



## OPEN ACCESS

## EDITED BY

Ananda L. Roy,  
National Institutes of Health (NIH),  
United States

## REVIEWED BY

Raki Sudan,  
Washington University in St. Louis,  
United States  
Alex Dent,  
Purdue University Indianapolis, United States

## \*CORRESPONDENCE

Patrick G. Hogan  
✉ phogan@lji.org

RECEIVED 14 February 2025

ACCEPTED 20 May 2025

PUBLISHED 17 July 2025

## CITATION

Hogan PG, Villalobos Reveles B and  
Arteaga Vazquez LJ (2025) TOX,  
through a glass, darkly.  
*Front. Immunol.* 16:1576468.  
doi: 10.3389/fimmu.2025.1576468

## COPYRIGHT

© 2025 Hogan, Villalobos Reveles and  
Arteaga Vazquez. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# TOX, through a glass, darkly

Patrick G. Hogan<sup>1,2,3\*</sup>, Bruno Villalobos Reveles<sup>1</sup>  
and Leo Josue Arteaga Vazquez<sup>1</sup>

<sup>1</sup>La Jolla Institute for Immunology, La Jolla, CA, United States, <sup>2</sup>Moore's Cancer Center, University of California–San Diego, La Jolla, CA, United States, <sup>3</sup>Program in Immunology, University of California–San Diego, La Jolla, CA, United States

The transcription factor TOX has attracted attention in recent years for its role in CD8<sup>+</sup> T cell exhaustion. In fact, TOX was known historically for its diverse roles in immune cell biology. Here, we inquire into the basis for this versatility, and propose that one main consideration is that TOX is an HMG-box transcription factor. We discuss some mechanisms that other HMG-box transcription factors employ to perform their cellular functions, as examples of the range of mechanisms TOX might employ in furthering T cell exhaustion. This inquiry begins with the literature placing TOX as a central player in CD8<sup>+</sup> T cell exhaustion and in other immune cell processes. An understanding of TOX as a transcription factor has to be organized around its binding to relevant target sites in DNA. Thus, we next cover the reasons that TOX is classified as an HMG-box protein, the well-defined but narrow scope of what TOX shares with other HMG-box proteins, and the unequivocal evidence that binding of HMG-box proteins stabilizes kinked or bent DNA. We consider the constant features and some variables in DNA recognition by HMG-box proteins. Since binding and bending DNA is not in isolation an explanation of any biological process, we look at biological examples highlighting specific ways that HMG-box proteins drive cellular processes. Finally, we outline some lines of research that could be informative in understanding the cellular mechanisms of TOX in T cell exhaustion.

## KEYWORDS

TOX, transcription factor, mechanism, HMG domain, HMG-box protein, T cell, CD8<sup>+</sup> T cell, exhaustion

## TOX in exhausted CD8<sup>+</sup> T cells

### Data that prompted attention to TOX

Converging threads of evidence made TOX a candidate regulator of T cell exhaustion. TOX mRNA was elevated in CD8<sup>+</sup> T cells when hyporesponsiveness was induced by a constitutively active version of NFAT1 (1), in CD8<sup>+</sup> tumor-infiltrating lymphocytes (TILs) undergoing exhaustion (2), in mouse tumor antigen-specific CD8<sup>+</sup> TILs or CD8<sup>+</sup> CAR TILs (3, 4), and in exhausted TILs recovered from human melanomas (5). TOX mRNA was similarly elevated when exhaustion was caused by chronic LCMV infection (6, 7). The collected evidence was summarized in (8, [Supplementary Figure S1](#)). The mRNA findings were backed up by the progressive elevation of TOX protein as exhaustion proceeded in

TILs infiltrating mouse B16 melanomas or tamoxifen-inducible liver cancers (8, 9), and by the expression of TOX protein in CD8<sup>+</sup> TILs from human tumors (8–10). There was high sustained expression of TOX protein in LCMV-specific T cells from chronic viral infections (10).

## TOX is involved in implementing T cell exhaustion

Several laboratories explored the role of TOX in T cell exhaustion independently. One experimental design used CD8<sup>+</sup> T cells from *TOX*<sup>−/−</sup> mice (9–12). Another design, sensitive to a concern about the effects of TOX absence on T cell development in the thymus, delivered shRNAs targeting *TOX* to CD8<sup>+</sup> CAR T cells that had been isolated from *TOX2*<sup>−/−</sup> mice (8). *TOX*<sup>−/−</sup> TILs exhibited very little upregulation of inhibitory receptors, but were nevertheless poorly functional, as assessed by IFN $\gamma$  and TNF expression and by target cell lysis (9). Tumor control was not examined in these experiments, because the tumor model involved the induction of widespread liver cancer. CAR TILs prepared with shTOX in a *TOX2*<sup>−/−</sup> background showed reduced expression of inhibitory receptors, along with better preservation of the ability to make IFN $\gamma$  and TNF and to lyse target cells (8). These CAR T cells were also more effective in controlling a tumor than control CAR T cells, completely clearing the tumor in many cases. There were characteristic changes in TIL gene expression in the absence of TOX, with a few notable differences between the two datasets that might reflect contributions from TOX2 in one case and contributions from residual TOX in the other (8, 9). Gene expression in LCMV-specific P14 T cells early in a chronic infection with LCMV clone 13 was distinguished by a strong effector-like signature (10).

A consistent observation in both tumors and chronic infection was that complete loss of TOX resulted in failure to maintain a responding T cell population beyond 1–3 weeks after T cell adoptive transfer (9–11). Thus TOX is crucial to the long term survival of exhausted CD8<sup>+</sup> T cells. This has been traced to a pathology of progenitor-exhausted T cells in the absence of TOX (11, 13). In contrast, TOX is dispensable for the initial activation and expansion of CD8<sup>+</sup> effector T cells and differentiation of memory T cells in the primary response to acute infections (9, 10).

## Level of TOX expression

Haploinsufficient P14 CD8<sup>+</sup> T cells—presumably with about half the normal level of TOX protein—controlled B16-gp33 tumor growth better than wildtype P14 cells (10), and bulk RNA-seq data showed that there was residual *Tox* mRNA on day 8 after transfer in the shTOX-treated CAR TILs that were effective in controlling tumor growth (8). It would be of major biological and practical importance if there were different threshold levels of TOX for TOX-dependent T cell survival and TOX-dependent T cell

hyporesponsiveness, but current evidence is not determinative on this point. Signaling interventions that modulate the *TOX* locus opening characteristic of exhausted CD8<sup>+</sup> TILs (1–3, 9) might offer a path forward. For example, BATF overexpression in TILs attenuates *TOX* locus opening and TOX expression (14), and CD4<sup>+</sup> T cell help in CD8<sup>+</sup> T cell–dendritic cell–CD4<sup>+</sup> T cell triads within a tumor prevents *TOX* locus opening and ameliorates CD8<sup>+</sup> T cell exhaustion (15). It has not been determined yet to what extent reduced TOX expression is causative for the lower expression of inhibitory receptors and improved effector function, but these cases might provide an avenue to test for a causal link.

## Unanswered questions

Defining the contributions of TOX as a transcription factor in T cell exhaustion requires connecting TOX to individual target genes. A strong case can be made that the *Pdcd1* locus is a direct transcriptional target of TOX. The exhaustion-specific -23 kb open-chromatin region upstream of the *Pdcd1* transcription start site is a functional enhancer in the EL4 cell line, a model for T cells with constitutive high expression of PD1 (16); and binding of TOX to that region has been demonstrated by ChIP-qPCR in CD8<sup>+</sup> T cells ectopically expressing FLAG-tagged TOX (8). TOX might conceivably regulate other genes in the exhaustion program in a similar way. On the other hand, a plausible reading of the Alfei et al. paper (11) is that TOX is a restraint on the CD8<sup>+</sup> T cell effector program in virally infected mice. The data there show that T cells lacking functional TOX have more KLRG1 expression and greater effector function during chronic LCMV infection, whereas forced expression of TOX under conditions where cells are on the borderline for developing exhaustion leads to markedly fewer cells expressing KLRG1 and an attenuated effector program. If a primary role of TOX lies in shifting the balance between effector and exhaustion programs, or in promoting the survival and expansion of specific differentiating T cell subsets, TOX might accomplish these objectives by controlling a relatively small set of direct target genes, and many genes whose expression is sensitive to TOX might not be direct targets.

## TOX in other immune cells

### TOX in memory T cells

A counterpoint to the studies of TOX and T cell exhaustion came in a report that mouse CD8<sup>+</sup> memory T cells can rapidly upregulate TOX in response to either TCR stimulation or inflammatory cytokine stimulation, at the same time upregulating the effector proteins interferon- $\gamma$  and granzyme B (17). Since expression of TOX was not monitored for an extended period in that report, it is not clear how the data should be related to the sustained TOX levels in exhausted T cells (8–10) or to the early transient expression of TOX at modest levels in effector T cells in acute infections (10). In the latter acute-infection experiments,

effector T cells were differentiating from naïve T cells, and one simple possibility is that memory CD8<sup>+</sup> T cells are more efficient than naïve CD8<sup>+</sup> T cells at the rapid transient induction of TOX. TOX is also typically present in human CD8<sup>+</sup> effector-memory cells (18). *Tox* mRNA levels are also elevated in tissue-resident memory CD8<sup>+</sup> T cells in the small intestine (19). The observation of TOX in memory T cells is reminiscent of findings from an earlier analysis of T cell gene coexpression networks in resolving LCMV Armstrong infection and in chronic LCMV clone 13 infection (20). There, TOX expression was detected both in CD8<sup>+</sup> memory T cells and in CD8<sup>+</sup> exhausted T cells, but TOX was connected into different transcriptional submodules in the two cases. The three papers (17, 18, 20) articulated, at a minimum, a clear caution against conflating TOX expression with exhaustion.

## Earliest studies: TOX in thymus

In fact, the caution was just a reminder of older evidence that TOX is a versatile actor in the immune system (Figure 1). TOX was initially discovered as a transcription factor contributing to T cell development in the thymus (21). TOX is upregulated transiently in thymocytes at the time of  $\beta$ -selection and again in DP thymocytes undergoing positive selection, and cells lacking functional TOX are blocked in CD4<sup>+</sup> lineage development at the CD4<sup>low</sup>CD8<sup>low</sup> > CD4<sup>+</sup>CD8<sup>low</sup> transition (21–24).

## Other cases

The prominent role of TOX in thymocyte development is not an isolated instance. TOX is crucial to the development of lymphoid tissue inducer (LTi) cells and conventional NK cells (25), innate lymphoid cells (ILCs) (26), and follicular helper T (Tfh) cells (27). In mouse oligodendrocyte-OVA or oligodendrocyte-LCMV glycoprotein models of pathogen-triggered CNS autoimmune

disease, TOX supports CD8<sup>+</sup> T cell persistence in the CNS and restrains terminal effector differentiation (28, 29). In both acute and chronic viral infections, TOX is expressed early in a population of TCF1<sup>+</sup>TOX<sup>+</sup> CD8<sup>+</sup> T cells that have been characterized as multipotential progenitors (30, 31). The unequivocal message is that TOX has critical functions in diverse immune-cell contexts. Remaining questions are whether and to what extent TOX invokes common mechanisms and gene targets in carrying out these diverse functions, and, conversely, whether TOX relies on distinct mechanisms and gene targets for some individual cases.

## TOX as an HMG-box protein

### Thymocyte selection-associated high mobility group box protein

As recognized in its formal designation—thymocyte selection-associated high mobility group (HMG) box protein—TOX is grouped in the broad class of HMG box proteins. These proteins share one or more copies of an ~80-residue stretch of recognizable sequence homology with the canonical examples of the class, the proteins designated HMGB1–HMGB4 in humans. Distinct subfamilies of HMG box share have limited or no sequence homology outside the HMG box(es) themselves. Correspondingly, their common characteristic is involvement in DNA binding, whereas they otherwise participate in diverse cellular processes and may operate by distinct biochemical mechanisms. The common feature of DNA binding is despite considerable sequence variability in the HMG box(es) themselves.

The HMG box is the sequence signature by which the class is recognized. ‘HMG box’ is also an accepted term for the corresponding folded protein domains. Here we will use the term ‘HMG box’ interchangeably with ‘HMG domain’ in the latter meaning, often referring to HMG domains when the intent is to emphasize their identity as physical subdomains of the protein.

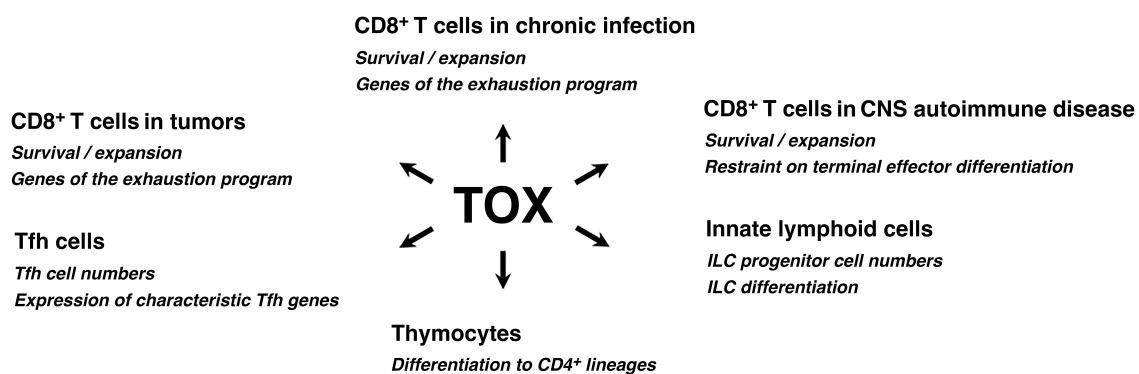


FIGURE 1

TOX acts in diverse processes in immune cells. TOX has documented roles in the development of adaptive and innate immune cells, and steers the differentiation of mature CD4<sup>+</sup> and CD8<sup>+</sup> T cells in specific contexts. Refer to the text for further discussion. TOX is also prominently expressed in thymocytes prior to  $\beta$ -selection, transiently expressed in some CD8<sup>+</sup> T cells at the outset of an effector response, and upregulated in some CD8<sup>+</sup> memory T cells and resident memory T cells of the small intestine (not depicted), but its biological effects in those cases have not been defined.

## Canonical HMGB protein, HMGB1

The defining member of the class is HMGB1, a relatively small protein of 215 residues in humans, comprising two HMG domains connected by a short linker region and exhibiting only modest pairwise identity; two nuclear localization sequences; and an acidic C-terminal tail (Figure 2A). We will focus on the way HMGB1 carries on an ancient role of similar proteins in DNA architecture and in processes that involve DNA flexibility and DNA rearrangements (DNA replication, DNA repair, transcription). Other established cellular functions of HMGB1 will not be addressed here.

## HMG domain alignments

The single HMG domain in human TOX can be aligned, by definition, with either HMG domain of HMGB1, but it has more sequence identity with the second HMG domain (Figure 2B). TOX exhibits 47% and 26% identity with HMGB1-(domain 2) and HMGB1-(domain 1), respectively, in the region covering the

respective HMG boxes. For comparison, the alignments for human TCF1 (percent identity 23%), human LEF1 (22%), human SRY (26%), and human SOX2 (27%)— other proteins that possess a single HMG box and that will be discussed below— are shown in (Figure 2B). The TOX-HMGB1 homology extends to the linker region of HMGB1 on the N-terminal side of the second HMG domain, resulting in 48% identity over the region comprising TOX residues 246–330 (Figure 2C). Thus TOX adheres more closely to the HMGB1 sequence in its DNA-binding domain than these other familiar HMG-box transcription factors. The protein alignment also brings to the fore the fact that there is more to TOX than an HMG domain: Its N-terminal region of ~240 residues and C-terminal region of ~200 residues account for the bulk of the TOX protein, and exhibit no informative homology with other human or vertebrate proteins.

## Anopheles 'TOX'

TOX homology with the predicted *Anopheles gambiae* protein now annotated as XP\_061513506 was reported in an early description of TOX-family proteins (33). The conserved sequence

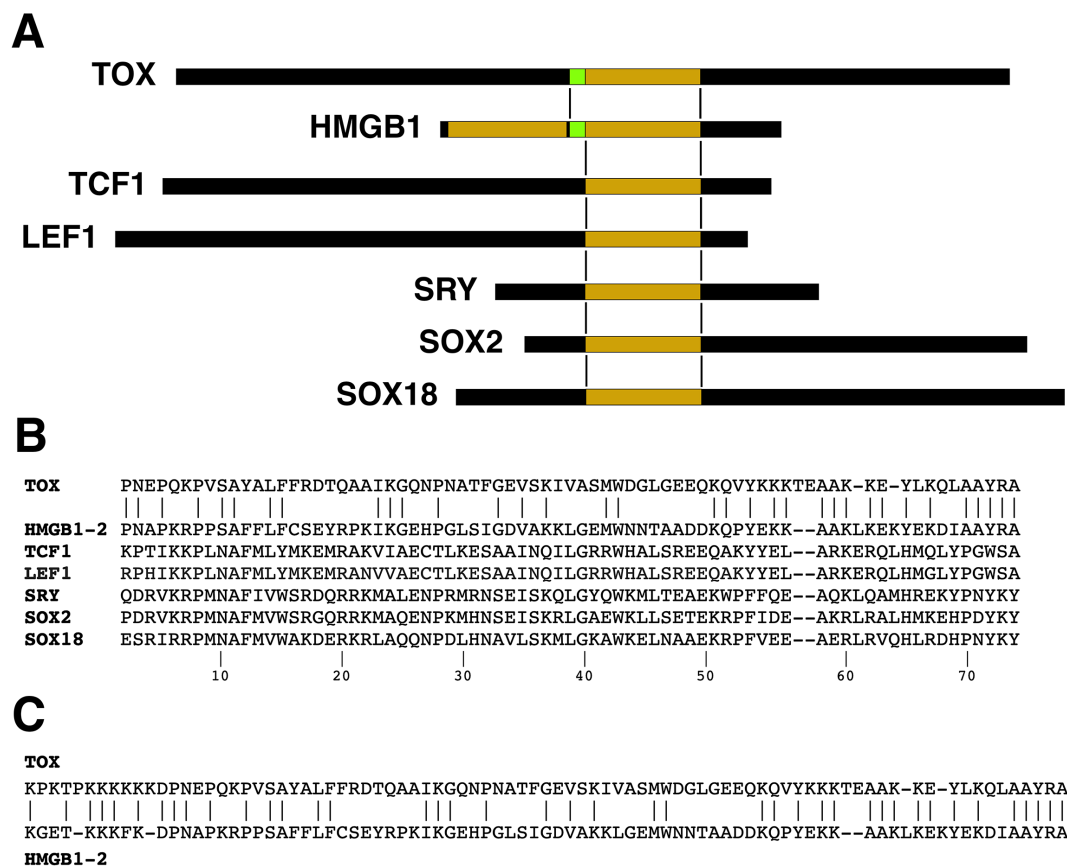


FIGURE 2

HMG domain alignments. (A) Schematic alignment of TOX with the second HMG domain of HMGB1 and a short segment of flanking sequence. Corresponding alignments of some other HMG-box proteins are also shown. The proteins lack substantial sequence homology with HMGB1 outside the HMG domain. (B) Sequence alignment of the HMG domains of human TOX, human HMGB1 (domain 2), human TCF1, human LEF1, human SRY, human SOX2, and human SOX18. The numbering of positions within the HMG domain follows (32). Note that other 'canonical' numbering systems have been used in the literature. (C) The TOX-HMGB1(domain 2) sequence alignment extends a short distance into the N-terminal region flanking the HMG domain.

matches the part of TOX that aligns with HMGB1 in the linker region and the second HMG box, and it extends five residues beyond the end of the TOX-HMGB1 alignment. Proteins with a similar region of alignment have since been predicted from some other insect genomes. These are frequently annotated as TOX, although homology to mammalian TOX proteins is lacking outside the indicated region, and there is currently no functional evidence linking this set of predicted proteins to TOX. The best tentative conclusion is that this constitutes an example of the reuse of a successful DNA-binding motif for different purposes during the evolution of some insect lineages and of vertebrates.

## Lessons applicable to vertebrate TOX

Recapitulating, the broad class of HMG box proteins does not form a protein family in the usual sense of having a common cellular function or functions. The HMG box has been reused often in evolution as a modular element incorporated into varied DNA-binding proteins and transcription factors. Therefore we can look to the HMG box of TOX for some understanding of its DNA binding characteristics, and we will do so here. The TOX HMG box in association with well conserved TOX N-terminal and C-terminal regions is a specialized adaptation of vertebrates. It will be necessary to look to these other regions of TOX for a more complete understanding of its biological and biochemical functions in vertebrate cells. We will touch only briefly on the questions that need to be addressed in this area.

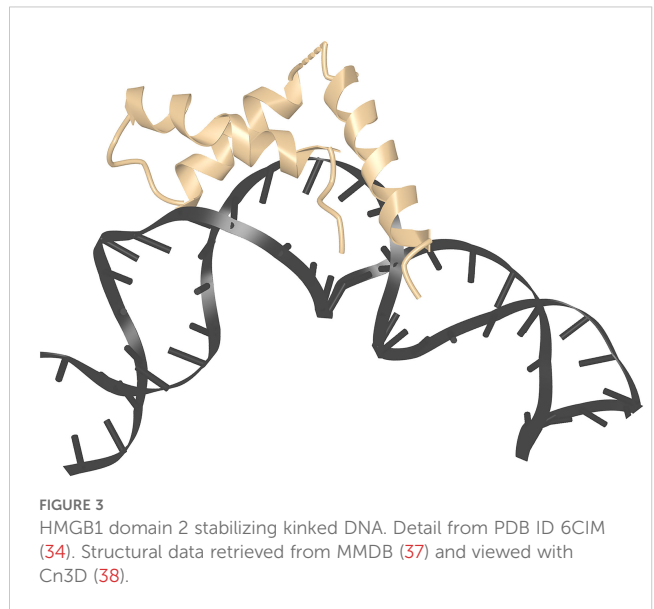
## HMGB1-(domain 2) stabilizes bent DNA

### HMGB1-(domain 2)

The structure of an HMGB1-(domain 2) complex with DNA has been determined as part of the RAG1/2 pre-reaction complex that forms on DNA prior to V(D)J recombination (34, 35). HMGB1 plays a supporting role in this process, since it is not essential for the coupled cleavage reaction catalyzed by recombinant RAG1/2 *in vitro*, but it can substantially increase the efficiency of the reaction (34, 36). HMGB1-(domain 2) contacts DNA in the minor groove and along the phosphate backbone. Its binding to DNA is associated with a localized distortion of DNA geometry, visible as widening of the minor groove and axial bending of the DNA helix by 90° (Figure 3). The local bending of DNA is supported indirectly by the linked HMGB1-(domain 1) and by interactions with RAG1/2 that specify the orientations of the distal parts of the recombining DNA segments, but only HMGB1-(domain 2) is in direct contact with DNA in the region of the bend.

### HMGB-D, SRY, SOX-family proteins, LEF1

X-ray crystal structures or NMR structures of the HMG domains of a *Drosophila* homolog of HMGB1, HMGB-D; SRY; the



SRY-related proteins SOX2, SOX9, and SOX18; and LEF1 exhibit similar sharp axial bending of the DNA centered on the protein binding site (39–45). The estimated extent of bending varies with the precise protein construct and experimental conditions, and can be influenced by lattice packing in protein-DNA crystals. Values observed in SRY-DNA crystal structures ranged from 69°–84°, and values for published SOX-DNA crystal structures, from 54°–76° (41, 44); measurements on protein-DNA complexes in solution indicated similar or even more pronounced bending (LEF1, 117°; SRY, 54°; and SOX, 101°) (46). These cases represent diverse HMG box sequences, each comprising a single HMG domain, with less identity than TOX to the defining HMGB1-(domain 2) sequence. All these proteins, and others not cited here, can stabilize the bending of otherwise linear DNA in solution or in crystals of simple protein-DNA complexes, under conditions that ensure no other proteins are involved, even indirectly, in DNA bending. These findings do not, of course, prove that the HMG-box proteins act alone to bend DNA in cells and at their physiological protein concentrations.

### Minor groove distortion and energetics of binding

Sharp bending of DNA is required in the course of many cellular processes such as DNA replication, DNA repair, and mRNA transcription, yet it runs counter to certain ingrained expectations. The DNA double helix is typically portrayed in the literature as a rather rigid rod, with a stated persistence length ~50 nm, or ~150 bp. On the other hand, the fact that DNA wraps tightly around the histone octamer of a nucleosome is a reminder that DNA can bend locally under the right conditions (47). Local stiffness is not, in fact, an intrinsic property of the DNA helix itself. At least in AT-rich segments of the helix, it is imparted by the presence of highly ordered water molecules in the double-helix



minor groove (48). The HMG domain is among several distinct protein adaptations that have evolved to bind in the minor groove, displace structured water, and enable the DNA flexibility needed for a variety of biological processes (49).

## Descriptions of DNA complexes

SRY serves as a textbook example of HMG-domain–DNA binding. The structural features underpinning specific binding are lucidly explained in (41). Overall, there is widening of the minor groove to accommodate the complexed protein, and narrowing of the adjacent major groove. Binding is supported by hydrogen bonds in the minor groove to DNA bases and by intraprotein hydrogen bonds that structure the protein–DNA interface, by van der Waals contacts in the minor groove, and by partial intercalation of an isoleucine side chain between DNA bases. Contacts on the DNA phosphate backbone attenuate the repulsion between negatively charged phosphates and thus support a narrowing of the major groove. The orientation of a few affected bases is twisted to accommodate the change from B-DNA geometry, but Watson–Crick base pairing is preserved. A high-resolution (1.75 Å) view of the closely related protein SOX18 bound to DNA shows a similar geometry and very similar hydrogen bonds (44). A slight difference is that arginine-20 of the SRY structure (termed arginine-18 in the SOX18 structure) retains the intraprotein hydrogen bonds that structure the interface, but has only van der Waals contact with DNA. The high resolution allows visualization of ordered water between that arginine side chain and DNA. Klaus et al. (44) discuss subtle rearrangements of protein–DNA contacts in other SOX-family crystal structures that accommodate variations in the sequence of the core DNA binding site. Other HMG domain–DNA structures display a similar or lower number of hydrogen bonds to DNA, and sometimes more reliance on van der Waals contacts, even to potential hydrogen-bond acceptors (39, 50). In

some cases, binding is supported by intercalation of protein side chains into DNA at a second neighboring dinucleotide site (39, 51).

## DNA bending is confirmed by physical measurements

DNA bending has been confirmed independently in many cases by physical measurements on HMG domain–DNA complexes in solution. The simplest approach conceptually has been to make FRET measurements on oligonucleotides with a centrally located binding site and FRET donor and acceptor fluorophores placed at opposite ends of the DNA. Distances are estimated for unbound linear DNA and protein-bound bent DNA, and are consistent with the structures. ‘Permutation gel electrophoresis’ has also been informative. DNA fragments complexed with an HMG-domain protein that induces a centrally located kink or bend migrate more slowly on an agarose gel than protein-bound linear DNA of the same length. As the HMG-domain binding site is moved from the center toward either end of the oligonucleotide, the mass of the protein–DNA complex is unchanged, but the DNA geometry changes, and electrophoretic migration reverts gradually to the pattern for the linear oligonucleotide. Migration of oligonucleotide standards with a known degree of bending can be compared to determine bend angles.

## TOX in this context

An NMR structure of mouse TOX in the absence of DNA is available in the Protein Data Bank (52), although the structure has never been published in a journal article. TOX has the same fold as other HMG box proteins discussed above (Figures 4A–C). Although it has been reported that the isolated TOX HMG domain does not bend DNA when it forms a complex with an unspecified oligonucleotide (53), this negative finding could



**FIGURE 4**

TOX closely resembles LEF1 in three-dimensional structure. **(A)** NMR structure of mouse TOX. From PDB ID 2CO9 (52). Structural data retrieved from MMDB (37) and viewed with Cn3D (38). **(B)** NMR structure of mouse LEF1, from PDB ID 2LEF (45). LEF1 was complexed with target DNA, but DNA is omitted from this view for direct comparison with the TOX structure in Figure 3A, Structural data retrieved from MMDB (37) and viewed with Cn3D (38). **(C)** NMR structure of mouse LEF1 in complex with DNA, from PDB ID 2LEF (45). Structural data retrieved from MMDB (37) and viewed with Cn3D (38).

indicate merely that DNA bending requires sequences outside the TOX HMG domain itself, as has been found with some other HMG box proteins, or it could mean that the oligonucleotide used was not a physiological target that fully engaged the TOX HMG domain. The question bears future investigation with authentic TOX sites. However— taking into account the high sequence conservation relative to HMGB1 domain 2, the conserved protein fold, and the consistent finding that HMG domains bind to bent DNA— the strong expectation is that TOX will bend DNA or will stabilize bent DNA when bound to its physiological target sites.

Recognition of target DNA sites

‘Sequence-specific’ versus ‘structure-specific’ HMG domains

HMG-domain proteins are commonly classified either as ‘sequence-specific’, that is, showing preference for binding a recognizable DNA consensus sequence; or as ‘structure-specific’, that is, having no experimentally recognizable DNA consensus sequence. The latter proteins are considered to select sites based mainly on the ability of DNA to assume a preferred kinked or bent geometry. This is an artificial distinction to an extent, first, because the consensus DNA sequences for the ‘sequence-specific’ factors SRY, SOX2, LEF1, and TCF1 center on AT-rich DNA that is intrinsically amenable to bending; and, second, because in binding the ‘sequence-specific’ factors do bend DNA. Conversely, ‘structure-specific’ factors have been shown to bind in certain cases at defined sites in DNA, rather than randomly, as exemplified in the RAG1/2 pre-reaction complex (34), or in the footprint of HMG-D on a DNA fragment from the *Drosophila ftz* locus upstream regulatory element (54), or in the footprints of the RNA polymerase I transcription factor UBF and the mitochondrial transcription factor TFAM on their target DNAs (55–57). It is therefore important that the frequent shorthand description of the structure-specific subset as ‘non-sequence-specific’ should not be read as conveying that the proteins are indifferent to DNA sequence.

Current practice is to assign these classifications based on the amino-acid residues present at a few telltale positions in the HMG domain (50), rather than from an experimental determination of DNA binding sites. Cases that have been closely examined indicate that there are subtle structural differences in the protein-DNA contacts of structure-specific *versus* sequence-specific proteins. The structure-specific group shows less reliance on direct hydrogen bonds to DNA bases, more reliance on water-mediated hydrogen bonds and van der Waals contacts, and a ‘looser’ protein-DNA interface (39). These structural differences come with corresponding differences in the thermodynamics of protein-DNA binding that are consistent with the looser interface (48). It is open to interpretation whether the two modes of protein-DNA interaction serve as two ways of achieving the same outcome of DNA minor-groove binding, as alternatives allowing a graded relaxation of DNA site specificity, or as a combination of those possibilities.

Highly abundant HMG-domain proteins

A grouping that is arguably more relevant to the cellular functions of HMG-domain proteins is according to their expression level. Mammalian HMGB1 and HMGB2 stand out in this regard. They are highly abundant proteins, with several million copies of HMGB1 per cell (58, 59) (Table 1), and are members of the structure-specific class. Evolution may have favored their ability to bind a range of linear DNA sequences and their availability at high protein levels in the cell nucleus because these features contribute to the efficiency of numerous processes that call for transient kinking of DNA. Examples are support for Ig locus recombination by the RAG1/2 complex, cited above, and action in concert with DNA topoisomerase to maintain the speed of transcription in long gene loci; as well as general cellular processes like DNA replication, DNA repair, and chromatin reorganization. The high abundance of HMGB1 and its rapid exchange at most binding sites in the cell nucleus (63) will help to ensure its ready availability at sites where it is needed.

HMG-domain proteins with similar high-level expression are found across species, typified by HMG-D in *Drosophila* and NHP6A and NHP6B in yeast. NHP6A is present at ~60,000 copies per yeast cell (64), which translates to a concentration in the yeast nucleus comparable to that of mammalian HMGB1 in the

TABLE 1 Protein abundances from PaxDb.

Human CD8 <sup>+</sup> T cells (60–62)	
HMGB1	5638 ppm
HMGB2	2071
LEF1	47.2
TCF1 (TCF7)	115
TOX	22.3
RUNX3 *	157
NFκB p65 (RELA) *	139
NFAT1 (NFATC2) *	148
<i>Drosophila</i> (61, 62)	
HMG-D	907 ppm
Yeast (61, 62)	
NHP6A	320 ppm
NHP6B	172
GCN4 *	10.6
GAL4 *	17.0
PHO4 *	8.97

\*Binds in the major groove of DNA.  
Values retrieved from PaxDb: Protein Abundance Database (<https://pax-db.org>) 31 December 2024. Ppm values are estimated representation of the indicated protein as a fraction of all protein molecules in the sample. Gene symbols are noted where they do not directly mirror the protein nomenclature. The database entry for human CD8<sup>+</sup> T cells is labelled ‘Cell line, Cd8’, but the source publication states clearly that the sample consisted of CD8<sup>+</sup> T cells isolated from human peripheral blood.

larger mammalian cell nucleus. The independent mass-spectrometric estimates of protein abundance in [Table 1](#) are consistent with the values just cited, after taking into account the typical volumes of a yeast cell ( $40\ \mu\text{m}^3$ ) and a human lymphocyte ( $200\ \mu\text{m}^3$ ) and the expected total protein density ( $\sim 3 \times 10^6$  proteins per  $\mu\text{m}^3$ ) ([65](#)). The seemingly depressed ppm values for NHP6A and NHP6B in [Table 1](#) compared to values for HMGB1 and HMGB2 reflect primarily the rather low nucleus/cytoplasm volume of *S. cerevisiae*— the nucleus is  $\sim 7\%$  of the cell volume ([66](#)) compared to more than half the cell volume in a T cell— and the consequent high representation of cytoplasmic proteins.

HMG-D, NHP6A, and NHP6B have single HMG domains, not a pair of domains as in HMGB1, but are thought to perform physiological functions akin to those of HMGB1 and HMGB2. Like HMGB1, they lack the signature residues of sequence-specific HMG-box proteins. Delivery of NHP6A to its sites of action may be assisted by one-dimensional facilitated diffusion of NHP6A along DNA ([67](#)), which could allow efficient scanning of DNA for relevant binding sites.

## Less abundant HMG-domain proteins

Most HMG-box proteins contrast with HMGB1 in having relatively low abundance in the cell. For example, LEF1, TCF1, and TOX are found at 50-fold to 250-fold lower protein levels than HMGB1 in human CD8<sup>+</sup> T cells ([Table 1](#)). These levels are comparable to the levels of major-groove-binding transcription factors ([Table 1](#)). Thus, most human HMG-box transcription factors share with the familiar major-groove-binding factors a vexing quantitative problem of finding their physiologically relevant binding sites in a sea of chromatin.

A rough calculation illustrates the dimensions of the problem. LEF1 and TCF1 have core consensus DNA sequence (A/T)(A/T)CAAAG (complement CTTTG(A/T)(A/T)) ([68](#), [69](#)), and SRY has the core consensus sequence ACAATG (complement CATTGT) ([41](#)). Calculating from the average GC content of 0.41 in human DNA, an exact match to each of these core sites would occur at 0.03–0.04% of positions in the human genome. Even allowing for some binding to inexact matches, target sites are still sparsely distributed in chromatin. The less abundant HMG-box transcription factors cannot afford to tarry on nonspecific sites. Two strategies they might borrow from the repertoire of HMGB and NHP6A are relatively rapid release from interactions with non-target sites and a DNA-scanning mechanism.

Although the sequence-specific factors do find certain target sites reliably and drive predictable physiological outputs, as illustrated in some of the examples below, it should not be expected that all sites matching the consensus sequence are robustly occupied in cells. A recent breakdown of observed TCF1 binding sites in CD8<sup>+</sup> T cells counted  $\sim 19,000$  high-confidence TCF1 ChIP-seq peaks, with  $\sim 5,000$  of those lacking TCF1/LEF1 motifs and thought to recruit TCF1 through its interactions with other transcription factors ([70](#)). The core sequence specifies an asymmetric site, so it is necessary to consider locations on either

strand of DNA, and the expected number of core TCF1 sites in the haploid human genome is  $\sim 0.04\%$  of  $6 \times 10^9$  bp, or  $\sim 2.4$  million sites. If attention is restricted to accessible chromatin— a few percent of the genome— the calculation indicates that there are, minimally, on the order of 50,000 accessible exact-match sites. There will be a considerably larger number of candidate binding sites after including sites with acceptable deviations from the consensus. The discrepancy can be accounted for to an extent by the presence of multiple copies of the core sequence in some peaks, but a visual scan through the genome confirms that isolated copies of the exact-match core sequence are not rare. The implication is that TCF1 and other sequence-specific HMG-box factors discriminate among potential binding sites based on DNA features beyond the core consensus sequence or cooperation with DNA-binding protein partners.

## Proposed TOX consensus site

TOX is assigned to the structure-specific group of HMG box proteins based on the signature distinguishing residues in its HMG domain ([33](#)). Nevertheless, Artegiani et al. stated that TOX binds preferentially to a GC-rich motif in HEK293T cells, based on labelling DNA sites using a TOX-DNA adenine methyltransferase (Dam) fusion protein and bioinformatic analysis ([71](#)). DamID-seq has the strength that it labels regions adjoining sites of protein-DNA interaction in live cells, and the limitation that labelling is integrated over a period of many hours— in this case 48h after infection with TOX-Dam lentivirus— and so could include repeated transient interactions in addition to stable TOX-Dam-DNA interaction sites ([72](#), [73](#)). TOX-Dam labelling was substantially increased at promoters ([71](#)), where GC-rich segments are common, and hence an alternative interpretation is that the bioinformatic enrichment merely reflected the abundance of promoter-proximal sequences in the dataset. Crucially, no data were offered showing preferential binding of TOX to the proposed consensus site. The recombinant TOX HMG domain did not bind to the site *in vitro* ([74](#)).

## Experimentally mapped TOX sites

There have been further attempts to define TOX binding sites experimentally. One of these utilized CUT&RUN mapping ([75](#), [76](#)) to delineate TOX binding sites in CD8<sup>+</sup> TILs that had infiltrated B16 melanomas ([77](#)). It is notable that the observed TOX signal often closely paralleled the H3K27Ac signal, even in housekeeping gene loci ([Figure 5](#)). TOX has been shown to coimmunoprecipitate with chromatin-binding and chromatin-remodeling proteins ([10](#)), so one plausible interpretation would be that TOX is prepositioned at promoters and enhancers to facilitate its normal cellular functions. A more skeptical view would be that TOX binding at its physiological DNA target sites is not stable during the prolonged CUT&RUN incubations of unfixed permeabilized cells, and that the observed binding at promoters and enhancers reflects a



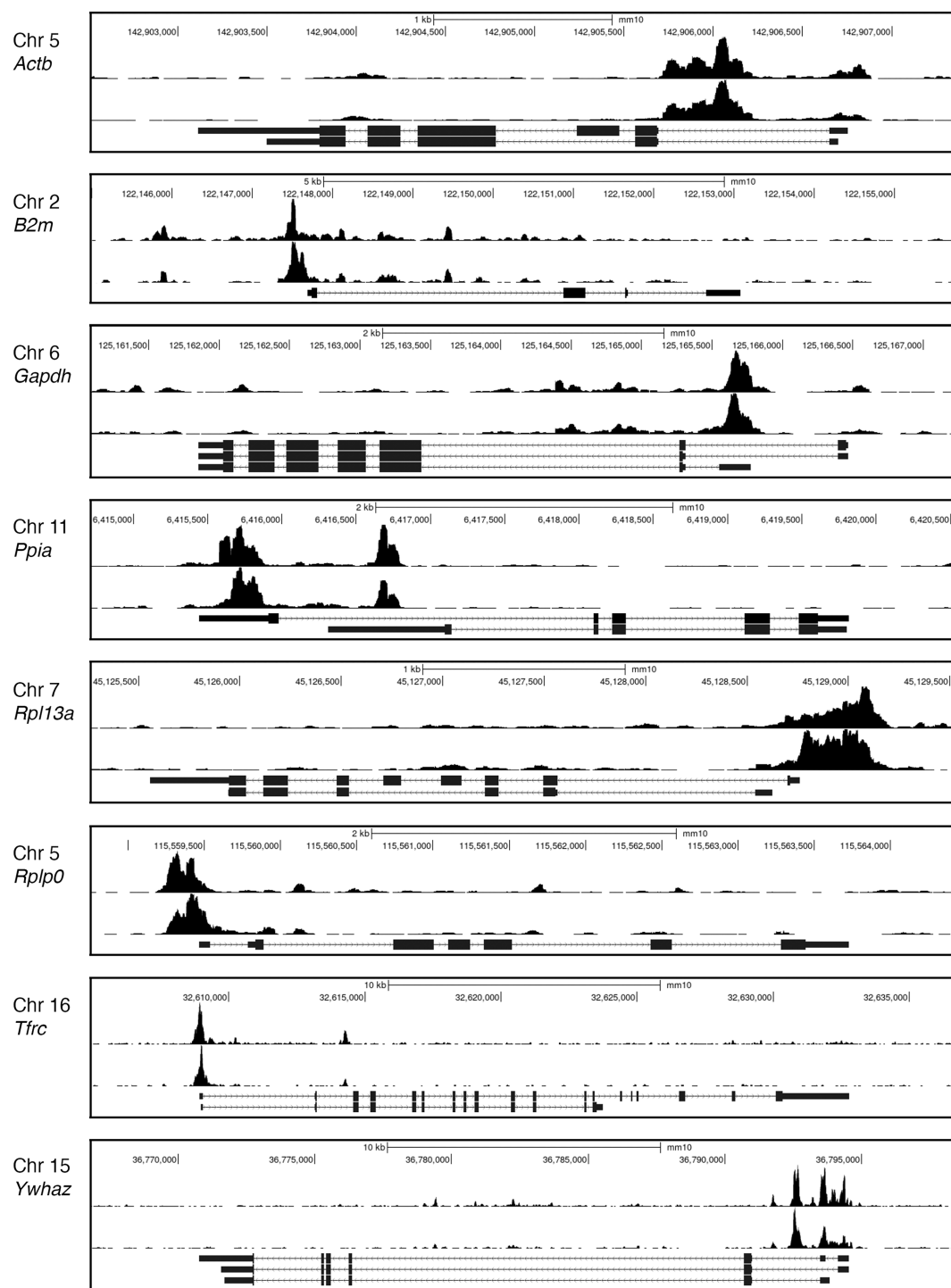


FIGURE 5

TOX CUT&RUN data. Representative Genome Browser views of the CUT&RUN signals for H3K27Ac (*upper track* in each panel) and TOX (*lower track*) at several housekeeping gene loci. The pronounced TOX CUT&RUN signals at promoters and enhancers of housekeeping genes in CD8<sup>+</sup> TILs raise two questions— whether these signals faithfully reflect TOX positioning in living TILs and, if so, whether this distribution of TOX can be related to control of the exhaustion transcriptional program. Data from (77, GEO accession GSE175443). The examples shown are for PD1<sup>high</sup>TIM3<sup>+</sup> TILs.

redistribution of TOX driven either by specific interactions with chromatin proteins or by nonspecific interactions with highly accessible chromatin.

Conventional TOX ChIP-seq identified candidate TOX peaks in CD8<sup>+</sup> T cells from a CNS autoimmune disease model (28). A

serious reservation in this case is that only single samples were analyzed for wildtype and for TOX<sup>-/-</sup> T cells. There are many instances where an apparent TOX ChIP peak from wildtype cells is mirrored by a substantial but somewhat smaller peak from TOX<sup>-/-</sup> cells, and, in the absence of experimental replicates, there is no way

to assess whether the higher signal in wildtype cells is statistically significant.

## Implications for TOX

What can be said confidently of TOX is that it is a low-abundance HMG box protein and has the protein-sequence fingerprints of a structure-selective binder. What is expected is that its HMG domain by itself will exhibit moderately weak binding to DNA sites, but that regions outside the HMG domain might make a further contribution to TOX-DNA affinity. One such region is the basic stretch immediately preceding the HMG domain, which bears sequence homology to a comparable stretch in HMGB1 that is thought to increase the affinity of rat HMGB1 for DNA (46). Like the more abundant structure-specific HMG-box factors, the TOX HMG domain is likely to be adaptable in its DNA site preferences, by minor rearrangements of its side chain contacts in the minor groove. As noted already for SOX-family proteins, this adaptability is a feature shared with at least some sequence-specific factors, although it may come to the fore in structure-specific factors due to their less-tight contacts on DNA.

TOX, like other transcription factors, has the biological imperative to recognize useful sites. In light of the low protein copy number, the need to discriminate relevant sites from irrelevant sites becomes particularly acute. In some respects, this is not a completely different situation from that of the sequence-specific factors, which in practice also select a subset from among possible binding sites. However, with no guidance from DNA sequence, the delineation of relevant TOX binding sites will rely entirely on experimental data. TOX target sites in DNA are likely to vary with the cell type and developmental context, and, correspondingly, TOX protein partners may differ in different contexts.

## Examples of varied HMG-box protein functions

### HMGB1 with RAG1/2 in V(D)J recombination

HMGB1 has been visualized binding to DNA along with RAG1/2 recombinase in the run-up to V(D)J recombination (34, 35) (Figure 3). The recombination signal sequences '12-RSS' and '23-RSS' mark sites for recombination in V, D, and J gene segments. HMGB1 is seen bound to a loop comprising the 23-bp spacer of the 23-RSS in the X-ray crystal structure. Despite its lack of DNA sequence selectivity and the consequent theoretical possibility of its binding anywhere, HMGB1-(domain 2) here chooses a specific site, implementing the biological imperative that the longer stretch of DNA in the 23-RSS spacer has to bend to position it correctly in the RAG1/2-DNA complex. HMGB1-(domain 1) binds adjacent— itself stabilizing a local bend in DNA in conjunction with DNA contacts made by RAG1/2— and is required for efficient positioning of domain 2. The 23-RSS spacer sequences themselves are not

strongly conserved in the 23-RSS segments of germline human and mouse immunoglobulin (Ig) loci, but Kim et al. note the frequent occurrence of bendable CA (complement TG) dinucleotides in a comparable position in these Ig gene segments, suggesting that evolution has provided an amenable local sequence appropriately positioned within this loop (34). This is a straightforward example of how the abundant HMG-box protein HMGB1 stabilizes a bent DNA conformation and enables efficient implementation of a cellular process.

### SOX2 in a transcription factor complex with OCT1

Low-abundance HMG box proteins can act as conventional site-specific transcription factors by forming complexes with protein partners. Drawing an example from the SOX family of transcription factors, whose roles in development have been reviewed in (78, 79), SOX-family proteins work together with partner proteins, including OCT-family proteins, to orchestrate diverse developmental processes (79). An X-ray crystal structure of the SOX2-OCT1 transcription factor complex on an *FGF4* enhancer and an NMR structure of the complex on a *Hoxb1* element (42, 80) (Figure 6A) portray the expected SOX-dependent bending of the DNA site, and protein-protein interfaces that are dictated by the relative geometries of bound SOX2 and OCT1 on the *FGF4* and *Hoxb1* DNAs. DNA bending is not directly involved in facilitating this protein-protein contact, but serves to narrow the minor groove beyond the SOX2 site, thus specifying the precise positioning of the C-terminal sequence flanking the SOX2 HMG domain. This placement of the C-terminal flanking sequence anchors the third  $\alpha$ -helix of the HMG domain, and either the C-terminal flanking sequence or the third  $\alpha$ -helix then presents as the binding surface for OCT1, depending on the geometry specified by the DNA site.

### LEF1 in a transcription factor complex

In another similar case, LEF1 acts at a defined TCR $\alpha$  minimal enhancer to promote mRNA expression (83). Here there is compelling biochemical evidence that LEF1 binding to the TCR $\alpha$  enhancer site acts in part by bringing together nearby binding sites for an Ets-1/RUNX1 complex and ATF/CREB, fostering assembly of a stable transcriptional complex (81, 82) (Figure 6B). Experimental evidence in these papers shows that DNA bending by LEF1 is a key to assembly of the full transcriptional complex. LEF1 also separately promotes transcription in this case via an action independent of DNA bending or CREB engagement in the complex (81, 82).

### SRY binding at a SOX9 enhancer

The sex-determining region Y protein, SRY, initiates the development of male gonads in mammals. Mechanistically, SRY

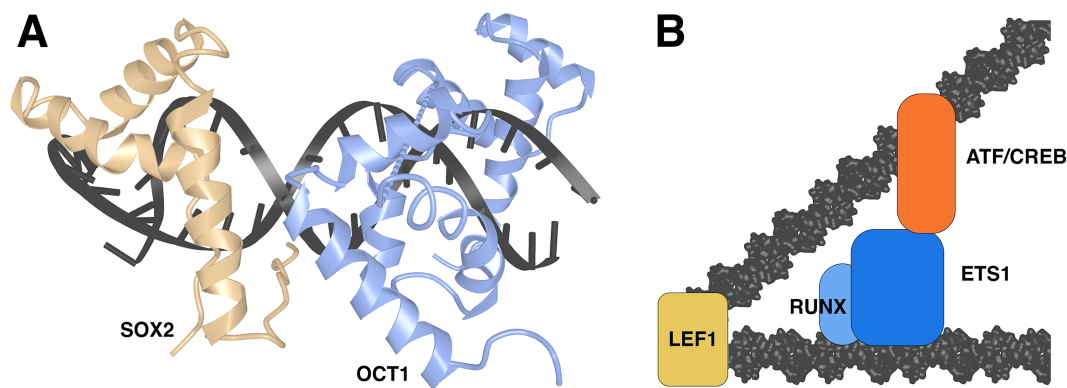


FIGURE 6

HMG-domain proteins in transcriptional complexes. (A) SOX2-OCT1 complex on DNA in the crystal structure PDB ID 1O4X (42). Structural data retrieved from MMDB (37) and viewed with Cn3D (38). (B) Schematic view of LEF1-ETS1-ATF/CREB complex. Based on (81, 82).

drives induction of the SRY-box transcription factor 9 gene (*Sox9*) in cells of the genital ridges in mouse embryos, during a defined time window between embryonic days 10.5 and 11.5, provided that a threshold level of SRY is achieved in those cells (reviewed in 84, 85). Subsequently, SOX9— an SRY homolog, as its name indicates — takes over to drive its own continuing expression and the differentiation of cells expressing SOX9 as Sertoli cells, which then orchestrate the differentiation of other classes of testicular cells. Several enhancers in the mouse *Sox9* locus can bind SRY, but one key enhancer site for the early induction of *Sox9* has been defined, *Enh13*, and testicular development is abrogated when this enhancer is deleted (86, 87). SRY occupies sites in *Enh13* at day 11.5, as shown by ChIP-qPCR of a MYC-tagged SRY transgene (86). SRY acts as a conventional transcription factor in this case, and whether DNA bending is central to its action has not been determined. However, the example highlights the two key points that timing can be essential to the action of an HMG-box transcription factor and that a faithful biological output may require that a threshold level of the transcription factor is reached (reviewed in 84). Further underscoring the existence of threshold concentrations, SOX9 is expressed naturally, but at lower levels, in XX-genotype genital ridges that go on to develop into ovaries (86); and overexpression of SOX9 redirects XX genital ridges to develop into testes (88, 89).

## SOX2 as a pioneer factor

The role of SOX-family transcription factors in development has been mentioned above. One specific mechanism through which they direct cellular developmental choices is as pioneer factors that initiate the opening of nucleosome-bound chromatin. Readers are referred to (90, 91) for an overview of pioneer transcription factors, and to (92) for a thoughtful review of the biology and mechanisms of SOX-family transcription factors. The pioneering function of SOX2 has begun to be captured in cryo-EM structures of SOX2 or its close sibling SOX11, or of a SOX2-OCT4 complex, engaging a nucleosome (93, 94). The two cases differ in the location of SOX2 binding relative

to the DNA entry and exit points on the nucleosome, raising the possibility that the detailed geometry of pioneer factor interactions may be specific to individual DNA sites, but a common feature is a local kink distorting the DNA to facilitate an initial partial release of DNA from its interaction with core histones.

## TCF1 in chromatin organization

TCF1 has a central role in the T cell lineage, working in conjunction with certain other transcription factors to enforce T cell-specific chromatin opening and T cell-specific gene expression in developing thymocytes (95, 96). TCF1 continues to ensure the normal functioning of mature T cells, not just by regulating the activity of promoters and enhancers in individual gene loci, but by specifying the three-dimensional organization of chromatin in the nucleus (70, 97, 98). Some mechanistic insight into this process has come from an investigation of how TCF1 ‘preprograms’ gene expression in CD8<sup>+</sup> central memory T cells (T<sub>CM</sub>) (97). In those experiments, mice engineered to express Cre recombinase under the control of a *Gzmb* promoter rapidly deleted *Tcf7* upon activation of mature *Tcf7*<sup>FL/FL</sup> CD8<sup>+</sup> T cells. The T cells mounted a normal effector response to a first LCMV Armstrong infection, even in the face of the concurrent deletion of *Tcf7*, and produced normal initial numbers of memory-precursor cells. However, the resulting TCF1-deficient memory cells exhibited a diminished response upon restimulation. The data suggested strongly that the recall response involved chromatin rearrangements within topologically associating domains that were ‘preprogrammed’ in resting memory cells by TCF1 occupancy of binding sites in gene promoters and distal regulatory regions, including specific TCF1 binding sites in glycolytic-pathway gene loci and in the *Id3* locus. Editing the latter sites, individually or in combinations, would provide a definitive test of whether they are essential to specify a TCF1-determined chromatin architecture in resting memory T cells and to prefigure the chromatin reorganization and gene expression during a recall response.

## The path to TOX mechanisms

This discussion holds several lessons for research on the cellular mechanisms of TOX in T cell exhaustion. Since TOX is not uniquely concerned with exhaustion, the first clear lesson is that it will be crucial to conduct the experiments during the initiation and progression of CD8<sup>+</sup> T cell exhaustion or under conditions that mirror the progression to exhaustion. The correct conditions will ensure an amenable chromatin state and the presence of relevant protein partners. Mechanistic experiments will need to give attention to the timing and cellular concentration of TOX, in light of the perturbing effect of partial knockdown or haploinsufficiency of TOX in T cell exhaustion models (8, 10), and the quantitative rather than qualitative difference in TOX expression in CNS-infiltrating CTLs that cause disease compared to CTLs that do not cause disease (28, Figure 2D). A central task will be to catalog TOX binding sites in chromatin that contribute to the exhaustion program. Then, for individual sites in or adjacent to exhaustion gene loci, it will be a priority to define the TOX protein partners at those sites, keeping in mind that both the HMG domain and TOX regions beyond the HMG domain may participate in recruiting partner transcription factors or chromatin modulators. An ancillary question will be whether DNA kinking or bending is central to assembly of protein complexes at specific sites, or, conversely, how DNA bending might thwart binding of some transcription factors. The default expectation has been that TOX acts as a conventional transcription factor, enhancing or suppressing gene expression at individual gene loci. Current evidence is not decisive on whether TOX controls the overall transcriptional program of exhaustion by acting at many loci, or controls expression of a small number of key gene targets that then secondarily implement the program, or is concerned primarily with the expression of genes that promote survival and expansion of the exhausted T cell population. TOX might have roles other than as a conventional transcription factor, including an unexplored pioneer function or a role in orchestrating the overall three-dimensional architecture of chromatin as cells progress into the exhausted state. There is no shortage of work still to be done exploring TOX mechanisms.

## Author contributions

PH: Conceptualization, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing.

BV: Formal analysis, Writing – review & editing. LA: Formal analysis, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The authors declare that financial support was received from the National Institutes of Health for the research and/or publication of this article. The laboratory receives funding from Kyowa Kirin North America in support of a project unrelated to this publication. No financial support was received from Kyowa Kirin North America for the research and/or publication of this article.

## Acknowledgments

The author's work in this area is supported by National Institutes of Health grants AI040127 and AI109842 and by an Institutional allocation from the La Jolla Institute for Immunology.

## Conflict of interest

PH is a founder and scientific advisor of CalciMedica, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

1. Martinez GJ, Pereira RM, Äijö T, Kim EY, Marangoni F, Pipkin ME, et al. The transcription factor NFAT promotes exhaustion of activated CD8<sup>+</sup> T cells. *Immunity*. (2015) 42:265–78. doi: 10.1016/j.immuni.2015.01.006
2. Mogno GP, Spreafico R, Wong V, Scott-Browne J, Togher S, Hoffmann A, et al. Exhaustion-associated regulatory regions in CD8<sup>+</sup> tumor-infiltrating T cells. *Proc Natl Acad Sci USA*. (2017) 114:E2776–85. doi: 10.1073/pnas.1620498114
3. Philip M, Fairchild L, Sun L, Horste EL, Camara S, Shakiba M, et al. Chromatin states define tumor-specific T cell dysfunction and reprogramming. *Nature*. (2017) 545:452–6. doi: 10.1038/nature22367
4. Chen J, López-Moyado I, Seo H, Lio C-WJ, Hempelman LJ, Sekiya T, et al. NR4A transcription factors limit CAR T cell function in solid tumors. *Nature*. (2019) 567:530–4. doi: 10.1038/s41586-019-0985-x



5. Tirosh I, Izar B, Prakadan SM, Wadsworth MH2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. (2016) 352:189–96. doi: 10.1126/science.aad0501
6. Scott-Browne JP, López-Moyado IF, Trifari S, Wong V, Chavez L, Rao A, et al. Dynamic changes in chromatin accessibility occur in CD8<sup>+</sup> T cells responding to viral infection. *Immunity*. (2016) 45:1327–40. doi: 10.1016/j.immuni.2016.10.028
7. Pauken KE, Sammons MA, Odorizzi PM, Manne S, Godec J, Khan O, et al. Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science*. (2016) 354:1160–5. doi: 10.1126/science.aaf2807
8. Seo H, Chen J, González-Avalos E, Samaniego-Castruita D, Das A, Wang YH, et al. TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose CD8<sup>+</sup> T cell exhaustion. *Proc Natl Acad Sci USA*. (2019) 116:12410–5. doi: 10.1073/pnas.1905675116
9. Scott AC, Dündar F, Zumbo P, Chandran SS, Klebanoff CA, Shakiba M, et al. TOX is a critical regulator of tumour-specific T cell differentiation. *Nature*. (2019) 571:270–4. doi: 10.1038/s41586-019-1324-y
10. Khan O, Giles JR, McDonald S, Manne S, Ngiew SF, Patel KP, et al. TOX transcriptionally and epigenetically programs CD8<sup>+</sup> T cell exhaustion. *Nature*. (2019) 571:211–8. doi: 10.1038/s41586-019-1325-x
11. Alfei F, Kanev K, Hofmann M, Wu M, Ghoneim HE, Roelli P, et al. TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. *Nature*. (2019) 571:265–9. doi: 10.1038/s41586-019-1326-9
12. Yao C, Sun HW, Lacey NE, Ji Y, Moseman EA, Shih HY, et al. Single-cell RNA-seq reveals TOX as a key regulator of CD8<sup>+</sup> T cell persistence in chronic infection. *Nat Immunol*. (2019) 20:890–901. doi: 10.1038/s41590-019-0403-4
13. Utzschneider DT, Gabriel SS, Chisanga D, Gloury R, Gubser PM, Vasanthakumar A, et al. Early precursor T cells establish and propagate T cell exhaustion in chronic infection. *Nat Immunol*. (2020) 21:1256–66. doi: 10.1038/s41590-020-0760-z
14. Seo H, Gonzalez-Avalos E, Zhang W, Ramchandani P, Yang C, Lio C-WJ, et al. BATF and IRF4 cooperate to counter exhaustion in tumor-infiltrating CAR T cells. *Nat Immunol*. (2021) 22:983–95. doi: 10.1038/s41590-021-00964-8
15. Espinosa-Carrasco G, Chiu E, Scivo A, Zumbo P, Dave A, Betel D, et al. Intratumoral immune triads are required for immunotherapy-mediated elimination of solid tumors. *Cancer Cell*. (2024) 42:1202–16. doi: 10.1016/j.ccell.2024.05.025
16. Sen DR, Kaminski J, Barnitz RA, Kurachi M, Gerdemann U, Yates KB, et al. The epigenetic landscape of T cell exhaustion. *Science*. (2016) 354:1165–9. doi: 10.1126/science.aae0491
17. Maurice NJ, Berner J, Taber AK, Zehn D, Prlic M. Inflammatory signals are sufficient to elicit TOX expression in mouse and human CD8<sup>+</sup> T cells. *JCI Insight*. (2021) 6:e150744. doi: 10.1172/jci.insight.150744
18. Sekine T, Perez-Potti A, Nguyen S, Gorin J-B, Wu VH, Gostick E, et al. TOX is expressed by exhausted and polyfunctional human effector memory CD8<sup>+</sup> T cells. *Sci Immunol*. (2020) 5:eaba7918. doi: 10.1126/sciimmunol.aba7918
19. Kurd NS, He Z, Louis TL, Milner JJ, Omilusik KD, Jin W, et al. Early precursors and molecular determinants of tissue-resident memory CD8<sup>+</sup> T lymphocytes revealed by single-cell RNA sequencing. *Sci Immunol*. (2020) 5:eaa6894. doi: 10.1126/sciimmunol.aa6894
20. Doering TA, Crawford A, Angelosanto JM, Paley MA, Ziegler CG, Wherry EJ. Network analysis reveals centrally connected genes and pathways involved in CD8<sup>+</sup> T cell exhaustion versus memory. *Immunity*. (2012) 37:1130–44. doi: 10.1016/j.immuni.2012.08.021
21. Wilkinson B, Chen JY-F, Han P, Rufner KM, Goularte OD, Kaye J. TOX: and HMG box protein implicated in the regulation of thymocyte selection. *Nat Immunol*. (2002) 3:272–80. doi: 10.1038/ni767
22. Aliahmad P, O'Flaherty E, Han P, Goularte OD, Wilkinson B, Satake M, et al. TOX provides a link between calcineurin activation and CD8 lineage commitment. *J Exp Med*. (2004) 199:1089–99. doi: 10.1084/jem.20040051
23. Aliahmad P, Kaye J. Development of all CD4 T lineages requires nuclear factor TOX. *J Exp Med*. (2008) 205:245–56. doi: 10.1084/jem.20071944
24. Aliahmad P, Kadavallore A, de la Torre B, Kappes D, Kaye J. TOX is required for development of the CD4 T cell lineage gene program. *J Immunol*. (2011) 187:5931–40. doi: 10.4049/jimmunol.1101474
25. Aliahmad P, de la Torre B, Kaye J. Shared dependence on the DNA-binding factor TOX for the development of lymphoid issue-inducer cell and NK cell lineages. *Nat Immunol*. (2010) 11:945–52. doi: 10.1038/ni.1930
26. Seehus CR, Aliahmad P, de la Torre B, Iliev ID, Spurka L, Funari VA, Kaye J. The development of innate lymphoid cells requires TOX-dependent generation of a common innate lymphoid cell progenitor. *Nat Immunol*. (2015) 16:599–608. doi: 10.1038/ni.3168
27. Wu C-J, Cho S, Huang H-Y, Lu C-H, Russ J, Cruz LO, et al. MiR-23~27~24-mediated control of humoral immunity reveals a TOX-driven regulatory circuit in follicular helper T cell differentiation. *Sci Adv*. (2019) 5:eaa1715. doi: 10.1126/sciadv.aaw1715
28. Page N, Klimek B, De Roo M, Steinbach K, Soldati H, Lemeille S, et al. Expression of the DNA-binding factor TOX promotes the encephalitogenic potential of microbe-induced autoreactive CD8<sup>+</sup> T cells. *Immunity*. (2018) 48:937–50. doi: 10.1016/j.immuni.2018.04.005
29. Page N, Lemeille S, Vincenti I, Klimek B, Mariotte A, Wagner I, et al. Persistence of self-reactive CD8<sup>+</sup> T cells in the CNS requires TOX-dependent chromatin remodeling. *Nat Commun*. (2021) 12:1009. doi: 10.1038/s41467-021-21109-3
30. Chu T, Wu M, Hoellbacher B, de Almeida GP, Wurmser C, Berner J, et al. Precursors of exhausted T cells are preemptively formed in acute infection. *Nature*. (2025). 640:782–92. doi: 10.1038/s41586-024-08451-4
31. McManus DT, Valanparambil RM, Medina CB, Scharer CD, McGuire DJ, Sobierajska E, et al. An early precursor CD8 T cell that adapts to acute or chronic viral infection. *Nature*. (2025). 640:772–81. doi: 10.1038/s41586-024-08562-y
32. Laudet V, Stehelin D, Clevers H. Ancestry and diversity of the HMG box superfamily. *Nucleic Acids Res*. (1993) 21:2493–2501. doi: 10.1093/nar/21.10.2493
33. O'Flaherty E, Kaye J. TOX defines a conserved subfamily of HMG-box proteins. *BMC Genomics*. (2003) 4:13. doi: 10.1186/1471-2164-4-13
34. Kim MS, Chuenchor W, Chen X, Cui Y, Zhang X, Zhou ZH, et al. Cracking the DNA code for V(D)J recombination. *Mol Cell*. (2018) 70:358–70. doi: 10.1016/j.molcel.2018.03.008
35. Chen X, Cui Y, Best RB, Wang H, Zhou ZH, Yang W, et al. Cutting antiparallel DNA strands in a single active site. *Nat Struct Mol Biol*. (2020) 27:119–26. doi: 10.1038/s41594-019-0363-2
36. van Gent DC, Hiom K, Paull TT, Gellert M. Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO J*. (1997) 16:2665–70. doi: 10.1093/emboj/16.10.2665
37. Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, et al. MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucl Acids Res*. (2014) 42:D297–303. doi: 10.1093/nar/gkt1208
38. Wang Y, Geer LY, Chappay C, Kans JA, Bryant SH. Cn3D: sequence and structure views for Entrez. *Trends Biochem Sci*. (2000) 25:300–2. doi: 10.1016/S0968-0004(00)01561-9
39. Murphy FVIV, Sweet RM, Churchill MEA. The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition. *EMBO J*. (1999) 18:6610–8.
40. Murphy EC, Zhurkin VB, Louis JM, Cornilescu G, Clore GM. Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation. *J Mol Biol*. (2001) 312:481–99. doi: 10.1006/jmbi.2001.4977
41. Racca JD, Chen Y-S, Brabender AR, Battistin U, Weiss MA, Georgiadis MM. Role of nucleobase-specific interactions in the binding and bending of DNA by human male sex determination factor SRY. *J Biol Chem*. (2024) 300:107683. doi: 10.1016/j.jbc.2024.107683
42. Williams DC, Cai M, Clore GM. Molecular basis for synergistic transcriptional activation by Oct1 and Sox2 revealed from the solution structure of the 42-kDa Oct1-Sox2-Hoxb1-DNA ternary transcription factor complex. *J Biol Chem*. (2004) 279:1449–57. doi: 10.1074/jbc.M309790200
43. Vivekanandan S, Moovarkumudalvan B, Lescar J, Kolatkar PR. PDB Entry 4S2Q. Crystal structure of HMG domain of the chondrogenesis master regulator, Sox9 in complex with ChIP-seq identified DNA element. (2015). doi: 10.2210/pdb4S2Q/pdb.
44. Klaus M, Prokoph N, Girbig M, Wang X, Huang YH, Srivastava Y, et al. Structure and decoy-mediated inhibition of the SOX18/Prox1-DNA interaction. *Nucleic Acids Res*. (2016) 44:3922–35. doi: 10.1093/nar/gkw130
45. Love JJ, Li X, Case DA, Giese K, Grosschedl R, Wright PE. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*. (1995) 376:791–5. doi: 10.1038/376791a0
46. Dragan AI, Read CM, Makeyeva EN, Milgotina EI, Churchill ME, Crane-Robinson C, et al. DNA binding and bending by HMG boxes: energetic determinants of specificity. *J Mol Biol*. (2004) 343:371–93. doi: 10.1016/j.jmb.2004.08.035
47. Cutter A, Hayes JJ. A brief review of nucleosome structure. *FEBS Lett*. (2015) 589:2914–22. doi: 10.1016/j.febslet.2015.05.016
48. Privalov PL, Dragan AI, Crane-Robinson C. The cost of DNA bending. *Trends Biochem Sci*. (2009) 34:464–70. doi: 10.1016/j.tibs.2009.05.005
49. Bewley CA, Gronenborn AM, Clore GM. Minor groove-binding architectural proteins: structure, function, and DNA recognition. *Annu Rev Biophys Biomol Struct*. (1998) 27:105–31. doi: 10.1146/annurev.biophys.27.1.105
50. Murphy FV IV, Churchill MEA. Nonsequence-specific DNA recognition: a structural perspective. *Structure*. (2000) 8:R83–9. doi: 10.1016/S0969-2126(00)00126-X
51. Churchill MEA, Klass J, Zoetewey DL. Structural analysis of HMGD-DNA complexes reveal influence of intercalation on sequence selectivity and DNA bending. *J Mol Biol*. (2011) 403:88–102.
52. Li H, Saito K, Koshiba S, Inoue M, Kigawa T, Yokoyama S, et al. PDB Entry 2CO9. Solution structure of the HMG-box domain of thymus high mobility group box protein TOX from mouse. (2005). doi: 10.2210/pdb2co9/pdb.
53. Aliahmad P, Seksenyan A, Kaye J. The many roles of TOX in the immune system. *Curr Opin Immunol*. (2011) 24:173–177. doi: 10.1016/j.coi.2011.12.001
54. Churchill MEA, Jones DNM, Glaser T, Hefner H, Searles MA, Travers AA. HMG-D is an architecture-specific protein that preferentially binds to DNA containing the dinucleotide TG. *EMBO J*. (1995) 14:1264–75. doi: 10.1002/j.1460-2075.1995.tb07110.x



55. Bell SP, Pikaard CS, Reeder RH, Tjian R. Molecular mechanisms governing species-specific transcription of ribosomal RNA. *Cell*. (1989) 59:489–97. doi: 10.1016/0092-8674(89)90032-9
56. Fisher RP, Parisi MA, Clayton DA. Flexible recognition of rapidly evolving promoter sequences by mitochondrial transcription factor 1. *Genes Dev*. (1989) 3:2202–17. doi: 10.1101/gad.3.12b.2202
57. Parisi MA, Clayton DA. Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. *Science*. (1991) 252:965–9. doi: 10.1126/science.2035027
58. Wu L, Rechsteiner M, Kuehl L. Comparative studies on microinjected high-mobility-group chromosomal proteins, HMG1 and HMG2. *J Cell Biol*. (1981) 91:488–96. doi: 10.1083/jcb.91.2.488
59. Kuehl L, Salmond B, Tran L. Concentrations of high-mobility-group proteins in the nucleus and cytoplasm of several rat tissues. *J Cell Biol*. (1984) 99:648–54. doi: 10.1083/jcb.99.2.648
60. Kim M-S, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map of the human proteome. *Nature*. (2014) 509:575–81. doi: 10.1038/nature13302
61. Wang M, Herrmann CJ, Simonovic M, Szklarczyk D, von Mering C. Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics*. (2015) 15:3163–8. doi: 10.1002/pmic.201400441
62. Huang Q, Szklarczyk D, Wang M, Simonovic M, von Mering C. PaxDb 5.0: curated protein quantification data suggests adaptive proteome changes in yeasts. *Mol Cell Proteomics*. (2023) 22:100640. doi: 10.1016/j.mcpro.2023.100640
63. Sapojnikova N, Maman J, Myers FA, Thorne AW, Vorobyev VI, Crane-Robinson C. Biochemical observation of the rapid mobility of nuclear HMG1. *Biochim Biophys Acta*. (2005) 1729:57–63. doi: 10.1016/j.bbexp.2005.03.002
64. Paull TT, Carey M, Johnson RC. Yeast HMG proteins NHP6A/B potentiate promoter-specific transcriptional activation *in vivo* and assembly of preinitiation complexes *in vitro*. *Genes Dev*. (1996) 10:2769–81. doi: 10.1101/gad.10.21.2769
65. Milo R. What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays*. (2013) 35:1050–5. doi: 10.1002/bies.201300066
66. Jorgenson P, Edgington NP, Schneider BL, Rupeš I, Tyers M, Fletcher B. The size of the nucleus increases as yeast cells grow. *Mol Biol Cell*. (2007) 18:3523–32. doi: 10.1091/mbc.e06-10-0973
67. Kamagata K, Ouchi K, Tan C, Mano E, Mandal S, Wu Y, et al. The HMG chromatin protein Nhp6A can bypass obstacles when traveling on DNA. *Nucleic Acids Res*. (2020) 48:10820–31. doi: 10.1093/nar/gkaa799
68. Waterman ML, Fischer WH, Jones KA. A thymus-specific member of the HMG protein family regulates the human T cell receptor C $\alpha$  enhancer. *Genes Dev*. (1991) 5:656–69. doi: 10.1101/gad.5.4.656
69. van de Wetering M, Clevers H. Sequence-specific interaction of the HMG-box factor TCF-1 occurs within the minor groove of a Watson/Crick double helix. *EMBO J*. (1992) 11:3039–44. doi: 10.1002/j.1460-2075.1992.tb05374.x
70. Shan Q, Li X, Chen X, Zeng Z, Zhu S, Gai K, et al. Tcf1 and Lef1 provide constant supervision to mature CD8<sup>+</sup> T cell identity and function by organizing genomic architecture. *Nat Commun*. (2021) 12:5863. doi: 10.1038/s41467-021-26159-1
71. Artegiani B, de Jesus Domingues AM, Bragado Alonso S, Brandl E, Massalini S, Dahl A, et al. Tox: a multifunctional transcription factor and novel regulator of mammalian corticogenesis. *EMBO J*. (2015) 34:896–910. doi: 10.15252/embj.201490061
72. van Steensel B, Henikoff S. Identification of *in vivo* DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol*. (2000) 18:424–8. doi: 10.1038/74487
73. Vogel MJ, Peric-Hupkes D, van Steensel B. Detection of *in vivo* protein-DNA interactions using DamID in mammalian cells. *Nat Protoc*. (2007) 2:1467–78. doi: 10.1038/nprot.2007.148
74. Seehus CR, Kaye J. The role of TOX in the development of innate lymphoid cells. *Mediators Inflammation*. (2015) 2015:243868. doi: 10.1155/2015/243868
75. Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*. (2017) 6:e21856. doi: 10.7554/eLife.21856
76. Skene PJ, Henikoff JG, Henikoff S. Targeted *in situ* genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc*. (2018) 13:1006–19. doi: 10.1038/nprot.2018.015
77. Rhodes Ford B, Vignali PDA, Rittenhouse NL, Scharping NE, Peralta R, Lontos K, et al. Tumor microenvironmental signals reshape chromatin landscapes to limit the functional potential of exhausted T cells. *Sci Immunol*. (2022) 7:eabj9123. doi: 10.1126/sciimmunol.abj9123
78. Sarkar A, Hochedlinger K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell*. (2013) 12:15–30. doi: 10.1016/j.stem.2012.12.007
79. Kamachi Y, Kondoh H. Sox proteins: regulators of cell fate specification and differentiation. *Development*. (2013) 140:4129–44. doi: 10.1242/dev.091793
80. Reményi A, Lins K, Nissen LJ, Reinbold R, Schöler HR, Wilmanns M. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev*. (2003) 17:2048–59. doi: 10.1101/gad.269303
81. Giese K, Kingsley C, Kirshner JR, Grosschedl R. Assembly and function of a TCR $\alpha$  enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev*. (1995) 9:995–1008. doi: 10.1101/gad.9.8.995
82. Mayall TP, Sheridan PL, Montminy MR, Jones KA. Distinct roles for P-CREB and LEF-1 in TCR $\alpha$  enhancer assembly and activation on chromatin templates *in vitro*. *Genes Dev*. (1997) 11:887–99. doi: 10.1101/gad.11.7.887
83. Ho I, Leiden J. Regulation of the human T-cell receptor  $\alpha$  gene enhancer: Multiple ubiquitous and T-cell-specific nuclear proteins interact with four hypomethylated enhancer elements. *Mol Cell Biol*. (1990) 10:4720–7.
84. Kashimada K, Koopman P. Sry: the master switch in mammalian sex determination. *Development*. (2010) 137:3921–30. doi: 10.1242/dev.048983
85. Svingen T, Koopman P. Building the mammalian testis: origins, differentiation, and assembly of the component cell populations. *Genes Dev*. (2013) 27:2409–26. doi: 10.1101/gad.228080.113
86. Gonen N, Futtner CR, Wood S, Garcia-Moreno SA, Salamone IM, Samson SC, et al. Sex reversal following deletion of a single distal enhancer of Sox9. *Science*. (2018) 360:1469–73. doi: 10.1126/science.aas9408
87. Ridnik M, Abberbock E, Alipov V, Lhermann SZ, Kaufman S, Lubman M, et al. Two redundant transcription factor binding sites in a single enhancer are essential for mammalian sex determination. *Nucleic Acids Res*. (2024) 52:5514–28. doi: 10.1093/nar/gkae178
88. Bishop CE, Whitworth DJ, Qin Y, Agoulnik AI, Agoulnik IU, Harrison WR, et al. A transgenic insertion upstream of sox9 is associated with dominant XX sex reversal in the mouse. *Nat Genet*. (2000) 26:490–4. doi: 10.1038/82652
89. Vidal VP, Chaboissier MC, de Rooij DG, Schedl A. Sox 9 induces testis development in XX transgenic mice. *Nat Genet*. (2001) 28:216–7. doi: 10.1038/90046
90. Zaret KS. Pioneer transcription factors initiating gene network changes. *Annu Rev Genet*. (2020) 54:367–85. doi: 10.1146/annurev-genet-030220-015007
91. Orsetti A, van Oosten D, Vasarhelyi R-G, Danescu T-M, Huertas J, van Ingen H, et al. Structural dynamics in chromatin unraveling by pioneer transcription factors. *Biophys Rev*. (2024) 16:365–82. doi: 10.1007/s12551-024-01205-6
92. Julian LM, McDonald ACH, Stanford WL. Direct reprogramming with SOX factors: masters of cell fate. *Curr Opin Genet Dev*. (2017) 46:24–36. doi: 10.1016/j.gde.2017.06.005
93. Dodonova SO, Zhu F, Dienemann C, Taipale J, Cramer P. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature*. (2020) 580:669–72. doi: 10.1038/s41586-020-2195-y
94. Michael AK, Grand RS, Isbel L, Cavadini S, Kozicka Z, Kempf G, et al. Mechanisms of OCT4-SOX2 readout on nucleosomes. *Science*. (2020) 368:1460–5. doi: 10.1126/science.abb0074
95. Johnson JL, Georgakilas G, Petrovic J, Kurachi M, Cai S, Harly C, et al. Lineage-determining transcription factor TCF-1 initiates the epigenetic identity of T cells. *Immunity*. (2018) 48:243–57. doi: 10.1016/j.immuni.2018.01.012
96. Emmanuel AO, Arnovitz S, Haghi L, Mathur PS, Mondal S, Quandt J, et al. TCF-1 and HEB cooperate to establish the epigenetic and transcriptional profiles of CD4<sup>+</sup>CD8<sup>+</sup> thymocytes. *Nat Immunol*. (2018) 19:1366–78. doi: 10.1038/s41590-018-0254-4
97. Shan Q, Hu SS, Zhu S, Chen X, Badinovac VP, Peng W, et al. Tcf1 preprograms the mobilization of glycolysis in central memory CD8<sup>+</sup> T cells during recall responses. *Nat Immunol*. (2022) 23:386–98. doi: 10.1038/s41590-022-01131-3
98. Shan Q, Zhu S, Chen X, Liu J, Yuan S, Li X, et al. Tcf1-CTCF cooperativity shapes genomic architecture to promote CD8<sup>+</sup> T cell homeostasis. *Nat Immunol*. (2022) 23:1222–35. doi: 10.1038/s41590-022-01263-6