PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
XXXX, XX(XX): 1~20
www.pibb.ac.cn

旋转定位调控先锋转录因子与核小体结合的体内外差异*

刘国庆[1,2)**] 郭星悦[1,2)] 苍 婧[1,2)] 张 智[1,2)] 刘国君[1,2)]

([1)] 内蒙古科技大学生命科学与技术学院，包头 014010；[2)] 内蒙古科技大学内蒙古自治区生命健康与生物信息学重点实验室，包头 014010）

摘要 目的 先锋转录因子（pioneer transcription factor，PTF）能够识别并结合核小体DNA，启动染色质开放和基因表达，在胚胎发育、细胞重编程及肿瘤发生等过程中发挥关键作用。然而，核小体旋转定位调控PTF与核小体的相互作用机制目前尚不明确。**方法** 本研究基于DNA形变能模型，探究DNA旋转定位在转录因子与核小体相互作用中的调控作用。**结果** 发现体外环境中，SOX7和P53等转录因子的结合强烈依赖于其识别基序在核小体上的旋转方位。然而，对8种PTF在体内环境中的分析表明，PTF结合的基序与未结合的基序在核小体上呈现出总体一致的旋转定位倾向，提示在体内环境下旋转方位并非调控PTF结合的关键决定因素。此现象在细胞重编程和胚胎干细胞分化过程中同样存在。PTF在体内能够结合被核小体包埋的基序上，可能是因为借助PTF的结构特性和核小体呼吸作用等因素来克服结合表面的空间位阻。**结论** 本研究揭示了DNA旋转定位在体内外环境中对转录因子结合的差异化调控，强调了PTF通过超越旋转定位的机制来主导染色质开放性的独特能力。

关键词 先锋转录因子，核小体，旋转定位，DNA形变能，染色质可及性
中图分类号 Q61 **DOI**：10.3724/j.pibb.2025.0434 **CSTR**：32369.14.pibb.20250434

先锋转录因子（pioneer transcription factors，PTFs）可以识别封闭异染色质中的目标DNA序列，初步打开核小体为非PTFs和染色质重塑复合物等提供可接近性，并可以触发该区域染色质结构的重塑，最终为特定基因的选择性表达提供必要条件[1]。PTFs与核小体中的DNA靶位点相互作用，触发DNA-核小体相互作用的松弛，从而影响染色质的开放状态和基因表达。通常而言，启动子或增强子区域的染色质开放性会受到PTFs的调控[2-3]。凭借该功能，PTFs在诸多生物学过程中扮演着关键角色。尤其，它在发育与肿瘤进展中的作用日益受到关注。

胚胎发育过程中，特定的转录因子（transcription factor，TF）通过调控基因表达来决定发育的方向和形态。PTFs的提出为此提供了一种潜在的调控机制。第一个被发现具有先锋作用的TF是叉头框蛋白（forkhead box protein，FOX）A，它可以结合核小体DNA中的靶序列[4]。这种能力使得PTFs能够与那些不易接近的DNA序列结合，与其他TF形成鲜明对比。在FOXA之后，其他具有先锋潜能的TF也逐渐被发现，例如八聚体结合转录因子4（octamer-binding transcription factor 4，OCT4）、SRY盒转录因子2（SRY-box transcription factor 2，SOX2）和Krüppel样因子4（Krüppel-like factor 4，KLF4）等[5-6]。这些因子在胚胎干细胞的自我更新和多能性维持中起重要的作用。在胚胎发育早期，OCT4和SOX2能够结合并主动打开封闭染色质，启动基因转录，从而促进多能性的建立和维持。胚胎发育的后期，SOX2与染色质的相互作用促进了胚胎干细胞从原始态多能性向形成态多能性的转变，为后续的细胞分化奠定了基

* 国家自然科学基金（62161043），内蒙古自然科学基金（2025MS06029）和2025年内蒙古自治区生命健康与生物信息学重点实验室项目（2025KYPT0135）资助。

** 通讯联系人。
Tel：15148991105，E-mail：gqliu1010@163.com

收稿日期：2025-10-02，接受日期：2025-12-03

础[5, 7]。特别是在诱导多能干细胞（induced pluripotent stem cells，iPSCs）的研究中，OCT4、SOX2和KLF4可以将成纤维细胞转化为iPSCs或神经元。SOX2和OCT4与KLF4与其核小体中的DNA靶点之间相互作用，能够产生稳定的复合物，使得DNA与核小体组蛋白之间的相互作用变形。而且，SOX2-OCT4二聚体的核小体结合能力比单个PTF更强[5]。PTFs之间的相互作用，如SOX2和OCT4与GATA3的结合，已通过体外研究得到广泛证实[8]。此外，PTFs在胚胎器官发育中也起到决定性作用。研究表明，FOXA TF家族的成员在内胚层衍生器官（如肝脏、胰腺、肺和前列腺）的正常发育中起着至关重要的作用[9]。在造血系统中，ETS家族转录因子PU.1（E26 transformation-specific family transcription factor PU.1）和GATA结合蛋白（GATA binding protein，GATA）1为关键PTFs，分别参与调控白细胞分化、免疫反应以及红细胞分化过程中的基因表达[10]。另有研究表明，在不同的细胞中，PTFs会被招募到不同的目标结合位点，而在某些目标结合位点的募集并不总能改变染色质的结构，这类位点被称为“PTF抗性位点”。例如，成对框转录因子7（paired box transcription factor 7，PAX7）不仅有垂体发育和促黑色素细胞发育的先锋活性，还在肌肉中也具有关键的发育活性。在垂体细胞中，尽管PAX7调控肌肉发育的结合位点是已知的，但这些位点的结合并不能发挥PTFs的作用[11]。更特别的，在肌源性祖细胞中PAX7促进垂体发育的结合位点都不能作为PTF抗性位点进行募集[11]。这一现象突显了细胞环境在调控PTFs活性中的重要作用。

研究表明，有些PTFs在肿瘤发生发展中也扮演着重要角色。例如，FOXA1在前列腺癌中发生高频突变，通过改变雄激素受体信号、诱导上皮-间质转化等机制影响肿瘤进展与治疗抵抗[12]。FOXA1与FOXA2在缺乏雄激素受体（androgen receptor，AR）的前列腺癌中共同定位于染色质，并协同驱动肿瘤的AR非依赖性生长[13]。FOXA2促进小细胞肺癌（small cell lung cancer，SCLC）多器官转移，其在转移相关肿瘤中高表达；FOXA2通过激活胎儿肺神经内分泌基因表达程序，赋予肿瘤细胞广泛的转移能力；ASCL1作为SCLC肿瘤发生的关键调控因子，直接结合FOXA2启动子和增强子区域，调控其表达，形成ASCL1－FOXA2轴驱动转移[14]。GATA6通过结合特异性增强子、与CCCTC结合因子（CCCTC-binding factor，CTCF）等蛋白质相互作用，调控三维基因组结构（如影响染色质环结构和增强子-启动子的相互作用），进而调控致癌基因的表达，驱动癌细胞增殖[15]。c-Myc作为一个在肿瘤发生中广泛参与的PTF[16]，它的存在与否直接影响其结合位点区域的染色质结构[17]。

综上所述，PTFs在肿瘤发生和细胞命运决定中的作用不容忽视。PTFs通过精确调控染色质的状态和基因的表达，对胚胎发育的各个阶段产生影响。深入理解PTFs的作用机制，不仅有助于揭示细胞命运决定和癌症等疾病的发展过程中的复杂基因调控网络，而且有可能为相关疾病开发新的治疗策略提供重要的分子靶点。本文重点探究PTFs与核小体的相互作用机制。

利用结构生物学和多组学（如基因组学、转录组学、表观遗传学等）的多种方法，研究TF与核小体结合的机制和功能，取得了丰富进展[18-21]。例如，研究人员开发了一种新方法——核小体连续亲和纯化-指数富集配体的系统进化技术（nucleosome consecutive affinity purification－systematic evolution of ligands by exponential enrichment，NCAP-SELEX），用于分析核小体对TF-DNA结合的影响，探索了220种代表不同结构家族的TF与核小体之间的相互作用[18]，发现大多数TF与核小体DNA的结合能力低于自由DNA。尽管核小体在一般情况下对TF与DNA的结合有抑制作用，但这种影响在不同TF之间有很大差异：一些TF能够更有效地结合到核小体DNA，尤其是靠近核小体DNA末端或在溶剂暴露侧的周期性位置上。一些TF能够同时结合核小体DNA的两个螺旋。核小体的存在打破了DNA的旋转对称性，导致TF在核小体DNA上的结合取向存在偏好性。**该**研究揭示了TF与核小体之间的多样化相互作用，包括跨越两个DNA螺旋的结合、取向偏好、末端偏好、周期性偏好和对二分位点的偏好。这些发现为理解TF如何在染色质环境中调控基因表达提供了新的视角，并为未来的研究提供了基于生化原理的转录调控机制的基础。然而，该研究并没有深入探究TF结合位点处的旋转定位对TF结合能力的影响。

PTF与染色质的相互作用过程中，核小体定位情况非常重要。核小体DNA的旋转定位描述的是DNA双螺旋在核小体上的取向，即DNA双螺旋表

面与组蛋白之间的相对位形。Cui等[22]发现，当P53结合位点嵌入核小体时，调控细胞周期停滞基因（cell cycle arrest genes， CCA）的结合位点和细胞凋亡相关基因（apoptosis associated genes，Apo）的结合位点在核小体表面上的可及性不同：CCA结合位点更容易暴露在核小体表面，而Apo结合位点则倾向于被埋在核小体内部，提示TF结合位点的旋转定位使得它们容易被P53识别并激活与细胞周期停滞相关的基因，而Apo结合位点则可能需要额外的步骤（如招募染色质重塑复合体）来暴露。体外实验也表明，核小体上的P53结合位点的旋转定位是决定P53是否与其有效结合的关键[23]。

虽然PTFs与核小体结合的研究已经取得了重要进展，但不同的PTFs与核小体的相互作用可能具有很高的差异性。例如，在核小体的DNA上，DNA柔性、弯曲、基序等多个因素对核小体的旋转定位产生不同的影响，进而调控TF与核小体的相互作用。DNA旋转定位究竟如何影响TF与核小体的相互作用，目前尚缺乏深入而广泛的研究。我们在前期工作中，开发了核小体定位的DNA形变能模型，该模型能非常准确地预测核小体旋转定位[24-25]。本文利用该模型[24]，系统研究DNA序列特异性和旋转定位究竟如何影响PTFs与核小体的结合。

1　数据与方法

1.1　NCAP–SELEX数据处理

从文献［18］获得体外TF与核小体结合的数据（https：//www. ebi. ac. uk/ena/browser/view/PRJEB22684）。NCAP-SELEX技术首先构建一个包含大量随机DNA序列的文库（称之为DNA配体）。其次将整个随机DNA序列文库与提纯的组蛋白八聚体（含链霉亲和素结合肽标签的组蛋白2A（histone 2A， H2A））混合，组装核小体，然后与纯化的TF孵育。通过链霉亲和素磁珠捕获核小体，再通过镍磁珠捕获TF结合物。进行5轮选择，每轮后对TF结合的DNA配体进行PCR扩增和高通量测序。分析第四轮选择后的DNA文库的测序结果（第四次循环纯化后达到最大结合程度），它代表与TF结合亲和力和特异性最强的核小体DNA。本文利用这套体外实验数据，研究TF与核小体相互作用中核小体旋转定位的影响。这套数据包含2个DNA文库，分别是147 bp（lig147）和200 bp（lig200）的随机合成序列，它们分别包含101或154 bp的随机序列，其两端是长度为24 bp和22 bp的固定配体序列：CCCTACACGACGCTCTTCCGATCT和AGATCGGAAGAGCACACGTCTG。最终获得193个TFs与核小体DNA（147 bp）相互作用的数据和154个TFs与核小体DNA（200 bp）相互作用的数据。每种TF对应一套双端测序后DNA文库，代表与该TF结合的核小体DNA序列。

从ENA数据库下载的147 bp（lig 147）和200 bp（lig 200）的序列为二代测序数据，本文采用Pandaseq工具进行序列的拼接工作[26]。在操作中，将最小重合度设定为5 bp，以确保拼接结果的可靠性和准确性。同时，为了避免数据质量的影响，还进行了含N序列的去除操作。使用MEME套件中的单基序查找软件（find individual motif occurrences，FIMO）[27]从拼接后的序列搜索确定TF结合基序的位置。为了确保FIMO软件能够正常工作，把从文献获取的pfm格式的基序矩阵[18]通过Jaspar2meme工具转换为了meme格式，以满足FIMO软件的分析需求。用FIMO搜索完特定TF对应的DNA文库中的基序后，计算长度为200 bp的序列上每个核苷酸位点被基序覆盖的次数，将其定义为基序富集分数。

在NCAP-SELEX技术多轮选择后的DNA文库测序基础上，研究人员通过计算每两个非重叠位置（3-mer宽）的序列分布之间的互信息（enriched 3-mer pair based mutual information， EMI）来检测TF结合事件。TF结合会导致特定3-mer对（如连续或间隔的6-mer）的分布相关，从而在EMI热图中显示信号。本文用文献［18］中提供的EMI指标来定量描述TF在DNA序列上结合的富集程度。EMI值越高，则对应的TF在该位置上结合的富集程度越高。

为了检测200 bp序列的不同位置上基序富集度与真实的TF结合的关联性，首先用峰值检测法来识别基序富集度的峰以及TF结合的峰。其次，用R语言findOverlaps函数检查两种峰之间的吻合情况：若两种峰之间的重叠区间在任一峰区间中的所占分值（长度百分比）大于70%，则定义二者之间“重叠”；若两类峰之间无任何重叠区域，则定义二者之间“不重叠”。

1.2 峰值检测方法

1.2.1 数据预处理

峰值检测前首先分别将基序富集度和TF富集度分值，用最大最小值归一化。其次，用平滑函数smooth.spline（x，spar=0.3）平滑归一化后的数据，这有助于过滤掉随机波动造成的峰。最后用平滑数据作为输入，识别峰。

1.2.2 峰值模式识别

使用正则表达式"［+］{2，5}［0］{0,}［-］{2，5}"定义峰值形态，其中各项含义如下：

［+］{2，5}：峰值上升阶段需要2~5个连续上升的数据点

［0］{0,}：允许0个或多个平稳阶段（峰值平台）

［-］{2，5}：峰值下降阶段需要2~5个连续下降的数据点

1.2.3 连续性要求

upN=3：峰值前至少连续上升3个点

downN=3：峰值后至少连续下降3个点

确保检测到的是真实的峰值，而非噪声波动

1.2.4 高度筛选

min峰height=0.2：峰值高度必须≥0.2

排除幅度过小的波动

1.2.5 距离筛选

min峰distance=10：相邻峰值最小间距10个数据点

避免在同一个峰附近重复检测

1.2.6 数量限制

n峰s=5：最多返回5个最显著的峰值

sortstr=T：按显著性排序输出

以上参数，是反复人工检查识别的峰图谱后确定的优化参数，能够有效识别真实的峰。

1.3 体内先锋因子DNA结合位点数据来源

本文中还研究了8种先锋因子（OCT4、SOX2、KLF4、GATA4、肌分化因子1（myogenic differentiation 1，MYOD1）、FOXA1、CCAAT/增强子结合蛋白α（CCAAT/enhancer-binding protein alpha，CEBPA）以及ASCL1）在体内的结合情况。从jaspar数据库（https：//jaspar.elixir.no/）下载了这些TF DNA结合基序在人类基因组（hg38-version）上的位点。从ReMap数据库（https：//remap.univ-amu.fr）获得这8种先锋因子在人类基因组（hg38）上的结合位点（基于ChIP-seq的TF结合峰值位点峰信息）。这些结合位点是用染色质免疫沉淀测序（chromatin immunoprecipitation sequencing，ChIP-seq）等实验技术确定的TF在多种人类细胞类型中DNA结合位点的合集。获得TF的基序在人类基因组上的位点和真实结合位点之后，将TF基序位点根据是否与其真实结合位点重叠分为两组：TF结合的基序位点（TF-bound）和TF不结合的基序位点（TF-unbound），前者是完全落入ChIP-seq峰区间的基序位点，后者是与ChIP-seq峰没有任何重叠的基序位点。在此基础上，根据基序位点信息，从人类基因组上截取以基序为中心的长度为500 bp的DNA序列，用于DNA弯曲能计算。

1.4 实验核小体定位数据处理

从基因表达综合数据库（Gene Expression Omnibus，GEO）获得IMR90细胞的微球菌核酸酶测序（micrococcal nuclease sequencing，MNase-seq）原始数据（GSM543311）。该数据包含7个样本的MNase-seq单端测序数据。用bowtie2[28]将测序fastq文件比对到GRCh38基因组上，获得sam文件。用samtools获得比对后的bam文件。用samtools[29]质控（samtools view -q 30 -F 4 -F 256），过滤掉低质量读段，包括去除低质量比对（映射质量（mapping quality，MAPQ）< 30）、去除未比对序列、去除非唯一比对。samtools markdup等函数去除PCR重复读段。最后，将7个样本对应的bam文件用samtools merge函数整合成一个bam文件。用deeptools[30]的bamCoverage函数将bam文件转化为bigwig文件，此过程中使用关键参数（extendReads 147）将单端测序的读段延长至147 bp。最后用deeptools的computeMatrix reference-point函数获得TF结合峰周围的核小体占据信号，并用plotHeatmap函数可视化。plotHeatmap函数的参数中可设置聚类参数（kmeans 4），将TF结合峰位点聚成4个簇，分别为被核小体占据、上游被核小体占据、下游被核小体占据、无核小体占据四类。

获得人胚胎干细胞（human embryonic stem cell，hESC）和人神经外胚层细胞（human neuroectodermal cell，hNEC）细胞的MNase-seq单端测序数据（GSM1973978、GSM1973979），各有2个生物学重复。获得原始序列读段归档（sequence read archive，SRA）文件后，用fastq-dump转换为fastq文件。后续的质控、比对等过程如上所述，最后获得整合生物学重复的bam文件。

基于bam文件，利用DANPOS工具[31]分别获得hESC和hNEC中核小体在染色体上位置。

在hESC向hNEC分化的过程中，SOX2分别结合hESC和hNEC的结合位点信息从公共数据（GSE76082）获得。我们将hg19参考基因组版本的SOX2结合位点用LiftOver（默认参数）转换为hg38版本的位点，以此来保证与基序、核小体位置的版本相统一。

为了探究在hESC向hNEC分化的过程中SOX2在核小体上的结合是否依赖于基序的旋转定位，我们进行如下分析：a. 用MNase-seq数据分析工具DANPOS分别获得hESC和hNEC中的核小体位置（数据来源：GSE76083和GSE76084），以及分化过程中核小体定位发生动态变化的相关信息（如核小体位置滑动、核小体占据率变化、核小体定位模糊度发生变化等）；b. 其次，获得落入hESC和hNEC中的核小体区域的SOX2基序位点（从jasper数据库获得，基于hg38版本）；c. 进一步将核小体区域的基序根据是否被SOX2结合（由ChIP-seq的结合峰判断）分成“SOX2结合”和“SOX2不结合”两组；d. 最后分析这两组基序的DNA弯曲能，比较二者的旋转定位是否有差异。

1.5　DNA弯曲能的计算

DNA弯曲能在核小体旋转定位的预测方面表现出非常优异的性能，而且DNA弯曲能在核小体占据率、核小体组装自由能的预测中同样有不错的表现[24-25]。DNA弯曲能模型以DNA双螺旋几何结构的准确表示方法和形变能的计算原理为两大主要基础。DNA双螺旋结构采用碱基对阶梯模型（base-pair step model）描述[32]。假定DNA是一条连续弯曲的弹性杆，则DNA弯曲能用胡克定律计算。DNA的弯曲主要由描述DNA结构的转角（roll）和倾角（tilt）的变化导致，因此假定一段DNA序列形成核小体时，其核小体中心129 bp区域的弯曲能用如下公式计算[24]：

$$\boldsymbol{E}_b = \sum_{i=1}^{128}\left\{\frac{1}{2}\boldsymbol{k}_\rho(\boldsymbol{i})\left[\boldsymbol{\rho}(\boldsymbol{i})-\boldsymbol{\rho}_0(\boldsymbol{i})\right]^2+\frac{1}{2}\boldsymbol{k}_\tau(\boldsymbol{i})\left[\boldsymbol{\tau}(\boldsymbol{i})-\boldsymbol{\tau}_0(\boldsymbol{i})\right]^2\right\} \quad (1)$$

式中，$\rho_0(i)$和$\tau_0(i)$是从DNA-蛋白质复合物晶体结构估算的平衡参数；$\rho(i)$和$\tau(i)$是假定DNA序列缠绕组蛋白八聚体形成核小体时该序列的结构参数，该结构参数利用核小体DNA晶体结构的结构约束（如全局曲率）估算；$k_\rho(i)$和$k_\tau(i)$是基于蛋白质-DNA复合物结构估算的二核苷酸依赖性力常数，通过6个碱基对阶梯参数的协方差矩阵的求逆计算获得。最后，将129 bp的DNA序列（对应128个碱基对阶梯）的弯曲能除以128后，得到每个碱基对步长的平均弯曲能。

1.6　DNA弯曲能的傅里叶变换分析

为了分析旋转定位强度和相位，对DNA弯曲能进行傅里叶变换，计算10-bp周期性的幅度和相位。例如，读取TF结合区域和非结合区域的DNA弯曲能数据，并以每条序列的中心点为基准，截取两侧各50 bp的区域（共101 bp），确保分析窗口对称。用R语言fft函数对每条序列的弯曲能数据进行快速傅里叶变换（fast Fourier transform， FFT），将信号从时域转换到频域。计算频域信号的幅度和相位，提取对应于10-bp周期性的频率成分。准确而言，在频域中识别每条序列在10-bp周期附近的最大频率成分，记录其对应的周期、幅度和相位。该相位代表每条序列弯曲能数据的首个数据对应的相位，因此对获得的相位进行校正，使其代表序列中心位置（如基序中心位置）的相位，并将相位标准化到0~2π范围内。

基于标准化后的相位值，采用两种聚类方案对序列的旋转定位模式进行分类。

四分类方案：

mode 1：［0， π/4）∪［7π/4， 2π）

mode 2：［3π/4， 5π/4）

mode 3：［π/4， 3π/4）

mode 4：［5π/4， 7π/4）

二分类方案：

MG-outward：［0， π/2）∪［3π/2， 2π），DNA小沟背向组蛋白

MG-inward：［π/2， 3π/2），DNA小沟面向组蛋白

为了检验DNA弯曲能相位（FFT标准化后的相位）是否在TF结合和不结合的两组之间存在显著差异，本文使用置换检验（permutation test）。置换检验时，首先对对两组样本进行顺序上的随机置换（即重排），并重新计算两组之间的相位均值差，把上述过程重复多遍（本文重复1 000遍），构造出统计量（即相位均值差）的经验分布，然后对比两组样本的实际统计量和构造出的统计量的经验分布，也就是计算构造出的统计量中不小于实际统计量的比例，即为P值。这个比例越小于0.05，表

明实际统计量大于随机置换统计量的可能性越高，即越显著。值得注意的是，相位是圆形数据，因此标准化后的相位首先用R语言circular包的circular函数将其转变为圆形数据再计算相位均值。而且，计算两组之间的相位均值差时须将输出值通过周期性变换限定在0~π之间。

相位平均合成向量长度（mean resultant length）表示圆形数据在圆周上的集中程度，其计算公式如下：

$$Rho=\sqrt{(\frac{1}{n}\sum_{i=1}^{n}\cos\theta_i)^2+(\frac{1}{n}\sum_{i=1}^{n}\sin\theta_i)^2} \quad (2)$$

其中θ_i是各个数据点的相位，n是样本大小。*Rho*=1说明所有相位数据都集中在同一个方向上；*Rho*=0说明所有相位数据随机均匀分布在圆周上。用R语言circular包的rho.circular函数计算相位平均合成向量长度。

2　结果

2.1　体外TF与核小体结合对旋转定位的依赖性

为了探究体外TF-核小体结合是否受核小体旋转定位的影响，我们分析了体外NCAP-SELEX技术产生的数据。简言之，该技术产生的原始数据是TF结合亲和力和特异性很强的核小体DNA序列，包括近200个TF结合的核小体DNA序列文库，每种TF对应一套双端测序后DNA文库，详细信息见“数据与方法”部分。

用DNA弯曲能模型判断核小体旋转定位[24]，其判断规则如图1所示。通俗而言，弯曲能极小值的位置是DNA小沟暴露于核小体（即DNA大沟面向组蛋白）的位置[24-25]。

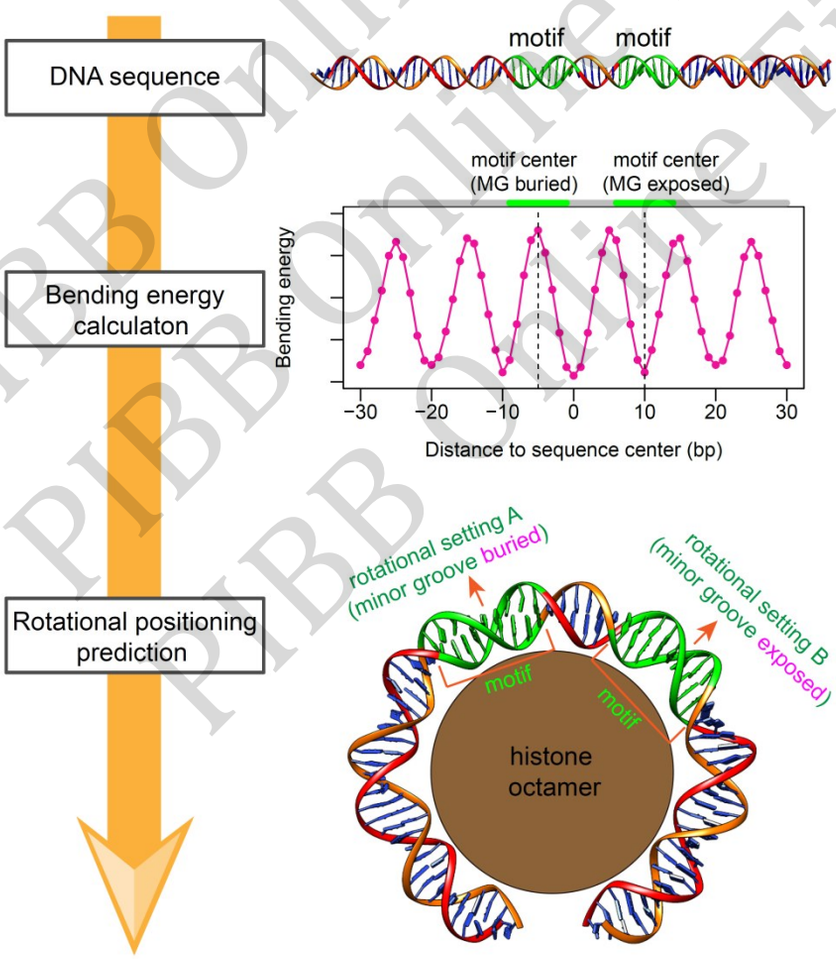


Fig.1　Schematic for determining the rotational orientation of TF−bound motifs using DNA bendability

A motif center is determined as having a “preference for minor groove exposed on the nucleosome” when its DNA bendability exhibits a local minimum. Conversely, it is determined as having a “preference for minor groove buried in the nucleosome” when the bendability exhibits a local maximum.

如果TF与DNA的结合受到模体旋转方位的干扰，则可以预期基序富集的峰位置上不会出现TF结合的EMI峰（即TF较少结合）。基于nuc-lig200 bp的数据，本文确实筛选到SOX7属于这一情形（图2a），即SOX7的基序富集的一个峰位置上并没有SOX7结合（没有EMI峰出现）。同时，SOX7的另外两个基序峰位置上有SOX7结合。通过计算这两类基序峰周围的DNA弯曲能并比较，发现两者之间的弯曲能图谱差异明显：没被SOX7结合的基序位置上平均DNA弯曲能呈现较强的10-bp周期性震荡（图2b），提示这些基序在核小体DNA上的相位一致性较高；在SOX7结合的基序峰位置上，虽然基序的相位并不十分统一（振幅较小），但能够清楚地观察到TF结合基序和TF未结合基序的平均相位相反（图2b），即旋转方位相反。这说明，SOX7的DNA基序在核小体DNA上的旋转方位的不同确实影响SOX7与基序的有效结合。

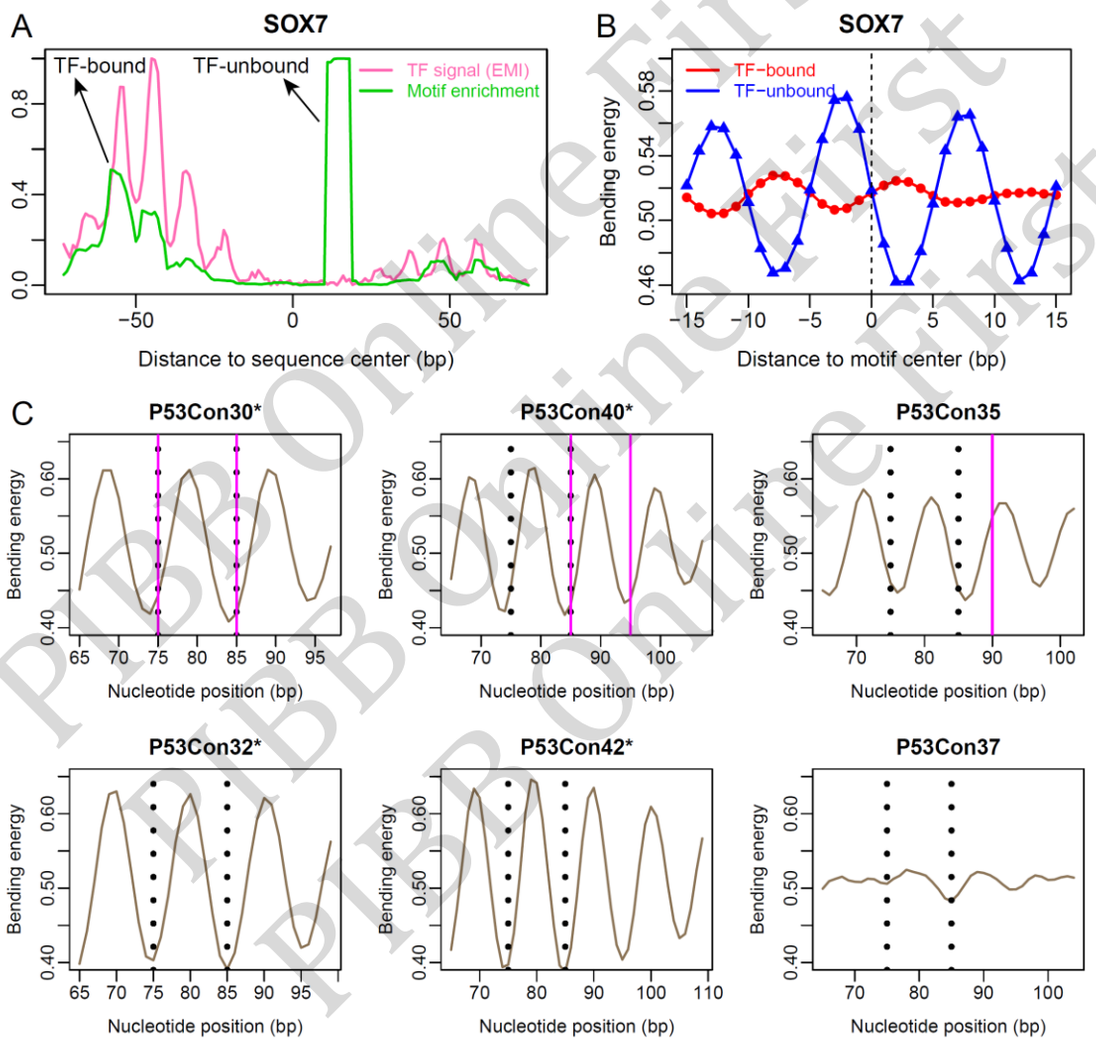


Fig. 2　TF−DNA binding is influenced by the rotational positioning of motifs on nucleosomal DNA

（a） Motif enrichment values and TF-binding signals （EMI） were normalized to 0 – 1 and plotted for comparison. SOX7 motif peaks were categorized as TF-bound or TF-unbound based on whether they overlapped （>70%） with an EMI peak. （b） Comparison of DNA bendability profiles around the TF-bound and TF-unbound groups of motifs. （c） The binding of P53 to DNA is affected by nucleosomal rotational positioning. The dashed lines indicate the center positions of two P53 motifs. An asterisk （*） in the sequence name indicates that the sequence remained bound by P53 after nucleosome reconstitution in the *in vitro* assembly assay （P53Con42*， P53Con30*， P53Con40*， P53Con32*）， whereas sequences without an asterisk were not bound by P53 after nucleosome assembly. The magenta lines represent the center positions of the reconstituted nucleosomes. The precise position of the nucleosome formed on P53Con32* was not experimentally determined.

另外，我们还分析了P53与核小体DNA结合的体外实验数据，发现基于DNA弯曲能判断的P53基序在核小体DNA上的旋转方位（如P53基序中心位置处DNA小沟暴露在核小体上）与P53的结合情况相吻合（图2c）。具体而言，核小体体外组装实验结果表明，P53Con30组装成核小体后其中心位于75/85 bp处（两个位置上都能形成核小体，但这两个核小体不是在同一个分子上形成的），P53Con40上形成的核小体的中心位于85/95 bp。核小体的晶体结构表明，核小体DNA的中心位置上DNA大沟面向组蛋白。DNA弯曲能计算结果表明，DNA弯曲能极小值与核小体中心位置基本一致（误差在2 bp以内），这说明DNA弯曲能模型能够较准确地预测这两条序列上形成的核小体的旋转定位。更重要的是，由于P53的二聚体基序单体长度为10 bp）的中心碱基位于75/85 bp处，根据我们的模型可以推断，P53是结合到暴露于核小体上的DNA小沟处。这与实验上观察到的P53与DNA小沟结合的事实相吻合[33]。P53Con35上组装的核小体的中心位于90 bp处，据此推断P53二聚体基序的两个中心位置上的DNA大沟朝外，这可能是实验上并没有看到P53与之结合的原因。也就是说，TF的基序在核小体DNA上的旋转方位能够影响TF的结合。然而，我们的模型未能准确预测P53Con35上组装的核小体的旋转定位（核小体中心位置上的弯曲能并不是局部极小值），且P53基序中心位置上的弯曲能仍然是倾向于极小值，即DNA弯曲能模型预测该位置上的DNA大沟面向组蛋白，小沟朝外，这有助于P53的结合。这种预测与实验结果相反，这种矛盾是因为实验确定的核小体位置有误，还是我们的模型能力不足导致的，还有待商榷。另外，P53Con32、P53Con42与其对应的P53Con30、P53Con40的弯曲能图谱相似，核小体形成和P53结合情况也无明显差别，与实验结果相吻合。P53Con37在核小体组装实验中虽然形成了核小体，但信号较弱。这条序列的弯曲能图谱不具有显著的10-bp周期性，表明该序列形成核小体的能力较弱，与实验吻合。总体而言，我们的模型能够较好地推断核小体DNA的旋转定位以及P53与核小体DNA结合的旋转定位依赖性。

2.2 旋转定位不是调控体内PTF与核小体结合的关键因素

根据他人文献报道[34]，同时考虑相关数据的可获得性，我们聚焦8个PTFs（OCT4、SOX2、KLF4、GATA4、MYOD1、FOXA1、CEBPA以及ASCL1）在体内基因组上的真实结合情况是否受到DNA旋转方位的影响。将TF基序在基因组上的位点根据是否与其真实TF结合位点重叠分为两组（详见“数据与方法部分”）：TF结合的基序位点和TF未结合的基序位点。计算这两组基序区域的DNA弯曲能，比较其旋转定位情况。结果表明（图3），DNA弯曲能在这8个PTFs的基序位点周围140 bp的范围内呈现明显的10-bp周期性震荡，提示这些TF结合核小体的先锋因子潜能。其中OCT4、GATA4、MYOD1、SOX2结合稳定核小体的潜能较大（DNA弯曲能振幅较大），而KLF4、ASCL1和CEBPA则较小。SOX2偏好结合到核小体组装能力更强的DNA区域，而GATA4和CEBPA可能倾向于结合核小体组装能力弱的DNA区域（图3）。重要的是，无论是TF结合的基序还是TF未结合的基序，其周围DNA变形能均呈现10-bp周期性变化，表明无论TF结合与否，这些基序均表现出旋转定位依赖性。具体而言，若同一个TF的多个基序的DNA弯曲能呈现周期性变化，且基序位点的相位相同（即出现在与核小体二分轴相距整数倍螺旋周期N*10bp的位置上），则对多个基序位点的弯曲能取平均后仍能观察到图中所示周期性信号。据此，我们认为：不论这些PTFs的基序是否被TF结合，漫长的进化过程中TF的DNA结合基序倾向于被设计在“相位相同的旋转方位位点”上。当然，如果基序相对于核小体二分轴位置的相位不同，则可能会影响基序在核小体上的暴露情况，进而影响TF与其结合的亲和力（如ASCL1）。

有两种情况可能会影响本文的结论：a. 如果TF结合基序位于非核小体占据区，此时无从考量旋转定位对TF结合的影响。换句话说，上面分析的TF结合基序为中心的基因组序列片段中，如果非核小体占据区较多，则用DNA形变能判断的核小体旋转定位信号就会受到噪声干扰。b. 有些DNA区域，即使形成核小体，其DNA弯曲能也有可能不具有强烈的10-bp周期性变化规律。这种情况一般会发生在易弯曲（弯曲能低）且弯曲能力呈各向同性（弯曲能的10-bp周期性很弱）DNA区域。为了排除此类干扰，我们进一步筛选出同时满足以下2个条件的基因组基序区域：a. DNA弯曲能的10-bp周期性位列前25%（或50%），即FFT的10周期性幅度较大的25%（或50%）的序列；b. 基

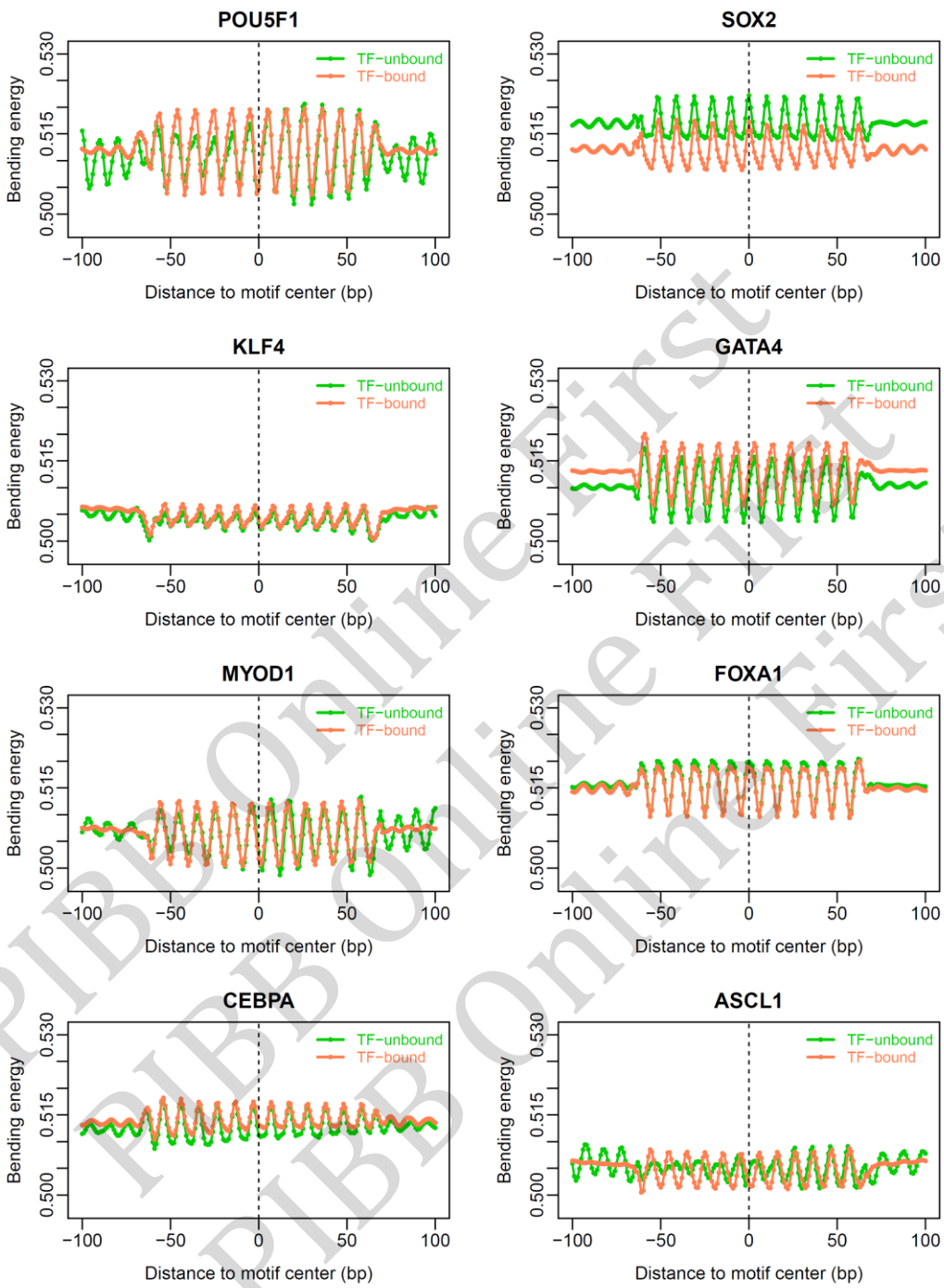


Fig.3　The comparison of rotational positioning between TF−bound and TF−unbound motif sites

For direct visual comparison， the vertical axis range was set to be the same for all subplots.

序完全落入实验测定的人的稳定核小体[35]的中心100 bp区域（https：//generegulation. org/NGS/stable_nucs/hg38/）。用上述条件，分别筛选出TF结合和未结合的基序区域，再比较二者的旋转定位。结果表明（图S1）：以弯曲能的10-bp周期性位列前50%的序列进行分析时，其总体上的旋转定位相同（ASCL1除外）。以弯曲能的10-bp周期性位列前25%的序列进行分析时，基序旋转定位增强，而且TF结合和TF未结合的基序的旋转定位仍然相同（ASCL1除外，图S2）。这些结果进一步

支持上述结论：TF结合的和未结合基序在核小体上总体上采取“相位相近的旋转定位”。换言之，在基因组上对应同一个PTF的众多基序中，采用图中所示旋转定位的基序占比较多（图S1）。这提示，体内TF的结合与否，在总体上并不是由基序在核小体上的旋转方位所调控的。值得注意的是，这只是全基因组水平的总体倾向性，并不意味着旋转定位不影响TF在任何基序位点上的结合。

上述结果是观察特定PTF在基因组上的众多基序位点的平均弯曲能图谱来判断两类基序的旋转定位的总体倾向性。为了进一步挖掘不同旋转定位方式的组成，我们根据FFT的10-bp周期性对应的相位，将基序的中心位置的旋转方位分成4组（图4a）：模式1的基序中心位置上DNA小沟背向组蛋白；模式2的基序中心位置上DNA小沟面向组蛋白；模式3的基序中心位置上游第3个核苷酸位置的DNA小沟背向组蛋白；模式4的基序中心位置下游第3个核苷酸位置的DNA小沟背向组蛋白。模式3和模式4的基序中心位置上DNA大沟与核小体DNA超螺旋轴方向平行，指向相反。这如同把DNA双螺旋以DNA大沟为基准方位，把旋转定位分为DNA的大沟朝向“内侧”（即指向组蛋白八聚体核心，模式1）、“外侧”（模式2）、“超螺旋轴反向”（模式3）和“超螺旋轴正向”（模式4）。同理，将基序的中心位置的旋转方位也可粗略分成以下两组：小沟朝外（DNA小沟背朝组蛋白）和小沟朝内（DNA小沟面朝组蛋白）。

通过分析4种旋转定位的基序所占比例发现（图4b），在TF结合和TF不结合之间变化较大的是ASCL1：模式1在ASCL1结合的一类中的占比大于与其相位相反（旋转定位相反）的模式2，但是在ASCL1不结合的一类中正好相反。另外，相比OCT4不结合的一类，OCT4结合的一类中模式2明显增多。统计检验结果也表明，ASCL1的基序的旋转定位相位在ASCL1结合和ASCL1不结合两组之间存在显著差异（图S3）。而且，ASCL1结合的基序的相位平均合成向量长度（mean resultant length）大于ASCL1不结合的一组（0.159 *vs.* 0.027），提示ASCL1结合的基序的相位较集中。SOX2的基序的旋转定位相位在SOX2结合和SOX2不结合两组之间虽然有显著差异（图S3），但相位平均值（即平均方向）的差别较小（18.3º），不足以对基序的旋转方位产生明显的改变。没有发现其他显著变化。

为了进一步考察这些PTFs与核小体的相互作用能力及其与旋转定位的依赖性，首先分析了人胚肺成纤维细胞（IMR90）[36]中TF结合基序位置周围核小体的占据情况。结果显示（图S4），8种先锋因子中，SOX2、GATA4、MYOD1的DNA结合基序位置上核小体最富集，其次是FOXA1、CEBPA、ASCL1的基序。SOX2、GATA4、MYOD1、FOXA1的结合基序中多数被核小体覆盖，提示在人胚肺成纤维细胞中这些TF表达水平低。这些TF通常与特定的细胞类型和功能相关：SOX2与多能干细胞和神经发育有关，GATA4与心脏发育和中胚层分化有关，MYOD1是肌肉分化的主导调节因子，FOXA1则与内胚层分化和肝脏发育相关。而成纤维细胞是来源于中胚层的已分化细胞，通常不会高表达这些与多能性或特定谱系分化相关的TF。这些TF的基序在IMR90细胞中广泛被核小体覆盖的情况也与其低表达活性相一致。然而，IMR90细胞中需要打开染色质时（如细胞重编程），这些TF是最有潜力的广泛作用因子。有意思的是，OCT4的DNA结合基序被核小体包埋的情况并不十分明显，提示除了人们所熟知的胚胎干细胞中的作用外，OCT4在分化成熟的成纤维细胞中可能与其裸露基序结合，发挥广泛的基因转录调控作用。

人胚肺成纤维细胞（IMR90）中，被核小体占据的基序是否倾向于采用特定的旋转方位，从而帮助细胞需要TF时（如细胞重编程时）促进TF与其结合呢？换句话说，被核小体占据的基序的旋转定位是否较强而且相位固定（相同）？为了回答这个问题，我们将每一种TF在基因组上的基序位点根据核小体占据情况[36]聚为4类：被核小体占据、上游被核小体占据、下游被核小体占据和没被核小体占据（图5a），并比较了这4类之间的旋转定位信号的强度和相位。结果表明（图5b），OCT4、SOX2、GATA4、FOXA1、CEBPA的旋转定位信号在四类之间并没有太大的差别，然而KLF4、ASCL1和MYOD1的四类基序的旋转定位信号存在明显的差别：KLF4、ASCL1和MYOD1的基序越被核小体占据（簇1），其DNA弯曲能越大；核小体越缺乏的基序位点，DNA弯曲能越小。DNA弯曲能的极小值和均值越小，越容易形成核小体[24-25]，因此上述结果说明，在DNA序列水平上不易形成核小体的KLF4、ASCL1和MYOD1的基序区域（簇1）恰恰形成和富集了核小体，DNA序

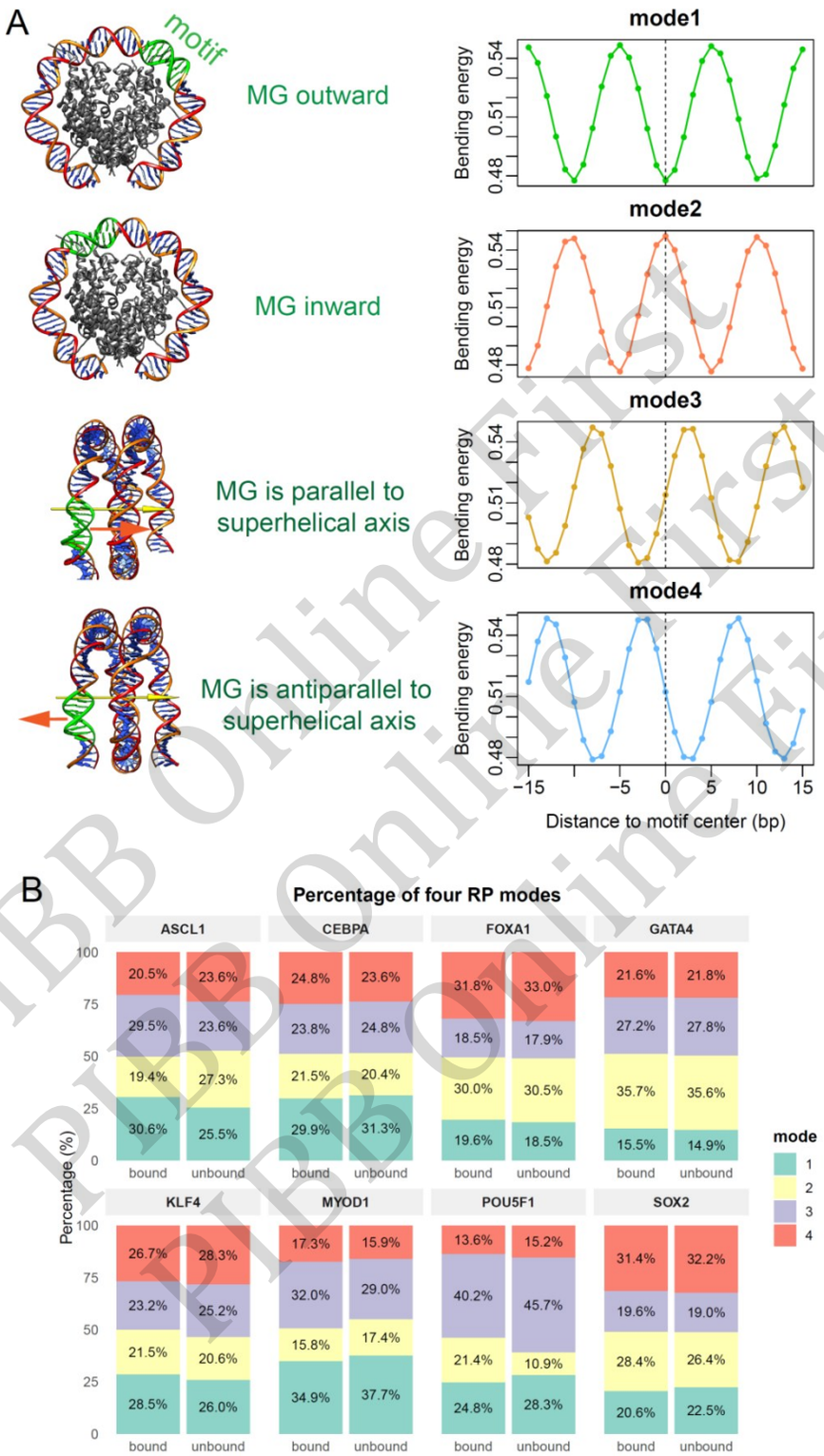


Fig.4　Four rotational orientations classified by FFT phase

（a） The four rotational orientations and their corresponding bending-energy profiles. （b） Proportion of TF-bound motifs among the top 50% of 10-bp periodicity in bending energy and located within stable nucleosome regions for each of the four rotational orientations.

列因素无法解释这一结果。最重要的是，如果基序　　位点上的旋转定位信号越强且相位越一致，则平均

弯曲能图谱上就会出现加强的10-bp周期性振荡，但被核小体覆盖的PTF基序区域（簇1）的弯曲能图谱并没有显示出明显的加强的旋转定位信号（图5b），即在IMR90细胞中没有发现核小体占据率越高的PTF基序上旋转定位越强和相位越固定的证据。

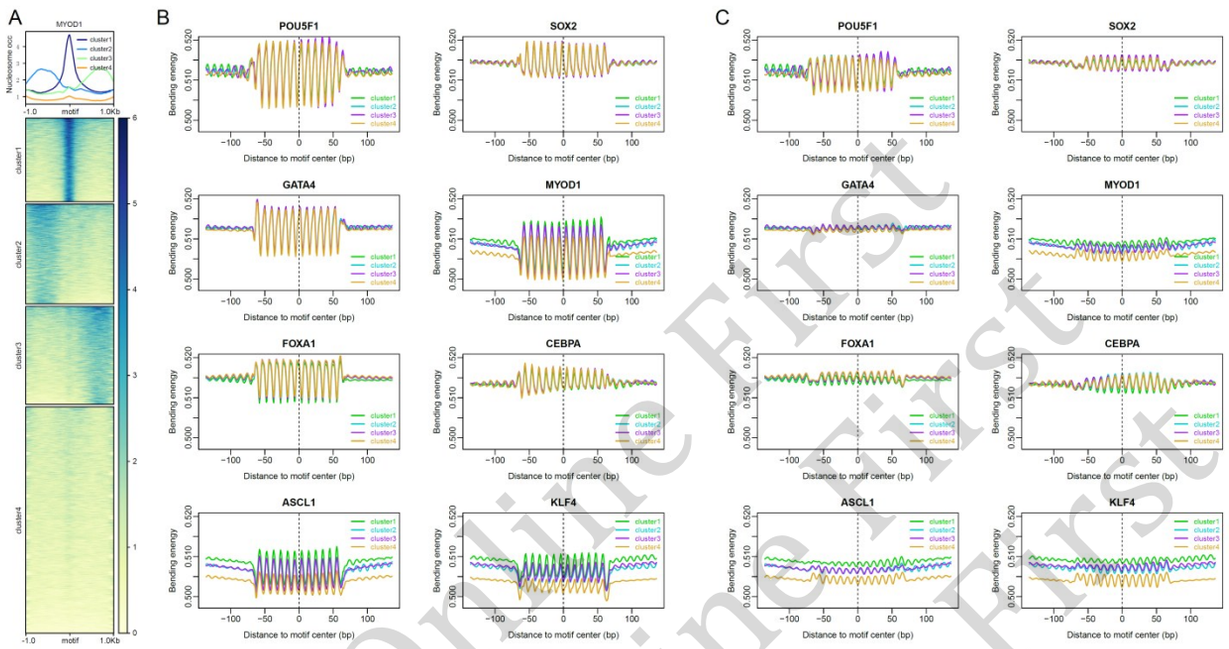


Fig.5　Nucleosome enrichment patterns around TF motifs and corresponding bending energy profiles for IMR90 cells

（a） TF binding motif sites are categorized into four clusters according to nucleosome enrichment pattern in IMR90 cells. K-mean clustering algorithm embedded in deepTools was used. （b） Bending energy profiles aligned at the center of TF motif sites across the human genome. （c） Bending energy profiles of DNA sequences with motif segments replaced with random sequences.

PTF基序周围的DNA弯曲能的强烈10-bp周期性振荡提示着基序位点的强烈的旋转定位信号。有意思的是，究竟是基序的存在导致了DNA序列很强的旋转定位倾向性还是即使没有基序的情况下该区域DNA本身就具备很强的旋转定位信号？为了回答该问题，我们把原有DNA序列中的基序寡聚核苷酸序列用随机序列替换（每条序列中都用不同的随机的序列片段替换），以此来探究不存在基序时序列的旋转定位信号。结果表明，基序被随机序列替换时，DNA弯曲能的平均10-bp周期性信号大大减弱（图5c）。然而，通过FFT分析发现，所有PTF的基序被随机序列替换后弯曲能的10-bp周期性信号（FFT幅度）全部显著上升（图S5），提示基序的存在使得DNA的弯曲各向异性减弱，降低局部DNA的弯曲能力，总体而言抑制核小体的形成。单独分析被核小体占据的簇1时，上述结论同样成立（图S6）。然而，我们结果的另外一个重要提示是：基序的存在会使局部DNA片段倾向于向基序的特定的方向弯曲，即相位一致性增强；虽然不存在基序时DNA弯曲能力更大，但此时弯曲方向不一致，相位一致性较弱。这是导致基序位点DNA弯曲能的平均10-bp周期性信号较强（图5b）的主要原因。同时，我们也注意到，OCT4、SOX2、CEBPA结合基序的侧翼序列本身也具有一定的向特定方向弯曲的旋转定位信号（基序被随机序列替换后仍表现出一定10-bp周期性振荡），但与存在基序时的旋转定位信号的相位不完全相同：OCT4结合基序的存在加强了原有特定方位的旋转定位信号，而SOX2结合基序的存在对侧翼局部序列的旋转定位信号的相位有所微调作用（改变弯曲方向），CEBPA结合基序的存在则完全翻转了侧翼序列原来的旋转方位（图5b和图5c中弯曲能波峰波谷完全翻转）。

OCT4、SOX2、KLF4和c-MYC（简写为OSKM）的表达具有诱导体细胞重新编程和产生iPSC的能力。iPSC类似于胚胎干细胞，它们能分

化成人身体所有细胞类型。人胚肺成纤维细胞（IMR90）被OSKM诱导重编程为iPSC的过程中，OSKM能够作用于关闭的染色质区上的核小体上，启动染色质重塑并激活下游相关基因的表达。我们获得了OSKM被诱导48 h时候在IMR90细胞的基因组上结合的位点（hg18）[37]，并将其转换为hg38位点，获得了IMR90细胞核小体覆盖区内OCT4、SOX2、KLF4的基序（如图5中簇1），并区分其中是否被相应PTF结合。这些基序同样是被核小体覆盖，为什么有些被OSKM结合，而有些却不被结合呢？DNA基序的旋转定位是否决定这些PTFs与其结合？为了回答这个问题，我们分析了这两组基序的旋转定位，发现其总体旋转方位保持一致（图6a），而且4种旋转定位方式的占比在TF结合和不结合的两组中并没有明显的翻转（图6b，c）。例如，在SOX2结合的一组中，SOX2的基序中心位置的DNA大沟暴露在核小体上（弯曲能极大）的情况多于DNA大沟被包埋的情况（模式2/模式1=1 228/1 083），在SOX2未结合的一组中仍然是DNA大沟暴露在核小体上的基序占多数（模式2/模式1=2 780/2 540）。统计检验也表明，旋转定位相位在PTF结合和不结合的两组基序之间不存在显著差异（图S7）。这些结果提示，DNA基序的旋转定位不像是调控这些PTFs能否与其结合的关键因素。

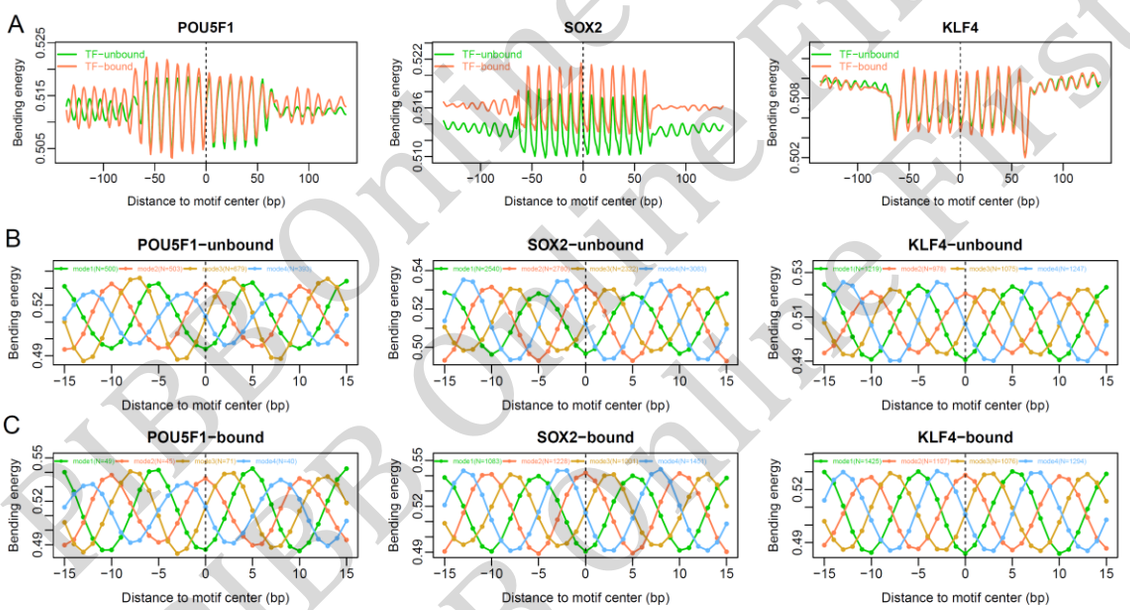


Fig.6　Average bending-energy profiles centered on nucleosomal TF-motif sites，comparing regions of TF-bound with TF-unbound

（a） Comparison of average bending energy profiles. （b） Bending energy profiles and quantities of four rotational positioning modes in TF-unbound motifs. （c） Bending energy profiles and quantities of four rotational positioning modes in TF-bound motifs. A motif was deemed TF-bound if fully overlapped by a ChIP-seq peak， and TF-unbound if entirely outside any peak.

早期胚胎发育的细胞分化过程中存在剧烈的染色质重塑和表观遗传重编程。在这过程中也定会有PTFs与染色质的相互作用协助染色质重塑并激活分化相关基因的表达。为了探究TF结合基序的旋转定位是否调控这一过程，我们还分析了hESC向hNEC分化的过程中PTF与核小体的相互作用[38]。从核小体富集结果看，hESC向hNEC分化过程中OCT4的DNA结合基序从核小体缺乏状态转变为核小体占据状态（图S8），提示hNEC中OCT4的活性大大降低，其结合基序大多处于被核小体占据的低活性状态。其他7个PTFs的基序位点被核小体占据的总体情况没有明显的变化。

为了探究在hESC向hNEC分化的过程中SOX2在核小体上的结合是否依赖于基序的旋转定位，我们将每种细胞类型中核小体区域的SOX2基序根据是否被SOX2结合分成“SOX2结合”和“SOX2

不结合”两组（数据处理过程见“1.4 实验核小体定位数据处理”），并通过比较两组基序区域的DNA弯曲能图谱判断其旋转定位差异。结果表明（图7），在ESC中核小体上的SOX2基序不论是否被SOX2结合，其旋转方位的总体倾向是一致的（图7a），但显然被SOX2结合的基序弯曲能的10-bp周期性信号更强，说明其在核小体上的旋转定位相位更统一（多个基序向着同一个方向弯曲的整齐划一的程度较高）。然而，在分化后的hNEC中，SOX2结合和不结合的基序的旋转定位正好相反（图7b），而且，比较意外的是，被SOX2结合的基序的旋转方位与之前的所有的SOX2基序的总体旋转方位相反。相比SOX2不结合的一组（图7c），旋转方位模式1在SOX2结合一组中增加，模式4则减少是导致SOX2结合和不结合的基序的旋转定位正好相反的直接原因。将核小体旋转方位粗略分成两大模式时（见“1.6 DNA弯曲能的傅里叶变换分析”），hNEC中motif的DNA沟槽在核小体上的暴露或包埋的占比差异更加明显（图7D）。置换检验结果也表明，hNEC中SOX2结合和不结合的基序的旋转定位相位差异显著（图S9）。结构生物学证据表明，SOX2基序的中心位置上DNA大沟朝向组蛋白弯曲，SOX2的高迁移率族（high-mobility group，HMG）结构域结合DNA小沟[39]，这与hNEC中SOX结合基序的旋转定位相一致。进一步分析发现，NEC中SOX2结合的核小体上的基序无一是hESC中SOX2结合的基序，也就是说hESC分化为hNEC后SOX2的结合位点发生了彻底重排。

我们进一步将核小体上的SOX2基序分成两组：启动子区域（转录起始位点上游2 000 bp）的基序和基因体区域（转录起始位点下游2 000 bp）的基序，并比较两者的旋转定位。发现无论是启动子还是基因体区域，SOX2基序的旋转定位倾向性（图S10）与总体倾向性（图7）保持一致。

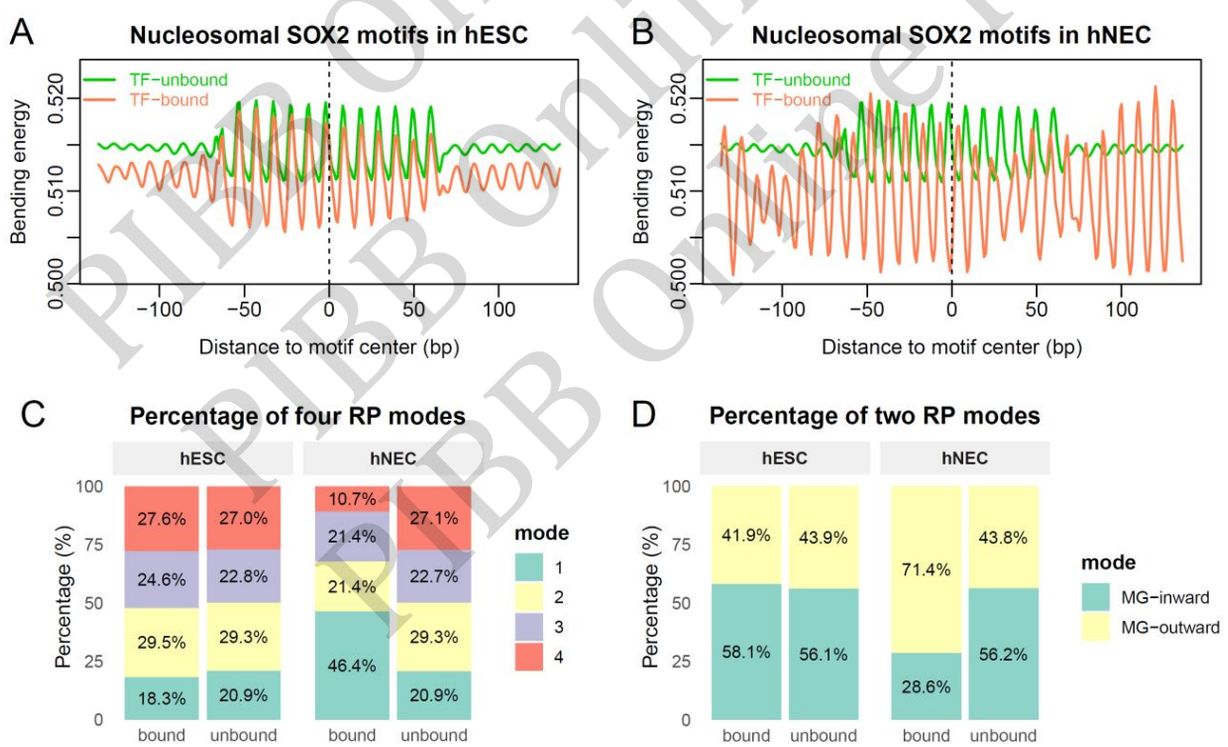


Fig.7　Rotational positioning of SOX2−bound versus unbound motifs on nucleosomes

（a） Rotational positioning of SOX2-bound and unbound motifs on nucleosomes in hESC. （b） Same as （a） but for hNEC. （c） Proportion of the four rotational positioning modes. （d） Proportion of the two rotational positioning modes.

3　讨论

本文研究结果显示，在体外，SOX7的结合受其DNA识别模体在核小体上旋转方位的影响：在某些模体富集位点，因其不利的旋转方位而未观察到SOX7结合。通过比较结合与未结合位点周围的DNA弯曲能图谱，发现未结合位点的弯曲能呈现规律且统一的强周期性，暗示其处于不利的结合相位。与之相反，成功结合的位点其弯曲能图谱虽周期性较弱，但相位与未结合位点恰好相反，表明旋转方位的差异是决定SOX7能否有效结合的关键因素。基于DNA弯曲能模型预测的核小体旋转定位，与P53结合实验相符：当模型预测P53识别基序中心处DNA小沟暴露时，P53能够结合。该模型成功解释了P53在不同序列（如P53Con30/40）上的结合差异。总体而言，DNA弯曲能模型能有效推断核小体定位及P53结合的旋转方位依赖性，与大部分实验数据吻合。

对8个PTFs的分析表明，其DNA识别基序周围的DNA弯曲能普遍呈现10-bp周期性，提示它们具有结合核小体的潜能。关键发现是，对于大多数因子，已被TF结合的模体与未被结合的模体，在核小体上倾向于采用相位相同的旋转方位。这表明在全基因组水平上，旋转方位并非调控其结合的主要因素。这一现象被认为是进化上的“设计”，使模体相位保持一致，可能有助于动态调控染色质可及性，从而在适当时机激活基因。ASCL1是个例外，ASCL1结合和不结合的基序的总体旋转定位倾向性有较大差异。一种可能性是：ASCL1的bHLH结构域是一个相对常见和简单的DNA结合域，而且其DNA结合依赖于异源二聚体化，导致ASCL1对旋转定位较敏感，表现出较弱的核小体结合能力，即擅长利用核小体上天然暴露的位点，但缺乏OCT4的“柔性适配”或SOX2的“强力重塑”能力去攻克被严重包埋的位点。在IMR90细胞中，研究发现对于大多数PTFs（如OCT4、SOX2），其模体位点的核小体占据情况与DNA序列所固有的旋转定位信号强度无关。然而，KLF4、ASCL1和MYOD1表现出反常现象：核小体占据率高的基序区域，其DNA序列本身反而更不利于核小体形成，这表明核小体在这些位置的富集并非由DNA序列因素主导。最关键的是，分析并未发现核小体占据率高的PTF模体具有更强或更一致的旋转定位信号，这反驳了“核小体通过强化特定旋转方位来调控PTF结合”的假设。

研究发现，PTF结合基序的存在，并非单纯地增强或减弱DNA的弯曲能力，而是显著增强了DNA朝向特定方向弯曲的“相位一致性”，从而在平均弯曲能图谱上产生强烈的10-bp周期性信号。当用随机序列替换基序后，DNA整体的弯曲能力（柔性）反而增加，但其弯曲方向变得不一致，导致平均周期性信号减弱，这揭示了基序在规范核小体旋转定位中的主导作用。不同PTF的基序对其侧翼序列固有的旋转定位信号影响各异：OCT4的基序强化了原有相位，SOX2的基序对其进行了微调，而CEBPA的基序则完全翻转了原有的弯曲方向，表明它们可能通过不同的机制调控核小体定位。

在IMR90细胞重编程过程中，对于被核小体覆盖的相同PTF（如SOX2）模体，其能否被OSKM结合并不主要取决于模体在核小体上的旋转方位。分析发现，已被结合的模体与未被结合的模体，其整体的旋转定位倾向保持一致，且4种典型旋转方位的比例在两组间未出现显著翻转。因此，DNA基序的旋转定位并非调控这些PTFs在早期重编程阶段与核小体DNA结合的关键因素，可能存在其他更重要的机制决定了结合的差异性。

结合先验知识和本文研究结果可知，SOX2在hESC分化为hNEC的过程中结合到一些核小体区域，打开染色质。SOX2在分化过程中会从其多能性相关的靶位点上解离，但同时会结合到新的、与神经外胚层命运相关的基因调控元件上，帮助打开这些区域的染色质，激活这些神经谱系基因的表达[38]。这是一个动态的、全局性的结合位点“重编程”过程，伴随着SOX2功能的“转岗”。SOX2在多能干细胞中表达最高，在向神经外胚层分化时，其表达水平会从峰值下降，但仍会维持在一个显著高于非神经细胞的水平上。它在神经前体/干细胞中持续表达，对维持神经谱系的身份至关重要。分化后hNEC细胞中仍有一定的SOX2结合到染色质上[38]，这些剩余的SOX2结合位点不再是维持“多能性”，而是转变为驱动和维持“神经外胚层身份”的核心。我们的结果表明，在胚胎干细胞中，SOX2结合和不结合的基序在核小体上的旋转方位总体一致。然而，在分化后的神经外胚层细胞中，SOX2倾向于结合那些旋转方位与胚胎干细胞中总体倾向完全相反的基序，这与已知的SOX2结合DNA小沟的结构生物学证据相符。这种结合

位点和旋转方位的彻底重排，提示细胞类型特异的辅助因子（而非单纯的DNA序列信息）可能是调控SOX2与核小体相互作用机制的关键。例如，协助SOX2与核小体相互作用的其他辅助蛋白质在两类细胞中的表达差异影响了SOX2与核小体上基序的相互作用机制。

总体而言，本研究结果表明：TF基序的旋转定位调控其在体外与核小体的结合，但在体内PTF结合核小体的过程很大程度上并不受控于基序的旋转定位。在胚胎干细胞的分化以及诱导体细胞重编程为多能干细胞的过程中都是如此。这种反差可能与两种因素有关：转录因子的先锋性和体内外环境。

在体内，即使结合表面（如DNA小沟）部分包埋在核小体上，先锋因子仍然能够结合。这可能正是它们区别于经典TF的关键能力。常规TF通常结合在裸露的DNA或染色质高度开放的区域，与关闭染色质区和核小体的结合能力很弱。PTFs（如SOX2、PAX7、OCT4等）拥有一系列独特的结构特性，使其能够识别核小体上的部分遮蔽的基序，并且在结合后主动去稳定核小体，为染色质重塑复合物打开通道，从而“先锋性地”打开染色质。如何结合包埋的基序？有研究提示，PTF并非粗暴地撕开核小体，而是采用更精巧的策略：即使主要结合表面被包埋，DNA在核小体上并非完全僵化的结构，而是存在着热波动，即DNA会自发地发生轻微的“呼吸”作用，暂时性地暴露出一些原本被遮蔽的区域[40-42]。PTFs具有足够高的浓度和与DNA结合的超高亲和力，使其能够利用这些瞬间的机会“抓住”DNA，即使初始结合可能很弱或不完全。另外，PTFs可以通过以下几种结构特性来应对核小体的挑战。a. 内在无序区（intrinsically disordered regions，IDRs）：许多PTFs（如SOX2）含有长的、非结构化的柔性区域。这些IDRs可以像“触手”一样先与核小体（包括组蛋白和DNA）发生非特异性相互作用，帮助将因子锚定在附近，为其结构域最终找到并结合目标基序争取时间和机会。b. 协同结合与二聚化：许多PTFs以二聚体形式发挥作用（如OCT4-SOX2）。其中一个单体可能先与核小体上某个更容易接触的位点结合，从而帮助另一个单体定位并结合到邻近的、包埋更深的基序上，提高结合效率[43]。c. DNA扭曲：一些PTFs在结合时能够主动地扭曲DNA。这种施加在DNA上的力可以部分地抵消核小体对DNA的包裹力，从而“撬开”一点空间，使被包埋的碱基更多地暴露出来。例如，SOX2可以通过其HMG结构域插入DNA小沟并引起DNA大幅弯曲[20，44]。

在体外实验环境中，上述协助PTF结合包埋的基序的条件未必充分。例如，在单一PTF结合的体外实验中，PTFs之间的协同作用彻底消失。即使在2~3个PTFs共同存在的体外环境中，协同作用的多样性和效果也可能会大打折扣。核小体的呼吸作用在体外环境中可能会受限，因为体外环境通常缺乏促进呼吸作用的关键因子。核小体的呼吸作用（DNA局部瞬时的解旋与重绕）在细胞内受多种因素调控，如组蛋白修饰酶、染色质重塑复合物、转录因子、组蛋白变体等[45-46]。体外实验体系（如纯化的核小体与DNA）往往缺乏这些生物活性成分，可能导致呼吸作用频率或幅度降低。另外，我们的结果提示，P53和SOX7在体外与核小体的结合的先锋潜能的大小还有待实验证实。若二者的先锋潜能较弱，则可以预期它们倾向于通过依赖于其基序旋转定位的方式与核小体结合。

4　结论

本研究基于DNA形变能模型与核小体定位实验数据，揭示了DNA旋转定位在TF-核小体相互作用中的调控作用。研究表明：a. 在体外，SOX7和P53的基序的旋转定位差异直接决定其是否能够有效结合核小体DNA；b. PTF的基序在基因组中普遍表现出相对优势的旋转定位特征，无论是否被结合，PTF的基序均倾向于出现在相位一致的核小体位点，提示其旋转定位在进化上被优化以利于TF结合调控；c. 在hESC向hNEC分化过程中，SOX2结合位点的旋转定位与结合关系可发生重塑，提示细胞环境因素可调节旋转定位与TF结合的关联；d. 最重要的是，本文结果表明，PTF在体内结合核小体的过程很大程度上不依赖于基序的旋转定位，即使其基序的结合表面（如DNA小沟）被核小体包埋，PTF仍然能够结合，这与其固有的结构特性、核小体呼吸作用等因素有关。综上，DNA旋转定位在PTF与核小体相互作用过程的功能解析有助于深入揭示PTF在表观调控、细胞命运重编程等过程分子机制。

附件见本文网络版（http：//www. pibb. ac. cn，http：//www.cnki.net）：

PIBB_20250434_Figure_S1.pdf
PIBB_20250434_Figure_S2.pdf
PIBB_20250434_Figure_S3.pdf
PIBB_20250434_Figure_S4.pdf
PIBB_20250434_Figure_S5.pdf
PIBB_20250434_Figure_S6.pdf
PIBB_20250434_Figure_S7.pdf
PIBB_20250434_Figure_S8.pdf
PIBB_20250434_Figure_S9.pdf
PIBB_20250434_Figure_S10.pdf

参考文献

[1] Mayran A, Drouin J. Pioneer transcription factors shape the epigenetic landscape. J Biol Chem, 2018, **293**(36): 13795-13804

[2] Zaret K S, Mango S E. Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr Opin Genet Dev, 2016, 37: 76-81

[3] Jacobs J, Atkins M, Davie K, *et al*. The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. Nat Genet, 2018, **50**(7): 1011-1020

[4] Cirillo L A, Zaret K S. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. Mol Cell, 1999, **4**(6): 961-969

[5] Sridharan R, Tchieu J, Mason M J, *et al*. Role of the murine reprogramming factors in the induction of pluripotency. Cell, 2009, **136**(2): 364-377

[6] 李令杰, 金颖. 调控胚胎干细胞自我更新的关键转录因子研究进展. 生命科学, 2009, **21**(5): 631-638
Li L J, Jin Y. Chin Bull Life Sci, 2009, **21**(5): 631-638

[7] Avilion A A, Nicolis S K, Pevny L H, *et al*. Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev, 2003, **17**(1): 126-140

[8] Shu J, Wu C, Wu Y, *et al*. Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell, 2013, **153**(5): 963-975

[9] Friedman J R, Kaestner K H. The Foxa family of transcription factors in development and metabolism. Cell Mol Life Sci CMLS, 2006, **63**(19): 2317-2328

[10] Arinobu Y, Mizuno S I, Chong Y, *et al*. Reciprocal activation of GATA-1 and PU. 1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. Cell Stem Cell, 2007, **1**(4): 416-427

[11] Mayran A, Khetchoumian K, Hariri F, *et al*. Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. Nat Genet, 2018, **50**(2): 259-269

[12] Teng M, Zhou S, Cai C, *et al*. Pioneer of prostate cancer: past, present and the future of FOXA1. Protein Cell, 2021, **12**(1): 29-38

[13] Formaggio N, Sgrignani J, Thillaiyampalam G, *et al*. Targeting FOXA1 and FOXA2 disrupts the lineage-specific oncogenic output program in prostate cancer. Cell Rep, 2025, **44**(10): 116324

[14] Kawasaki K, Salehi S, Zhan Y A, *et al*. FOXA2 promotes metastatic competence in small cell lung cancer. Nat Commun, 2025, **16**(1): 4865

[15] Lyu H, Chen X, Cheng Y, *et al*. Pioneer factor GATA6 promotes colorectal cancer through 3D genome regulation. Sci Adv, 2025, **11**(6): eads4985

[16] Dang C V. MYC on the path to cancer. Cell, 2012, **149**(1): 22-35

[17] Knoepfler P S, Zhang X Y, Cheng P F, *et al*. Myc influences global chromatin structure. EMBO J, 2006, **25**(12): 2723-2734

[18] Zhu F, Farnung L, Kaasinen E, *et al*. The interaction landscape between transcription factors and the nucleosome. Nature, 2018, **562**(7725): 76-81

[19] Morgunova E, Taipale J. Structural insights into the interaction between transcription factors and the nucleosome. Curr Opin Struct Biol, 2021, 71: 171-179

[20] Fernandez Garcia M, Moore C D, Schulz K N, *et al*. Structural features of transcription factors associating with nucleosome binding. Mol Cell, 2019, **75**(5): 921-932.e6

[21] Ozden B, Boopathi R, Barlas A B, *et al*. Molecular mechanism of nucleosome recognition by the pioneer transcription factor sox. J Chem Inf Model, 2023, **63**(12): 3839-3853

[22] Cui F, Zhurkin V B. Rotational positioning of nucleosomes facilitates selective binding of p53 to response elements associated with cell cycle arrest. Nucleic Acids Res, 2014, **42**(2): 836-847

[23] Sahu G, Wang D, Chen C B, *et al*. p53 binding to nucleosomal DNA depends on the rotational positioning of DNA response element. J Biol Chem, 2010, **285**(2): 1321-1332

[24] Liu G, Xing Y, Zhao H, *et al*. A deformation energy-based model for predicting nucleosome dyads and occupancy. Sci Rep, 2016, **6**: 24133

[25] Liu G, Zhao H, Meng H, *et al*. A deformation energy model reveals sequence-dependent property of nucleosome positioning. Chromosoma, 2021, **130**(1): 27-40

[26] Masella A P, Bartram A K, Truszkowski J M, *et al*. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinform, 2012, **13**(1): 31

[27] Grant C E, Bailey T L, Noble W S. FIMO: scanning for occurrences of a given motif. Bioinformatics, 2011, **27**(7): 1017-1018

[28] Langmead B, Salzberg S L. Fast gapped-read alignment with bowtie 2. Nat Methods, 2012, **9**(4): 357-359

[29] Li H, Handsaker B, Wysoker A, *et al*. The sequence alignment/map format and SAMtools. Bioinformatics, 2009, **25**(16): 2078-2079

[30] Ramírez F, Dündar F, Diehl S, *et al*. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res, 2014, 42 (web server issue): **W187**-W191

[31] Chen K, Xi Y, Pan X, *et al*. DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. Genome Res, 2013, **23**(2): 341-351

[32] Olson W K, Bansal M, Burley S K, *et al*. A standard reference

frame for the description of nucleic acid base-pair geometry. J Mol Biol, 2001, **313**(1): 229-237

[33] Kitayner M, Rozenberg H, Kessler N, *et al*. Structural basis of DNA recognition by p53 tetramers. Mol Cell, 2006, **22**(6): 741-753

[34] Balsalobre A, Drouin J. Pioneer factors as master regulators of the epigenome and cell fate. Nat Rev Mol Cell Biol, 2022, **23**(7): 449-464

[35] Gaffney D J, McVicker G, Pai A A, *et al*. Controls of nucleosome positioning in the human genome. PLoS Genet, 2012, **8**(11): e1003036

[36] Kelly T K, Liu Y, Lay F D, *et al*. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res, 2012, **22**(12): 2497-2506

[37] Soufi A, Donahue G, Zaret K S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell, 2012, **151**(5): 994-1004

[38] Du Y, Liu Z, Cao X, *et al*. Nucleosome eviction along with H3K9ac deposition enhances Sox2 binding during human neuroectodermal commitment. Cell Death Differ, 2017, **24**(6): 1121-1131

[39] Dodonova S O, Zhu F, Dienemann C, *et al*. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. Nature, 2020, **580**(7805): 669-672

[40] Mishra S K, Bhattacherjee A. How do nucleosome dynamics regulate protein search on DNA? J Phys Chem B, 2023, **127**(25): 5702-5717

[41] Mondal A, Felipe C, Kolomeisky A B. Nucleosome breathing facilitates the search for hidden DNA sites by pioneer transcription factors. J Phys Chem Lett, 2023, **14**(17): 4096-4103

[42] Hungyo K, Audit B, Vaillant C, *et al*. Thermodynamics of nucleosome breathing and positioning. J Chem Phys, 2025, **162**(2): 025101

[43] Mondal A, Mishra S K, Bhattacherjee A. Nucleosome breathing facilitates cooperative binding of pluripotency factors Sox2 and Oct4 to DNA. Biophys J, 2022, **121**(23): 4526-4542

[44] Hou L, Srivastava Y, Jauch R. Molecular basis for the genome engagement by Sox proteins. Semin Cell Dev Biol, 2017, 63: 2-12

[45] Buning R, van Noort J. Single-pair FRET experiments on nucleosome conformational dynamics. Biochimie, 2010, **92**(12): 1729-1740

[46] Wen Z, Zhang L, Ruan H, *et al*. Histone variant H2A.Z regulates nucleosome unwrapping and CTCF binding in mouse ES cells. Nucleic Acids Res, 2020, **48**(11): 5939-5952

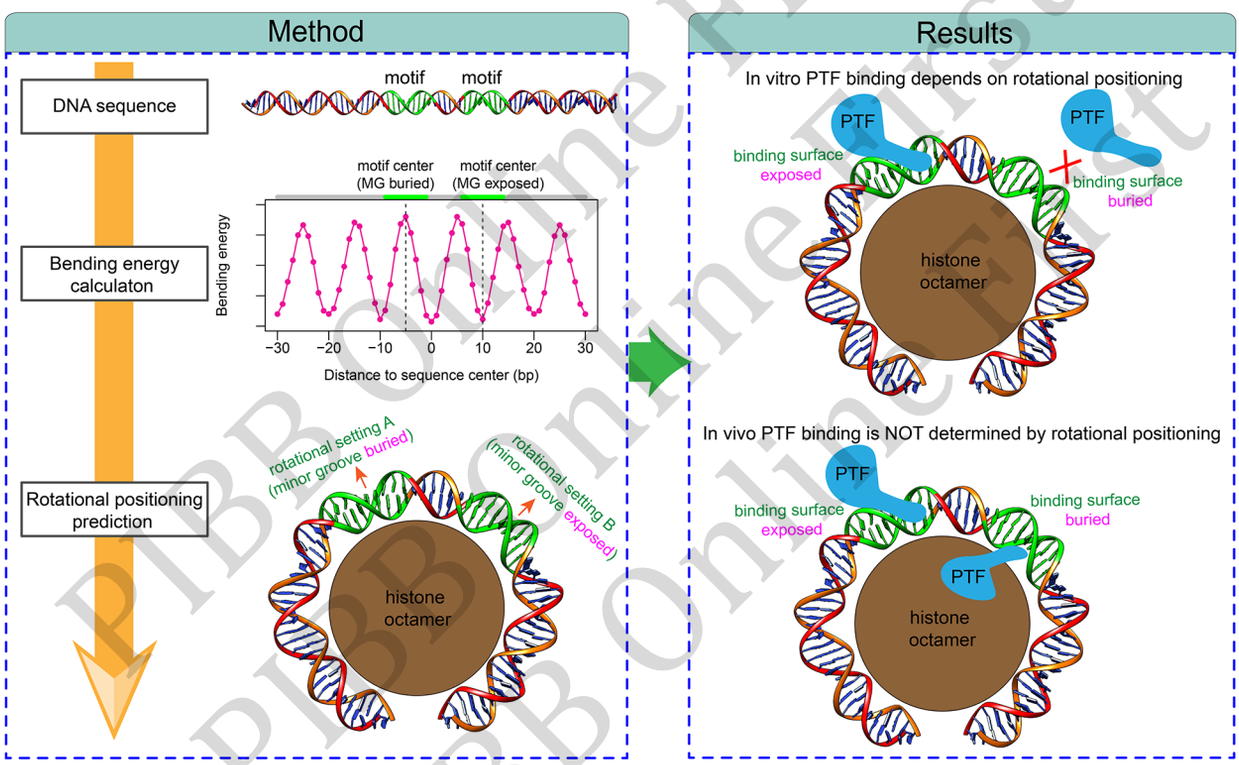
Differential Role of Rotational Positioning in Pioneer Transcription Factor Binding to Nucleosomes *In vivo vs. In vitro**

LIU Guo-Qing[1,2)**], GUO Xing-Yue[1,2)], CANG Jing[1,2)], ZHANG Zhi[1,2)], LIU Guo-Jun[1,2)]

([1)]*School of Life Science and Technology, Inner Mongolia University of Science and Technology, Baotou* 014010, China;

[2)]*Inner Mongolia Key Laboratory of Life Health and Bioinformatics, Inner Mongolia University of Science and Technology, Baotou* 014010, China)

Graphical abstract



Abstract **Objective** Pioneer transcription factors (PTFs) possess the unique ability to recognize and bind their target DNA sequences within compacted nucleosomal DNA, thereby initiating chromatin opening and gene expression. They play pivotal roles in fundamental biological processes such as embryonic development, cellular reprogramming, and tumorigenesis. The specific regulatory mechanism by which nucleosomal rotational positioning governs PTF-nucleosome interactions remains inadequately elucidated. This study aims to systematically investigate the role of the rotational orientation of motifs in PTF-nucleosome binding. **Methods** We employed a DNA deformation energy model to predict the rotational positioning of DNA on nucleosomes. We analyzed high-throughput in vitro data from the NCAP-SELEX assay, which profiles the binding landscapes of numerous transcription factors to nucleosomal DNA. For in vivo analysis, we integrated genome-wide binding data (ChIP-seq) and nucleosome positioning data (MNase-seq) for eight well-characterized pioneer factors (OCT4, SOX2, KLF4, GATA4, MYOD1, FOXA1, CEBPA, and ASCL1) in human cells. Binding motifs were classified as "TF-bound" if they overlapped with ChIP-seq 峰 s and "TF-unbound" otherwise. DNA

bendability profiles and Fast Fourier Transform (FFT) analysis were used to assess rotational positioning patterns around these motif sites. This analytical framework was further applied to specific biological contexts, including cellular reprogramming from IMR90 fibroblasts to induced pluripotent stem cells (iPSCs) and the differentiation of human embryonic stem cells (hESCs) to human neuroectodermal cells (hNECs). **Results** Our *in vitro* analysis revealed a strong dependence of transcription factor binding on the rotational orientation of TF-binding motifs. For SOX7, the unbound motifs at specific enrichment 峰s exhibited a rotational phase clearly opposite to that of the SOX7-bound motifs. Similarly, analysis of P53 binding sequences confirmed that successful binding in vitro correlated with model-predicted exposure of the DNA minor groove at the motif center, consistent with P53's binding mode. Genome-wide *in vivo* analysis of the eight PTFs showed that their DNA binding motifs were generally associated with DNA sequences exhibiting significant 10-bp periodicity in bendability, suggesting an inherent potential for nucleosome association. Crucially, for most factors (except ASCL1), the average rotational positioning preferences were remarkably similar between TF-bound and TF-unbound motifs. This indicates that, at a global genomic level, rotational positioning is not the primary determinant dictating whether a nucleosomal motif is bound by its cognate PTF *in vivo*. This phenomenon persisted during cellular reprogramming (IMR90 to iPSC), where the rotational positioning of OSKM factor motifs bound versus unbound in nucleosomal regions showed no significant overall difference. Interestingly, during hESC differentiation to hNECs, SOX2 binding sites underwent comprehensive reprogramming. In hNECs, the rotational positioning of nucleosomal SOX2-bound motifs was significantly different and, unexpectedly, opposite to the general preference observed in hESCs and for unbound motifs in hNECs, suggesting a cell context-dependent rewiring of binding mechanisms. **Conclusion** This study suggests a distinction in the role of DNA rotational positioning in TF-nucleosome binding between in vitro and in vivo environments. While rotational positioning critically governs the binding efficiency of factors like SOX7 and P53 in simplified in vitro systems, PTFs *in vivo* appear to overcome this steric hindrance at the binding interface. The ability of PTFs to bind nucleosomal motifs, even when key interaction surfaces are partially buried, might stem from their unique structural properties (*e.g.*, intrinsically disordered regions, DNA distortion/binding domains), nucleosome breathing which transiently exposes DNA, and potential cooperativity with other factors. Our results highlight the unique capacity of pioneer factors to drive chromatin openness through mechanisms beyond rotational positioning.

Key words pioneer transcription factors,nucleosome,rotational positioning,DNA deformation energy,chromatin accessibility

DOI：10.3724/j.pibb.2025.0434 **CSTR**：32369.14.pibb.20250434

* This work was supported by grants from The National Natural Science Foundation of China (62161043), Inner Mongolia Natural Science Foundation of China (2025MS06029), and the 2025 Inner Mongolia Key Laboratory of Life Health and Bioinformatics Project (2025KYPT0135).

** Corresponding author.
Tel：86-15148991105，E-mail：gqliu1010@163.com

Received：October 2，2025 Accepted：December 3，2025