Supplement Material

A Kinetic scheme

Different steps in transcription initiation can be presented by the following reactions[26]:

$$\left[RNAP\right]+\left[P\right]\underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}}\left[RNAP-P\right]_c\xrightarrow{k_f}\left[RNAP-P\right]_o\xrightarrow{k_e}\left[RNAP\right]_e+\left[P\right] \tag{S.1}$$

In the above scheme [$RNAP$], [P], [$RNAP$-P]$_c$ and [$RNAP$-P]$_o$ denote, respectively, RNAP, promoter DNA, RNAP-promoter closed complex and RNAP-promoter open complex; k_{on} and k_{off} are on and off rates of closed promoter-RNAP complex formation; k_f is the transition rate from closed to open complex, while k_e is the rate of promoter escape, which incorporates abortive transcription initiation and promoter clearance. Binding affinity of RNAP to dsDNA is defined as K_B=k_{on}/k_{off}.

The first reaction in Eq. (S.1) is reversible binding of RNAP to dsDNA of the core promoter, which is referred to as the closed complex formation. This binding leads to promoter melting, i.e. a transcription bubble-corresponding roughly to positions -12 to +2 is formed[5] (+1 corresponds to transcription start) and the open complex is formed, which is presented by the second reaction in Eq. (S.1). The last reaction in Eq. (S.1) is promoter escape, during which RNAP clears the promoter and enters the elongation.

Kinetic parameters K_B and k_f can be used to estimate the rate of transcription initiation for most promoters, since there is typically a separation of time-scales to fast binding and unbinding of RNAP to dsDNA (~1s), and slow transition from closed to open complex (~100s)[26]. It is also generally assumed that the open complex formation is rate limiting in transcription initiation which has been valid in all cases where explicitly checked[26, 32]. Under these assumptions, from the kinetic scheme presented by Eq. (S.1) follows that the transcription initiation rate is given by the following expression[7]:

$$\varphi=\frac{\left[RNAP\right]K_B}{1+\left[RNAP\right]K_B}k_f\,, \tag{S.2}$$

where $\left[RNAP\right]$ is concentration of free RNAP in cell. Furthermore, in the widely used unsaturated approximation ($\left[RNAP\right] \ll 1/K_B$)[18a, 23]the transcription rate reduces to $\varphi \approx \left[RNAP\right] K_B k_f$, so that it becomes directly proportional to $K_B\, k_f$; note that this product of the binding affinity and the transition rate corresponds to a common measure of the promoter strength[26].

B Calculation of the kinetic parameters

We here explicitly summarize equations that are used to calculate the kinetic parameters. Binding affinity K_B depends on interactions of σ[70] with, respectively, dsDNA of -35 element and -10 element, and on the length of the spacer between these two elements[5, 7]:

$$\log\left(K_B\left(S\right)\right) \sim c - \frac{\Delta G_{ds}\left(S_{(-35)}\right) + \Delta G\left(\gamma\right) + \Delta G_{ds}\left(S_{(-10)}\right)}{k_B T} \qquad \text{(S.3)}$$

In the equation above, $S_{(-35)}$, $S_{(-10)}$ and γ denote, respectively, sequences of -35 box, of -10 box, and the spacer length; $\Delta G_{ds}\left(S_{(-35)}\right)$ and $\Delta G_{ds}\left(S_{(-10)}\right)$ are interaction energies of σ4.2 and σ2.4 with, respectively, -35 box and -10 box dsDNA; $\Delta G\left(\gamma\right)$ are energy differences associated with variable spacer length between the -35 box and the -10 box; *c* is a sequence independent constant that does not enter the analysis since we calculate the binding affinities relative to the sequence of lacUV5 promoter.

To obtain the relationship between k_f and the interaction energies we use an explicit mechanistic model of the open complex formation[7]:

$$\log\left(k_f\left(S_{(-10)}\right)\right) = c + \frac{\Delta G_m\left(S^*_{(-10)}\right) + \Delta G_{ds}\left(S^*_{(-10)}\right) - \Delta G_{ss}\left(S^*_{(-10)}\right)}{k_B T} \qquad \text{(S.4)}$$

In the expression above, $S^*_{(-10)}$ denotes the sequence corresponding to positions from -11 to -7, which is the portion of the -10 box that is melted during the open complex formation. One should note that the most upstream base of the -10 region (-12) remains double stranded in the open complex, which is why $S^*_{(-10)}$ does not include base -12[12]. The energy terms are denoted as follows: $\Delta G_m\left(S^*_{(-10)}\right)$ is the melting energy of -10 region of promoter DNA in the absence of RNAP, which originates from Watson-Crick base-pairing and stacking interactions; $\Delta G_{ds}\left(S^*_{(-10)}\right)$ is the sequence-specific interaction energy of σ subunit with -10 region dsDNA in the closed complex; the interaction energy of σ subunit with the non-template strand of -10 region in the open complex is denoted by $\Delta G_{ss}\left(S^*_{(-10)}\right)$. Similarly as in Eq. (S.3), c is a sequence independent additive constant, which can be set by either measuring k_f for a provisional sequence, or eliminated by calculating the transition rates relative to a reference sequence (in our analysis we calculate all the kinetic parameters relative to lacUV5, which has consensus -10 element sequence). One should note that the signs of all energy terms in Eq. (S.4) are such that more negative terms correspond to stronger interactions. Therefore, stronger interaction energy of σ with -10 box dsDNA and larger energy needed to melt the -10 region in the absence of RNAP decreases k_f, while the stronger interaction energy of σ with -10 box ssDNA increases k_f.

Consequently, in the unsaturated approximation, transcription activity of sequence S can be expressed in terms of the interaction energies in the following way:

$$\log\left(\varphi(S)\right)=c+\frac{-\Delta G_{ds}\left(S_{(-35)}\right)-\Delta G(\gamma)-\Delta G_{ds}\left(S_{-12}\right)+\Delta G_m\left(S^*_{(-10)}\right)-\Delta G_{ss}\left(S^*_{(-10)}\right)}{k_B T} \qquad \text{(S.5)}$$

The terms in Eq. (S.5) are defined above, with the exception of $\Delta G_{ds}\left(S_{-12}\right)$, which is the interaction energy of σ2.4 with the base appearing at position -12; note that $S^*_{(-10)}$ does not include base -12, and $\Delta G_{ds}\left(S_{-12}\right)$ appears separately in Eq. (S.5), since, as noted above, the most upstream base of -10 region (-12) remains double-stranded in the open complex. Similarly as in Eqs. (S.3) and (S.4), c is a sequence independent additive constant, which does not enter our analysis since we calculate all the kinetic parameters (including $\varphi(S)$) relative to the sequence of lacUV5 promoter. Note that interactions of σ^{70} with dsDNA of -35 element ($\Delta G_{ds}\left(S_{(-35)}\right)$), and the energies associated with changing the spacer length $\Delta G(\gamma)$ were, to our knowledge, not measured until now. However, since we vary only the sequence of -10 element, these two terms remain constant, and do not enter our analysis.

C Model parameterization

To parameterize Eqs. (S.3), (S.4) and (S.5), we use a widely used independent nucleotide approximation[19b, 28], according to which the interaction energies are given by the sum of the terms that correspond to different bases at different positions. Furthermore, in this study we vary only the sequence of -10 element, so that all the terms that do not contain -10 element sequence in Eqs. (S.3), (S.4) and (S.5) become sequence independent additive constants (see above). Consequently, by using Eqs. (S.3), (S.4) and (S.5), the relevant kinetic parameters (binding affinity, the rate of transition from closed to open complex, and the rate of transcription initiation), can be expressed in the following way[7]:

$$\log\left(K_B\left(S\right)\right) \sim c - \sum_{i=1}^{6}\sum_{\alpha=1}^{4}\left(\Delta G_{i,\alpha}^{(ds)} / k_B T\right) S_{i,\alpha} \qquad \text{(S.6)}$$

$$\log\left(k_f\left(S_{(-10)}\right)\right) = c + \sum_{i=2}^{6}\sum_{\alpha=1}^{4}\left(\Delta G_{\alpha}^{m} / k_B T + \Delta G_{i,\alpha}^{(ds)} / k_B T - \Delta G_{i,\alpha}^{(ss)} / k_B T\right) S_{i,\alpha} \qquad \text{.(S.7)}$$

$$\log\left(\varphi\left(S\right)\right) = c + \sum_{i=1}^{6}\sum_{\alpha=1}^{4} \Delta G_{i,\alpha}^{(eff)} \, S_{i,\alpha} \ , \qquad \text{(S.8)}$$

where in the last equation we introduced the effective binding energy $\Delta G_{i,\alpha}^{(eff)}$:

$$\Delta G_{i,\alpha}^{(eff)} \equiv \begin{cases} \left(-\Delta G_{i,\alpha}^{(ss)} + \Delta G_{\alpha}^{(m)}\right) / k_B T \ \ for\ i \in \left(2,6\right) \\ -\Delta G_{i,\alpha}^{(ds)} / k_B T \ \ for\ i = 1 \end{cases} \qquad \text{(S.9)}$$

In the equations above, the index i denotes different positions within the -10 box, so that $i = 1$ corresponds to the position -12, while $i = 6$ corresponds to the position -7, relative to the transcription start site. Further, α denotes the four different bases (A, T, C or G), while $S_{i,\alpha}$ is equal to one if base α is present at position i in sequence S, and is equal to zero otherwise. The notation for the interaction energies ($\Delta G_{\alpha}^{(m)}$, $\Delta G_{i,\alpha}^{(ss)}$, $\Delta G_{i,\alpha}^{(ds)}$) is the same as in Eqs. (S.3), (S.4) and (S.5), in particular: *i)* $\Delta G_{\alpha}^{(m)}$ denotes the melting energies of different bases *ii)* $\Delta G_{i\alpha}^{(ss)}$ denotes the interaction energies of σ with different bases at different positions of the non-template strand. in the open complex *iii)* $\Delta G_{i\alpha}^{(ds)}$ denotes the interaction energies of σ with different bases at different positions of duplex DNA for both -10 box and -35 box.

Parameters of DNA melting have been extensively experimentally measured, which allows inferring $\Delta G_{\alpha}^{(m)}$. Note that due to the symmetry of the two DNA strands $\Delta G_{A}^{(m)} = \Delta G_{T}^{(m)}$ and $\Delta G_{C}^{(m)} = \Delta G_{G}^{(m)}$, so that there are effectively two parameters that determine melting energy in the single nucleotide approximation[9]. Furthermore, measurements of RNAP binding to -10 region DNA in both duplex form, and in the form that mimics the intermediate open complex, were done for all 3*6 single-base mutants of the consensus -10 box[33]. These measurements allow inferring interaction energies $\Delta G_{i\alpha}^{(ss)}$ and $\Delta G_{i\alpha}^{(ds)}$ for -10 box. Inference of the DNA interaction parameters is described in detail in[7], and these parameters are used for the analysis presented here.