



# The pattern and evolution of yeast promoter bendability

Itay Tirosh<sup>1</sup>, Judith Berman<sup>2,3</sup> and Naama Barkai<sup>1,4</sup>

<sup>1</sup> Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

<sup>2</sup> Department of Genetics, Cell Biology and Development, University of Minnesota, MN 55455, USA

<sup>3</sup> Department of Microbiology, University of Minnesota, MN 55455, USA

<sup>4</sup> Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel

**In *Saccharomyces cerevisiae*, transcription factor binding sites are found preferentially ~100–200 bp upstream of the start codon. Here, we show that this region is associated with rigid DNA in promoters lacking a TATA box, but not in TATA-containing promoters. The association of rigid DNA with transcription factor binding sites is conserved in TATA-less promoters from 11 yeast species, whereas the position of the rigid DNA varies substantially among species. Rigid DNA could influence nucleosome positioning and assist in the assembly of the transcriptional machinery at TATA-less promoters.**

## Unraveling promoter structure and function

The capacity of transcription factors to activate gene expression is encoded in the promoter sequences. Best characterized are short sequence motifs that function as binding sites for specific transcription factors [1]. However, the presence of a particular sequence motif is not sufficient to ensure regulation by the associated transcription factor. For example, binding sites must be accessible within the chromatin to enable their recognition [2,3]. They should also be properly positioned to enable their interaction with the general transcriptional machinery or other regulatory proteins. Promoter features that control these events are still poorly understood.

The exact nucleotide sequence of a promoter influences its three-dimensional structure. Although promoter structure cannot currently be predicted from its sequence, some sequence-dependent structural properties have been described. In particular, the propensity of each trinucleotide to bend (i.e. their bendability) was previously estimated from DNase I digestion studies [4]. Bendability of promoter sequences has been examined in several organisms [5–7] and was suggested to influence nucleosome positions [5,8], transcription factor binding [9,10] and DNA looping [11]. However, our knowledge of a general association between promoter ‘bendability’ and regulatory features remains sparse.

## Typical architecture of *S. cerevisiae* promoters

We examined the bendability pattern of yeast promoters. The promoters of many yeast genes were found to have

regions of ~50 bp with significantly low bendability compared with the entire promoter. For example, 44% of *Saccharomyces cerevisiae* promoters displayed a region of low bendability at a significance level of  $P < 0.05$  (Online Supplementary Figure S1). Regions of low bendability were greatly enriched 100–200 bp upstream of the start codon (Online Supplementary Figure S1). Examples of bendability patterns of specific promoters are given in Online Supplementary Figure S2.

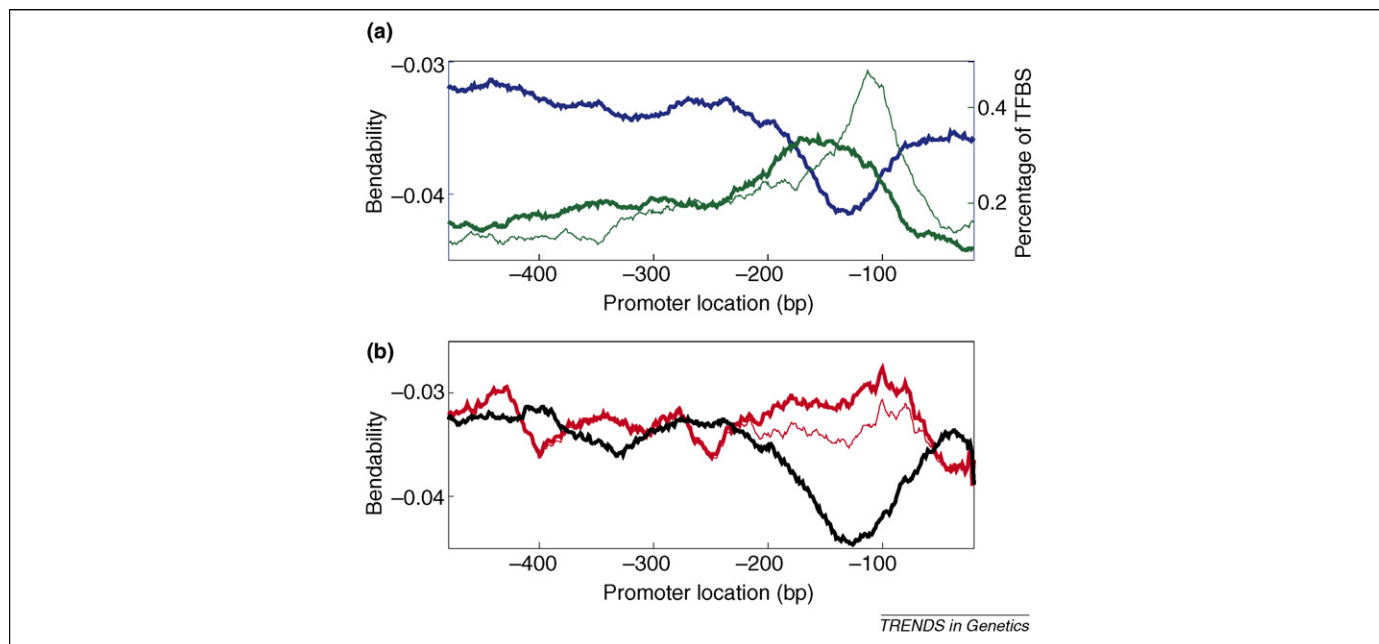
This region of low bendability is also apparent when averaging the bendability patterns over all promoters (Figure 1a). Thus, *S. cerevisiae* promoters typically contain a region of rigid DNA located 100–200 bp upstream of the start codon. The rigidity of this region is correlated with enrichment of A/T nucleotides. However, this A/T richness cannot fully account for the rigidity of this region, which is also dependent on di- and tri-nucleotide composition (Online Supplementary Figure S3).

Notably, transcription factor binding sites are also greatly enriched 100–200 bp upstream of the start codon [1,12] (Figure 1a). This proximity of binding sites to a region of rigid DNA is also apparent when individual promoters are analyzed (Online Supplementary Figure S4). Moreover, large-scale chromatin immunoprecipitation experiments [2] and computational analysis [13] have detected a nucleosome-free region (NFR) at the same location. Thus, DNA rigidity might prevent nucleosome packaging of this promoter region [5,8,14,15], thereby making it more accessible to transcription factors.

## Different architecture of TATA-containing and TATA-less promoters

We asked whether promoters that display a region of rigid DNA at 100–200 bp are distinct from those that do not. To this end, we compared the two sets of genes in terms of various functional and regulatory properties (see Online Supplementary Figure S5). Surprisingly, promoters without the rigid DNA region were greatly enriched with TATA boxes ( $P < 10^{-15}$ ). To explore this connection further, we separately analyzed promoters that contained or lacked a TATA-consensus sequence [16]. The pattern of DNA bendability differs markedly between the two promoter types (Figure 1b). Specifically, the bendability values at the typical location of rigid DNA were significantly lower at TATA-less compared with TATA-containing promoters ( $P = 6 \times 10^{-22}$ , two-sample *t*-test). Thus, the region of rigid

Corresponding author: Barkai, N. (naama.barkai@weizmann.ac.il).  
Available online xxxxxx.



**Figure 1.** Architecture of *Saccharomyces cerevisiae* promoters. TATA-less promoters contain a region of rigid DNA that is correlated with the position of transcription factor binding sites (TFBS). (a) The pattern of DNA bendability (based on arbitrary units, as previously defined [4]), averaged over all *S. cerevisiae* promoters, is shown with respect to the start codon (blue line). Also shown is the percentage of TFBS (thick green line) and TATA boxes (thin green line) at each promoter position. Each pattern was smoothed with a moving average of 30 bp. (b) Average DNA bendability of TATA-less (black line) and TATA-containing (red line) promoters. The bendability pattern of TATA-containing promoters after excluding their TATA elements is also shown (thin red line); the values at positions in which some of the promoters have a TATA box were calculated by averaging only over those promoters that do not contain a TATA box in that position.

DNA was seen only in TATA-less promoters, and this distinction remained when we controlled for the bendability of the TATA box (Figure 1b;  $P = 1 \times 10^{-13}$ , two-sample *t*-test). Furthermore, motifs that are associated with TATA-less genes, such as binding sites for the transcription factor RPN4, are highly localized to the region of rigid DNA. By contrast, other elements (e.g. binding sites for the transcription factors SWI5 and SKN7, which are primarily associated with TATA-containing promoters, show a significantly weaker localization bias (see Online Supplementary Figure S6).

An independent study has recently shown that TATA-less promoters contain a broader nucleosome-free region (NFR) compared with TATA-containing promoters [13]. Interestingly, the measure used to determine nucleosome positioning (a biased distribution of AA and TT dinucleotides within segments of 139 bp) is not correlated with our bendability measure when applied to random DNA sequences (see Online Supplementary Figure S7). This suggests that the two measures reflect complementary properties that affect nucleosome positions and maintain the differences between TATA-containing and TATA-less promoters [16,17].

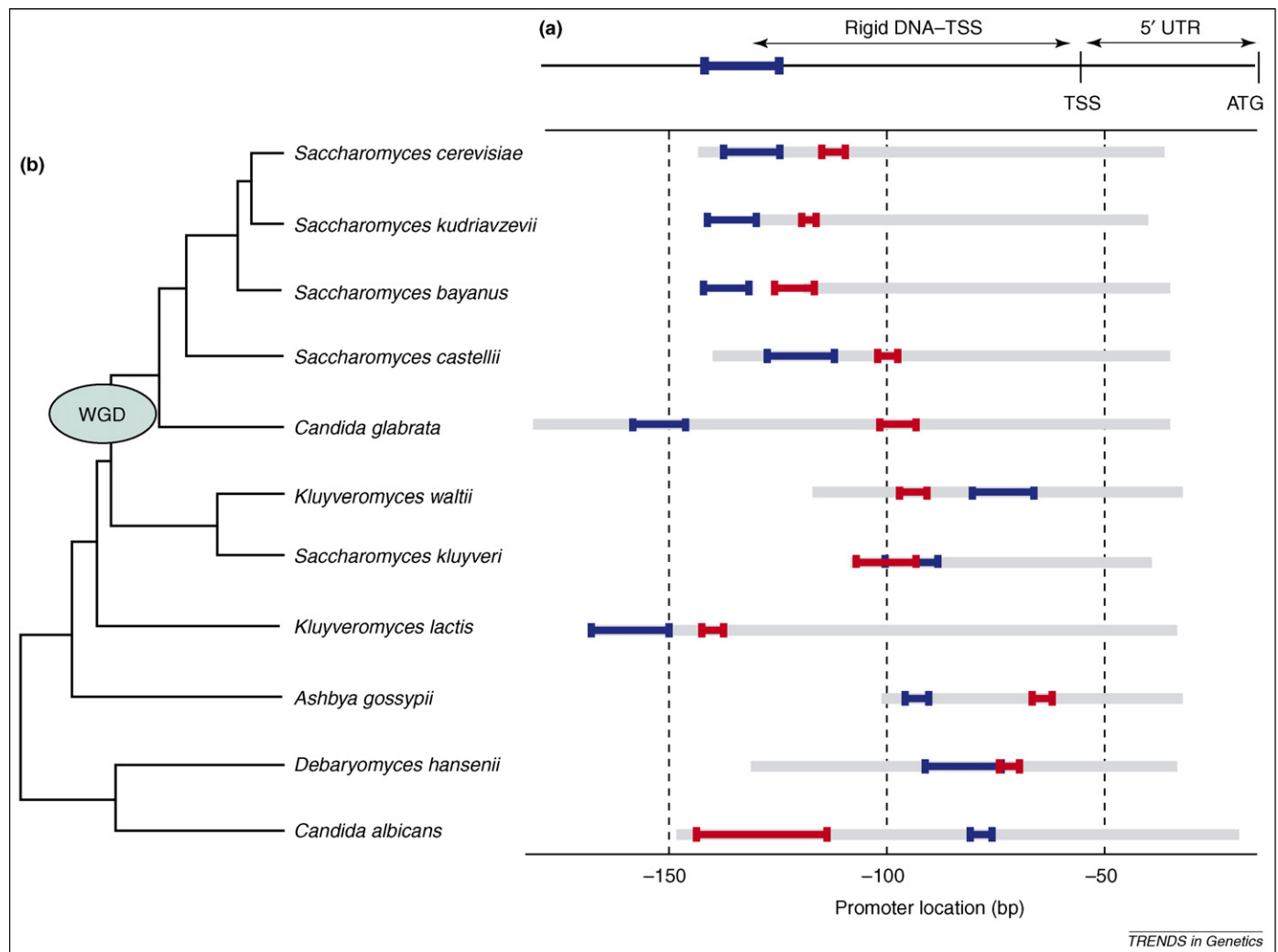
### Variation of promoter architecture among yeast species

The results above suggest that DNA rigidity has a distinct role in the regulation of genes with TATA-less promoters. To examine this association further, we analyzed the conservation of these sequence features in 11 yeast species (see Online Supplementary Material). Indeed, the general architecture was conserved (see Online Supplementary Figure S8); in all species examined, TATA-less promoters typically displayed a region of rigid DNA, and this region

was associated with enrichment of transcription factor binding sites. By contrast, typically no region of rigid DNA was observed in promoters that contained TATA boxes (see Online Supplementary Figure S9).

When we examined the general location of rigid DNA relative to the start codon, we found that it was enriched at a given position in each species, but that the distance of that position relative to the start codon varied greatly between the yeast species (Figure 2). For example, whereas in *S. cerevisiae* rigid DNA was positioned  $\sim 130$  bp upstream of the start codon, this distance ranged from  $\sim 160$  bp in *Candida glabrata* and *Kluyveromyces lactis* to  $\sim 80$  bp in *Kluyveromyces waltii*, *Candida albicans* and *Debaryomyces hansenii*. Analysis of NFRs [13] in each of these species predicts a similar interspecies variability (Figure 2; see Online Supplementary Material), suggesting that the NFRs are extended in some species (e.g. *C. glabrata*) and restricted in others (e.g. *Ashbya gossypii*). In TATA-containing genes, a similar interspecies variability was also observed in the location of the TATA box (Figure 2).

The interspecies variations that we observed could reflect either changes in the distance of the rigid DNA (or TATA box) from the transcription start site (TSS) or changes in the length of the 5'-UTR (see top of Figure 2). To distinguish between these possibilities, we examined the length distribution of experimentally determined 5'-UTRs [18] in *S. cerevisiae*, *Candida albicans* and *Schizosaccharomyces pombe*. The distributions appear to be indistinguishable between the three species (see Online Supplementary Figure S10), suggesting that the observed interspecies differences primarily reflect changes in the distance between the rigid DNA, or the TATA box, and the transcription start site.



**Figure 2.** Evolution of yeast promoter architecture. (a) The locations of rigid DNA (blue), the TATA box (red) and nucleosome-free regions (NFRs; gray), shown relative to the start codon, vary among 11 yeast species, whose predicted phylogenetic tree [23] is shown in (b). Error bars indicate standard error, calculated by bootstrapping. The whole-genome duplication (WGD) event [23] and the approximate location of the 5'-untranslated region (5'-UTR) are also indicated. TSS, transcription start site.

Thus, whereas the TATA box position is conserved at ~30 bp from the TSS in organisms ranging from plants to mammals [17,19], it is highly variable in the hemiascomycete yeasts. Furthermore, promoters in these yeasts also exhibit variability in the location of rigid DNA and NFRs and in the positions enriched in transcription factor binding sites. These global differences in the position of promoter elements among the different yeast species suggest that promoter structure is flexible and can adapt to fit the constraints of different species. Such adaptation might involve simultaneous changes in the sequence of thousands of promoters. The evolutionary pressures that could have selected for such species-specific promoter structures remain unclear.

*Saccharomyces cerevisiae* is thought to use a transcription mechanism in which RNA polymerase II first binds the promoter near the TATA box and then scans downstream until it reaches the appropriate TSS [20,21]. This scanning mechanism would relax the constraints on the distance between the TSS and either the TATA box or alternative sites such as the rigid DNA region. Thus, it is possible that this scanning mechanism enabled the rapid evolution of yeast promoter architecture that is reported here.

In summary, our results suggest that localized rigid DNA is a general sequence property of TATA-less yeast promoters. Because the TATA box serves as an anchor for the assembly of the general transcriptional machinery [22], the rigid DNA region could facilitate the binding and assembly of the transcriptional machinery in TATA-less promoters.

#### Acknowledgements

We thank Andreas Doncic, Noa Rappaport, Mark Kaganovich and Sagi Levy for helpful comments, and Ilya Ioshikhes for providing the nucleosome positioning sequence. This work was supported by NIH grant R01 DE/AI14666 (J.B.) and grants from the Kahn fund for Systems Biology at the Weizmann Institute of Science, the Tauber fund, the Israeli Ministry of Science and the Bi-national Science Foundation (BSF).

#### Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2007.03.015.

#### References

- 1 Harbison, C.T. *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 99–104

- 2 Yuan, G.C. *et al.* (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309, 626–630
- 3 Segal, E. *et al.* (2006) A genomic code for nucleosome positioning. *Nature* 442, 772–778
- 4 Brukner, I. *et al.* (1995) Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. *EMBO J.* 14, 1812–1818
- 5 Pedersen, A.G. *et al.* (1998) DNA structure in human RNA polymerase II promoters. *J. Mol. Biol.* 281, 663–673
- 6 Kanhere, A. and Bansal, M. (2005) Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes. *Nucleic Acids Res.* 33, 3165–3175
- 7 Florquin, K. *et al.* (2005) Large-scale structural analysis of the core promoter in mammalian and plant genomes. *Nucleic Acids Res.* 33, 4255–4264
- 8 Iyer, V. and Struhl, K. (1995) Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J.* 14, 2570–2579
- 9 Grove, A. *et al.* (1996) Localized DNA flexibility contributes to target site selection by DNA-bending proteins. *J. Mol. Biol.* 260, 120–125
- 10 Starr, D.B. *et al.* (1995) DNA bending is an important component of site-specific recognition by the TATA binding protein. *J. Mol. Biol.* 250, 434–446
- 11 Matthews, K.S. (1992) DNA looping. *Microbiol. Rev.* 56, 123–136
- 12 Hughes, J.D. *et al.* (2000) Computational identification of cis-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J. Mol. Biol.* 296, 1205–1214
- 13 Ioshikhes, I.P. *et al.* (2006) Nucleosome positions predicted through comparative genomics. *Nat. Genet.* 38, 1210–1215
- 14 Mai, X. *et al.* (2000) Preferential accessibility of the yeast his3 promoter is determined by a general property of the DNA sequence, not by specific elements. *Mol. Cell. Biol.* 20, 6668–6676
- 15 Sekinger, E.A. *et al.* (2005) Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* 18, 735–748
- 16 Basehoar, A.D. *et al.* (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell* 116, 699–709
- 17 Tirosh, I. *et al.* (2006) A genetic signature of interspecies variations in gene expression. *Nat. Genet.* 38, 830–834
- 18 Mignone, F. *et al.* (2005) UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.* 33, D141–D146
- 19 Struhl, K. (1995) Yeast transcriptional regulatory mechanisms. *Annu. Rev. Genet.* 29, 651–674
- 20 Kuehner, J.N. and Brow, D.A. (2006) Quantitative analysis of *in vivo* initiator selection by yeast RNA polymerase II supports a scanning model. *J. Biol. Chem.* 281, 14119–14128
- 21 Giardina, C. and Lis, J.T. (1993) DNA melting on yeast RNA polymerase II promoters. *Science* 261, 759–762
- 22 Smale, S.T. and Kadonaga, J.T. (2003) The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 72, 449–479
- 23 Wolfe, K. (2004) Evolutionary genomics: yeasts accelerate beyond BLAST. *Curr. Biol.* 14, R392–R394