ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE
1666

Comptes Rendus. Chimie
Supplementary material to the article available at:
https://doi.org/10.5802/crchim.272

Research article

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

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

Current addresses: Structural Motility, Institut Curie, Paris Université Sciences et Lettres, Sorbonne Université, CNRS UMR144, 75005 Paris, France (C. Bouguechtouli), Université de Paris, Institut Cochin, CNRS UMR8104, INSERM U1016, Paris, France (A. Alik)

E-mail: francois-xavier.theillet@cnrs.fr (F.-X. Theillet)

1. Experimental procedures

1.1. *OSNE peptides production*

The TEV (Tobacco Etch Virus) protease was produced in-house recombinantly in *E. coli* BL21(DE3)Star, from a construct containing a hexahistidine tag (His6).

All peptides were produced in *E. coli* (strain BL21(DE3)Star) transformed with the plasmids presented in the main text. Cells were grown in M9 medium containing $^{15}NH_4^+$ (0.5 g/L) and ^{13}C-glucose (2 g/L) as sole sources of nitrogen and carbon for producing samples used for NMR assignment, and natural abundance ^{12}C-glucose (2 g/L) otherwise. Media were supplemented with kanamycine at 50 µg/mL, and the expression was induced at an optical density OD_{600} = 0.8 by supplementing the medium with ITPG at 1 mM at 37 °C. Cells were harvested by centrifugation (5 min at 5000 g) 4 h later and cell pellets were stored at −20 °C. Cells were lysed using sonication in Tris 20 mM, NaCl 150 mM, at pH 7.4 (buffer called "Tris Buffer Saline", TBS) in presence of benzonase (E1014 Sigma-Aldrich), lysozyme, protease inhibitors 1× (EDTA-free cOmplete, Roche) and 10 mM DTT.

Soluble and insoluble fractions were separated by 15 min of centrifugation at 15,000 g. Oct4-, Nanog- and Esrrb-peptides were purified from the soluble fractions. The lysates were loaded on a His-Trap FF column (5 mL, Cytiva) and eluted using a gradient of imidazole (in TBS). The eluted fractions were concentrated, submitted to TEV treatment for 1 h in TBS+imidazole supplemented with 10 mM DTT, and then diluted in TBS and re-loaded on the His-Trap column. Fractions containing the peptide of interest were submitted to a size-exclusion chromatography (SEC) in a column (Superdex 16/60 75 pg, Cytiva) previously equilibrated with Hepes at 10 mM or 20 mM, or phosphate at 20 mM, and NaCl 50 mM or 150 mM, at pH 6.8 (low-salt samples for NMR assignments,

* Corresponding author.

high-salt samples for phosphorylation kinetics or pull-down assays). For the cysteine-containing peptides, the eluted fractions of interest were immediately supplemented with DTT or TCEP at 2 mM, concentrated and stored at −20 °C. Fresh DTT or TCEP was supplemented further after thawing before the NMR experiments.

The constructs containing Sox2(aa1-42), Sox2(aa115-187) and Sox2(aa234-317)-AviTag-His6 were also purified from the soluble fraction, as explained above. The other Sox2-constructs were recovered from the insoluble fractions of the lysates, and resolubilized in TBS supplemented with 8 M urea, loaded on a His-Trap FF column (5 mL, Cytiva) and eluted using a gradient of imidazole; the eluates were then supplemented with β-mercaptoethanol at 50 mM and incubated at room temperature for 15 min, before being dialyzed in TBS supplemented with DTT at 1 mM, in order to refold the GST domain. The samples were then submitted to TEV cleavage in 0.5 M urea, and a second His-Trap purification was carried out in TBS supplemented with urea at 2 M. The fractions of interest were concentrated and submitted to a SEC in Hepes at 10 mM or phosphate at 20 mM, and NaCl 50 mM or 150 mM, urea at 2 M, at pH 6.8. The samples were concentrated and stored at −20 °C. Before the NMR experiments, they were thawed and 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. We paid attention to avoid precipitation during the concentration steps, because these peptides had a limited solubility, about 100–150 μM.

We achieved some liquid–liquid phase separation assays, using DIC microscopy at room temperature in Ficoll-70 at 100 mg/mL. These were carried out with Sox2 peptides previously centrifuged during 10 min at 15,000 g to remove the aggregates: for example, coacervates were observed at 4 μM of Sox2(aa115-317_C265A), and some aggregates were rapidly forming under the microscope at 20 μM.

1.2. *Production of BirA and biotinylation of AviTag-peptide chimera*

Bacteria transformed with pET21a-BirA were precultured at 37 °C overnight in a Luria-Bertani (LB) culture medium supplemented with ampicillin at 100 μg/mL. Then, these were cultured in a larger volume of LB supplemented with ampicillin at 50 μg/mL at 37 °C, and at 30 °C when they reached an optical density (OD) of 0.4. At an OD = 0.8, the culture was transferred to 20 °C and the protein expression was induced by supplementing the medium with IPTG at 0.5 mM. The incubation was carried out overnight, the bacteria were harvested by centrifugation at 4500 g for 5 min and the pellets were stored at −20 °C.

The purification was carried out at 4 °C. Cells were lysed using sonication in TBS at pH 7.5 in presence of 0.5 μL of benzonase (E1014 Sigma-Aldrich), lysozyme, PMSF at 1 mM (Sigma-Aldrich) and DTT at 10 mM. The soluble and insoluble fractions were separated by 15 min centrifugation at 15,000 g. The lysate (supernatant, soluble fraction) was loaded on a His-Trap column (His-Trap FF 5 mL, Cytiva) and eluted in TBS using a gradient of imidazole. The eluted fractions of interest were concentrated in presence of DTT at 10 mM, and later submitted to a SEC in a column (Superdex 16/60 75 pg, Cytiva) previously equilibrated with TBS at pH 7.5, supplemented with 10% v/v glycerol. The fractions of interest were concentrated in presence of DTT at 2 mM. Final concentrations of BirA were about 100 μM. The obtained sample was aliquoted, flash-frozen and stored at −80 °C.

The primary sequence of the expressed construct is:

```
MKDNTVPLKLIALLANGEFHSGEQLGETLGMSRAAINKHIQTLRDWG
VDVFTVPGKGYSLPEPIQLLNAKQILGQLDGGSVAVLPVIDSTNQYL
LDRIGELKSGDACIAEYQQAGRGRRGRKWFSPFGANLYLSMFWRLEQ
GPAAAIGLSLVIGIVMAEVLRKLGADKVRVKWPNDLYLQDRKLAGIL
VELTGKTGDAAQIVIGAGINMAMRRVEESVVNQGWITLQEAGINLDR
NTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGD
KEIFGISRGIDKQGALLLEQDGIIKPWMGGEISLRSAEKKLAAALEH
HHHHH*
```

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

Almost all NMR spectra were recorded on a 700 MHz Bruker Avance Neo spectrometer or a 600 MHz Bruker Avance II, equipped with cryogenically cooled triple resonance ^{1}H[^{13}C/^{15}N] probes optimized for ^{1}H-detection, a TCI and a TXI, respectively. Assignment spectra of Sox2_aa115-317 were recorded on a 950 MHz Bruker Avance III

spectrometer, equipped with a cryogenically cooled triple resonance $^{1}H[^{13}C/^{15}N]$ probe (TCI). All spectra were processed in Topspin 3 or Topspin 4. 3D spectra analysis was carried out using CccpNmr 2.4.2. DSS at 100 μM and 7.5% D_2O were added in all samples.

NMR assignments of backbone amide resonances of uniformly-labeled peptides ($^{13}C/^{15}N$) was achieved using BEST-HNCO, -HN(CA)CO, -HNCACB,[4] and (H)N(CA)NH 3D experiments, in HEPES at 10 mM, NaCl at 50 mM, DTT or TCEP at 2 to 5 mM, at pH 6.8 and 283 K, and at peptide concentrations ranging from 150 to 900 μM in 5 mm diameter Shigemi tubes.

Assignments of Oct4_aa286-360, Sox2_aa1-42, Nanog_aa1-85, Esrrb_aa1-102 were carried out at 700 MHz; those of Oct4_aa1-145, Sox2_aa115-236, His6-AviTag-Sox2_aa234-317_C265A at 600 MHz; those of Sox2_aa115-317_C265A at 950 MHz.

Oct4_aa286-360: interscan delay: 0.5 s

B-HNCO and were carried out with 1024 (^{1}H) × 96 (^{13}C) × 80 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 8 ppm (^{13}C) and 22 ppm (^{15}N),

B-HN(CA)CO were carried out with 1024 (^{1}H) × 96 (^{13}C) × 64 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 8 ppm (^{13}C) and 22 ppm (^{15}N),

B-HNCACB 1024 (^{1}H) × 128 (^{13}C) × 64 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 60 ppm (^{13}C) and 22 ppm (^{15}N),

B-(H)N(CA)NH with 1024 (^{1}H) × 64 (^{15}N) × 64 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), and 22 ppm (^{15}N).

Sox2_aa1-42: interscan delay: 0.5 s

B-HNCO and was carried out with 2048 (^{1}H) × 88 (^{13}C) × 88 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 8 ppm (^{13}C) and 26 ppm (^{15}N),

HN(CA)CO was carried out with 2048 (^{1}H) × 72 (^{13}C) × 72 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 8 ppm (^{13}C) and 26 ppm (^{15}N),

B-HNCACB 2048 (^{1}H) × 80 (^{13}C) × 80 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 60 ppm (^{13}C) and 26 ppm (^{15}N),

B-(H)N(CA)NH with 2048 (^{1}H) × 64 (^{15}N) × 64 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), and 26 ppm (^{15}N).

Esrrb_aa1-102_3Cys->3Ala: interscan delay: 0.5 s

B-HNCO and was carried out with 2048 (^{1}H) × 92 (^{13}C) × 92 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 8 ppm (^{13}C) and 26 ppm (^{15}N),

HN(CA)CO was carried out with 2048 (^{1}H) × 922 (^{13}C) × 72 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 8 ppm (^{13}C) and 26 ppm (^{15}N),

B-HNCACB 2048 (^{1}H) × 128 (^{13}C) × 72 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 60 ppm (^{13}C) and 26 ppm (^{15}N),

B-(H)N(CA)NH with 2048 (^{1}H) × 72 (^{15}N) × 72 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), and 26 ppm (^{15}N).

Nanog_aa1-85: interscan delay: 0.5 s

B-HNCO was carried out with 1024 (^{1}H) × 96 (^{13}C) × 92 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 8 ppm (^{13}C) and 26 ppm (^{15}N),

B-HN(CA)CO: 1024 (^{1}H) × 88 (^{13}C) × 72 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 8 ppm (^{13}C) and 26 ppm (^{15}N),

B-HNCACB: 1024 (^{1}H) × 96 (^{13}C) × 72 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), 60 ppm (^{13}C) and 22 ppm (^{15}N),

B-(H)N(CA)NH: 1024 (^{1}H) × 72 (^{15}N) × 72 (^{15}N) complex points and sweep widths of 12.98 ppm (^{1}H), and 26 ppm (^{15}N).

Oct4_aa1-145: interscan delay: between 0.12 and 0.2 s

B-HNCO was carried out with 1536 (^{1}H) × 92 (^{13}C) × 72 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), 10 ppm (^{13}C) and 24 ppm (^{15}N),

B-HN(CA)CO: 1536 (^{1}H) × 92 (^{13}C) × 64 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), 10 ppm (^{13}C) and 24 ppm (^{15}N),

B-HNCACB: 1536 (^{1}H) × 128 (^{13}C) × 64 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), 65 ppm (^{13}C) and 24 ppm (^{15}N),

B-(H)N(CA)NH: 1536 (^{1}H) × 64 (^{15}N) × 64 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), and 24 ppm (^{15}N).

Sox2_aa115-236 (together with Sox2_aa1-187): interscan delay: between 0.12 and 0.2 s.

B-HNCO was carried out with 1536 (^{1}H) × 92 (^{13}C) × 64 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), 10 ppm (^{13}C) and 25 ppm (^{15}N),

B-HN(CA)CO: 1536 (^{1}H) × 64 (^{13}C) × 64 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), 10 ppm (^{13}C) and 26 ppm (^{15}N),

B-HNCACB: 1536 (^{1}H) × 84 (^{13}C) × 64 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), 65 ppm (^{13}C) and 26 ppm (^{15}N),

B-(H)N(CA)NH: 1536 (^{1}H) × 32 (^{15}N) × 32 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), and 24 ppm (^{15}N).

His6-AviTag-Sox2_aa234-317: interscan delay: between 0.12 and 0.2 s.

B-HNCO was carried out with 1536 (^{1}H) × 92 (^{13}C) × 72 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), 11 ppm (^{13}C) and 24 ppm (^{15}N),

B-HN(CA)CO: 1536 (^{1}H) × 92 (^{13}C) × 64 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), 11 ppm (^{13}C) and 24 ppm (^{15}N),

B-HNCACB: 1536 (^{1}H) × 128 (^{13}C) × 64 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), 65 ppm (^{13}C) and 24 ppm (^{15}N),

B-(H)N(CA)NH: 1536 (^{1}H) × 64(^{15}N) × 64 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H), and 24 ppm (^{15}N).

Sox2_aa115-317_C265A: d1 = 0.2 s.

B-HNCO: 2126 (^{1}H) × 128 (^{13}C) × 128 (^{15}N) complex points and sweep widths of 14 ppm (^{1}H), 7 ppm (^{13}C) and 22 ppm (^{15}N).

B-HN(CA)CO: 2126 (^{1}H) × 92 (^{13}C) × 92 (^{15}N) complex points and sweep widths of 14 ppm (^{1}H), 7 ppm (^{13}C) and 22 ppm (^{15}N). Non-uniform sampling at 35%.

B-HNCACB: 2126 (^{1}H) × 128 (^{13}C) × 128 (^{15}N) complex points and sweep widths of 14 ppm (^{1}H), 60 ppm (^{13}C) and 22 ppm (^{15}N). Non-uniform sampling at 35%.

B-(H)N(CA)NH: 18218(^{1}H) × 92 (^{15}N) × 92 (^{15}N) complex points and sweep widths of 14 ppm (^{1}H), and 22 ppm (^{15}N). Non-uniform sampling at 35%.

Spectra were processed with linear prediction of 16 or 32 complex points in both ^{13}C, and ^{15}N dimensions, cosine apodization in ^{1}H and ^{15}N dimensions, no apodization in ^{13}C dimension, and zero filling to 2048, 512 and 256 complex points in ^{1}H, ^{13}C, and ^{15}N dimensions, respectively. Assignment 2D ^{1}H–^{15}N HSQC spectra were recorded using at least 1536 (^{1}H) × 256 (^{15}N) complex points and sweep widths of 16.23 ppm (^{1}H) and 30 ppm (^{15}N), and processed with zero filling to 4 K and 1 K in the proton and nitrogen dimensions, respectively.

1.4. *NMR monitoring of phosphorylation reactions and production of phosphorylated peptides*

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 on a 600 MHz Bruker Avance II or a 700 MHz Bruker Avance Neo spectrometer, both equipped with cryogenically cooled triple resonance ^{1}H[^{13}C/^{15}N] probes optimized for ^{1}H detection.

The kinase was spiked in the IDR sample on ice just before filling the NMR tube, which was immediately placed in the spectrometer. About 2 min were necessary to reach a temperature equilibrium. The automatic shimming procedure was then executed, and short 1D ^{1}H and ^{1}H(^{15}N-filtered)-SOFAST-HMQC spectra were recorded before the 2D spectra.

We recorded 2D spectra during the phosphorylation reactions as follows: 2D ^{1}H–^{15}N SOFAST-HMQC experiments were recorded using 2048 (^{1}H) × 96 (^{15}N) complex points and sweep widths of 16.6 ppm (^{1}H) and 26 ppm (^{15}N), 128 scans and interscan delays of 0.04 s; hence, the acquisition of one spectrum took 30 min. All spectra were processed zero filling to 2 K and 1 K in the direct and indirect dimensions, respectively. No apodization was applied for ^{1}H–^{15}N SOFAST-HMQC spectra.

After processing spectra in Topspin3, we measured peak intensities in NMRFAM-SPARKY [1]. Peaks were centered in every spectrum to follow peak shifting because of pH drifts. Progress curves were plotted and fitted in Kaleidagraph 4.5. In the case of the phosphorylation reactions that were not complete, we used decay curves to normalize phosphorylation build-up curves. At the opposite, for the phosphorylation reactions reaching ~100%, we used the phospho-peaks intensities for normalizing the build-up curves. Detailed descriptions of the methods can be found in previous reports and published protocols [2–5].

1.5. *Pull-down assays for interactomic analysis*

Mouse Embryonic Stem Cells (mESCs) were harvested using a classical trypsin treatment (2×150 cm^2 cell culture dishes, 70% confluency). The production of nuclear extracts was inspired by the procedures previously published by Gingras and colleagues [6]. Trypsin was blocked, and cell pellets were washed three times in PBS, before being resuspended in a first gentle lysis buffer containing HEPES at 10 mM, KCl at 10 mM, EDTA at 0.5 mM, DTT at 1 mM, PMSF at 0.5 mM, 1% v/v NP40. After 10 min on ice, the cells were centrifuged 10 min at 15,000 g. The supernatant containing the cytosolic fraction was discarded, and the pellets containing the nuclei were resuspended in a second lysis buffer containing HEPES at 20 mM, KCl at 250 mM, EDTA at 0.5 mM, DTT at 1 mM, PMSF at 0.5 mM, phosphatase inhibitors (2× PhosSTOP, Roche), 5% v/v glycerol, supplemented with 2 µL of benzonase (>250 units/µL, E1014 Sigma-Aldrich), before being sonicated on ice using a microtip sonicator and 3 pulses of 10 seconds. The lysis of nuclei was verified visually under the microscope using a cell counting chamber. The extract concentration used for the pull-downs was about 5 mg/mL, as measured by Bradford protein assay.

The pull-down assays were executed using 25 µL of streptavidin-coated magnetic beads (Magbeads streptavidine, Genscript), i.e. 50 µL of resuspended beads in the storing buffer. After every step described below, the tubes were placed on a magnetic rack to collect the beads, while the supernatant was removed with a pipette. The fresh beads were washed 3 times during 5 min in 500 µL of a PBS buffer. 1 nmol of biotinylated (using BirA, see above) AviTag-chimera peptides (either AviTag-Sox2(aa115-240), AviTag-Sox2(aa234-317_C265A), phospho-AviTag-Sox2(aa115-240), or phospho-AviTag-Sox2(aa234-317_C265A)) were diluted in 500 µL of PBS and incubated with the beads during one hour at room temperature under rotary agitation. The supernatant was then removed, the beads were washed 3 times during 5 min in 500 µL of a PBS buffer.

The mESCs extract (200 µL, generated from 15 million cells) was mixed with the beads, and then incubated during one hour at room temperature under rotary agitation. The supernatant was removed and the beads were washed 3 times during 5 min in 500 µL of a buffer containing HEPES at 20 mM, KCl at 250 mM, EDTA at 0.5 mM, DTT at 1 mM, PMSF at 0.5 mM + PhosSTOP 1×.

1.6. *Mass spectrometry-based proteomics analysis of pull-down assays*

1.6.1. *Sample preparation*

The beads were resuspended in 100 µL of 25 mM NH_4HCO_3 and digested by adding 0.2 µg of trypsin/LysC (Promega) for 1 h at 37 °C. Samples were then loaded into custom-made C18 StageTips packed by stacking three AttractSPE® disk (#SPE-Disks-Bio-C18-100.47.20 Affinisep) into a 200 µL micropipette tip for desalting. Peptides were eluted using a ratio of 40:60 CH_3CN:H_2O + 0.1% formic acid and vacuum concentrated to dryness with a SpeedVac apparatus. Peptides were reconstituted in 10 µL of injection buffer in 0.3% trifluoroacetic acid (TFA) before liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

1.6.2. *LC-MS/MS analysis*

Online chromatography was performed with an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled to an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific). Peptides were trapped on a C18 column (75 µm inner diameter × 2 cm; nanoViper Acclaim PepMap^{TM} 100, Thermo Scientific) with buffer A (2/98 MeCN/H_2O in 0.1% formic acid) at a flow rate of 4.0 µL/min over 4 min. Separation was performed on a 50 cm × 75 µm C18 column (nanoViper Acclaim PepMap^{TM} RSLC, 2 µm, 100 Å, Thermo Scientific) regulated to a temperature of 55 °C with a linear gradient of 5% to 25% buffer B (100% MeCN in 0.1% formic acid) at a flow rate of 300 nL/min over 100 min. Peptides were ionized by a nanospray ionization (NSI) ion source at 2.2 kV. Full-scan MS in the Orbitrap was set at a scan range of 400–1500 with a resolution at 120,000 and ions from each full scan were fragmented in higher-energy collisional dissociation mode (HCD) and analyzed in the linear ion trap in rapid mode. The fragmentation was set top speed mode in data-dependent analysis (DDA). We selected ions with charge state from 2+ to 7+ for screening. Normalized collision energy (NCE) was set to 30, AGC target to 20,000 and the dynamic exclusion to 30 s.

1.6.3. *Data analysis*

For identification, the datasets were searched against the *Mus Musculus* (UP000000589) UniProt database using Sequest HT through proteome discoverer (version 2.2). Enzyme specificity was set to trypsin and a maximum of two miss cleavages sites were allowed. Oxidized methionine, phosphorylation of serines, threonines and tyrosines, carbamidomethylation of cysteines and N-terminal acetylation were set as variable modifications. 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. [7]). False-discovery rate (FDR) was calculated using Percolator [8] 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 [9] v.2.2.1. For protein quantification, XICs from proteotypic peptides shared between compared conditions (TopN matching) were used, missed cleavages and peptide modifications were not allowed. Median and scale normalization at peptide level was applied on the total signal to correct the XICs for each biological replicate ($N = 2$). To estimate the significance of the change in protein abundance, a linear model (adjusted on peptides and biological replicates) was performed, and p-values were adjusted using the Benjamini–Hochberg FDR procedure.

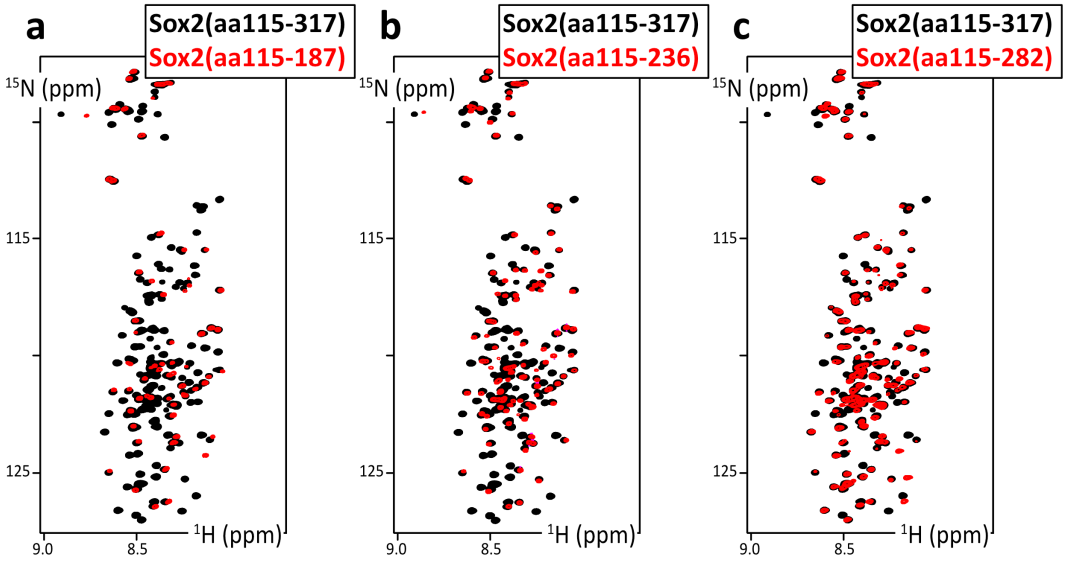
The mass spectrometry proteomics raw data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository dataset [10]: identifier PXD 040573 (reviewer_pxd040573@ebi.ac.uk and **Password:** sVM686zJ).

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

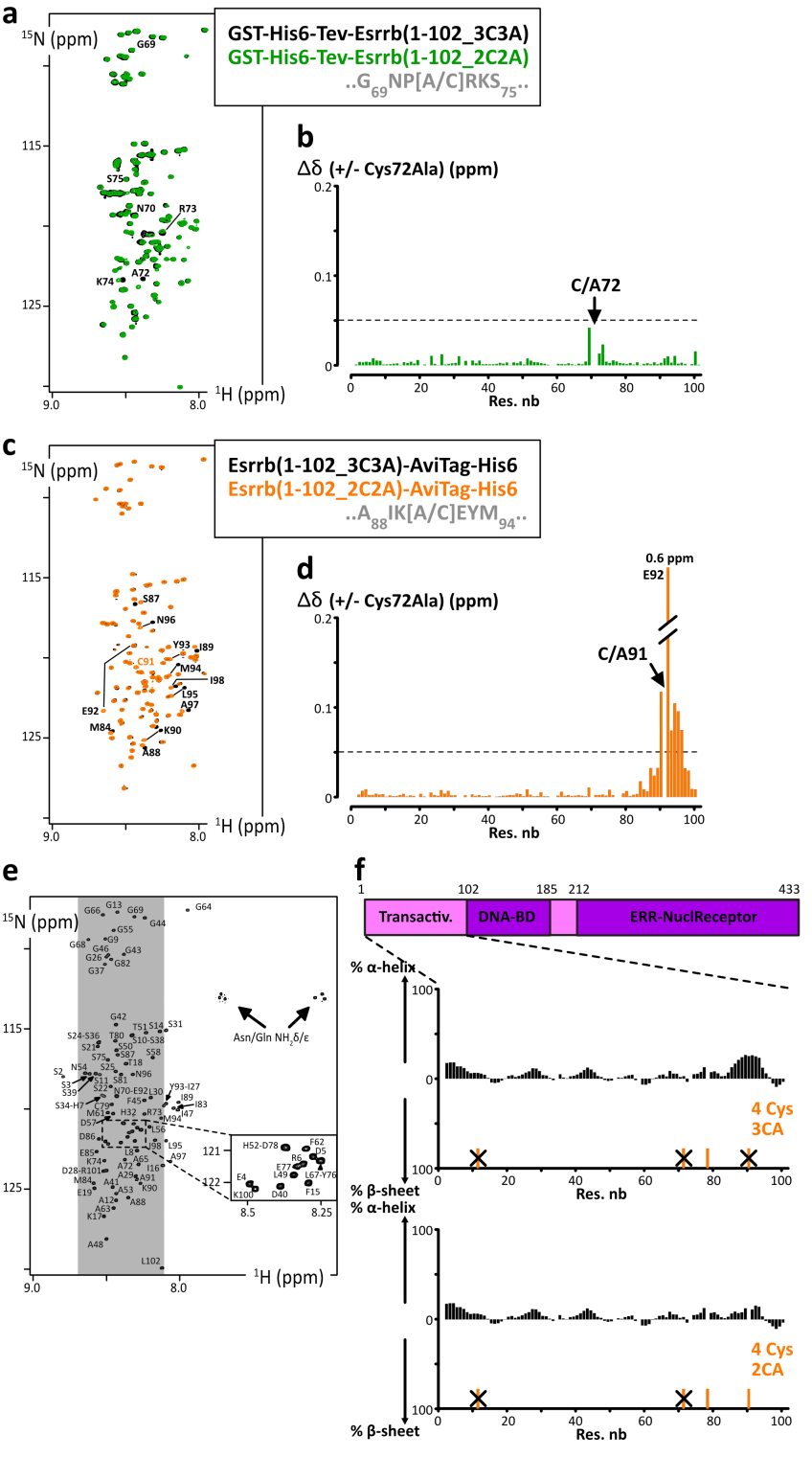
The plasmid containing the gene coding for the Pin1-WW domain was a kind gift from Isabelle Landrieu. The production was executed according to the previously published protocol [11]. The NMR analysis of binding with phosphoSox2(aa115-240) were performed with the GST-Pin1-WW construct and ^{15}N-labeled Sox2(aa115-240) mixed in stoichiometric proportions, either at 50 or 10 µM for non-phospho and phosphoSox2, respectively. The solution contained Hepes at 20 mM, NaCl at 50 mM, urea at 0.25 mM (left-overs from Sox2(aa115-240) stock, stored at 2 M urea for solubility, see above), 5% D_2O and DSS at 0.1 mM, at pH = 7.0. The 2D ^{1}H–^{15}N SOFAST-HMQC spectra were recorded at 283 K, using a 600 MHz Bruker Avance II equipped with a cryogenically cooled triple resonance ^{1}H[^{13}C/^{15}N] probe and a 5 mm diameter Shigemi tube.

The experiments were recorded using 1536 (^{1}H) × 128 (^{15}N) complex points and sweep widths of 16.0 ppm (^{1}H) and 264 ppm (^{15}N), 64 or 128 scans and interscan delays of 0.04 s. The spectra were processed with zero filling to 2 K and 1 K in the direct and indirect dimensions, respectively. Cosine apodization was applied in both dimensions. After processing spectra in Topspin3, we measured peak intensities in NMRFAM-SPARKY [1], and plotted the intensity ratios in Kaleidagraph 4.5.

2. Supplementary figures

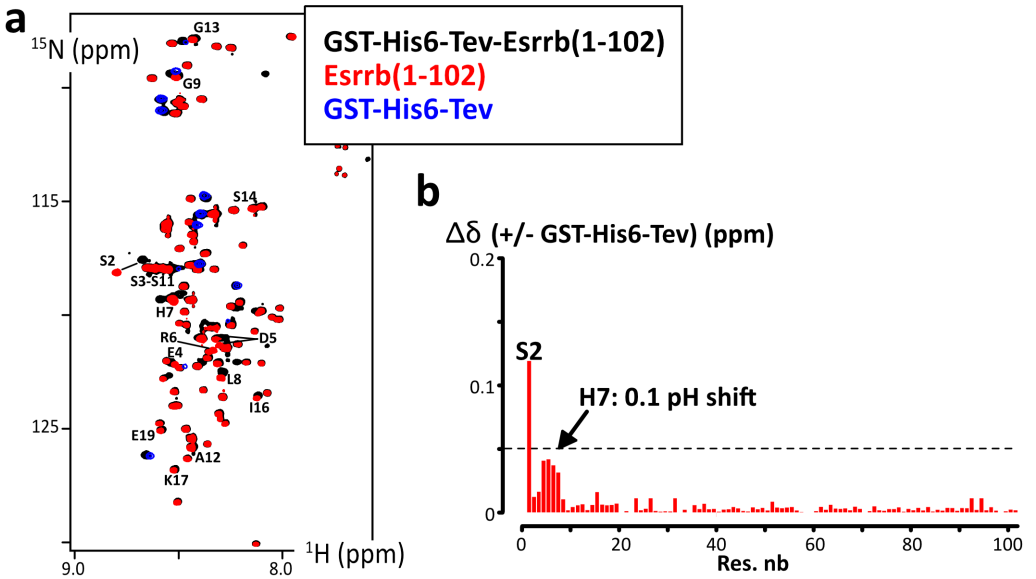


Supplementary Figure S1. Overlays of 2D ^{1}H–^{15}N HSQC spectra of Sox2(aa115-317_C265A) and Sox2(aa115-187), Sox2(aa115-236) and Sox2(aa115-282_C265A). These spectra have been recorded in a buffer containing urea at 0.25 M, except for Sox2(aa115-187), at 283 K and 700 MHz.

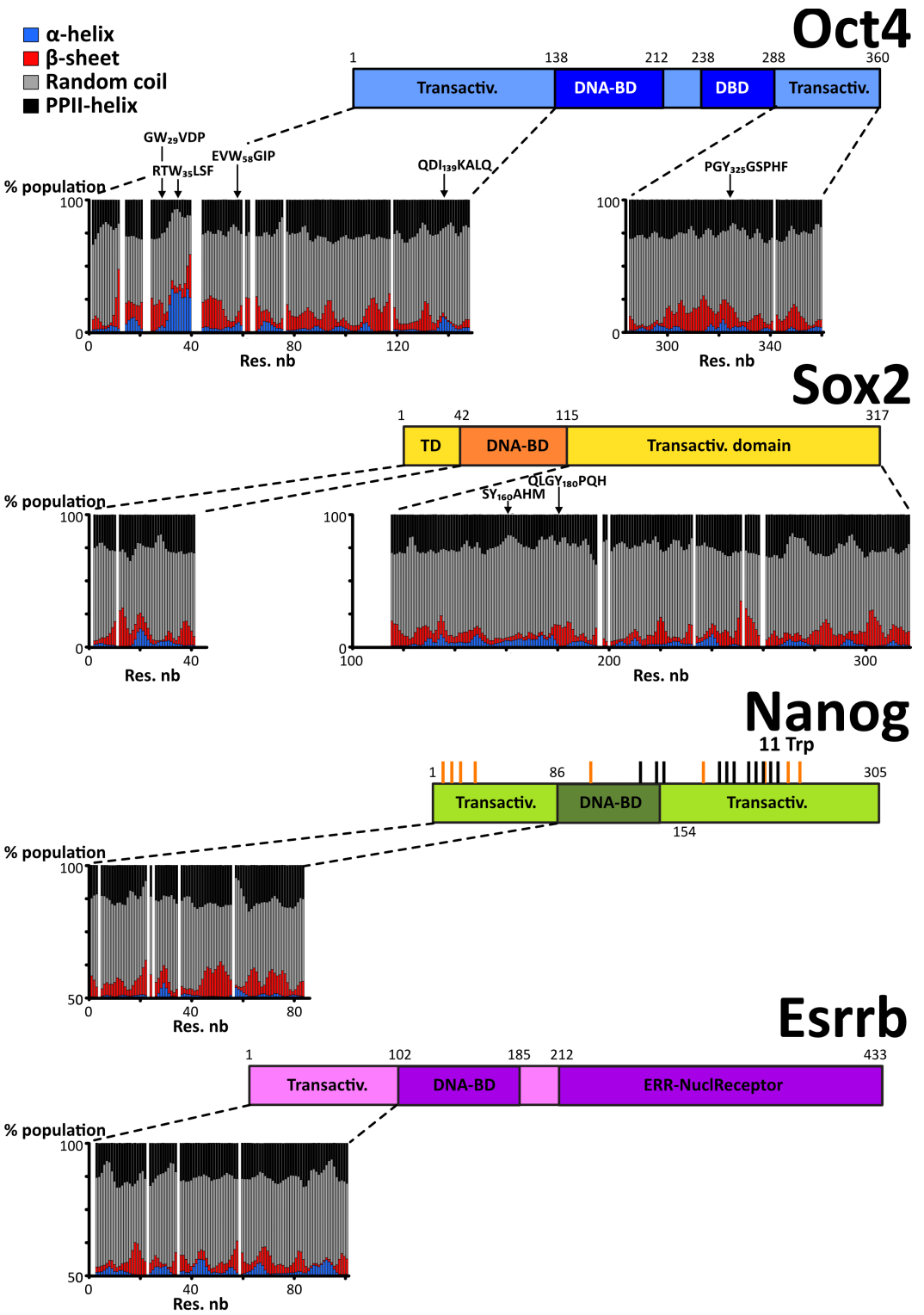


Supplementary Figure S2. Caption continued on next page.

Supplementary Figure S2. (**cont.**) (a) Overlay of 2D ^{1}H–^{15}N HSQC spectra of Esrrb(aa1-102_C12A-C72A-C91A) (black) and Esrrb(aa1-102_C12A-C91A) (green); (b) chemical shift perturbations between the two constructs in (a) using $\Delta\delta = [(\Delta\delta^2_{1H} + (\Delta\delta_{15N}/5)^2)/2]^{1/2}$; (c) overlay of 2D ^{1}H–^{15}N HSQC spectra of Esrrb(aa1-102_C12A-C72A-C91A) (black) and Esrrb(aa1-102_C12A-C72A) (orange); (d) chemical shift perturbations between the two constructs in (a); (e) 2D ^{1}H–^{15}N HSQC spectrum of the N-terminal IDRs of human Esrrb(aa1-102_C12A-C72A-C91A), the labels indicating the assignments; (f) primary structure of human Esrrb, and Secondary structure propensities of Esrrb(aa1-102_C12A-C72A-C91A) and Esrrb(aa1-102_C12A-C72A) calculated from the experimental chemical shifts of the peptide backbone Cα and Cβ, using the ncSPC algorithm [12,13].

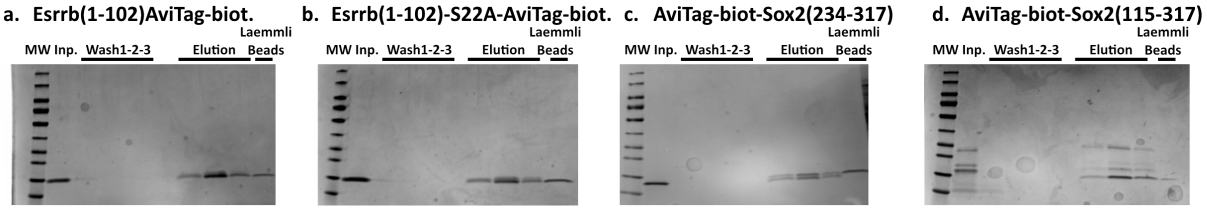


Supplementary Figure S3. (a) Overlay of 2D ^{1}H–^{15}N HSQC spectra of GST-His6-Tev-Esrrb(aa1-102_C12A-C72A-C91A) (black), Esrrb(aa1-102_C12A-C72A-C91A) (red) and GST-His6-Tev (blue); (b) chemical shift perturbations between the GST-His6-Tev-Esrrb(aa1-102_C12A-C72A-C91A) and Esrrb(aa1-102_C12A-C72A-C91A) after TEV cleavage, using $\Delta\delta = [(\Delta\delta^2_{1H} + (\Delta\delta_{15N}/5)^2)/2]^{1/2}$.



Supplementary Figure S4. Secondary structure propensities calculated from the experimental chemical shifts of the peptide backbone H_N, N_H, CO, Cα and Cβ, using the δ2D algorithm [14].

Supplementary Figure S5. Liquid–liquid phase separation of Sox2_aa115-317 at 4 µM in PBS supplemented with Ficoll70 at 100 g/L, observed under differential interference contrast (DIC) (TCS SP8-X inversed FALCON, Leica, 63× PLAN oil immersion implemented with DIC, Numerical Aperture: 1.4, Leica).



Supplementary Figure S6. (a–c) SDS-PAGE analysis of binding assays: 1 nmol of biotinylated AviTag-IDRs (Input: Inp.) were incubated with 25 µL of streptavidin-coated magnetic beads; the 3 washing steps showed the stability of the binding; the elutions were performed with a Laemmli buffer, which provokes also the release of the streptavidine, whose band is shown in the last lane and unfortunately overlaps with the AviTag-IDRs constructs here. (d) We show here one of our tests with a batch of AviTag-Sox2(aa115-317_C265A), which was partially proteolyzed; this is to show that one of the peptides that do not contain the AviTag is removed by the first wash.

3. Supplementary material 3: protein sequences

3.1. *GB1*

QYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEGG

3.2. *Oct4*

3.2.1. *Sequence alignment: mammals*

S: Serine
P : Proline
SP : phosphorylation motif for MAPKand Cdk
FILVYW : hydrophobic
DE : Asp/Glu
KR : Lys/Arg
T : Thr
XXX : DND-BD in crystal (3L1P)

```
oct4-human         MAGHLASDFAFSPPPGGGGDGPGGPEPGWVDPRTWLSFQGPPGGPGIGPGWGPGSEVWGI 60
oct4-mice          MAGHLASDFAFSPPPGGG-DGSAGLEPGWVDPRTWLSFQGPPGG----PGIGPGSEVLGI 55
oct4-cow           MAGHLASDFAFSPPPGGGGDGPGGPEPGWVDPRTWMSFQGPPGGSGIGPGWVPGAEVWGL 60
oct4-dog           MAGHLASDLAFSPSPGGGGDGPGGPDPGWGDPRAWLSFPGPPGGGHALGPGWGPGAEVWGL 60
Oct4-sheep         MAGHLASDFAFSPPPGGGGDGPGGPEPGWVDPRTWMSFQGPPGGSGIGPGWVPGAEVWGL 60
pou5f1-spermwhale MAGHLASDFAFSPPPGGGGDGPGGPEPGWVDPRTWMSFQGPPGGSGIGPGWGPGAEVWGL 60
                   ********:****.**** *...* :*** ***:*:** ***.*    **:  **:** *:

oct4-human         PPCPPPYEFCGGMAYCGPQVGVGLVPQGGLETSQPEGEAGVGVESNSDGASPEPCTVTPG 120
oct4-mice          SPCPPAYEFCGGMAYCGPQVGLGLVPQVGVETLQPEGQAGARVESNSEGTSSEPCADRPN 115
oct4-cow           PPCPPPYDLCGGMAYCAPQVGVGPVPPGGLETPQPEGEAGAGVESNSEGASPDPCAAPAG 120
oct4-dog           PPCPPPYEFCGGMAYCGPQVGVGLLPQGGLDTSQPEGERGAGLEGSSEGASPEPCAAPPG 120
Oct4- sheep        PPCPPPYDLCGGMAYCAPQVGVGPVPPGGLETPQPEGEAGAGVESNSEGASPDPCAAPAG 120
pou5f1-spermwhale PACPPPYDLCGGMAYCAPQVGVGLVPQGGLETPQPEGEAGAGVESKSEGASPEPCAAPAG 120
                   ..***.*::****:**.****:* :*  *::* ****: *. :*..*:*:*. **:  ..

oct4-human         AVKLEKEKLEQNPEESQDIKALQKELEQFAKLLKQKRITLGYTQADVGLTLGVLFGKVFS 180
oct4-mice          AVKL--EKVEPTPEESQDMKALQKELEQFAKLLKQKRITLGYTQADVGLTLGVLFGKVFS 173
oct4-cow           APKLDKEKLEPNPEESQDIKALQKDLEQFAKLLKQKRITLGYTQADVGLTLGVLFGKVFS 180
oct4-dog           VVKPDKEKLEQNPEESQDIKALQKDLEQFAKLLKQKRITLGYTQADVGLTLGVLFGKVFS 180
Oct4-sheep         AAKLDKEKLEPNPEESQDIKALQKDLEQFAKLLKQKRITLGYTQADVGLTLGVLFGKVFS 180
pou5f1-spermwhale AEKVDKEKLEPNPEESQDIKALQKDLEQFAKLLKQKRITLGYTQADVGLTLGVLFGKVFS 180
                   . *   **:* ..*****:*****:************************************

oct4-human         QTTICRFEALQLSFKNMCKLRPLLQKWVEEADNNENLQEICKAETLVQARKRKRTSIENR 240
oct4-mice          QTTICRFEALQLSLKNMCKLRPLLEKWVEEADNNENLQEICKSETLVQARKRKRTSIENR 233
oct4-cow           QTTICRFEALQLSFKNMCKLRPLLQKWVEEADNNENLQEICKAETLVQARKRKRTSIENR 240
oct4-dog           QTTICRFEALQLSFKNMCKLRPLLQKWVEEADNNENLQEICKAETLVQARKRKRTSIENR 240
Oct4-sheep         QTTICRFEALQLSFKNMCKLRPLLQKWVEEADNNENLQEICKAETLVQARKRKRTSIENR 240
pou5f1-spermwhale QTTICRFEALQLSFKNMCKLRPLLQKWVEEADNNENLQEICKAETLMQARKRKRTSIENR 240
                   ************:**********:*****************:***:**************

oct4-human         VRGNLENLFLQCPKPTLQQISHIAQQLGLEKDVVRVWFCNRRQKGKRSSSDYAQREDFEA 300
oct4-mice          VRWSLETMFLKCPKPSLQQITHIANQLGLEKDVVRVWFCNRRQKGKRSSIEYSQREEYEA 293
oct4-cow           VRGNLESMFLQCPKPTLQQISHIAQQLGLEKDVVRVWFCNRRQKGKRSSSDYSQREDFEA 300
oct4-dog           VRGNLENMFLQCPKPTLQQISHIAQQLGLEKDVVRVWFCNRRQKGKRSSSDYSQREDFEA 300
Oct4-sheep         VRGNLESMFLQCPKPTLQQISHIAQQLGLEKDVVRVWFCNRRQKGKRSSSDYSQREDFEA 300
pou5f1-spermwhale VRGNLESMFLQCPKPTLQQISHIAQQLGLEKDVVRVWFCNRRQKGKRSSSDYSQREDFEA 300
                   ** .**.:**:****:****: **:*****:****************** :*:***::**

oct4-human         AGSPFSGGPVSFPLAPGPHFGTPGYGSPHFTALYSSVPFPEGEAFPPVSVTTLGSPMHSN 360
oct4-mice          TGTPFPGGAVSFPLPPGPHFGTPGYGSPHFTTLYS-VPFPEGEAFPSVPVTALGSPMHSN 352
oct4-cow           AGSPFTGGPVSSPLAPGPHFGTPGYGGPHFTTLYSSVPFPEGEVTPSVSVTALGSPMHAN 360
oct4-dog           AGSPFSGAPVSFPLAPGPHFGTPGYGGPHFTTLYSSVPLPEGEGFPSVSVTTLGSPMHSN 360
Oct4-sheep         AGSPFAGGPVSFPLAPGPPFGTPGYGGPHFTTLYSSVPFPEGEAFPSVSVTALGSPMHAN 360
pou5f1-spermwhale AGSPFSGGPVSFPLAPGPHFGTPGYGGPHFTTLYSSVPFPEGEAFPSVSVTTLGSPMHSN 360
                   :*.**.*..** **.*** **:****.****:*** **:***** **.*.**.*******:*
```

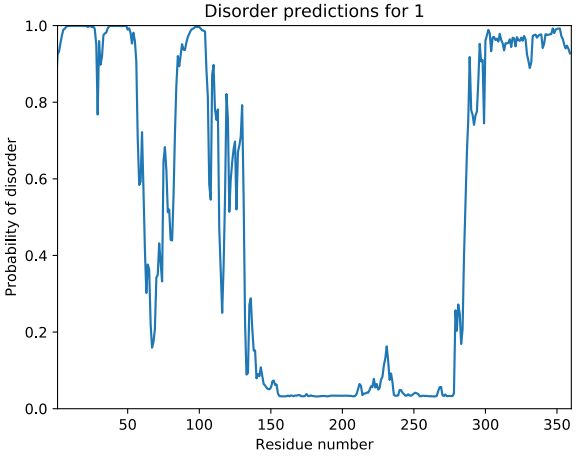
```
>oct4-human
MAGHLASDFAFSPPPGGGGDGPGGPEPGWVDPRTWLSFQGPPGGPGIGPGVGPGSEVWGIPPCPPPYEFC
GGMAYCGPQVGVGLVPQGGLETSQPEGEAGVGVESNSDGASPEPCTVTPGAVKLEKEKLEQNPEESQDIK
ALQKELEQFAKLLKQKRITLGYTQADVGLTLGVLFGKVFSQTTICRFEALQLSFKNMCKLRPLLQKWVEE
ADNNENLQEICKAETLVQARKRKRTSIENRVRGNLENLFLQCPKPTLQQISHIAQQLGLEKDVVRVWFCN
RRQKGKRSSSDYAQREDFEAAGSPFSGGPVSFPLAPGPHFGTPGYGSPHFTALYSSVPFPEGEAFPPVSV
TTLGSPMHSN
```

3.2.2. *Sequence alignment: vertebrates*

```
oct4-human        MAGHLASDFAFS----------PPPGGGGDGPGGPEPGWVDP------------------ 32
pou5f1-Danio      MTERAQSPTAADCRPYEVNRAMYPQAAGLDGLGGASLQFAHGMLQDPSLIFNKAHFNGIT 60
oct4-Gallus       MHVKAKN--------------LLRMCKWLKGLRNA------------------------- 21
                  *  :  .                        .*  ..

oct4-human        ----RTWLSFQGPPGGPGIGPG-------------VGPGSEVWGIPPCPPPYEFCGG--- 72
pou5f1-Danio      PATAQTFFPFSGDFKTNDLQGGDFTQPKHWYPFAAPEFTGQVAGATAATQPANISPPIGE 120
oct4-Gallus       ------------------------------------RGSTWGRSGGRKPMRSSG---- 39
                                                      ... * .    * . .

oct4-human        -----MAYCGPQVGVGLVPQGGLETSQP--------EGEAGVGVESNSDG---------- 109
pou5f1-Danio      TREQIKMPSEVKTEKDVEEYGNEENKPPSQYHLTAGTSSVPTGVNYYTPWNPNFWPGLSQ 180
oct4-Gallus       -----RLPRSADPG-----WGNHANR----------AAVVTRGISSHSPR---------- 69
                             .        *.  .            . .  *:.  :

oct4-human        ----------------------------------------ASPEPCTVTPGAVKLEKEKL 129
pou5f1-Danio      ITAQANISQAPPTPSASSPSLSPSPPGNGFGSPGFFSGGTAQNIPSAQAQSAPRSSGSSS 240
oct4-Gallus       ----------------------------------------VCLCLCQDAP------- 79
                                                           .*

oct4-human        EQNPEESQDIKALQKELEQFAKLLKQKRITLGYTQADVGLTLGVLFGKVFSQTTICRFEA 189
pou5f1-Danio      GGCSDSEEEETLTTEDLEQFAKELKHKRITLGFTQADVGLALGNLYGKMFSQTTICRFEA 300
oct4-Gallus       ------------TSEELEQFAKDLKHKRIMLGFTQADVGLALGTLYGKMFSQTTICRFEA 127
                            ::****** **:*** **:*******:** *:**:***********

oct4-human        LQLSFKNMCKLRPLLQKWVEEADNNENLQEICKAE-TLVQARKRKR-TSIENRVRGNLEN 247
pou5f1-Danio      LQLSFKNMCKLKPLLQRWLNEAENSENPQDMYKIERVFVDTRKRKRRTSLEGTVRSALES 360
oct4-Gallus       LQLSFKNMCKLKPLLQRWLNEAENTDNMQEMCNAEQVLAQARKRKRRTSIETNVKGTLES 187
                  ***********:****:*::**:*.:* *:: : * .:.::***** **:*  *:. **.

oct4-human        LFLQCPKPTLQQISHIAQQLGLEKDVVRVWFCNRRQKGKRSSSDYAQREDFEAAGSPFSG 307
pou5f1-Danio      YFVKCPKPNTLEITHISDDLGLERDVVRVWFCNRRQKGKRLALPFDDECVEAQYYEQSPP 420
oct4-Gallus       FFRKCVKPSPQEISQIAEDLNLDKDVVRVWFCNRRQKGKRLLLPFGNESEGVMYDMNQSL 247
                   * :* **.  :*::*:::*.*::**************    : :.          .

oct4-human        GPVSFPLAPGPHFGTPGYGSPHFTALYSSVPFPEGEAFPPVSVTTLGSPMHSN 360
pou5f1-Danio      PPPHMGGTVLPGQGYPGPAHPGGAPALYMPSLHRPDVFKNGLHPGLVGHLTS- 472
oct4-Gallus       VPPGLP-IPVTSQGYS----LAPSPPVYMPPFHKAEMTPPPLQPGTSMNNSSH 295
                   *  :     .  * .       :.     .: . : *     . :     *
```

3.2.3. *Disorder prediction*

https://st-protein.chem.au.dk/odinpred
https://www.nature.com/articles/s41598-020-71716-1



3.2.4. *Coding DNA sequences, produced protein constructs*

Oct4-aa1-145

Synthesized sequence:

```
Ccgcgggtgagaacctgtacttccaaggcatggcgggtcacctggcgagcgattttgcgtttagcccgccgccgggtggtggtggtgacggtcc
gggtggcccggaaccgggttgggtggatccgcgtacctggctgagcttccaaggtccgccgggtggcccgggtattggtccgggtgtgggcccg
ggtagcgaggtttggggtattccgccgtgcccgccgccgtacgaattttgcggtggcatggcgtattgcggtccgcaagtgggcgttggtctgg
ttccgcaaggtggcctggaaaccagccagccggagggtgaagcgggcgtgggtgttgagagcaacagcgatggtgcgagcccggaaccgtgcac
cgtgaccccgggtgcggttaagctggagaaggaaaaactggagcagaacccggaggaaagccaagatatcaaggcgctgcagaaagaaaacctg
tactttcaaggtggcgcgggtggcgcgggtggccagtataaactgattctgaacggcaagaccctgaaaggtgaaaccaccaccgaagcggtgg
atgcggcgaccgctgagaaggttttcaaacagtacgcgaacgacaacggcgtggatggcgagtggacctatgacgatgcgaccaagacctttac
cgttaccgaaggtggctaaaagctt
```

Translates into:

```
AGENLYFQGMAGHLASDFAFSPPPGGGGDGPGGPEPGWVDPRTWLSFQGPPGGPGIGPGVGPGSEVWGIPPCPPPYEFCGGMAYCGPQVGVGLV
PQGGLETSQPEGEAGVGVESNSDGASPEPCTVTPGAVKLEKEKLEQNPEESQDIKALQKENLYFQGGAGGAGGQYKLILNGKTLKGETTTEAVD
AATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEGG*
```

After TEV-cleavage *(leaving 1 Gly in N-ter, and ENLYFQ in C-ter)*:

```
          10          20          30          40          50          60
 G MAGHLASDFA  FSPPPGGGGD  GPGGPEPGWV  DPRTWLSFQG  PPGGPGIGPG  VGPGSEVWGI

         70          80          90         100         110         120
 PPCPPPYEFC  GGMAYCGPQV  GVGLVPQGGL  ETSQPEGEAG  VGVESNSDGA  SPEPCTVTPG

        130         140
 AVKLEKEKLE  QNPEESQDIK  ALQKENLYFQ
```

Oct4-aa286-360

Synthesized sequence:

```
Ccgcgggtgagaacctgtactttcagggcaagcgtagcagcagcgactatgcgcaacgtgaggatttcgaagcggcgggtagcccgtttagcgg
tggcccggtgagcttcccgctggcgccgggtccgcactttggtaccccgggttatggcagcccgcacttcaccgcgctgtatagcagcgttccg
ttcccggagggtgaagcgtttccgccggtgagcgttaccaccctgggcagcccgatgcacagcaacgaaaatctgtactttcagggtggcgcgg
gtggcgcgggtggccaatataagctgatcctgaacggcaagaccctgaaaggcgaaaccaccaccgaagcggtggatgcggcgaccgctgagaa
ggtttttaaacagtacgcgaacgacaacggtgtggatggcgagtggacctatgacgatgcgaccaaaaccttcaccgttaccgaaggtggctaa
aagctt
```

Translates into:

```
AGENLYFQGKRSSSDYAQREDFEAAGSPFSGGPVSFPLAPGPHFGTPGYGSPHFTALYSSVPFPEGEAFPPVSVTTLGSPMHSNENLYFQGGAG
GAGGQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEGG*
```

After TEV-cleavage *(leaving 1 Gly in N-ter, and ENLYFQ in C-ter)*:

```
    290         300
G KRSSS  DYAQREDFEA

       310          320          330          340          350          360
AGSPFSGGPV  SFPLAPGPHF  GTPGYGSPHF  TALYSSVPFP  EGEAFPPVSV  TTLGSPMHSN  ENLYFQG
```

3.3. *Sox2*

3.3.1. *Sequence alignment: mammals*

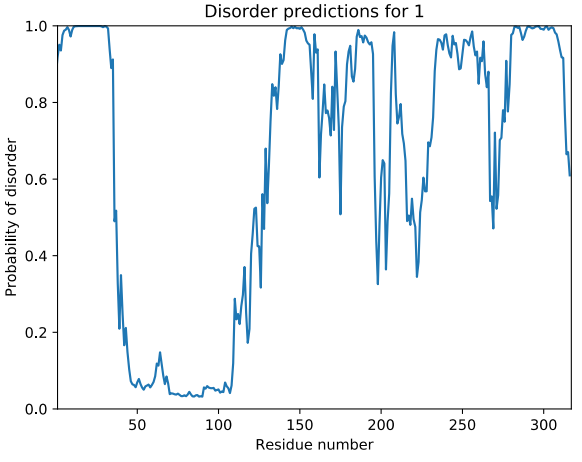
```
S: Serine
P : Proline
SP : phosphorylation motif for MAPKand Cdk
FILVYW : hydrophobic
DE : Asp/Glu
KR : Lys/Arg
T : Thr
XXX : DND-BD in crystal (1GT0)

sox2-human              MYNMMETELKPPGPQQTSGGGGG-------NSTAAAAGGNQKNSPDRVKRPMNAFMVWSR 53
sox2-mus                MYNMMETELKPPGPQQASGGGGGGG-----NATAAATGGNQKNSPDRVKRPMNAFMVWSR 55
sox2-Bos                MYNMMETELKPPGPQQTSGGGGGGG----NSTAAAAGGNQKNSPDRVKRPMNAFMVWSR 56
sox2-Canis              MYNMMETELKPPGPQQTSGGGGGGGGGGGGNSTAAAAGGNQKNSPDRVKRPMNAFMVWSR 60
sox2-Capra              MYNMMETELEQPGLQHNSGGGGGGG----NSTAAAAGGNQKNSPDRVKRPMNAFMVWSR 56
sox2-Balaenoptera       ---------------------------------AAAAGGNQKNSPDRGKRPMNAFMVWSR 27
                                                          ***:********** ************

sox2-human              GQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPR 113
sox2-mus                GQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPR 115
sox2-Bos                GQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPR 116
sox2-Canis              GQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPR 120
sox2-Capra              GQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPR 116
sox2-Balaenoptera       GQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPR 87
                        ********************:****************************************

sox2-human              RKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSM 173
sox2-mus                RKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSM 175
sox2-Bos                RKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSM 176
sox2-Canis              RKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSM 180
sox2-Capra              RKTKTLMKKDKYTLPGGLLAPGGNIMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSM 176
sox2-Balaenoptera       RKTKTLMKKDKYRRAGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSM 147
                        ************  .********* ***************************** ****

sox2-human              MQDQLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTP 233
sox2-mus                MQEQLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTP 235
sox2-Bos                MQDQLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTP 236
sox2-Canis              MQDQLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTP 240
sox2-Capra              MQDQLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTP 236
sox2-Balaenoptera       MQDQLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTP 207
                        **:*************.******************************************

sox2-human              GMALGSMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRDMISMYLPGAEVPEPAAPSRLH 293
sox2-mus                GMALGSMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRDMISMYLPGAEVPEPAAPSRLH 295
sox2-Bos                GMALGSMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRDMISMYLPGAEVPEPAAPSRLH 296
sox2-Canis              GMALGSMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRDMISMYLPGAEVPEPAAPSRLH 300
sox2-Capra              GMALGSMGSVVKSEASSSPPVVTSSSHSRPPCQAGDLRDMISMYLPGDEVPEPAAPSRLH 296
sox2-Balaenoptera       GMALGSMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRDMISMYLPGAEVPEPAAPSRLH 267
                        **:*************************.****************** ************

sox2-human              MSQHYQSGPVPGTAINGTLPLSHM 317
sox2-mus                MAQHYQSGPVPGTAINGTLPLSHM 319
sox2-Bos                MSQHYQSGPVPGTAINGTLPLSHM 320
sox2-Canis              MSQHYQSGPVPGTAINGTLPLSHM 324
sox2-Capra              MSQHYQSGAVPGTAINGILPLSHM 320
sox2-Balaenoptera       MSQHYQSGPVPGTAINGTLPLSHM 291
                        *:*******.******** ******
```

```
>sox2-human
MYNMMETELKPPGPQQTSGGGGGNSTAAAAGGNQKNSPDRVKRPMNAFMVWSRGQRRKMAQENPKMHNSE
ISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKTLMKKDKYTLPGGLLAPGGNSMA
SGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLGYPQHPGLNAHGAAQMQPMHRYDVSALQYNSM
TSSQTYMNGSPTYSMSYSQQGTPGMALGSMGSVVKSEASSSPPVVTSSSHSRAPCQAGDLRDMISMYLPG
AEVPEPAAPSRLHMSQHYQSGPVPGTAINGTLPLSHM
```

3.3.2. *Sequence alignment: vertebrates*

```
sox2-human      MYNMMETELKPPGPQQTSGGGGGNSTAAAAGGNQKNSPDRVKRPMNAFMVWSRGQRRKMA 60
sox2-Danio      MYNMMETELKPPAPQPNTGG-TGNTNSSGN--NQKNSPDRIKRPMNAFMVWSRGQRRKMA 57
sox2-Gallus     MYNMMETELKPPAPQQTSGGGTGNSNSAAN--NQKNSPDRVKRPMNAFMVWSRGQRRKMA 58
                ************.** .:**  **:.:::.   ********:******************

sox2-human      QENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKTLM 120
sox2-Danio      QENPKMHNSEISKRLGAEWKLLSESEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKTLM 117
sox2-Gallus     QENPKMHNSEISKRLGAEWKLLSEAEKRPFIDEAKRLRALHMKEHPDYKYRPRRKTKTLM 118
                ************************:************************************

sox2-human      KKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLGY 180
sox2-Danio      KKDKYTLPGGLLAPGGNGMGAGVGVGAGLGAGVNQRMDSYAHMNGWTNGGYGMMQEQLGY 177
sox2-Gallus     KKDKYTLPGGLLAPGTNTMTTGVGVGATLGAGVNQRMDSYAHMNGWTNGGYGMMQEQLGY 178
                *************** * * :****** ******************:**.*.***:****

sox2-human      PQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTPGMALGSM 240
sox2-Danio      PQHPSLNAHNTAQMQPMHRYDMSALQYNSMTNSQTYMNGSPTYSMSYSQQSTPGMTLGSM 237
sox2-Gallus     PQHPGLNAHNAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTPGMALGSM 238
                ****.****.:**********:*********.******************.****:****

sox2-human      GSVVKSEASSSPPVVTSSSHSRA-PCQAGDLRDMISMYLPGAEVPEPAAPSRLHMSQHYQ 299
sox2-Danio      GSVVKSESSSSPPVVTSSSHSRAGQCQTGDLRDMISMYLPGAEVQDQSAQSRLHMSQHYQ 297
sox2-Gallus     GSVVKTESSSSPPVVTSSSHSRA-PCQAGDLRDMISMYLPGAEVPEPAAPSRLHMSQHYQ 297
                *****:*:***************  **:**************** : :* **********

sox2-human      SGPVPGTAINGTLPLSHM 317
sox2-Danio      SAPVPGTTINGTIPLSHM 315
sox2-Gallus     SAPVPGTAINGTLPLSHM 315
                *.*****:****:*****
```

3.3.3. *Disorder prediction*



3.3.4. *Coding DNA sequences, produced protein constructs*

Sox2-aa1-42

Synthesized sequence :

ccgcgggtgagaacctgtacttccagggcatgtataacatgatggaaaccgaactgaagccgccgggtccgcagcaaaccagcggtggcggtgg
cggtaacagcaccgctgcggcggcgggtggtaaccaaaagaacagcccggaccgtgtgaaataaaagctt

Translates into :

AGENLYFQGMYNMMETELKPPGPQQTSGGGGGNSTAAAAGGNQKNSPDRVK*

After TEV-cleavage *(leaving 1 Gly in N-ter)*:

```
         10          20          30          40
G MYNMMETEL  KPPGPQQTSG  GGGGNSTAAA  AGGNQKNSPD  RVK
```

Sox2-aa115-317_C265A

Synthesized sequence :

cgcgggtgagaacctgtacttccagggcaagaccaaaaccctgatgaagaaagacaagtataccctgccgggtggcctgctggcgccgggtggc
aacagcatggcgagcggtgtgggcgttggtgcgggcctgggtgcgggcgtgaaccagcgtatggacagctacgcgcacatgaacggttggagca
acggcagctacagcatgatgcaggatcaactgggttatccgcaacatccgggtctgaacgcgcatggtgcggcgcagatgcaaccgatgcaccg
ttacgatgttagcgcgctgcagtataacagcatgaccagcagccaaacctatatgaacggcagcccgacctacagcatgagctatagccaacaa
ggtaccccgggtatggcgctgggtagcatgggcagcgtggttaaaagcgaagcgagcagcagcccgccggtggttaccagcagcagccacagcc
gtgcgccggcgcaagcgggtgacctgcgtgatatgatcagcatgtacctgccgggtgcggaagtgccggaaccggctgcgccgagccgtctgca
catgagccagcactatcaaagcggtccggttccgggcaccgcgattaacggtaccctgccgctgagccacatgtaaaagctt

Translates into :

AGENLYFQGKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLGYPQHPGLNAHGAAQMQPMHR
YDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTPGMALGSMGSVVKSEASSSPPVVTSSSHSRAPAQAGDLRDMISMYLPGAEVPEPAAPSRLH
MSQHYQSGPVPGTAINGTLPLSHM*

After TEV-cleavage *(leaving 1 Gly in N-ter)*:

```
     120
G KTKTLM

       130         140         150         160         170         180
KKDKYTLPGG  LLAPGGNSMA  SGVGVGAGLG  AGVNQRMDSY  AHMNGWSNGS  YSMMQDQLGY

       190         200         210         220         230         240
PQHPGLNAHG  AAQMQPMHRY  DVSALQYNSM  TSSQTYMNGS  PTYSMSYSQQ  GTPGMALGSM

       250         260         270         280         290         300
GSVVKSEASS  SPPVVTSSSH  SRAPAQAGDL  RDMISMYLPG  AEVPEPAAPS  RLHMSQHYQS

       310
GPVPGTAING  TLPLSHM
```

Sox2-aa115-187

Coding sequence - mutation from Sox2-aa115-317_C265A :

Ccgcgggtgagaacctgtacttccagggcaagaccaaaaccctgatgaagaaagacaagtataccctgccgggtggcctgctggcgccgggtgg
caacagcatggcgagcggtgtgggcgttggtgcgggcctgggtgcgggcgtgaaccagcgtatggacagctacgcgcacatgaacggttggagc
aacggcagctacagcatgatgcaggatcaactgggttatccgcaacatccgggtctgaactaaaagctt

Translates into :

AGENLYFQGKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLGYPQHPGLN*

After TEV-cleavage *(leaving 1 Gly in N-ter)*:

```
     120
G KTKTLM

       130          140          150          160          170          180
KKDKYTLPGG   LLAPGGNSMA   SGVGVGAGLG   AGVNQRMDSY   AHMNGWSNGS   YSMMQDQLGY

       190
PQHPGLN
```

Sox2-aa115-236

Coding sequence - mutation from Sox2-aa115-317_C265A :

```
ccgcgggtgagaacctgtacttccagggcaagaccaaaaccctgatgaagaaagacaagtataccctgccgggtggcctgctggcgccgggtgg
caacagcatggcgagcggtgtgggcgttggtgcgggcctgggtgcgggcgtgaaccagcgtatggacagctacgcgcacatgaacggttggagc
aacggcagctacagcatgatgcaggatcaactgggttatccgcaacatccgggtctgaacgcgcatggtgcggcgcagatgcaaccgatgcacc
gttacgatgttagcgcgctgcagtataacagcatgaccagcagccaaacctatatgaacggcagcccgacctacagcatgagctatagccaaca
aggtaccccgggtatggcgtaaaagctt
```

Translates into :

```
AGENLYFQGKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLGYPQHPGLNAHGAAQMQPMHR
YDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTPGMA*
```

After TEV-cleavage *(leaving 1 Gly in N-ter)*:

```
     120
G KTKTLM

       130          140          150          160          170          180
KKDKYTLPGG   LLAPGGNSMA   SGVGVGAGLG   AGVNQRMDSY   AHMNGWSNGS   YSMMQDQLGY

       190          200          210          220          230          240
PQHPGLNAHG   AAQMQPMHRY   DVSALQYNSM   TSSQTYMNGS   PTYSMSYSQQ   GTPGMA
```

Sox2-aa115-282_C265A

Coding sequence - mutation from Sox2-aa115-317_C265A :

```
ccgcgggtgagaacctgtacttccagggcaagaccaaaaccctgatgaagaaagacaagtataccctgccgggtggcctgctggcgccgggtgg
caacagcatggcgagcggtgtgggcgttggtgcgggcctgggtgcgggcgtgaaccagcgtatggacagctacgcgcacatgaacggttggagc
aacggcagctacagcatgatgcaggatcaactgggttatccgcaacatccgggtctgaacgcgcatggtgcggcgcagatgcaaccgatgcacc
gttacgatgttagcgcgctgcagtataacagcatgaccagcagccaaacctatatgaacggcagcccgacctacagcatgagctatagccaaca
aggtaccccgggtatggcgctgggtagcatgggcagcgtggttaaaagcgaagcgagcagcagcccgccggtggttaccagcagcagccacagc
cgtgcgccggcgcaagcgggtgacctgcgtgatatgatcagcatgtacctgccgggtgcggaataaaagctt
```

Translates into :

```
AGENLYFQGKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLGYPQHPGLNAHGAAQMQPMHR
YDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTPGMALGSMGSVVKSEASSSPPVVTSSSHSRAPAQAGDLRDMISMYLPGAE*
```

After TEV-cleavage *(leaving 1 Gly in N-ter)*:

```
     120
G KTKTLM

       130         140         150         160         170         180
KKDKYTLPGG  LLAPGGNSMA  SGVGVGAGLG  AGVNQRMDSY  AHMNGWSNGS  YSMMQDQLGY

       190         200         210         220         230         240
PQHPGLNAHG  AAQMQPMHRY  DVSALQYNSM  TSSQTYMNGS  PTYSMSYSQQ  GTPGMALGSM

       250         260         270         280
GSVVKSEASS  SPPVVTSSSH  SRAPAQAGDL  RDMISMYLPG   AE
```

AviTag-Sox2-115-317_C265A

Coding sequence - mutation from Sox2-aa115-317_C265A :

```
Ggtaccggcctgaacgacatttttgaagcgcagaagatcgagtggcacgagggcgcgggcaagaccaagaccctgatgaagaaggacaagtata
ccctgccgggtggcctgctggcgccgggtggcaacagcatggcgagcggtgtgggcgttggtgcgggcctgggtgcgggcgtgaaccagcgtat
ggacagctacgcgcacatgaacggttggagcaacggcagctacagcatgatgcaggatcaactgggttatccgcaacatccgggtctgaacgcg
catggtgcggcgcagatgcaaccgatgcaccgttacgacgttagcgcgctgcagtataacagcatgaccagcagccaaacctatatgaacggta
gcccgacctacagcatgagctatagccaacagggcaccccgggtatggcgctgggtagcatgggcagcgtggttaaaagcgaggcgagcagcag
cccgccggtggttaccagcagcagccacagccgtgcgccggcgcaggcgggtgacctgcgtgatatgatcagcatgtacctgccgggtgcggaa
gtgccggaaccggcggcgccgagccgtctgcacatgagccaacactatcagagcggtccggttccgggcaccgcgattaacggcaccctgccgc
tgagccatatgtaaaagctt
```

Translates into :

```
GTGLNDIFEAQKIEWHEGAGKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLGYPQHPGLNA
HGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTPGMALGSMGSVVKSEASSSPPVVTSSSHSRAPAQAGDLRDMISMYLPGAE
VPEPAAPSRLHMSQHYQSGPVPGTAINGTLPLSHM*
```

Expressed peptide:

```
MAHHHHHHVGTGLNDIFEAQKIEWHEGAGKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLG
YPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTPGMALGSMGSVVKSEASSSPPVVTSSSHSRAPAQAGDLRDM
ISMYLPGAEVPEPAAPSRLHMSQHYQSGPVPGTAINGTLPLSHM*
```

```
                                    120
MAHHHHHHVGTGLNDIFEAQKIEWHEGAGG  KTKTLM

       130         140         150         160         170         180
KKDKYTLPGG  LLAPGGNSMA  SGVGVGAGLG  AGVNQRMDSY  AHMNGWSNGS  YSMMQDQLGY

       190         200         210         220         230         240
PQHPGLNAHG  AAQMQPMHRY  DVSALQYNSM  TSSQTYMNGS  PTYSMSYSQQ  GTPGMALGSM

       250         260         270         280         290         300
GSVVKSEASS  SPPVVTSSSH  SRAPAQAGDL  RDMISMYLPG  AEVPEPAAPS  RLHMSQHYQS

       310
GPVPGTAING  TLPLSHM
```

AviTag-Sox2-115-240

Coding sequence - mutation from Sox2-aa115-317_C265A :

```
Ggtaccggcctgaacgacatttttgaagcgcagaagatcgagtggcacgagggcgcgggcaagaccaagaccctgatgaagaaggacaagtata
ccctgccgggtggcctgctggcgccgggtggcaacagcatggcgagcggtgtgggcgttggtgcgggcctgggtgcgggcgtgaaccagcgtat
ggacagctacgcgcacatgaacggttggagcaacggcagctacagcatgatgcaggatcaactgggttatccgcaacatccgggtctgaacgcg
catggtgcggcgcagatgcaaccgatgcaccgttacgacgttagcgcgctgcagtataacagcatgaccagcagccaaacctatatgaacggta
gcccgacctacagcatgagctatagccaacagggcaccccgggtatggcgctgggtagcatgtaaaagctt
```

Translates into :

```
GTGLNDIFEAQKIEWHEGAGKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLGYPQHPGLNA
HGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTPGMALGSM*
```

Expressed peptide:

```
MAHHHHHHVGTGLNDIFEAQKIEWHEGAGKTKTLMKKDKYTLPGGLLAPGGNSMASGVGVGAGLGAGVNQRMDSYAHMNGWSNGSYSMMQDQLG
YPQHPGLNAHGAAQMQPMHRYDVSALQYNSMTSSQTYMNGSPTYSMSYSQQGTPGMALGSM*
```

```
                                  120
MAHHHHHHVGTGLNDIFEAQKIEWHEGAGG  KTKTLM

       130          140          150          160          170          180
KKDKYTLPGG  LLAPGGNSMA  SGVGVGAGLG  AGVNQRMDSY  AHMNGWSNGS  YSMMQDQLGY

       190          200          210          220          230          240
PQHPGLNAHG  AAQMQPMHRY  DVSALQYNSM  TSSQTYMNGS  PTYSMSYSQQ  GTPGMALGSM
```

AviTag-Sox2-234-317_C265A

Coding sequence - mutation from Sox2-aa115-317_C265A :

```
ggtaccggcctgaacgacatttttgaagcgcagaagatcgagtggcacgagggcgcgggtatggcgctgggtagcatgggcagcgtggttaaaa
gcgaggcgagcagcagcccgccggtggttaccagcagcagccacagccgtgcgccggcgcaggcgggtgacctgcgtgatatgatcagcatgta
cctgccgggtgcggaagtgccggaaccggcggcgccgagccgtctgcacatgagccaacactatcagagcggtccggttccgggcaccgcgatt
aacggcaccctgccgctgagccatatgtaaaagctt
```

Translates into :

```
GTGLNDIFEAQKIEWHEGAGMALGSMGSVVKSEASSSPPVVTSSSHSRAPAQAGDLRDMISMYLPGAEVPEPAAPSRLHMSQHYQSGPVPGTAI
NGTLPLSHM*
```

Expressed peptide:

```
MAHHHHHHVGTGLNDIFEAQKIEWHEGAGMALGSMGSVVKSEASSSPPVVTSSSHSRAPAQAGDLRDMISMYLPGAEVPEPAAPSRLHMSQHYQ
SGPVPGTAINGTLPLSHM*
```

```
                                  140
MAHHHHHHVGTGLNDIFEAQKIEWHEGAG  MALGSM

       250          260          270          280          290          300
GSVVKSEASS  SPPVVTSSSH  SRAPAQAGDL  RDMISMYLPG  AEVPEPAAPS  RLHMSQHYQS

       310
GPVPGTAING  TLPLSHM
```

3.4. *Nanog*

3.4.1. *Sequence alignment: mammals*

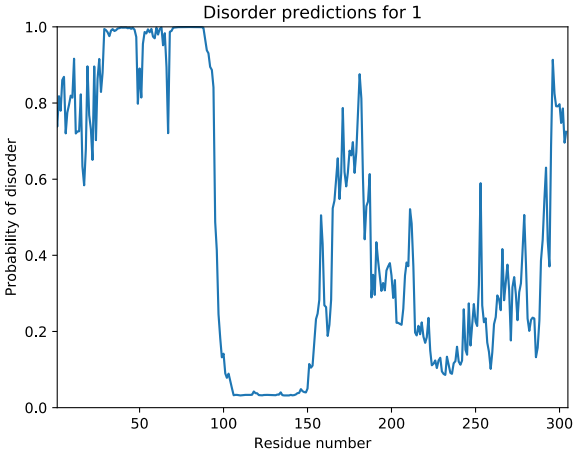
S: Serine
P : Proline
SP : phosphorylation motif for MAPKand Cdk
FILVYW : hydrophobic
DE : Asp/Glu
KR : Lys/Arg
T : Thr
XXX: DND-BD in crystal (4RBO)

```
Nanog-human           MSVDPACPQSLPCFEASDCKESSPMPVICGPEENYPSLQMSSAEMPHTETVSPLPSSMDLLIQDSPDSST 70
nanog-Mus             MSVGLPGPHSLPSSEEASNSGNASSMPAVFHP-ENYSCLQGSATEMLCTEAASPRPS-SE 58
NANOG-Bos             MSVGPACPQSLLCGPEASNSRESSPMP-----EESYVSLQTSSADTLDTDTVSPLPSS-SM 53
Nanog-Canis           ---------MPA-GPQAPNSRDPSPMPEVYGPRGNPASLPMSSAETPHAETVSPLPSS-SM 49
Nanog-Capra           MSVDPACPQSLLCGPEASNSGESSPMP-----EESYASLQMSSADTLDTDTVSPLPSS-SM 53
NANOG-Balaenoptera    MSVDPACPQSLRCGPEASNSRESSPMPEIYGPEENYVSLQMSSVETHDMETVSPLPSFSM 59
                                 :*. . :.*.:*       .   *  *:.:    ::.** ** .

Nanog-human           KGKQPTSAEKSVAKKEDKVPVKKQKTRTVFSSTQLCVLNDRFQRQKYLSLQQMQELSNILNLSYKQVK 140
nanog-Mus             KGPEEEE-NKVLARKQKMRTVFSQAQLCALKDRFQK 117
NANOG-Bos             KGPEEEE-STEK-EETVPVKKQKIRTVFSQTQLCVLNDRFQR 111
Nanog-Canis           KGEE-RTARKEDATQGKKQKMRTVFSQTQLYVLNDRFQR 108
Nanog-Capra           KGEE-STEK-EEKVPVKKQKIRTVFSQTQLCVLNDRFQR 111
NANOG-Balaenoptera    KGPDSSTSPRVKLLATAADK-STEKKEEKVLLKKQKTRTVFSQTQLCVLNDRFQR 118
                      *:****:.*: *   .   .    .      :*** *****.:** .*:****:

Nanog-human           TWFQNQRMKSKRWQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQT 210
nanog-Mus             QKYLSLQQMQELSSIINLLSYKQVKTWFQNQRMKCKRWQKNQWLKTSNGLLQKGSAPVEYP 177
NANOG-Bos             QKYLSLQQMQELSNILNLSYKQVKTWFQNQRMKCKKWQKNNWPRNSNGMPQGP-AMAEYP 170
Nanog-Canis           QKYLSLQQMQELSNILNLSYKQVKTWFQNQRMKSKRWQKSNWPKESNSVTQNSSATTEYA 168
Nanog-Capra           QKYLSLQQMQELSNILNLSYKQVKTWFQNQRMKCKKWQKNNWPRNSNDVPQDP-ATAEYP 170
NANOG-Balaenoptera    QKYLSLQQMQELSNILNLSYKQVKTWFQNQRMKCKRWQKNNWPRNSNTVTQGP-ATTEYP 177
                      *:********:**.**.****************.*:***.:* : .* : *   :   *.

Nanog-human           QNIQSWSNHSWNTQTWCTQSWNNQAWNSPFYNCGEESLQSCMQFQPNSPASDLEAALEAAGEGLNVIQQT 280
nanog-Mus             QSIHCSYPQGYLVNASGSLSMWGSQTWTNPTWSSQTWTNPTWNNQTWTNPTWSSQAWTAQS 237
NANOG-Bos             GFYS-YHQGCLVNSPGNLPMWG----------NQTWNNPTWSNQSWNSQSWSNHSWNSQA 219
Nanog-Canis           GFYP-CRQGYLLNPSGNLPLW----------SSQAWNNPNWSSQTWNSQSWSSHSWNSQT 217
Nanog-Capra           SFYS-YHQGCLVNSPRNMPMWG----------NQTWNNPTWSNQNWNSQSWSNHSWNSQA 219
NANOG-Balaenoptera    GFYS-YHQGCLANSSGNLPMWG----------NQTWNNPTWSNQSWNSQSWSNHPWNNQT 226
                      .::    ** * *.. .:.:*           .*:*.. .*..*. .  :**.:.*. *:

Nanog-human           QQTTRYFSTPQTMDLFLNYSMNMQPEDV 308
nanog-Mus             AWTAQSQAPVEYP 250
NANOG-Bos             AWNSQSWSNHSWNTQTWCTQSWNNQAWNSPFYNCGEESLQSCMQFQPNSPVSDLEATLETAGESYNIIQQTAKYF 285
Nanog-Canis           SWNSQTWNSQSWSSHSWNSQTWCTQSWNNQAWNNPFYNCGEESLQSCMQFQPNSPASDLEAALEAAGEGLNVIQQT 280
Nanog-Capra           AWNSQSWSNHSWNTQTWCTQSWNNQAWNSPFYNCGEESLQSCMQFQPNSPASDLEAALEAAGEGLNVIQQT 278
NANOG-Balaenoptera    QWNNQTWCTQSWNNQAWNSPFYNCGEESLQSCMQFQPNSPASDLEAALEAAGEGLNVIQQT 285
                      *:  *  ::* **.
```

>Nanog-human
MSVDPACPQSLPCFEASDCKESSPMPVICGPEENYPSLQMSSAEMPHTETVSPLPSSMDLLIQDSPDSST
SPKGKQPTSAEKSVAKKEDKVPVKKQKTRTVFSSTQLCVLNDRFQRQKYLSLQQMQELSNILNLSYKQVK
TWFQNQRMKSKRWQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQT
QNIQSWSNHSWNTQTWCTQSWNNQAWNSPFYNCGEESLQSCMQFQPNSPASDLEAALEAAGEGLNVIQQT
TRYFSTPQTMDLFLNYSMNMQPEDV

3.4.2. *Sequence alignment: vertebrates*

3.4.3. *Disorder prediction*



3.4.4. *Coding DNA sequences, produced protein constructs*

Nanog-aa1-85

Name Genscript : TEV-NanogNter

Coding sequence (mutated from Nanog-aa1-85-Tev):

Caattggtgaaaatctgtacttccagggcatgtccgtcgatccggcgtgtccgcagagcctgccgtgctttgaagcgagcgactgtaaagaatc
gagcccgatgccggtcatttgcggcccggaagaaaactatccgtctctgcagatgagctctgcagaaatgccgcatacggaaaccgtgagcccg
ctgccgagttccatggatctgctgatccaggatagtccggactcatcgacgtccccgaaaggtaaacaaccgaccagcgcggaaaaatctgtgg
cctaaaagctt

Translates into :

IGENLYFQGMSVDPACPQSLPCFEASDCKESSPMPVICGPEENYPSLQMSSAEMPHTETVSPLPSSMDLLIQDSPDSSTSPKGKQPTSAEKSVA
*

Before TEV-cleavage:

GST-His-Tev_NanogCter_aa1-85

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISM
LEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQ
IDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDGSTSGSGHHHHHHSAGLVPRGSTAIGENLYFQGMSVDPACPQSLPCFEASDCKESSPMPVIC
GPEENYPSLQMSSAEMPHTETVSPLPSSMDLLIQDSPDSSTSPKGKQPTSAEKSVA*

Number of amino acids: 338
Molecular weight: 38016.43
Theoretical pI: 5.54
Total number of negatively charged residues (Asp + Glu): 47
Total number of positively charged residues (Arg + Lys): 35
Ext. coefficient 46340
Abs 0.1% (=1 g/l) 1.219, assuming all pairs of Cys residues form cystines
Ext. coefficient 45840
Abs 0.1% (=1 g/l) 1.206, assuming all Cys residues are reduced

After Tev-cleavage *(leaving 1 Gly in N-ter)*:

GMSVDPACPQSLPCFEASDCKESSPMPVICGPEENYPSLQMSSAEMPHTETVSPLPSSMDLLIQDSPDSSTSPKGKQPTSAEKSVA

```
          10          20          30          40          50          60
 G MSVDPACPQS  LPCFEASDCK  ESSPMPVICG  PEENYPSLQM  SSAEMPHTET  VSPLPSSMDL

         70          80
 LIQDSPDSST  SPKGKQPTSA  EKSVA
```

Nanog-aa154-305-Tev

Synthesized sequence :

Ccgcgggtgagaacctgtacttccaaggcatggcgggtcacctggcgagcgattttgcgtttagcccgccgccgggtggtggtggtgacggtcc
gggtggcccggaaccgggttgggtggatccgcgtacctggctgagcttccaaggtccgccgggtggcccgggtattggtccgggtgtgggcccg
ggtagcgaggtttggggtattccgccgtgcccgccgccgtacgaattttgcggtggcatggcgtattgcggtccgcaagtgggcgttggtctgg
ttccgcaaggtggcctggaaaccagccagccggagggtgaagcgggcgtgggtgttgagagcaacagcgatggtgcgagcccggaaccgtgcac
cgtgaccccgggtgcggttaagctggagaaggaaaaactggagcagaacccggaggaaagccaagatatcaaggcgctgcagaaagaaaacctg
tactttcaaggtggcgcgggtggcgcgggtggccagtataaactgattctgaacggcaagaccctgaaaggtgaaaccaccaccgaagcggtgg
atgcggcgaccgctgagaaggttttcaaacagtacgcgaacgacaacggcgtggatggcgagtggacctatgacgatgcgaccaagacctttac
cgttaccgaaggtggctaaaagctt

Translates into :

ENLYFQGQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWCTQSWNNQAWNSPF
YNCGEESLQSCMQFQPNSPASDLEAALEAAGEGLNVIQQTTRYFSTPQTMDLFLNYSMNMQPEDVENLYFQGGAGGAGGQYKLILNGKTLKGET
TTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEGG*

Bfore TEV-cleavage:

GST-His-Tev_NanogCter_aa154-305_Tev-GB1 from pET41a+

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISM
LEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQ
IDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDGSTSGSGHHHHHHSAGLVPRGSTAIGMKETAENLYFQGQKNNWPKNSNGVTQKASAPTYPSL
YSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWCTQSWNNQAWNSPFYNCGEESLQSCMQFQPNSPASDLEAALEAAG
EGLNVIQQTTRYFSTPQTMDLFLNYSMNMQPEDVENLYFQGGAGGAGGQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDD
ATKTFTVTEGG

After Tev-cleavage *(leaving 1 Gly in N-ter and ENLYFQ in C-ter)*:

GQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWCTQSWNNQAWNSPFYNCGEE
SLQSCMQFQPNSPASDLEAALEAAGEGLNVIQQTTRYFSTPQTMDLFLNYSMNMQPEDVENLYFQ

```
       160         170         180
 G QKNNWPK   NSNGVTQKAS  APTYPSLYSS

        190         200         210         220         230         240
 YHQGCLVNPT  GNLPMWSNQT  WNNSTWSNQT  QNIQSWSNHS  WNTQTWCTQS  WNNQAWNSPF

        250         260         270         280         290         300
 YNCGEESLQS  CMQFQPNSPA  SDLEAALEAA  GEGLNVIQQT  TRYFSTPQTM  DLFLNYSMNM

QPEDV ENLYFQ
```

Nanog-aa154-305

Coding sequence (mutated from Nanog-aa154-305-Tev):

```
ccgcgggtgaaaacctgtacttccagggtcaaaagaacaactggccgaaaaacagcaacggtgtgacccaaaaggcgagcgcgccgacctatcc
gagcctgtacagcagctatcaccagggttgcctggttaacccgaccggcaacctgccgatgtggagcaaccaaacctggaacaacagcacctgg
agcaaccagacccaaaacatccagagctggagcaaccacagctggaacacccagacctggtgcacccaaagctggaacaaccaggcgtggaaca
gcccgttctacaactgcggcgaggaaagcctgcaaagctgcatgcagtttcaaccgaacagcccggcgagcgacctggaggcggcgctggaagc
ggcgggtgaaggcctgaacgtgatccagcaaaccacccgttacttcagcaccccgcaaaccatggacctgtttctgaactatagcatgaacatg
cagccggaggatgtttaaaagctt
```

Translates into :

```
ENLYFQGQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWCTQSWNNQAWNSPF
YNCGEESLQSCMQFQPNSPASDLEAALEAAGEGLNVIQQTTRYFSTPQTMDLFLNYSMNMQPEDVENLYFQGGAGGAGGQYKLILNGKTLKGET
TTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEGG*
```

After Tev-cleavage *(leaving 1 Gly in N-ter)*:

```
GQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWCTQSWNNQAWNSPFYNCGEE
SLQSCMQFQPNSPASDLEAALEAAGEGLNVIQQTTRYFSTPQTMDLFLNYSMNMQPEDV*
```

```
      160         170         180
G QKNNWPK  NSNGVTQKAS  APTYPSLYSS

       190         200         210         220         230         240
YHQGCLVNPT  GNLPMWSNQT  WNNSTWSNQT  QNIQSWSNHS  WNTQTWCTQS  WNNQAWNSPF

       250         260         270         280         290         300
YNCGEESLQS  CMQFQPNSPA  SDLEAALEAA  GEGLNVIQQT  TRYFSTPQTM  DLFLNYSMNM

 QPEDV
```

Nanog-aa154-215-Tev

Coding sequence (mutated from Nanog-aa154-305-Tev):

```
Ccgcgggtgaaaacctgtacttccagggtcaaaagaacaactggccgaaaaacagcaacggtgtgacccaaaaggcgagcgcgccgacctatcc
gagcctgtacagcagctatcaccagggttgcctggttaacccgaccggcaacctgccgatgtggagcaaccaaacctggaacaacagcacctgg
agcaaccagacccaaaacatccagagcgaaaacctgtactttcaaggtggcgcgggtggcgcgggtggccagtataagctgattctgaacggca
agaccctgaaaggcgaaaccaccaccgaggcggtggatgcggcgaccgctgagaaggttttcaaacagtacgcgaacgacaacggtgtggatgg
cgaatggacctatgacgatgcgaccaaaaacctttaccgttaccgagggtggctaaaagctt
```

Translates into:

```
AGENLYFQGQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQTQNIQSENLYFQGGAGGAGGQYKLILNGK
TLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEGG
```

After Tev-cleavage *(leaving 1 Gly in N-ter and ENLYFQ in C-ter)*:

```
GQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQTQNIQSENLYFQ
```

```
      160         170         180
G QKNNWPK  NSNGVTQKAS  APTYPSLYSS

       190         200         210
YHQGCLVNPT  GNLPMWSNQT  WNNSTWSNQT  QNIQS  ENLYFQ
```

Nanog-aa154-272-Tev

Coding sequence (mutated from Nanog-aa154-305-Tev):

Ccgcgggtgaaaacctgtacttccagggtcaaaagaacaactggccgaaaaacagcaacggtgtgacccaaaaggcgagcgcgccgacctatcc
gagcctgtacagcagctatcaccagggttgcctggttaacccgaccggcaacctgccgatgtggagcaaccaaacctggaacaacagcacctgg
agcaaccagacccaaaacatccagagctggagcaaccacagctggaacacccagacctggtgcacccaaagctggaacaaccaggcgtggaaca
gcccgttctacaactgcggcgaggaaagcctgcaaagctgcatgcagtttcaaccgaacagcccggcgagcgacctggaggcggcgctggaagc
ggcgggtgaagaaaacctgtactttcaaggtggcgcgggtggcgcgggtggccagtataagctgattctgaacggcaagaccctgaaaggcgaa
accaccaccgaggcggtggatgcggcgaccgctgagaaggttttcaaacagtacgcgaacgacaacggtgtggatggcgaatggacctatgacg
atgcgaccaaaaccctttaccgttaccgagggtggctaaaagctt

Translates into:

AGENLYFQGQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWCTQSWNNQAWNS
PFYNCGEESLQSCMQFQPNSPASDLEAALEAAGEENLYFQGGAGGAGGQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDD
ATKTFTVTEGG*

After Tev-cleavage *(leaving 1 Gly in N-ter and ENLYFQ in C-ter)*:

GQKNNWPKNSNGVTQKASAPTYPSLYSSYHQGCLVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWCTQSWNNQAWNSPFYNCGEE
SLQSCMQFQPNSPASDLEAALEAAGEENLYFQ

```
      160         170         180
 G QKNNWPK  NSNGVTQKAS  APTYPSLYSS

       190         200         210         220         230         240
 YHQGCLVNPT  GNLPMWSNQT  WNNSTWSNQT  QNIQSWSNHS  WNTQTWCTQS  WNNQAWNSPF

       250         260         270
 YNCGEESLQS  CMQFQPNSPA  SDLEAALEAA  GE ENLYFQG
```

Nanog-aa154-305_4C4A

Coding sequence (mutated from Nanog-aa154-305-Tev):

Ccgcgggtgagaatctgtatttccaaggcaaaaacaactggccgaagaacagcaatggtgtgacccaaaaagcgagcgcgccgacctatccgag
cctgtacagcagctatcaccaaggtgcgctggtgaacccgaccggtaacctgccgatgtggagcaaccaaacctggaacaacagcacctggagc
aaccagacccaaaacatccagagctggagcaaccacagctggaacacccagacctgggcgacccaaagctggaacaaccaggcgtggaacagcc
cgttctacaacgcgggcgaggaaagcctgcagagcgcgatgcagtttcaaccgaacagcccggcgagcgacctggaggcggcgctggaagcggc
gggtgaaggcctgaacgttattcagcaaaccacccgttatttcagcaccccgcaaaccatggacctgtttctgaattatagcatgaatatgcag
ccggaggatgtgtaaaagctt

Translates into:

AGENLYFQGKNNWPKNSNGVTQKASAPTYPSLYSSYHQGALVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWATQSWNNQAWNSP
FYNAGEESLQSAMQFQPNSPASDLEAALEAAGEGLNVIQQTTRYFSTPQTMDLFLNYSMNMQPEDV*

After Tev-cleavage *(leaving 1 Gly in N-ter)*

```
GKNNWPKNSNGVTQKASAPTYPSLYSSYHQGALVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWATQSWNNQAWNSPFYNAGEES
LQSAMQFQPNSPASDLEAALEAAGEGLNVIQQTTRYFSTPQTMDLFLNYSMNMQPEDV
```

```
      160         170         180
G QKNNWPK  NSNGVTQKAS  APTYPSLYSS

       190         200         210         220         230         240
YHQGALVNPT  GNLPMWSNQT  WNNSTWSNQT  QNIQSWSNHS  WNTQTWATQS  WNNQAWNSPF

       250         260         270         280         290         300
YNAGEESLQS  AMQFQPNSPA  SDLEAALEAA  GEGLNVIQQT  TRYFSTPQTM  DLFLNYSMNM

 QPEDV
```

Nanog-aa154-272_4C4A

Coding sequence (mutated from Nanog-aa154-305_4C4A):

```
Ccgcgggtgagaatctgtatttccaaggcaaaaacaactggccgaagaacagcaatggtgtgacccaaaaagcgagcgcgccgacctatccgag
cctgtacagcagctatcaccaaggtgcgctggtgaacccgaccggtaacctgccgatgtggagcaaccaaacctggaacaacagcacctggagc
aaccagacccaaaacatccagagctggagcaaccacagctggaacacccagacctgggcgacccaaagctggaacaaccaggcgtggaacagcc
cgttctacaacgcgggcgaggaaagcctgcagagcgcgatgcagtttcaaccgaacagcccggcgagcgacctggaggcggcgctggaagcggc
gggtgaataaaagctt
```

Translates into:

```
AGENLYFQGKNNWPKNSNGVTQKASAPTYPSLYSSYHQGALVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWATQSWNNQAWNSP
FYNAGEESLQSAMQFQPNSPASDLEAALEAAGE
```

After Tev-cleavage *(leaving 1 Gly in N-ter)*:

```
GKNNWPKNSNGVTQKASAPTYPSLYSSYHQGALVNPTGNLPMWSNQTWNNSTWSNQTQNIQSWSNHSWNTQTWATQSWNNQAWNSPFYNAGEES
LQSAMQFQPNSPASDLEAALEAAGE
```

```
      160         170         180
G QKNNWPK  NSNGVTQKAS  APTYPSLYSS

       190         200         210         220         230         240
YHQGALVNPT  GNLPMWSNQT  WNNSTWSNQT  QNIQSWSNHS  WNTQTWATQS  WNNQAWNSPF

       250         260         270
YNAGEESLQS  AMQFQPNSPA  SDLEAALEAA  GE
```

3.5. *Esrrb*

3.5.1. *Sequence alignment: mammals*

S: Serine
P : Proline
SP : phosphorylation motif for MAPKand Cdk
FILVYW : hydrophobic
DE : Asp/Glu
KR : Lys/Arg
T : Thr
C : Cys

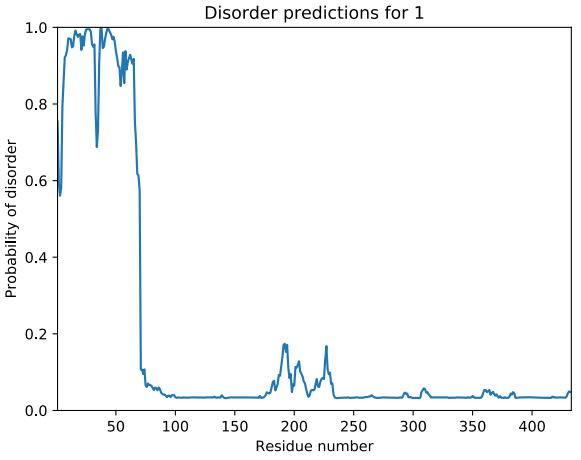
```
human             ----------------------MSSEDRHLGSSCGSFIKTEPSSPSSGIDALSHHSPSGSS
mouse             ----------------------MSSEDRHLGSSCGSFIKTEPSSPSSGIDALSHHSPSGSS
capra             MDVSELCVPDPLGYHNQLLNRMSADDRHLSSSCGSFIKTEPSSPSSGIDALSHHSPSGSS
bos               MDVSELCVPDPLGYHNQLLNRMSADDRHLGSSCGSFIKTEPSSPSSGIDALSHHSPSGSS
balaenoptera      MDVSELCIPDPLGYHNQLLNRMSADDRHLVSSCGSFIKTEPSSPSSGIDALSHHSPRGSS
                                       **::**** ************************** ***

human             DASGGFGIALSTHANGLDSPPMFAGAGLGGNPCRKSYEDCTSGIMEDSAIKCEYMLNAIP
mouse             DASGGFGIALSTHANGLDSPPMFAGAGLGGNPCRKSYEDCTSGIMEDSAIKCEYMLNAIP
capra             DASGGFGLALGAHANGLDSPPMFAGAGLGGTPCRKGYEDCAGGLMEDSAIKCEYMLNAIP
bos               DASGGFGLALGAHANGLDSPPMFAGAGLGGTPCRKGYEDCAGGLMEDSAIKCEYMLNAIP
balaenoptera      DASGGFGLALGAHANGLDSPPMFAGAGLGGTPCRKGYEDCASGIMEDSAIKCEYMLNAIP
                  *******:**.:******************.****.****:.*:****************

human             KRL
mouse             KRL
capra             KRL
bos               KRL
balaenoptera      KRL
                  ***
```

```
>sp|O95718|ERR2_HUMAN Steroid hormone receptor ERR2 OS=Homo sapiens GN=ESRRB PE=1 SV=2
MSSDDRHLGSSCGSFIKTEPSSPSSGIDALSHHSPSGSSDASGGFGLALGTHANGLDSPP
MFAGAGLGGTPCRKSYEDCASGIMEDSAIKCEYMLNAIPKRLCLVCGDIASGYHYGVASC
EACKAFFKRTIQGNIEYSCPATNECEITKRRRKSCQACRFMKCLKVGMLKEGVRLDRVRG
GRQKYKRRLDSESSPYLSLQISPPAKKPLTKIVSYLLVAEPDKLYAMPPPGMPEGDIKAL
TTLCDLADRELVVIIGWAKHIPGFSSLSLGDQMSLLQSAWMEILILGIVYRSLPYDDKLV
YAEDYIMDEEHSRLAGLLELYRAILQLVRRYKKLKVEKEEFVTLKALALANSDSMYIEDL
EAVQKLQDLLHEALQDYELSQRHEEPWRTGKLLLTLPLLRQTAAKAVQHFYSVKLQGKVP
MHKLFLEMLEAKVGQEQLRGSPKDERMSSHDGKCPFQSAAFTSRDQSNSPGIPNPRPSSP TPLNERGRQISPSTRTPGGQGKHLWLTM
```

3.5.2. *Sequence alignment: vertebrates*

```
C: cysteines in N-ter
XX: DNA-BD ordered in NMR structure 1LO1 (construct: human aa96-194_C163A)
XX: folded: aa235-432 in Ligand-Binding Domain crystal structures 6LIT, 6LN4 (construct: human aa204-
433_Y215H, mutation for solubility/stability)


mouse         ---------------------MSSEDRHLGSSCGSFIKTEPSSPSSGIDALSHHSPSGSS   39
human         ---------------------MSSDDRHLGSSCGSFIKTEPSSPSSGIDALSHHSPSGSS   39
goat          MDVSELCVPDPLGYHNQLLNRMSADDRHLSSSCGSFIKTEPSSPSSGIDALSHHSPSGSS   60
whale         MDVSELCIPDPLGYHNQLLNRMSADDRHLVSSCGSFIKTEPSSPSSGIDALSHHSPRGSS   60
chicken       MDISELCISDPLGYHNQLLGRMATEERHLSSSCGSFIKTEPSSPSSGIDAISHHSPSGSS   60
fish          ---------------------MAADERHLPSSCGSYIKTEPSSPSSVIDTVSHHSPSGNS   39
                                   *:::::*** *****:********** **::***** *.*

mouse         DASGGFGIALSTHANGLDSPPMFAGAGLGGNPCRKSYEDCTSGIMEDSAIKCEYMLNAIP   99
human         DASGGFGLALGTHANGLDSPPMFAGAGLGGTPCRKSYEDCASGIMEDSAIKCEYMLNAIP   99
goat          DASGGFGLALGAHANGLDSPPMFAGAGLGGTPCRKGYEDCAGGLMEDSAIKCEYMLNAIP   120
whale         DASGGFGLALGAHANGLDSPPMFAGAGLGGTPCRKGYEDCASGIMEDSAIKCEYMLNAIP   120
chicken       DASGGYGIAMGGHPNGLDSPPMFNGTGIGGGSCRKRYDDCASAIMEDSPTKCEYMLNAIP   120
fish          DASGGYVSTMNSHSNGLDSPPMFTPSGLGAGTCRKRYDDCSSTIMEDSSIKCEYMLNSLP   99
              *****:  ::. * *********  :*:*.  *** *:**:. :****  *******::*

mouse         KRLCLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYNCPATNECEITKRRRKSCQACR   159
human         KRLCLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPATNECEITKRRRKSCQACR   159
goat          KRLCLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPATNECEITKRRRKSCQACR   180
whale         KRLCLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPATNECEITKRRRKSCQACR   180
chicken       KRLCLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPATNECEITKRRRKSCQACR   180
fish          KRLCLVCGDIASGYHYGVASCEACKAFFKRTIQGNIEYSCPATNECEITKRRRKSCQACR   159
              ************************************.**********************

mouse         FMKCLKVGMLKEGVRLDRVRGGRQKYKRRLDSENSPYLNLPISPPAKKPLTKIVSNLLGV   219
human         FMKCLKVGMLKEGVRLDRVRGGRQKYKRRLDSESSPYLSLQISPPAKKPLTKIVSYLLVA   219
goat          FMKCLKVGMLKEGVRLDRVRGGRQKYKRRLDSESSPYLSLQISPPAKKPLTKIVSYLLVA   240
whale         FMKCLKVGMLKEGVRLDRVRGGRQKYKRRLDSESSPYLSLQISPPAKKPLTKIVSYLLVA   240
chicken       FMKCLKVGMLKEGVRLDRVRGGRQKYKRRLDSESSTYLSLQIPPPAKKPLTKIVSHLLVA   240
fish          FMKCLKVGMLKEGVRLDRVRGGRQKYKRRLDSENNPYLGLTLPPPTKKPLTKIVSHLLVA   219
              **********************************.. **.* : **:********* ** .

mouse         EQDKLYAMPPNDIPEGDIKALTTLCELADRELVFLINWAKHIPGFPSLTLGDQMSLLQSA   279
human         EPDKLYAMPPPGMPEGDIKALTTLCDLADRELVVIIGWAKHIPGFSSLSLGDQMSLLQSA   279
goat          EPDKLYAMPPPGMPEGDIKALTTLCDLADRELVVIIGWAKHIPGFSNLSLGDQMSLLQSA   300
whale         EPDKLYAMPPPGMPEGDIKALTTLCDLADRELVVIIGWAKHIPGFSNLSLGDQMSLLQSA   300
chicken       EPEKIYAMPDPTMPESDIKALTTLCDLADRELVVIIGWAKHIPGFSNLSLGDQMSLLQSA   300
fish          EPEKIYAMPDPTMPESDIKALTTLCDLADRELVVIIGWAKHIPGFSTLSLGDQMSLLQSA   279
              * :*:****   :**.*********:*******.:*.******** .*:***********

mouse         WMEILILGIVYRSLPYDDKLAYAEDYIMDEEHSRLVGLLDLYRAILQLVRRYKKLKVEKE   339
human         WMEILILGIVYRSLPYDDKLVYAEDYIMDEEHSRLAGLLELYRAILQLVRRYKKLKVEKE   339
goat          WMEILILGIVYRSLPYDDKLVYAEDYIMDEEHSRLAGLLELYRAILQLVRRYKKLKVEKE   360
whale         WMEILILGIVYRSLPYDDKLVYAEDYIMDEEHSRLAGLLELYRAILQLVRRYKKLKVEKE   360
chicken       WMEILILGIVYRSLPYEDKLVYAEDYIMDEEHSRLTGLLELYLAILQLVRRYKKLKVEKE   360
fish          WMEILILSIVFRSLPYEDELVYAEDYIMDEEHSRLTGLLDLYVSILQLVRKYKKLKVEKE   339
              *******.**:*****:*:*.***************.***:** :******:*********

mouse         EFMILKALALANSDSMYIENLEAVQKLQDLLHEALQDYELSQRHEEPRRAGKLLLTLPLL   399
human         EFVTLKALALANSDSMYIEDLEAVQKLQDLLHEALQDYELSQRHEEPWRTGKLLLTLPLL   399
goat          EFVTLKALALANSDSMYIEDLEAVQKLQDLLHEALQDYELSQRHEEPRRTGKLLLTLPLL   420
whale         EFVTLKALALANSDSMYIEDLEAVQKLQDLLHEALQDYELSQRHEEPRRTGKLLLTLPLL   420
chicken       EFVTLKALALANSDSMHIEDMDAVQKLQDLLHEALQDYELSQRNEEPRRAGKLLLTLPLL   420
fish          EFVTLKAIALANSDSMHIEDMEAVQKLQDALHEALQDFECSQHQEDPRRAGKLLMTLPLL   399
              **: ***:********:**:::******* *******:* **::*:* *:****:*****

mouse         RQTAAKAVQHFYSVKLQGKVPMHKLFLEMLEAKV    433
human         RQTAAKAVQHFYSVKLQGKVPMHKLFLEMLEAKV    433
goat          RQTAAKAVQHFYSVKLQGKVPMHKLFLEMLEAKV    454
whale         RQTAAKAVQHFYSIKLQGKVPMHKLFLEMLEAKV    454
chicken       RQTAAKAVQHFYSIKLQGKVPMHKLFLEMLEAKV    454
fish          RQTATKAVQHFYSIKVQGKVPMHKLFLEMLEAKV    433
              ****:********:*:******************
```

3.5.3. *Disorder prediction*



3.5.4. *Coding DNA sequences, produced protein constructs*

Esrrb-h_aa1-102_C12A-C72A-C91A

Synthesized sequence:

Ccgcgggtgagaacctgtacttccagggcatgagcagcgaagatcgtcacctgggtagcagcgcgggcagctttattaaaaccgagccgagcag
cccgagcagcggtattgatgcgctgagccaccatagcccgagcggtagcagcgatgcgagcggtggcttcggtattgcgctgagcacccatgcg
aacggtctggatagcccgccgatgtttgcgggtgcgggcctgggtggcaacccggcgcgtaaaagctacgaagactgcaccagcggcatcatgg
aggatagcgcgattaaggcggaatatatgctgaacgcgattccgaaacgtctgtaaaagctt

Translates into:

AGENLYFQGMSSEDRHLGSSAGSFIKTEPSSPSSGIDALSHHSPSGSSDASGGFGIALSTHANGLDSPPMFAGAGLGGNPARKSYEDCTSGIME
DSAIKAEYMLNAIPKRL*

Expressed peptide:

GST-His-Tev-Esrrb(aa1-102)_C12A-C72A-C91A

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISM
LEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQ
IDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDGSTSGSGHHHHHHSAGENLYFQGMSSEDRHLGSSAGSFIKTEPSSPSSGIDALSHHSPSGSS
DASGGFGIALSTHANGLDSPPMFAGAGLGGNPARKSYEDCTSGIMEDSAIKAEYMLNAIPKRL

After Tev-cleavage *(leaving 1 Gly in N-ter)*:

GMSSEDRHLGSSAGSFIKTEPSSPSSGIDALSHHSPSGSSDASGGFGIALSTHANGLDSPPMFAGAGLGGNPARKSYEDCTSGIMEDSAIKAEY
MLNAIPKRL

```
          10          20          30          40          50          60
 G MSSEDRHLGS  SAGSFIKTEP  SSPSSGIDAL  SHHSPSGSSD  ASGGFGIALS  THANGLDSPP

         70          80          90         100
 MFAGAGLGGN  PARKSYEDCT  SGIMEDSAIK  AEYMLNAIPK  RL
```

Esrrb-h_aa1-102_C12A-C91A

Coding sequence (mutated from Esrrb-h_aa-102_C12A-C72A-C91A):

Ccgcgggtgagaacctgtacttccagggcatgagcagcgaagatcgtcacctgggtagcagcgcgggcagctttattaaaaccgagccgagcag
cccgagcagcggtattgatgcgctgagccaccatagcccgagcggtagcagcgatgcgagcggtggcttcggtattgcgctgagcacccatgcg
aacggtctggatagcccgccgatgtttgcgggtgcgggcctgggtggcaacccgtgccgtaaaagctacgaagactgcaccagcggcatcatgg
aggatagcgcgattaaggcggaatatatgctgaacgcgattccgaaacgtctgtaaaagctt

Translates into:

AGENLYFQGMSSEDRHLGSSAGSFIKTEPSSPSSGIDALSHHSPSGSSDASGGFGIALSTHANGLDSPPMFAGAGLGGNPCRKSYEDCTSGIME
DSAIKAEYMLNAIPKRL*

After Tev-cleavage *(leaving 1 Gly in N-ter)*:

GMSSEDRHLGSSAGSFIKTEPSSPSSGIDALSHHSPSGSSDASGGFGIALSTHANGLDSPPMFAGAGLGGNPCRKSYEDCTSGIMEDSAIKAEY
MLNAIPKRL

```
          10          20          30          40          50          60
 G MSSEDRHLGS  SAGSFIKTEP  SSPSSGIDAL  SHHSPSGSSD  ASGGFGIALS  THANGLDSPP

        70          80          90         100
 MFAGAGLGGN  PCRKSYEDCT  SGIMEDSAIK  AEYMLNAIPK  RL
```

Esrrb-h_aa1-102_C12A-C72A-C91A_AviTag-His6

Synthesized sequence:

Ccgcgggtgagaacctgtacttccagggcatgagcagcgaagatcgtcacctgggtagcagcgcgggcagctttattaaaaccgagccgagcag
cccgagcagcggtattgatgcgctgagccaccatagcccgagcggtagcagcgatgcgagcggtggcttcggtattgcgctgagcacccatgcg
aacggtctggatagcccgccgatgtttgcgggtgcgggcctgggtggcaacccggcgcgtaaaagctacgaagactgcaccagcggcatcatgg
aggatagcgcgattaaggcggaatatatgctgaacgcgattccgaaacgtctgtaaaagctt

Translates into:

MSSEDRHLGSSAGSFIKTEPSSPSSGIDALSHHSPSGSSDASGGFGIALSTHANGLDSPPMFAGAGLGGNPARKSYEDCTSGIMEDSAIKAEYM
LNAIPKRLGLNDIFEAQKIEWHEGAGLE

Expressed peptide:

MSSEDRHLGSSAGSFIKTEPSSPSSGIDALSHHSPSGSSDASGGFGIALSTHANGLDSPPMFAGAGLGGNPARKSYEDCTSGIMEDSAIKAEYM
LNAIPKRLGLNDIFEAQKIEWHEGAGLEHHHHHH*

Expressed peptide:

MSSEDRHLGSSAGSFIKTEPSSPSSGIDALSHHSPSGSSDASGGFGIALSTHANGLDSPPMFAGAGLGGNPARKSYEDCTSGIMEDSAIKAEYM
LNAIPKRLGLNDIFEAQKIEWHEGAGLEHHHHHH

```
          10          20          30          40          50          60
 MSSEDRHLGS  SAGSFIKTEP  SSPSSGIDAL  SHHSPSGSSD  ASGGFGIALS  THANGLDSPP

        70          80          90         100
 MFAGAGLGGN  PARKSYEDCT  SGIMEDSAIK  AEYMLNAIPK  RL  GLNDIFEAQKIEWHEGAGLEHHHHHH
```

Esrrb-h_aa1-102_C12A-C72A_AviTag-His6

Coding sequence (mutated from Esrrb-h_aa-102_C12A-C72A-C91A_AviTag-His6):

```
Catatgagcagcgaagaccgtcacctgggtagcagcgcgggtagctttattaagaccgaaccgagcagcccgagcagcggcattgatgcgctga
gccatcatagcccgagcggtagcagcgatgcgagcggtggcttcggtattgcgctgagcacccatgcgaacggtctggatagcccgccgatgtt
tgcgggtgcgggcctgggtggcaacccggcgcgtaagagctacgaggactgcaccagcggcatcatggaggatagcgcgattaagtgcgaatat
atgctgaacgcgattccgaaacgcctgggcctgaacgacatttttgaagcgcagaagattgagtggcatgagggtgcgggcctcgag
```

Translates into:

```
MSSEDRHLGSSAGSFIKTEPSSPSSGIDALSHHSPSGSSDASGGFGIALSTHANGLDSPPMFAGAGLGGNPARKSYEDCTSGIMEDSAIKCEYM
LNAIPKRLGLNDIFEAQKIEWHEGAGLE
```

Expressed peptide:

```
MSSEDRHLGSSAGSFIKTEPSSPSSGIDALSHHSPSGSSDASGGFGIALSTHANGLDSPPMFAGAGLGGNPARKSYEDCTSGIMEDSAIKCEYM
LNAIPKRLGLNDIFEAQKIEWHEGAGLEHHHHHH
```

```
        10          20          30          40          50          60
 MSSEDRHLGS  SAGSFIKTEP  SSPSSGIDAL  SHHSPSGSSD  ASGGFGIALS  THANGLDSPP

        70          80          90         100
 MFAGAGLGGN  PARKSYEDCT  SGIMEDSAIK  CEYMLNAIPK  RL  GLNDIFEAQKIEWHEGAGLEHHHHHH
```

Esrrb-h_aa1-102_C12A-C72A-C91A-S22A_AviTag-His6

Coding sequence (mutated from Esrrb-h_aa-102_C12A-C72A-C91A_AviTag-His6):

```
Catatgagcagcgaagaccgtcacctgggtagcagcgcgggtagctttattaagaccgaaccgagcgcgccgagcagcggcattgatgcgctga
gccatcatagcccgagcggtagcagcgatgcgagcggtggcttcggtattgcgctgagcacccatgcgaacggtctggatagcccgccgatgtt
tgcgggtgcgggcctgggtggcaacccggcgcgtaagagctacgaggactgcaccagcggcatcatggaggatagcgcgattaaggcggaatat
atgctgaacgcgattccgaaacgcctgggcctgaacgacatttttgaagcgcagaagattgagtggcatgagggtgcgggcctcgag
```

Translates into:

```
MSSEDRHLGSSAGSFIKTEPSAPSSGIDALSHHSPSGSSDASGGFGIALSTHANGLDSPPMFAGAGLGGNPARKSYEDCTSGIMEDSAIKAEYM
LNAIPKRLGLNDIFEAQKIEWHEGAGLE
```

Expressed peptide:

```
MSSEDRHLGSSAGSFIKTEPSAPSSGIDALSHHSPSGSSDASGGFGIALSTHANGLDSPPMFAGAGLGGNPARKSYEDCTSGIMEDSAIKAEYM
LNAIPKRLGLNDIFEAQKIEWHEGAGLEHHHHHH*
```

```
        10          20          30          40          50          60
 MSSEDRHLGS  SAGSFIKTEP  SAPSSGIDAL  SHHSPSGSSD  ASGGFGIALS  THANGLDSPP

        70          80          90         100
 MFAGAGLGGN  PARKSYEDCT  SGIMEDSAIK  AEYMLNAIPK  RL  GLNDIFEAQKIEWHEGAGLEHHHHHH
```

References

[1] W. Lee, M. Tonelli, J. L. Markley, *Bioinformatics*, 2015, **31**, 1325-1327.

[2] F.-X. Theillet, C. Smet-Nocca, S. Liokatis, R. Thongwichian, J. Kosten, M.-K. Yoon, R. W. Kriwacki, I. Landrieu, G. Lippens, P. Selenko, *J. Biomol. NMR*, 2012, **54**, 217-236.

[3] F.-X. Theillet, H. M. Rose, S. Liokatis, A. Binolfi, R. Thongwichian, M. Stuiver, P. Selenko, *Nat. Protoc.*, 2013, **8**, 1416-1432.

[4] A. Mylona, F.-X. Theillet, C. Foster, T. M. Cheng, F. Miralles, P. A. Bates, P. Selenko, R. Treisman, *Science*, 2016, **354**, 233-237.

[5] M. Julien, C. Bouguechtouli, A. Alik, R. Ghouil, S. Zinn-Justin, F.-X. Theillet, in *Intrinsically Disordered Proteins: Methods and Protocols* (B. B. Kragelund, K. Skriver, eds.), Springer US, New York, NY, 2020, 793-817.

[6] J.-P. Lambert, M. Tucholska, T. Pawson, A.-C. Gingras, *J. Proteomics*, 2014, **100**, 55-59.

[7] P. Poullet, S. Carpentier, E. Barillot, *Proteomics*, 2007, **7**, 2553-2556.

[8] M. The, M. J. MacCoss, W. S. Noble, L. Käll, *J. Am. Soc. Mass Spectrom*, 2016, **27**, 1719-1727.
[9] B. Valot, O. Langella, E. Nano, M. Zivy, *Proteomics*, 2011, **11**, 3572-3577.
[10] Y. Perez-Riverol, J. Bai, C. Bandla, D. García-Seisdedos, S. Hewapathirana, S. Kamatchinathan, D. J. Kundu, A. Prakash, A. Frericks-Zipper, M. Eisenacher, M. Walzer, S. Wang, A. Brazma, J. A. Vizcaíno, *Nucleic Acids Res.*, 2022, **50**, D543-D552.
[11] C. Smet-Nocca, H. Launay, J.-M. Wieruszeski, G. Lippens, I. Landrieu, *J. Biomol. NMR*, 2013, **55**, 323-337.
[12] K. Tamiola, F. A. A. Mulder, *Biochem. Soc. Trans.*, 2012, **40**, 1014-1020.
[13] J. T. Nielsen, F. A. A. Mulder, *J. Biomol. NMR*, 2018, **70**, 141-165.
[14] C. Camilloni, A. De Simone, W. F. Vranken, M. Vendruscolo, *Biochemistry*, 2012, **51**, 2224-2231.