inoue, N. Nagata, H. Kurokawa, S. Yamanaka, iPS cells: a game changer for future medicine, EMBO J. 33 (2014) 409–417. https://doi.org/10.1002/embj.201387098.

[2] Y. Shi, H. Inoue, J.C. Wu, S. Yamanaka, Induced pluripotent stem cell technology: a decade of progress, Nat Rev Drug Discov. 16 (2017) 115–130.

https://doi.org/10.1038/nrd.2016.245.
[3] R.G. Rowe, G.Q. Daley, Induced pluripotent stem cells in disease modelling and drug discovery, Nat Rev Genet. 20 (2019) 377–388. https://doi.org/10.1038/s41576-019-0100-z.
[4] W. Deng, E.C. Jacobson, A.J. Collier, K. Plath, The transcription factor code in iPSC reprogramming, Curr. Opin. Genet. Dev. 70 (2021) 89–96. https://doi.org/10.1016/j.gde.2021.06.003.
[5] T.W. Theunissen, R. Jaenisch, Molecular Control of Induced Pluripotency, Stem Cell. 14 (2014) 720–734. https://doi.org/10.1016/j.stem.2014.05.002.
[6] M. Li, J.C.I. Belmonte, Ground rules of the pluripotency gene regulatory network, Nat. Rev. Genet. 18 (2017) 180–191. https://doi.org/10.1038/nrg.2016.156.
[7] C. Chronis, P. Fiziev, B. Papp, S. Butz, G. Bonora, S. Sabri, J. Ernst, K. Plath, Cooperative Binding of Transcription Factors Orchestrates Reprogramming, Cell. 168 (2017) 442-459.e20. https://doi.org/10.1016/j.cell.2016.12.016.
[8] N. Festuccia, I. Gonzalez, N. Owens, P. Navarro, Mitotic bookmarking in development and stem cells, Development. 144 (2017) 3633–3645. https://doi.org/10.1242/dev.146522.
[9] I. Gonzalez, A. Molliex, P. Navarro, Mitotic memories of gene activity, Curr. Opin. Cell Biol. 69 (2021) 41–47. https://doi.org/10.1016/j.ceb.2020.12.009.
[10] P. Mu, Z. Zhang, M. Benelli, W.R. Karthaus, E. Hoover, C.-C. Chen, J. Wongvipat, S.-Y. Ku, D. Gao, Z. Cao, N. Shah, E.J. Adams, W. Abida, P.A. Watson, D. Prandi, C.-H. Huang, E. de Stanchina, S.W. Lowe, L. Ellis, H. Beltran, M.A. Rubin, D.W. Goodrich, F. Demichelis, C.L. Sawyers, SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer, Science. 355 (2017) 84–88. https://doi.org/10.1126/science.aah4307.
[11] A.C. Hepburn, R.E. Steele, R. Veeratterapillay, L. Wilson, E.E. Kounatidou, A. Barnard, P. Berry, J.R. Cassidy, M. Moad, A. El-Sherif, L. Gaughan, I.G. Mills, C.N. Robson, R. Heer, The induction of core pluripotency master regulators in cancers defines poor clinical outcomes and treatment resistance, Oncogene. 38 (2019) 4412–4424. https://doi.org/10.1038/s41388-019-0712-y.
[12] S. Mirzaei, M.D.A. Paskeh, M. Entezari, S. reza Mirmazloomi, A. Hassanpoor, M. Aboutalebi, S. Rezaei, E.S. Hejazi, A. Kakavand, H. Heidari, S. Salimimoghadam, A. Taheriazam, M. Hashemi, S. Samarghandian, SOX2 function in cancers: Association with growth, invasion, stemness and therapy response, Biomed. Pharmacother. 156 (2022) 113860. https://doi.org/10.1016/j.biopha.2022.113860.
[13] E.-H. Ervin, R. French, C.-H. Chang, S. Pauklin, Inside the stemness engine: Mechanistic links between deregulated transcription factors and stemness in cancer, Semin Cancer Biol. 87 (2022) 48–83. https://doi.org/10.1016/j.semcancer.2022.11.001.
[14] A. Chaudhary, S.S. Raza, R. Haque, Transcriptional factors targeting in cancer stem cells for tumor modulation, Seminars in Cancer Biology. 88 (2023) 123–137. https://doi.org/10.1016/j.semcancer.2022.12.010.
[15] D. Esch, J. Vahokoski, M.R. Groves, V. Pogenberg, V. Cojocaru, H. vom Bruch, D. Han, H.C.A. Drexler, M.J. Araúzo-Bravo, C.K.L. Ng, R. Jauch, M. Wilmanns, H.R. Schöler, A unique Oct4 interface is crucial for reprogramming to pluripotency, Nat Cell Biol. 15 (2013) 295–301. https://doi.org/10.1038/ncb2680.
[16] A. Reményi, K. Lins, L.J. Nissen, R. Reinbold, H.R. Schöler, M. Wilmanns, Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers, Genes Dev. 17 (2003) 2048–2059. https://doi.org/10.1101/gad.269303.
[17] M.D. Gearhart, S.M.A. Holmbeck, R.M. Evans, H.J. Dyson, P.E. Wright, Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface

mediates extended half-site recognition., J. Mol. Biol. 327 (2003) 819–832. https://doi.org/10.1016/s0022-2836(03)00183-9.

[18] R. Jauch, C.K.L. Ng, K.S. Saikatendu, R.C. Stevens, P.R. Kolatkar, Crystal Structure and DNA Binding of the Homeodomain of the Stem Cell Transcription Factor Nanog, J. Mol. Biol. 376 (2008) 758–770. https://doi.org/10.1016/j.jmb.2007.11.091.

[19] Y. Hayashi, L. Caboni, D. Das, F. Yumoto, T. Clayton, M.C. Deller, P. Nguyen, C.L. Farr, H.-J. Chiu, M.D. Miller, M.-A. Elsliger, A.M. Deacon, A. Godzik, S.A. Lesley, K. Tomoda, B.R. Conklin, I.A. Wilson, S. Yamanaka, R.J. Fletterick, Structure-based discovery of NANOG variant with enhanced properties to promote self-renewal and reprogramming of pluripotent stem cells, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) 4666–4671. https://doi.org/10.1073/pnas.1502855112.

[20] B. Yao, S. Zhang, Y. Wei, S. Tian, Z. Lu, L. Jin, Y. He, W. Xie, Y. Li, Structural Insights into the Specificity of Ligand Binding and Coactivator Assembly by Estrogen-Related Receptor β, J Mol Biol. 432 (2020) 5460–5472. https://doi.org/10.1016/j.jmb.2020.08.007.

[21] S.O. Dodonova, F. Zhu, C. Dienemann, J. Taipale, P. Cramer, Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function, Nature. 580 (2020) 669–672. https://doi.org/10.1038/s41586-020-2195-y.

[22] A.K. Michael, R.S. Grand, L. Isbel, S. Cavadini, Z. Kozicka, G. Kempf, R.D. Bunker, A.D. Schenk, A. Graff-Meyer, G.R. Pathare, J. Weiss, S. Matsumoto, L. Burger, D. Schübeler, N.H. Thomä, Mechanisms of OCT4-SOX2 motif readout on nucleosomes, Science. 368 (2020) 1460–1465. https://doi.org/10.1126/science.abb0074.

[23] K. Echigoya, M. Koyama, L. Negishi, Y. Takizawa, Y. Mizukami, H. Shimabayashi, A. Kuroda, H. Kurumizaka, Nucleosome binding by the pioneer transcription factor OCT4, Sci Rep. 10 (2020) 11832. https://doi.org/10.1038/s41598-020-68850-1.

[24] G.A. Roberts, B. Ozkan, I. Gachulincová, M.R. O'Dwyer, E. Hall-Ponsele, M. Saxena, P.J. Robinson, A. Soufi, Dissecting OCT4 defines the role of nucleosome binding in pluripotency, Nat Cell Biol. 23 (2021) 834–845. https://doi.org/10.1038/s41556-021-00727-5.

[25] E. Morgunova, J. Taipale, Structural insights into the interaction between transcription factors and the nucleosome, Curr. Opin. Struct. Biol. 71 (2021) 171–179. https://doi.org/10.1016/j.sbi.2021.06.016.

[26] W. Kagawa, H. Kurumizaka, Structural basis for DNA sequence recognition by pioneer factors in nucleosomes, Current Opinion in Structural Biology. 71 (2021) 59–64. https://doi.org/10.1016/j.sbi.2021.05.011.

[27] E. Luzete-Monteiro, K.S. Zaret, Structures and consequences of pioneer factor binding to nucleosomes, Current Opinion in Structural Biology. 75 (2022) 102425. https://doi.org/10.1016/j.sbi.2022.102425.

[28] B.D. Sunkel, B.Z. Stanton, Pioneer factors in development and cancer, IScience. 24 (2021) 103132. https://doi.org/10.1016/j.isci.2021.103132.

[29] F.C.M. Gadea, E.N. Nikolova, Structural Plasticity of Pioneer Factor Sox2 and DNA Bendability Modulate Nucleosome Engagement and Sox2-Oct4 Synergism, J. Mol. Biol. 435 (2023) 167916. https://doi.org/10.1016/j.jmb.2022.167916.

[30] B. Jagga, M. Edwards, M. Pagin, K.M. Wagstaff, D. Aragão, N. Roman, J.D. Nanson, S.R. Raidal, N. Dominado, M. Stewart, D.A. Jans, G.R. Hime, S.K. Nicolis, C.F. Basler, J.K. Forwood, Structural basis for nuclear import selectivity of pioneer transcription factor SOX2, Nat Commun. 12 (2021) 28. https://doi.org/10.1038/s41467-020-20194-0.

[31] B. Xue, C.J. Oldfield, Y.-Y. Van, A.K. Dunker, V.N. Uversky, Protein intrinsic disorder and induced pluripotent stem cells, Mol. BioSyst. 8 (2012) 134–150. https://doi.org/10.1039/C1MB05163F.

[32] M.M. Babu, The contribution of intrinsically disordered regions to protein function,

cellular complexity, and human disease, Biochemical Society Transactions. 44 (2016) 1185–1200. https://doi.org/10.1042/BST20160172.
[33] P.E. Wright, H.J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation., Nat. Struct. Mol. Biol. 16 (2015) 18–29. https://doi.org/10.1038/nrm3920.
[34] D. Piovesan, M. Necci, N. Escobedo, A.M. Monzon, A. Hatos, I. Mičetić, F. Quaglia, L. Paladin, P. Ramasamy, Z. Dosztanyi, W.F. Vranken, N.E. Davey, G. Parisi, M. Fuxreiter, S.C.E. Tosatto, MobiDB: intrinsically disordered proteins in 2021., Nucleic Acids Res. 49 (2021) D361–D367. https://doi.org/10.1093/nar/gkaa1058.
[35] P. Kulkarni, S. Bhattacharya, S. Achuthan, A. Behal, M.K. Jolly, S. Kotnala, A. Mohanty, G. Rangarajan, R. Salgia, V. Uversky, Intrinsically Disordered Proteins: Critical Components of the Wetware, Chem. Rev. 122 (2022) 6614–6633. https://doi.org/10.1021/acs.chemrev.1c00848.
[36] D. Piovesan, A. Del Conte, D. Clementel, A.M. Monzon, M. Bevilacqua, M.C. Aspromonte, J.A. Iserte, F.E. Orti, C. Marino-Buslje, S.C.E. Tosatto, MobiDB: 10 years of intrinsically disordered proteins, Nucleic Acids Res. 51 (2023) D438–D444. https://doi.org/10.1093/nar/gkac1065.
[37] Y. Buganim, D.A. Faddah, R. Jaenisch, Mechanisms and models of somatic cell reprogramming, Nat. Struct. Mol. Biol. 14 (2013) 427–439. https://doi.org/10.1038/nrg3473.
[38] A. Rizzino, Concise Review: The Sox2-Oct4 Connection: Critical Players in a Much Larger Interdependent Network Integrated at Multiple Levels, Stem Cells. 31 (2013) 1033–1039. https://doi.org/10.1002/stem.1352.
[39] S. Jerabek, F. Merino, H.R. Schöler, V. Cojocaru, OCT4: Dynamic DNA binding pioneers stem cell pluripotency, Biochim Biophys Acta Gene Regul Mech. 1839 (2014) 138–154. https://doi.org/10.1016/j.bbagrm.2013.10.001.
[40] A. Saunders, F. Faiola, J. Wang, Concise Review: Pursuing Self-Renewal and Pluripotency with the Stem Cell Factor Nanog, Stem Cells. 31 (2013) 1227–1236. https://doi.org/10.1002/stem.1384.
[41] S.E. Bondos, A.K. Dunker, V.N. Uversky, Intrinsically disordered proteins play diverse roles in cell signaling, Cell Commun Signal. 20 (2022) 20. https://doi.org/10.1186/s12964-022-00821-7.
[42] K.-P. Kim, Y. Wu, J. Yoon, K. Adachi, G. Wu, S. Velychko, C.M. MacCarthy, B. Shin, A. Röpke, M.J. Arauzo-Bravo, M. Stehling, D.W. Han, Y. Gao, J. Kim, S. Gao, H.R. Schöler, Reprogramming competence of OCT factors is determined by transactivation domains, Sci. Adv. 6 (2020) eaaz7364. https://doi.org/10.1126/sciadv.aaz7364.
[43] I. Aksoy, R. Jauch, V. Eras, W.A. Chng, J. Chen, U. Divakar, C.K.L. Ng, P.R. Kolatkar, L.W. Stanton, Sox Transcription Factors Require Selective Interactions with Oct4 and Specific Transactivation Functions to Mediate Reprogramming, Stem Cells. 31 (2013) 2632–2646. https://doi.org/10.1002/stem.1522.
[44] A. Boija, I.A. Klein, B.R. Sabari, A. Dall'Agnese, E.L. Coffey, A.V. Zamudio, C.H. Li, K. Shrinivas, J.C. Manteiga, N.M. Hannett, B.J. Abraham, L.K. Afeyan, Y.E. Guo, J.K. Rimel, C.B. Fant, J. Schuijers, T.I. Lee, D.J. Taatjes, R.A. Young, Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains, Cell. (2018) 1–31. https://doi.org/10.1016/j.cell.2018.10.042.
[45] L. Staby, C. O'Shea, M. Willemoës, F. Theisen, B.B. Kragelund, K. Skriver, Eukaryotic transcription factors: paradigms of protein intrinsic disorder, Biochem. J. 474 (2017) 2509–2532. https://doi.org/10.1042/BCJ20160631.
[46] C.N. Ravarani, T.Y. Erkina, G. De Baets, D.C. Dudman, A.M. Erkine, M.M. Babu, High-throughput discovery of functional disordered regions: investigation of transactivation domains, Mol Syst Biol. 14 (2018) e8190-14. https://doi.org/10.15252/msb.20188190.
[47] S. Brodsky, T. Jana, K. Mittelman, M. Chapal, D.K. Kumar, M. Carmi, N. Barkai,

Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity, Mol. Cell. 79 (2020) 459-471.e4. https://doi.org/10.1016/j.molcel.2020.05.032.
[48] A. Erijman, L. Kozlowski, S. Sohrabi-Jahromi, J. Fishburn, L. Warfield, J. Schreiber, W.S. Noble, J. Söding, S. Hahn, A High-Throughput Screen for Transcription Activation Domains Reveals Their Sequence Features and Permits Prediction by Deep Learning, Molecular Cell. 78 (2020) 890-902.e6. https://doi.org/10.1016/j.molcel.2020.04.020.
[49] G. Næs, J.O. Storesund, P. Udayakumar, M. Ledsaak, O.S. Gabrielsen, Dissecting the transactivation domain (tAD) of the transcription factor c-Myb to assess recent models of tAD function, FEBS Open Bio. 10 (2020) 2329–2342. https://doi.org/10.1002/2211-5463.12978.
[50] A.L. Sanborn, B.T. Yeh, J.T. Feigerle, C.V. Hao, R.J. Townshend, E. Lieberman Aiden, R.O. Dror, R.D. Kornberg, Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator, ELife. 10 (2021) e68068. https://doi.org/10.7554/eLife.68068.
[51] L.M. Tuttle, D. Pacheco, L. Warfield, D.B. Wilburn, S. Hahn, R.E. Klevit, Mediator subunit Med15 dictates the conserved "fuzzy" binding mechanism of yeast transcription activators Gal4 and Gcn4, Nat Commun. 12 (2021) 2220. https://doi.org/10.1038/s41467-021-22441-4.
[52] L.F. Soto, Z. Li, C.S. Santoso, A. Berenson, I. Ho, V.X. Shen, S. Yuan, J.I. Fuxman Bass, Compendium of human transcription factor effector domains, Mol. Cell. 82 (2022) 514–526. https://doi.org/10.1016/j.molcel.2021.11.007.
[53] M.V. Staller, E. Ramirez, S.R. Kotha, A.S. Holehouse, R.V. Pappu, B.A. Cohen, Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains, Cell Systems. 13 (2022) 334-345.e5. https://doi.org/10.1016/j.cels.2022.01.002.
[54] B. Bourgeois, T. Gui, D. Hoogeboom, H.G. Hocking, G. Richter, E. Spreitzer, M. Viertler, K. Richter, T. Madl, B.M.T. Burgering, Multiple regulatory intrinsically disordered motifs control FOXO4 transcription factor binding and function, Cell Rep. 36 (2021) 109446. https://doi.org/10.1016/j.celrep.2021.109446.
[55] K. Teilum, J.G. Olsen, B.B. Kragelund, On the specificity of protein–protein interactions in the context of disorder, Biochemical Journal. 478 (2021) 2035–2050. https://doi.org/10.1042/BCJ20200828.
[56] F.-X. Theillet, A. Binolfi, T. Frembgen-Kesner, K. Hingorani, M. Sarkar, C. Kyne, C. Li, P.B. Crowley, L. Gierasch, G.J. Pielak, A.H. Elcock, A. Gershenson, P. Selenko, Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs)., Chem. Rev. 114 (2014) 6661–6714. https://doi.org/10.1021/cr400695p.
[57] N.E. Davey, The functional importance of structure in unstructured protein regions, Curr. Opin. Struct. Biol. 56 (2019) 155–163. https://doi.org/10.1016/j.sbi.2019.03.009.
[58] T.M. Filtz, W.K. Vogel, M. Leid, Regulation of transcription factor activity by interconnected post- translational modifications, Trends Pharmacol. Sci. 35 (2014) 76–85. https://doi.org/10.1016/j.tips.2013.11.005.
[59] D. Han, M. Huang, T. Wang, Z. Li, Y. Chen, C. Liu, Z. Lei, X. Chu, Lysine methylation of transcription factors in cancer, Cell Death Dis. 10 (2019) 290. https://doi.org/10.1038/s41419-019-1524-2.
[60] M. Qian, F. Yan, T. Yuan, B. Yang, Q. He, H. Zhu, Targeting post-translational modification of transcription factors as cancer therapy, Drug Discov Today. 25 (2020) 1502–1512. https://doi.org/10.1016/j.drudis.2020.06.005.
[61] P. Weidemüller, M. Kholmatov, E. Petsalaki, J.B. Zaugg, Transcription factors: Bridge between cell signaling and gene regulation, Proteomics. 21 (2021) 2000034. https://doi.org/10.1002/pmic.202000034.
[62] N. Cai, M. Li, J. Qu, G.-H. Liu, J.C. Izpisua Belmonte, Post-translational modulation

of pluripotency, J. Mol. Cell Biol. 4 (2012) 262–265. https://doi.org/10.1093/jmcb/mjs031.
[63] L. Fang, L. Zhang, W. Wei, X. Jin, P. Wang, Y. Tong, J. Li, J.X. Du, J. Wong, A Methylation-Phosphorylation Switch Determines Sox2 Stability and Function in ESC Maintenance or Differentiation, Mol. Cell. 55 (2014) 537–551. https://doi.org/10.1016/j.molcel.2014.06.018.
[64] D.S. Yoon, Y. Choi, Y. Jang, M. Lee, W.J. Choi, S.-H. Kim, J.W. Lee, SIRT1 Directly Regulates SOX2 to Maintain Self-Renewal and Multipotency in Bone Marrow-Derived Mesenchymal Stem Cells, Stem Cells. 32 (2014) 3219–3231. https://doi.org/10.1002/stem.1811.
[65] H. Jang, T.W. Kim, S. Yoon, S.-Y. Choi, T.-W. Kang, S.-Y. Kim, Y.-W. Kwon, E.-J. Cho, H.-D. Youn, O-GlcNAc Regulates Pluripotency and Reprogramming by Directly Acting on Core Components of the Pluripotency Network, Stem Cell. 11 (2012) 62–74. https://doi.org/10.1016/j.stem.2012.03.001.
[66] J. Brumbaugh, Z. Hou, J.D. Russell, S.E. Howden, P. Yu, A.R. Ledvina, J.J. Coon, J.A. Thomson, Phosphorylation regulates human OCT4., Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 7162–7168. https://doi.org/10.1073/pnas.1203874109.
[67] S. Dan, B. Kang, X. Duan, Y.-J. Wang, A cell-free system toward deciphering the post-translational modification barcodes of Oct4 in different cellular contexts, Biochem. Biophys. Res. Commun. 456 (2015) 714–720. https://doi.org/10.1016/j.bbrc.2014.12.043.
[68] Y. Cho, H.G. Kang, S.-J. Kim, S. Lee, S. Jee, S.G. Ahn, M.J. Kang, J.S. Song, J.-Y. Chung, E.C. Yi, K.-H. Chun, Post-translational modification of OCT4 in breast cancer tumorigenesis, Cell Death Differ. (2018) 1781–17951. https://doi.org/10.1038/s41418-018-0079-6.
[69] C.A.C. Williams, A. Soufi, S.M. Pollard, Post-translational modification of SOX family proteins: key biochemical targets in cancer?, Semin Cancer Biol. 67 (2019) 30–38. https://doi.org/10.1016/j.semcancer.2019.09.009.
[70] X. Abulaiti, H. Zhang, A. Wang, N. Li, Y. Li, C. Wang, X. Du, L. Li, Phosphorylation of Threonine343 Is Crucial for OCT4 Interaction with SOX2 in the Maintenance of Mouse Embryonic Stem Cell Pluripotency, Stem Cell Rep. 9 (2017) 1630–1641. https://doi.org/10.1016/j.stemcr.2017.09.001.
[71] D.K. Kim, B. Song, S. Han, H. Jang, S.-H. Bae, H.Y. Kim, S.-H. Lee, S. Lee, J.K. Kim, H.-S. Kim, K.-M. Hong, B.I. Lee, H.-D. Youn, S.-Y. Kim, S.W. Kang, H. Jang, Phosphorylation of OCT4 Serine 236 Inhibits Germ Cell Tumor Growth by Inducing Differentiation, Cancers. 12 (2020) 2601. https://doi.org/10.3390/cancers12092601.
[72] N.P. Mullin, J. Varghese, D. Colby, J.M. Richardson, G.M. Findlay, I. Chambers, Phosphorylation of NANOG by casein kinase I regulates embryonic stem cell self-renewal, Febs Lett. 595 (2021) 14–25. https://doi.org/10.1002/1873-3468.13969.
[73] K.T.G. Rigbolt, T.A. Prokhorova, V. Akimov, J. Henningsen, P.T. Johansen, I. Kratchmarova, M. Kassem, M. Mann, J.V. Olsen, B. Blagoev, System-Wide Temporal Characterization of the Proteome and Phosphoproteome of Human Embryonic Stem Cell Differentiation, Sci. Signal. 4 (2011). https://doi.org/10.1126/scisignal.2001570.
[74] Y. Kamachi, H. Kondoh, Sox proteins: regulators of cell fate specification and differentiation., Development. 140 (2013) 4129–4144. https://doi.org/10.1242/dev.091793.
[75] J. Shin, T.W. Kim, H. Kim, H.J. Kim, M.Y. Suh, S. Lee, H.-T. Lee, S. Kwak, S.-E. Lee, J.-H. Lee, H. Jang, E.-J. Cho, H.-D. Youn, Aurkb/PP1-mediated resetting of Oct4 during the cell cycle determines the identity of embryonic stem cells., Elife. 5 (2016) e10877. https://doi.org/10.7554/elife.10877.
[76] P.N. Malak, B. Dannenmann, A. Hirth, O.C. Rothfuss, K. Schulze-Osthoff, Novel AKT phosphorylation sites identified in the pluripotency factors OCT4, SOX2 and KLF4., Cell Cycle. 14 (2015) 3748–3754. https://doi.org/10.1080/15384101.2015.1104444.

[77] T. Schaefer, C. Lengerke, SOX2 protein biochemistry in stemness, reprogramming, and cancer: the PI3K/AKT/SOX2 axis and beyond, Oncogene. 39 (2020) 278–292. https://doi.org/10.1038/s41388-019-0997-x.

[78] J. Ouyang, W. Yu, J. Liu, N. Zhang, L. Florens, J. Chen, H. Liu, M. Washburn, D. Pei, T. Xie, Cyclin-dependent Kinase-mediated Sox2 Phosphorylation Enhances the Ability of Sox2 to Establish the Pluripotent State, J. Biol. Chem. 290 (2015) 22782–22794. https://doi.org/10.1074/jbc.M115.658195.

[79] S. Lim, A. Bhinge, S. Bragado Alonso, I. Aksoy, J. Aprea, C.F. Cheok, F. Calegari, L.W. Stanton, P. Kaldis, Cyclin-Dependent Kinase-Dependent Phosphorylation of Sox2 at Serine 39 Regulates Neurogenesis, Mol. Cell. Biol. 37 (2017) e00201-17–24. https://doi.org/10.1128/MCB.00201-17.

[80] H.J. Kim, J. Shin, S. Lee, T.W. Kim, H. Jang, M.Y. Suh, J.-H. Kim, I.-Y. Hwang, D.S. Hwang, E.-J. Cho, H.-D. Youn, Cyclin-dependent kinase 1 activity coordinates the chromatin associated state of Oct4 during cell cycle in embryonic stem cells, Nucleic Acids Res. 46 (2018) 6544–6560. https://doi.org/10.1093/nar/gky371.

[81] M. Moretto-Zita, H. Jin, Z. Shen, T. Zhao, S.P. Briggs, Y. Xu, Phosphorylation stabilizes Nanog by promoting its interaction with Pin1., Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 13312–13317. https://doi.org/10.1073/pnas.1005847107.

[82] S.-H. Kim, M.-O. Kim, Y.-Y. Cho, K. Yao, D.J. Kim, C.-H. Jeong, D.H. Yu, K.B. Bae, E.-J. Cho, S.K. Jung, M.H. Lee, H. Chen, J.Y. Kim, A.M. Bode, Z. Dong, ERK1 phosphorylates Nanog to regulate protein stability and stem cell self-renewal, Stem Cell Research. 13 (2014) 1–11. https://doi.org/10.1016/j.scr.2014.04.001.

[83] J. Brumbaugh, J.D. Russell, P. Yu, M.S. Westphall, J.J. Coon, J.A. Thomson, NANOG Is Multiply Phosphorylated and Directly Modified by ERK2 and CDK1 In Vitro, Stem Cell Rep. 2 (2014) 18–25. https://doi.org/10.1016/j.stemcr.2013.12.005.

[84] A. Saunders, D. Li, F. Faiola, X. Huang, M. Fidalgo, D. Guallar, J. Ding, F. Yang, Y. Xu, H. Zhou, J. Wang, Context-Dependent Functions of NANOG Phosphorylation in Pluripotency and Reprogramming, Stem Cell Reports. 8 (2017) 1115–1123. https://doi.org/10.1016/j.stemcr.2017.03.023.

[85] L. Liu, W. Michowski, H. Inuzuka, K. Shimizu, N.T. Nihira, J.M. Chick, N. Li, Y. Geng, A.Y. Meng, A. Ordureau, A. Kołodziejczyk, K.L. Ligon, R.T. Bronson, K. Polyak, J.W. Harper, S.P. Gygi, W. Wei, P. Sicinski, G1 cyclins link proliferation, pluripotency and differentiation of embryonic stem cells, Nat Cell Biol. 19 (2017) 177–188. https://doi.org/10.1038/ncb3474.

[86] L. Liu, W. Michowski, A. Kolodziejczyk, P. Sicinski, The cell cycle in stem cell proliferation, pluripotency and differentiation, Nat Cell Biol. 21 (2019) 1060–1067. https://doi.org/10.1038/s41556-019-0384-4.

[87] S. Jirawatnotai, S. Dalton, M. Wattanapanitch, Role of cyclins and cyclin-dependent kinases in pluripotent stem cells and their potential as a therapeutic target, Semin. Cell Dev. Biol. 107 (2020) 63–71. https://doi.org/10.1016/j.semcdb.2020.05.001.

[88] R. Spelat, F. Ferro, F. Curcio, Serine 111 Phosphorylation Regulates OCT4A Protein Subcellular Distribution and Degradation, J. Biol. Chem. 287 (2012) 38279–38288. https://doi.org/10.1074/jbc.m112.386755.

[89] K.B. Bae, D.H. Yu, K.Y. Lee, K. Yao, J. Ryu, D.Y. Lim, T.A. Zykova, M.-O. Kim, A.M. Bode, Z. Dong, Serine 347 Phosphorylation by JNKs Negatively Regulates OCT4 Protein Stability in Mouse Embryonic Stem Cells, Stem Cell Rep. 9 (2017) 2050–2064. https://doi.org/10.1016/j.stemcr.2017.10.017.

[90] Y. Hao, X. Fan, Y. Shi, C. Zhang, D. Sun, K. Qin, W. Qin, W. Zhou, X. Chen, Next-generation unnatural monosaccharides reveal that ESRRB O-GlcNAcylation regulates pluripotency of mouse embryonic stem cells, Nat. Commun. (2019) 1–13.

https://doi.org/10.1038/s41467-019-11942-y.
[91] N.S. Sharma, V.K. Gupta, P. Dauer, K. Kesh, R. Hadad, B. Giri, A. Chandra, V. Dudeja, C. Slawson, S. Banerjee, S.M. Vickers, A. Saluja, S. Banerjee, O-GlcNAc modification of Sox2 regulates self-renewal in pancreatic cancer by promoting its stability, Theranostics. 9 (2019) 3410–3424. https://doi.org/10.7150/thno.32615.
[92] D.K. Kim, J.-S. Lee, E.Y. Lee, H. Jang, S. Han, H.Y. Kim, I.-Y. Hwang, J.-W. Choi, H.M. Shin, H.J. You, H.-D. Youn, H. Jang, O-GlcNAcylation of Sox2 at threonine 258 regulates the self-renewal and early cell fate of embryonic stem cells, Exp Mol Med. 53 (2021) 1759–1768. https://doi.org/10.1038/s12276-021-00707-7.
[93] S. Constable, J.-M. Lim, K. Vaidyanathan, L. Wells, O-GlcNAc transferase regulates transcriptional activity of human Oct4, Glycobiology. 27 (2017) 927–937. https://doi.org/10.1093/glycob/cwx055.
[94] T. Miura, S. Nishihara, The Functions of O-GlcNAc in Pluripotent Stem Cells, (2019).
[95] L. Ciraku, E.M. Esquea, M.J. Reginato, O-GlcNAcylation regulation of cellular signaling in cancer, Cell. Signal. 90 (2022) 110201. https://doi.org/10.1016/j.cellsig.2021.110201.
[96] J. Ma, C. Hou, C. Wu, Demystifying the O-GlcNAc Code: A Systems View, Chem. Rev. 122 (2022) 15822–15864. https://doi.org/10.1021/acs.chemrev.1c01006.
[97] A.M. Gronenborn, D.R. Filpula, N.Z. Essig, A. Achari, M. Whitlow, P.T. Wingfield, G.M. Clore, A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G., Nature. 253 (1991) 657–661.
[98] C.K. Smith, J.M. Withka, L. Regan, A Thermodynamic Scale for the .beta.-Sheet Forming Tendencies of the Amino Acids, Biochemistry. 33 (1994) 5510–5517. https://doi.org/10.1021/bi00184a020.
[99] A. Alik, C. Bouguechtouli, M. Julien, W. Bermel, R. Ghouil, S. Zinn-Justin, F.-X. Theillet, Sensitivity-Enhanced 13C-NMR Spectroscopy for Monitoring Multisite Phosphorylation at Physiological Temperature and pH, Angew. Chem. Int. Ed. 59 (2020) 10411–10415. https://doi.org/10.1002/anie.202002288.
[100] M. Howarth, K. Takao, Y. Hayashi, A.Y. Ting, Targeting quantum dots to surface proteins in living cells with biotin ligase, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 7583–7588. https://doi.org/10.1073/pnas.0503125102.
[101] M. Fairhead, M. Howarth, Site-specific biotinylation of purified proteins using BirA., Methods Mol. Biol. 1266 (2015) 171–184. https://doi.org/10.1007/978-1-4939-2272-7_12.
[102] F.-X. Theillet, A. Binolfi, B. Bekei, A. Martorana, H.M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb, P. Selenko, Structural disorder of monomeric α-synuclein persists in mammalian cells, Nature. 530 (2016) 45–50. https://doi.org/10.1038/nature16531.
[103] R. Dass, F.A.A. Mulder, J.T. Nielsen, ODiNPred: comprehensive prediction of protein order and disorder, Sci Rep. 10 (2020) 14780. https://doi.org/10.1038/s41598-020-71716-1.
[104] K. Tamiola, F.A.A. Mulder, Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins, Biochem. Soc. Trans. 40 (2012) 1014–1020. https://doi.org/10.1042/BST20120171.
[105] J.T. Nielsen, F.A.A. Mulder, POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins., J. Biomol. NMR. 70 (2018) 141–165. https://doi.org/10.1007/s10858-018-0166-5.
[106] W. Borcherds, F.-X. Theillet, A. Katzer, A. Finzel, K.M. Mishall, A.T. Powell, H. Wu, W. Manieri, C. Dieterich, P. Selenko, A. Loewer, G.W. Daughdrill, Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells, Nat Chem Biol. 10 (2014) 1000–1002. https://doi.org/10.1038/nchembio.1668.
[107] F.-X. Theillet, C. Smet-Nocca, S. Liokatis, R. Thongwichian, J. Kosten, M.-K. Yoon,

R.W. Kriwacki, I. Landrieu, G. Lippens, P. Selenko, Cell signaling, post-translational protein modifications and NMR spectroscopy, J. Biomol. NMR. 54 (2012) 217–236. https://doi.org/10.1007/s10858-012-9674-x.

[108] F.-X. Theillet, H.M. Rose, S. Liokatis, A. Binolfi, R. Thongwichian, M. Stuiver, P. Selenko, Site-specific NMR mapping and time-resolved monitoring of serine and threonine phosphorylation in reconstituted kinase reactions and mammalian cell extracts, Nat. Protoc. 8 (2013) 1416–1432. https://doi.org/10.1038/nprot.2013.083.

[109] A. Mylona, F.-X. Theillet, C. Foster, T.M. Cheng, F. Miralles, P.A. Bates, P. Selenko, R. Treisman, Opposing effects of Elk-1 multisite phosphorylation shape its response to ERK activation., Science. 354 (2016) 233–237. https://doi.org/10.1126/science.aad1872.

[110] M. Julien, C. Bouguechtouli, A. Alik, R. Ghouil, S. Zinn-Justin, F.-X. Theillet, Multiple Site-Specific Phosphorylation of IDPs Monitored by NMR, in: B.B. Kragelund, K. Skriver (Eds.), Intrinsically Disordered Proteins: Methods and Protocols, Springer US, New York, NY, 2020: pp. 793–817. https://doi.org/10.1007/978-1-0716-0524-0_41.

[111] N. Festuccia, N. Owens, A. Chervova, A. Dubois, P. Navarro, The combined action of Esrrb and Nr5a2 is essential for murine naïve pluripotency, Development. 148 (2021) dev199604. https://doi.org/10.1242/dev.199604.

[112] J.-P. Lambert, M. Tucholska, T. Pawson, A.-C. Gingras, Incorporating DNA shearing in standard affinity purification allows simultaneous identification of both soluble and chromatin-bound interaction partners, J. Proteomics. 100 (2014) 55–59. https://doi.org/10.1016/j.jprot.2013.12.022.

[113] C. Smet-Nocca, H. Launay, J.-M. Wieruszeski, G. Lippens, I. Landrieu, Unraveling a phosphorylation event in a folded protein by NMR spectroscopy: phosphorylation of the Pin1 WW domain by PKA, J. Biomol. NMR. 55 (2013) 323–337. https://doi.org/10.1007/s10858-013-9716-z.

[114] W. Peti, R. Page, Molecular basis of MAP kinase regulation, Protein Sci. 22 (2013) 1698–1710. https://doi.org/10.1002/pro.2374.

[115] D.J. Wood, J.A. Endicott, Structural insights into the functional diversity of the CDK–cyclin family, Open Biol. 8 (2018) 180112–26. https://doi.org/10.1098/rsob.180112.

[116] D.L. Sheridan, Y. Kong, S.A. Parker, K.N. Dalby, B.E. Turk, Substrate Discrimination among Mitogen-activated Protein Kinases through Distinct Docking Sequence Motifs, Journal of Biological Chemistry. 283 (2008) 19511–19520. https://doi.org/10.1074/jbc.M801074200.

[117] S. Liokatis, A. Stützer, S.J. Elsässer, F.-X. Theillet, R. Klingberg, B. van Rossum, D. Schwarzer, C.D. Allis, W. Fischle, P. Selenko, Phosphorylation of histone H3 Ser10 establishes a hierarchy for subsequent intramolecular modification events., Nat. Struct. Mol. Biol. 19 (2012) 819–823. https://doi.org/10.1038/nsmb.2310.

[118] S.A. Lambert, A. Jolma, L.F. Campitelli, P.K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T.R. Hughes, M.T. Weirauch, The Human Transcription Factors, Cell. 172 (2018) 650–665. https://doi.org/10.1016/j.cell.2018.01.029.

[119] J. Liu, N.B. Perumal, C.J. Oldfield, E.W. Su, V.N. Uversky, A.K. Dunker, Intrinsic Disorder in Transcription Factors, Biochemistry. 45 (2006) 6873–6888. https://doi.org/10.1021/bi0602718.

[120] I. Yruela, C.J. Oldfield, K.J. Niklas, A.K. Dunker, Evidence for a Strong Correlation Between Transcription Factor Protein Disorder and Organismic Complexity., Genome Biol Evol. 9 (2017) 1248–1265. https://doi.org/10.1093/gbe/evx073.

[121] M. Nishi, H. Akutsu, S. Masui, A. Kondo, Y. Nagashima, H. Kimura, K. Perrem, Y. Shigeri, M. Toyoda, A. Okayama, H. Hirano, A. Umezawa, N. Yamamoto, S.W. Lee, A. Ryo, A Distinct Role for Pin1 in the Induction and Maintenance of Pluripotency, J. Biol. Chem. 286 (2011) 11593–11603. https://doi.org/10.1074/jbc.M110.187989.

[122] X.Z. Zhou, K.P. Lu, The isomerase PIN1 controls numerous cancer-driving pathways

and is a unique drug target, Nat. Struct. Mol. Biol. 16 (2016) 463–478. https://doi.org/10.1038/nrc.2016.49.
[123] Y. Chen, Y. Wu, H. Yang, X. Li, M. Jie, C. Hu, Y. Wu, S. Yang, Y. Yang, Prolyl isomerase Pin1: a promoter of cancer and a target for therapy, Cell Death Dis. 9 (2018) 883. https://doi.org/10.1038/s41419-018-0844-y.
[124] J. Gebel, M. Tuppi, A. Chaikuad, K. Hötte, M. Schröder, L. Schulz, F. Lohr, N. Gutfreund, F. Finke, E. Henrich, J. Mezhyrova, R. Lehnert, F. Pampaloni, G. Hummer, E.H.K. Stelzer, S. Knapp, V. Dötsch, p63 uses a switch-like mechanism to set the threshold for induction of apoptosis., Nat. Chem. Biol. 16 (2020) 1078–1086. https://doi.org/10.1038/s41589-020-0600-3.
[125] W. Qin, K.F. Cho, P.E. Cavanagh, A.Y. Ting, Deciphering molecular interactions by proximity labeling, Nat. Meth. (2021) 1–11. https://doi.org/10.1038/s41592-020-01010-5.
[126] J.-P. Lambert, M. Tucholska, C. Go, J.D.R. Knight, A.-C. Gingras, Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes, J. Proteomics. 118 (2015) 81–94. https://doi.org/10.1016/j.jprot.2014.09.011.
[127] A.-C. Gingras, K.T. Abe, B. Raught, Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles, Curr. Opin. Chem. Biol. 48 (2019) 44–54. https://doi.org/10.1016/j.cbpa.2018.10.017.
[128] X. Liu, K. Salokas, R.G. Weldatsadik, L. Gawriyski, M. Varjosalo, Combined proximity labeling and affinity purification−mass spectrometry workflow for mapping and visualizing protein interaction networks, Nat. Protoc. (2020) 1–32. https://doi.org/10.1038/s41596-020-0365-x.
[129] D.-P. Minde, M. Ramakrishna, K.S. Lilley, Biotin proximity tagging favours unfolded proteins and enables the study of intrinsically disordered regions, Commun Biol. (2020) 1–13. https://doi.org/10.1038/s42003-020-0758-y.
[130] H. Göös, M. Kinnunen, K. Salokas, Z. Tan, X. Liu, L. Yadav, Q. Zhang, G.-H. Wei, M. Varjosalo, Human transcription factor protein interaction networks, Nat Commun. 13 (2022) 766. https://doi.org/10.1038/s41467-022-28341-5.
[131] C.P. Wigington, J. Roy, N.P. Damle, V.K. Yadav, C. Blikstad, E. Resch, C.J. Wong, D.R. Mackay, J.T. Wang, I. Krystkowiak, D.A. Bradburn, E. Tsekitsidou, S.H. Hong, M.A. Kaderali, S.-L. Xu, T. Stearns, A.-C. Gingras, K.S. Ullman, Y. Ivarsson, N.E. Davey, M.S. Cyert, Systematic Discovery of Short Linear Motifs Decodes Calcineurin Phosphatase Signaling, Mol. Cell. 79 (2020) 342-358.e12. https://doi.org/10.1016/j.molcel.2020.06.029.
[132] Y. Ueki, T. Kruse, M.B. Weisser, G.N. Sundell, M.S.Y. Larsen, B.L. Mendez, N.P. Jenkins, D.H. Garvanska, L. Cressey, G. Zhang, N. Davey, G. Montoya, Y. Ivarsson, A.N. Kettenbach, J. Nilsson, A Consensus Binding Motif for the PP4 Protein Phosphatase, Mol. Cell. 76 (2019) 953-964.e6. https://doi.org/10.1016/j.molcel.2019.08.029.
[133] C. Benz, M. Ali, I. Krystkowiak, L. Simonetti, A. Sayadi, F. Mihalic, J. Kliche, E. Andersson, P. Jemth, N.E. Davey, Y. Ivarsson, Proteome-scale mapping of binding sites in the unstructured regions of the human proteome, Mol. Syst. Biol. 18 (2022). https://doi.org/10.15252/msb.202110584.
[134] N.E. Davey, L. Simonetti, Y. Ivarsson, ProP-PD for proteome-wide motif-mediated interaction discovery, Trends Biochem. Sci. 47 (2022) 547–548. https://doi.org/10.1016/j.tibs.2022.01.005.
[135] G. Krainer, T.J. Welsh, J.A. Joseph, J.R. Espinosa, S. Wittmann, E. de Csilléry, A. Sridhar, Z. Toprakcioglu, G. Gudiškytė, M.A. Czekalska, W.E. Arter, J. Guillén-Boixet, T.M. Franzmann, S. Qamar, P.S. George-Hyslop, A.A. Hyman, R. Collepardo-Guevara, S. Alberti, T.P.J. Knowles, Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions, Nat Commun. 12 (2021) 1085. https://doi.org/10.1038/s41467-

021-21181-9.
[136] K. Tsafou, P.B. Tiwari, J.D. Forman-Kay, S.J. Metallo, J.A. Toretsky, Targeting Intrinsically Disordered Transcription Factors: Changing the Paradigm, J. Mol. Biol. 430 (2018) 2321–2341. https://doi.org/10.1016/j.jmb.2018.04.008.
[137] A. Chen, A.N. Koehler, Transcription Factor Inhibition: Lessons Learned and Emerging Targets, Trends Mol Med. 26 (2020) 508–518. https://doi.org/10.1016/j.molmed.2020.01.004.

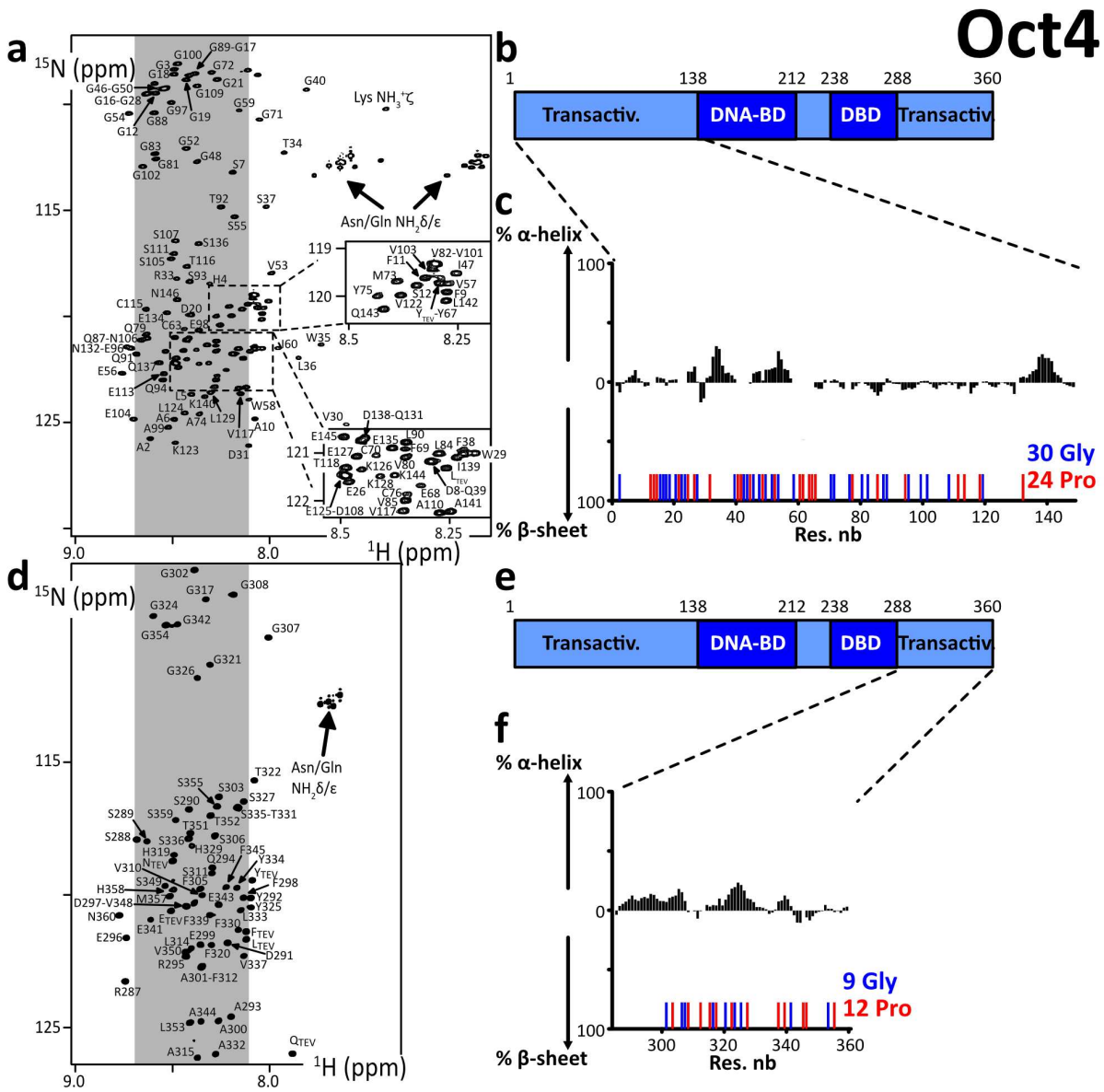

Figure 1 : ***a. d.*** *2D ^{1}H-^{15}N HSQC spectra of the N- and C-terminal IDRs of human Oct4, the labels indicating the assignments; the grey areas show the spectral regions where random coil amino acids resonate usually;* ***b. c.*** *Primary structures of Oct4; dark and light colors indicate the folded and disordered domains, respectively; blue and red sticks indicate the positions of Glycines and Prolines, respectively;* ***c. e.*** *Secondary structure propensities calculated from the experimental chemical shifts of the peptide backbone Cα and Cβ, using the ncSPC algorithm [104,105].*

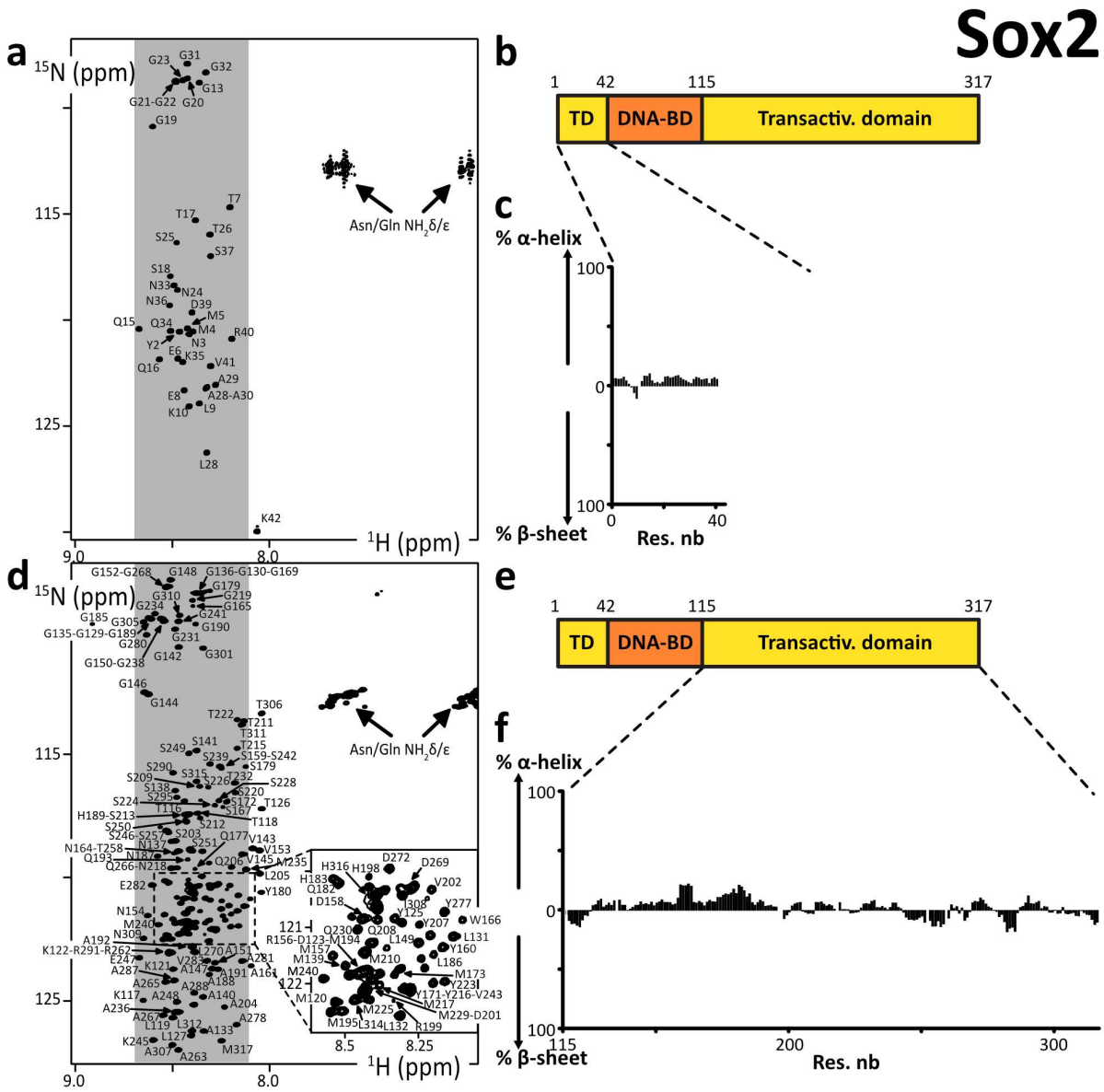

Figure 2: ***a. d.*** *2D ^{1}H-^{15}N HSQC spectra of the N- and C-terminal IDRs of human Sox2, the labels indicating the assignments; the grey areas show the spectral regions where random coil amino acids resonate usually;* ***b. c.*** *Primary structures of human Sox2; dark and light colors indicate the folded and disordered domains, respectively;* ***c. e.*** *Secondary structure propensities calculated from the experimental chemical shifts of the peptide backbone Cα and Cβ, using the ncSPC algorithm [104,105].*

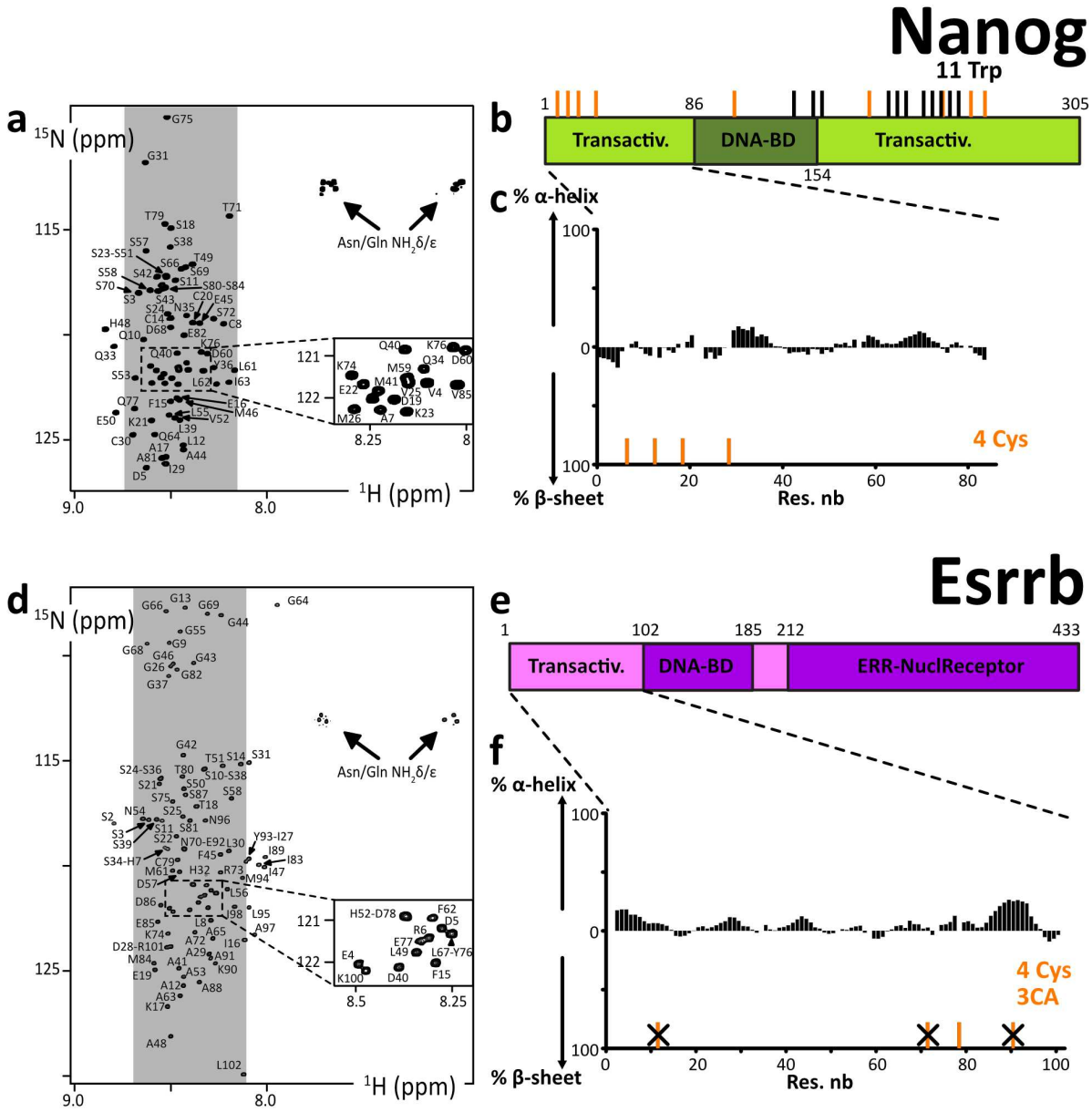

Figure 3: ***a. d.*** *2D ^{1}H-^{15}N HSQC spectra of the N-terminal IDRs of human Nanog and Esrrb, the labels indicating the assignments; the grey areas show the spectral regions where random coil amino acids resonate usually;* ***b. c.*** *Primary structures of human Nanog and Esrrb; dark and light colors indicate the folded and disordered domains, respectively; orange and black sticks indicate the positions of cysteines and tryptophanes, respectively;* ***c. e.*** *Secondary structure propensities calculated from the experimental chemical shifts of the peptide backbone Cα and Cβ, using the ncSPC algorithm [104,105].*

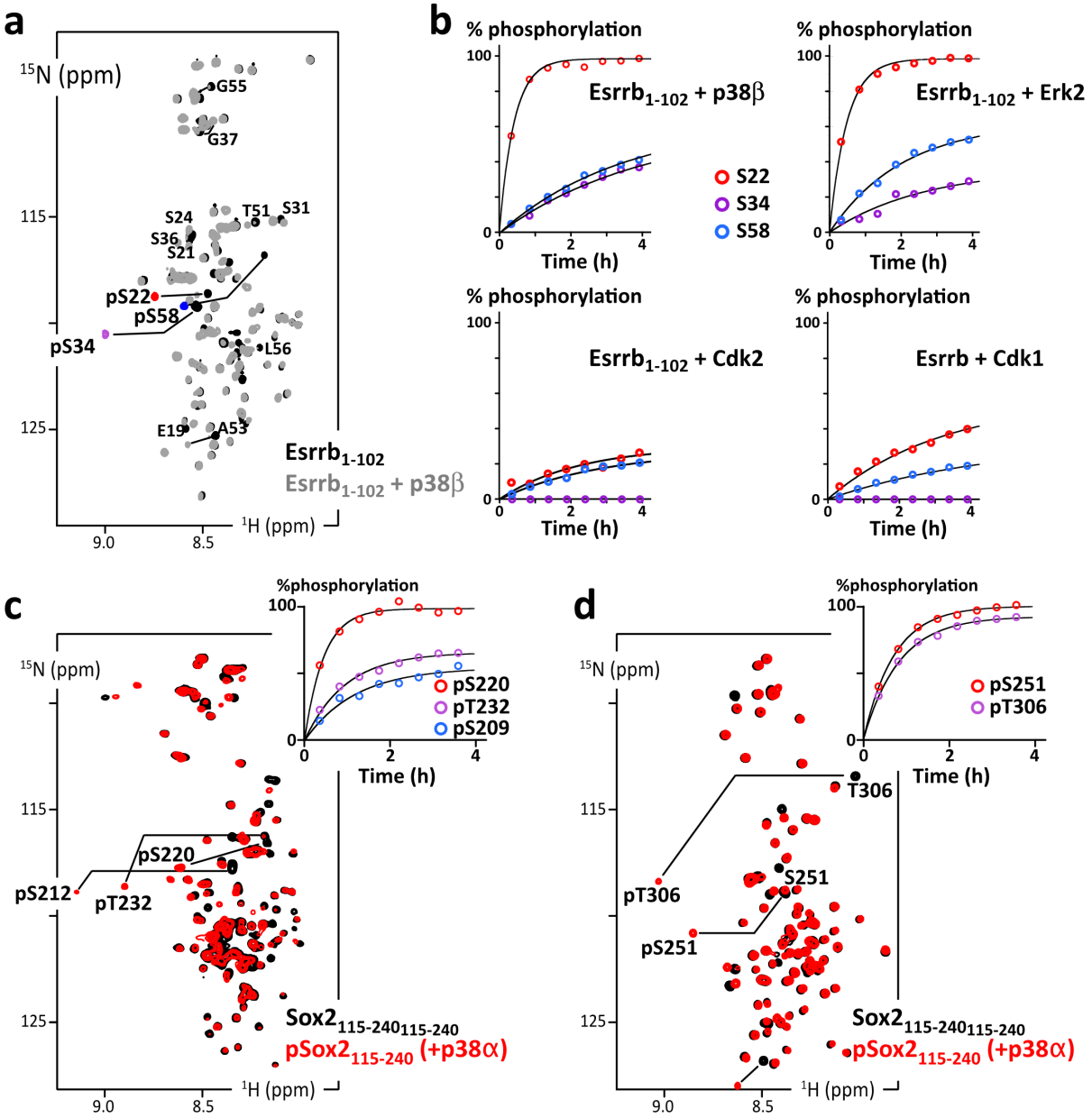

Figure 4: a. *Overlay of 2D ^{1}H-^{15}N HSQC spectra of Esrrb(aa1-102_C12A-C72A-C91A) before (black) and after (grey and phosphosites colored in red/purple/blue) phosphorylation by p38β;* ***b.*** *Residue specific time courses of the phosphorylation of Esrrb(aa1-102) executed by commercial kinases p38β, Erk2, Cdk2 or Cdk1, as measured in time series of 2D ^{1}H-^{15}N SOFAST-HMQC spectra recorded during the reaction;* ***c.*** *Overlay of 2D ^{1}H-^{15}N HSQC spectra of AviTag-Sox2(aa115-240) before (black) and after (red) phosphorylation by p38a; the inset at the top-right shows the residue specific phosphorylation build-up curves, as measured in time series of 2D ^{1}H-^{15}N SOFAST-HMQC spectra; d. Same as* ***c.*** *with AviTag-Sox2(aa234-317_C265A).*

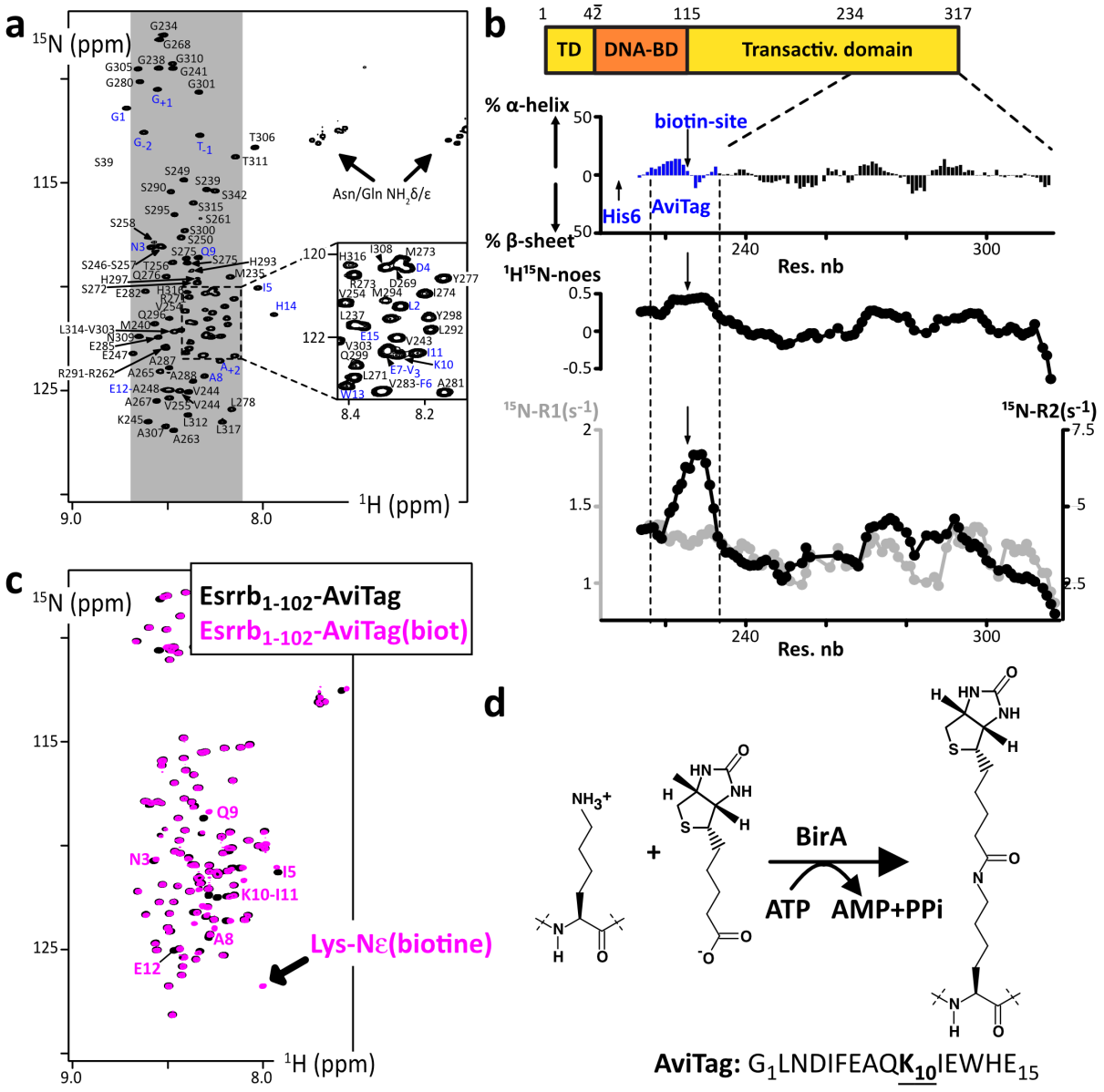

Figure 5 : ***a.*** *2D ^{1}H-^{15}N HSQC spectrum of ^{15}N-His6-AviTag-Sox2(aa234-317_C265A), the blue labels indicated the assigned signals from the AviTag residues ;* ***b.*** *Secondary structure propensities calculated from the experimental chemical shifts of the peptide backbone Cα and Cβ, using the ncSPC algorithm [104,105]; the residue specific ^{1}H-^{15}N-noes, ^{15}N-R1 (grey) and ^{15}NR2 (black) measured at 600 MHz are shown below (the profiles show values averaged over three consecutive residues);* ***c.*** *Overlay of 2D ^{1}H-^{15}N HSQC spectra of the Esrrb(aa1-102_C12A-C72A-C91A)-AviTag-His6 before (black) et after (magenta) biotinylation by BirA ; the NMR signals from the residues neighboring the biotinylation site are indicated, which permit the quantification of the biotinylated population ;* ***d.*** *Scheme of the reaction of Avi-Tag biotinylation executed by the ATP-dependent BirA.*

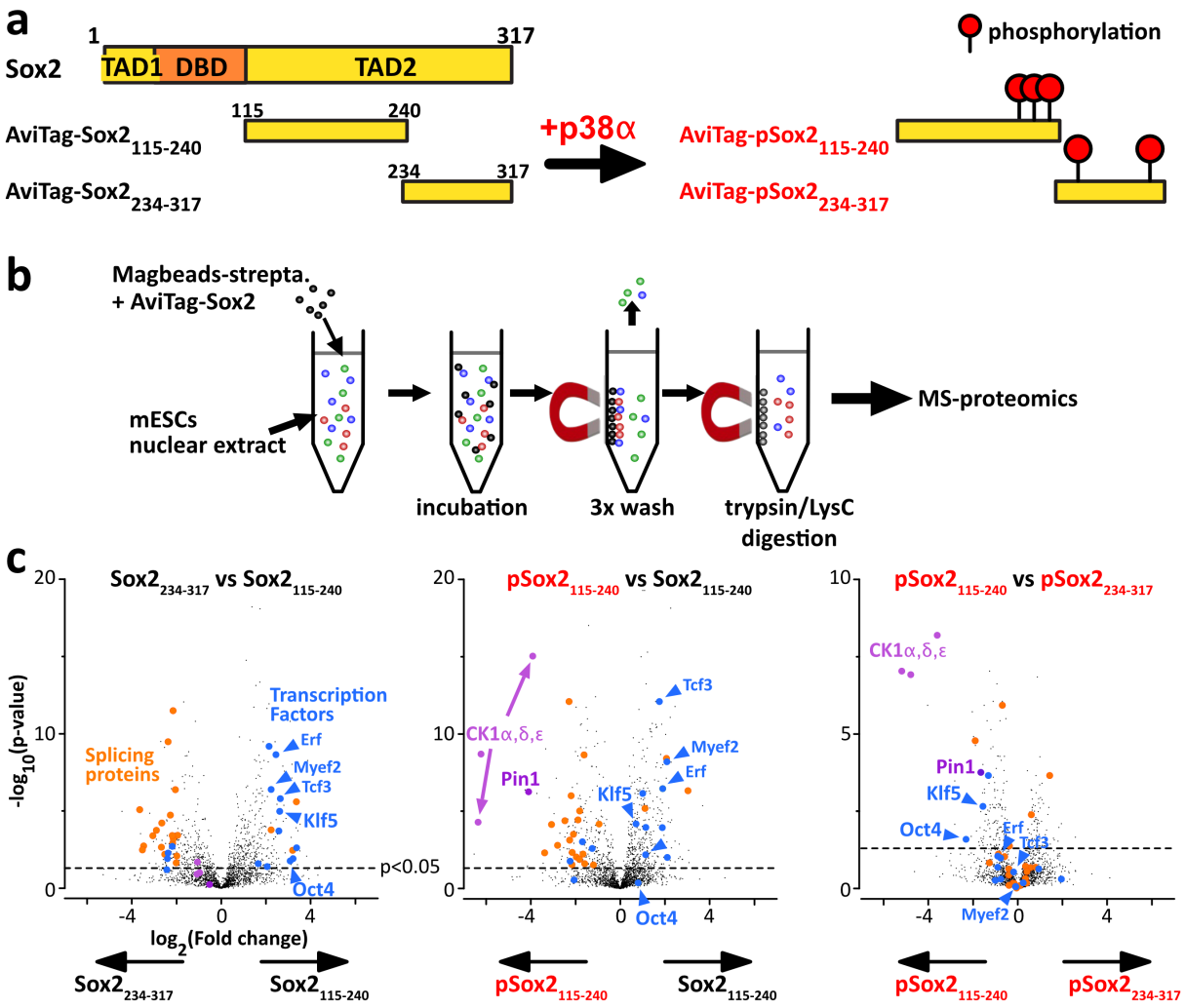

Figure 6: Differential interactomics of Sox2 constructs upon p38α phosphorylation. a. *We have produced unmodified/phospho-Sox2 truncations carrying N-ter biotinylation on the AviTag, and later attached on streptavidin-magnetic beads.* ***b.*** *We have generated mESCs nuclear extracts for pull-down assays using our Sox2 constructs as baits.* ***c.*** *The volcano plots of the log2 ratios, showing a quantitative analysis of the proteins present at the end of the pull-down assays; the dashed line indicates the threshold of p-value<0.05; we highlighted interesting partners in: <u>blue:</u> transcription factors associating with Sox2(aa115-240); <u>magenta:</u> phospho-dependent partners of Sox2(aa115-240) and or Sox2(aa234-317). Pull-downs have been performed in duplicates, using 15 million cells per sample (extract protein conc.: 5 mg/mL), and 1 nmol of bait protein. Experimental conditions may be improved (higher number of replicates, cells, washing conditions, ...).*

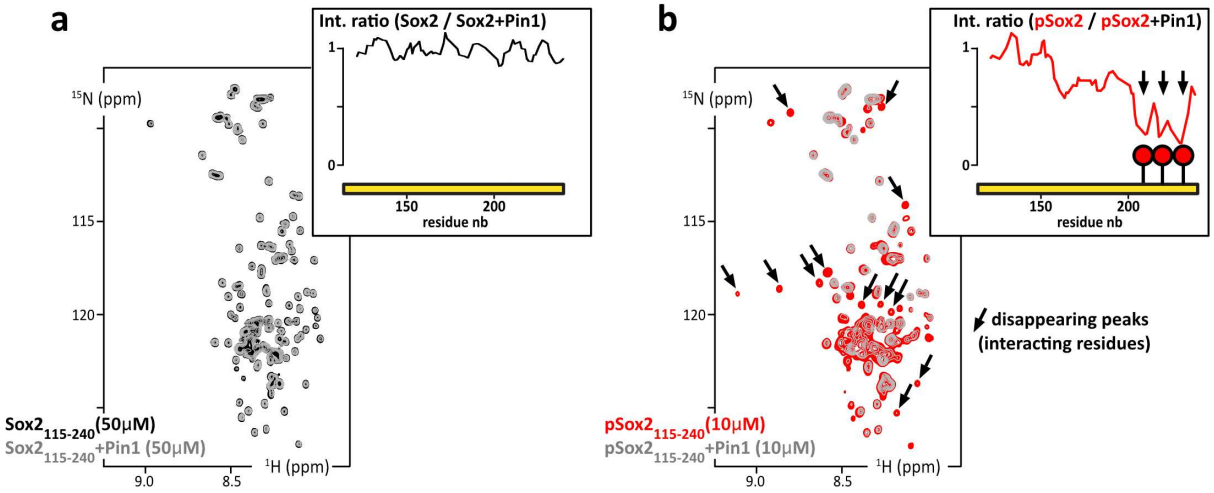

Figure 7: a. *Overlay of 2D ^{1}H-^{15}N HSQC spectra of ^{15}N-Sox2(aa115-240) alone at 50 µM (black) or mixed with Pin1 in isotopic natural abundance and in stoichiometric amounts (grey) ; inset up-right : residue specific NMR signal intensity ratios as measured in the two HSQC spectra.* ***b.*** *Overlay of 2D ^{1}H-^{15}N HSQC spectra of ^{15}N-phosphoSox2(aa115-240) alone at 10 µM (red) or mixed with the Pin1-WW domain in stoichiometric amounts (grey); insets, up-right : residue specific NMR signal intensity ratios as measured in the two HSQC spectra (in absence/presence of Pin1-WW domain).*