

Article

In Silico Identification of SOX1 Post-Translational Modifications Highlights a Shared Protein Motif

Azaz Ahmad ¹, Stephanie Strohbuecker ¹, Claudia Scotti ² , Cristina Tufarelli ³  and Virginie Sottile ^{1,2,*} 

¹ School of Medicine, The University of Nottingham, Nottingham NG7 2RD, UK; azazkhan47@gmail.com (A.A.); StephanieStrohbuecker@web.de (S.S.)

² Department of Molecular Medicine, The University of Pavia, 27100 Pavia, Italy; claudia.scotti@unipv.it

³ Department of Genetics and Genome Biology/Leicester Cancer Research Centre, The University of Leicester, Leicester LE2 7LX, UK; cristina.tufarelli@leicester.ac.uk

* Correspondence: virginie.sottile@unipv.it

Received: 26 October 2020; Accepted: 10 November 2020; Published: 13 November 2020



Abstract: The transcription factor SOX1 is a key regulator of neural stem cell development, acting to keep neural stem cells (NSCs) in an undifferentiated state. Postnatal expression of Sox1 is typically confined to the central nervous system (CNS), however, its expression in non-neural tissues has recently been implicated in tumorigenesis. The mechanism through which SOX1 may exert its function is not fully understood, and studies have mainly focused on changes in SOX1 expression at a transcriptional level, while its post-translational regulation remains undetermined. To investigate this, data were extracted from different publicly available databases and analysed to search for putative SOX1 post-translational modifications (PTMs). Results were compared to PTMs associated with SOX2 in order to identify potentially key PTM motifs common to these SOXB1 proteins, and mapped on SOX1 domain structural models. This approach identified several putative acetylation, phosphorylation, glycosylation and sumoylation sites within known functional domains of SOX1. In particular, a novel SOXB1 motif (xKSExSxxP) was identified within the SOX1 protein, which was also found in other unrelated proteins, most of which were transcription factors. These results also highlighted potential phospho-sumoyl switches within this SOXB1 motif identified in SOX1, which could regulate its transcriptional activity. This analysis indicates different types of PTMs within SOX1, which may influence its regulatory role as a transcription factor, by bringing changes to its DNA binding capacities and its interactions with partner proteins. These results provide new research avenues for future investigations on the mechanisms regulating SOX1 activity, which could inform its roles in the contexts of neural stem cell development and cancer.

Keywords: SOX1; post-translational modification; in silico; transcription factors; regulatory motifs

1. Introduction

SOX1 is a transcription factor mainly known as an early neural stem cell marker, which has also been implicated in tumorigenesis outside of its neural context. Studies have demonstrated that the SOX family of genes can serve as either tumour suppressor genes or oncogenes in different cancers types [1,2]. Autoantibodies to SOX1 are common in small-cell lung carcinoma (SCLC) and serve as a serological tumour marker for SCLC, while the role of a SOX1-related autoimmune response in SCLC is still elusive [3,4]. SOX1 function in neural stem cell maintenance has been studied widely [5,6], but little is known about the role of SOX1 in cancer.

In addition to mechanisms regulating gene expression at a transcriptional and epigenetic level, molecular function can also be regulated by modulating the properties of the encoded protein through

different types of post-translational modifications (PTMs) [7]. There are several PTMs that can occur in a protein sequence depending on the identity of its amino acids. There is currently no information with regards to PTMs affecting SOX1 protein function in either neural stem cells or cancer. In this study, an in-silico analysis was performed [8] on the SOX1 protein sequence to identify PTMs using different bioinformatics tools and data sources, in order to provide new information hinting to potential mechanisms regulating SOX1 function in different cellular contexts.

2. Materials and Methods

PTM identification: Evidence of potential PTM residues in the human SOX1 protein was surveyed using the protein name human SOX1 (ACC#O00570) or SOX1 amino acid sequence in a FASTA format. The PhosphoSitePlus® database [9] was searched to identify existing experimentally verified data about SOX1 PTMs (<http://www.phosphosite.org/homeAction.do>). The NetPhos 3.1 server (<http://www.cbs.dtu.dk/services/NetPhos>) was used to predict phosphorylation sites at serine, threonine and tyrosine residues for SOX1 [10]. YinOYang 1.2 (<http://www.cbs.dtu.dk/services/YinOYang>) was used for prediction of YinOYang sites [11,12]. GSP-SUMOv2 (<http://sumosp.biocuckoo.org>) [13] and JASSAv4 (<http://www.jassa.fr>) [14] were used as sumoylation prediction tools.

Multiple sequence alignment (MSA) tools: Protein BLAST [15] was used to identify homologous sequences for the Human SOX1 protein in the refseq collection database (refseq_protein) [16], selecting Mouse (NP_033259), Chimpanzee (XP_016781055), Rabbit (XP_002724136), Whale (XP_004273645), Monkey (XP_017357713), Dog (XP_849239), and Chicken (NP_989664) sequences to perform multiple sequence alignment (MSA) [17] using Clustal-Omega from the EMBL-EBI resources [18].

Consensus motif search: ScanProsite [19] was used to scan for the SOXB1 consensus motif XKSExSxxP for matches against other proteins, selecting UniProtKB and searching against human proteins only, excluding the option of splice variants. The result obtained was matched against UniProt protein entries and was analysed by retrieving the Gene ontology (molecular) in the UniProt database [20]. InterPro (<https://www.ebi.ac.uk/interpro/search/sequence/>) and Eukaryotic Linear Motif (ELM, <http://elm.eu.org/>) databases were used to search against recognised protein motifs.

SNPs analysis within SOX1 coding locus: NCBI-dbSNP [21] and Uniprot feature searches [20] were used for SNP analysis.

IBS Illustrator for Biological Sequences: The IBS webserver was used for schematic illustration of SOX1 protein sequence figures [22].

Protein disorder and modelling: SOX1 disorder was evaluated by IUPred and ANCHOR2 [23]. Protein domain modelling was performed by Modeller [24] for SOX1 HMG, residues 49-120, using the structure of SOX2 HMG transcription factor as a template (PDB ID: 1o4x). I-TASSER [25] was instead used for modelling of SOXp (residues 129-194) and the C-terminal (residues 263-366) domains, based on its excellent performance in all recent CASP competitions [26,27]. The structure for intrinsically disordered regions (IDRs) was build using the same strategy as for structured protein [28], and the secondary structure prediction was based on the PSSpred algorithm, a combination of 7 neural network predictors from different profile data and parameters [29,30]. All models were geometry optimized by Phenix [31] and validated by ProSA Web [32], Verify 3D [33] and Rampage [34]. PTMs were modelled by Vienna-PTM 2.0 [35] and molecules visualised by Pymol [36].

3. Results

3.1. Survey of PTMs of the SOX1 Protein

Databases from PhosphoSitePlus, YinOYang 1.2 server and NetPhosK 3.1 server were surveyed for predicted PTMs. This analysis showed the presence of a variety of putative sites on SOX1, such as phosphorylation, sumoylation, acetylation and glycosylation (Figure 1) (Figures S1–S3). Predicted PTM sites were primarily found within functional SOX1 domains, such as the HMG-domain and C-terminal region [9] (Figure 1A). Both predicted and experimentally verified acetylation of lysine (K)

residues was found at HMG-BOX and SOXp domain of SOX1 (Figure 1A). Residues K83 and K132 in human SOX1 were experimentally verified to be acetylated, while K131 was predicted to be acetylated (Figure 1). K131 and K132 acetylation was also experimentally verified in mouse SOX1 (Figure 1A) [9].

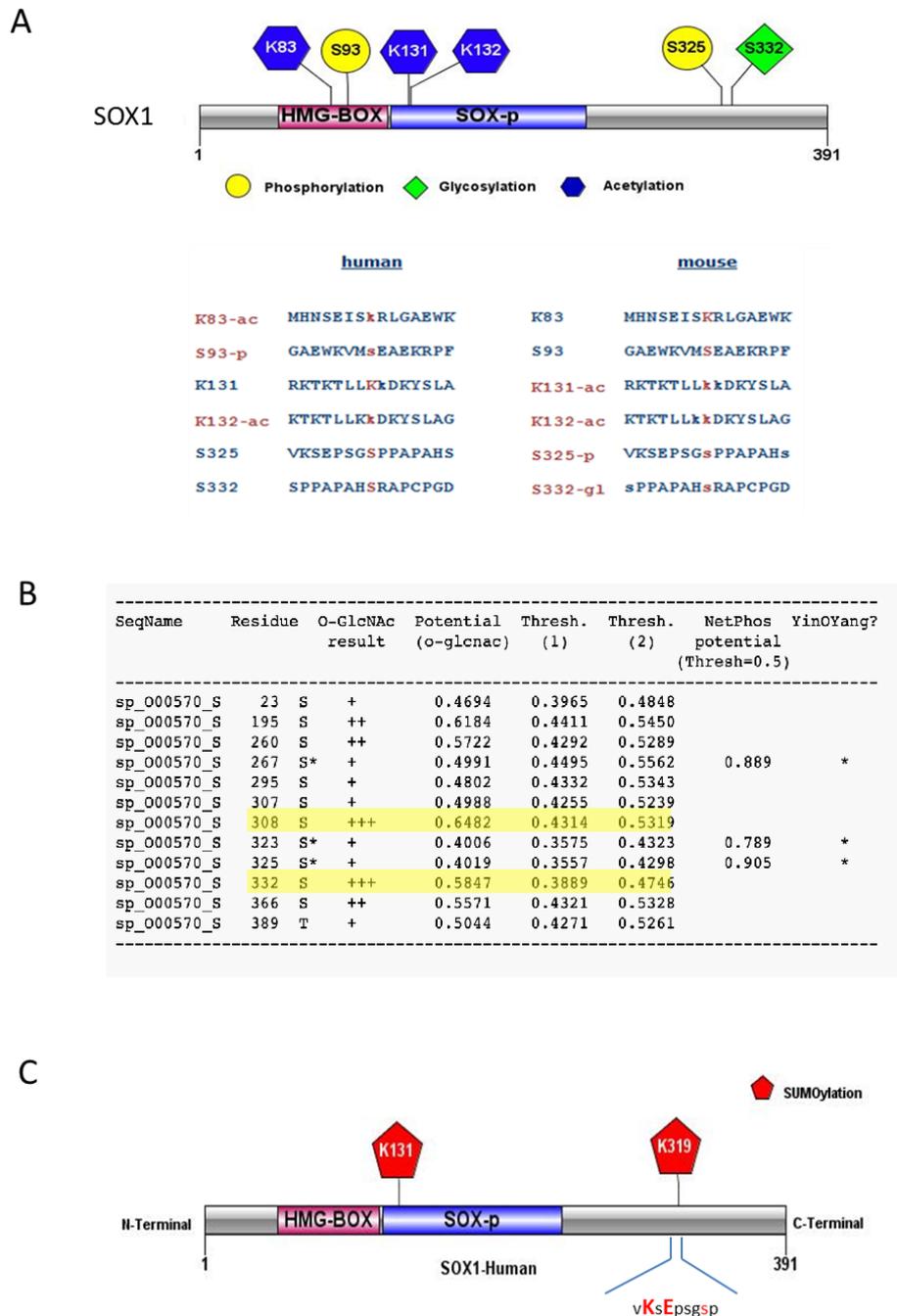


Figure 1. Post-translational modifications (PTMs) predicted for the SOX1 protein. (A) Schematic representation of different PTM types identified from the PhosphoSitePlus database with multiple sequence alignment for both human and mouse SOX1 protein showing validated (red) and predicted (blue) PTM residues. Phosphorylation (-p), Acetylation (-ac) and Glycosylation (-gl). (B) Glycosylation (O-GlcNAc), phosphorylation (-p) and predicted residues with YinOYang effect (asterisk) from the YinOYang 1.2 Server. (C) Sumoylated residues within SOX1 predicted by GSP-SUMOV2.0 and JASSAv4 analysis tools. Diagrams in (A) and (C) generated using the IBS illustrator webserver [22].

PhosphoSitePlus indicated phosphorylation sites at serine (S) residues at positions S93 and S325 (Figure 1A), that were also predicted by NetPhosK server 3.1, with phosphorylation potential scores of 0.829 and 0.904, respectively (Figure 1B; Figures S1–S3). Site S93 was the only experimentally verified phosphorylation site reported for the human SOX1 protein, found phosphorylated in tumour tissue affected by ischemia [37]. S93 is found within the HMG-Box of SOX1 (Figure 1A), a DNA binding domain whose phosphorylation might be crucial to SOX1 DNA binding and transcription factor activities. Residue S325 is present within the C-terminal region of SOX1, which is known for its transcriptional activation [38]. PhosphoSitePlus reported S325 as an experimentally verified phosphorylated site in mouse SOX1 [39,40], suggesting it might be phosphorylated in human SOX1 as well.

The survey of glycosylation sites within the SOX1 protein was performed using YinOYang 1.2 server tools [11,12], and indicated SOX1 residues S325, S332 and S308 (Figure 1A,B) were predicted to be glycosylated. S332 was also experimentally verified in mouse SOX1 by PhosphositePlus [39] (Figure 1A). A predicted Yin-Yang effect [9,39] was found at residue S325, which was also predicted to be phosphorylated (Figure 1B), while residues S308 and S332 showed high probability of O-GlcNAc but much less probable Yin-Yang effect (Figure 1B, highlighted in yellow).

Sumoylation analyses performed with GSP-SUMOv2.0 and JASSAv4 showed two potential sites for human SOX1 sumoylation, at residues K131 and K319 (Figure 1C). K319 is present within the C-terminal domain, and this domain is known for the transactivation ability of SOX1. Residue K131 was predicted sumoylated and also acetylated, and was found located close to the DNA binding domain of SOX1 (HMG-domain), where PTMs may exert a regulatory role for SOX1 DNA binding activity.

3.2. Phosphorylation-Dependent Sumoylation Sites Within SOX1

The canonical sumoylation consensus motif in a protein sequence is ψ KxE, where ' ψ ' can be any hydrophobic amino acid such as A, I, L, M, P, F, V or W, while ' x ' can be any amino acid [41]. In SOX1, residue K319 was found to lie within the sumoylation consensus motif ${}_{318}\psi$ KsE ${}_{321}$ (Figure 1C). The K319 sumoylation consensus motif is in close proximity to S325 (${}_{318}\psi$ KsEpsgSp ${}_{326}$), which is a predicted phosphorylation site (Figure 1A,C). Since studies have shown that phosphorylation may regulate sumoylation of a substrate [41], further analysis was carried out on this specific motif (${}_{318}\psi$ KSEPSGSP ${}_{326}$) containing sumoylation and phosphorylation sites, to investigate whether it could provide additional information about the regulation of SOX1 function.

3.3. Identification of a New SOXB1 Consensus Motif Within the SOX1 Protein

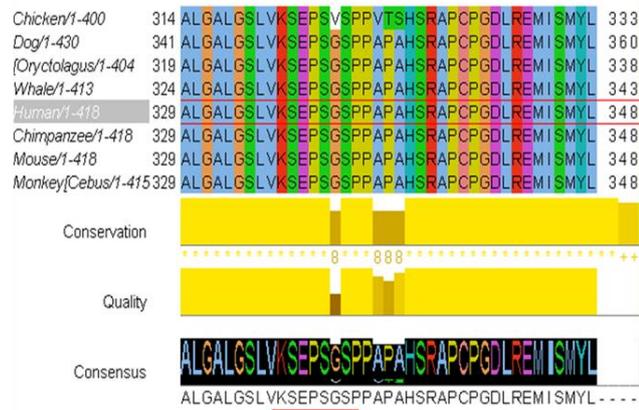
Evolutionary conserved regions generally point to key functional parts of a protein. Multiple Sequence Alignment (MSA) of SOX1 across different species revealed that the region ${}_{318}\psi$ KSEPSGSP ${}_{326}$ is highly conserved across different species (Figure 2A). The transcription factors SOX1, SOX2 and SOX3 belong to the SOXB1 sub family, sharing high sequence similarities and relatively similar expression patterns [6]. MSA investigation of the homology in the ${}_{318}\psi$ KSEPSGSP ${}_{326}$ region among transcription factors belonging to the SOXB1 sub-family showed that it is highly conserved across this sub-family (Figure 2B). Based on the level of conservation, the consensus sequence xKSEsxxP was identified as SOXB1 consensus motif (Figure 2C), where ' x ' can be any amino acid.

The consensus motif was not recognised to be associated with any specific protein family listed in the InterPro platform, while a search in the Eukaryotic Linear Motif (ELM) database did not yield a full homology but only returned partial matches to (i) a CK1 phosphorylation site ('SEPSGSP', with a probability score of $1.70e-02$), (ii) to a GAG attachment site ('(E)PSGS', with a probability of $1.92e-02$) and (iii) to a SUMO-1 modification site ('VKSE', with a probability score of $1.91e-03$) (Table S1).

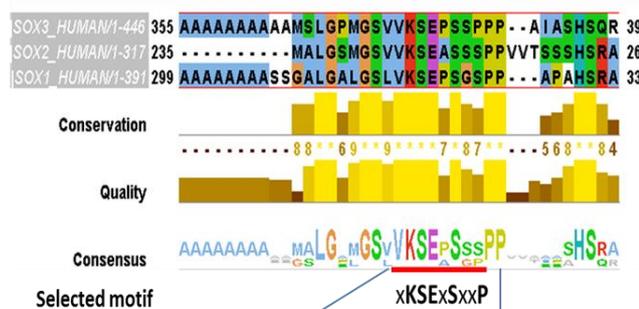
To further investigate whether the SOXB1 consensus motif is present in other unrelated proteins, a collection of human protein motifs from the ScanProsite database was surveyed. Interestingly, the SOXB1 consensus motif highlighted in Figure 2 was found to be shared by 32 unrelated proteins (Table S2). Gene ontology analysis of these proteins (Figure 3) showed that the majority are transcription

regulators acting as transcriptional activators/repressors with sequence-specific DNA binding, and are involved in developmental processes [42].

A Multiple sequence alignment of SOX1 protein across different species



B Multiple sequence alignment between Human SOX1, SOX2 and SOX3 protein



C

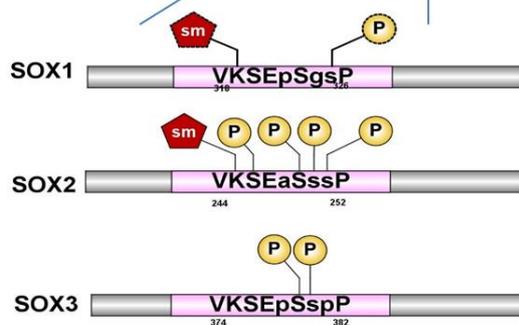


Figure 2. Multiple Sequence alignment of SOX1 protein. (A) Multiple Sequence alignment (MSA) for SOX1 protein focusing on the highly conserved motif $_{318}VKSEPSGSP_{326}$ (underlined in red). (B) MSA between members of the SOXB1 subgroup, focusing on the consensus motif named as ‘SOXB1 consensus motif’ used to further search protein databases. Figure generated using the EMBL-EBI resources, ClustalOmega alignment output page [17]. (C) Predicted PTMs within SOXB1 consensus motif. Conserved amino acids are shown in capital letters. P; Phosphorylation, sm; Sumoylation. IBS illustrator was used to produce the diagram [22].

The level of conservation assessed in a subset of proteins sharing this motif (Figure S4) showed conservation, supporting the hypothesis of a functional role.

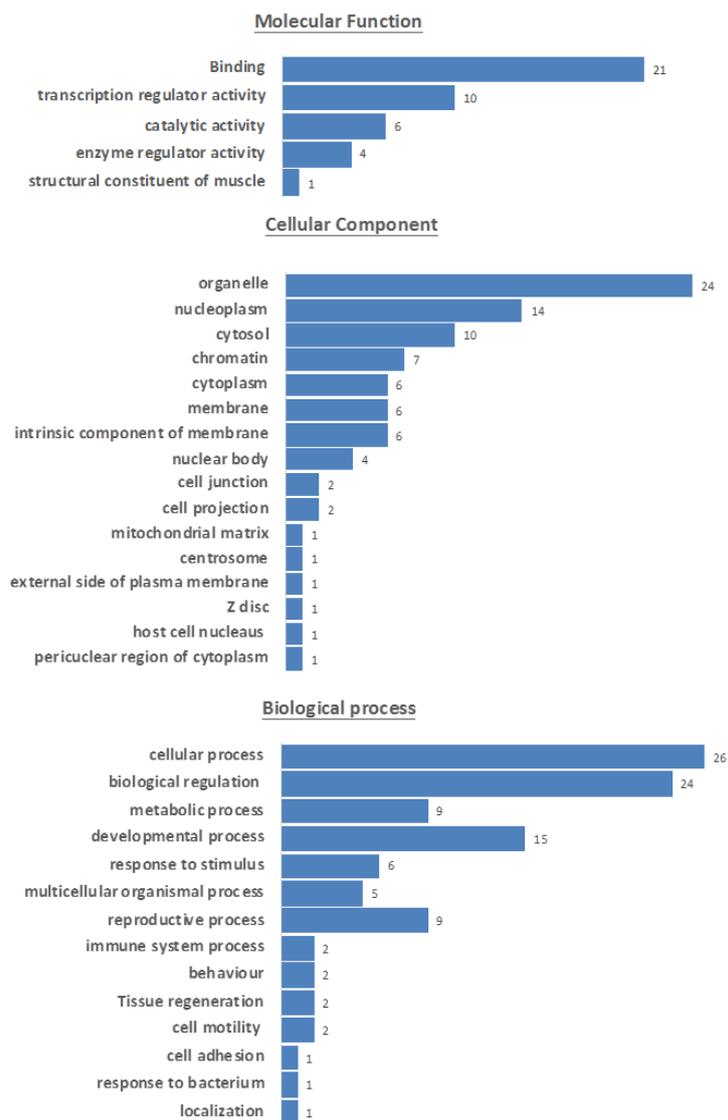


Figure 3. Gene ontology information of proteins sharing the SOXB1 motif highlighting involvement in common processes represented by blue horizontal bars, based on information (Table S3) retrieved from UniProt [20].

3.4. SOX1 Variants Triggering Changes to PTMs

An analysis was conducted to identify Single Nucleotide Polymorphisms (SNPs) within the SOX1 gene locus causing mutations in the SOX1 coding region. NCBI-dbSNP and Uniprot feature searches [21] highlighted many SNPs present at the SOX1 coding locus. SNPs triggering amino acid changes in the regions of interest containing HMG-domain, SOXp and C-terminal region of SOX1 protein were selected (Figure 4A). This analysis revealed that SNP rs1419694769 (A > T), which causes a missense mutation at K132 (K > E), leads to loss of acetylation at this crucial position (Table S4). Other relevant variants identified were rs1389486372, rs1178459411, rs1294049725, rs1161821962, rs1382937308 and rs1362549824. These SNPs are all linked to amino acid sequence changes within the SOXB1 consensus motif and in its close proximity, triggering loss/changes to the PTMs identified at this crucial motif. Moreover, they imply in three cases out of five the replacement of a Pro residue, which has relevant structural roles in loop rigidification and binding sites bracketing [43].

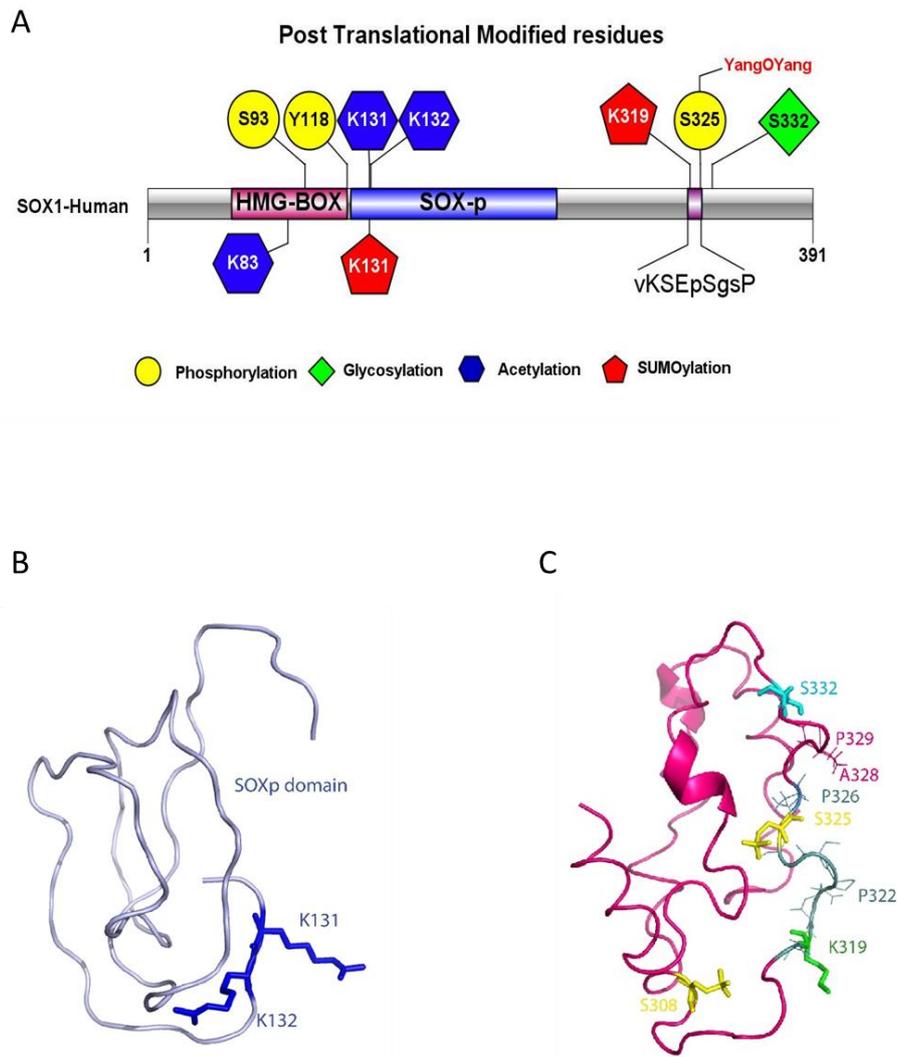


Figure 4. Models of SOX1 PTMs and structure. (A) SOX1 structure with predicted post translational modification sites and SOXB1 consensus motif sequence. The IBS illustrator was used to generate the diagram [22]. (B) SOXp domain model (residues 129–194). In blue: acetylated lysine residues K131 and K132. Residue K131 is also predicted to be sumoylated. (C) C-terminal domain model (residues 263–366). Sticks represent PTMs. In yellow: residues S308 and S325 are predicted to be phosphorylated. In green: residue K319 is predicted to be sumoylated. In light blue: residue S332 is predicted to be glycosylated. Thin lines show the residues forming the SOXB1 motif beyond S325 and V318. Labels are also provided for SNPs locations (P322, P326, A328, P329). Models in (B) and (C) were generated using I-TASSER [25], geometry optimized by Phenix [31] and validated by ProSA Web [32], Verify 3D [33] and Rampage [34]. PTMs were modelled by Vienna-PTM 2.0 [35] and molecules visualised by Pymol [36].

3.5. SOX1 Structural Modelling

Analysis of SOX1 sequence by IUPred and ANCHOR2 highlighted several regions with a high intrinsic disorder (Figure S5A). An attempt was made to model the full-length protein structure with I-TASSER, and the final structure evidenced three folded domains separated by disordered sequences: the HMG Box (residues 49–120), the SOXp (residues 129–194) and the C-terminal domains (residues 263–366), which turned out to cover all of the identified PTMs (Figure S5B). Further modelling work was then performed on each of these three individual domains. In particular, the model for the C-terminal domain is to be considered as a structurally plausible solution among the ensemble of structures potentially assumed by the IDR. In this respect, the validation of the model provided by three different servers (Table S5) confirmed its reliability. The α -helix predicted to include residues 288–290

has a low confidence score (<1), while the one including the REMIS sequence (residues 342–346) has confidence scores of eight or nine for each residue (Figure 4B).

The models of the three domains were then used to introduce phosphorylation and acetylation in the predicted positions by VIENNA-PTM 2.0, while, for other PTMs, residues involved were highlighted on the models and visualised by Pymol. The three-dimensional models of the SOXp domain and of the C-terminal domain (Figure 4 panels B and C, respectively) with the locations of the predicted PTMs and of the novel SOXB1 motif (identified by residues shown as sticks) show the PTMs all clustered together, especially in the C-terminal domain, where also several SNPs map, suggesting a close functional interaction of the relevant residues and modifications. SOX1 HMG Box is a well-characterized DNA binding domain, which, from the structural point of view, includes three α -helices, numbered from one to three, that are highly conserved throughout the SOX family of transcription factors [44]. PDB ID 6T7B describes the structure of human SOX2 transcription factor (HMG box domain) in complex with a nucleosome. Superposition of the highly homologous SOX1 HMG Box domain model (RMSD: 0.20 Å) showed that phosphorylation of residue S93 affects a position located in the loop comprised between the second and third α -helices, while acetylation affects K83, a residue located very close to the DNA molecule and potentially interacting with it (Figure 5). Interestingly, the only three amino acid differences which distinguish the SOX1 and SOX2 HMG Box sequences (positions 91, 92 and 95) [45] are the main components of the five-residue connection loop between α -helices two and three. This sequence tends to be specific to each SOX family member. Its location, pointing outside the nucleosome, suggests it may be involved in the recruitment of specific molecular partners.

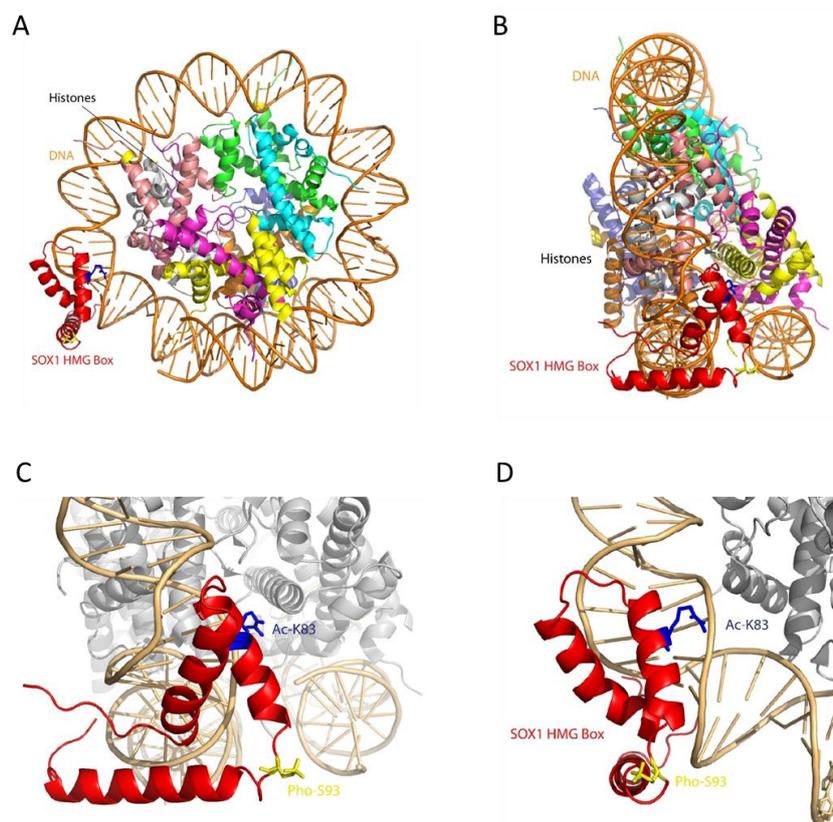


Figure 5. Structural model of SOX1 HMG Box domain (residues 49–120) generated by Modeller [24], using the structure of SOX2 HMG transcription factor as a template (PDB ID: 1o4x). (A) Model of SOX1 interaction with a nucleosome, based on SOX2 (PDB ID: 6T7B). (B) 90° rotation of (A). (C) Magnification of the SOX1 DNA binding domain (HMG box). (D) Rotation by 90 degrees of (C). In red: SOX1 HMG Box domain. In blue: acetylated K83. In yellow: phosphorylated S93.

4. Discussion

The PTM data uncovered in this study (summarised in Figure 4) strongly suggest that SOX1 might be regulated at the post-translational level, which could impact its function. Depending on its post-translational modifications, a transcription factor might change its DNA binding activities and interactions with partner proteins within transcription regulatory complexes, consequently altering its function [45,46]. This paradigm is well exemplified by SOX2, a SOX1-related transcription factor belonging to the same SOXB1 subfamily, and given SOX1 and SOX2 show functional redundancy [47], mechanisms regulating SOX2 provide useful information on the potential role of PTMs in SOX1 function. Liu et al. have shown that different PTMs confer SOX2 variable functions that regulate stem cell pluripotency and differentiation [48]. SOX2 sumoylation at K247 impairs its binding to the Fgf4 enhancer, which results in negative regulation of SOX2 transcriptional activities [49]. In mouse embryonic stem cells, interaction of SOX2 with the nuclear export machinery by acetylation at K75 within the HMG domain retains SOX2 inside the nucleus to regulate its target genes [50]. Consistent with this, the present study indicates that SOX1 is predicted to harbour sites of sumoylation at K319 and acetylation at K83.

The HMG domain of SOX1 is an evolutionarily conserved and functionally important region that binds to specific DNA sequences to bring conformational changes within chromatin structure [51]. The activity of the HMG Box domain is both to bind to and to bend DNA, and this happens thanks to residue M59, absolutely conserved throughout the SOX family of transcription factors, which indents within the DNA double helix [44]. The predicted post-translationally modified residues identified at K83 (acetylated) and S93 (phosphorylated) lie within the HMG domain of SOX1, making these modifications significant to SOX1 function. In fact, evaluation of the HMG box SOX-1 domain model in the structural context of the nucleosome suggests that these PTMs could interfere with its direct interaction with nucleosomal DNA (Ac-K83) or with other molecular partners (Pho-S93).

Another important SOX1 domain is SOXp, as binding site for the nestin neural enhancer [52] which is in close proximity to the HMG domain. It has been documented that binding of SOX2 to the nestin enhancer upregulates the expression of Nestin [52]. In SOX1, the acetylated and structurally adjacent residues K131 and K132 might similarly influence SOX1 binding to the nestin enhancer. K131 is also predicted as a sumoylation site, therefore interplay between acetylation and sumoylation is likely to occur at this site. The interplay between sumoylation and acetylation at K131 might be involved in regulating the transcriptional and DNA binding activity of SOX1, in line with evidence of cross-talk between sumoylation and acetylation regulating transcription and DNA binding activity of the tumour suppressor p53 [53].

Similarities between the PTMs predicted in SOX1 and those reported for SOX2 highlight the need to further investigate whether they may affect protein function in a comparable way. Interestingly, the SOXB1 motif identified in this study is not unique to the SOXB1 family of proteins, but is in fact present among other unrelated proteins, the majority of which are transcription factors involved in biological processes such as transcription regulation, DNA binding and transcription factor activity. The sequence structure of the SOXB1 motif identified in SOX1 fits with reports in the literature of sumoylation motifs being adjacent to sites of proline-directed phosphorylation [41]. Indeed, the SOXB1 motif identified in SOX1 includes the predicted phosphorylation site S325, which precedes proline residue P326, and includes the predicted sumoylated site K319. This suggests that the SOXB1 motif within SOX1 is a phosphorylation-dependent sumoylation motif. The majority of transcriptional regulators, such as heat-shock factors (HSFs) and transcription factors such as GATA-1 and myocyte enhancer factor 2 contain a phosphorylation-dependent sumoylation motif [41]. Phosphorylation-dependent sumoylation has been shown to repress transactivation capacities of HSF family proteins [41]. The SOXB1 motif identified here is conserved among SOX1 and SOX2. Since phospho-sumoyl switches have been reported at a similar conserved region of SOX2 [54] and sumoylation within this region was seen to affect SOX2 transcriptional activity [55], our findings

suggest that SOX1 the transcription factor function may be regulated by phosphorylation-dependent sumoylation at this specific motif.

From the structural point of view, SOX-1 includes several predicted disordered amino acid regions, especially at the C-terminal in the HMG domain. It is reasonable to assume that the full-length protein reaches the correct functional conformation upon specific partner binding and that PTMs can act either as activity or affinity modifiers. Selection of specific protein partners is therefore likely to play a major role in regulating and generating cell and developmental stage-specific functions for SOX transcription factors [56]. There is no information currently available regarding the details of the tertiary structure of SOX1, but the fact that PTM modification sites are clustered together, especially in the modelled C-terminal domain, suggests that alternative or simultaneous modifications can contribute to switching on and off functions even by a structural hindrance-mediated mechanism and/or by modifying inter-domain interactions.

Changes in the genomic coding sequence that can alter the amino acid sequence may change protein function either through direct structural changes or through alterations in PTMs. Here, evidence for single nucleotide polymorphisms (SNPs) within SOX1 protein was reviewed, in particular for the existence of mutations within the SOXB1 consensus motif, which might affect PTMs in this target region. An SNP causing a missense mutation (K132 > E132) was linked to loss of acetylation at K132, which could affect the DNA binding activity of SOX1 as a transcription factor [52] or its nuclear import [50]. Future mutational analyses experiments will be needed to better evaluate the role of these residues and their link to SOX1 regulation, both as a neural marker and for its role in cancer development. In addition, wider comparative genomic analyses between pathological and control samples could highlight point mutations in upstream sequences, including the promoter region, which could inform on critical regulatory regions.

SOX1 PTMs will need to be experimentally validated in order to clarify the mechanisms regulating SOX1 transcriptional function. For example, SOX1 plays a regulatory role during neuronal cell fate determination and differentiation as it binds to the Hes1 promoter, attenuating Notch signalling that suppresses neurogenesis [57]. The SOX1 HMG and C-terminal domains are both required for interaction with the Hes1 promoter [57]. Therefore, PTMs predicted within these domains could have a possible effect on SOX1 binding and activity at the Hes1 promoter. SOX1 is also known to bind to β -catenin to attenuate Wnt signalling, while dysregulated Wnt signalling has been reported in many cancer types [40,58]. Acetylation of SOX1 at K83, as raised by the present analysis, might be involved in retaining SOX1 inside the nucleus [59]. Further research will determine whether blocking K83 acetylation can modify SOX1 function in neurodevelopment and cancer.

5. Conclusions

In conclusion, our analysis has identified a set of putative PTMs within SOX1 suggested to regulate SOX1 function as a transcription factor. Previous studies have shown how the identification of functional regulatory regions through computational analysis has been validated experimentally. For example, putative HMG and POU motifs identified *in silico* were subsequently confirmed experimentally to exert a functional role in the regulation of DPPA4, a gene that plays an important role in ES cell self-renewal [60]. *In-silico* analysis can thus help address biological questions more efficiently by identifying new potential regulatory regions. Results presented here will inform future experimental investigations on the differential regulation of SOX1 transcriptional function in neural stem cell fate and in cancer.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4409/9/11/2471/s1>, Figure S1: SOX1 PTM result from the analysis with the NetPhosK 3.1 server. Figure S2: Graphical output from the SOX1 PTM analysis with NetPhosK 3.1. Figure S3: Graphical output from the SOX1 PTM analysis with PhosphoPlus. Figure S4: Multiple sequence alignment (MSA) of SOXB1 consensus motif. MSA of SOXB1 consensus motif among different organisms, performed for 5 randomly selected proteins from the list in Table S2 (#5, 7, 8, 13, 23, respectively) analysed with Clustal Omega alignment (EMBL-EBI resources [17]) showing SOXB1 consensus motif region (yellow), positions which have a single, fully conserved residue (“*”), conservation between

groups of strongly similar properties (‘.’), and conservation between groups of weakly similar properties (‘.’). Figure S5: Analysis of SOX1 sequence. (A) Analysis with IUPred and ANCHOR2 highlighted several regions with a high intrinsic disorder. (B) I-TASSER model of full-length SOX1 produced with I-TASSER. Table S1: Results of an Eukaryotic Linear Motif (ELM) database search for the consensus motif ‘VKSEPSGSP’. Table S2: List of transcription factors sharing the SOXB1 consensus motif (xKSExSxxP) identified using the ScanProsite database. Table S3: Gene ontology information from UniProt for the proteins identified. Table S4: SOX1 SNPs collected from NCBI-dbSNP and Uniprot. Table S5: Statistical report for the 3 validation servers for the 3 modelled domains.

Author Contributions: Methodology, A.A. and S.S.; Data acquisition and analysis, A.A. and C.S.; Manuscript writing, review and editing, A.A., C.T., C.S., and V.S.; Project supervision, C.T. and V.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Alzheimer’s Society (SS) and the University of Nottingham Vice Chancellor’s scholarship for research excellence (AA). A kind donation from R. Jones in the memory of A. Stacey is gratefully acknowledged. This work was supported by a Wellcome Trust Research Career Re-entry Fellowship to CT [WT210911/Z/18/Z]. VS and CS are supported by a grant from the Italian Ministry of Education, University and Research (MIUR) to the Department of Molecular Medicine of the University of Pavia under the initiative ‘Dipartimenti di Eccellenza (2018–2022)’.

Acknowledgments: The authors are grateful to C. Talbot (University of Leicester) for useful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kormish, J.D.; Sinner, D.; Zorn, A.M. Interactions between SOX factors and Wnt/ β -catenin signaling in development and disease. *Dev. Dyn.* **2010**, *239*, 56–68. [[CrossRef](#)] [[PubMed](#)]
2. Song, L.; Liu, D.; He, J.; Wang, X.; Dai, Z.; Zhao, Y.; Kang, H.; Wang, B. SOX1 inhibits breast cancer cell growth and invasion through suppressing the Wnt/ β -catenin signaling pathway. *APMIS* **2016**, *124*, 547–555. [[CrossRef](#)] [[PubMed](#)]
3. Kazarian, M.; Laird-Offringa, I.A. Small-cell lung cancer-associated autoantibodies: Potential applications to cancer diagnosis, early detection, and therapy. *Mol. Cancer* **2011**, *10*, 33. [[CrossRef](#)] [[PubMed](#)]
4. Lipka, A.F.; Verschuuren, J.J.G.M.; Titulaer, M.J. SOX1 antibodies in Lambert-Eaton myasthenic syndrome and screening for small cell lung carcinoma. *Ann. N. Y. Acad. Sci.* **2012**, *1275*, 70–77. [[CrossRef](#)]
5. Feng, N.; Han, Q.; Li, J.; Wang, S.; Li, H.; Yao, X.; Zhao, R.C. Generation of highly purified neural stem cells from human adipose-derived mesenchymal stem cells by Sox1 activation. *Stem Cells Dev.* **2014**, *23*, 515–529. [[CrossRef](#)]
6. Archer, T.C.; Jin, J.; Casey, E.S. Interaction of Sox1, Sox2, Sox3 and Oct4 during primary neurogenesis. *Dev. Biol.* **2011**, *350*, 429–440. [[CrossRef](#)]
7. Wang, Y.C.; Peterson, S.E.; Loring, J.F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* **2014**, *24*, 143–160. [[CrossRef](#)]
8. Blom, N.; Sicheritz-Pontén, T.; Gupta, R.; Gammeltoft, S.; Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **2004**, *4*, 1633–1649. [[CrossRef](#)]
9. Hornbeck, P.V.; Zhang, B.; Murray, B.; Kornhauser, J.M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* **2015**, *43*, D512–D520. [[CrossRef](#)]
10. Blom, N.; Gammeltoft, S.; Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **1999**, *294*, 1351–1362. [[CrossRef](#)]
11. Gupta, R. Prediction of Glycosylation Sites in Proteomes: From Post-Translational Modifications to Protein Function. Technical University of Denmark (DTU). 2001. Available online: [http://orbit.dtu.dk/en/publications/prediction-of-glycosylation-sites-in-proteomes-from-posttranslational-modifications-to-protein-function\(38d496a0-8ff0-4fb7-8e74-a84bb78eb707\).html](http://orbit.dtu.dk/en/publications/prediction-of-glycosylation-sites-in-proteomes-from-posttranslational-modifications-to-protein-function(38d496a0-8ff0-4fb7-8e74-a84bb78eb707).html) (accessed on 21 July 2020).
12. Gupta, R.; Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* **2002**, *7*, 310–322.
13. Zhao, Q.; Xie, Y.; Zheng, Y.; Jiang, S.; Liu, W.; Mu, W.; Liu, Z.; Zhao, Y.; Xue, Y.; Ren, J. GPS-SUMO: A tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids. Res.* **2014**, *42*, W325–W330. [[CrossRef](#)] [[PubMed](#)]
14. Beauclair, G.; Bridier-Nahmias, A.; Zagury, J.F.; Saïb, A.; Zamborlini, A. JASSA: A comprehensive tool for prediction of SUMOylation sites and SIMs. *Bioinformatics* **2015**, *31*, 3483–3491. [[CrossRef](#)] [[PubMed](#)]

15. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
16. O’Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Bhanu, R.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [[CrossRef](#)] [[PubMed](#)]
17. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.N.; Potter, S.C.; Finn, R.D.; et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641. [[CrossRef](#)]
18. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [[CrossRef](#)]
19. de Castro, E.; Sigrist, C.J.A.; Gattiker, A.; Bulliard, V.; Langendijk-Genevaux, P.S.; Gasteiger, E.; Bairoch, A.; Hulo, N. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **2006**, *34*, 362–365. [[CrossRef](#)]
20. UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [[CrossRef](#)]
21. Sherry, S.T.; Ward, M.H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E.M.; Sirotkin, K. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* **2001**, *29*, 308–311. [[CrossRef](#)]
22. Liu, W.; Xie, Y.; Ma, J.; Luo, X.; Nie, P.; Zuo, X.; Lahrmann, U.; Zhao, Q.; Zheng, Y.; Zhao, Y.; et al. IBS: An illustrator for the presentation and visualization of biological sequences. *Bioinformatics* **2015**, *31*, 3359–3361. [[CrossRef](#)] [[PubMed](#)]
23. Mészáros, B.; Erdos, G.; Dosztányi, Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **2018**, *46*, W329–W337. [[CrossRef](#)] [[PubMed](#)]
24. Sali, A.; Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815. [[CrossRef](#)]
25. Yang, J.; Zhang, Y. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43*, W174–W181. [[CrossRef](#)] [[PubMed](#)]
26. Moulton, J.; Pedersen, J.T.; Judson, R.; Fidelis, K. A large-scale experiment to assess protein structure prediction methods. *Proteins* **1995**, *23*, ii–v. [[CrossRef](#)] [[PubMed](#)]
27. Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R.W.H.; Blackledge, M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17002–17007. [[CrossRef](#)] [[PubMed](#)]
28. Yang, J.; Zhang, Y. Protein Structure and Function Prediction Using I-TASSER. *Curr. Protoc. Bioinform.* **2015**, *52*, 5.8.1–5.8.15. [[CrossRef](#)] [[PubMed](#)]
29. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7–8. [[CrossRef](#)]
30. Yan, R.; Xu, D.; Yang, J.; Walker, S.; Zhang, Y. A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. *Sci. Rep.* **2013**, *3*, 2619. [[CrossRef](#)]
31. Adams, P.D.; Grosse-Kunstleve, R.W.; Hung, L.W.; Ioerger, T.R.; McCoy, A.J.; Moriarty, N.W.; Read, R.J.; Sacchettini, J.C.; Sauter, N.K.; Terwilliger, T.C. PHENIX: Building new software for automated crystallographic structure determination. *Acta Cryst. D. Biol. Crystallogr.* **2002**, *58*, 1948–1954. [[CrossRef](#)]
32. Wiederstein, M.; Sippl, M.J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **2007**, *35*, W407–W410. [[CrossRef](#)] [[PubMed](#)]
33. Bowie, J.U.; Lüthy, R.; Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **1991**, *253*, 164–170. [[CrossRef](#)] [[PubMed](#)]
34. Lovell, S.C.; Davis, I.W.; Arendall, W.B., 3rd; de Bakker, P.I.W.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by Calpha geometry: Phi, psi and Cbeta deviation. *Proteins* **2003**, *50*, 437–450. [[CrossRef](#)] [[PubMed](#)]
35. Margreitter, C.; Petrov, D.; Zagrovic, B. Vienna-PTM web server: A toolkit for MD simulations of protein post-translational modifications. *Nucleic Acids Res.* **2013**, *41*, W422–W426. [[CrossRef](#)]
36. DeLano, W.L. Pymol: An open-source molecular graphics tool. *Ccp4 Newsl. Protein Crystallogr.* **2002**, *40*, 82–92.

37. Mertins, P.; Yang, F.; Liu, T.; Mani, D.R.; Petyuk, V.A.; Gillette, M.A.; Moore, R.J.; Levine, D.A.; Townsend, R.; Erdmann-Gilmore, P.; et al. Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels. *Mol. Cell Proteom.* **2014**, *13*, 1690–1704. [[CrossRef](#)]
38. Kamachi, Y.; Cheah, K.S.; Kondoh, H. Mechanism of regulatory target selection by the SOX high-mobility-group domain proteins as revealed by comparison of SOX1/2/3 and SOX9. *Mol. Cell Biol.* **1999**, *19*, 107–120. [[CrossRef](#)]
39. Trinidad, J.C.; Barkan, D.T.; Gullledge, B.F.; Thalhhammer, A.; Sali, A.; Schoepfer, R.; Burlingame, A.L. Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. *Mol. Cell Proteom.* **2012**, *11*, 215–229. [[CrossRef](#)]
40. Anastas, J.N.; Moon, R.T. WNT signalling pathways as therapeutic targets in cancer. *Nat. Rev. Cancer* **2013**, *13*, 11–26. [[CrossRef](#)]
41. Hietakangas, V.; Anckar, J.; Blomster, H.A.; Fujimoto, M.; Palvimo, J.J.; Nakai, A.; Sistonen, L. PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 45–50. [[CrossRef](#)]
42. Huntley, R.P.; Sawford, T.; Mutowo-Meullenet, P.; Shypitsyna, A.; Bonilla, C.; Martin, M.J.; O'Donovan, C. The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Res.* **2015**, *43*, D1057–D1063. [[CrossRef](#)] [[PubMed](#)]
43. Kini, R.M.; Evans, H.J. A hypothetical structural role for proline residues in the flanking segments of protein-protein interaction sites. *Biochem. Biophys. Res. Commun.* **1995**, *212*, 1115–1124. [[CrossRef](#)] [[PubMed](#)]
44. Dodonova, S.O.; Zhu, F.; Dienemann, C.; Taipale, J.; Cramer, P. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature* **2020**, *580*, 669–672. [[CrossRef](#)] [[PubMed](#)]
45. Duan, G.; Walther, D. The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput. Biol.* **2015**, *11*, e1004049. [[CrossRef](#)]
46. Hill, K. Post-translational modifications of hormone-responsive transcription factors: The next level of regulation. *J. Exp. Bot.* **2015**, *66*, 4933–4945. [[CrossRef](#)]
47. Ekonomou, A.; Kazanis, I.; Malas, S.; Wood, H.; Alifragis, P.; Denaxa, M.; Karagogeos, D.; Constanti, A.; Lovell-Badge, R.; Episkopou, V. Neuronal Migration and Ventral Subtype Identity in the Telencephalon Depend on SOX1. *PLoS Biol.* **2005**, *3*, e186. [[CrossRef](#)]
48. Liu, K.; Lin, B.; Zhao, M.; Yang, X.; Chen, M.; Gao, A.; Liu, F.; Que, J.; Lan, X. The multiple roles for Sox2 in stem cell maintenance and tumorigenesis. *Cell Signal.* **2013**, *25*, 1264–1271. [[CrossRef](#)]
49. Tsuruzoe, S.; Ishihara, K.; Uchimura, Y.; Watanabe, S.; Sekita, Y.; Aoto, T.; Saitoh, H.; Yuasa, Y.; Niwa, H.; Kawasuji, M.; et al. Inhibition of DNA binding of Sox2 by the SUMO conjugation. *Biochem. Biophys. Res. Commun.* **2006**, *351*, 920–926. [[CrossRef](#)]
50. Baltus, G.A.; Kowalski, M.P.; Zhai, H.; Tutter, A.V.; Quinn, D.; Wall, D.; Kadam, S. Acetylation of Sox2 Induces its Nuclear Export in Embryonic Stem Cells. *Stem Cells* **2009**, *27*, 2175–2184. [[CrossRef](#)]
51. Thomas, J.O. HMG1 and 2: Architectural DNA-binding proteins. *Biochem. Soc. Trans.* **2001**, *29*, 395–401. [[CrossRef](#)]
52. Tanaka, S.; Kamachi, Y.; Tanouchi, A.; Hamada, H.; Jing, N.; Kondoh, H. Interplay of SOX and POU Factors in Regulation of the Nestin Gene in Neural Primordial Cells. *Mol. Cell Biol.* **2004**, *24*, 8834–8846. [[CrossRef](#)] [[PubMed](#)]
53. Wu, S.Y.; Chiang, C.M. Crosstalk between sumoylation and acetylation regulates p53-dependent chromatin transcription and DNA binding. *EMBO J.* **2009**, *28*, 1246–1259. [[CrossRef](#)] [[PubMed](#)]
54. Yang, X.J.; Grégoire, S. A Recurrent Phospho-Sumoyl Switch in Transcriptional Repression and Beyond. *Mol. Cell.* **2006**, *23*, 779–786. [[CrossRef](#)] [[PubMed](#)]
55. Tahmasebi, S.; Ghorbani, M.; Savage, P.; Yan, K.; Gocevski, G.; Xiao, L.; You, L.; Yang, X.J. Sumoylation of Krüppel-like Factor 4 Inhibits Pluripotency Induction but Promotes Adipocyte Differentiation. *J. Biol. Chem.* **2013**, *288*, 12791–12804. [[CrossRef](#)]
56. Chew, L.J.; Gallo, V. The Yin and Yang of Sox proteins: Activation and repression in development and disease. *J. Neurosci. Res.* **2009**, *87*, 3277–3287. [[CrossRef](#)]
57. Kan, L.; Israsena, N.; Zhang, Z.; Hu, M.; Zhao, L.R.; Jalali, A.; Sahni, V.; Kessler, J.A. Sox1 acts through multiple independent pathways to promote neurogenesis. *Dev. Biol.* **2004**, *269*, 580–594. [[CrossRef](#)]
58. Clevers, H. Wnt/ β -Catenin Signaling in Development and Disease. *Cell* **2006**, *127*, 469–480. [[CrossRef](#)]

59. Filtz, T.M.; Vogel, W.K.; Leid, M. Regulation of transcription factor activity by interconnected post-translational modifications. *Trends Pharm. Sci.* **2014**, *5*, 76–85. [[CrossRef](#)]
60. Chakravarthy, H.; Boer, B.; Desler, M.; Mallanna, S.K.; McKeithan, T.W.; Rizzino, A. Identification of DPPA4 and other genes as putative Sox2:Oct-3/4 target genes using a combination of in silico analysis and transcription-based assays. *J. Cell Physiol.* **2008**, *216*, 651–662. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).