# Studies on the function of SRY with an A-to-S substitution in the HMG domain in the Okinawa spiny rat

## A-to-S アミノ酸置換が生じたオキナワトゲネズミ SRY タンパク質 HMG ドメインの機能解析

**A DISSERTATION**

**Submitted to the Graduate School of Life Science,**

**Hokkaido University**

**In partial fulfilment of the requirements for the degree**

**DOCTOR OF LIFE SCIENCE**

**URUNANONT Puntakarn**

**September 2025**

# Contents

# Abbreviations

| | |
|---|---|
| A | Alanine |
| AMH | anti-Müllerian hormone |
| A-to-S | Alanine to Serine |
| Avg | Average |
| BSA | Bovine serum albumin |
| C | Cysteine |
| D | Aspartic Acid |
| DiG | Digoxigenin |
| DNA | Deoxyribonucleic acid |
| DSD | disorders of sex development |
| E | Glutamic Acid |
| *E. coli* | *Escherichia coli* |
| EDTA | Ethylenediaminetetraacetic acid |
| EMSA | Electrophoretic Mobility Shift Assay |
| Enh | Enhancer |
| F | Phenylalanine |
| FA | Fluorescent Anisotropy |
| FAM | Fluorescein amidites |
| Fmol | Femtomole |
| G | Glycine |
| H | Histidine |
| H-bond | Hydrogen bond |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HMG | high mobility group |
| HPLC | High-performance liquid chromatography |
| I | Isoleucine |
| IPTG | isopropyl β-D-thiogalactopyranoside |
| IUCN | International Union for Conservation of Nature |
| K | Lysine |
| Kb | Kilobase |
| KCI | Potassium chloride |
| $K_d$ | dissociation constants |
| kDa | Kilodalton |
| L | Leucine |
| M | Methionine |
| ml | Millilitre |
| mM | Milli Molar |
| MgCl2 | Magnesium chloride |
| µM | Micro Molar |
| MnCl2 | Manganese (II) chloride |
| N | Asparagine |
| NaCl | Sodium Chloride |

| | |
|---|---|
| NaOH | Sodium hydroxide |
| NCBI | National Center for Biotechnology Information |
| OD | optical density |
| P | Proline |
| PAGE | Polyacrylamide Gel Electrophoresis |
| PAR | pseudo-autosomal region |
| PCR | Polymerase Chain Reaction |
| PDB | Protein Data Bank |
| pH | potential of hydrogen |
| PurR | purine repressor protein |
| Q | Glutamine |
| R | arginine |
| RMSD | Root Mean Square Deviation |
| S | Serine |
| SD | standard deviation |
| SDS-PAGE | sodium dodecyl sulfate–polyacrylamide gel electrophoresis |
| *SOX9* | SRY-Box Transcription Factor 9 (gene) |
| *Sry/SRY* | sex-determining region Y (gene) |
| SRY | sex-determining region Y (protein) |
| T | Threonine |
| TBP | TATA box-binding protein TBP |
| TF | trigger factor |
| TG | transgenic |
| TMU | *Tokudaia muenninki* |
| Tris-HCI | Tris (hydroxymethyl) aminomethane (THAM) hydrochloride |
| V | Valine |
| W | Tryptophan |
| WT | Wide Type |
| Y | Tyrosine |

# General Introduction

Therian mammals (including marsupials and placental mammals) possess heteromorphic sex chromosomes, the X and Y chromosomes, which emerged approximately 181 to 175 million years ago (Cortez et al., 2014). Males are generally categorized by XY karyotyped configuration, whereas females show a representation of XX with one X inactivated to maintain the dosage compensation (Robert Finestra and Gribnau, 2017). This means at least one X is considered to be necessary for survive in both sexes. The X and Y chromosomes are believed to have originated from the same pair of autosomes, as the ancestral Y was once homologous to the ancestral X (Charlesworth, 2000). Over time, the progressive degradation of the Y chromosome led to a significant loss of homology between the two, resulting in major differences in their size and the number of genes they carry. For example, in humans, the Y chromosome is about three times smaller than the X chromosome (Wallis et al., 2008) and retains firstly discovered only around 45 protein-coding genes (Graves, 2004), or up to 106 coding genes reported by the recent Telomere-to-Telomere (T2T) consortium (Rhie et al., 2023) whereas the X chromosome contains up to 1,300 genes (Ross et al., 2005). Despite numerous proposed theories, the mechanisms driving Y chromosome degeneration remain unclear (Charlesworth, 2020). One of the most well-known explanations suggests that the suppression of recombination between the X and Y chromosomes caused the Y chromosome to lose its homologous partner, weakening natural selection against deleterious mutations (Muller, 1950). This suppression of recombination was primarily driven by inversions in the pseudo-autosomal region

(PAR), which occurred multiple times throughout evolution. As a result, the X-Y recombining region progressively diminished, accelerating the degeneration of the Y chromosome (Lahn and Page, 1999).

To determine the sex of XY-possessing species, the process is divided into two stages called primary and secondary sex determination. While secondary sex determination focuses on the development of sexual phenotype characteristic regulated by hormones secreted from those differentiated gonads, the primary sex determination is contingent on the differential development of bipotential embryonic gonads into the testis or ovary. With that said, the promotion of bipotential gonads, testis differentiation is triggered by the activation of a sex-specific gene called sex-determining region Y (*Sry/SRY*), a transcription factor which is located on the Y chromosome (Berta et al., 1990). The novel mechanisms of *Sry/SRY* are to encode SRY proteins, which bind to its specific target upstream of SRY-Box Transcription Factor 9 (*SOX9*) via the high mobility group (HMG) domain and promote testis differentiation (Kashimada and Koopman, 2010). In human, mutations or deletions of *SRY* could results in such dramatic effects as its potential to cause disorders of sex development (DSD) where an individual's chromosomal sex (XY) might not align with their phenotypic sex.; for example, 20% of patients with Swyer syndrome (pure gonadal dysgenesis) harbor an *SRY* mutation and/or deletion (Culha et al., 2012). Most SRY mutations in DSD patients are found in the high-mobility group (HMG) domain, which is essential for DNA-binding (Graves et al., 1999). Recent studies have identified single amino acid substitutions in this domain, such as N65H (Assumpção et al., 2002), R76S (Ambulkar

et al., 2021), and A66T (Wang et al., 2018), which result in a female phenotype in genetically male individuals (46, XY). Some of these mutations have been shown to disrupt HMG-DNA binding activity, impairing testis development (Werner et al., 1995). All these findings highlight the crucial role of the HMG domain in male sex determination and emphasize how even a single mutation can cause a complete shift in sexual differentiation.

The importance of the Y chromosome and its fate-to-be degeneration has sparked the worrisome concern regarding the future of Y chromosome evolution, which is also lying ahead for humanity. However, despite the widespread degeneration of the Y chromosome in many species, certain exceptional mammals have managed to preserve their sex determination mechanisms by developing alternative systems independent of the Y chromosome. These systems involve the evolution of new sex-determining genes or chromosomal rearrangements that allow sex determination to function without reliance on the Y chromosome. Such adaptations highlight the dynamic nature of sex chromosome evolution and the potential for species to develop novel mechanisms to maintain reproductive viability. Among those exceptions, some native Japanese murine rodent species belonging to the genus *Tokudaia* exhibit unique characteristics in their sex chromosomes and sex determination mechanisms. The genus consists of three species that inhabit small regions of the Ryukyu Islands, Japan. All species which are now classified as endangered (or critically endangered), being protected by the Japanese government as natural monuments (Yamada et al., 2011) due to threats such as deforestation, logging, predation by domestic feral cats and dogs, and competition

with the common black rat.

Two of the three species, the Tokunoshima spiny rat (*Tokudaia tokunoshimensis*) and the Amami spiny rat (*Tokudaia osimensis*), have already lost their Y chromosome (Honda et al., 1977). Males and females possess only a single X chromosome (XO/XO), resulting in an odd number of chromosomes which *T. tokunoshimensis* has 2n = 45, while *T. osimensis* has 2n = 25 (Honda et al., 1977; Kobayashi et al., 2007). Moreover, the SRY gene is also completely lost in these species, suggesting the evolution of a new sex-determining mechanism independent of SRY (Kuroiwa et al., 2010; Kimura et al., 2014; Otake & Kuroiwa, 2016). The recent study shed some light for new sex determination key factor in *T. osimensis* as a 17-kb duplication located 430 kb upstream of Sox9 on an autosome, which might play some important role as substitution mechanisms in replacement of SRY-dependent system (Terao et al., 2023).

On the other hand, the third species, the Okinawa spiny rat *Tokudaia muenninki* (TMU), is one only *Tokudaia* species that retains an XX/XY sex chromosome system (Tsuchiya et al., 1989). This species was once believed to be extinct in the wild, as there had been no recorded sightings in its natural habitat for 30 years until a habitat investigation in September 2008 (Yamada et al., 2011). A research team successfully rediscovered the species, capturing 24 individual animals. While all animals were released at the capture sites, small tissue samples were collected from the tips of their tails for fibroblast cell culture (Yamada et al., 2010; Murata et al., 2011). Cultured cells and their extracted DNA were then used for cytogenetic and molecular biological analysis, leading to the discovery of the species' unique sex-determining characteristics.

Unlike those Y chromosome-lacking *Tokudaia* species, TMU has exceptionally large sex chromosomes, formed by sex chromosome-autosome fusions (Murata et al., 2012). Furthermore, multiple copies of the SRY gene are located along the long arm of the Y chromosome (Murata et al., 2010). However, the function of the SRY gene in TMU remains unclear, as all SRY copies share a specific amino acid substitution, Alanine (A) to Serine (S) at position 21. This mutation is believed to be critical for SRY-DNA binding activity within the HMG domain, suggesting that the SRY gene in TMU may be dysfunctional (Murata et al., 2010). Thus, TMU could be a key species for studying mammalian SRY evolution and the potential future of sex determination mechanisms.

In a previous study, transgenic (TG) mice carrying the TMU's SRY gene were generated, but the function of sex determination could not be examined because the SRY protein was not stably expressed in transgenic mouse cells (Ogata et al., 2019). The mouse SRY gene contains a long polyglutamine (Q-rich) domain, a species-specific structure (Bowles et al., 1999). By contrast, the *Sry* of TMU lacks the long Q-rich domain and its protein is fragile in mouse cells, indicating the long Q-rich domain plays an essential role in protein stabilization in mice. Therefore, the function of sex determination in TMU's *Sry* has remained unclear. Understanding the evolution and degeneration of the Y chromosome remains a crucial area of study in genetics, as it provides insight into broader mechanisms of chromosome evolution and genetic diversity among mammals, especially when some exceptional cases have been reported. In this study, I hereby to pursue and defining what lied within the secret of TMU's SRY. I hypothesize that the mutation might decrease the function of SRY, however, might be

subsidized by multiple intact copies instead.

In chapter one, I compared the HMG domain sequences of three intact SRY copies with the registered SRY available on NCBI database. I analysed the substitution-rate of amino acids within HMG structure of all NCBI's SRY. I predict the three-dimensional structure of TMU's HMG. I stimulated the HMG-DNA binding activity *in silico*, evaluated the changing in molecular properties, protein-DNA contacted compared to mouse.

In chapter two, I cloned HMG-coding genes of TMU's SRY along with the S to A substitution as well as mouse and its vice versa A-to-S. I analyzed the DNA-binding ability *in vitro*, using the recombinant protein produced by *E.coli* to perform the Electrophoretic Mobility Shift Assays and Fluorescence Anisotropy as the comparative methods.

In chapter three, in order to provide more details of TMU's SRY as the sex determination factor, I analyzed the DNA-bending ability *in vitro* through gel permutation assays using Polyacrylamide Gel Electrophoresis (PAGE).

Based on the results obtained through these experiments, I discussed the possibility of how TMU's SRY might be able to retain its function in sex determination even if A-to-S substitutions appeared in the HMG domain.

# Chapter I

# *In silico* analysis of the molecular properties in the SRY

# 1.1 ABSTRACT

The mammalian sex-determining gene *SRY* is highly conserved across species, with only a few exceptions. The Okinawa spiny rat, *Tokudaia muenninki* (TMU), acquired neo-sex chromosomes with multiple *Sry* copies by sex chromosome-autosome fusions. All SRY copies in TMU have a substitution from alanine (A) to serine (S) at position 21 in the high-mobility group (HMG) box, a critical DNA-binding domain, suggesting that they are nonfunctional. However, the sex determination system in TMU remains unclear, in part because the species is endangered, and it is therefore extremely difficult to obtain experimental samples. In this study, I investigate the molecular properties and function of SRY *in silico*. A comparison of SRY sequences from 225 species showed that TMU is the only species with a substitution at the 21st position. This result highlights the rarity and specificity of this substitution, suggesting that the locus is critical for sex determination. Structural predictions and DNA docking simulations showed that DNA-binding ability was retained, at least *in silico.*

## 1.2 Introduction

In mammalian species, the *SRY* gene located on the Y chromosome is a crucial genetic element responsible for initiating male sex determination. Its encoded SRY plays its pivotal role by activating SOX9, a key transcription factor, which drives testis development of embryonic bipotential gonads. The SRY protein consists of three important parts, the N-terminal, C-terminal, and high -mobility group (HMG) domain, which is an approximately 80 residues (Laudet et al., 1993), L-shaped transcriptional activation domain consist of N-terminal β-strand (consensus residues 1–11) and three α-helices (Weir et al., 1993). Packing of the β-strand against the C-terminal segment of α-3 defines the minor wing, whereas α-1, α-2, and the N-terminal portion of α-3 comprise the major wing.

Via HMG domain (Fig. 1), SRY binds to any DNA target contain A/T AACAA A/T/G core binding sequences (Harley et al., 1992; Dubin and Ostrer, 1994; Harley et al., 1994), with having the highest affinity for the (A/T) AACAA (T/A) sequence (Harley et al., 1994) which is believed to be directly activated by SOX9 (Bridgewater et al., 2003; Bhandari et al., 2012) that ensures the continued production of testosterone and anti-Müllerian hormone (AMH). Consequences as the promotion of male reproductive tract formation and suppress the development of female structures such as the Müllerian ducts. The structure of HMG-DNA binding complexes in SRY and their SOX-related proteins has been well determined, at least in humans. The novel structure of the complexes has been confirmed by multidimensional NMR spectroscopy (Murphy et al., 2001) and x-ray crystallography (Palasingam et al., 2009; Reményi et al., 2003).

Any mutation that changes the physical structure could be critical for the HMG–DNA binding complex, resulting in sex reversal. In individuals with an XY chromosome configuration, the patient may show a female phenotype, which might arise from the protein losing specificity in its DNA-binding domain due to changes in the folding structure.

Naturally, proteins tend to fold into specific compact structures despite having a vast number of possible configurations (Chakrabarti and Bhattacharyya, 2007). Folding process of protein is actually solely rely on thermodynamic properties that proximal to intramolecular interactions (Shortle, 2009), with the burial of hydrophobic groups serves as the primary source of stabilization energy in the folded structure (Kauzmann, 1959; Dill, 1990). Physicochemical properties of each amino acid residue, including steric volume and hydrophobicity, are critical for maintaining proper folding and function in the "native" state, as they preserve protein structure and stability (Zhang and Kong, 2020). This phenomenon, known as enthalpy-entropy compensation, highlights the crucial role of intramolecular hydrogen (H)-bonds in stabilizing protein structures, which could be described by using the Gibbs free energy equation (Li et al., 2021). Substitution of amino acids, even if single, but occurs at the right position, might lead to dramatic changes in protein folding structure. By eliminating favorable intramolecular contacts and increasing the enthalpy of the native state, the stability and abundance of protein would also decrease (Predki et al., 1996; Buric et al., 2025). For the protein function, because of the changes in structure, amino acid substitutions can alter the physicochemical properties of DNA-binding domains. Depending on the

substituted position and the properties of the substituted residues, changes in electric charge and molecular bonding can affect the interaction with DNA. This has been observed in well-known human DSD studies, where loss of SRY function led to complete gonadal dysgenesis, as described by Harley et al., 1992 and Werner et al., 1995. The other notable study is that the inherited mutation compatible with either male or female somatic phenotypes have been observed in both an XY father and individuals with XY Swyer syndrome. One such example is the substitution of F109S in the HMG domain (corresponding to consensus box position 54), which leads to weaker HMG–DNA binding complexes. This results in an ambiguous developmental stage, where some patients exhibit a male phenotype while others develop a female phenotype. The phenotypic outcome may depend on the availability of environmental free hydrogen atoms (free water), which can assist the mutated SRY protein in forming its bond with DNA (Racca et al., 2016).

Even though the amino acid sequences of the HMG domain in SRY are highly conserved among most mammalian species (Pontiggia et al., 1995) and are thought to carry out a similar function in all mammals in which it is present (Ross et al., 2008), compared to humans, there appear to be only a few studies on HMG–DNA complexes in other species. Especially in TMU, the endangered and protected species status makes it impossible to obtain NMR-crystallized structure of SRY or HMG-DNA complexes. To compensate for the lacking of specimens, the first chapter of my study regarding the SRY-DNA complexes in TMU has been performing with *in silico* analysis, which recently plays an important role in protein-DNA binding studies, as bioinformatic tools

provide considerably cheaper, reliable, and faster results when compared to the conservative *in vitro* or *in vivo* experimental methods. In this study, I compare three intact TMU's SRY copies with all mammalian intact SRY with completely cover the HMG domain registered on NCBI database. This analysis showed the conservation of amino acid sequences, which showed the characteristic of these multiple SRY copies and might provide a glimpse of its function. Next, I modelled the SRY2, which shares most similarity to mouse SRY, rendered it from the amino acid sequences along with its sequence editing version (StoA), Mouse WT, and Mouse A-to-S. I tried docking the 3D structure of modelled HMG with the enhancer 13 (Enh13) upstream region of Sox9 gene locus, the DNA binding target common in mouse (Gonen et al., 2018), rat, and TMU's SRY.

## 1.3 Materials and Methods

The sequence comparison study between each SRY copy of TMU and the novel mammalian species has been reported by our laboratory before in the past. However, this is the first time in several years that using all available identical sequences on the updated NCBI database in order to confirm and prove the significance of A-to-S substitution that might be unique to TMU.

**HMG sequence analysis**

All SRY sequences from mammalian species in the NCBI database were downloaded as of September 2023, trimmed to obtain the HMG domain, and filtered to remove duplicates. An alignment was generated using AliView (Larsson, 2014), an alignment viewer released under the Open-Source Software License GNU General Public License, version 3.0 (GPLv3); the source code is available at GitHub (https://github.com/AliView/AliView). MEGA X (Kumar et al., 2018) was used for statistical model prediction and to generate a substitution matrix of the HMG domain via maximum likelihood methods based on the JTT model and gamma-distributed rates with 1000 bootstrap replicates (Jones et al., 1992)

**Three dimensional structure prediction of HMG domain**

Previously reported TMU *Sry* sequences, SRY2 (BAJ08420) (Murata et al., 2010), were used for structure prediction based on the highest alignment score for HMG sequences

compared with mouse, the model, and species with structural annotations. Along with TMU, mouse SRY (NP_035694) was saved as a FASTA file and the amino acid sequence was trimmed to retain the HMG domain. A-to-S substituted sequences were generated and then submitted to Robetta, a protein structure prediction server developed by the Baker lab at the University of Washington (Kim et al., 2004). Three-dimensional (3D) structures were predicted using a deep learning-based method, RoseTTAFold (Baek et al., 2021).

**DNA-binding complex simulation and structure comparison**

The predicted structure of the HMG domain protein together with the SRY-binding DNA sequences were subsequently analyzed using pyDockDNA, a web server for predicting protein-DNA interactions based on pyDock and FTDOCK algorithms (Rodríguez-Lumbreras et al., 2022). We selected a 30-bp DNA sequence from the testis-specific enhancer of *Sox9*, Enh13 (706 bp), which is conserved among rodent species, including rat and mouse, as the HMG-binding DNA (Gonen et al., 2018). The PDB files were downloaded and the model with highest prediction score was visualized. The molecular structures were compared between each sample using UCSF Chimera and Chimera X (Petterson et al., 2004). Protein-DNA contact maps were generated by using Mapiya (Badaczewska-Dawid et al., 2022) with default parameters.

## 1.4 Results

### A-to-S substitution is unique in TMU SRY

We evaluated the A-to-S substitution at the 21st position of HMG domain in the TMU *Sry* gene through a comparative sequence analysis. We obtained 537 SRY sequences from various mammalian species available in the National Center for Biotechnology Information (NCBI) database. We trimmed these sequences to retain only the first 76 amino acids of the HMG domain, which is highly conserved across mammalian species. To screen for intra-specific variation, 228 sequences from 225 species were finally retained (Fig.1). This dataset included three previously identified TMU SRY sequences: SRY1, SRY2, and SRY3 (Murata et al., 2010). A multiple sequence alignment of the HMG domain was generated using the ClustalW algorithm, revealing strict conservation at position 21 of the HMG domain across mammalian species, with the exception of TMU (showing an A-to-S substitution in Fig.1 and Fig. 3). We also employed the Maximum Likelihood Estimate of Substitution Matrix to further elucidate the nature of amino acid substitutions in the HMG domain. A matrix was generated with 1,000 bootstrap iterations. Substitution patterns and rates were estimated under the Jones-Taylor-Thornton (1992) model (+G). A discrete gamma distribution was used to model evolutionary rate differences among sites (5 categories, (+G), parameter = 1.1882), with relative values of instantaneous rates (r values) simplified by setting the sum of r values to 100. The amino acid frequencies were 7.69% (A), 5.11% (R), 4.25% (N), 5.13% (D), 2.03% (C), 4.11% (Q), 6.18% (E), 7.47% (G), 2.30% (H), 5.26% (I), 9.11% (L), 5.95% (K), 2.34% (M), 4.05% (F), 5.05% (P), 6.82% (S), 5.85% (T), 1.43%

(W), 3.23% (Y), and 6.64% (V), and A was most likely to be replaced with threonine (T) (r = 1.82), followed by serine (r = 1.55) and valine (r = 1.16) (Fig. 4). These results suggest that although A can be easily replaced with S, the residue is highly conserved across mammals other than TMU. These results indicate that the 21st amino acid in the HMG domain is functionally important.

**A-to-S substitution does not have a major impact on the HMG domain structure**

To understand the potential functional implications of the A-to-S substitution, we analyzed the 3D structure of the HMG domain. Four models with the highest prediction scores, corresponding to wild-type mouse HMG, mouse HMG with an A-to-S substitution, wild-type TMU HMG, and TMU HMG with a S-to-A substitution, were generated using the Robetta computational platform. An alignment of these 3D models revealed structural similarities across all samples, regardless of whether A or S was present at the 21st position (Fig. 5-1A and B, Fig. 5-2). Remarkably, angstrom error estimates were low for all of the predicted models (Fig. 6). For residues just after A or S at position 21 of HMG, specifically beginning from glutamine (Q) at position 23, the error exceeded 1.0 angstroms, peaking at 3.4 angstroms around S 26th and decreasing to less than 1.0 angstroms by Q 31st. Subsequent significant angstrom error estimates were observed after tyrosine (Y) at position 71, reaching over 18 angstroms at position 80 of the HMG domain, indicating that the lack of the Q-rich domains in the C-terminus region of the SRY protein may contribute to the difficulties in model prediction. However, the Q-rich domain varies considerably among species and is not highly

conserved, even within the rodent family; its inclusion in the structural model could interfere with the interpretation of the A-to-S substitution in the HMG-DNA complex. Therefore, we focused exclusively on the HMG domain, which is the critical DNA-binding domain of the SRY protein.

**DNA-HMG domain interaction remains intact in A and S variants**

To assess the impact of the A-to-S substitution on the DNA-HMG domain interaction, we performed *in silico* DNA-binding predictions using the pyDockDNA platform. The selected HMG models in PDB format were used as receptors, while the conserved rodent Enh13 nucleotide sequence (5′-TCCACTTCTAAACAAACAGCTGAGGGGA GT-3′) was used as the ligand. The docking analyses revealed that the DNA-binding patterns of the TMU HMG were similar to those of mouse HMG, regardless of whether A or S was present at the 21st position (Fig. 5-1, Fig. 5-2). The A-to-S substitution was positioned near the protein surface; however, it did not exhibit direct contact with the Enh13 DNA model in the docking results when the interaction bond length was set at 4 angstroms. This suggests that the substitution does not impact the DNA-HMG domain interaction substantially. Of note, the hydroxyl group in S resulted in the formation of an additional hydrogen bond within the protein-protein folding structure between S at position 21 and arginine (R) at position 17. The molecular distance of this hydrogen bond was 2.767 angstroms in TMU HMG and 2.773 angstroms in the mouse homolog (Fig. 5-1 A and B). These subtle structural differences suggest that the A-to-S substitution influences intramolecular stability, although the effect is weak and unlikely

to alter or disrupt the DNA-HMG domain interaction, as evidenced by the lack of significant change between the A/S and S/A transitions. The protein contact map results retrieved from Mapiya revealed that rather than A/S at position 21 of the HMG domain, in TMU the nearby area R17 exhibited obviously strong interactions with the nucleotide molecules (Fig. 7) in alignment with the structural change in the surface area of R17 (Fig. 8). Notably, the bonds were weaker and some contact was lost in TMU HMG compared with that of mouse, suggesting that the additional hydrogen bond between S at position 21 and R at position 17 might detach R17 or widen the area of contact with DNA.

## 1.5 DISCUSSION

The number of known protein sequences has increased rapidly in recent years, improving such databases is important, as it confirms the significance of our studies. The more NMR structures that are submitted to the database, the more accurately the native state of protein structures and protein–nucleic acid interactions can be predicted. The one only A-to-S substitution in TMU once again confirms its significance as species-specific characteristic. The A at this (21$^{st}$) position is highly conserved across mammals and was detected in 228 SRY sequences from 225 species evaluated in this study; this high conservation suggests that the site is critical in sex determination. The discovery of the substitution in TMU raises important questions about the evolutionary forces driving this unique substitution in the species. Given that the A-to-S substitution was not observed in any other mammalian species, even if A is easily substituted to S, it likely represents an isolated evolutionary event specific to TMU, occurring before SRY duplication events.

However, except for I10, which was M in rodent species, and R4, which was C in TMU SRY1 (BAJ08419) and H in TMU SRY3 (BAJ08421), other amino acids involved in binding to the minor groove of DNA were strictly conserved. The predicted structure also indicated that the HMG of TMU might be able to fold closely similar to those native stage in mouse, meaning there are not so much difference in Gibbs free energy between these samples.

The docking simulation results apparently could bind to the Enh13, resembling to those in mouse, and vice versa, to the substitution edited version, regardless of A or

S. Moreover, the additional hydrogen bonds between S21 and R17 have been confirmed between HMG molecules. This bond is likely to be responses why the bond between DNA and R17 could be considered as relatively weak when compare to mouse showed in the protein contact map.

This *in silico* results indicated that the SRY of TMU retains its DNA-binding ability, at least via the HMG domain. There was a slight reduction, but not complete loss, of the DNA-binding activity of SRY. However, since there is limited information regarding DNA-binding ability that could be archive from *in silico* prediction, I would like to validate and confirm this hypothesize using *in vitro* experiments as the next chapter.

## 1.6 Figures

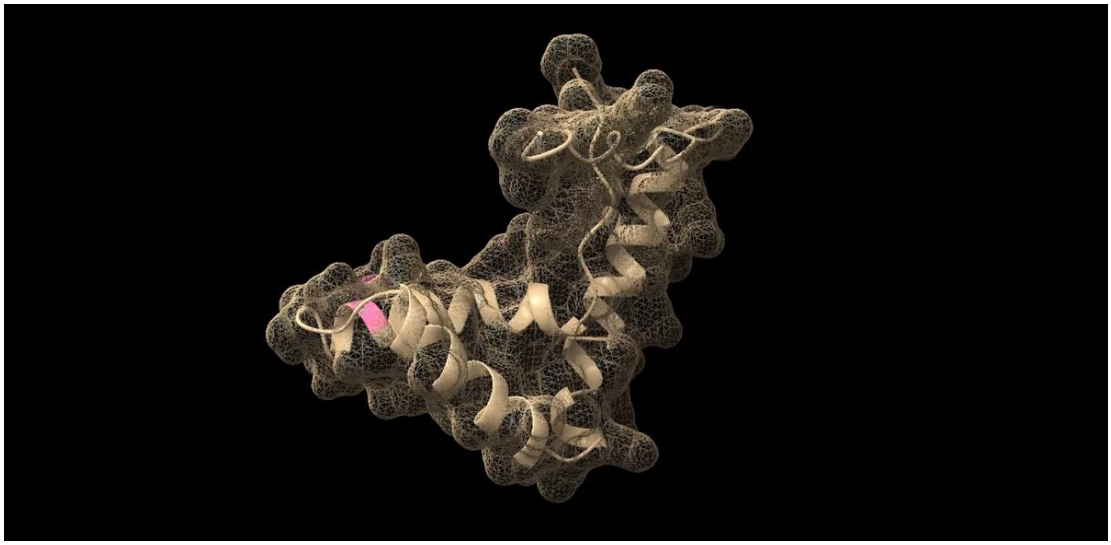

**Fig. 1** The 3D structure of the HMG domain was generated from the obtained mouse SRY sequence. Folded amino acids are shown as a ribbon, while the molecular surface is represented as a mesh. The 21st alanine residue is labeled in pink.

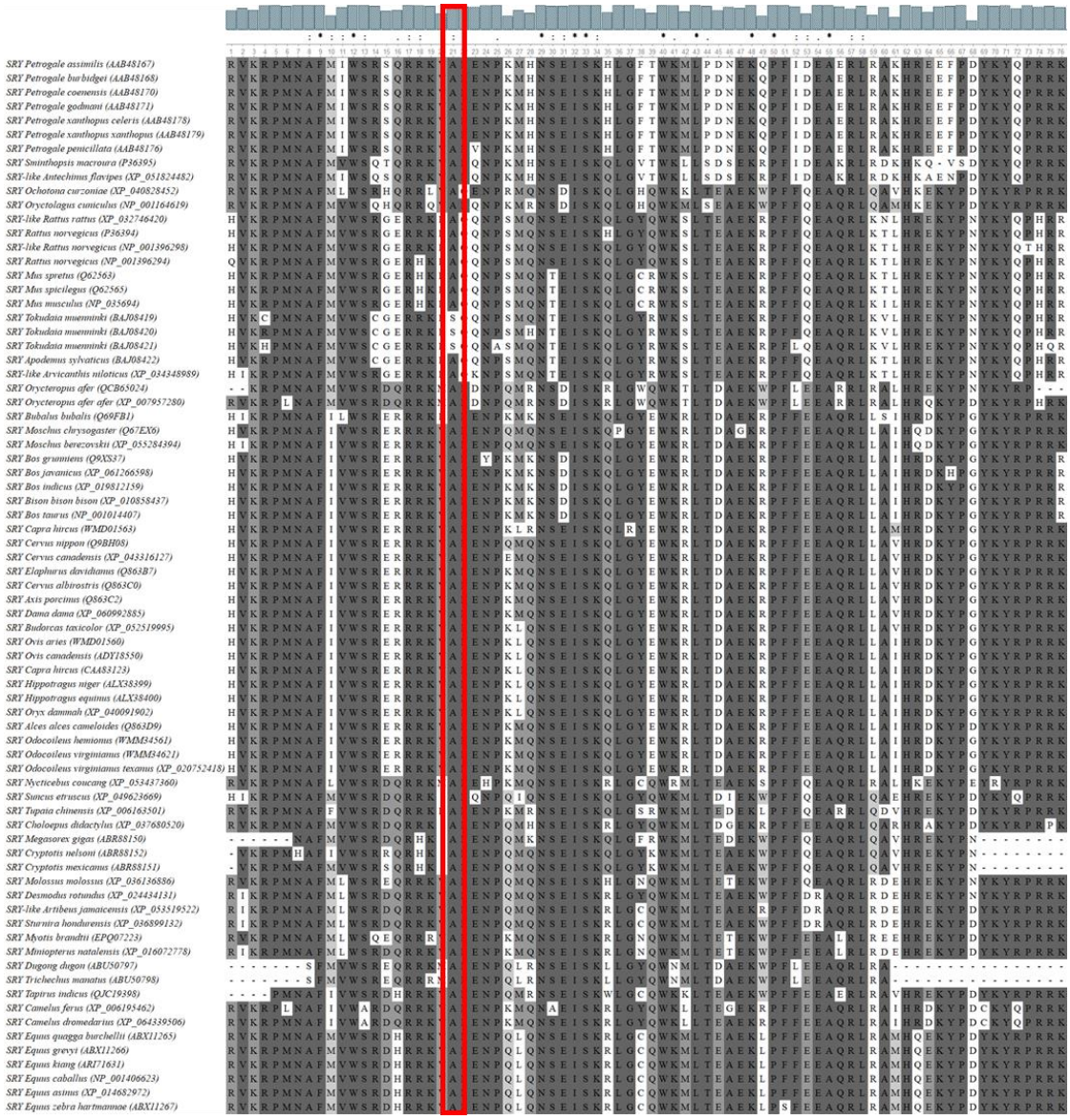

**Fig. 2** The completed 228 individual HMG sequences alignment retrieved from 225 mammalian species showed that the Okinawa spiny rat is the only one with an alanine-to-serine substitution at the 21st position of the HMG domain (Red box).

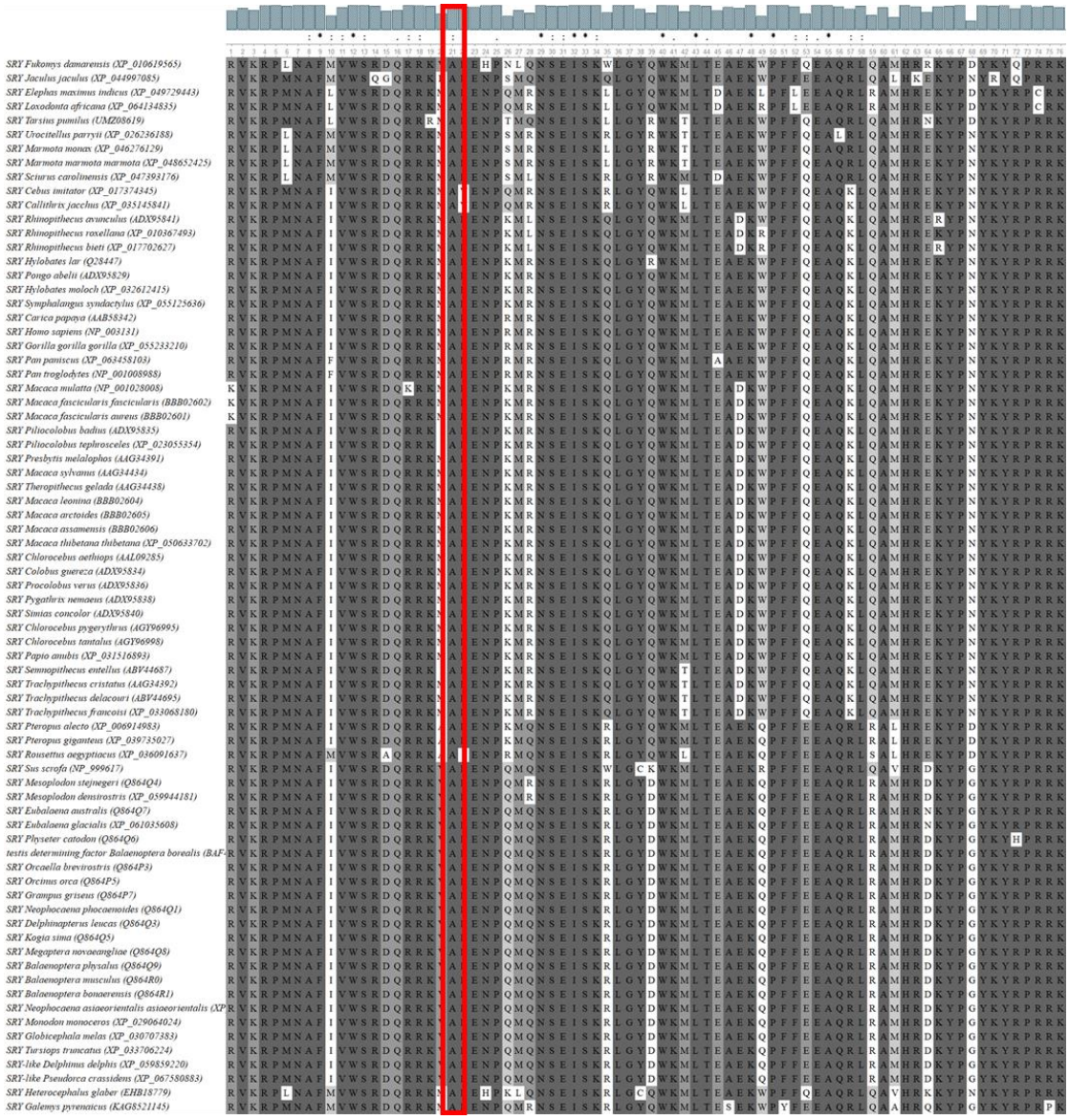

**Fig. 2** The completed 228 individual HMG sequences alignment retrieved from 225 mammalian species showed that the Okinawa spiny rat is the only one with an alanine-to-serine substitution at the 21st position of the HMG domain.

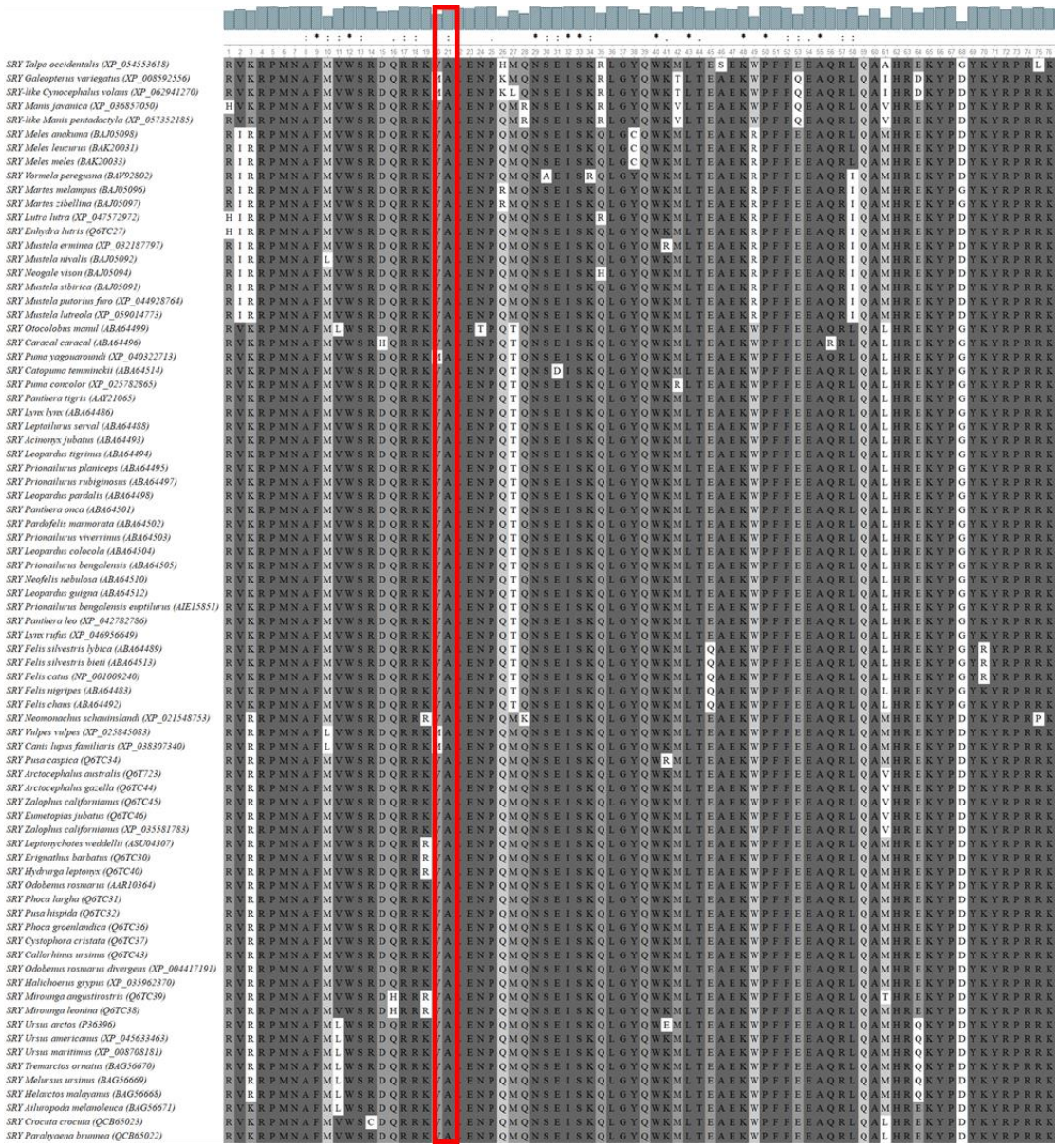

**Fig. 2** The completed 228 individual HMG sequences alignment retrieved from 225 mammalian species showed that the Okinawa spiny rat is the only one with an alanine-to-serine substitution at the 21st position of the HMG domain.

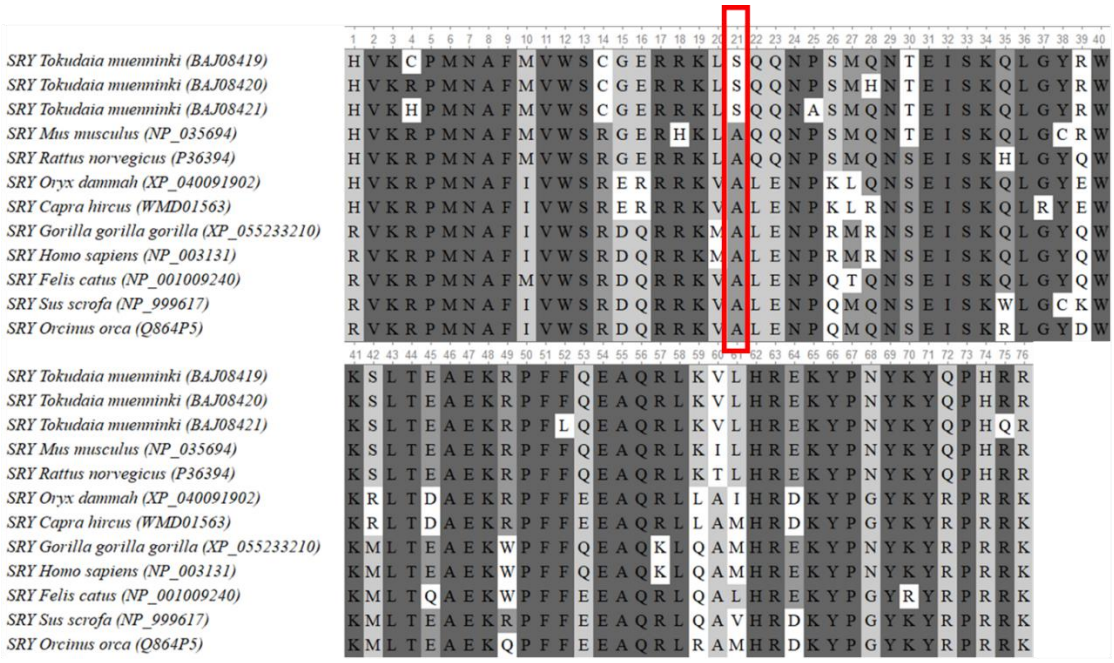


**Fig. 3** Representative sequence alignment of the HMG domain in various mammalian species. The Okinawa spiny rat is the only species with the A-to-S substitution at position 21 of the HMG domain (red box).

| Maximum Likelihood Estimate of Substitution Matrix | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | R | N | D | C | Q | E | G | H | I | L | K | M | F | P | S | T | W | Y | V |
| A | - | 0.1407 | 0.1230 | 0.2198 | 0.0604 | 0.1185 | 0.3417 | 0.6737 | 0.0262 | 0.0985 | 0.1464 | 0.1139 | 0.0570 | 0.0290 | 0.5131 | 1.3742 | 1.3896 | 0.0063 | 0.0234 | 1.0058 |
| R | 0.2117 | - | 0.0994 | 0.0411 | 0.1071 | 0.6428 | 0.1020 | 0.5262 | 0.3822 | 0.0651 | 0.1757 | 2.0124 | 0.0523 | 0.0137 | 0.1860 | 0.3540 | 0.1971 | 0.0934 | 0.0394 | 0.0591 |
| N | 0.2226 | 0.1195 | - | 1.4767 | 0.0330 | 0.1638 | 0.1855 | 0.2999 | 0.4802 | 0.1340 | 0.0649 | 0.7811 | 0.0402 | 0.0155 | 0.0319 | 1.7910 | 0.7141 | 0.0021 | 0.1175 | 0.0567 |
| D | 0.3295 | 0.0410 | 1.2234 | - | 0.0111 | 0.1110 | 2.4877 | 0.4926 | 0.1229 | 0.0316 | 0.0290 | 0.0871 | 0.0231 | 0.0068 | 0.0333 | 0.2083 | 0.1289 | 0.0043 | 0.0760 | 0.1084 |
| C | 0.2287 | 0.2697 | 0.0690 | 0.0280 | - | 0.0194 | 0.0173 | 0.2114 | 0.0863 | 0.0410 | 0.0777 | 0.0151 | 0.0496 | 0.1424 | 0.0324 | 0.7616 | 0.1424 | 0.0820 | 0.3538 | 0.2136 |
| Q | 0.2216 | 0.7992 | 0.1694 | 0.1385 | 0.0096 | - | 1.0944 | 0.0895 | 0.6766 | 0.0213 | 0.3346 | 0.9143 | 0.0554 | 0.0096 | 0.4209 | 0.1939 | 0.1588 | 0.0128 | 0.0426 | 0.0618 |
| E | 0.4252 | 0.0843 | 0.1276 | 2.0650 | 0.0057 | 0.7278 | - | 0.4323 | 0.0291 | 0.0305 | 0.0461 | 0.5343 | 0.0213 | 0.0092 | 0.0503 | 0.1106 | 0.1006 | 0.0085 | 0.0106 | 0.1602 |
| G | 0.6936 | 0.3600 | 0.1706 | 0.3383 | 0.0575 | 0.0492 | 0.3576 | - | 0.0240 | 0.0147 | 0.0328 | 0.0833 | 0.0158 | 0.0106 | 0.0545 | 0.6631 | 0.0962 | 0.0405 | 0.0088 | 0.1618 |
| H | 0.0876 | 0.8493 | 0.8873 | 0.2742 | 0.0762 | 1.2091 | 0.0781 | 0.0781 | - | 0.0495 | 0.2552 | 0.1619 | 0.0400 | 0.0952 | 0.2990 | 0.2628 | 0.1447 | 0.0095 | 0.9787 | 0.0419 |
| I | 0.1440 | 0.0633 | 0.1082 | 0.0308 | 0.0158 | 0.0167 | 0.0358 | 0.0208 | 0.0216 | - | 1.1024 | 0.0624 | 0.5862 | 0.1632 | 0.0258 | 0.1432 | 0.7743 | 0.0100 | 0.0508 | 3.2788 |
| L | 0.1236 | 0.0986 | 0.0303 | 0.0163 | 0.0173 | 0.1510 | 0.0312 | 0.0269 | 0.0644 | 0.6365 | - | 0.0452 | 0.4682 | 0.5254 | 0.2779 | 0.2096 | 0.0827 | 0.0394 | 0.0404 | 0.6062 |
| K | 0.1472 | 1.7283 | 0.5579 | 0.0751 | 0.0052 | 0.6315 | 0.5550 | 0.1045 | 0.0626 | 0.0552 | 0.0692 | - | 0.0758 | 0.0052 | 0.0567 | 0.1678 | 0.2930 | 0.0066 | 0.0147 | 0.0427 |
| M | 0.1872 | 0.1142 | 0.0730 | 0.0505 | 0.0430 | 0.0973 | 0.0561 | 0.0505 | 0.0393 | 1.3176 | 1.8229 | 0.1928 | - | 0.0917 | 0.0430 | 0.1011 | 0.6420 | 0.0150 | 0.0318 | 1.0462 |
| F | 0.0551 | 0.0173 | 0.0162 | 0.0087 | 0.0714 | 0.0097 | 0.0141 | 0.0195 | 0.0541 | 0.2119 | 1.1819 | 0.0076 | 0.0530 | - | 0.0389 | 0.3341 | 0.0422 | 0.0400 | 0.9192 | 0.2044 |
| P | 0.7814 | 0.1882 | 0.0269 | 0.0338 | 0.0130 | 0.3426 | 0.0616 | 0.0807 | 0.1362 | 0.0269 | 0.5013 | 0.0668 | 0.0199 | 0.0312 | - | 0.9869 | 0.3573 | 0.0052 | 0.0191 | 0.0728 |
| S | 1.5495 | 0.2652 | 1.1161 | 0.1567 | 0.2267 | 0.1169 | 0.1002 | 0.7263 | 0.0886 | 0.1105 | 0.2800 | 0.1464 | 0.0347 | 0.1984 | 0.7308 | - | 1.4500 | 0.0231 | 0.1053 | 0.1406 |
| T | 1.8267 | 0.1722 | 0.5188 | 0.1130 | 0.0494 | 0.1115 | 0.1063 | 0.1228 | 0.0569 | 0.6962 | 0.1288 | 0.2980 | 0.2568 | 0.0292 | 0.3084 | 1.6904 | - | 0.0060 | 0.0337 | 0.3938 |
| W | 0.0337 | 0.3338 | 0.0061 | 0.0153 | 0.1164 | 0.0368 | 0.0368 | 0.2113 | 0.0153 | 0.0368 | 0.2511 | 0.0276 | 0.0245 | 0.1133 | 0.0184 | 0.1103 | 0.0245 | - | 0.1256 | 0.0827 |
| Y | 0.0556 | 0.0624 | 0.1546 | 0.1207 | 0.2224 | 0.0542 | 0.0203 | 0.0203 | 0.6969 | 0.0827 | 0.1139 | 0.0271 | 0.0231 | 1.1525 | 0.0298 | 0.2224 | 0.0610 | 0.0556 | - | 0.0569 |
| V | 1.1648 | 0.0455 | 0.0363 | 0.0838 | 0.0653 | 0.0383 | 0.1491 | 0.1820 | 0.0145 | 2.5974 | 0.8317 | 0.0383 | 0.3687 | 0.1247 | 0.0554 | 0.1444 | 0.3469 | 0.0178 | 0.0277 | - |

**Fig. 4** HMG substitution matrix analysis conducted in MEGA X showed the amino acid substitution probability within the HMG domain among mammalian species. Gradient scales of which amino acid (row), is likely to be substituted to (column) range from lowest chance to be substituted (blue) until highest chance to be substituted (red). For estimating ML values, a tree topology was automatically computed. The maximum Log likelihood for this computation was -2580.654. A total of 83 positions in the final dataset.

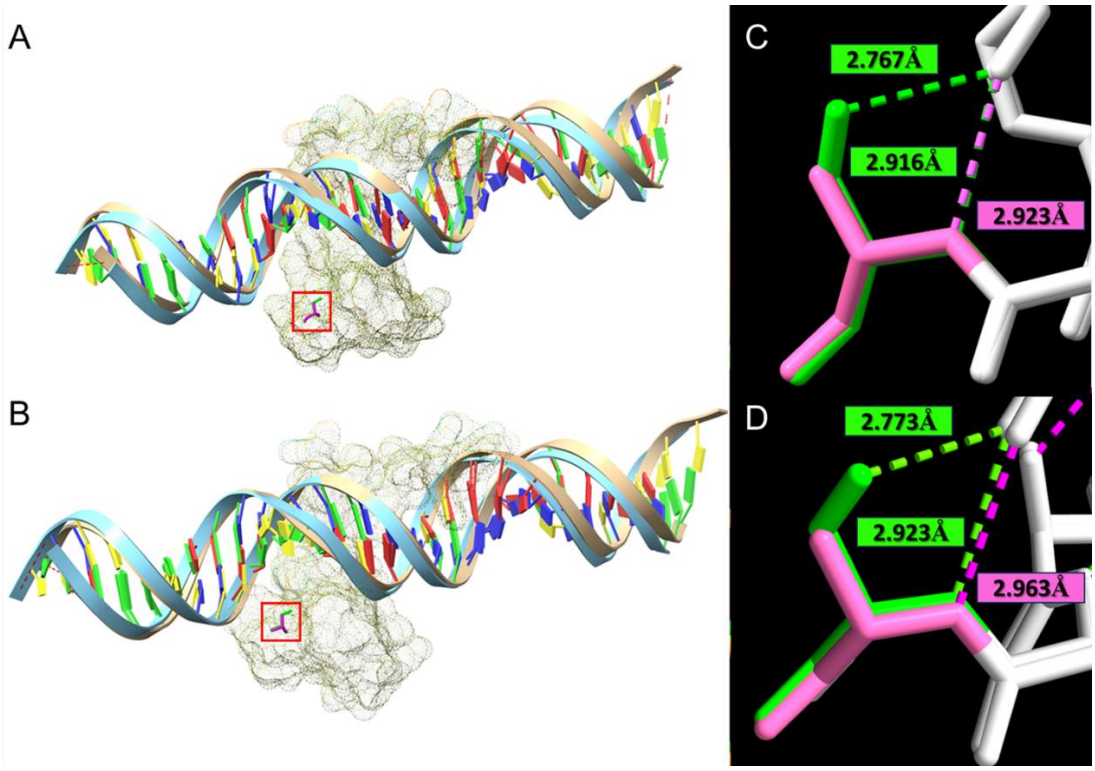


**Fig. 5-1** 3D docking results showing the similarity in HMG-DNA binding between TMU (A) and mouse (B). The magnified positions are indicated by red boxes. Molecular properties of the S variant (green) indicate more hydrogen bonds within the protein-protein folding structure compared with those for the A variant (pink) in both TMU (C) and mouse (D).

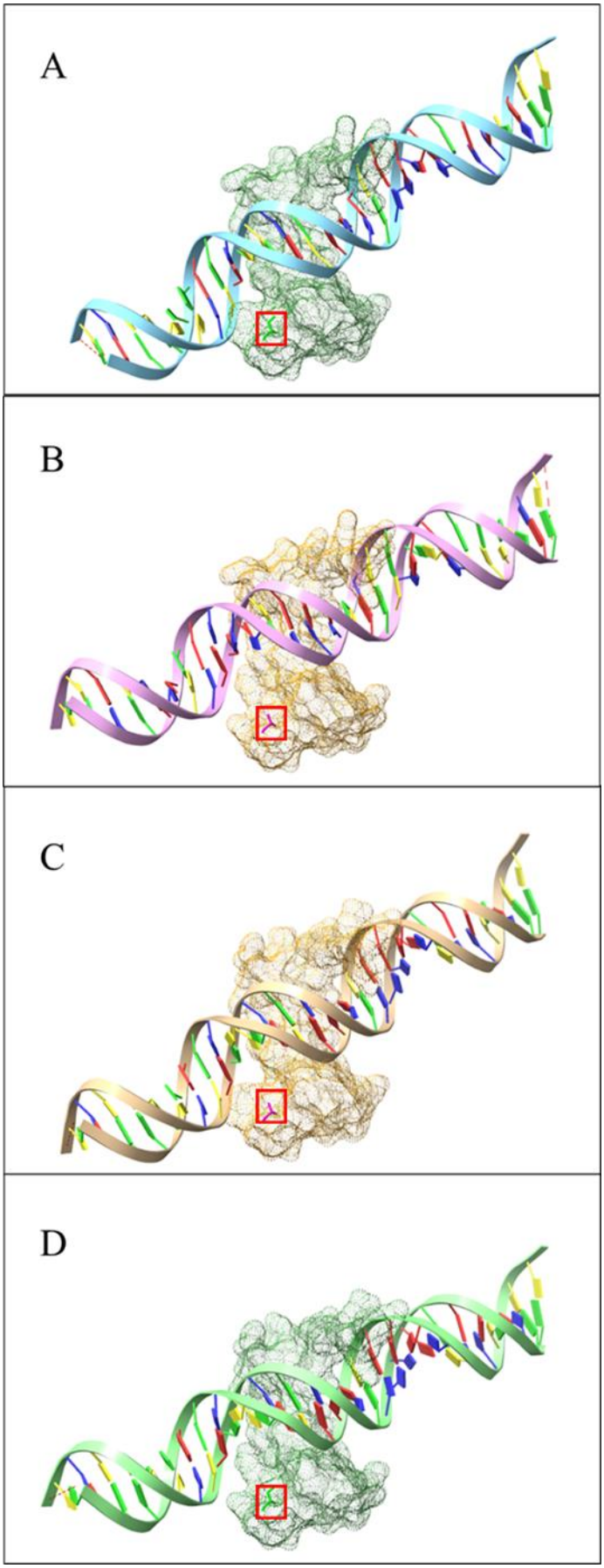


**Fig. 5-2** The 3D docking results showed similarity of the HMG-DNA binding complexed between TMU WT (S), TMU S to A (B), mouse (C), and mouse A-to-S (D). The position of the substitution (red box) is proximally close to the surface of the HMG protein, but does not make direct contact with the DNA.

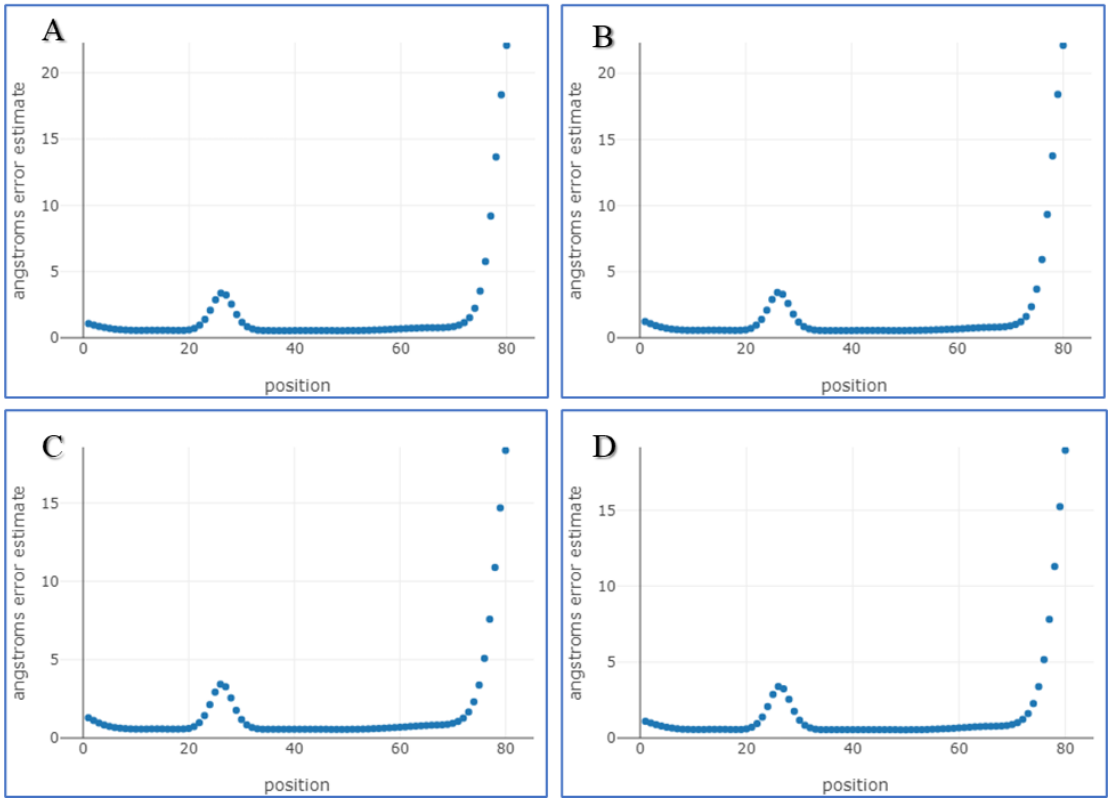


**Fig. 6** Angstrom error, per residue local error estimates as predicted distances in angstrom (Å) between the Cα positions of the model compared to the native structure in mouse WT (A), mouse A-to-S (B), TMU WT (C), and TMU S to A (D). All models showed average errors between 1-5 Å at 1st to 75th positions of HMG domain, which are very conserved across species.

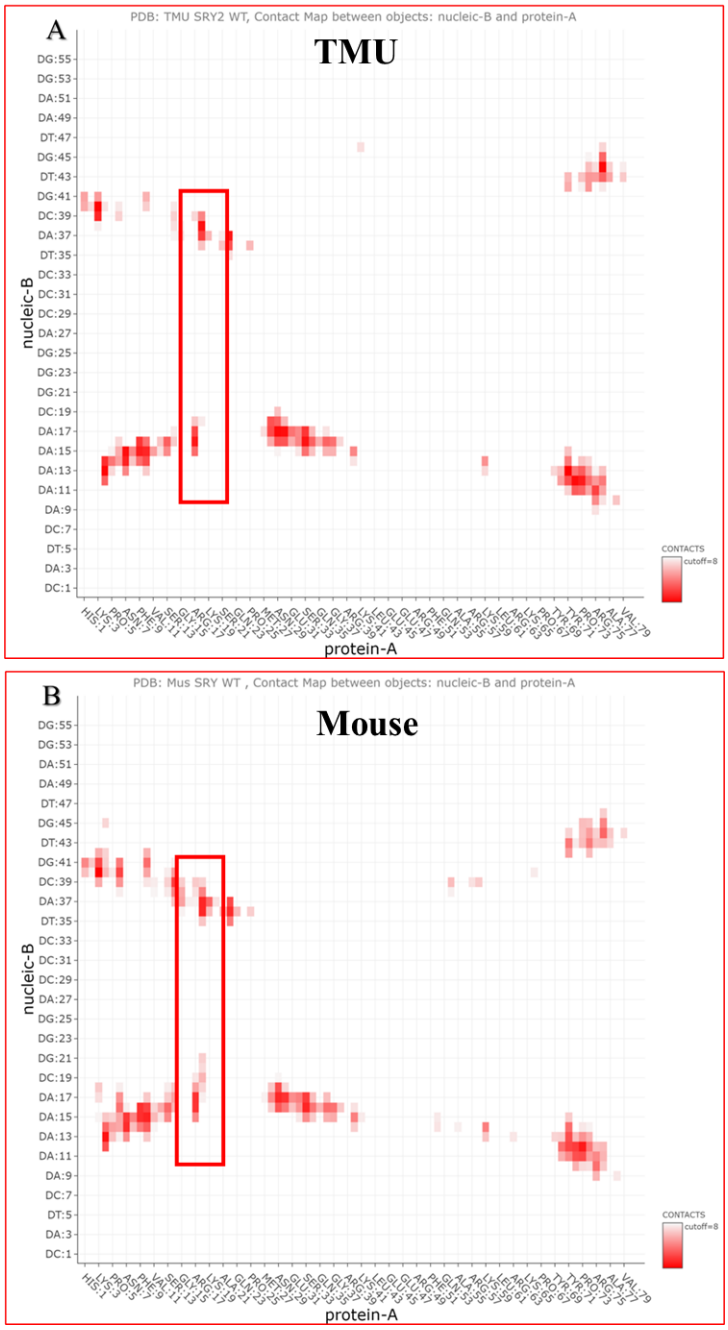


**Fig. 7** The protein contact maps with default cutoff at 8 Angstrom showed HMG-DNA binding complexed between TMU (A) and mouse (B) to the Enh13 core binding sequence of SRY. A-to-S substitution might reduce the binding affinity of R17 and R18 (red boxes) to the DNA.

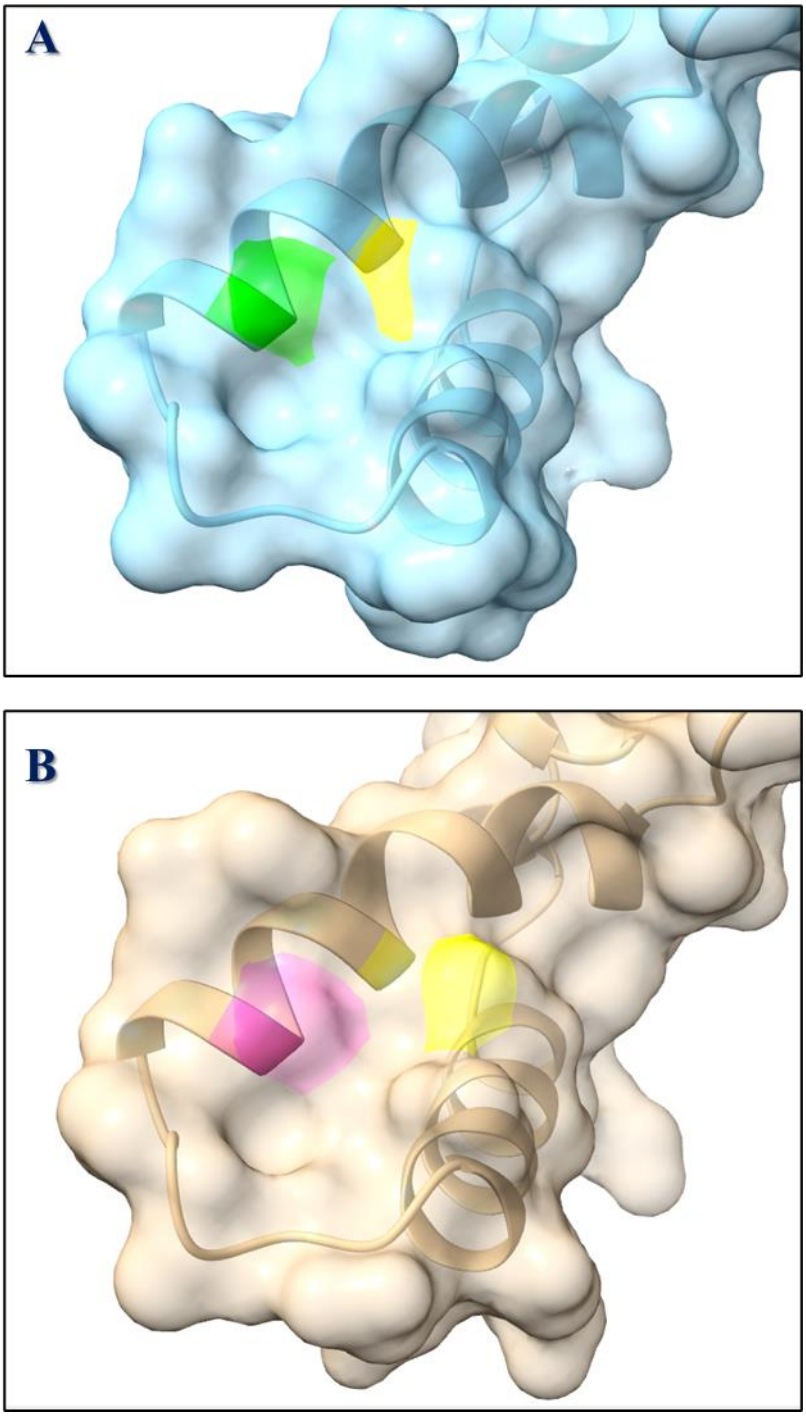


**Fig. 8** The 3D HMG-domain structures of TMU (A) and Mouse (B) show changes in the surface area around residue R17 (yellow), caused by an A-to-S substitution. In TMU, the presence of S21 (green) brings R17 closer through additional hydrogen bonds, compared to A21 (pink) in the mouse.

# Chapter II

# Analysis of the DNA-binding activity and affinity of the SRY

## 2.1 Abstract

One of the most well-known functions of the SRY is its transcriptional activity. As the transcription factor, in order to upregulate its relative, SOX9 activities, SRY needs to precisely bind to its DNA target, Enh13. In the last chapter, I identified the molecular properties of TMU, the multi SRY copies of critically endangered species using *in silico* methods. The results showed that it might possibly be bound to the Enh13 DNA. However, the amount of SRY required to be bound to the DNA (DNA binding affinity) is yet to be confirmed. In this chapter, *in vitro* experiments included western blotting, electrophoretic mobility shift assays, and fluorescence anisotropy, which were conducted using recombinant protein from *E. coli*. The results showed that although the affinity was slightly lower than that of the mouse homolog, DNA-binding ability was retained. These findings provide insight into the effects of the A-to-S substitution in SRY on sex determination.

## 2.2 Introduction

To activate a gene, a transcription factor generally interacts with DNA through a combination of direct and indirect readout mechanisms (Vonderfecht et al., 2008). Direct readout consists of specific chemical interactions between protein and nucleic-acid bases (Pabo and Sauer, 1992) primarily around two-thirds to hydrogen bonds, similar with van der Waals forces, and one-sixth each of hydrogen and water-mediated bonds (Luscombe et al., 2001). On the other hand, Indirect Readout, protein senses the overall conformation of the DNA helix by perturbing to the recognition of DNA structure, included shape, intrinsic curvature, topology, ordered water structures, local geometry of backbone phosphates, and flexibility, influenced by the nucleotide sequence (Aeling et al., 2006). In addition, Hydrophobic interactions from aromatic residues like phenylalanine and tyrosine can also engage in stacking interactions with nucleotide bases, as have been reported (Anjana et al., 2012; Chakrabarti and Bhattacharyya, 2007), as well as provide significant stability to DNA-protein complexes (Wilson et al., 2014). These interactions are indicated to be heavily influenced by amino acid substitutions. For example, amino acids with ionizable side chains positively charged residues like arginine and lysine (Zhou et al., 2018) often interact with the negatively charged phosphate backbone of DNA (Martin et al., 2022). Substituting these residues with neutral or negatively charged amino acids can weaken the interactions, reducing binding affinity. In similar terms, alterations in residues included glutamine and asparagine responses for hydrogen bonds forming with DNA bases will disrupt the born formation, weakening both specificity and affinity (Corona

and Guo, 2016).

SRY, as a transcription factor, recognizes DNA through its HMG domain by partially intercalating into the minor groove of the DNA helix (King and Weiss, 1993), using an α-helix at the 3′ end of the binding site and a flexible, cationic C-terminal tail at the 5′ end, which contains four lysines and three arginines in close proximity to the DNA (Etheve et al., 2016). Therefore, the topological mechanism of sequence specificity of SRY could also be drastically changed when some amino acid residues are replaced by others, which can significantly alter its DNA-binding activity, impacting its role in male sex determination. In humans, the Y129N mutation, in which Tyrosine is replaced by Asparagine, is associated with partial gonadal dysgenesis, has been shown to reduce DNA-binding affinity and alter the salt dependence of the SRY-DNA interaction (Baud et al., 2002). Similarly, the A66T (Alanine to Threonine) mutation with 46 XY-pure gonadal dysgenesis from a G to A transition within the HMG domain, results of this substitution have been reported as it significantly reduces SRY's DNA-binding capacity and impairs its nuclear import. Indicated that the 66th alanine of the HMG domain is important for both DNA interaction and proper intracellular localization (Wang et al., 2018). These findings remind us of the critical influence of specific amino acid residues on the structural and functional integrity of SRY, with mutations especially in the HMG binding domain leading to significant disruption in DNA binding and, consequently, male sexual development.

In this chapter, to determine the effect of amino acid substitution on TMU's SRY in HMG-DNA binding activity, I have employed two techniques, the electrophoretic

mobility shift assay (EMSA) and the fluorescence anisotropy (FA), which could provide qualitative and quantitative data, respectively. For EMSA, it has long been recognized as one of the most prominent cornerstone methods for analyzing protein–nucleic acid interactions (Moxley et al., 2003). This method utilizes the principle that protein–DNA or protein–RNA complexes migrate more slowly through a non-denaturing gel during electrophoresis compared to free nucleic acids, due to the increased size and altered charge of the complexes (Hellman and Fried, 2007). In addition, FA, also known as fluorescence polarization, is a powerful and sensitive technique to assess molecular complex formation by measuring the degree of polarization of emitted light after excitation (Lacowicz, 2006). To study DNA–protein interactions, this method leverages the principle that when a fluorescently labeled DNA molecule binds to a protein, the resulting complex exhibits altered rotational motion, leading to changes in fluorescence anisotropy that can be quantitatively analyzed to determine binding affinities (Weaver and Whelan, 2021).

## 2.3 Materials and Methods

### Ethical consideration

The Okinawa spiny rat (*Tokudaia muenninki*, 2*n*=46, XX/XY) is endangered (The IUCN Red List of Threatened Species, DOI: e.T21972A22409515), and has been protected by the Japanese government as a natural monument and national endangered species of wild fauna and flora since 1972 and 2016, respectively. No animal samples were used in this study.

### Preparation of the plasmid carrying the HMG sequence

For HMG domain recombinant protein constructs, primers were designed to specifically amplify the HMG region. (F- AAA CTCGAG CAT GTC AAG CGC /R- GGG AAGCTT TGA CAC TTT AGC CC for mouse, and F- AAA CTCGAG CAT GTC AAG CGC / R- GGG AAGCTT TGG CAC TTT AGC TC for TMU). All products were double digested with restriction enzymes (TaKaRa, Kusatsu, Japan) and ligated into the XhoI/HindIII restriction site of pCold TF, a fusion cold shock expression vector (TaKaRa). Ligation products were electrophoresed using 1% agarose gels stained with MIDORI Green (Nippon Genetics, Tokyo, Japan) and purified using the NucleoSpin Gel and PCR Clean-up Kit (MACHEREY-NAGEL GmbH & Co, KG, Düren, Germany). The products were confirmed by DNA sequencing using the Hitachi Applied Biosystems ABI 3100 Avant DNA Sequencer (HITACHI, Tokyo, Japan) after transformation into the JM109 *Escherichia coli* strain (Nippon Gene Ecos) according to the manufacturer's instructions. After colony selection using antibiotics (50 μg/ml

ampicillin in 1.5% LB agarose media) and plasmid extraction using the FastGene DNA Extraction Kit (Nippon Genetics), inserted DNA was confirmed again using the Colony PCR using SimpliAmp Thermal Cycler (Thermo Fisher, Japan) and KAPA Taq PCR Kit (Nippon Genetics).

**Recombinant HMG domain protein production**

To produce the recombinant protein, recombinant plasmid vectors, pCold.TF carrying the HMG sequence were transformed into *E. coli* BL21 (Nippon Gene ECOS), a bacterial recombinant protein expression system, using the heat shock method. The transfected BL21 *E. coli* was precultured overnight in 10 ml of 2× Yeast Extract Tryptone liquid medium containing 50 μg/ml ampicillin, incubated at 37°C, and shaken at 200 rpm in a shaking chamber. On the following day, large-scale recombinant protein production was performed by transferring the precultured *E. coli* into a flask with 2× YT medium (Yeast extract 20 g, Tryptone 20 g, NaCl 10 g, 2 M NaOH 3.2 ml). Additional incubation was performed until the cell concentrations measured by the optical density at 600 nm (OD600) reached 0.6 and samples were immediately cooled with ice water for 15 minutes. After cold shock, the cultures were shaken at 15°C for 30 minutes before the addition of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG), incubation for another 24 hours, and collection by centrifugation (10,000 *g*, 4°C, 5 minutes per 500 ml of culture). Cell lysis was performed by mixing the collected suspension with 35 ml of lysis buffer (50 mM Tris-HCI pH.8.0, 0.5 M NaCl, 0.1% NP-40), 0.05 g of lysozyme, 200 μl of 1 M $MgCl_2$, 20 μl of 1 M $MnCl_2$, and 200 μl of 1

mg/ml DNase I, with mixing using a magnetic stirrer at 4°C for 30 minutes. The lysed samples then were sonicated using the Branson SONIFIER 250 for 10 minutes and centrifuged at 18,000 rpm for 30 minutes to remove the cell debris. The HMG-TF recombinant proteins were purified using the ÄKTA™ prime plus HPLC system on a HisTrap HP column (Cytiva, Uppsala, Sweden) following the instructions provided. The fusion tag (trigger factor [TF] chaperone) was removed from the HMG-TF recombinant protein by incubating the column-purified HMG-TF proteins overnight at 4°C with the HRV 3C protease expressed and purified from *E. coli* BL21 (DE3). The mixture was purified again using the Nickel resin column (TakaRa) according to the manufacturer's instructions. The expression of purified recombinant proteins was confirmed through sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and western blotting.

**EMSA**

Specific double-stranded DNA sequences (CACTTCTAAACAAACAGCTGA) for SRY-binding activity were designed, stepwise annealed, and labeled with Digoxigenin (DIG) according to the DIG Gel Shift Kit, 2nd generation protocol (Roche, Mannheim, Germany). HMG-DNA binding reactions were performed according to the kit instructions. Briefly, for 10 µl reaction, 16 fmol of the labeled DNA and 0.5 µl purified HMG proteins were mixed with EMSA binding buffer, Poly [d(I-C)] as a non-specific competitor, and Poly L-lysine and incubated for 15 minutes at 25°C. Samples were then immediately cooled on ice to stop the reaction. The DNA-protein complex mixture was

separated on an 8% polyacrylamide gel for 80 minutes at 70 V in 5% TBE buffer. Samples were electro-transferred to a 5% TBE presoaked Amersham Hybond™-N+ nylon membrane (GE Healthcare UK Limited, United Kingdom). The membrane was then baked at 120°C for 15 minutes before treated with filter papers presoaked with 2× SSC solution. DNA fixation was performed following the crosslinking method using a UV transilluminator (Vilber Lourmat, France). Then, the membrane was incubated in blocking solution provided in the EMSA kit for 30 minutes at room temperature and further incubated with a 1/20,000-diluted DIG-POD (150 U Anti-Digoxigenin-POD Fab fragments) (Roche) antibody for 2 hours. Chemiluminescent detection was performed using the LAS-3000 Imager (Fujifilm, Tokyo, Japan) using a chemiluminescent agent, following the instrument's protocol.

**FA**

HPLC-grade fluorescent label oligonucleotides (CACTTCTAAACAAACAGCTGA) were synthesized to resemble those used in EMSAs with an additional fluorescent group, 6-carboxyfluorescein (6-FAM) bound to the 5′ end (Fasmac, Kanagawa, Japan), and dissolved in TEN buffer (10 mM Tris, 0.1 M NaCl, and 1 mM EDTA (pH 8.0)). Together with the complementary strand, the oligomers were annealed at 95°C for 5 minutes and cooled at room temperature. The DNA-protein complexes were obtained using premixed 50 nM annealed oligomers with HEPES binding buffer (40 mM HEPES, 40 mM KCI, 20% glycerol, 100 mg/ml BSA, pH 7.5) incubated at 37°C for 15 minutes, and the HMG protein was titrated from 10 μM with the total volume of the reaction at

1 ml at 25°C. To avoid non-specific binding, 10 μg/ml sheared salmon sperm DNA (BioDynamics Laboratory, Plainview, NY, USA) was added. Fluorescence anisotropy of serial titrations of HMG-DNA complexes (0.002 to 10 μM HMG) was performed after incubation using an FP-8500 fluorometer (JASCO, Tokyo, Japan). The measurements were taken at excitation and emission wavelengths of 495 and 515 nm, respectively, with spinning at 250 rpm. The average values from three replications were calculated using the anisotropy equation (Equation 1) (Nam et al., 2020). $K_d$ values were compared among samples.

$$Anisotropy\ (r) = \frac{I_{VV} - GI_{VH}}{I_{VV} - 2GI_{VH}} \quad (1)$$

where $I$ is the fluorescence intensity, the first character of the subscript is the polarization direction of the excitation light, and the second character is the direction of the emission. V and H indicate the vertical and horizontal directions, respectively, and $G$ is an instrument correction factor determined by excitation of the protein solution with horizontally polarized light. $K_d$ for the HMG-DNA complex was determined by using the Hill equation and the nonlinear least-squares method (IGOR Pro, Wavemetrics, Lake Oswego, OR, USA):

$$Anisotropy(r) = base + \frac{max - base}{1 + \left(\frac{K_d}{[Pt]}\right)^n} \quad (2)$$

where *base* is the baseline fluorescence anisotropy in the absence of binding, *max* is the maximum fluorescence anisotropy, [Pt] is the total protein concentration, and *n* is the Hill coefficient.

## 2.4 Results

### HMG with both S and A exhibited DNA-binding ability

To validate the presence and expression of recombinant SRY proteins of mouse and TMU, we performed sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and western blotting. The expression of recombinant proteins corresponding to the HMG proteins with the A-to-S or S-to-A substitution in a bacterial system was confirmed by Coomassie Blue staining after SDS-PAGE. The electrophoresis results showed clear bands between the 10 kDa and 15 kDa molecular weight markers at the predicted size of the HMG protein, which is approximately 12.9 kDa (Fig. 10A). Western blotting using an anti-SRY antibody showed single bands at positions consistent with those observed in the Coomassie Blue-stained gels (Fig. 10B). Together, these findings confirmed the validity of the recombinant HMG proteins for *in vitro* HMG-DNA binding analyses.

To explore the binding ability of the DNA-HMG domain with the A-to-S substitution, we employed electrophoretic mobility shift assays (EMSA) using digoxigenin (Dig) labeled oligonucleotide probes. We designed DNA probes to simulate the HMG-binding region on Enh13, a testis-specific *Sox9* enhancer sequence in the mouse genome (Gonen et al., 2018); this region contains the essential SRY/SOX9 binding motif (5′-A/T AACAA A/T-3′) (Ferrari et al., 1992; King and Weiss, 1993; Giese et al., 1994; Murphy et al., 2001; Phillips et al., 2004). We confirmed the binding motif sequence was conserved in the upstream region of *Sox9* in TMU (Fig. 9).

As shown in Fig. 11, the EMSA results revealed shifted bands corresponding to

the formation of DNA-protein complexes in all HMG samples derived from both mouse and TMU. The presence of these shifted bands in all Enh13 BS(+) binding complexes indicated that both A- and S-containing HMG domains were capable of binding to the DNA probes. Importantly, no bands were observed in the negative control lanes, Enh13 BS(-) with deletion of the SRY core binding motif, confirming the specificity of the interactions. The HMG-DNA binding intensity was comparable for the A or S variant within both mouse and TMU, suggesting that this specific amino acid substitution did not critically impair the HMG-DNA interaction. This finding aligned with computational predictions, which indicated that the A-to-S substitution might not play a critical role in modulating the DNA-binding activity of TMU SRY, at least when considering the HMG domain in isolation. The HMG-DNA complexes formed by TMU proteins, for both the A and S variants seems to exhibited weaker shifted-band intensity than those of the corresponding mouse HMG-DNA complexes by observation. However, to precisely confirm the change in binding affinity, an alternative measurement has been adopted and will be described later. This suggested that while TMU SRY retained the ability to bind DNA through its HMG domain, its binding affinity was lower than that of the mouse SRY protein.

**A-to-S substitution exhibited a slight reduction in DNA-binding affinity**

To complement the qualitative EMSA results, we performed fluorescence anisotropy assays to quantify the effects of the A-to-S substitution on DNA-binding affinity. By diluting the HMG proteins over a range of concentrations, we traced the extent of

complex formation from the anisotropy values. The highest anisotropy values observed at 10 μM HMG were 0.178 for the S-substituted mouse variant, 0.179 for both the S and A variants of TMU, and 0.181 for the wild-type mouse SRY, which indicates that all HMGs form a similar DNA complex at a certain DNA concentration.

As the dilution progressed and protein concentrations decreased, the anisotropy values declined gradually, until reaching approximately 0.067 at 0.0019 μM protein (Tables 1 and 2). An anisotropy plot revealed notable differences in binding affinity patterns between species and variants. For concentrations ranging between 0.01 μM and 1 μM, the reduction in fluorescence anisotropy initiates at higher concentration levels in the S mutants of mice and TMU, compared to the wild-type mouse SRY, indicating a decreased affinity for the target DNA. This trend was more pronounced in the S variant of TMU, suggesting a weaker interaction with DNA than that for the A variant (Fig. 12).

To quantify the binding affinities further, we calculated the dissociation constants ($K_d$) for each HMG-DNA complex. The calculated $K_d$ values revealed a gradient in binding affinity across the different variants. The lowest $K_d$ value, indicating the highest affinity, was observed for the wild-type mouse HMG-DNA complex ($K_d$ = 0.12 ± 0.02 μM) followed by the serine-substituted mouse HMG protein with a slightly higher $K_d$ of 0.14 ± 0.03 μM. For TMU HMG, both variants displayed higher $K_d$ values, with the A variant showing a $K_d$ of 0.16 ± 0.02 μM and the S variant exhibiting the weakest binding affinity with a $K_d$ of 0.22 ± 0.03 μM (Table 3). These findings supported the conclusion that in TMU the A-to-S substitution resulted in a slight

reduction in DNA-binding affinity, particularly in the TMU HMG protein. However, the A-to-S-substituted mouse HMG showed slightly different affinity within the margin of error compared to wild-type, which both still exhibited a higher binding affinity than the wild-type of TMU (S-variant). This observation suggests that the differences in binding complexes and affinities may be species-specific, influenced by other residues within the HMG domain and their interactions with neighboring amino acids.

## 2.5 Discussion

In this chapter, I investigated the DNA-binding properties of recombinant HMG proteins from mouse and TMU with the substituted versions of both. In addition to the *in silico* experiment validated in the Chapter I, the results provided more insights into the structural and functional characteristics of these proteins.

As a continuation to *in silico* results, we confirm the possibility that the protein could bind to the DNA, and further verified that the additional hydrogen bond in protein-protein folding structure found in HMG domain might actually give some effect to the DNA binding affinity. The binding patterns have been conserved in both mouse and TMU, with a slight difference in *Kd* value. The result aligned with the concept that the changes in a number of bonds result in changes in the pattern of hydrogen bond donors and acceptors presented by the DNA bases and the complementary groups in the protein (Luscombe et al., 2001), which largely determines the binding affinity through direct readout mechanisms.

Serine has been known as one of the popular amino acids compatible with phosphorylation. When the phosphate group of the phosphoprotein reacts with serine, threonine, or tyrosine, it results in the -OH group of their sidechain in an esterification reaction (Ghosh and Adams, 2011; Kalantari et al., 2015). The reversible phosphorylation of serine is well recognized as a switche for DNA-binding, playing a vital role in transcriptional activity. Because of that, substitutions to serine can have diverse effects on DNA-binding affinity, depending on the specific protein context and the nature of the substitution. The comparative study conducted by Lin and Guo (2019),

where substitution with serine or threonine residues possessing hydroxyl groups can strengthen protein-DNA interactions when appropriately positioned. In my study, on the other hand, the fluorescent anisotropy results show that the serine variant of both species exhibits slightly lower DNA binding affinity. I hypothesize that if SRY could be expressed normally, the binding activity and the target DNA recognition should not be a critical problem in A21S. Unfortunately, the expression patterns of TMU's SRY have yet to be confirmed in embryonic gonads, but at least both *in silico* and *in vitro* results align with or reflect the same concept as the previously mentioned study on the F109S mutation in SRY. In that study, almost no differences in DNA-binding activity were observed, though the mutation may have slightly altered in vivo activity. The resulting sex phenotype appeared to depend on genetic background or environmental factors (Jäger et al., 1992; Racca et al., 2016).

Nevertheless, this is the first documentation of the HMG-DNA binding activity in TMU that against the "supposed to be dysfunction due to A-to-S substitution in all SRY copies". To properly confirm this hypothesis, it is necessary to investigate the other function as a transcription factor of HMG-domain, the DNA bending activity, which would be validated on Chapter III.

## 2.6 Tables

**Table 1** Anisotropy ( *r* ) value observed in Mouse HMG-DNA complexes in HEPES buffer at each different concentration with Average binding affinity and standard deviation compared to Mouse with A-to-S substitution.

| | Mouse wt | | | | | Mouse A-to-S | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Protein(μM)** | **Anisotropy (r)** | | | **Avg** | **SD** | **Anisotropy (r)** | | | **Avg** | **SD** |
| 10 | 0.1788 | 0.1832 | 0.1821 | 0.18137 | 0.00187 | 0.1757 | 0.1788 | 0.1795 | 0.178 | 0.00165 |
| 5 | 0.1838 | 0.1836 | 0.1844 | 0.18393 | 0.00034 | 0.1766 | 0.1812 | 0.1804 | 0.1794 | 0.00201 |
| 2.5 | 0.1804 | 0.1803 | 0.1819 | 0.18087 | 0.00073 | 0.1737 | 0.1788 | 0.1778 | 0.17677 | 0.00221 |
| 1.25 | 0.1734 | 0.1753 | 0.1765 | 0.17507 | 0.00128 | 0.1712 | 0.174 | 0.1719 | 0.17237 | 0.00119 |
| 0.625 | 0.167 | 0.1679 | 0.169 | 0.16797 | 0.00082 | 0.1653 | 0.1681 | 0.1652 | 0.1662 | 0.00134 |
| 0.3125 | 0.1602 | 0.1624 | 0.1635 | 0.16203 | 0.00137 | 0.1524 | 0.1635 | 0.1612 | 0.15903 | 0.00478 |
| 0.1562 | 0.1208 | 0.142 | 0.1462 | 0.13633 | 0.01112 | 0.1103 | 0.1427 | 0.1229 | 0.1253 | 0.01334 |
| 0.078 | 0.0967 | 0.1127 | 0.1121 | 0.10717 | 0.00741 | 0.0886 | 0.1091 | 0.0991 | 0.09893 | 0.00837 |
| 0.039 | 0.0791 | 0.0905 | 0.091 | 0.08687 | 0.0055 | 0.0773 | 0.0891 | 0.0818 | 0.08273 | 0.00486 |
| 0.019 | 0.0719 | 0.0803 | 0.0776 | 0.0766 | 0.0035 | 0.0722 | 0.0788 | 0.0734 | 0.0748 | 0.00287 |
| 0.0019 | 0.0666 | 0.0697 | 0.0682 | 0.06817 | 0.00127 | 0.068 | 0.0686 | 0.0685 | 0.06837 | 0.00026 |

**Table 2** Anisotropy ( *r* ) value observed in TMU HMG-DNA complexes in HEPES buffer at each different concentration with Average binding affinity and standard deviation compared to TMU with S to A substitution.

| | TMU Wt | | | | | TMU S to A | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Protein (μM)** | **Anisotropy (r)** | | | **Avg** | **SD** | **Anisotropy (r)** | | | **Avg** | **SD** |
| 10 | 0.1811 | 0.1776 | 0.1792 | 0.1793 | 0.00143 | 0.1778 | 0.1794 | 0.1808 | 0.17933 | 0.00123 |
| 5 | 0.1805 | 0.1794 | 0.1797 | 0.17987 | 0.00046 | 0.1774 | 0.1806 | 0.1808 | 0.1796 | 0.00156 |
| 2.5 | 0.1793 | 0.1785 | 0.1786 | 0.1788 | 0.00036 | 0.1764 | 0.1791 | 0.1798 | 0.17843 | 0.00147 |
| 1.25 | 0.1781 | 0.1774 | 0.1774 | 0.17763 | 0.00033 | 0.1742 | 0.1765 | 0.1776 | 0.1761 | 0.00142 |
| 0.625 | 0.1723 | 0.1766 | 0.1743 | 0.1744 | 0.00176 | 0.1748 | 0.1755 | 0.1763 | 0.17553 | 0.00061 |
| 0.3125 | 0.1397 | 0.1535 | 0.1345 | 0.14257 | 0.00802 | 0.1602 | 0.1662 | 0.1551 | 0.1605 | 0.00454 |
| 0.1562 | 0.1028 | 0.113 | 0.0991 | 0.10497 | 0.00588 | 0.1182 | 0.1282 | 0.1116 | 0.11933 | 0.00682 |
| 0.078 | 0.0822 | 0.0873 | 0.0789 | 0.0828 | 0.00346 | 0.0913 | 0.0987 | 0.0873 | 0.09243 | 0.00472 |
| 0.039 | 0.0734 | 0.0746 | 0.0707 | 0.0729 | 0.00163 | 0.0771 | 0.0828 | 0.0741 | 0.078 | 0.00361 |
| 0.019 | 0.0708 | 0.0706 | 0.0686 | 0.07 | 0.00099 | 0.0711 | 0.0752 | 0.071 | 0.07243 | 0.00196 |
| 0.0019 | 0.0659 | 0.0667 | 0.0672 | 0.0666 | 0.00054 | 0.0648 | 0.0681 | 0.0673 | 0.06673 | 0.00141 |

**Table 3** *Kd* value for HMG-DNA complexes in Mouse WT, Mouse A-to-S, TMU WT, and TMU with S to A substitution.

| **Protein** | $K_d$ **(μM)** |
|---|---|
| Mouse WT (A) | 0.12 ± 0.02 |
| Mouse A-to-S | 0.14 ± 0.03 |
| TMU WT (S) | 0.22 ± 0.03* |
| TMU S to A | 0.16 ± 0.02 |

## 2.7 Figures

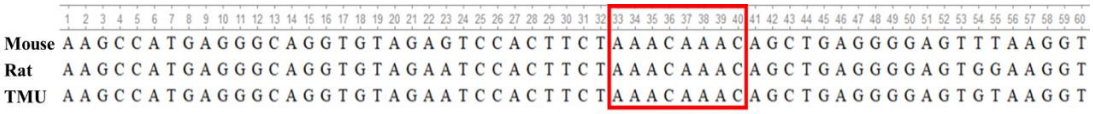


**Fig. 9** Sequence alignment of Enh13 shows strictly conservation of SRY DNA-binding site “AAACAAA” (red box) among rodent species including mouse, rat, and TMU.

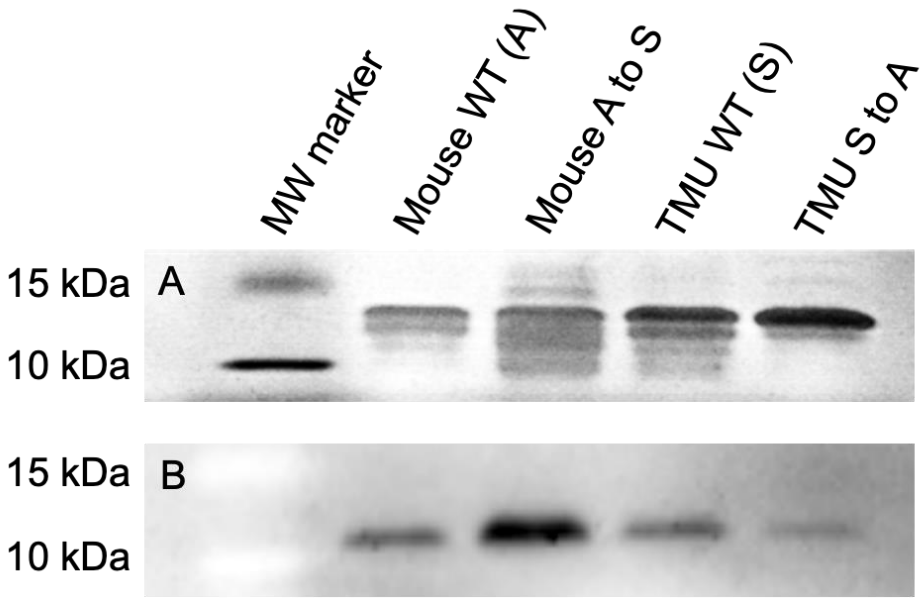


**Fig. 10** SDS-PAGE (A) and western blotting (B) results for recombinant HMG, the prediction size for recombinant protein is 12.9, which appeared between 10 and 15 kDa bands in the ladder (left).

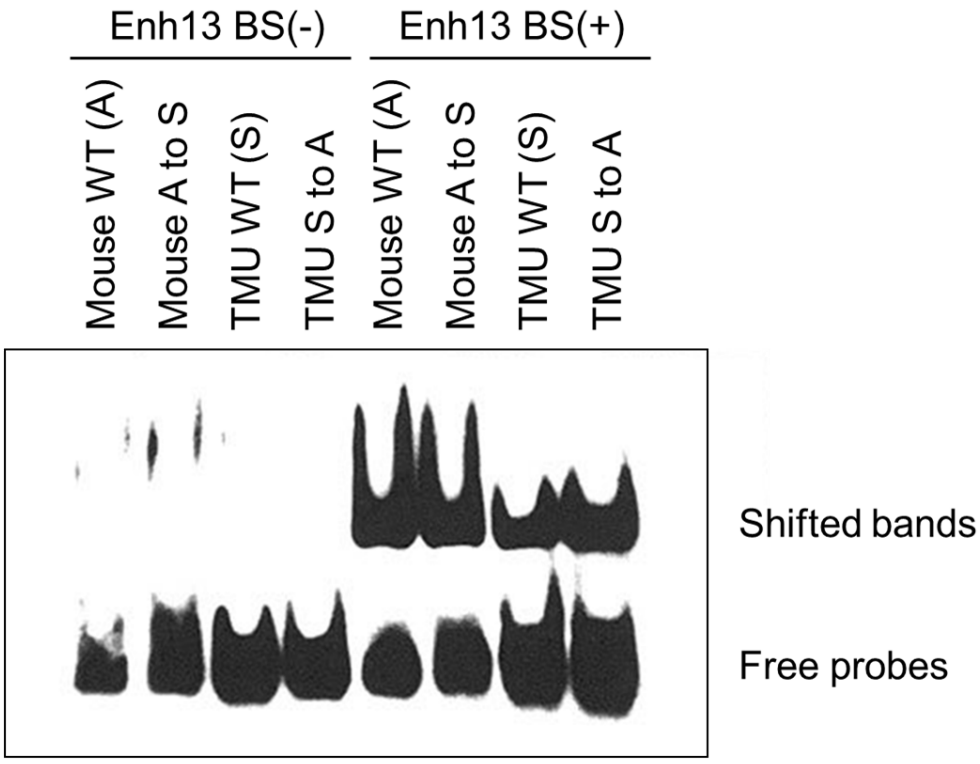


**Fig. 11** EMSA results for mouse WT (A), mouse A-to-S, TMU WT (S), and TMU S to A. The negative control, Enh13 BS (-), showed no band, unlike the HMG-Enh13 binding complex Enh13 BS (+).

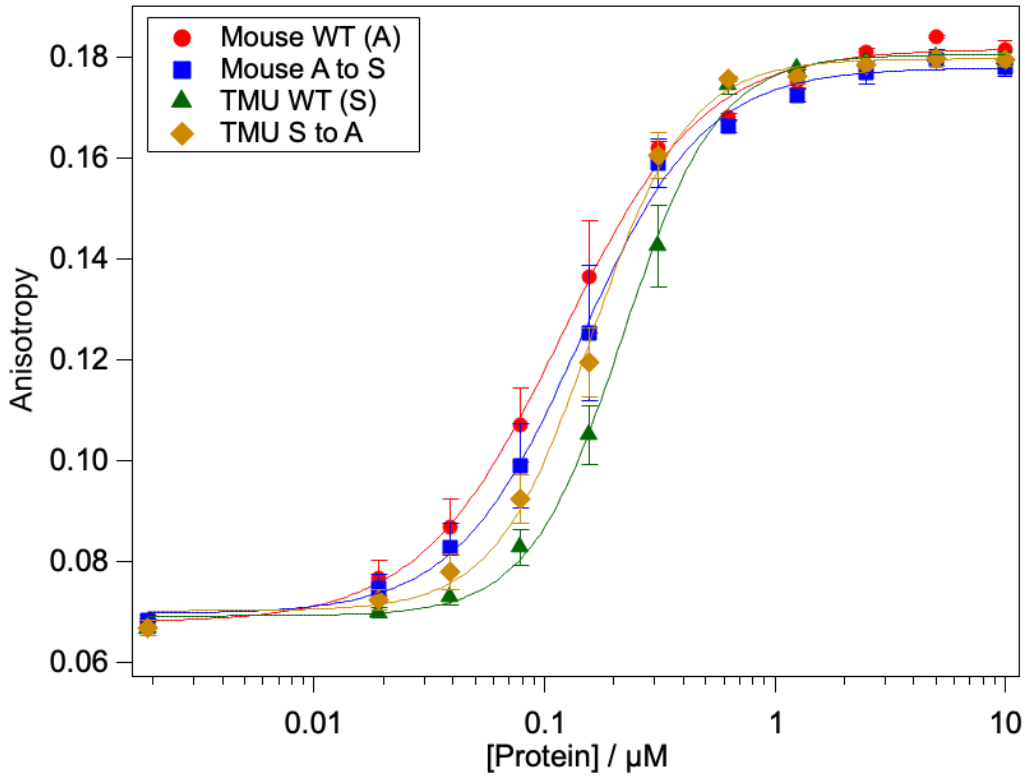


**Fig. 12** Fluorescence anisotropy results on a log scale for comparisons of binding affinity among mouse WT (A), mouse A-to-S, TMU WT (S), and TMU S to A with the reaction solution contained 0-10 µM of HMG protein. The fluorescence anisotropy was measured after exactly 15 minutes of incubation with excitation and emission at 495 and 515 nm, respectively. Error bars represent the standard deviation of three independent experiments. The solid lines represent fit to the Hill equation.

# Chapter III

# Analysis of the DNA-bending activity in the SRY

## 3.1 Abstract

DNA bending could be critical for functional proteins. The changes in protein structure due to some amino acid substitution proximal to those specific residues of the DNA-binding domain might not only affect the DNA binding ability, but also concomitantly reduce the transcriptional activity. In the last chapter, I discovered that the HMG domain of TMU's SRY, which employed an A-to-S substitution, still maintained the DNA binding ability, although with slightly reduced binding affinity. In this chapter, I performed Permutation Gel Electrophoresis (PGE) with the recombinant HMG protein from mouse and TMU and five labeled DNA to measure the DNA bending capacity. The results showed that both mouse and TMU exhibit similar DNA bending angles, 86.3° in Mouse, and 79.3° in TMU, indicating that TMU's SRY should work as sex determination properly if it could be expressed in the animals.

## 3.2 Introduction

The naked DNA's intrinsic bending reflects the natural flexibility or rigidity of the double helix strand. However, the pronounced bending induced by specific DNA-binding proteins, especially transcription factors, is the one that holds particular biological significance in the processing of genetic information (Privalov et al., 2009). Since many genes are regulated by enhancers, which could be located thousands of base pairs away from the promoter (the site where transcription begins), DNA bending by proteins or transcription factors brings these distant regions into close proximity. As a protein bends DNA, it causes looping that allows enhancers to contact the promoter, resulting in the spatial organization that facilitates recruitment of RNA polymerase II and other co-activators to the transcription start site, thereby initiating transcription (Van der Vliet and Verrijzer, 1993).

DNA bending, alongside DNA binding, is widely recognized as a critical feature of transcription factors (TFs) necessary for gene activation. This complex process involves a structural alteration of the DNA double helix from its standard linear conformation into a bent or kinked structure upon protein interaction. Such conformational changes can either promote or hinder the assembly of transcriptional machinery, thereby directly affecting the efficiency and outcome of gene expression (Van der Vliet and Verrijzer, 1993). DNA bending facilitates the spatial arrangement of regulatory elements, aiding in the formation of higher-order nucleoprotein complexes required for transcription initiation. Mechanistically, DNA bending primarily occurs through two well-characterized methods, which are the asymmetric

neutralization of DNA phosphate groups and the insertion of protein side chains between base pairs, commonly referred to as side chain intercalation (Privalov et al., 2009). These two modes represent fundamental strategies by which proteins physically deform the DNA double helix to achieve regulatory effects.

For asymmetric phosphate neutralization, proteins with positively charged surfaces could neutralize the phosphates on one face of the DNA helix and reduce the repulsive forces on that side, causing an unbalanced compression force that deforms the DNA toward the protein (Strauss and Maher, 1994). This hypothesis has later been tested with the studies on nucleosomal DNA, where histone atoms asymmetrically shield phosphate groups along the DNA helix, leaving one side of the DNA more negatively charged and bent toward the histone-occupied side due to a change in electrostatic repulsion (Mirzabekov and Rich,1997). Additional proof is that when the catabolite activator protein (CAP) in *Escherichia coli* was chemically neutralized using methyl-phosphonate substitution, the results showed a relative increase in DNA curvature of about 23% compared to the unmodified CAP–DNA half-site complexes (Hardwidge et al., 2002). Along with these experimental confirmations, computational simulations have also provided additional insights into this phenomenon. By all atom energy simulations of DNA dodecamers with selectively neutralized phosphate groups, DNA structure induced by the binding of positively charged proteins and other tightly associated cationic species, demonstrated significant bending toward the neutralized face (Kosikov et al., 2002). A key aspect of DNA binding and bending via asymmetric neutralization of phosphate charges valid by the entropy-driven,

neutralizing the phosphates displaces bound counterions, and their release into the surrounding solution increases entropy, providing a favorable thermodynamic contribution to the interaction (Manning, 1978; Manning, 2003).

The other prevalent mechanism, where proteins achieve DNA deformation, is through the side chain intercalation (Werner et al., 1996). The insertion of specific amino acid side chains between DNA base pairs has been observed in the structure studies of protein-DNA complexes (Kim et al., 1993; Schumacher et al., 1994; Werner et al., 1995; Love et al., 1995). The insertion of protein side chains into the DNA helix primarily targets the major groove (sometimes minor), disrupting the standard base stacking interactions when the protein is bound to the DNA (Chakraborty et al., 2017), often combined with hydrogen and covalent bonding (Reha et al., 2002). Completed intercalation refers to a case where the side chain is fully stacked over a base pair, penetrating to the helix axis similar to what is found in DNA-binding drugs (Roh et al., 2009). Partial intercalation refers to the situation in which the intruding side chain unstacks two adjacent base pairs but does not itself stack over them (Das et al., 2020**)**. In gene regulation, intercalation of one or more amino acids to the minor groove of DNA has been observed for several proteins involved with transcriptional activities. For example, *Escherichia coli* purine repressor protein PurR induces DNA bending via the intercalation of two L56 residues (Nagadoi et al., 1993), The TATA box-binding protein TBP in *Arabidopsis thaliana*, where molecular dynamics simulations of the TBP–DNA complex showed cohesion in partial intercalation of two Phe residues which seems to stabilize the kinked and opened-up DNA conformation

(Mondal et al., 2014), or in human oncogene product ETS1, where true intercalation of the side chain of Trp-28 exhibited approximately 60 degrees, widening of the minor of the DNA (Werner et al., 1995). The studies all indicated how important intercalation is to protein-DNA bending, as the bending angle is critical, as each protein exhibits a unique topology, with distinct structural regions engaging in DNA interactions (Werner et al., 1996**)**. For example, basic leucine zipper (bZIP) domains of the Fos and Jun proteins demonstrated that exchanging single amino acid residues near the basic regions influenced DNA bending in direct proportion to the change in net charge for all heterodimeric combinations between these proteins (Leonard et al., 1997). Studies have shown the impact of amino acid mutations on DNA bending and the function of genes that are regulated. However, the relationship between induced bent angles and the energetic characteristics of protein-DNA complex remains unclear. Fortunately, for the HMG domain, the thermodynamic and enthalpy energy information are currently available as almost complete dataset, make it easier for interpretation (Privalov et al., 2009).

In SRY, the functional HMG domain consists of three α-helices that form an L-shaped structure, recognizes the minor-groove of DNA binding site (King and Weiss, 1993) in contrast to many other TF families like zinc fingers (McBryant et al., 1996), and Basic Helix-Loop-Helix (bHLH) Proteins (Slattery et al., 2014). By partial intercalation of a nonpolar side chain (isoleucine) upon binding, HMG disrupts base stacking but not base pairing of the insertion site (King and Weiss, 1993). Structural studies of the *Drosophila* HMG-D protein reveal the DNA bending angles of ~111°,

which are critical for enabling the formation of nucleoprotein complexes (Murphy et al., 1999). On the other hand, validation by permutation gel electrophoresis (PGE), the HMG-protein complexed typically bent the DNA, employed sharp twist with angles >70° (Pontiggia et al., 1994). The alternative method where the binding complexes observed by using NMR of SRY- *MIS,* showed that *MIS* promoter DNA (synergistically activated by the WT1, SRY box protein 9 (SOX9), and steroidogenic factor 1 (SF1) (Hossain and Saunders, 2003) has changed it profound structure. These included an inter-base helical twist of approximately 26° ± 6°, resulting in an overall DNA bending angle of about 70° to 80° in the HMG–DNA complex (Bewley et al., 1998).

A recent study provided deeper insights into how specific amino acid substitutions within the HMG domain affect DNA bending angles. For instance, mutations such as N32, S36, and Y74 substituted with alanine did not significantly alter DNA bending, maintaining angles around 70–73°, which are comparable to the wild-type value of approximately 79°. In contrast, substitutions at R7, N10, or R20 resulted in a substantial reduction in DNA bending angles, ranging from 53° to 65°, indicating a marked loss in bending efficiency (Racca et al., 2024). These changes are particularly important because the DNA bending angle plays a critical role in transcriptional activation. A sharp bend is thought to provide kinetic stability to the SRY–DNA complex, facilitating effective transcription initiation. Therefore, even moderate alterations in bending angle may lead to significant changes in transcriptional activity, which might be able to alter overall transcriptional activity

dramatically (by several fold) (Ukiyama et al., 2001; Li et al., 2006), which may contribute to developmental disorders such as de novo sex reversal (Pontiggia et al., 1994).

In this chapter, I performed DNA-bending ability assay of TMU's HMG protein *in vitro* through gel permutation assays using Polyacrylamide Gel Electrophoresis (PAGE). By using different locations of the SRY core binding site on the set of labeled DNA probes, the DNA bending properties of each complex could be determined. By comparing with the bending angle between mouse WT and TMU WT, it is possible to investigate my hypothesis that SRY of TMU might still retain its sex determining function or not. The results as well provided more information and interpret the effects of the amino acid substitution from A-to-S to DNA bending ability of this critically endangered rodent species.

## 3.3 Materials and Methods

### Plasmid design

For PGE analysis, the plasmid was designed by using template from pbendAT, a DNA segment containing five pairs of restriction endonuclease recognition sites which can produce five DNA fragments of identical length, either through cleavage by restriction enzymes or via PCR (Leng, 2013). A 239 bp DNA segment containing the SRY core binding site 'TAAACAAA' inserted in the middle was synthesized (Fig. 13) and obtained from Fasmac (Kanagawa, Japan) in conjugation with the plasmid, mutated pUC19 with Amp resistance. The plasmid then amplified by transformation into the JM109 *Escherichia coli* strain (Nippon Gene Ecos) according to the manufacturer's instructions. After colony selection using antibiotics (50 μg/ml ampicillin in 1.5% LB agar media) and plasmid extraction using the FastGene DNA Extraction Kit (Nippon Genetics), inserted DNA was confirmed again using the Colony PCR using SimpliAmp™ Thermal Cycler (Thermo Fisher, Japan) and KAPA Taq PCR Kit (Nippon Genetics).

### DNA labeling

To obtain five pairs of nucleotide probes with identical lengths but different binding positions, PCR was performed on the synthesized plasmid using five pairs of specific primers included 1F (5'-TCGTCGTCCTGGCGGTAC-3'), 1R (5'-GCGGCCGGTCGACGTTTG-3'), 2F(5'-ACCACGCTATCTGTGCAA-3'), 2R(5'-

TGCCTGTTGACCTGGTGC-3'), 3F(5'-AGCCACCGGACACGTGCT-3'), 3R(5'-AGGTCGCATGAAGAGGCC-3'), 4F(5'-TCCGAGTGTCCACCACCG-3'), 4R(5'-TCTGCGATGCTGAAGGTC-3), 5F(5'-ACCGAGGTGAAGGAGCTC-3'), 5R(5'-TCCAGAGGGACAGATCTG-3') (Fasmac, Kanagawa, Japan) by using SimpliAmp™ Thermal Cycler (Thermo Fisher, Japan) and KAPA Taq PCR Kit (Nippon Genetics). Around 120 bps, 5 identical length PCR products were electrophoresed using 2% agarose gels stained with MIDORI Green (Nippon Genetics, Tokyo, Japan), and purified using the NucleoSpin® Gel and PCR Clean-up Kit (MACHEREY-NAGEL GmbH & Co, KG, Düren, Germany). The DNA concentration have been measured, and titrated using nanodrop 8000 spectrophotometer (Thermo Scientific, Germany). The products were then labeled with Digoxigenin (DIG) according to the DIG Gel Shift Kit, 2nd Generation protocol (Roche, Mannheim, Germany), and stored at –30 °C until further used.

**Permutation Gel Electrophoresis (PGE)**

DIG Gel Shift Kit, 2nd generation protocol (Roche, Mannheim, Germany), was again used in this analysis. HMG-DNA binding reactions were performed according to the kit instructions. Five labeled DNA probes and purified HMG proteins, similar to those used in the previous chapter, were individually mixed with EMSA binding buffer, Poly [d(I-C)] as a non-specific competitor, and Poly L-lysine and incubated for 15 minutes at 25°C. Samples were then immediately cooled on ice to stop the reaction. The DNA-protein complex mixture was separated on an 8% polyacrylamide gel for 120 minutes

at 70 V in 5% TBE buffer. Samples were electro-transferred to a 5% TBE presoaked Amersham Hybond™-N+ nylon membrane (GE Healthcare UK Limited, United Kingdom). The membrane was then baked at 120°C for 15 minutes before treated with filter papers presoaked with 2× SSC solution. DNA fixation was performed following the crosslinking method using a UV transilluminator (Vilber Lourmat, France). Then, the membrane was incubated in blocking solution provided in the EMSA kit for 30 minutes at room temperature and further incubated with a 1/20,000-diluted DIG-POD (150 U Anti-Digoxigenin-POD Fab fragments) (Roche) antibody for 2 hours. Chemiluminescent detection was performed using the LAS-3000 Imager (Fujifilm, Tokyo, Japan) using a chemiluminescent agent, following the instrument's protocol.

**HMG-DNA bending angles calculation**

Claude 3.7 Sonnet, the upgraded version of Claude 3.5 Sonnet (released on October 22, 2024), was used to visualize and analyze the relative distance migration of the shifted bands. This model had undergone a joint pre-deployment evaluation by the UK Artificial Intelligence Safety Institute (UK AISI) and the U.S. Artificial Intelligence Safety Institute (US AISI). The relative mobility has been calculated by applied Identified Minimum and Maximum Mobility Values with the equation 3 described by Thompson and Landy (1988), and Kim et al. (1989):

$$\mathbf{RL} = \frac{\mu \mathbf{M}}{\mathbf{E}} \qquad (\mathbf{3})$$

where RL is the relative mobility values in the binding complexes, μM is the minimum mobility, and μE as the Maximum mobility.

The bending angle then has been calculated by equation 4 described by Thompson and Landy (1988), and Kim et al. (1989), which could be reconstructed as equation 5:

$$\mathbf{RL} = cos\frac{a}{2} \qquad (4)$$

$$a = 2arccos(\mathbf{RL}) \qquad (5)$$

Where *a* is the DNA bending angles in the binding complexes, and RL is the relative mobility values. The simplified image showing changes in the DNA bending angle was generated using Python programming in JupyterLab. The original code used is described in Fig. 13.

## 3.4 Results

**Both mouse and TMU show "functionable" bending angles.**

To further validate the bending ability of the DNA-HMG domain with the A-to-S substitution, we did the permutation gel assay by again employing electrophoretic mobility shift assays (EMSA) using digoxigenin (Dig) labeled oligonucleotide probes. Five pairs of probes, each approximately 120 base pairs in length but differing in the location of the core binding sequence, were generated (Fig. 14) by PCR amplification of the synthesized plasmid using specifically designed primers

As shown in (Fig. 15A-B), the PGE results revealed shifted bands corresponding to the formation of DNA-protein complexes with different migration of the shifted bands, indicating that the location of core binding sequences mattered to the migration speed of HMG-DNA in both mouse and TMU. The shifted pattern and the bands migratory formation as the multiple-shifted bands appeared only in Mouse HMG-DNA complex but not in TMU, indicating that the binding affinity of mouse SRY to the core DNA binding sequence is by far higher than TMU, as suggested in the Chapter II. With the assistance of bar-marks draw at the bottom line of each band (Fig. 15C-D), the observed migration bands and bending pattern in both samples are resemble each other, where the visualize detected mobility of mouse shifted bands are 84, 73, 62, 74, 85 (arbitrary units) respectively, while in TMU are 87, 76, 66, 75, and 86 by using Claude 3.7 Sonnet. The calculated relative mobility by applied equation 3 to the mobility shifted band revealed that Mouse have 0.7176 RL, which is slightly lower when compared to TMU as 0.7586 RL (Table 4). The calculated DNA bending

angles of HMG-DNA complexed by apply equation 4 and 5 indicated that the DNA bending angle in mouse is 88.3°, which is considerably higher than TMU which showed as 81.3° (Table 4). The results reveal that mouse HMG domain retained more sharp-bending angle than TMU (Fig. 16).

This is the first report to demonstrate that both Mouse and TMU HMG proteins retain their DNA sharp-bending ability. Although some interspecies differences are present, both proteins maintain bending angles above 70°, as expected for functional SRY. These findings suggest that if SRY is properly expressed in the embryonic gonads of TMU, it may be capable of determining sex in this multiple-SRY rodent species.

## 3.5 Discussion

In Chapter II, we have known that the A-to-S substitution might not actually give a critical impact on SRY function on TMU, since it at least maintains the DNA binding ability to some extent. The substitution of an amino acid from A-to-S at the 21st position of the HMG domain slightly reduced the DNA binding affinity, but has been proven to be insignificant in both *in silico* and *in vitro* experiments. In this chapter, I confirmed this theory as the results suggested that the DNA bending angle of both mouse and TMU's HMG domain still maintained its sharp >70° bent angles, which is necessary for SRY to activate its downstream gene as described by Bewley et al. (1998).

Actually, the bending angle reported by the previous PGE assay could be vary from times to times depending on gel system, labeling methods, and the length of probes they used, which could be somewhat lower from 78.8° ± 2.1 (Racca et al., 2024) to 86° ± 2 (Chen et al., 2022) or more in WT of human HMG domain. In this chapter, the approximate bending angle of the Mouse WT HMG is 88.3°, and 81.3° in TMU, both of which fall within the acceptable "bending-capable" range. A comparable decrease of approximately 13° (or 15° as visualized by NMR) associated with XY gonadal dysgenesis were reported before in human (Pontiggia et al., 1994; Phillips et al., 2004). However, in this study, less than 10° is shown in the difference between mouse and TMU WT bending angle, suggesting that there's merely some difference between species traits.

It is important to note that determining whether differences in DNA bending angles between two species are statistically or biologically significant remains challenging, especially when experiments are performed using different gel systems. The calibration of PGE results is not absolute and is heavily dependent on the model and experimental parameters used, which can sometimes lead to overestimation or underestimation of the bending angle. For example, some studies have observed that the mouse SRY–DNA complex exhibits sharper bending than the human SRY–DNA complex, with differences reported at approximately 25–30° (Giese et al., 1994). Contrarily, other studies suggest the opposite, indicating that the human complex bends more acutely by about 6–10° (Phillips et al., 2004). Nonetheless, both results fall within the novelty expected range of significant DNA bending, which is considered to be typically around 70° or more as a certain degree required to facilitate transcriptional regulation through DNA looping.

Previous research involving transgenic mice has provided further insight into the functional relevance of interspecies differences in DNA bending. In these studies, some attempts aimed to replace the entire HMG domain of mouse SRY with that of other HMG-box proteins such as SOX3, SOX9, or human SRY, each containing multiple amino acid differences and producing distinct DNA bending angle profiles. Surprisingly, the results showed that these replacements did not disrupt SRY-mediated male sex determination (Bergstrom et al., 2000; Lovell-Badge et al., 2002), suggesting a degree of functional flexibility. The observed 9° difference in DNA bending between the Mouse and TMU HMG domains showed in the context proved

that both HMG domain might be replaceable and still be functionally tolerated. This aligns with the findings of Ogata et al. (2019), who proposed that male developmental failure in TMU may be attributed more to the absence of a long glutamine-rich (Q-rich) domain compared to that in mouse, which results in impaired protein stability, rather than DNA bending capacity alone.

To more precisely assess the impact of DNA bending, future experiments should include additional controls, such as performing the DNA-bending assay using recombinant mouse HMG protein carrying the A-to-S amino acid substitution. Validating the DNA bending angle within the same species might offer a more accurate measure of how a single amino acid change affects the DNA-bending activity of SRY. In contrast, cross-species comparisons introduce variability from multiple sequence differences, making it difficult to isolate the effect of a single substitution. By using an intra-species control, the influence of the A-to-S substitution can be examined without confounding factors, allowing for a clearer interpretation of its structural and functional significance in SRY-mediated gene regulation.

Nevertheless, the currently validated experimental results gave confidence to the previous study in both Chapters I and II where the results suggested that for TMU, if in any case SRY could be expressed properly, the A-to-S substitution might not actually be critical for its SRY protein as a transcriptional factor in determining sex. This left the fact behind the reason for the availability of multiple SRY copies in TMU, which would become an interesting topic for further investigation in the future.

## 3.6 Table

**Table 4** Band migrations (arbitrary units), Relative mobility, and calculated DNA bending angles of HMG-DNA complexes in Mouse WT in compared with TMU WT.

| | Bands migration (lane) | | | | | Relative mobility | DNA-bending angles |
|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | | |
| Mouse WT | 85 | 72 | 61 | 73 | 86 | 0.7176 | 88.3° |
| TMU WT | 87 | 76 | 66 | 75 | 86 | 0.7586 | 81.3° |

## 3.7 Figures

```python
import numpy as np
import matplotlib.pyplot as plt

def rotate(coords, angle, axis, center):
    axis = axis / np.linalg.norm(axis)
    theta = np.radians(angle)
    a, b, c, d = np.cos(theta/2), *(-axis*np.sin(theta/2))
    R = np.array([
        [a*a + b*b - c*c - d*d, 2*(b*c - a*d),     2*(b*d + a*c)],
        [2*(b*c + a*d),         a*a + c*c - b*b - d*d, 2*(c*d - a*b)],
        [2*(b*d - a*c),         2*(c*d + a*b),     a*a + d*d - b*b - c*c]
    ])
    return np.array([R @ (p - center) + center if p[2] > center[2] else p for p in coords])

def make_dna(bp):
    z = np.linspace(0, 3.4*(bp-1), bp)
    angle = 2 * np.pi * z / (10.5 * 3.4)
    x, y = np.cos(angle), np.sin(angle)
    return np.column_stack((x, y, z)), np.column_stack((-x, -y, z))

def plot_dna(ax, s1, s2, c_main, c_pair, label):
    ax.plot(*s1.T, c=c_main, lw=3, label=label)
    ax.plot(*s2.T, c=c_main, lw=3)
    for a, b in zip(s1, s2):
        ax.plot(*np.array([a, b]).T, c=c_pair, lw=1.8)

bp = 16
s1, s2 = make_dna(bp)
center = np.array([0, 0, s1[bp//2, 2]])

fig = plt.figure(figsize=(9,7))
ax = fig.add_subplot(projection='3d')

plot_dna(ax, rotate(s1, 88.3, [1,0,0], center), rotate(s2, 88.3, [1,0,0], center), 'hotpink', 'pink', 'Mouse WT (88.3°)')
plot_dna(ax, rotate(s1, 81.3, [1,0,0], center), rotate(s2, 81.3, [1,0,0], center), 'limegreen', 'lightgreen', 'TMU WT (81.3°)')

ax.set_box_aspect([1,1,2])
ax.legend()
plt.tight_layout()
plt.show()
```

**Fig. 13** Original code used to create a 3D image showing the changes in DNA bending angle. With X, Y, and Z axis, the proper 3D rotatable figure has been generated.

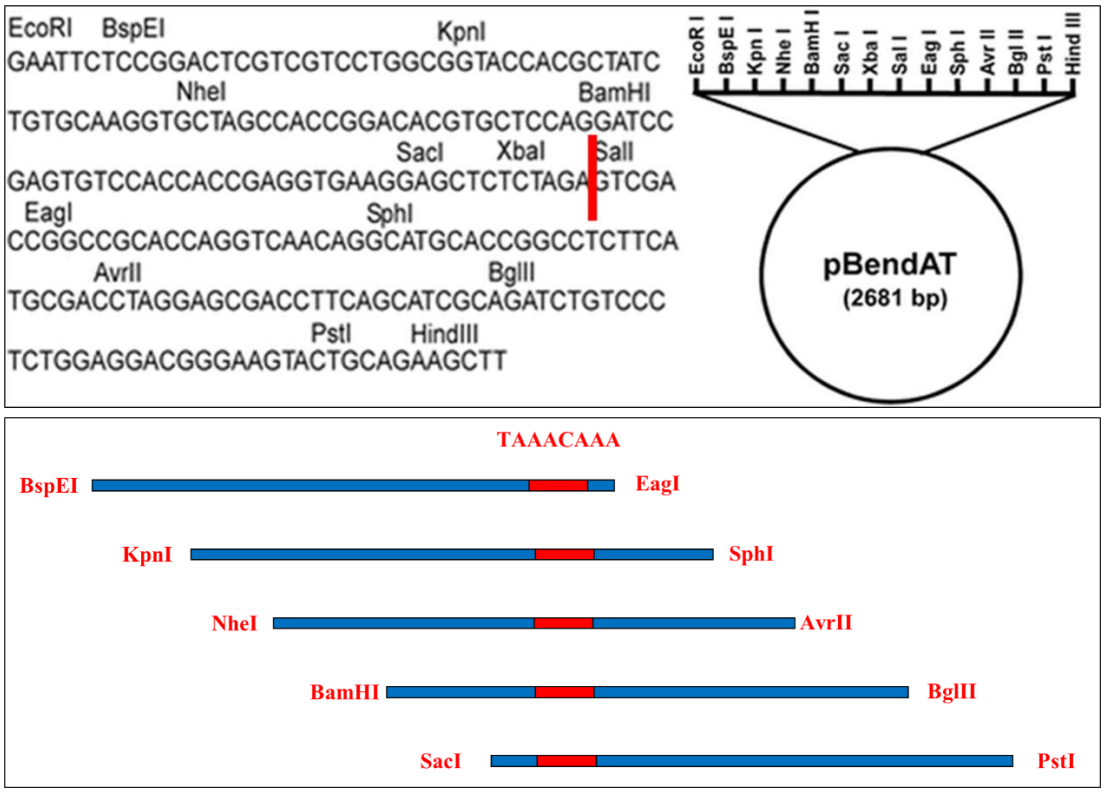


**Fig. 14** Upper, the construction of pBendAT-SRY core, the SRY core binding sequence has been inserted between Xbal and Sall restriction enzyme sites (red box) in the original insertion sequences of pBendAT plasmid designed by Leng (2013). Lower, the amplified double strand DNA for PGE assay which included 5 identical length of DNA, but different core binding position (red boxes) for SRY.

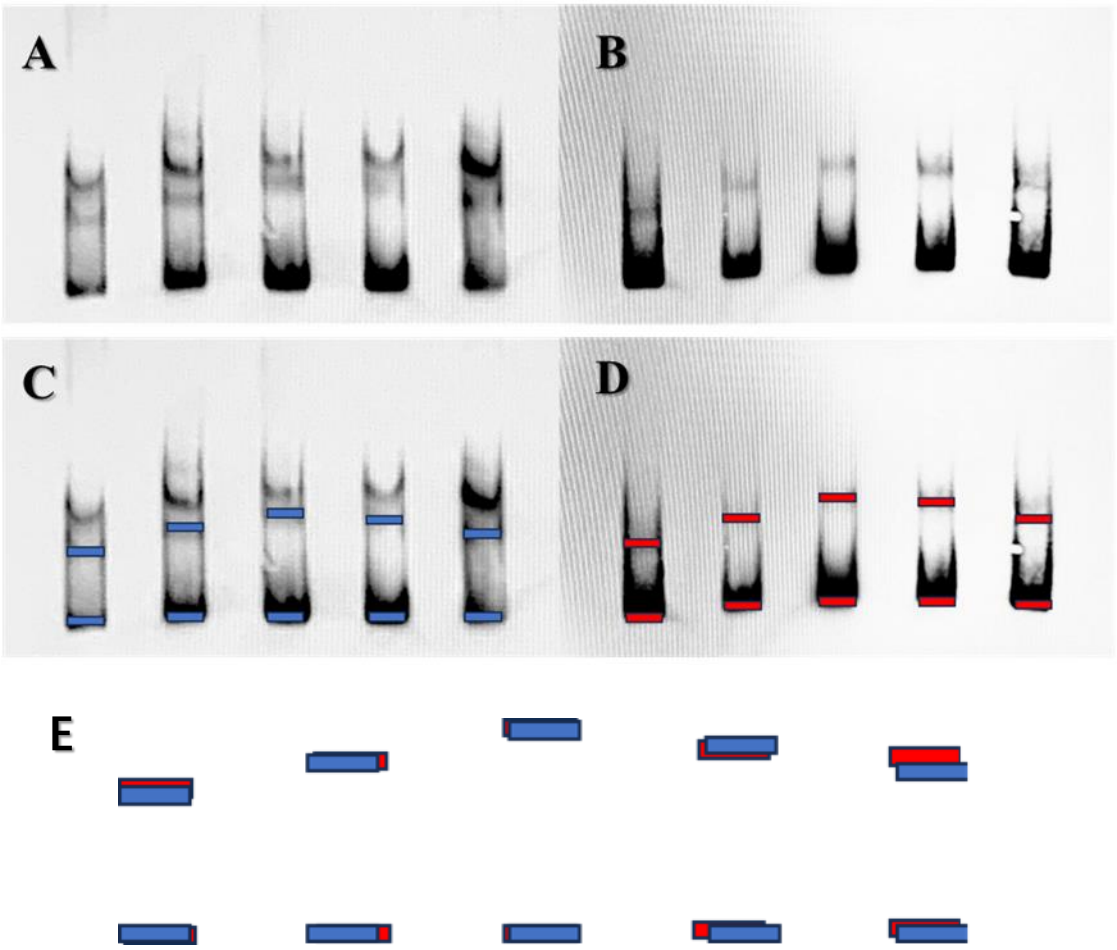


**Fig. 15** The PGE results for mouse WT (A and C), and TMU WT (B and D), The assist bands for image analysis have been showed in blue and red box (C and D), the migrated bands appear not that much difference between both sample with the adjustment of the free probes (E).

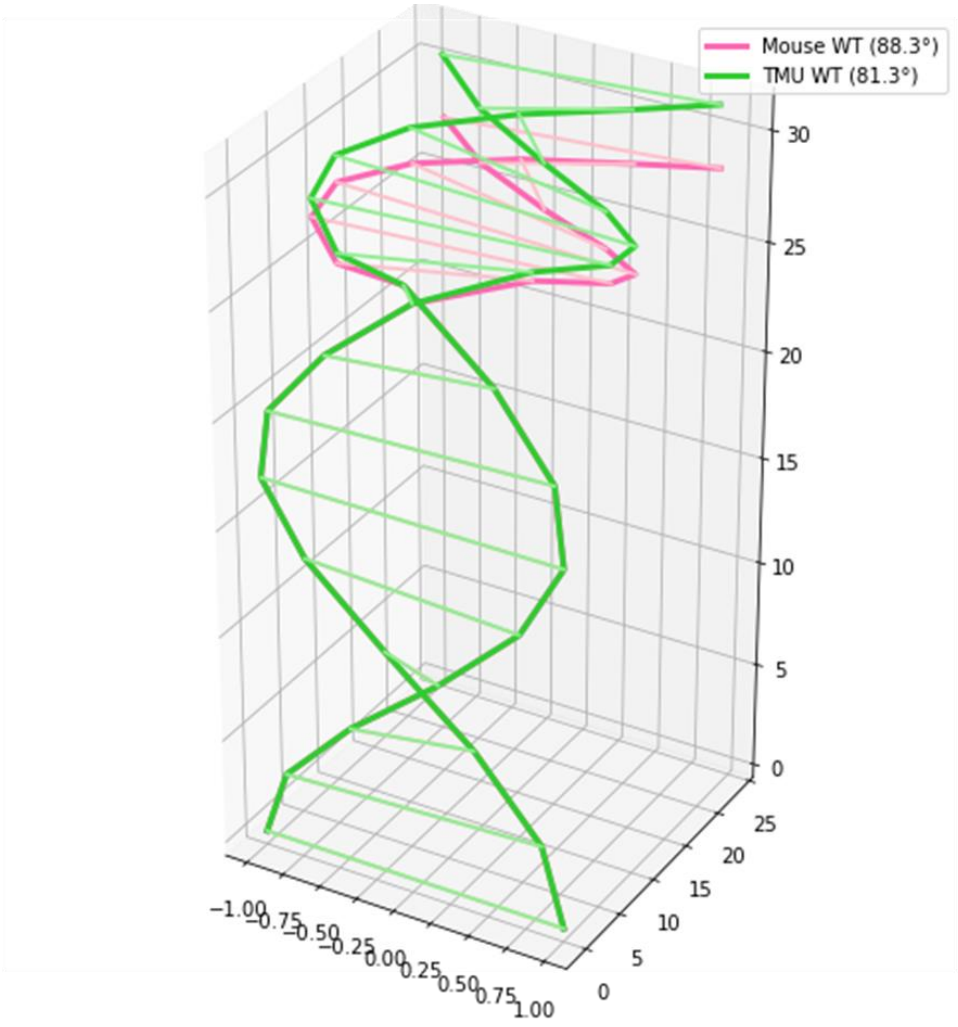


**Fig. 16** The simplified image shows the estimated differences in DNA bending angles between the Mouse (pink) and TMU (green) HMG proteins. Both variants still exhibit a sharp bend of more than 80°.

# General discussion

This study explores a compelling and important question in the field of molecular genetics, biochemistry, and evolutionary biology about how resilient the sex-determining gene, SRY, is to mutations in its highly conserved HMG domain. It is well known that alterations or changes in the HMG region of SRY typically affect its central role in sex determination, often resulting in profound phenotypic consequences such as sex reversal, infertility (Kellermayer et al., 2005), and high risk of developing gonadoblastoma (Shahid et al., 2004). The study focused on A-to-S substitution in the SRY gene of TMU, a species already recognized for its unique XY sex determination system and evolutionary divergence from other *Tokudaia* species. It is truly remarkable that all male TMU are known to carry the A-to-S substitution, yet the population with normal fertile individuals exist. Our study, step by step, investigated a suspected species-specific mutation in a vital HMG domain.

SRY typically functions via two mechanisms, DNA-binding and DNA bending, both of which are critical for *Sox9* activation and lead to the promotion of testis development from bipotential gonads (Weiss, 2005). As a transcription factor, SRY binds to the target DNA via the HMG domain and recognizes the minor groove containing AACAA (Murphy et al., 2001). In human SRY, 10 hydrogen bonds between the DNA and amino acid have been confirmed in R7, N10, T74 on one side with I13 as an axle, and the other side with R20, N32, and S36, corresponding to R4, N7, T71, I10, R17, N29, and S33, respectively (Racca et al., 2024), in the HMG domain. All of these positions are highly conserved among mammalian species (Fig. 1-1, 1-2, 1-3).

In Chapter I**,** results confirmed the A-to-S substitution at the 21st position of the HMG domain only in TMU. The A at this position is highly conserved across mammals and was detected in 228 SRY sequences from 225 species evaluated in this study; this high conservation is, however, by exception, a species-specific characteristic in TMU. Computational insights evaluation of the mutated SRY protein reveals that the A-to-S mutation does not significantly distort the tertiary structure of the protein by homology modeling. This is crucial because the HMG domain relies on specific folding patterns to insert into the minor groove of DNA and facilitate bending (Racca et al., 2024). With molecular docking, and dynamic contact map simulations, key residues responsible for DNA contact were conserved in their spatial orientation, although the additional hydrogen bonds were found between protein-protein molecules at R17, one of the important DNA binding molecules described by Racca et al. (2024). However, the docking results suggest that the mutation does not obstruct the ability of SRY to recognize its DNA target sequence. The ability to retain DNA-binding potential, even with an altered amino acid residue, suggests that the HMG domain might possess structural plasticity. This opens up fascinating implications for evolutionary biology, as such plasticity could permit species to adapt their sex-determination mechanisms in response to chromosomal changes or environmental pressures.

To validate the computational findings, Chapter II transitions into biochemical experimentation using recombinant protein techniques. In addition to the *in silico* results, the bacterial-expressed recombinant HMG of TMU retained its DNA-binding ability. A shifted band of the HMG–DNA complex and a clear binding affinity curve

were observed in TMU wild-type (WT), although it showed the weakest binding among the tested groups compared to modified TMU (S-to-A), mouse A-to-S, and mouse WT. This indicates that if TMU can properly express its multiple SRY copies, the expressed SRY proteins are still capable of binding to the DNA target.

In Chapter III, the results additionally confirmed that the bending of DNA, a core requirement for transcription factor integral to the initiation of transcriptional events, was effectively preserved. The HMG domain induces the sharp bending of DNA, facilitating chromatin remodeling for gene regulation through the hydrophobic intercalation of amino acid residues into the DNA (Racca et al., 2024). This sharp bend may provide kinetic stability for SRY-DNA transcriptional activation, and a change in the angle caused by an amino acid substitution is expected to alter overall transcriptional activity dramatically (by several fold) (Ukiyama et al., 2001; Li et al., 2006), potentially leading to de novo sex reversal (Pontiggia et al., 1994). In this finding, with less than 10° bending angles different, almost similar to those previously reported “switchable” mouse and human SRY (Bergstrom et al., 2000; Lovell-Badge et al., 2002) define that A-to-S is more like a species’s specific character than a vital mutation. These findings are the first to collectively support the idea that the A-to-S mutation is functionally neutral, at least under laboratory conditions, against the previous studies that suggested that A-to-S might cause the SRY gene in TMU to be dysfunctional (Murata et al., 2010).

The Okinawa spiny rat represents a unique model organism for studying the fluidity of sex-determination systems. Prior studies have noted that other *Tokudaia*

species, like the Tokunoshima spiny rat (*T. tokunoshimensis*) and the Amami spiny rat (*T. osimensis*) have lost their Y chromosome entirely, and might instead rely on autosomal mechanisms for determining sex. The retention of SRY functionality in a species that still carries a Y chromosome but harbors a point mutation within the HMG domain suggests a transitional phase in evolutionary adaptation. One possibility is that this mutation is part of a larger genomic reorganization aimed at phasing out dependency on SRY, like that Y and W chromosome losses foreshadow events (Charlesworth and Charlesworth, 2020; Lenormand et al., 2020**)**. Alternatively, it might be a benign mutation that reflects natural variability within a small population. The ability of a critical gene to tolerate mutations without losing functionality indicates a potential mechanism for evolutionary resilience. I hypothesize that the reduction in binding affinity of TMU might actually subsidize the effect by their multiple intact SRY copies to balance the dosage compensation (Birchler et al., 2022**)**.

In this finding, a key strength lies in its dual approach, integrating computational predictions with empirical validation, which enhances the reliability of the conclusions of the studies as it demonstrates a comprehensive understanding of both bioinformatics and molecular biology techniques. However, the study is not without limitations. The *in silico* assay is surely dependent on the size of the database (Mirdita et al., 2022). On the other hand, the use of recombinant protein assays, while informative, may not fully replicate the *in vivo* cellular environment. Factors such as chromatin accessibility, co-factor availability, and post-translational modifications could influence the actual functionality of the recombinant protein (Hingorani and

Gierasch, 2014). For future direction, I personally suggest that it is worth to carry this work for further investigation by *in vivo* experiment. By testing switches, replacing the HMG domain of TMU into mouse SRY with Q-rich domain intact would help solidify this finding, as it is currently impossible to obtain the living specimen of TMU to observe the SRY expression in embryonic gonads.

In conclusion, this finding provides a first report, in detail demonstrating that the SRY gene can tolerate an A-to-S substitution in its HMG domain without losing its DNA-binding and bending capabilities. The study challenges previous assumptions about the fragility of sex-determination pathways, resulting in the loss of SRY dependence properties. It invites further inquiry into the molecular and evolutionary flexibility of these systems and stands as a valuable contribution to our understanding of genetic regulation, development, and adaptation.

# Acknowledgements

I would like to pay my deepest gratitude to my supervisor, Professor KUROIWA Asato, Ph.D. for providing me a chance to study abroad in Japan with MEXT's scholarship. This is truly my lifetime precious experience.

I am indebted to Associate Professor MIZUSHIMA Shusei, Ph.D., for most of my research experimental guidelines and instruments.

I express my deep gratitude to Professor ISHIMORI Koichiro, Ph.D. and Associate Professor UCHIDA Takeshi, Ph.D. for their contributions and support to the anisotropy experiments.

Many thanks to my family, who always giving deepest supported, and patiently waiting for me, to Aj. Teerasak E-KOBON, Aj. Narongsak PITCHAYAPISUT, Aj. Narongrit MEAUNGMAI, and Aj. Kornsorn SRIKULNART from KU who inspired me to pursue further study in Japan. And as OSK134, to our long live standing friendship, which reminds me who I am. As well as Chokotto sushi members for irreplaceable experiences.

Finally, a special thanks are given to 11th floor, and the original GAIJIN-CORNER members who are always fated the same hardship and happiness together.

## References

Aeling, K. A., Opel, M. L., Steffen, N. R., Tretyachenko-Ladokhina, V., Hatfield, G. W., Lathrop, R. H., & Senear, D. F. (2006). Indirect recognition in sequence-specific DNA binding by Escherichia coli integration host factor: the role of DNA deformation energy. The Journal of biological chemistry, 281(51), 39236–39248. https://doi.org/10.1074/jbc.M606363200

Anjana R, Vaishnavi MK, Sherlin D, Kumar SP, Naveen K, Kanth PS, Sekar K. Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance. Bioinformation. 2012;8(24):1220-4. doi: 10.6026/97320630081220.

Ambulkar, P. S., Waghmare, J. E., Verma Shivkumar, P., Narang, P., and Pal, A. K. (2021). A missense mutation (c.226C>A) in HMG box SRY gene affects nNLS function in 46, XY sex reversal female. Andrologia, **53**, e14011. https://doi.org/10.1111/and.14011

Assumpção, J. G., Benedetti, C. E., Maciel-Guerra, A. T., Guerra, G., Jr, Baptista, M. T., Scolfaro, M. R., and de Mello, M. P. (2002). Novel mutations affecting SRY DNA-binding activity: the HMG box N65H associated with 46,XY pure gonadal dysgenesis and the familial non-HMG box R30I associated with variable phenotypes. J Mol Med (Berl), **80**, 782–790. https://doi.org/10.1007/s00109-002-0376-9.

Badaczewska-Dawid, A. E., Nithin, C., Wroblewski, K., Kurcinski, M., and Kmiecik,

S. (2022). MAPIYA contact map server for identification and visualization of molecular interactions in proteins and biological complexes. Nucleic Acids Res, **50**, W474–W482. https://doi.org/10.1093/nar/gkac307

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., et al. (2021) Accurate prediction of protein structures and interactions using a three-track neural network. Science **373**, 871-876. https://doi.org/10.1126/science.abj8754

Baud, S., Margeat, E., Lumbroso, S., Paris, F., Sultan, C., Royer, C., & Poujol, N. (2002). Equilibrium binding assays reveal the elevated stoichiometry and salt dependence of the interaction between full-length human sex-determining region on the Y chromosome (SRY) and DNA. *The Journal of biological chemistry*, *277*(21), 18404–18410. https://doi.org/10.1074/jbc.M112366200

Bergstrom, D. E., Young, M., Albrecht, K. H., & Eicher, E. M. (2000). Related function of mouse SOX3, SOX9, and SRY HMG domains assayed by male sex determination. *Genesis (New York, N.Y. : 2000)*, *28*(3-4), 111–124.

Berta, P., Hawkins, J. R., Sinclair, A. H., Taylor, A., Griffiths, B. L., Goodfellow, P. N., and Fellous, M. (1990). Genetic evidence equating SRY and the testis-determining factor. Nature, **348**, 448–450. https://doi.org/10.1038/348448A0

Bewley, C. A., Gronenborn, A. M., & Clore, G. M. (1998). Minor groove-binding architectural proteins: structure, function, and DNA recognition. *Annual review of biophysics and biomolecular structure*, *27*, 105–131.

https://doi.org/10.1146/annurev.biophys.27.1.105

Bhandari, R. K., Haque, M. M., & Skinner, M. K. (2012). Global genome analysis of the downstream binding targets of testis determining factor SRY and SOX9. *PloS one*, *7*(9), e43380. https://doi.org/10.1371/journal.pone.0043380

Birchler JA, Yang H. The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell*. 2022;34(7):2466-2474. doi:10.1093/plcell/koac076

Bowles, J., Cooper, L., Berkman, J., and Koopman, P. (1999). Sry requires a CAG repeat domain for male sex determination in *Mus musculus*. Nat. Genet*.,* **22 (4)**, 405–408. https://doi.org/10.1038/11981

Bridgewater LC, Walker MD, Miller GC, Ellison TA, Holsinger LD, Potter JL, Jackson TL, Chen RK, Winkel VL, Zhang Z, McKinney S, de Crombrugghe B. Adjacent DNA sequences modulate Sox9 transcriptional activation at paired Sox sites in three chondrocyte-specific enhancer elements. Nucleic Acids Res. 2003 Mar 1;31(5):1541-53. doi: 10.1093/nar/gkg230. PMID: 12595563; PMCID: PMC149823.

Buric F, Viknander S, Fu X, Lemke O, Carmona OG, Zrimec J, Szyrwiel L, Mülleder M, Ralser M, Zelezniak A. Amino acid sequence encodes protein abundance shaped by protein stability at reduced synthesis cost. Protein Sci. 2025 Jan;34(1):e5239. doi: 10.1002/pro.5239. PMID: 39665261; PMCID:

PMC11635393.

Chakrabarti, P., & Bhattacharyya, R. (2007). Geometry of nonbonded interactions involving planar groups in proteins. *Progress in biophysics and molecular biology*, *95*(1-3), 83–137. https://doi.org/10.1016/j.pbiomolbio.2007.03.016

Chakraborty, A., Lyonnais, S., Battistini, F., Hospital, A., Medici, G., Prohens, R., Orozco, M., Vilardell, J., & Solà, M. (2017). DNA structure directs positioning of the mitochondrial genome packaging protein Abf2p. *Nucleic acids research*, *45*(2), 951–967. https://doi.org/10.1093/nar/gkw1147

Charlesworth, B., & Charlesworth, D. (2020). Evolution: A New Idea about the Degeneration of Y and W Chromosomes. *Current biology : CB*, *30*(15), R871–R873. https://doi.org/10.1016/j.cub.2020.06.008

Chen YS, Racca JD, Weiss MA. Tenuous Transcriptional Threshold of Human Sex Determination. I. SRY and Swyer Syndrome at the Edge of Ambiguity. *Front Endocrinol (Lausanne)*. 2022;13:945030. Published 2022 Jul 26. doi:10.3389/fendo.2022.945030

Churchill MEA, Klass J, Zoetewey DL. Structural analysis of HMGD-DNA complexes reveals influence of intercalation on sequence selectivity and DNA bending. J Mol Biol. 2010 Oct 15;403(1):88-102. doi: 10.1016/j.jmb.2010.08.031. Epub 2010 Aug 25. PMID: 20800069; PMCID: PMC2962916.

Corona RI, Guo JT. Statistical analysis of structural determinants for protein-DNA-

binding specificity. Proteins. 2016 Aug;84(8):1147-61. doi: 10.1002/prot.25061. Epub 2016 Jun 15. PMID: 27147539; PMCID: PMC4945413.

Clépet, C., Schafer, A. J., Sinclair, A. H., Palmer, M. S., Lovell-Badge, R., and Goodfellow, P. N. (1993) The human SRY transcript. Hum Mol Genet **2(12)**, 2007-2012. https://doi.org/10.1093/hmg/2.12.2007

Culha, C., Ozkaya, M., Serter, R., Sahin, I., Aydin, B., and Aral, Y. (2012) Swyer's Syndrome: In a Fifty-Year-Old Female. J Obstet Gynaecol India **62**, 571-574. https://doi.org/10.1007/s13224-011-0100-1

Das, S., Roy, S., & Bhattacharyya, D. (2020). DNA base sequence specificity through partial intercalation: DFT-D based energy analysis of molecular dynamics snapshots. *Journal of molecular graphics & modelling*, *101*, 107722. https://doi.org/10.1016/j.jmgm.2020.107722

Dewing, P., Chiang, C. W., Sinchak, K., Sim, H., Fernagut, P. O., Kelly, S., Chesselet, M. F., Micevych, P. E., Albrecht, K. H., Harley, V. R., et al. (2006) Direct regulation of adult brain function by the male-specific factor SRY. Curr Biol **16**, 415-420. https://doi.org/10.1016/j.cub.2006.01.017

Dill K. A. (1990). Dominant forces in protein folding. *Biochemistry*, *29*(31), 7133–7155. https://doi.org/10.1021/bi00483a001

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics **29**, 15-21. https://doi.org/10.1093/bioinformatics/bts635

Dubin, R. A., & Ostrer, H. (1994). Sry is a transcriptional activator. *Molecular endocrinology (Baltimore, Md.)*, *8*(9), 1182–1192. https://doi.org/10.1210/mend.8.9.7838151

Etheve, L., Martin, J., & Lavery, R. (2016). Protein-DNA interfaces: a molecular dynamics analysis of time-dependent recognition processes for three transcription factors. *Nucleic acids research*, *44*(20), 9990–10002. https://doi.org/10.1093/nar/gkw841

Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell-Badge R, and Bianchi M. E. (1992) SRY, like HMG1, recognizes sharp angles in DNA. EMBO J **11**, 4497-4506. https://doi.org/10.1002/j.1460-2075.1992.tb05551.x

Ghosh G, Adams JA. Phosphorylation mechanism and structure of serine-arginine protein kinases. FEBS J. 2011 Feb;278(4):587-97. doi: 10.1111/j.1742-4658.2010.07992.x. Epub 2011 Jan 12. PMID: 21205204; PMCID: PMC3079193

Giese, K., Pagel, J., and Grosschedl, R. (1994) Distinct DNA-binding properties of the high mobility group domain of murine and human SRY sex-determining factors. Proc Natl Acad Sci U S A. **91**, 3368-3372. https://doi.org/10.1073/pnas.91.8.3368

Gonen, N., Futtner, C. R., Wood, S., Garcia-Moreno, S. A., Salamone, I. M., Samson, S. C., Sekido, R., Poulat, F., Maatouk, D. M., and Lovell-Badge, R. (2018) Sex reversal following deletion of a single distal enhancer of Sox9. Science **360**, 1469-1473. https://doi.org/10.1126/science.aas9408

Graves, P. E., Davis, D., Erickson, R. P., Lopez, M., Kofman-Alfaro, S., Mendez, J. P., and Speer, I. E. (1999) Ascertainment and mutational studies of SRY in nine XY females. Am J Med Genet **83**, 138-139.

Hacker, A., Capel, B., Goodfellow, P., and Lovell-Badge, R. (1995) Expression of Sry, the mouse sex determining gene. Development **121**, 1603-1614. https://doi.org/10.1242/dev.121.6.1603

Hardwidge, P. R., Zimmerman, J. M., & Maher, L. J., 3rd (2002). Charge neutralization and DNA bending by the Escherichia coli catabolite activator protein. *Nucleic acids research*, *30*(9), 1879–1885. https://doi.org/10.1093/nar/30.9.1879

Harley, V. R., Jackson, D. I., Hextall, P. J., Hawkins, J. R., Berkovitz, G. D., Sockanathan, S., Lovell-Badge, R., & Goodfellow, P. N. (1992). DNA binding activity of recombinant SRY from normal males and XY females. *Science (New York, N.Y.)*, *255*(5043), 453–456. https://doi.org/10.1126/science.1734522

Harley VR, Lovell-Badge R, Goodfellow PN. Definition of a consensus DNA binding site for SRY. Nucleic Acids Res. 1994 Apr 25;22(8):1500-1. doi: 10.1093/nar/22.8.1500. PMID: 8190643; PMCID: PMC308012.

Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat Protoc. 2007;2(8):1849-61. doi: 10.1038/nprot.2007.249. PMID: 17703195; PMCID: PMC2757439

Hingorani, K. S., & Gierasch, L. M. (2014). Comparing protein folding in vitro and in

vivo: foldability meets the fitness challenge. *Current opinion in structural biology*, *24*, 81–90. https://doi.org/10.1016/j.sbi.2013.11.007 Honda, T., Suzuki, H., and Itoh, M. (1977) An unusual sex chromosome constitution found in the Amami spinous country-rat, Tokudaia osimensis osimensis. Japan J Genetics **52**, 247-249.

Hossain, A., & Saunders, G. F. (2003). Role of Wilms tumor 1 (WT1) in the transcriptional regulation of the Mullerian-inhibiting substance promoter. *Biology of reproduction*, *69*(6), 1808–1814. https://doi.org/10.1095/biolreprod.103.015826

Jäger, R. J., Harley, V. R., Pfeiffer, R. A., Goodfellow, P. N., & Scherer, G. (1992). A familial mutation in the testis-determining gene SRY shared by both sexes. *Human genetics*, *90*(4), 350–355. https://doi.org/10.1007/BF00220457 Cortez, D., Marin, R., Toledo-Flores, D., Froidevaux, L., Liechti, A., Waters, P. D., Grützner, F., & Kaessmann, H. (2014). Origins and functional evolution of Y chromosomes across mammals. Nature, 508(7497), 488–493. https://doi.org/10.1038/nature13151

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. CABIOS **8**, 275-282. https://doi.org/10.1093/bioinformatics/8.3.275

Kalantari, A., Derouiche, A., Shi, L., & Mijakovic, I. (2015). Serine/threonine/tyrosine phosphorylation regulates DNA binding of bacterial

transcriptional regulators. *Microbiology (Reading, England)*, *161*(9), 1720–1729. https://doi.org/10.1099/mic.0.000148

Kashimada, K., and Koopman, P. (2010) Sry: the master switch in mammalian sex determination. Development **137**, 3921-3930. https://doi.org/10.1242/dev.048983

KAUZMANN W. (1959). Some factors in the interpretation of protein denaturation. *Advances in protein chemistry*, *14*, 1–63. https://doi.org/10.1016/s0065-3233(08)60608-7

Kellermayer, R., Halvax, L., Czakó, M., Shahid, M., Dhillon, V. S., Husain, S. A., Süle, N., Gömöri, E., Mammel, M., & Kosztolányi, G. (2005). A novel frame shift mutation in the HMG box of the SRY gene in a patient with complete 46,XY pure gonadal dysgenesis. Diagnostic molecular pathology : the American journal of surgical pathology, part B, 14(3), 159–163. https://doi.org/10.1097/01.pas.0000176770.56541.dd

Kim, D. E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res **32**, 526-531. https://doi.org/10.1093/nar/gkh468

Kim, J. L., Nikolov, D. B., & Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, *365*(6446), 520–527. https://doi.org/10.1038/365520a0

Kimura, R., Murata, C., Kuroki, Y., and Kuroiwa, A. (2014) Mutations in the testis-specific enhancer of SOX9 in the SRY independent sex-determining mechanism

in the genus Tokudaia. PLoS One **9**. https://doi.org/10.1371/journal.pone.0108779

King, C. Y., and Weiss, M. A. (1993) The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity. Proc Natl Acad Sci U S A. **90**, 11990-11994. https://doi.org/10.1073/pnas.90.24.11990

Kobayashi, T., Yamada, F., Hashimoto, T., Abe, S., Matsuda, Y., and Kuroiwa, A. (2007) Exceptional minute sex-specific region in the X0 mammal, Ryukyu spiny rat. Chromosome Res **15**, 175-187. https://doi.org/10.1007/s10577-006-1093-y

Kosikov, K. M., Gorin, A. A., Lu, X. J., Olson, W. K., & Manning, G. S. (2002). Bending of DNA by asymmetric charge neutralization: all-atom energy simulations. *Journal of the American Chemical Society*, *124*(17), 4838–4847. https://doi.org/10.1021/ja0125218

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol **35**, 1547-1549. https://doi.org/10.1093/molbev/msy096

Kuroiwa, A., Ishiguchi, Y., Yamada, F., Shintaro, A., and Matsuda, Y. (2010) The process of a Y-loss event in an XO/XO mammal, the Ryukyu spiny rat. Chromosoma **119**, 519-526. https://doi.org/10.1007/s00412-010-0275-8

Lahr, G., Maxson, S. C., Mayer, A., Just, W., Pilgrim, C., and Reisert, I. (1995) Transcription of the Y chromosomal gene, Sry, in adult mouse brain. Brain Res

Mol Brain Res **33**, 179-182. https://doi.org/10.1016/0169-328x(95)00136-g

Lakowicz, J. R. (2006) Principles of Fluorescence Spectroscopy, 3rd ed., Springer, New York.

Larsson, A. (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics **30**, 3276-3278. https://doi.org/10.1093/bioinformatics/btu531

Laudet, V., Stehelin, D., and Clevers, H. (1993). Ancestry and diversity of the HMG box superfamily. Nucl. Acids Res. 21, 2493-2501.

Leng F. (2013). DNA bending by proteins: utilizing plasmid pBendAT as a tool. *Methods in molecular biology (Clifton, N.J.)*, *1054*, 267–282. https://doi.org/10.1007/978-1-62703-565-1_18

Leonard DA, Rajaram N, Kerppola TK. Structural basis of DNA bending and oriented heterodimer binding by the basic leucine zipper domains of Fos and Jun. Proc Natl Acad Sci U S A. 1997 May 13;94(10):4913-8. doi: 10.1073/pnas.94.10.4913. PMID: 9144164; PMCID: PMC24605.

Lenormand, T., Fyon, F., Sun, E., & Roze, D. (2020). Sex Chromosome Degeneration by Regulatory Evolution. *Current biology : CB*, *30*(15), 3001–3006.e5. https://doi.org/10.1016/j.cub.2020.05.052

Li, B., Phillips, N. B., Jancso-Radek, A., Ittah, V., Singh, R., Jones, D. N., Haas, E., and Weiss, M. A. (2006) SRY-directed DNA bending and human sex reversal:

reassessment of a clinical mutation uncovers a global coupling between the HMG box and its tail. J Mol Biol **360**, 310-328. https://doi.org/10.1016/j.jmb.2006.04.048

Li, B., and Dewey, C. N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12**. 10.1186/1471-2105-12-323. https://doi.org/10.1186/1471-2105-12-323

Li, J., Hou, C., Ma, X., Guo, S., Zhang, H., Shi, L., Liao, C., Zheng, B., Ye, L., Yang, L., & He, X. (2021). Entropy-Enthalpy Compensations Fold Proteins in Precise Ways. *International journal of molecular sciences*, *22*(17), 9653.

Lin, M., & Guo, J. T. (2019). New insights into protein-DNA binding specificity from hydrogen bond based comparative study. *Nucleic acids research*, *47*(21), 11103–11113. https://doi.org/10.1093/nar/gkz963

Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., & Wright, P. E. (1995). Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, *376*(6543), 791–795. https://doi.org/10.1038/376791a0

Lovell-Badge, R., Canning, C., and Sekido, R. (2002) in *The genetics and biology of sex determination* (Chadwick, D., and Good, J., Eds.) pp 4−22, John Wiley, West Sussex, U.K.

Luscombe NM, Laskowski RA, Thornton JM. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res. 2001 Jul 1;29(13):2860-74. doi: 10.1093/nar/29.13.2860. PMID:

11433033; PMCID: PMC55782.

Manning G. S. (1978). The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Quarterly reviews of biophysics*, *11*(2), 179–246. https://doi.org/10.1017/s0033583500002031

Manning G. S. (2003). Is a small number of charge neutralizations sufficient to bend nucleosome core DNA onto its superhelical ramp?. *Journal of the American Chemical Society*, *125*(49), 15087–15092. https://doi.org/10.1021/ja030320t

Martin B, Dans PD, Wieczór M, Villegas N, Brun-Heath I, Battistini F, Terrazas M, Orozco M. Molecular basis of Arginine and Lysine DNA sequence-dependent thermo-stability modulation. PLoS Comput Biol. 2022 Jan 10;18(1):e1009749. doi: 10.1371/journal.pcbi.1009749. PMID: 35007284; PMCID: PMC8782489.

Mayer, A., Lahr, G., Swaab, D. F., Pilgrim, C., and Reisert, I. (1998) The Y-chromosomal genes SRY and ZFY are transcribed in adult human brain. Neurogenetics **1**, 281-288. https://doi.org/10.1007/s100480050042

McBryant SJ, Gedulin B, Clemens KR, Wright PE, Gottesfeld JM. Assessment of major and minor groove DNA interactions by the zinc fingers of Xenopus transcription factor IIIA. Nucleic Acids Res. 1996;24(13):2567-2574. doi:10.1093/nar/24.13.2567

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature methods*, *19*(6), 679–682. https://doi.org/10.1038/s41592-022-01488-1

Mirzabekov, A. D., & Rich, A. (1979). Asymmetric lateral distribution of unshielded phosphate groups in nucleosomal DNA and its role in DNA bending. *Proceedings of the National Academy of Sciences of the United States of America*, *76*(3), 1118–1121. https://doi.org/10.1073/pnas.76.3.1118

Mondal, M., Mukherjee, S., & Bhattacharyya, D. (2014). Contribution of phenylalanine side chain intercalation to the TATA-box binding protein-DNA interaction: molecular dynamics and dispersion-corrected density functional theory studies. *Journal of molecular modeling*, *20*(11), 2499. https://doi.org/10.1007/s00894-014-2499-7

Moxley, R. A., Jarrett, H. W., & Mitra, S. (2003). Methods for transcription factor separation. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, *797*(1-2), 269–288. https://doi.org/10.1016/s1570-0232(03)00609-3

Murata, C., Yamada, F., Kawauchi, N., Matsuda, Y., and Kuroiwa, A. (2010). Multiple copies of SRY on the large Y chromosome of the Okinawa spiny rat, Tokudaia muenninki. Chromosome Res **18**, 623-634. https://doi.org/10.1007/s10577-010-9142-y

Murata, C., Yamada, F., Kawauchi, N., Matsuda, Y., and Kuroiwa, A. (2012) The Y chromosome of the Okinawa spiny rat, Tokudaia muenninki, was rescued through fusion with an autosome. Chromosome Res **20**, 111-125. https://doi.org/10.1007/s10577-011-9268-6

Murphy, F. V., 4th, Sweet, R. M., & Churchill, M. E. (1999). The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition. *The EMBO journal*, *18*(23), 6610–6618. https://doi.org/10.1093/emboj/18.23.6610

Murphy, E. C., Zhurkin, V. B., Louis, J. M., Cornilescu, G., and Clore, G. M. (2001) Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation. J Mol Biol. **312**, 481-99. https://doi.org/10.1006/jmbi.2001.4977

Nagadoi, A., Morikawa, S., Nakamura, H., Enari, M., Kobayashi, K., Yamamoto, H., Sampei, G., Mizobuchi, K., Schumacher, M. A., & Brennan, R. G. (1995). Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain. *Structure (London, England : 1993)*, *3*(11), 1217–1224. https://doi.org/10.1016/s0969-2126(01)00257-x

Nam, D., Matsumoto, Y., Uchida, T., O'Brian, M. R., and Ishimori, K. (2020) Mechanistic insights into heme-mediated transcriptional regulation via a bacterial manganese-binding iron regulator, iron response regulator (Irr). J Biol Chem **295**, 11316-11325. https://doi.org/10.1074/jbc.RA119.011855

Ogata, Y., Nishikata, M., Kitada, K., Mizushima, S., Jogahara, T., and Kuroiwa, A. (2019) Spiny rat SRY lacks a long Q-rich domain and is not stable in transgenic mice. Dev Dyn **248**, 784-794. https://doi.org/10.1002/dvdy.73

Okuno, M., Mochimaru, Y., Matsuoka, K., Yamabe, T., Matiz-Ceron, L., Jogahara, T.,

Toyoda, A., Kuroiwa, A., and Itoh, T. (2023) Chromosomal-level assembly of Tokudaia osimensis, Tokudaia tokunoshimensis, and Tokudaia muenninki genomes. Sci Data **10**(1):927. https://doi.org/10.1038/s41597-023-02845-1

Otake, T., and Kuroiwa, A. (2016) Molecular mechanism of male differentiation is conserved in the SRY-absent mammal, Tokudaia osimensis. Sci Rep **6**. https://doi.org/10.1038/srep32874.

Pabo, C. O., & Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annual review of biochemistry*, *61*, 1053–1095. https://doi.org/10.1146/annurev.bi.61.070192.005201

Palasingam, P., Jauch, R., Ng, C. K., & Kolatkar, P. R. (2009). The structure of Sox17 bound to DNA reveals a conserved bending topology but selective protein interaction platforms. *Journal of molecular biology*, *388*(3), 619–630. https://doi.org/10.1016/j.jmb.2009.03.055

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem **25**, 1605-1612. https://doi.org/10.1002/jcc.20084

Phillips, N. B., Nikolskaya, T., Jancso-Radek, A., Ittah, V., Jiang, F., Singh, R., Haas, E., and Weiss, M. A. (2004) Sry-directed sex reversal in transgenic mice is robust with respect to enhanced DNA bending: comparison of human and murine HMG boxes. Biochemistry **43**, 7066-7081. https://doi.org/10.1021/bi049920a

Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell-Badge, R., and Bianchi, M. E. (1994) Sex-reversing mutations affect the architecture of SRY-DNA complexes. EMBO J **13**, 6115-6124. https://doi.org/10.1002/j.1460-2075.1994.tb06958.x

Pontiggia, A., Whitfield, S., Goodfellow, P. N., Lovell-Badge, R., & Bianchi, M. E. (1995). Evolutionary conservation in the DNA-binding and -bending properties of HMG-boxes from SRY proteins of primates. *Gene*, *154*(2), 277–280. https://doi.org/10.1016/0378-1119(94)00853-k

Predki, P. F., Agrawal, V., Brünger, A. T., & Regan, L. (1996). Amino-acid substitutions in a surface turn modulate protein stability. *Nature structural biology*, *3*(1), 54–58. https://doi.org/10.1038/nsb0196-54

Privalov, P. L., Dragan, A. I., & Crane-Robinson, C. (2009). The cost of DNA bending. *Trends in biochemical sciences*, *34*(9), 464–470. https://doi.org/10.1016/j.tibs.2009.05.005

Prokop, J. W., Underwood, A. C., Turner, M. E., Miller, N., Pietrzak, D., Scott, S., Smith, C., & Milsted, A. (2013). Analysis of Sry duplications on the Rattus norvegicus Y-chromosome. *BMC genomics*, *14*, 792. https://doi.org/10.1186/1471-2164-14-792

RStudio: Integrated Development Environment for R. http://www.posit.co/ (2006). Accessed 10 Nov 2023.

Racca, J. D., Chen, Y. S., Yang, Y., Phillips, N. B., & Weiss, M. A. (2016). Human Sex

Determination at the Edge of Ambiguity: INHERITED XY SEX REVERSAL DUE TO ENHANCED UBIQUITINATION AND PROTEASOMAL DEGRADATION OF A MASTER TRANSCRIPTION FACTOR. *The Journal of biological chemistry*, *291*(42), 22173–22195. https://doi.org/10.1074/jbc.M116.741959

Racca, J. D., Chen, Y. S., Brabender, A. R., Battistin, U., Weiss, M. A., and Georgiadis, M. M. (2024) Role of nucleobase-specific interactions in the binding and bending of DNA by human male sex determination factor SRY. J Biol Chem **300(9).** https://doi.org/10.1016/j.jbc.2024.107683

Reha, D., Kabelác, M., Ryjácek, F., Sponer, J., Sponer, J. E., Elstner, M., Suhai, S., & Hobza, P. (2002). Intercalators. 1. Nature of stacking interactions between intercalators (ethidium, daunomycin, ellipticine, and 4',6-diaminide-2-phenylindole) and DNA base pairs. Ab initio quantum chemical, density functional theory, and empirical potential study. *Journal of the American Chemical Society*, *124*(13), 3366–3376. https://doi.org/10.1021/ja011490d

Reményi, A., Lins, K., Nissen, L. J., Reinbold, R., Schöler, H. R., & Wilmanns, M. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes & development*, *17*(16), 2048–2059. https://doi.org/10.1101/gad.269303

Rhie, A., Nurk, S., Cechova, M., Hoyt, S. J., Taylor, D. J., Altemose, N., Hook, P. W., Koren, S., Rautiainen, M., Alexandrov, I. A., Allen, J., Asri, M., Bzikadze, A. V.,

Chen, N. C., Chin, C. S., Diekhans, M., Flicek, P., Formenti, G., Fungtammasan, A., Garcia Giron, C., … Phillippy, A. M. (2023). The complete sequence of a human Y chromosome. Nature, 621(7978), 344–354. https://doi.org/10.1038/s41586-023-06457-y

Robert Finestra, T., & Gribnau, J. (2017). X chromosome inactivation: silencing, topology and reactivation. *Current opinion in cell biology*, *46*, 54–61. https://doi.org/10.1016/j.ceb.2017.01.007

Rodríguez-Lumbreras, L. A., Jiménez-García, B., Giménez-Santamarina, S., and Fernández-Recio, J. (2022) pyDockDNA: A new web server for energy-based protein-DNA docking and scoring. Front Mol Biosci **9**. https://doi.org/10.3389/fmolb.2022.988996.

Rohs, R., Bloch, I., Sklenar, H., & Shakked, Z. (2005). Molecular flexibility in ab initio drug docking to DNA: binding-site and binding-mode transitions in all-atom Monte Carlo simulations. *Nucleic acids research*, *33*(22), 7048–7057. https://doi.org/10.1093/nar/gki1008

Ross, D. G., Bowles, J., Koopman, P., & Lehnert, S. (2008). New insights into SRY regulation through identification of 5' conserved sequences. *BMC molecular biology*, *9*, 85. https://doi.org/10.1186/1471-2199-9-85

Salas-Cortés, L., Jaubert, F., Barbaux, S., Nessmann, C., Bono, M. R., Fellous, M., McElreavey, K., and Rosemblatt, M. (1999) The human SRY protein is present in fetal and adult Sertoli cells and germ cells. Int J Dev Biol **43**, 135-40.

Schumacher, M. A., Choi, K. Y., Zalkin, H., & Brennan, R. G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science (New York, N.Y.)*, *266*(5186), 763–770. https://doi.org/10.1126/science.7973627

Shahid, M., Dhillion, V. S., Jain, N., Hedau, S., Diwakar, S., Sachdeva, P., Batra, S., Das, B. C., & Husain, S. A. (2004). Two new novel point mutations localized upstream and downstream of the HMG box region of the SRY gene in three Indian 46,XY females with sex reversal and gonadal tumour formation. *Molecular human reproduction*, *10*(7), 521–526. https://doi.org/10.1093/molehr/gah071

Shortle D. (2009). One sequence plus one mutation equals two folds. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21011–21012. https://doi.org/10.1073/pnas.0912370107

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A. C., Gordân, R., & Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. *Trends in biochemical sciences*, *39*(9), 381–399. https://doi.org/10.1016/j.tibs.2014.07.002

Strauss, J. K., & Maher, L. J., 3rd (1994). DNA bending by asymmetric phosphate neutralization. *Science (New York, N.Y.)*, *266*(5192), 1829–1834. https://doi.org/10.1126/science.7997878

Terao, M., Ogawa, Y., Takada, S., Kajitani, R., Okuno, M., Mochimaru, Y., Matsuoka,

K., Itoh, T., Toyoda, A., Kono, T., Jogahara, T., et al. (2022) Turnover of mammal sex chromosomes in the Sry-deficient Amami spiny rat is due to male-specific upregulation of Sox9. Proc Natl Acad Sci U S A. **119**. https://doi.org/10.1073/pnas.2211574119

Tsuchiya, K., Wakana, S., Suzuki, H., Hattori, S., and Hayashi, Y. (1989) Taxonomic study of Tokudaia (Rodentia: Muridae): I. Genetic differentiation. Mem Natl Sci Museum Tokyo **22**, 227-34.

Tsuji-Hosokawa, A., Ogawa, Y., Tsuchiya, I., Terao, M., and Takada, S. (2022) Human SRY Expression at the Sex-determining Period is Insufficient to Drive Testis Development in Mice. Endocrinology **163**. https://doi.org/10.1210/endocr/bqab217

Turner, M. E., Martin, C., Martins, A. S., Dunmire, J., Farkas, J., Ely, D. L., and Milsted, A. (2007). Genomic and expression analysis of multiple Sry loci from a single Rattus norvegicus Y chromosome. BMC Genet **8,** 11. https://doi.org/10.1186/1471-2156-8-11.

Ukiyama, E., Jancso-Radek, A., Li, B., Milos, L., Zhang, W., Phillips, N. B., Morikawa, N., King, C. Y., Chan, G., Haqq, C. M., et al (2001) SRY and architectural gene regulation: the kinetic stability of a bent protein-DNA complex can regulate its transcriptional potency. Mol Endocrinol **15**, 363-377. https://doi.org/10.1210/mend.15.3.0621

van der Vliet, P. C., & Verrijzer, C. P. (1993). Bending of DNA by transcription

factors. *BioEssays : news and reviews in molecular, cellular and developmental biology*, *15*(1), 25–32. https://doi.org/10.1002/bies.950150105

Vonderfecht, T. R., Schroyer, D. C., Schenck, B. L., McDonough, V. M., & Pikaart, M. J. (2008). Substitution of DNA-contacting amino acids with functional variants in the Gata-1 zinc finger: a structurally and phylogenetically guided mutagenesis. *Biochemical and biophysical research communications*, *369*(4), 1052–1056. https://doi.org/10.1016/j.bbrc.2008.02.136

Wang, X., Xue, M., Zhao, M., He, F., Li, C., and Li, X. (2018) Identification of a novel mutation (Ala66Thr) of SRY gene causes XY pure gonadal dysgenesis by affecting DNA binding activity and nuclear import. Gene **651**, 143-151. https://doi.org/10.1016/j.gene.2018.01.076

Weaver, S. D., & Whelan, R. J. (2021). Characterization of DNA aptamer-protein binding using fluorescence anisotropy assays in low-volume, high-efficiency plates. *Analytical methods : advancing methods and applications*, *13*(10), 1302–1307. https://doi.org/10.1039/d0ay02256j

Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R., Laue, E. D., & Thomas, J. O. (1993). Structure of the HMG box motif in the B-domain of HMG1. *The EMBO journal*, *12*(4), 1311–1319. https://doi.org/10.1002/j.1460-2075.1993.tb05776.x

Weiss, J., Meeks, J. J., Hurley, L., Raverot, G., Frassetto, A., and Jameson, J. L. (2003) Sox3 is required for gonadal function, but not sex determination, in males and females. Mol Cell Biol **23**, 8084-8091.

https://doi.org/10.1128/MCB.23.22.8084-8091.2003

Weiss, M.A. (2005). Molecular Mechanisms of Male Sex Determination. In: DNA Conformation and Transcription. Molecular Biology Intelligence Unit. Springer, Boston, MA. https://doi.org/10.1007/0-387-29148-2_12

Werner, M. H., Clore, M., Fisher, C. L., Fisher, R. J., Trinh, L., Shiloach, J., & Gronenborn, A. M. (1995). The solution structure of the human ETS1-DNA complex reveals a novel mode of binding and true side chain intercalation. *Cell*, *83*(5), 761–771. https://doi.org/10.1016/0092-8674(95)90189-2

Werner, M. H., Gronenborn, A. M., & Clore, G. M. (1996). Intercalation, DNA kinking, and the control of transcription. *Science (New York, N.Y.)*, *271*(5250), 778–784. https://doi.org/10.1126/science.271.5250.778

Werner, M. H., Huth, J. R., Gronenborn, A. M., & Clore, G. M. (1995). Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, *81*(5), 705–714. https://doi.org/10.1016/0092-8674(95)90532-4

Wilson KA, Kellie JL, Wetmore SD. DNA-protein π-interactions in nature: abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. Nucleic Acids Res. 2014 Jun;42(10):6726-41. doi: 10.1093/nar/gku269. Epub 2014 Apr 17. PMID: 24744240; PMCID: PMC4041443.

Wu, T. D., and Watanabe, C. K. (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics **21**,1859–1875. https://doi.org/10.1093/bioinformatics/bti310

Xu, L., Jiang, S., Wu, J., & Zou, Q. (2021). An in silico approach to identification, categorization and prediction of nucleic acid binding proteins. *Briefings in bioinformatics*, *22*(3), bbaa171. https://doi.org/10.1093/bib/bbaa171

Yamada, F., Kawauchi, N., Nakata, K., Abe, S., Kotaka, N., Takashima, A., Murata, C., and Kuroiwa, A. (2011) Rediscovery After Thirty Years Since the Last Capture of the Critically Endangered Okinawa Spiny Rat Tokudaia muenninki in the Northern Part of Okinawa Island. Mammal Stu **35**, 243-255.

Zhang, L., & Kong, L. (2020). A Novel Amino Acid Properties Selection Method for Protein Fold Classification. *Protein and peptide letters*, *27*(4), 287–294. https://doi.org/10.2174/0929866526666190718151753

Zhou HX, Pang X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. Chem Rev. 2018 Feb 28;118(4):1691-1741. doi: 10.1021/acs.chemrev.7b00305. Epub 2018 Jan 10. PMID: 29319301; PMCID: PMC5831536.

# Publication

Urunanont, P., Mizushima, S., Uchida, T., Ishimori, K., Matiz-Ceron, L., & Kuroiwa, A. (2025). Multiple Sry genes with an A-to-S substitution in the HMG domain retain DNA-binding ability in the Okinawa spiny rat. Genes & genetic systems, 10.1266/ggs.25-00021. Advance online publication. https://doi.org/10.1266/ggs.25-00021

# List of participated conferences

## Oral presentation

Urunanont, P., Mizushima, S., Itoh, T., Kuroiwa, A. Eccentric SRY of the Okinawa spiny rat and its sex-determining ability revealed by *in silico* and *in vitro* analysis. Asia-Pacific Chromosome Colloquium. Tekirdag, Turkiye. September 18, 2023.

## Poster presentation

Urunanont, P., Mizushima, S., Itoh, T., Kuroiwa, A. SRY structure and function of the Okinawa spiny rat, *Tokudaia muenninki*, revealed by in silico and in vitro analysis. 73th Annual Conference of the Society of Chromosome Research. Online. November 16, 2022.