Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2614

Bind at First Sight?

Specificity, Stability, Inducibility and Search Kinetics in lac repressor-DNA binding

JINWEN YUAN

ACTA UNIVERSITATIS
UPSALIENSIS
2025

ISSN 1651-6214
ISBN 978-91-513-2671-9
urn:nbn:se:uu:diva-571679

UPPSALA
UNIVERSITET

Dissertation presented at Uppsala University to be publicly examined in B21, BMC, UPPSALA, Friday, 16 January 2026 at 09:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Björn Högberg (Karolinska Institutet).

Abstract

Yuan, J. 2025. Bind at First Sight? Specificity, Stability, Inducibility and Search Kinetics in *lac* repressor-DNA binding. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 2614. 69 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-2671-9.

T Transcription factors (TFs) must locate and selectively bind their DNA targets while remaining responsive to cellular signals. In my thesis project, I used the lac repressor (LacI) as a model to investigate the dominant microscopic parameter for sequence specificity and study how conformation switch bias influences TF-DNA interaction phenotypes. I also developed a new methodology for high-throughput mapping of these phenotypes across a vast protein sequence space. In **Paper I**, a three-state model was derived, which links macroscopic association (ka) and dissociation (kd) rates to microscopic determinants, and predicts an anti-correlation between ka and kd. High-throughput kinetics on protein-binding microarrays (HT-k-PBMs) across 2,479 Lac operator variants confirmed this anti-correlation. We found that the inferred variation in recognition probability (ptot) exceeded that of the microscopic off-rate (koff,μ) by ~1.7-fold, conclusively demonstrating that sequence specificity is governed primarily by association. In **Paper II**, we tested a mechanistic hypothesis that ptot is set by the conformational switch centered on the LacI hinge region, which was proposed as the mechanism to balance speed and stability of target binding. We engineered two hinge-helix mutants of LacI—V52A, which increases helical propensity, and Q55N, which decreases it—and quantified their kinetic phenotypes both in vitro using HT-k-PBM and in vivo by measuring repression strength with the Miller assay and target search and binding rates using single-molecule measurements in living bacterial cells. Relative to WT-LacI, the in vitro macroscopic kinetics (ka, kd, Kd) of the engineered mutants shifted in opposite directions. This translated to distinct in-cell performances: V52A strengthened repression but with reduced DNA specificity and loss of inducibility, while Q55N weakened repression but increased specificity and retained inducibility. Notably, neither variant measurably altered the target search speed in cells relative to WT-LacI. Finally, in **Paper III**, we present an Optical Pooled Screening (OPS) method that combines chromosomal dual barcodes with pooled λ-Red recombineering to scale single-molecule phenotyping to many LacI variants expressed from the chromosome. In the pilot study, 5/6 strains were correctly decoded, and the expected phenotypes were recovered. We also outlined key constraints needing design refinements prior to full-scale implementation.

Keywords: Specificity, Stability, Seach speed, Lac repressor, inducibility

Jinwen Yuan, Department of Cell and Molecular Biology, Molecular Systems Biology, Box 596, Uppsala University, SE-751 24 Uppsala, Sweden. Science for Life Laboratory, SciLifeLab, Box 256, Uppsala University, SE-75105 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-513-2671-9
URN urn:nbn:se:uu:diva-571679 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-571679)

致我的父母：郁建玉，袁玉成，
感谢你们一路的用心栽培与无尽的爱，
赋我以行向天地的自由。
谨以此文献给你们：愿知辽阔，亦知归处。

To friends and families,
For wonderful company during times of fun ,
for great support whenever.
I always cherished and will cherish every moment with you!

To Viktor,
Thank you for being my calm and constant;
with you, Nordic skies and still pine forests feel like home.

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. Marklund E., Mao G., **Yuan J.**, Zikrin S., Abdurakhmanov E., Deindl S., Elf J. (2022) Sequence specificity in DNA binding is mainly governed by association. *Science* 375(6579):442–445.

II. **Yuan J.**, Lüking M., Zikrin S., Sen BC., Marklund E., Fange D., Elf J. (2025) The LacI hinge region balances binding stability against inducibility. *bioRxiv (Preprint)*

III. **Yuan J**., Kandavalli V., Schirman D., Zikrin S., Elf J. An optical pooled screening platform for single-molecule measurements in live cells: A pilot study. *Manuscript*

Reprints were made with permission from the respective publishers.

Contents

Abbreviations

DNA	Deoxyribonucleic acid
EMSA	Electrophoretic Mobility Shift Assay
HT-k-PBM	High-throughput kinetics measurement on Protein-binding microarray
IPTG	Isopropyl **β**-D-1-thiogalactopyranoside
LacI	Lac Inhibitor / Repressor
MD	Molecular Dynamics
MFPT	Mean First Passage Time
mRNA	messenger RNA
O_1	Natural Lac Operator
O_2	Natural Lac Operator
O_{sym}	Synthetic Lac Operator
ODE	Ordinary Differential Equation
ONPG	O-Nitrophenyl-**β**-D-galactopyranoside
PBM	Protein-binding microarray
RNAP	RNA polymerase
V52A	LacI hinge variant with Valine-52 to Alanine substitution
Q55N	LacI hinge variant with Glutamine-55 to Asparagine substitution
Wt (or WT)	Wild-type LacI

1. Introduction

1.1. We Are Programmed Beings

Imagine a single fertilized egg cell blossoming into a complex human with trillions of cells. How does this one cell give rise to muscle, nerve, and blood cells, all with distinct shapes and functions? The secret lies in gene expression control; even though nearly every cell contains the same DNA blueprint, each cell type “reads” different parts of that blueprint. In a liver cell, one set of genes is active while others stay silent; in a skin cell, a different subset of genes is turned on.

We are, in a sense, programmed beings: cells execute distinct gene-expression pipelines from the same genome codebase, producing diverse phenotypes.

Gene expression must be exquisitely regulated in a coordinated manner to build and maintain a healthy organism. If this regulation breaks down, the consequences can be severe. Genes expressed when they shouldn’t be, or silenced when they should be active, are hallmarks of many diseases, including cancer[1].

1.2. Transcription Factors: Regulators of Gene Expression

How does a cell “decide” which genes to express? While gene-regulatory mechanisms differ across organisms, the underlying information flow remains conserved, following the central dogma of molecular biology[2], as shown in **Fig. 1.1.** Among these steps, transcription – the copying of DNA into messenger RNA (mRNA) – is the primary control point for gene expression in both simple and complex organisms.

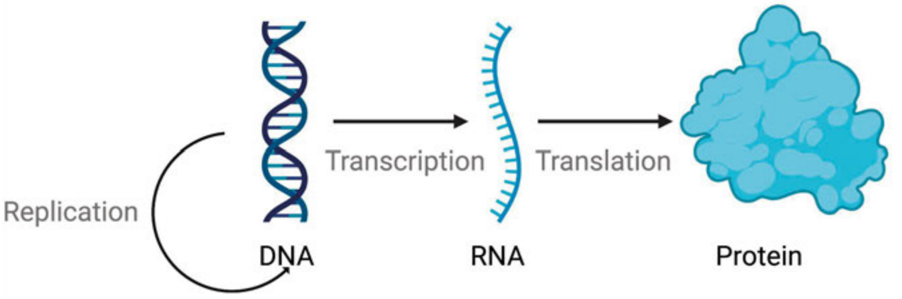

Figure 1.1. The central dogma in molecular biology (created in BioRender https://BioRender.com/ q45e576).

Transcription factors (TFs) are proteins that bind specific DNA sequences and act as switches to turn genes on or off for transcription. Some TFs are activators that boost a gene's transcription, helping to recruit or stabilize RNA polymerase (RNAP) at the promoter to copy DNA information into RNA. Others are repressors, which block transcription, for example, by preventing RNAP from binding or by recruiting proteins that compact the local DNA, making the specific DNA sequence inaccessible for RNAP[3]. TFs are essential for the regulation of gene expression and are, as a consequence, found in all living organisms.

It's important to note that transcriptional control is the primary means for regulating gene expression, but it's not the only level. Cells also control how RNA transcripts are processed, how much of them gets translated into protein, and how quickly proteins are degraded. Still, deciding whether to transcribe a gene or not is often the most significant regulatory step[4]. And because transcription factors must first and foremost find their target DNA to regulate, a fundamental question arises: How does a transcription factor find its target/specific gene? This question is the central topic of this thesis.

1.3. The target search problem

Every transcription factor faces a daunting challenge: locating its target DNA sequence with 1 bp precision, within an enormous expanse of genomic DNA (10^6 to 10^9 bp). This is often referred to as the target search problem. It is hard for us, humans, to imagine how a biomolecule searches for its target in the intracellular micro world. Unlike us, the molecules don't have eyes; no vision could help them scan one orientation for familiar faces and then move on to the next orientation to find their best buddy. One way to imagine the process is to picture yourself with a blindfold over your eyes. Your task is to find a specific person among 3.5 million people in a space station. One possible approach is to float around randomly, bump into people, feel their faces, and then decide whether to hold on to that person or move on to the next face you encounter. If a protein were to use such random 3D diffusion to search for its DNA site, it could take hours or even longer to find the target. Yet in living cells, gene regulation happens on much faster, biologically relevant timescales, often minutes or seconds, which tells us that cells have evolved efficient ways for DNA-binding proteins to locate their targets.

Over 50 years ago, researchers provided the first clues to how nature solves this puzzle. Riggs and colleagues (1970)[5] measured how quickly the lac repressor (LacI, a TF in *E. coli*) could bind to its DNA operator site, and they got a surprise: the binding rate was on the order of 10^{10} $M^{-1}\cdot s^{-1}$, roughly 100 times faster than a 3D-diffusion-limited protein-DNA reaction rate could give, explained below.

The association for protein(P) and its specific target site(S) in solution can be seen as:

$$P + S \rightleftharpoons PS$$

The macroscopic association rate of [PS] is defined as $\frac{d[PS]}{dt} = k_a[P][S]$, where [PS], [P], and [S] are the concentration of protein, target site, and bound protein on the target site, respectively. Using the theoretical diffusion-limited maximum association rate, obtained by Smoluchowski[6], with the knowledge of protein diffusion coefficient ($D \approx 5 \mu m^2 s^{-1}$) and reaction radius between protein and DNA($\rho \approx 5*10^{-9}$ m), we get

$$k_a = 4\pi D\rho \approx 2 \times 10^{-19}\ m^3\ s^{-1} \quad (1)$$

Multiplying by the Avogadro constant (6.02×10^{23} mol^{-1}) and converting to moles per liter, we get $k_a \approx 10^8 M^{-1}s^{-1}$, which is 100 times slower than what was measured by Riggs et al.,1970[5].

To resolve this discrepancy between the experimentally measured rate of 10^{10} $M^{-1}\cdot s^{-1}$ and the theoretical maximal rate limited by diffusion: $10^8 M^{-1}s^{-1}$, it was proposed that the target search process involves a shift in dimensionality, by Richter et al. 1974. This hypothesis was later formalized as facilitated diffusion by Berg et al. 1976&1981 and von Hippel et al. 1989[7–10]. Facilitated diffusion describes a hybrid strategy that TFs utilise during the target search process, where they alternate between three-dimensional (3D) diffusion in the cytoplasm and one-dimensional (1D) sliding on non-specific DNA mediated by electrostatic force. This conversion of movement was suggested to effectively increase the 'target size', ρ in equation (1), by the sliding distance.

Since the birth of the facilitated diffusion model theory, a sequence of *in vitro*[11–13] and *in vivo*[14–16] approaches have tested, verified, and complemented it. Particularly, Elf et al., 2007[14] developed *in vivo* single-molecule microscopy for resolving specifically bound versus transiently non-specifically bound LacI and quantifying residence times. Later, Hammar et al., 2012[15] quantified the sliding length, which is around 45bp, for LacI in *E. coli* and, crucially, observed that the LacI's binding probabilities differ across operator variants, indicateing sequence-dependent recognition already at the encounter stage[15]. Subsequently, *in vitro* single-molecule experiments revealed the hopping events of the Lac repressor during 1D sliding; LacI hops one or two grooves (10–20 bp) every 200–700 μs[16], providing a physical basis for operator bypass. This hopping study also directly showed that LacI can pass over even high-affinity sites multiple times before committing to a specific binding.

1.4. Gap in knowledge & study focus in Paper I

Previous studies and textbooks have viewed DNA-binding specificity as arising mainly after binding, with cognate sites forming long-lived complexes, i.e., differential dissociation dominates; association has been seen as effectively diffusion-limited and sequence-independent[17]. However, the single-molecule findings by Hammar et al. 2012[15] and Marklund et al. 2020[16] instead indicated a sequence-dependent association step. Several other studies have also suggested that the association is sequence-dependent[18–21]. What remained unclear was the relative contribution of association versus dissociation to overall DNA specificity?

This uncertainty provided the motivation for **Paper I** of this thesis. **Paper I** set out to determine, still using the LacI as the study subject, whether DNA-binding specificity is governed primarily by differences in dissociation kinetics or by differences in the association process. In other words, does LacI distinguish between operator sequences and other DNA mainly by how long it stays bound once attached, or by how likely it is to bind in the first place when it encounters each sequence?

1.5. The *lac* operon and Lac repressor

Owing to LacI's central role in early gene regulation research and the tractability of bacterial systems[22,23], it has anchored decades of work dissecting how transcription factors locate and recognize their operators. As mentioned above, many kinetic measurements and single-molecule tracking in living cells deployed LacI as the model subject for studying the search mechanisms[5,7,14–16,24]. This thesis intends to dig deeper into the search problem of TF proteins, also using LacI as a model, so before presenting the research chapters, I first present an overview of the lac operon it controls, the function and structure of the lac repressor, and its specific DNA target sites: the *lac* operators.

1.5.1. The lac operon and its function in a bacterial cell

In *E.coli*, the *lac* operon, as illustrated in **Fig. 1.2**, has three structural genes (*lacZ*, *lacY*, *lacA*) encoding a set of enzymes that are responsible for importing and degrading lactose, a single promoter that is for RNA polymerase (RNAP) to bind and transcribe the operon, and the operator sequences that is the regulatory unit for allowing or blocking the transcription. The transcriptional control is mediated by the binding of the regulatory TF: LacI. When lactose is not present in the surroundings of the *E.coli* cells, free LacI molecules have a high affinity for the operator sequence, thus blocking the production of lactose-processing enzymes. To conserve resources, cells restrict protein synthesis to conditions in which the product is actually needed or useful. When lactose is available, LacI binds allolactose (a degradation product of lactose), thereby losing its ability to bind stably to DNA. Then the RNAP can transcribe the three lactose-processing enzymes without the hindrance of LacI bound between the promoter and the genes, thereby allowing the uptake and metabolism of lactose. This regulatory circuit – discovered by Jacob and Monod in the 1960s – was the first illustration of gene regulation by DNA-binding proteins and operators[25,26].

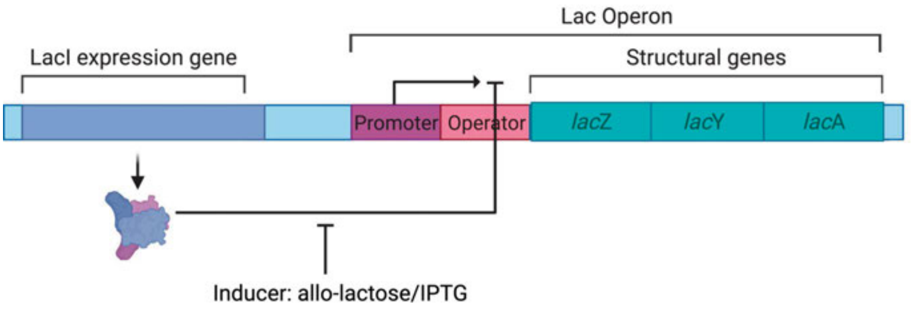

Figure 1.2 LacI operon circuit model. The *lacI* gene codes for the Lac repressor protein, which binds to the operator region to block transcription of the *lacZYA* structural genes (created in BioRender https://BioRender.com/ q45e576). Inducers like allolactose or IPTG bind to LacI, releasing it from the DNA and allowing gene expression

1.5.2. Structure of LacI

Each monomer of LacI consists of four distinct regions. The N-terminal DNA-binding domain (residues: 1-45)[27] is followed by a hinge region (residues: 52-60) that links the DNA binding domain to core domain (residues: 62-332). The core domain contains a binding pocket for the natural inducer allolactose or artificial inducer IPTG (Isopropyl β-D-1-thiogalactopyranoside[27,28]) and also includes a region responsible for dimerization (residues: 275–290). The dimer is the minimal unit for binding to lac operators[28,29]. Finally, the C-terminal tail (residues: 339–360) mediates the tetramerization[27,30].

1.5.3. lac operators

In the *E.coli* genome, the *lac* operator regulatory region is composed of three operator sites: O_1, O_2, and O_3. The primary operator, O_1, is situated directly at the transcription start site[31], while the additional operators, O_2 (located 401 bp downstream) and O_3 (located 93 bp up stream), stabilize repression by enabling the LacI tetramer to bind two sites simultaneously, thus forming a repressive DNA loop. All three lac operators are necessary for a maximum repression *in vivo*[32–35]. In addition to the natural *lac* operators, researchers often use the synthetic O_{sym} for experimental studies. This synthetic sequence[27,36] was designed to achieve the strongest possible binding affinity with the LacI dimer. The consensus ranking for the binding affinity with the LacI dimer is: $O_{sym} > O_1 > O_2 > O_3$.[15,36,37]

2. Paper I: DNA specificity is mainly governed by association (binding probability)

2.1. Relationship between k_a and k_d with the three-state model

In **Paper I** (Marklund et al., 2022), we addressed the question: What contributes to sequence specificity—the dissociation rate (which determines how long a TF remains bound), the association rate (which determines how likely a TF is to bind a given site), or both?

To answer this question, we adopted the classical model where the protein can slide into the target sequence through nonspecific interactions[7], and modelled this process as a three-state model (see supplementary info in **Paper I**), as shown in **Fig. 2.1A**.

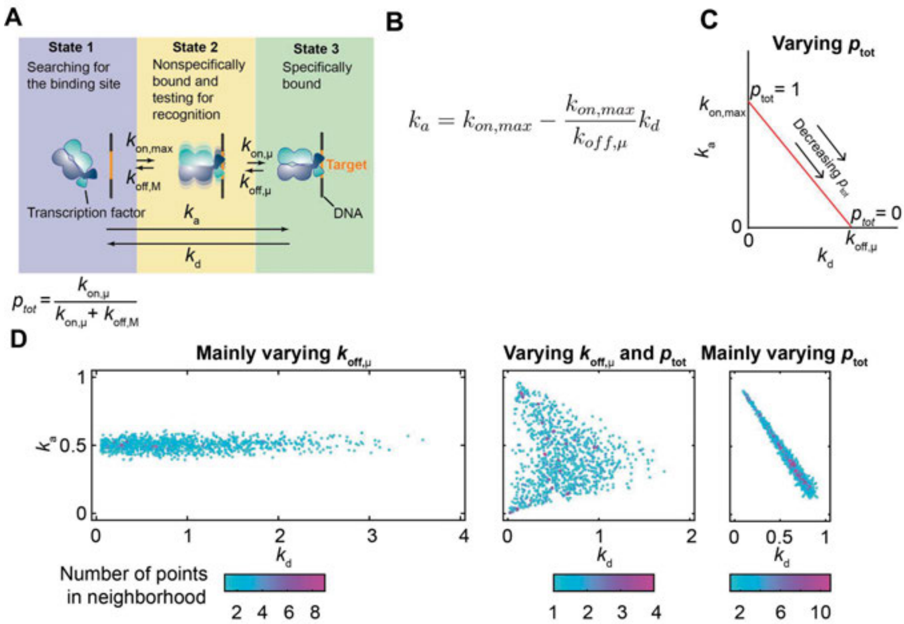

Figure 2.1. Bimolecular association and dissociation rates are inherently coupled as a result of target site testing. (Adapted from Paper I Fig 1) **A.** Schematic of a three-state kinetic model describing association and dissociation linked with microscopic rates. **B.** Derived equations revealing k_a and k_d are inherently coupled through microscopic rates (see detailed derivation process in the supplementary text of **Paper I**), assuming that State 2 is very short-lived. **C.** The relationship between k_a and k_d shown in B. becomes linearly anticorrelated when $k_{off,\mu}$ is constant and p_{tot} changes (red line). **D.** Effect on k_a and k_d when varying: $k_{off,\mu}$ 10 times more than p_{tot} (left), $k_{off,\mu}$ and p_{tot} to the same extent (center), and p_{tot} 10 times more than $k_{off,\mu}$ (right), in simulations of the model. Each plot contains 1.000 points, where each point represents one target site with a randomly sampled ($k_{off,\mu}$, p_{tot}) (See simulation method in Paper I's supplement)

In this three-state model, a TF's interaction with a DNA fragment is treated as follows: First, it diffuses freely in the 3D cellular space (state1). It can attach to a DNA fragment nonspecifically (with an sequence-independent association rate $k_{on,max}$) and start testing (state1→state2). When the TF is bound nonspecifically to the DNA fragment (state2), it can either

(1) go back to state1 (state2→state1) with a sequence-independent microscopic dissociation rate: $k_{off,M}$ or

(2) get stably bound on this DNA fragment (state2→state3) with a sequence-dependent microscopic association rate: $k_{on,\mu}$.

Recognition probability (p_{tot}) thus defines the chance of LacI choosing scenario (2) when in state 2, $p_{tot} = \frac{k_{on,\mu}}{k_{off,M} + k_{on,\mu}}$, and therefore, the chance of dissociating back into solution, scenario (1), is 1- p_{tot}. Once specifically bound (state 3), the TF can still leave the specifically bound state and return to the testing state (state 3→state2) with the microscopic unbinding rate $k_{off,\mu}$. With this model, we obtained an explicit equation (shown in **Fig. 2.1B**) showing that k_a and k_d are inherently coupled by two microscopic parameters ($k_{on,max}$ and $k_{off,\mu}$). Notably, if assuming $k_{off,\mu}$ is a constant, as shown in **Fig. 2.1C**, then an increase in the p_{tot} will increase the k_a and decrease k_d, which actually reflects the intuitive trade-off that easier recognition binding leads to longer retention on average.

As illustrated in **Figure 2.1 D**, different major sources contributing to DNA specificity yield characteristically different distributions of (k_a, k_d) values in computational simulation. If DNA specificity were dominated solely by differences in the dissociation step (a $k_{off,m}$-dominated scenario), one would expect a broad scatter of points with no strong correlation between k_a and k_d. Instead, if specificity were governed largely by differences in the recognition step (a p_{tot}-dominated scenario), the model predicts a near-linear anti-correlation between k_a and k_d (high k_a accompanied by low k_d, and vice versa) due to the coupling imposed by p_{tot}. An intermediate scenario in which p_{tot} and $k_{off,m}$ co-vary produces a cloud pattern. Crucially, a strong anticorrelation between observed k_a and k_d emerges only in the regime where p_{tot} varies much more than $k_{off,m}$ across sequences.

To determine which scenario actually occurs, we measured k_a and k_d across thousands of DNA sequences under the same experimental conditions. We achieved these high-throughput kinetics measurements by studying protein–DNA binding on a DNA microarray (HT-k-PBM) (see method set-up, method principle in **Chapter: Methodology**, and experiment details in **Paper I**'s methods).

2.2. *In vitro* high-throughput kinetic measurements on PBM proved k_a anticorrelates with k_d

Using the HT-k-PBM approach, we measured the k_a and k_d for ~2,479 different DNA sequences under the same *in vitro* conditions, see **Fig. 2.2A&B**. In this assay, fluorescently labeled LacI proteins were introduced over a DNA microarray slide containing thousands of DNA sequences (including the natural O_1 and O_2 operators, the high-affinity synthetic operator O_{sym}, and all single and numerous double O_{sym}, O_1, O_2 operator mutants). The microarray slide was prepared for protein binding according to the classic PBM protocol[38]. We monitored the time-traces of binding and unbinding of LacI (Labelled with Cy3, see labelling method in Paper I's methods) to each DNA sequence in movies recorded via fluorescence microscopy, see **Fig. 2.2 B**, and obtained the k_a and k_d rates from the recorded time-traces for each DNA sequence. The results showed a clear anticorrelation between k_a and k_d, averaged from two PBM replicates for each operator (see **Fig. 2.2D**). This outcome was exactly as the p_{tot}-dominated model scenario had predicted, indicating that differences in recognition probability (p_{tot}) were driving the sequence specificity. See quality of reproducibility between two replicates in **Fig 2.2C**.

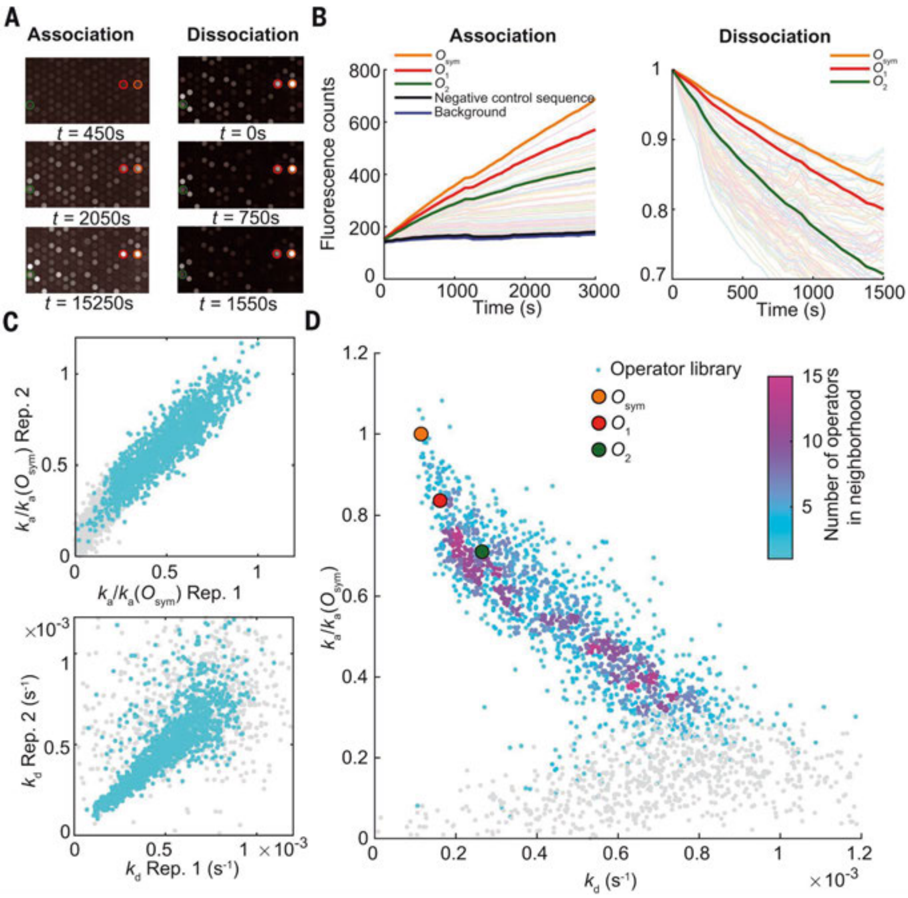

Figure 2.2. Macroscopic association (k_a) and dissociation (k_d) rates measured with high-throughput kinetics measurement using the PBM method (Taken from Paper I Fig.2). A. Example images taken during association (left) and dissociation (right) of LacI-Cy3 to spots on the DNA microarray. Orange circle, O_{sym} operator; red circle, O_1 operator; green circle, O_2 operator. **B.** Association and dissociation curves for the O_{sym}, O_1, and O_2 operators (thick lines), and 100 examples of their mutants (faint colors). The dissociation curves for each operator were normalized to the fluorescence count for the first frame of the corresponding dissociation movie **C.** Reproducibility of k_a (top) and k_d (bottom) for individual operators (cyan points) between replicates. The k_a values are all normalized to the k_a value of O_{sym} in that replicate. Sequences associated with weak binding (fluorescence signal at equilibrium <3% of signal for O_{sym}; gray points). **D**. Measured association and dissociation rates for wild-type operators (circles) and their single and double mutants [points are colored by operator density in that (k_a, k_a) neighborhood; gray points are the same as in C. The value for each operator is the mean from two PBM replicate experiments; see **Paper I**'s Fig. S1B for data from individual replicates.

Quantitative analysis was done to estimate the variation of the inferred microscopic rates (p_{tot} and $k_{off,\mu}$) from the measured k_a and k_d in each PBM replicate with equation (shown in Fig. 2.1B). The range of variation in p_{tot} across sequences was roughly 1.7-fold greater than the variation in $k_{off,m}$. This suggests that LacI discriminates between DNA sites mainly through differences in how effectively it recognizes each site, rather than through differences in how long it is retained on different DNA sites.

2.3. Additional evidence (brief summary)

Other studies were also done to test and validate the model and equation we mentioned above. The method principles and main conclusions are summarized here.

Published *in vivo* estimates of LacI association and dissociation (k_a, k_d) at O_{sym}, O_1, O_2, and O_3 were compiled and reanalyzed by plotting k_a (y-axis) versus k_d (x-axis). Most operators aligned along an anti-correlated trend consistent with a p_{tot}-dominated scenario; O_3 deviated due to large uncertainty in its single-molecule k_d measurement. Using the available *in vivo* $k_{on,max}$ [15] together with operators' (k_a, k_d), the microscopic off-rate $k_{off,\mu}$ was inferred with the equation shown in **Fig. 2.1B**. In agreement with the HT-k-PBM results, $k_{off,\mu}$ varied little across operators (again with O_3 being the exception). Because the published *in vivo* data set is small, this analysis alone was not sufficient to establish p_{tot} as the dominant source of specificity. To perturb p_{tot} directly without varying the target DNA sequence, additional *in vitro* single-molecule and SPR (Surface Plasma Resonance) with salt-titration experiments were performed with O_1. Changing ionic strength modulates the dwell time in nonspecific sliding [39], thereby altering operator-encounter frequency in one sliding event and, as expected, altering p_{tot}. Across measured k_a and k_d in different salt conditions, the k_a–k_d anti-correlation trend persisted.

Other than evidence shown in **Paper I**, Kandavalli et al. (2025)[40] inferred *in vivo* association and dissociation rates for 35 LacI binding sites by analyzing mRNA production from promoter-occlusion reporter constructs. In these constructs, a lac operator is placed immediately downstream of the constitutive apFAB120 promoter and controls expression of an mVenusNB reporter: when LacI is bound, it physically blocks transcription from the promoter, and when LacI is unbound, transcription proceeds and mRNA is produced. By fitting a two-state stochastic gene expression model to the single-cell mRNA copy-number distributions measured for each construct, they extracted the corresponding k_a and k_d values of LacI to each lac operator. Their analysis also revealed a clear anti-correlation between k_a and k_d across the variants of lac operators they tested, further supporting the p_{tot}–dominated scenario for sequence specificity.

3. Speed–stability paradox and the conformation-switch solution

In **Paper I**, we showed that for LacI, whose binding behaviour can be described with a three-state search scheme (3D diffusion, short-lived 1D search, and recognition), the sequence specificity is mainly governed by the recognition probability (p_{tot}) because the microscopic dissociation rate ($k_{off,\mu}$) changes comparatively little across DNA sequences. We now know that variation in p_{tot} primarily determines DNA sequence specificity, but what biophysical mechanism sets these differences in p_{tot}? A possible mechanism is the sequence-dependent modulation of conformational switch probability between the testing(search) state and the recognition state. Theoretical studies have suggested that such a conformational switch is needed for a TF protein to solve the rapid target search and stable target binding paradox[7,24]. The authors reasoned that a single rigid conformation of protein on DNA cannot deliver both fast sliding and stable yet sequence-selective binding, since fast search needs a smooth sequence landscape ($\sigma \approx 1–2$ kBT), whereas high equilibrium occupancy requires a rough landscape ($\sigma \gtrsim 5$ kBT)[24]. This speed–stability paradox can be resolved if the TF toggles between two states: a search (S) state enabling rapid sliding and a recognition (R) state enabling tight, sequence-specific binding. In this framework, recognition is affected by an S→R conformational transition probability that is coupled to binding.

3.1. Previous structural studies and MD studies confirmed conformation switch

Experimental and structural studies over the years have provided strong support for this conformational switching hypothesis. In particular, LacI has been shown to adopt two different conformational states when bound to DNA non-specifically versus specifically. Using NMR, Kalodimos et al. (2004) determined that the LacI headpiece (DNA binding domain + Hinge region) bound to nonspecific DNA forms a 'search-state' complex in which some regions, including the hinge, remain unstructured and the bound DNA stays essentially straight (B-form). In contrast, X-ray structures of LacI-operator complexes show DNA bending (≈40°) and the formation of the hinge helices. There structures are consistent with the theory that TFs adopt the search conformation when bound non-specifically, and undergoes a transition to adopt a tighter 'recognition' complex[41–43].

Importantly, the conformational-switching model is reinforced by the fact that many TFs' DNA binding ability can be stabilized or destabilized by small molecules or partner proteins. In bacteria, LacI/GalR family regulators exemplify ligand-tunable allostery, where effector binding remodels conformational equilibria and thus DNA affinity, thereby gating recognition vs. release[44]. In human proteins, MYC requires heterodimerization with MAX for sequence-specific DNA binding, an interaction that itself defines the recognition state[45]. Beyond dimerization, recent work[46] shows that intrinsically disordered regions (IDRs) within MYC/MAX mediate intramolecular interactions that modulate DNA binding. More broadly, this allosteric plasticity, and the presence of IDRs, serve as the universal molecular mechanism that allows TFs to rapidly and dynamically shift between functional conformations.

In addition to structures and experimental studies, molecular dynamics (MD) simulations have reinforced the importance of conformational switching. Lüking et al. (2022)[47] performed coarse-grained MD simulations on LacI and directly examined how its two conformations behave during one-dimensional diffusion along DNA. The simulations showed that LacI's "search conformation" is highly suited for rapid sliding: it allows agile movement and frequent micro-dissociation/re-association hops along DNA, facilitated by a looser protein–DNA interface. By contrast, the "recognition conformation" (in which certain regions of LacI, like the hinge, form helices against DNA) enforces tight groove-tracking and rotational alignment with the DNA helix, which is ideal for recognizing the target site but incompatible with fast sliding. The MD study vividly demonstrated that if LacI prematurely adopts the helical, tightly-bound conformation while scanning, its ability to slide and sample DNA rapidly is severely hindered. Conversely, retaining a flexible conformation allows quick exploration but cannot achieve high-affinity binding. These findings emphasise that a flexible state for facilitated diffusion and an ordered state for stable specific binding are necessary.

3.2. Hinge region: Mediating the switch between the recognition and search conformations

How do we probe this conformation switch process? Let's look at how the conformation of the LacI-DNA complex differs depending on if it's specifically or non-specifically bound. The LacI-DNA complex exhibits numerous structural (including DNA geometry and protein–DNA contacts) changes when transitioning from a non-specifically bound state to a specifically bound state[47]. We hypothesize that the hinge region constitutes the principal conformational distinction, undergoing a dramatic disorder-to-helix transition between the search and recognition state, as shown in **Fig. 3.1A**.

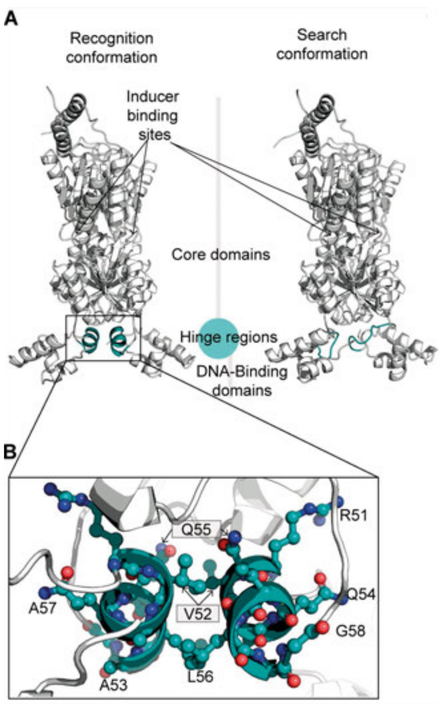

Figure 3.1. Conformation switch shows case with LacI (Adapted from Paper II Figure 1) A. Cartoon representation of LacI dimer conformations. (Left) Recognition conformation (PDB ID: 1EFA[43]) with hinge region highlighted in teal. (Right) Search conformation modeled by integrating the core domain from 1EFA with the DNA-binding domain from the search conformation structure (PDB ID: 1OSL[41]), as described in Lüking et al.[47], with hinge region similarly highlighted in teal. **B**. Cartoon and Ball-and-Stick representation of the hinge helix residues. In the Ball-and-Stick model: Carbon atoms are shown in teal, nitrogen atoms in dark blue, and oxygen atoms in red. Residue names are labeled.

Kinetic analyses further support a hinge-centric switching mechanism: the formation of the hinge helices appears to define a high-energy transition state in the LacI–operator binding pathway, marking the commitment to full specific binding[48]. Given these observations, we hypothesize that the hinge region in lacI might serve as a lever that can toggle the probability of LacI to be in search or recognition modes, balancing rapid sliding state with high-fidelity operator binding state.

Therefore, in **Paper II**, we directly probe this hypothesis by introducing point mutations in the hinge region and quantifying their impact on the DNA-LacI interaction. We quantified specificity, binding stability with both operator and non-operator sequences, allosteric inducibility, and the impact on search speed.

4. Paper II: The LacI hinge region balances binding stability against inducibility

Paper I left us with a concrete lead. By mapping association and dissociation of LacI across thousands of sequences, we saw that DNA specificity is largely governed by association; the recognition probability varied more than the microscopic dissociation rates across the operator library. This observation raises a mechanistic question: which structural element of LacI determines the recognition probability (p_{tot})?

Prior kinetic studies and structural works implicated the hinge region of LacI as the site of control. This region is disordered when LacI is non-specifically bound to DNA (search conformation), but helical in the specifically bound LacI-operator complex (recognition conformation). In **Paper II,** we tested the mechanistic hypothesis that tuning the hinge's helical propensity would directly bias the switch through observing the changes in binding stability and DNA specificity. We also examined the consequences for *in vivo* search speed and inducibility.

4.1. Rational design of hinge-helix variants to bias the switch

To bias the conformational switch, Molecular Dynamics (MD) simulations were conducted (performed by M. Lüking & D. Fange in Johan Elf's lab), which predicted the effects of specific single amino acid substitutions on the hinge region's α-helical propensity:

1. Val52 →Ala (V52A): Predicted to increase hinge helicity (biasing toward the recognition state).
2. Gln55 → Asn (Q55N): Predicted to decrease hinge helicity (biasing toward the search state).

(A third MD candidate, G58A, was excluded after showing weak

preliminary binding in an initial screen with HT-k-PBM, despite being predicted to increase hinge helicity through MD and helical propensity theory, see discussion session and Fig. S9 in ***Paper II****).*

4.2. HT-k-PBM results validate the switch bias

We began by screening WT, Q55N, and V52A *in vitro* on the same operator library used in **Paper I**, assaying these LacI variants in parallel with HT-k-PBM under matched conditions (setup described in the **Chapter: Methodology**; experimental details in **Paper II** methods). With the HT-k-PBM assay, each LacI variant's k_a and k_d were extracted for thousands of DNA sequences, allowing comparison of binding stability, sequence specificity, and inference of p_{tot}, $k_{off,\mu}$ across LacI variants.

As shown in **Fig. 4.1.A-D**, the kinetic measurements yielded a clear, converse pattern that validates the MD-predicted shifts:

Variant	MD-Predicted Shift in Helicity	Observed Kinetic Phenotype	Overall Effect
V52A	Increased (Recognition-Prone)	Associated faster (↑k_a) and dissociated slower (↓k_d)	↑Affinity (K_d) and ↓Specificity
Q55N	Decreased (Search-Prone)	Associated slower (↓k_a) and dissociated faster (↑k_d)	↓Affinity (K_d) and ↑Specificity

In summary, V52A, designed to be forced toward favoring the recognition conformation, indeed stabilized LacI binding and increased its overall affinity across the thousands of DNA sequences tested, but sacrificed its ability to discriminate against near-cognate sequences (reduced specificity). Conversely, Q55N, designed to be reluctant to adopt the recognition state, weakened the overall affinity of LacI but "sharpened" its preference for the highest-affinity sequences (increased specificity). This behavioural split echoed earlier observations by Zhan *et al.* 2006 of LacI with substitutions at

Val52, where increased helical propensity mutations strengthened binding while diminishing discrimination[49].

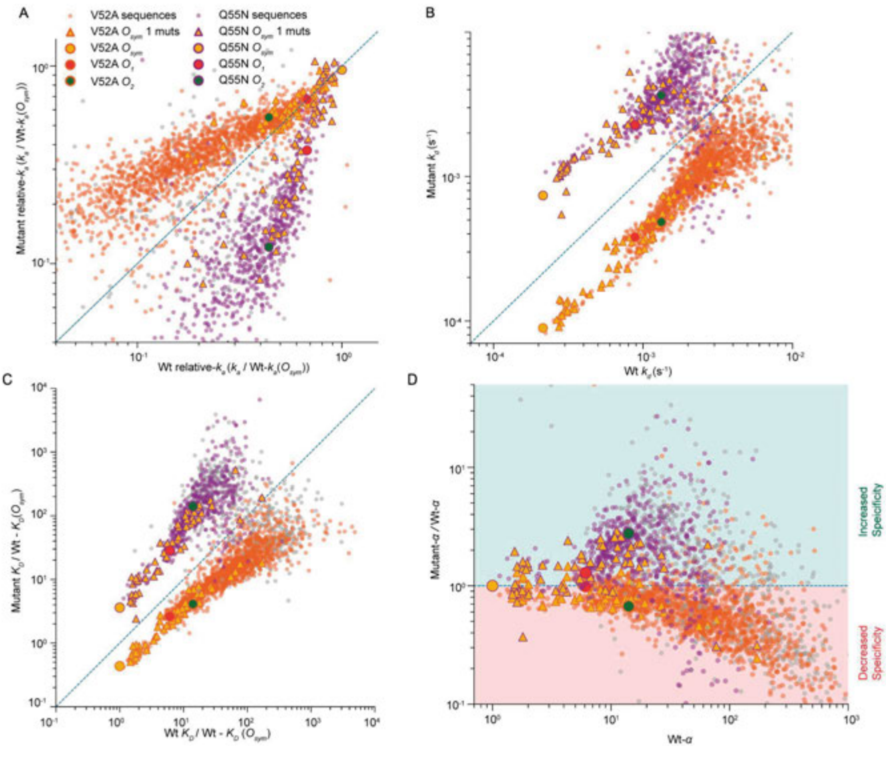

Figure 4.1. Measured affinity and estimated DNA specificity with operator library on the PBM (Adapted from Figure 2 in Paper II) A. Correlation of relative-k_a values (normalized with Wt-LacI $k_a(O_{sym})$ value) for unique DNA sequences between Wt-LacI (x-axis) and LacI mutants (y-axis), with V52A (orange) and Q55N (purple) LacI variants plotted in the same panel. Sequences with high variability in k_a measurements (CV > 0.3) across corresponding replicates are marked in grey. Each data point represents the mean relative-k_a value for a unique DNA sequence, calculated from all binding spots with the same sequence in 10 replicated experiments for Wt-LacI and 5 replicates for each mutant-LacI (See individual side-by-side Wt vs Mutant PBM experiment in Fig. S5A in **Paper II**). **B**. Correlation of dissociation rates, k_d, between Wt (x-axis) and its mutants (y-axis), Sequences with high variability in k_d measurements (CV > 0.3) are marked in grey (see individual side-by-side Wt vs Mutant PBM experiment in Fig. S5B). **C.** Correlation of equilibrium dissociation constants (K_D) between Wt (x-axis) and LacI mutants (y-axis). K_D values, calculated from k_d / k_a, of each protein to different sequences are normalized with the K_D value of Wt to the O_{sym} operator. **D.** Specificity changes were assessed using dissociation constant (K_D) estimates. For each DNA sequence (i), a LacI variant's K_D to it was defined as

the product of its K_D for the O_{sym} and a specificity factor, α(i), *i.e.* $K_D(\mathrm{i})=\alpha(\mathrm{i})\ K_D(O_{sym})$. If a mutation in LacI does not alter its DNA specificity to O_{sym}, the α values for Wt-LacI and Mutant-LacI would be identical for all sequences, *i.e* Wt-α(i) = Mutant-α(i) for all *i*. Normalized Mutant-α values (relative to Wt-α values) were plotted against Wt-α values. If a LacI mutant's normalized α values are generally bigger than 1 (green area), that means, compared to Wt, its specificity to O_{sym} is increased, while values below 1 (red area) indicate that its specificity to O_{sym} is decreased.

4.3. In-cell search kinetics by single-molecule imaging & Miller assay

The successful demonstration of the LacI hinge variants' effects *in vitro* then led to a question related to the stability-speed paradox[24]: Do these hinge-driven conformational biases alter how fast LacI finds and binds to its operator in a living cell?

To address this, we adapted the single-molecule search assay by Hammar et al., 2012[15] to measure and compare the k_{a_eff} (effective association rate) of LacI variants to a single chromosomal O_{sym} *in vivo*. Time for a target site O_{sym} to be searched by LacI molecules, $1/k_{a\_eff}$ is defined as:

$$\frac{1}{k_{a\_eff}} \propto \frac{T_{access}}{N \times p_{tot,Osym}} \quad (2),$$

Where T_{access} is the time for any LacI molecule to reach the target site but not necessarily bind, N is the number of searching LacI molecules, and $p_{tot,Osym}$ is the recognition probability to the O_{sym} site.

Details of the imaging set-up, the modifications introduced relative to the Hammar et al. (2012) protocol, and the considerations underlying these changes are provided in the **chapter: Methodology.** The specific imaging and experimental configurations are described in **Paper II**'s methods. Briefly, for each LacI variant, isogenic O_{sym} and ΔO_{sym} strains were loaded in parallel into a two-sided microfluidic device and pre-equilibrated in growth media with IPTG. At $t = 0$, the medium was switched to IPTG-free medium to initiate rebinding of LacI molecules, which were translationally fused to a the fluorescent protein mVenus on the C-terminal, replacing the tetramerization region. We the nrecorded time traces of diffraction-limited fluorescent spots ("foci"), representing stably bound LacI-Venus dimers; freely diffusing molecules blurred into the background. Examples of these images are shown in **Fig. 4.2.A** for Wt-LacI cells in steady states, with or without IPTG. The O_{sym}-specific trace was then obtained by subtracting background binding as determined by the time-trace of the ΔO_{sym} strain from the time-trace of the O_{sym} strain. Finally, the binding curve was fitted to an IPTG-integrated three-state model (see this model detailed in **Chapter: Methodology**). For anchoring kinetic fits, a normalized Miller β-galactosidase assay (O_{sym}–lacZ reporter) was performed for quantifying O_{sym} occupancy ($\frac{k_{a-eff}}{k_{a-eff} + k_d}$) in matched strain backgrounds (see the regression considerations and fitting in **Chapter: Methodology**)

Firstly, we observed that the *in vitro* properties were successfully validated *in vivo*. In the steady state without(w/o) IPTG in the ΔO_{sym} strain, which means only non-operator binding is possible, V52A showed the most binding and Q55N showed the least binding (**Fig. 4.2B**), matching their lowered and increased operator specificity phenotypes, respectively, as observed *in vitro*. Their binding stability to O_{sym} was also verified in the Miller assay, where V52A showed the strongest repression while Q55N showed the least repression on O_{sym}. (**Paper II Fig.3A**) In conclusion, we confirmed that V52A and Q55N exhibited the expected properties from biasing the hinge helicity also *in vivo*. We then looked at how the k_{a_eff} (effective association rate) was influenced.

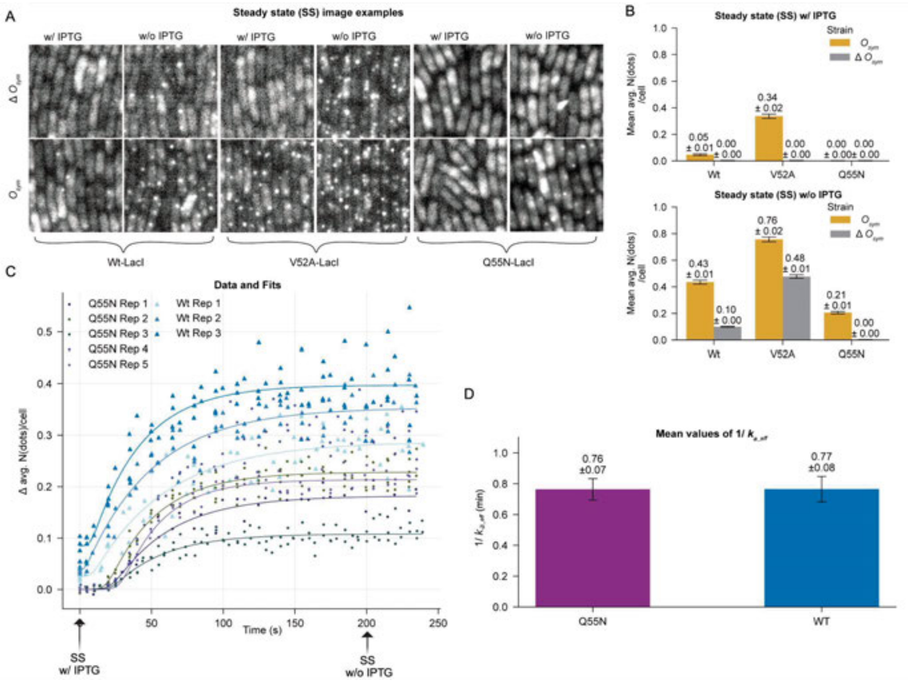

Figure 4.2. Single-molecule imaging result. (Adapted from Figure 3&4 in Paper II) A. Fluorescence images of cells expressing three LacI variants at steady state under two conditions: without IPTG (w/o IPTG) and with IPTG (w/ IPTG). Columns correspond to the IPTG condition, and rows correspond to two genetic backgrounds: one with the O_{sym} operator sequence integrated into the genome and one in which this operator is not present (ΔO_{sym}). For each LacI variant, the paired images of the O_{sym} and ΔO_{sym} strains are indicated by brackets. **B**. Detected average number of binding dots per cell for LacI variants in strains with or without O_{sym}, in the steady states with IPTG (*top*) or without (*bottom*). **C**. Binding is quantified by the difference in fluorescent spots per cell between O_{sym}-containing strains (O_{sym}) and control strains lacking O_{sym} (ΔO_{sym}). The binding curves were fitted using the IPTG-integrated three-state model (solid lines; see model and details about fitting in the Methodology chapter). A residual analysis (Fig. S8 in Paper II) shows no systematic deviations, indicating the fitting model sufficiently describes the data at the resolution of our experiments.

The fitted time-traces showed that the WT and Q55N searched the O_{sym} site at comparable rates (**Fig. 4.2D**), despite their markedly different specificities (**Fig. 4.1D**), see more discussion in the next session. For V52A, a reliable rebinding curve to O_{sym} could not be obtained because high O_{sym} occupancy persisted even in the presence of IPTG, indicating poor inducibility (**Fig. 4.2B**), thereby making the washout-based rebinding measurement impractical; this poor inducibility was also shown by measuring the repression with the Miller assay (**Paper II, Fig. 3A**).

4.4. Result summary and discussion

HT-k-PBM and *in vivo* readouts confirmed that LacI hinge mutants engineered to exhibit the opposite bias in hinge helicity produced predicted changes in LacI performance. For V52A, MD simulations predicted an increased hinge helicity and therefore a bias towards the recognition conformation, which increased the binding stability of V52A on the O_{sym} site *in vivo*, while reducing the DNA Specificity. Q55N, biased toward search, showed the opposite pattern (**Fig. 4.1; Fig. 4.2A–B**).

For target search, only Q55N could be evaluated by using the IPTG-washout single-molecule imaging assay. As stated in the above section, its k_{a_eff} to a single chromosomal O_{sym} was indistinguishable from WT (**Fig. 4.2D**). To interpret this, we utilized the relationship for the inverse of k_{a_eff} shown in eq.(2). Since our measurements showed WT-$k_{a\_eff} \approx$ Q55N-k_{a_eff}, and N(LacI) is similarly expressed for both variants (due to expression from the same promoter), eq.(2) implies that $\frac{T_{access}}{p_{tot,Osym}}$ must be similar for WT and Q55N.

The HT-k-PBM results (**Paper II, Fig. S1B**) showed that Q55N exhibits a decreased microscopic recognition probability ($p_{tot,Osym}\downarrow$) by ~30% compared to WT. For $\frac{T_{access}}{p_{tot,Osym}}$ to remain constant, this decrease in $p_{tot,Osym}$ has to be compensated by a proportional decrease in operator access time ($T_{access}\downarrow$) by 30% as well. Therefore, Q55N, assuming that the drop of $p_{tot,Osym}$ of Q55N compared to WT can be translated to *in vivo*, indeed needs to be a quicker accesser to the target site in the cell. (Explained more in the **Chapter: Conclusion & Future perspective**).

For V52A, its effective association rate could not be assayed by the IPTG-washout single-molecule assay because it retained a high occupancy on O_{sym} even with IPTG present (**Fig. 4.2A–B**), indicating loss of inducibility. Therefore, we could only infer its k_{a_eff} indirectly. The close correspondence between its ~2.4× slower dissociation *in vitro* (**PaperII Fig. 2G**), assuming that the WT:Mutant ratio is transferable between assays, and ~2.4 – 2.5× stronger repression *in vivo* (**Paper II Fig. 3A**) support that the repression strength difference *in vivo* between V52A and WT likely originated from their dissociation rate difference observed *in vitro*, which means V52A also has a similar k_{a_eff} as WT. However, V52A did not show an increased binding probability to O_{sym} compared with WT in vitro (**Paper II, Fig.S1B**). This means that V52A LacI-variant, while exhibiting high stability (due to k_d↓) and low inducibility, did not sacrifice its target search speed, suggesting the hinge selectively tuned $k_{off,\mu}$ (thus k_d) to achieve stability enhancement without incurring the search speed penalty predicted by the classic paradox.

Notably, the assumption that WT:mutant affinity ratios transfer between assays breaks for Q55N. Q55N dissociates from O_{sym} only ~3-fold faster than WT *in vitro*, and its *in vivo* association rate is indistinguishable from WT; yet its repression *in vivo* is ~20-fold weaker. Two possible factors were examined and found to be insufficient to explain this discrepancy. First, promoter/RNAP context: if RNAP competition at the reporter were the main driver, the Q55N:WT occupancy ratio would diverge between the Miller assay and the single-molecule readout (because the strains used for the Miller assay had a promoter for expressing LacZ while the strains used for the single-molecule assay did not). However, these ratios are in a similar range (≈0.78 vs ≈0.59), arguing that the presence of a promoter does not uniquely penalize Q55N at O_{sym}. Second, ionic strength: The EMSA showed that increasing NaCl from 50→200 mM weakened the binding for both proteins, but penalized the WT more than Q55N (WT KD rises ~10×; Q55N ~2×, as shown in **PaperII's Fig. S8 C&D**), which would shrink, not expand, the WT:Q55N gap at higher salt. Thus, neither factor explains the ≈20× loss of repression. Chromosomal supercoiling and local DNA dynamics may accelerate dissociation *in vivo* relative to short, relaxed oligos *in vitro*. If this preferentially inflates $k_{off,\mu}$, for lower-affinity complexes, Q55N (lower binding energy) would be disproportionately penalized, whereas V52A (higher binding energy) would remain close to *in vitro* expectations.

4.5. Summary

In summary, **Paper II** shows that biasing hinge-helix propensity reallocates LacI performance in terms of binding stability, DNA specificity, and inducibility without detectably altering the *in vivo* search time. So, **Paper II** results support a model in which the helicity of the LacI hinge region primarily tunes the stability–inducibility balance.

Given the limited mutant set, a definitive conclusion about the stability—inducibility trade-off and whether WT LacI sits at an evolutionary optimum—cannot yet be drawn. Reports of variants with both stronger repression and increased IPTG inducibility[50,51] suggest the WT may not be optimal; however, because IPTG is non-native, it remains essential to test whether this dual enhancement persists with allolactose *in vivo*. A systematic expansion of hinge-helix variants, coupled with comprehensive profiling of DNA specificity, binding stability, inducibility, and $k_{a,eff}$, could help clarify the evolutionary constraints shaping LacI—an aim pursued in **Paper III** (pilot study).

5. Paper III: On the way to building a large LacI mutant library

Paper III addresses the need for a scalable platform to bridge the gap between the detailed kinetic studies of a few LacI variants in **Papers I** and **II** and the goal of comprehensively mapping the large-scale sequence-function landscape. This manuscript describes the development and pilot validation of a high-throughput Optical Pooled Screening (OPS) platform for single-molecule readouts in live *E. coli* cells, with the ultimate goal of quantifying the precise kinetic and stability phenotypes of 10^3 to 10^4 LacI mutants.

The key technical development in the **Paper III** manuscript was based on the OPS workflow in Soares et al., 2025[52]. We integrated a high-throughput genotyping scheme with the design of a dual barcode on a single RNA transcript and combined it with a pooled strain construction workflow usingλ-Red recombineering. A pilot library of six known LacI genotypes and phenotypes was successfully constructed and screened to validate the integrated platform. This pilot confirmed the system's biological fidelity by correctly recovering the expected single-molecule phenotypes for all successfully identified strains, such as V52A displaying increased binding and Q55N showing the opposite trend, compared to WT.

The pilot study was instrumental in identifying technical constraints that must be overcome before scaling the Optical Pooled Screening (OPS) platform to its intended capacity of 10^3 - 10^4 strains. A primary limitation was the genotyping efficiency, as the dual barcode decoding success could not be reliably extrapolated from single-barcode systems, suggesting a potential loss of many unique genotypes due to detection failure. Furthermore, the pooled λ-Red recombineering introduced two sequence stability issues. First, the tandem Lac operator array (initially integrated to ensure strong single-focus binding) proved unstable; I propose mitigating this by replacing it with a single operator site (O_{sym}); Second, we observed a low rate of genotypebarcode linkage error—one case in 21 screened colonies—where the dual barcode and protein variant were swapped. This latter issue is addressable by implementing Next-Generation Sequencing (NGS) to verify library linkage fidelity prior to single-molecule phenotyping. These constraints and their identified solutions clarify the necessary design refinements required to fully unlock the scalable power of the OPS platform.

Ultimately, **Paper III** provides the necessary technical foundation and proof-of-principle for combining large-scale genomic engineering with single-molecule resolution phenotyping.

6. Methodology

In this section, I will summarise the key methods that I used for **Papers I and II**. **Paper III** is a pilot study focused on methodology development, and the techniques are best described in the manuscript itself.

6.1. High-throughput kinetics measurement on PBMs

The protein-binding microarray (PBM) protocol is an *in vitro* high-throughput assay for characterizing TF–DNA interactions. In early studies [38,53–56], PBMs was used for equilibrium specificity profiling of transcription factors (e.g., to determine relative equilibrium binding constants) by comparing end-point signals across the array. In a typical PBM experiment (PBM schematic as shown in **Fig. 6.1**), tens of thousands of DNA fragments are immobilized on the surface of a glass slide, and the purified, fluorescently labeled transcription factor of interest is applied to the array at one or more concentrations. After incubation and washing, the total binding of protein to each DNA spot is detected via the fluorescent tag of the TF protein, giving a fluorescence intensity that reflects the relative binding affinity of each DNA sequence. Single low-concentration TF on PBMs summarizes signals as rank-based scores (e.g., E-scores or normalized intensities) to report relative binding preferences across DNA sequences[56,57]. A series of TF concentrations can also be applied to obtain an apparent dissociation constant ($K_{d,app}$) per DNA sequence; these values rank affinities quantitatively but are "apparent" because surface immobilization and labeling stoichiometry of the PBMs can shift absolute scales[58].

In our work, we utilise PBMs for kinetic assays by time-lapse imaging of binding and dissociation. Specifically, we have ~2,479 *lac* operator variants (all single- and some double-base mutants of the natural O_1 and O_2 sequences plus a symmetric operator O_{sym}) printed on microarrays. (See details of microarray design in methods of **Paper I**).

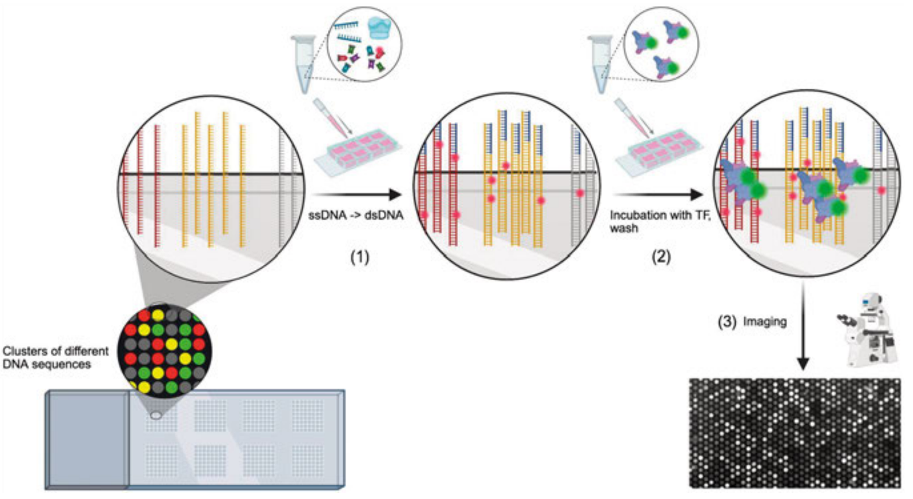

Figure 6.1: Overview of the classical Protein-Binding Microarray (PBM) Workflow. The PBM assay characterizes the binding affinity of a TF to thousands of unique DNA sequences (created in BioRender https://BioRender.com/ q45e576). (1) DNA Immobilization and Quantification: Single-stranded DNA (ssDNA) is printed on the microarray surface and converted to double-stranded DNA (dsDNA) using a fluorescently tagged dNTP. The resulting fluorescence at this stage is recorded to determine the relative dsDNA concentration at each printed DNA sequence spot for normalization. (2) Protein Binding and Imaging: The microarray is incubated with a fluorescently labeled transcription factor (e.g., LacI variant) until equilibrium is reached. (3) After washing away unbound protein, the bound complexes are imaged via fluorescence microscopy. The fluorescense image shows the signal from bound proteins on the array.

6.1.1. Adaptation of the PBM for kinetic measurement

To perform kinetic measurements, the PBM workflow was re-engineered (**Figure 6.2**). The major modification was that the Agilent DNA microarray containing the Lac operator variants was mounted in a flow cell and put on a fluorescence microscope. We continuously imaged both the association phase (flow of TF protein in buffer) and the dissociation phase (flow of buffer only) to extract macroscopic association and dissociation rates (k_a, k_d) for every cluster(spot) of DNA sequences under identical conditions. Because very weak binders lack reliable washout signals, analyses are restricted to sequences whose equilibrium fluorescence exceeds ~3% of that of O_{sym}.

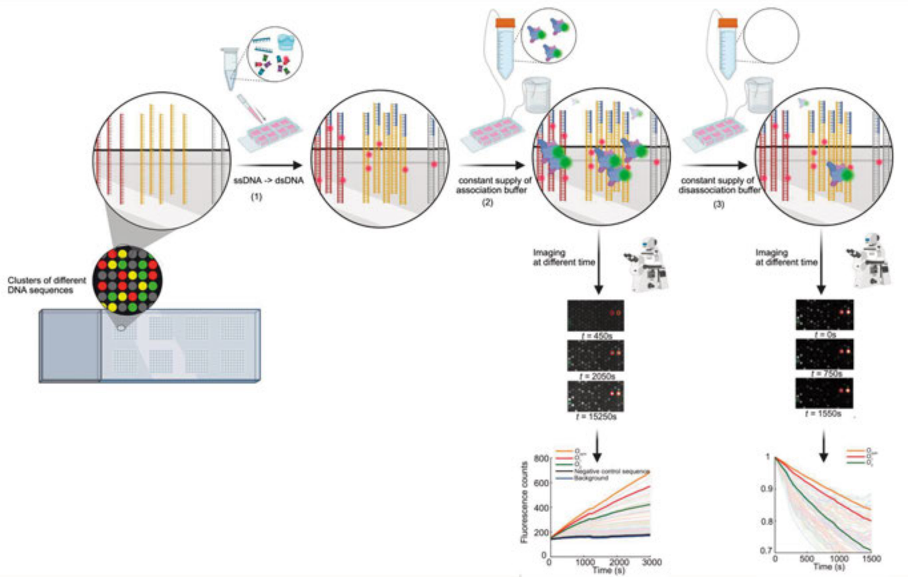

Figure 6.2: Workflow for high-throughput kinetic protein-binding microarray (HT-k-PBM) for real-time kinetic measurement. This adaptation extends the classical PBM (Figure 6.1) to continuously monitor TF-DNA interactions (created in BioRender https://BioRender.com/ q45e576). **(1)** ssDNA is immobilized and converted to dsDNA on the microarray surface, as in the classical PBM. **(2)** Real-time Association: A constant supply of fluorescently labeled TF in the binding buffer is flowed over the array. Fluorescence intensity is continuously imaged at different time points (e.g., t=450s, 2050s, 15250s), showing the accumulation of bound protein on DNA spots. The inset graph illustrates example association curves for different DNA sequences over time. **(3)** Real-time Dissociation: After association, the flow is switched to a buffer without TF. Fluorescence intensity is continuously imaged at different time points (e.g., t=0s, 750s, 1550s), showing the decrease in bound protein as TF dissociates from DNA spots. The inset graph illustrates example dissociation curves. From these real-time association and dissociation curves, microscopic kinetic rates (ka and kd) are extracted for thousands of unique DNA sequences, providing a comprehensive kinetic profile of TF-DNA binding.

In **Paper II**, parallel flow cells were installed on one microarray. This reduced experimental variation since WT-LacI and hinge-helix mutants (V52A-LacI, Q55N-LacI) could be analyzed in the same experiment. All computations, imaging cadence, and fits (including buffer-switch detection and exponential models) are detailed in the Methods of **Paper II**.

6.2. Search kinetics measurement with single-molecule experiments

6.2.1. Single-molecule imaging assay

Single-molecule fluorescence microscopy detects operator-bound fluorescently labelled LacI in living *E. coli* by detecting diffraction-limited spots; rapidly diffusing repressors blur into the background, whereas a LacI–mVenus molecule specifically bound at a chromosomal operator appears as bright, diffraction-limited spots (foci). Reliable detection requires keeping the number of labeled repressors at just a few dimers per cell and using long (>1s) exposures so that only the stationary, specifically bound molecules rise above the diffuse fluorescence of the unbound pool[14].

Building on this principle, time-resolved rebinding kinetics can be measured by combining imaging with a rapid media switch. In the Hammar et al. assay[15], cells are first incubated in IPTG-containing medium to prevent binding of LacI to the genome; IPTG is then removed by fast medium dilution with medium without IPTG (but with ONPF to prevent IPTG rebinding to LacI if IPTG has not diffused out of the cell). The fraction of cells exhibiting an operator-localized LacI spot is then recorded over time. Fitting this fractional binding curve with a single-exponential function yields the *in vivo* association rate to a single genomic operator. In **Paper II**, we followed these established practical criteria; we maintained a low copy number of the fluorescently labeled protein expressed in the cell, and used an appropriate acquisition timing (4s). In the following section, I describe how we applied and modified (see details in section 6.2.2.) the Hammar et al. rebinding protocol to compare search kinetics across LacI variants with modifications.

We did not use the autorepression schemes described in Elf et al., 2007 and Hammar et al., 2012 to maintain the low copy number of lacI expressed in cells, because in **Paper II**, we varied the LacI protein sequence; autoregulation of LacI would introduce variant-dependent expression differences and complicate kinetic comparisons. Instead, all LacI–Venus constructs were expressed chromosomally from the same low-expression promoter pFAB138[59], yielding comparably low copy number of LacI variants in all constructs while preserving single-molecule detectability (few dimers per cell; 4-s exposure reveals a single operator-bound dimer over the diffuse background), see example fluorescence images in **Fig. 6.3**.

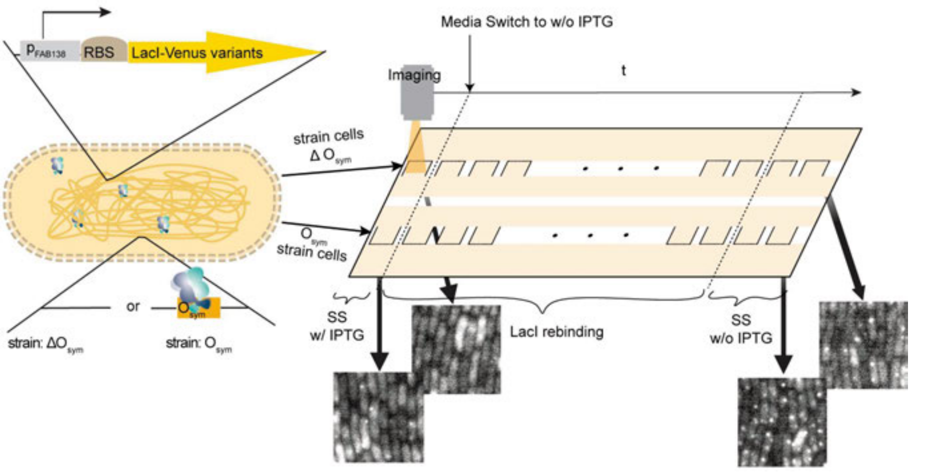

Figure 6.3. **A schematic overview of the single-molecule experimental workflow conducted on a microfluidic chip.** Two isogenic *E. coli* strains, both expressing a certain Venus-tagged LacI variant but differing in their genomic presence or absence of the O_{sym} operator sequence, were independently introduced into hydraulically interconnected cell chambers (labelled as squares open on one side), which are positioned on opposite sides of the chip. This setup allows strains to have simultaneous access to the shared growth media. Initially, chambers were exposed to media containing 0.3 mM IPTG. Fluorescence imaging was performed to capture the steady-state (SS) LacI binding in each strain under the IPTG-supplemented condition (SS w/ IPTG). Following this, the media was exchanged to IPTG-free media, and time-lapse fluorescence imaging subsequently captured downstream chambers on both sides of the chip to quantify LacI rebinding kinetics until a new steady-state equilibrium was established in media without IPTG (SS w/o IPTG). Additional fluorescence images show representative steady-state cells carrying Wt-LacI; arrows indicate from which time point and chamber each image was taken.

Also, to ensure that the time-resolved rebinding curve reflects LacI association to the O_{sym} operator rather than long-lived binding elsewhere in the genome, we utilised a microfluidic device that can run an isogenic control (ΔO_{sym}) in parallel with the O_{sym} strain (strain constructs are as shown in **Fig. 6.3**). Both strains carried the same chromosomal LacI–Venus construct and were exposed to identical flow paths, illumination, and acquisition settings. The ΔO_{sym} strain side therefore reported all signals not attributable to O_{sym} operator binding, i.e., stable binding to non-operator DNA. Cells were first equilibrated on-chip in IPTG to suppress binding. At $t = 0$, the medium was switched to IPTG-free medium to initiate rebinding. Imaging alternated between the two sides at a fixed interval (see details in Paper II's method); For every frame, cells were segmented and diffraction-limited foci were detected with a single, global threshold, computing the average number of detected foci per cell for a strain at a certain time point (avg.N(dots)/cell). Then the operator-specific signal was computed per time point (with interpolation) as:

$$\Delta \text{avg.N(dots)/cell} = \text{avg.N(dots)/cell}_{\text{Osym}} - \text{avg.N(dots)/cell}_{\Delta\text{Osym}}$$

(Data presented in **Fig. 6.4A**)

6.2.2. Why IPTG must be modelled explicitly for measuring search kinetics of LacI variants

As mentioned above, we adapted the rebinding assay of Hammar et al.[15], itself based on Elf et al.[14], in which cells are pre-equilibrated with IPTG and then subjected to a rapid medium exchange to initiate the rebinding process to the operator while recording the fraction of operator-localized LacI spots over time[14,15]. However, in both Elf's and Hammar's assays, an anti-inducer (ONPF) was added during IPTG washout so that IPTG could not rebind; this helped remove inducer rebinding due to slow removal of IPTG from the cell and helped yield a near single-exponential rise in fluorescence, dominated by target search.

In contrast, because we varied the LacI protein itself, we avoided introducing a second ligand (ONPF) with potential variant-dependent effects. Instead, we explicitly modeled inducer dynamics by extending the three-state search/recognition framework illustrated in **Paper I** to include IPTG binding to LacI in the free and nonspecific ("testing") states[60] (as shown in **Fig. 6.4**), and by describing the intracellular IPTG concentration as a first-order decay after the switch.

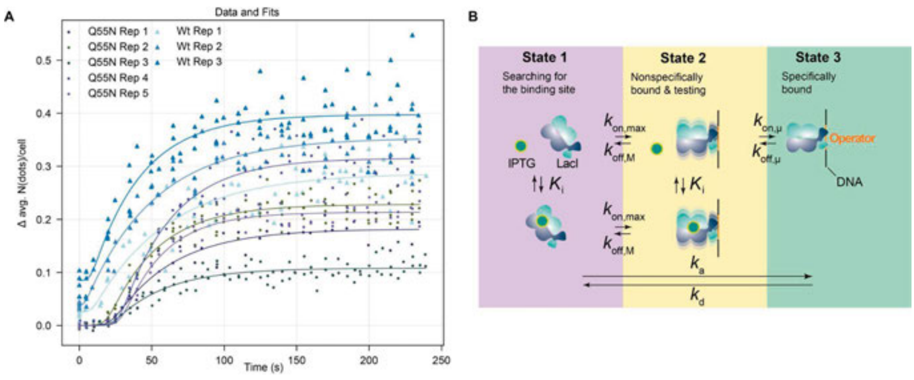

Figure 6.4 Three-state model integrated with IPTG. **A**. Binding is quantified by the difference in average detected foci per cell between O_{sym}-containing strains (O_{sym}) and control strains lacking O_{sym} (ΔO_{sym}). The binding curves were fitted using the IPTG-integrated three-state model (solid lines; see model in panel B and details about fitting method in the next session). The residual analysis (Paper II's Fig. S8) that was performed shows no systematic deviations, indicating the fitting model sufficiently describes the data at the resolution of our experiments. **B.** Microscopic rates of entering and exiting from one state to another are labelled at the cross sections of states: $k_{on,max}$ is the on rate into testing state from unbound state to operator site; $k_{off,M}$ is the dissociation rate from the testing (non-specifically bound to the operator site) state to the state of unbound with operator site; K_i: IPTG binding constant; $k_{on,\mu}$ is the on rate of testing state LacI transition into the specific-bound state on the operator; $k_{off,\mu}$ is the off rate of specific-bound state LacI go back into the testing state; K_i is the equilibrium dissociation constant for the interaction between IPTG and LacI that is either in state 1 or state 2.

6.2.3. ODE model for rebinding after IPTG washout

We modeled the operator-specific bound fraction $[OL_S](t)$ and the intracellular inducer = $[I](t)$ after the medium switch using an IPTG-integrated three-state scheme:

$$
\begin{array}{ccccc}
O + LI & \underset{k_{off,M}}{\overset{k_{on,max}}{\rightleftharpoons}} & OLI & & \\
\uparrow\downarrow K_i & & \uparrow\downarrow K_i & & \\
O + L + I & \underset{k_{off,M}}{\overset{k_{on,max}}{\rightleftharpoons}} & OL + I & \underset{k_{off,\mu}}{\overset{k_{on,\mu}}{\rightleftharpoons}} & OL_S
\end{array},
$$

The same scheme is shown in **Fig. 6.4B**. From the model, we got a set of ODEs:

$$\begin{cases} \dfrac{d[OL_S]}{dt} = \tilde{k}_{\text{on,max}}\, p_{\text{tot}}(I)\, \big(O_{\text{tot}} - [OL_S]\big) \; - \; k_{\text{off},\mu}\, \big(1 - p_{\text{tot}}(I)\big)\, [OL_S], \\ \dfrac{d[I]}{dt} = -\, \gamma\, [I]. \end{cases}$$

Here, O_{tot} is the intracellular concentration of the O_{sym} site (≈1 per cell, with modest replication-stage variation). [I]*γ represents the efflux rate of free intracellular IPTG after the media switch, $\tilde{k}_{\text{on,max}} = k_{\text{on,max}}\,[L]$ represents the arrival flux into the testing state at the operator locus, the free LacI concentration, [L], is assumed to be approximately constant throughout the experiment, and $p_{tot}([I])$ represents the probability that a LacI molecule, while in the testing state, converts to the specific state; IPTG blocks this conversion.

We assumed rapid IPTG binding/unbinding to LacI (fast equilibrium assumption, K_i, the IPTG affinity constant) and allowed IPTG to bind LacI only in the free and testing state. We knew the recognition probability ptot([I]=0)=$k_{on,\mu}/(k_{on,\mu}+k_{off,M})$ when [I]=0 (from **Paper I**). So, by normalizing K_i with i_0, $\tilde{K}_i = K_i/i_0$, where i_0 is the intracellular IPTG concentration at the time point when washout starts (at $t\to 0$), we could get the IPTG-dependent recognition probability as:

$$p_{\text{tot}}([I]) = \frac{k_{\text{on},\mu} \cdot \dfrac{K_i}{K_i + [I]}}{k_{\text{off},M} + k_{\text{on},\mu} \cdot \dfrac{K_i}{K_i + [I]}} = \frac{1}{1 + (\dfrac{1}{p_{tot}([I]=0)} - 1) \cdot (1 + \dfrac{[I]/i_0}{\tilde{K}_i})} \quad \text{(E.q 6.1)}$$

6.2.3.1. What the rebinding curve can (and cannot) determine

The binding curve, [OL_s](t), as shown in **Fig. 6.4A**, intuitively contains four "features": the initial slope, the plateau, the early-time sigmoidal curvature ('the delay' in rebinding curves), and the late-time (close to plateau) region. Our model, however, has more than four kinetic parameters to be determined; therefore, without external information, distinct parameter sets can result in the same curve (non-identifiability). Concretely:

Early-slope (scale): When $t\to 0$, the normalized inducer $[I]/i_0$ is 1 and the bound fraction is still small (valid for WT-LacI and Q55N-LacI)). The initial slope of the rebinding trace is therefore,

$$\left.\frac{d[OLS]}{dt}\right|_{t->0} \approx \tilde{k}_{on,max} O_{\text{tot}} p_{\text{tot}}(i_0)$$

with

$$p_{\text{tot}}(i_0) = \frac{1}{1 + \left(\frac{k_{\text{off},M}}{k_{\text{on},\mu}}\right)\left(1 + \frac{1}{\tilde{K}_i}\right)}$$

only the product: $\tilde{k}_{on,max} O_{\text{tot}} p_{\text{tot}}(i_0)$ is seen through fitting to the initial slope of the curve.

Plateau: When $t \to \infty$, IPTG has cleared ($[I] \to 0$) and the system is at steady state ($d[OL_S]/dt=0$), so

$0 = \tilde{k}_{on,max} p_{tot}(0)(O_{tot} - [OL_S]^*) - k_{off,u}(1 - p_{tot}(0))[OL_S]^*$ When IPTG is cleared, $k_a = \tilde{k}_{on,max} p_{tot}(0)$, $k_d = k_{off,\mu}(1 - p_{tot}(0))$, as in **Paper I**. So the plateau time derivative can be re-written as $0 = k_a(O_{tot} - [OL_S]^*) - k_d[OL_S]^*$, from which we can get $[OL_S]^* = O_{tot} * k_a/(k_a + k_d)$. So again, only the $O_{tot} * k_a/(k_a + k_d)$ or product is seen through the plateau constraint ($[OL_S]^*$).

Early time (Sigmoidal rise): The early sigmoidal rise shape of the binding curve arises from the factor $\frac{[I]/i_0}{\tilde{K}_i} = \frac{e^{-\gamma t}}{\tilde{K}_i}$, as shown in **E.q. 6.1**. Making washout faster (bigger γ) but also making IPTG binding weaker (bigger $\tilde{K}_i$) can yield a nearly identical sigmoid shape. Thus, γ and $\tilde{K}_i$ cannot be determined by the shape alone.

Late-time (near-plateau) region: After enough time has passed post-washout, we can assume intracellular IPTG is negligible $[(I] \to 0)$, so $p_{tot}([I]) \approx p_{tot}(0)$ and is constant. Then the ODE become:

$$\frac{d[OL_S]}{dt} \approx k_a(O_{tot} - [OL_S]) - k_d[OL_S] = k_a O_{tot} - (k_a + k_d)[OL_S]$$

By letting the residual $E(t) = [OL_S]^* - [OL_S](t)$ (so $E \geq 0$ and decays to 0), differentiating both sides, and substituting $[OL_S]$ with $[OL_S]^*$-E, we get:

$\frac{dE}{dt} = (k_a + k_d)([OL_S]^* - E) - k_a O_{tot} = (k_a + k_d)[OL_S]^* - k_a O_{tot} - (k_a + k_d)E$, and from the plateau information above, we know that $[OL_S]^* = \frac{O_{tot} * k_a}{k_a + k_d}$, we get $\frac{d[E]}{dt} = (k_a + k_d)(\frac{O_{tot} * k_a}{k_a + k_d} - E) - k_a O_{tot} = -(k_a + k_d)E$, which makes E(t) a single exponential decay, and $E(t) = E(t_0)e^{-\lambda(t-t_0)}$ with a single time constant $\lambda = k_a + k_d$. Thus, we can get the additive value of k_a and k_d from the late-time (near-plateau) region of a binding curve, but, again, not k_a and k_d separately.

Taken together, the initial slope, the sigmoidal rise, and the near-plateau tail of the rebinding curve provide, at most, (*i*) a product constraint on $\tilde{k}_{\mathrm{on,max}}$, O_{tot}, $p_{tot}([I]/i_0=1)$ i.e., $\frac{d[OLS]}{dt}\Big|_{t->0} \approx \tilde{k}_{on,max} O_{\mathrm{tot}} p_{\mathrm{tot}}(i_0)$ (*ii*)

If we define $P_{bound} = [OL_S]^* / O_{tot}$, meaning the chance of O_{sym} getting specifically bound at steady-state w/o IPTG, $P_{bound} = k_a/(k_a + k_d)$, we get a ratio constraint $\frac{k_a}{k_d} = \frac{P_{bound}}{1 - P_{bound}}$ via the plateau, and (*iii*) a constant $\lambda = k_a + k_d$ from the late-time residual *E* decay. Because these features (*i* - *iii*) do not help to uniquely separate k_a, the parameter we are most interested in. An external constraint is thus required. We therefore turned to a matched Miller assay in the same strain background to obtain that extra constraint.

6.2.4. Miller assay: method, principle, and how it constrains k_a

6.2.4.1. Principle (what the readout means)

The Miller assay[61] measures β-galactosidase (β-gal/LacZ) activity using ONPG as a substrate: LacZ hydrolyses ONPG to o-nitrophenol, which is quantified by its absorbance at 420 nm. Reaction rates are normalized for cell density and sample volume to yield Miller Units (see method details in **Paper II** methods). Thus, Miller Units is proportional to intracellular [β-gal], which in turn is proportional to the probability that the operator O_{sym} is free, P_{free}.

6.2.4.2. From occupancy to a constraint for anchoring the binding curve fit

In the LacI variants, $[\beta\text{-gal}] = \gamma \cdot P_{free}/\mu$ (where γ = maximal production rate, μ = degradation rate of β-gal), while in the ΔLacI strain (where the operator is always free, P_{free} = 1), $[\beta\text{-gal}] = \gamma/\mu$. Normalizing Miller Units (which is proportional to [β-gal] in cells) gives: $\frac{Miller\ Units\ (LacI)}{Miller\ Units\ (\Delta LacI)} = \frac{\gamma \cdot P_{free}/\mu}{\gamma/\mu} = P_{free}$, so P_{free} = 1- P_{bound} = normalized Miller Units. The inverse of normalized Miller Units (ϕ), ϕ = 1/(1- P_{bound}), O_{sym} occupancy can thus be quantified as $P_{bound} = 1 - \frac{1}{\phi}$, and in steady-state w/o IPTG, $P_{bound} = k_a/(k_a + k_d)$, therefore $k_a/k_d = \phi$ -1.

As discussed in the 6.2.3 section, for each binding curve, its E in the near-plateau late-time region decays single-exponentially with $\lambda = k_a + k_d$, which means that we can get an additive value for k_a and k_d. We implemented the curve fit with the k_a and k_d ratio fixed by $k_a/k_d = \phi$ -1, to get unique rates of k_a, k_d, and O_{tot} (plateau) from each curve fit.

6.2.5. Method summary and Scope of applicability

(*i*) We adopted the single-molecule localization and rapid media-switch assay pioneered by Elf *et al*. (2007)[14] and Hammar *et al*. (2012)[15]: LacI–Venus was kept at low copy number so that operator-bound dimers appeared as diffraction-limited spots above the background fluorescence from free diffusing LacI dimers, and a fast inducer washout initiated time-resolved rebinding (**Fig. 6.3**). A two-sided microfluidic device allowed parallel imaging of isogenic strains with and without the target operator (O_{sym} vs ΔO_{sym}), enabling operator-specific binding curves by subtraction.

(*ii*) We then developed a strategy to compare target-search kinetics across variants of an inducible DNA-binding TF by:

> (*a*) extracting the operator-specific time course $[OL_S](t)$ from interlaced imaging with O_{sym} & ΔO_{sym} strains ($[OL_S](t)$ = avg.N(dots)/cell$_{\mathrm{Osym}}$(t) − avg.N(dots)/cell$_{\Delta\mathrm{OsymI}}$(t));
>
> (b) explicitly modeling inducer efflux dynamics as a first-order intracellular decay $d[I]/dt = -\gamma[I]$ after the switch, and integrating inducer effect into the three-state recognition model from **Paper I** (ligand binds the free and testing states but not the specific state); and
>
> (c) performing regression on binding curve with orthogonal constraint so each trace yields unique kinetic parameters: the late-time curvature provides $\lambda = k_a + k_d$, and a matched bulk reporter (Miller) assay in the

same strain background fixes the rate ratio $k_a/k_a=\phi-1$. From these we compute $k_a=\lambda(\phi-1)$ (therefore $k_d=k_a/(\phi-1)$), while the plateau sets O_{tot} *v*ia $[OL_S]^*= k_a/(k_a + k_d)*O_{tot}$.

The framework above generalizes to inducible TFs whose target search can be coarse-grained by a three-state scheme: diffusion (unbound), testing (nonspecific DNA contact near the site), and recognition (specific complex), and for which the ligand primarily gates the testing→recognition transition (i.e., ligand binds free and/or testing states but not the specific state). Under these conditions, the same pipeline: operator-present/absent imaging, ligand washout, first-order intracellular ligand decay, and a matched occupancy assay to anchor k_a/k_d, yields a unique k_a value for comparison across TF mutated proteins that are still inducible.

7. Conclusion & Future perspective

Paper I shows that LacI's DNA specificity is largely operator recognition-gated: across thousands of operator variants measured by HT-k-PBM, the association and dissociation rates anti-correlate, indicating that variation in specificity is dominated by changes in the probability of committing to the recognition state when testing an operator (p_{tot}), with comparatively smaller variation in the microscopic off-rate from the specific state ($k_{off,\mu}$). In **Paper II**, we build on this mechanistic insight, and describe how perturbations of the hinge-helix, which mediates the search-to-recognition switch, alter the LacI search dynamics. We find that increasing hinge helicity (V52A) strengthens binding stability, but at the cost of reduced DNA specificity and inducibility. Conversely, decreasing helicity (Q55N) weakens binding stability yet sharpens DNA specificity. Importantly, within our experimental resolution, the cellular target association rate (k_{a_eff}) remains robust; the "search-biased" mutant and the "recognition-biased" mutant do not show drastically impaired search efficiency. This suggests that global target search is buffered, likely through compensatory effects such as faster operator access time but lower binding probability (p_{tot}) to the target for Q55N, and slower operator access time but higher p_{tot} to the target for V52A compared to WT. In conclusion, stability and inducibility emerge as the primary mechanistic limitations imposed by hinge helicity in LacI.

Reconciling the minimal sliding model with EMSA

The EMSA showed that Q55N's K_d varies less with ionic strength than the WT's. This was qualitatively consistent with a minimal sliding model in which higher salt primarily shortens the one-dimensional sliding length (detailed in supplementary text in paper II). Within this minimal sliding model, the recognition probability accumulates with a sliding distance N, $p_{tot}(\mathrm{N, p}) = 1-(1-p)^{\mathrm{N}}$, where p is the per-encounter commitment probability when LacI is positioned on the operator. We assumed that salt mainly changed N, not p, and that the salt→N mapping was similar for both variants. Now still with this minimal sliding model, let R= $\frac{p_{tot}(N_{low-salt},p)}{p_{tot}(N_{high-salt},p)} = \frac{1-(1-p)^{N_{low-salt}}}{1-(1-p)^{N_{high-salt}}}$, define *1-p* as α, R= $\frac{1-\alpha^{N_{low-salt}}}{1-\alpha^{N_{high-salt}}}$, for $N_{low-salt} > N_{high-salt} \geq 1$, R = $\prod_{k=1+N_{high-salt}}^{N_{low-salt}} \frac{1-\alpha^{k}}{1-\alpha^{k-1}}$, where k is an integer index. Each factor $\frac{1-\alpha^{k}}{1-\alpha^{k-1}}$ is strictly > 1 when $\alpha \in (0,1)$, thus R is strictly increasing with α. Because Q55N has a smaller p than WT (so Q55N-α > Wt-α, suggested by HT-k-PBM data), it follows that Q55N-R>WT-R, i.e., as salt increases (reducing N), p_{tot} for Q55N necessarily drops more steeply than for WT. This result then supports that Q55N is theoretically a quicker target accesser than WT-LacI (as discussed in **Chapter 4.4**).

What about V52A? The IPTG washout single-molecule assay was not practical to measure the k_{a_eff} of V52A with O_{sym} due the loss of inducibility when V52A is bound on O_{sym} . However V52A is still inducible when it is bound on other non-Osym genomic sites, which means if we replace O_{sym} as target site for V52A with the lower-affinity target site O_1, or O_2, we could be able to actually compare V52A-k_{a_eff} and Wt-k_{a_eff} to the O_1 or O_2 target site.

8. Popular science summary

Genes are like lines of code that program life, but each single cell has to decide when to run each line. Proteins called transcription factors (TFs) act like gatekeepers: they bind DNA, recognize a short targeted DNA "address" and then either block or enable the transcription of the nearby gene. Their job is a tricky balancing act: find the right DNA 'address' fast, and hold on firmly enough to control the gene, yet still be able to let go when the cell needs to respond to a new signal.

Every transcription factor faces a daunting challenge: locating its target DNA sequence with single-letter precision somewhere in an enormous genome, often containing 10^6–10^9 letters of DNA. This is often referred to as the target search problem. It is hard for us to imagine how a biomolecule searches for its target in the intracellular micro-world. Unlike us, molecules do not have eyes, maps, and neither GPS. One way to picture the problem is to imagine yourself blindfolded in a giant space station filled with 3.5 million people. Your task is to find one specific person. One strategy is to float around randomly, bump into people, feel their faces, and decide whether to hold on or move on. If a protein searched for its DNA site in this purely random three-dimensional (3D) way, it could take hours or longer to find the right target.

Yet in living cells, gene regulation happens on the timescale of seconds to minutes. This tells us that proteins must use a smarter strategy. Experiments over the past decades have shown that many DNA-binding proteins do not just float in 3D; once they hit the DNA, they can slide along it, checking many nearby DNA sites in one go. Returning to the space-station picture, imagine that people are divided into groups, and everyone in the same group is holding onto the same extra-long scarf that threads its way through the crowd. You no longer need to keep bumping into people, letting go, and then bumping into the next person to feel every face in the group: as soon as you bump into any one person, you can grab the scarf and slide along it, using a single bump-and-release sequence to "touch" many faces in one go. In the same way, sliding makes the protein's "target" effectively behave like a broad region in the genome, rather than a single point, greatly increasing the chance of finding it in time.

This thesis follows one classic TF model, the lac repressor (LacI) from *E. coli*, to ask three connected questions regarding a TF's search process.

1) **What actually makes recognition selective?**

Conventional wisdom says selectivity comes mainly from how long a TF stays bound at the right site while quickly slipping away on the wrong ones. But that ignores the first step: deciding to bind when the correct sequence is encountered. In Paper I, the search behaviour of LacI was modelled by dividing the process into three states: diffusing in 3D, testing for finding the binding site, and recognition on the binding site. I measured how quickly LacI associates with and dissociates from thousands of DNA sequences, including the natural sequences that LacI is looking for and also sequences that are slightly different, on a microarray, in real time. The surprising pattern was that sequences which the TF bound quickly also tended to let it go more slowly, and vice versa. That signature, together with the derived model, means the key difference between sequences is how frequently LacI commits to recognition upon encounter (the recognition step), not primarily how slowly it leaves once bound.

2) **What structure in the protein controls that decision? And if we change that, what consequences will it bring?**

Structures of LacI show a small hinge region that is flexible when scanning along DNA but folds into a helix when the TF is locking onto the correct sequence. In Paper II, I introduced tiny changes (single amino-acid substitutions) designed to bias this hinge toward or away from forming a helix. As predicted, making the hinge more helical stabilized binding and reduced DNA specificity (falls into recognition on non-targeted sites as well)—but at a cost: the protein became hard to switch off (less inducible). Tweaking the hinge the other way made LacI less good at holding onto the target site. In other words, the hinge acts as a conformational gate that balances stability with responsiveness. Notably, these changes did not measurably slow the protein's ability to find and bind to its right site in the cell for the variants tested.

3) Can we study thousands of variants directly in living cells?

Single-molecule microscopy in living cells is powerful but traditionally slow—one protein at a time. Paper III presents a pilot solution: a pooled approach that writes two tiny barcodes and a protein sequence variant into each strain's chromosome. After filming live cells to capture single-molecule behavior, those barcodes are read out directly on the microscope by genotyping, linking each phenotype behavior back to its genetic variant. In a small pilot of six genotypes, the method correctly identified five and reproduced the expected phenotype behaviors. We also defined what must be improved to scale to thousands of variants.

Take-home message:

LacI, and likely many TFs, achieves specificity mainly by controlling the chance of committing to recognition, rather than by "sticking longer." A small structural switch (the hinge) tunes the trade-off between firm control and quick responsiveness, but with limited impact on target search and binding efficiency. And with pooled, single-cell imaging, we are moving toward testing these ideas across entire landscapes of protein variants in their natural cellular context.

9. Populärvetenskaplig sammanfattning

Gener är kodsträngar som programmerar livet, men varje enskild cell måste bestämma när varje program ska köras. Proteiner som kallas transkriptionsfaktorer (TFer) fungerar som grindvakter. De känner igen korta, specifika DNA-”adresser” där de binder och blockerar eller möjliggör uttrycket av närliggande gener. Deras uppgift är en svår balansgång. De måste hitta rätt DNA-”adress” snabbt, hålla fast tillräckligt hårt för att styra genen, men ändå kunna släppa taget om cellen behöver svara på en ny signal.

Varje transkriptionsfaktor står inför den enorma utmaningen att hitta sin DNA-adress med en bokstavs noggrannhet någonstans i ett genetiskt material som ofta innehåller i storleksordningen 10^6–10^9 DNA-bokstäver. Detta brukar kallas target search-problemet (målsökningsproblemet). Det är svårt för oss att föreställa oss hur en biomolekyl letar efter sitt mål i cellens mikrovärld. Till skillnad från oss har molekyler varken ögon, kartor eller GPS. Ett sätt att visualisera problemet är att tänka sig att du är försedd med ögonbindel i en gigantisk rymdstation fylld med 3,5 miljoner människor. Din uppgift är att hitta en specifik person. En strategi är att flyta runt slumpmässigt i tre dimensioner, stöta in i människor, känna på deras ansikten och avgöra om du ska hålla fast eller gå vidare. Om ett protein letade efter sin DNA-adress på samma sätt skulle det kunna ta timmar eller längre att hitta rätt mål.

Ändå sker genreglering i levande celler på tidskalan sekunder eller minuter, vilket säger oss att proteinerna måste använda en smartare strategi. Experiment som utförts de senaste decennierna har visat att många DNA-bindande proteiner inte bara flyter runt i 3D utan även glider längs DNAt. På så vis kan de kontrollera många närliggande DNA-adresser i ett svep. Om vi återvänder till tankeexperimentet med rymdstationen kan vi föreställa oss att människorna delas in i grupper, och att alla i samma grupp håller i samma extra långa halsduk som slingrar sig genom folkmassan. Du behöver inte längre gång på gång stöta in i människor, släppa, och sedan stöta in i nästa för att utforska allas ansikten i gruppen. Så fort du stöter in i någon kan du greppa halsduken, glida längs den, och i en enda ”stöta-till-och-släppa”-rörelse utvärdera många ansikten på en gång. Det är också lättare att slumpmässigt stöta in i en stor grupp av människor jämfört med en enda, vilket dramatiskt ökar chansen att hitta målet i tid.

I den här avhandlingen följer jag transkriptionsfaktorn LacI i bakterien *E. coli,* en klassisk modell för att studera genreglering. Syftet är att besvara tre relaterade frågor om hur DNA-bindande proteiner hittar sitt mål.

1) **Vad är det egentligen som gör igenkänningen selektiv?**

Det är en vedertagen uppfattning att selektiviteten främst beror på att en TF sitter kvar länge på rätt plats och snabbt släpper de felaktiga. Men denna hypotes bortser från det första steget i bindningsprocessen, att över huvud taget bestämma sig för att binda när den korrekta sekvensen påträffas. I Artikel I modellerade vi LacI:s sökbeteende genom att dela upp processen i tre tillstånd; diffusion i 3D, utvärdering av DNA för att hitta bindningsstället och igenkänning av bindningsstället. Jag mätte hur snabbt LacI binder till och släpper från tusentals DNA-sekvenser på en microarray i realtid, både de naturliga sekvenser som LacI letar efter och sekvenser som är något förändrade. Det överraskande mönstret var att sekvenser som TF:n band snabbt också tenderade att släppa den långsammare – och tvärtom. Denna signatur, tillsammans med den härledda modellen, innebär att den avgörande skillnaden är hur ofta LacI väljer att gå in i igenkännandetillståndet när den stöter på en viss sekvens (igenkänningssteget), inte i första hand hur lång tid det tar innan proteinet lämnar platsen när det väl sitter fast.

2) **Vilken struktur i proteinet styr det beslutet? Och om vi ändrar den, vilka konsekvenser får det?**

Strukturer av LacI visar en liten gångjärnsregion som är flexibel när proteinet skannar DNAt, men som viker ihop sig till en helix när TF:n fastnar på den korrekta sekvensen. I Artikel II införde jag mycket små förändringar (enstaka aminosyrasubstitutioner) som var utformade för att göra det antingen lättare eller svårare för gångjärnsregion att bilda en helix. Som förväntat stabiliserade en mer helikal gångjärnsregion bindningen och minskade DNA-specificiteten (proteinet går oftare in i igenkänning även på icke-målsekvenser) – men till ett pris: proteinet blev svårt att stänga av (mindre inducerbart). Att justera gångjärnet åt andra hållet gjorde LacI sämre på att hålla fast vid målsekvensen. Med andra ord fungerar gångjärnet som en konformationsport som balanserar stabilitet mot snabb respons. Det är viktigt att påpeka att dessa förändringar inte mätbart förändradeproteinets förmåga att snabbt hitta och binda sin rätta plats i cellen.

3) Kan vi studera tusentals varianter direkt i levande celler?

Enmolekylmikroskopi i levande celler är kraftfullt men traditionellt långsamt – ett protein i taget. Artikel III presenterar en pilotlösning: ett kombinerat angreppssätt där två små streckkoder och en proteinsekvensvariant skrivs in i DNAt hos tusentals bakterier samtidigt. Efter att levande celler har filmats för att fånga beteendet hos de olika proteinvarianterna, läser vi av de genetiska streckkoderna under mikroskopet. På så vis kan vi koppla varje beteende till rätt genetisk variant. I en liten pilot med sex varianter identifierade metoden fem korrekt och återskapade de förväntade beteendena. Vi definierade också vad som måste förbättras för att kunna skala upp till tusentals varianter.

Sammanfattande budskap:

LacI – och sannolikt många andra TF:er – uppnår specificitet främst genom att styra sannolikheten att gå in i igenkänningstillståndet, snarare än genom att ”sitta kvar längre” på rätt ställa. En liten strukturell omkopplare (gångjärnet) finjusterar avvägningen mellan snabb kontroll och responsivitet, men påverkar inte målsökning och bindningseffektivitet nämnvärt. Med kombinerad avbildning av enskilda molekyler i levande celler är vi dessutom på väg mot att testa dessa idéer för tusentals proteinvarianter i deras naturliga cellulära miljö.

10. 科普摘要

基因就像编写生命程序的代码行，但每一个细胞都必须自己决定什么时候“运行”哪一行代码。叫作转录因子（transcription factors, TF）的蛋白质就像门卫：它们结合到DNA上，识别一段很短的、特定的DNA“地址”，然后要么阻断、要么允许附近基因的转录。它们的工作是一场微妙的平衡：一方面要尽快找到正确的 DNA“地址”，另一方面又要抓得足够牢才能控制基因，同时在细胞需要响应新的信号时，还得能及时松手。

每一个转录因子都面临着同样艰巨的挑战：在庞大的基因组中，用单个碱基的精度找到自己的目标DNA序列；而基因组往往包含 10^6–10^9 个DNA“字母”。这挑战通常被称为“靶点搜索问题”（target search problem）。很难想象，在细胞内部这样一个微观世界里，一个生物大分子是如何寻找自己的靶点的。和我们不同，分子没有眼睛、没有地图，更没有GPS。一个形象的比喻是：想象你被蒙上眼睛，置身于一个挤满 350 万人的巨大空间站，你的任务是找到其中一个特定的人。一种策略是随意漂浮，随机撞到别人身上，用手摸他们的脸，然后决定是要抓住这个人还是离开继续去找下一个。如果蛋白质完全靠这种在三维空间中随机“乱撞”的方式去寻找自己的 DNA 结合位点，那么它找到正确目标可能需要数小时甚至更久。

然而，在真实的活细胞中，基因调控往往发生在几秒到几分钟的时间尺度上。这说明转录因子蛋白质一定采用了更聪明的策略。过去几十年的实验表明，许多DNA搜索蛋白并不是只在三维空间里乱漂；一旦它们碰到 DNA，就可以沿着 DNA“滑动”，在一次接触中检查附近的许多DNA位置。回到空间站的比喻，就好像人们被分组，分到同一组的人们手里共同拉着一条穿过人群的超长围巾。你不再需要重复撞到-分离-撞到来感知这个组的每个人的脸——只要撞到其中一个人，抓住这

条围巾，沿着它滑过去，就能在一次撞到分离的动作中“摸到”很多人的脸。类似地，这种滑动使得蛋白质的“靶点”在效果上好像变成了基因组里的一大片区域，而不是一个单一的小点，从而大大提高了在有限时间内找到目标的机会。

本论文以一个经典的转录因子模型——来自大肠杆菌的乳糖阻遏蛋白（lac repressor, LacI）——为主线，围绕转录因子的搜索过程提出三个相互关联的问题。

1）究竟是什么让DNA识别具有特异性？

传统观点认为，选择性主要来自这样一个事实：转录因子在正确的位点上停留得更久，而在错误位点上很快就脱离。但这种说法忽略了第一步：当遇到正确序列时，要不要“决定”真正结合上去。在文章 I 中，我们通过把 LacI 的搜索过程分成三个状态来建立模型：在三维空间中的扩散、在结合位点附近进行“测试”、以及在结合位点上进入识别状态。我利用芯片（microarray）实时测量了 LacI 与成千上万段 DNA 序列的结合和解离速率，这些序列包括 LacI 天然要寻找的序列，也包括与之略有差别的序列。令人意外的模式是：那些 LacI 结合得越快的序列，往往也把它“放走”得越慢，反之亦然。结合这一特征和模型分析，可以看出，不同序列之间的关键差异在于：当 LacI 遇到它们时，有多大概率会真正“下决心”进入识别状态（识别步骤），而不是主要体现在一旦已经结合上之后离开的速度有多慢。

2）蛋白质中哪一部分结构在控制这个决策？如果我们去改变它，会带来什么后果？

LacI 的结构显示，它有一个很小的“铰链”区域：在沿着 DNA 扫描时，这个区域是柔软灵活的；而当转录因子锁定正确序列时，它会折叠成一段 α 螺旋。在文章II中，我通过引入极小的改变（单个氨基酸替换），来人为地偏向这个铰链更容易或更不容易形成螺旋。正如预测的那样，让铰链更“爱成螺旋”会增强结合的稳定性，同时降低 DNA 特异性（也更容易在非目标位点上进入识别态）——但代价是：这种蛋白更难被“关掉”（诱导性变差）。反过来，把铰链调到不那么容易成螺旋，则会让 LacI 更不善于牢牢抓住目标位点。换句话说，这个铰链就

像一个构象“闸门”，用来在稳定性和响应速度之间做权衡。值得注意的是，在我们测试的变体中，这些改变并没有明显减慢蛋白在细胞内寻找并结合其正确位点的能力。

3）我们能不能在活细胞里一次性研究成千上万种变体？

活细胞中的单分子显微成像非常有威力，但传统上非常耗时——通常一次只能看一种蛋白。论文 III 提出了一种初步的解决思路：采用“混池”（pooled）的方式，在每个菌株的染色体上写入两个短条形码和一个蛋白序列变体。先对活细胞进行拍摄，记录单分子的行为，然后在显微镜上直接通过基因分型方法读出这些条形码，从而把每一种表型行为重新对应回它的遗传变体。在一个包含六种基因型的小规模试验中，这种方法成功地正确识别了其中五种，并且重现了预期的表型行为。我们也据此界定了如果要把该方法推广到成千上万种变体，还需要在哪些方面进一步改进。

核心结论：

LacI——很可能也包括许多其他转录因子——主要是通过“控制进入识别态的概率”来获得特异性，而不是简单地通过“黏得更久”。一个很小的结构开关（铰链）在“牢牢控制目标”和“快速响应信号”之间调节折衷，而对靶点搜索和结合效率本身的影响却相对有限。借助混池的单细胞单分子追踪成像，我们正朝着在天然细胞环境中、在整片蛋白变体“景观”上系统检验这些观点的方向迈进。

11. Acknowledgement

Firstly, I would like to thank my supervisor, **Johan Elf**, thank you for the guidance and support throughout my PhD journey. I deeply appreciate the resources, professional direction, and prompt responses whenever I needed help. As the last standing member of the LacI project for now, I hope the LacI tradition continues after I leave... **Emil Marklund**, thank you for the fascinating discussions during our work on my first project in Lab and for being an incredible co-supervisor, who would just prompt answering things in Slack even when you were thousands of miles away and in different time zone. I also cherish the fun memories from the beer clubs you hosted. **Malin Lüking**, thank you for being my "LacI-sister." I wish I had cherished our discussions more while you were in the lab; I took them for granted at the time. I miss your bright energy and contagious laughs you always bring to the room. I wish you the best with your industry career. **David Fange**, thank you for your patience with my endless questions regarding analysis methods, kinetic modeling, and math. I appreciate your insightful suggestions whenever I ask. **Spartak Zikrin**, thank you for optimizing my image analysis pipeline—turning a week's work into 8 minutes was impressive! Your advice on coding and pipelines saved me from many struggles. **Beer Chakra Sen**, thank you for your advice on strain construction and gene editing. I learned a lot from your organized working style, and I believe we made a great team during the Miller assays. **Jimmy Larsson**, it was always delightful interacting with you. Thank you for your positive energy and for helping me loading bacterial cells into the chips, which ensured my experiments ran smoothly. **Irmeli Barkefors**, thank you for editing my reports, papers, and thesis. I also appreciate our fun lunch talks and your career advice whenever I ask. **Dvir Schirman**, thank you for inspiring me with your scientific curiosity and passion. You also inspired my culinary life—I only started making my own bread and hummus after tasting yours. **Vinodh Kandavalli**, we entered the lab around the same time, and I have always valued your scientific advice, kindness, and patience. You were always my go-to person for biological and experimental questions. Our trip to Durham was a highlight—a beautiful city

with great indian food! Not with Chinese food, ewww. **Elias Amselem**, thank you for fixing the flow control for my microarray experiments and for the great recommendations for Chinese restaurants in Stockholm. **Johan Öhman**, thank you for baking tasty buns while you were in Elf lab; the smell from the oven always lifted the mood during group meetings. **Petter Knagge**, thank you for the insightful chats when we shared an office. It was interesting to learn about the business development side of academic projects. **Irfan Ahmad**, thank you for your kind advice and insightful suggestions regarding the questions I brought to your office. **Oscar Broström**, thank you for the fun lunch talks and for always being patient when I needed help with report, thesis questions, or just figuring out the printer color problem. **Konrad Gras**, I appreciate your company in the beer club and our shared iGEM connection. Thank you for your kindness and patience whenever I asked for help. **Henar Marino**, it is always fun to hang out with you, nice to have our common interests in food and travel! **Buu Minh Tran**, thank you for the cultural exchange talks and Vietnam travel tips. Seeing how you write your manuscript was inspiring, if I get an application oriented project I should also do that, if ever… Anway, I wish the best for you and your family, it must be exciting to have a new born! **Lisa Johansson**, once I graduate, you will be both the most senior and junior PhD student in the Elf lab! I am sure you will go far with your hard work and all the guidance available to you. Good luck! **Praneeth Karempudi**, I am glad you successfully transitioned from biology to IT as you wanted. Thank you for the talks in my early PhD years, and best of luck in software development field. **Vatsala Chauhan**, you were a legend in Vinodh and Henar's stories until I finally met you! I appreciate your cheerful energy. I enjoyed swimming with you and hope we can manage to go dancing sometime. **Daniel Jones**, thank you for being my co-supervisor and for the kind help and suggestions regarding my half-time and presentations whenever I ask. **Luke Richards**, thank you for the delightful lunch talks and fun conversations; you always bring great energy to the daily work environment. **Filippa Nilsson**, I appreciate the positive energy and fun lunch talks you always bring. **Magnus Johansson**, thank you for the insightful scientific advice and fun lunch time conversations. **Ivan Volkov**, I am impressed by your curiosity and detail-oriented attitude. I wish the best for you and Viktoriia. **Mikhail Metelev**, thank you for all the fun lunch talks, and I wish the best for you and your family. **Anneli Borg**, thank you for accompanying me during the microarray experiment when we were both puzzled, and for the nice lunch talks we had. **Erik Lundin**, I enjoyed our challenging and thought-

provoking conversations regarding culture and traditions. I appreciate your straightforwardness. **Filip Ilievski**, thank you for introducing me to TNDR, which I got involved in quite a bit during my 2nd and 3rd years. **Tora Hävermark**, I admire how you maintained good science while managing excellent instrument performance in your family band. Also thank you for the heartwarming conversations. **Linnea Wikström**, thank you for the great company during lunch and game nights; you always light up the room. **Anthony Forster**, thank you for your guidance during my Master's. It was intense but fruitful, and I appreciate the opportunity to be included on a research paper as a master student.

I would also like to thank the people outside BMC, who are not part of my everyday life but still hold a very special place in my heart.

Viktor Öhrner, you hold a soft spot in my heart. You made my life in Uppsala exciting yet cozy. I deeply appreciate your supportive and calm presence in my life. You are truly the best travel companion I could ever wish for, and I hope we can continue exploring the world together once a year forever. **Yuanxi Zou**, I vividly remember picking up you on the bus, haha, glad we became friends since then. Thank you for being the best companion for nations, parties, gasques, hotpot, fika, teak, game nights….so many fun moments! Most importantly, if it is not for your advice during my graduation year, I might not even be in Sweden today. **Heyin Chen**, we've known each other since day one in 2018. I cherish our 9-hour "fika" marathons. Best of luck with your new life and marriage in Beijing. **Songyi Ma**, it was a pleasure getting to know you and your independent spirit. Thank you for the sailing trip—it opened a new world map for me—and for our talks on financial responsibility. **Zhaorui Zhang**, it's funny how we get more familiar with each other, since the night in nation we randomly bumbed into each other. Thank you for the makeup & cocktail nights, hotpot, and beer club hangouts. There is many more to come in Uppsala for us. **Katherin Vallejos,** thank you for being my iGEM project leader. That iGEM summer bring me so many friends. I really appreciate our hangouts, and those memories still make me smile. **Vanja Höynälä**, thank you for the cozy family dinners, cocktails, and games with you and Emil. Hopefully we can make an Asia trip together later, I can be your local guide for China part at least. **Linnéa Yuan Andersson**, thank you for initiating our language exchange "dates." I really cherish our long talks and board game nights. I am glad you found your passion in Muay Thai, it is

inspiring to see how passion can bring such big change to people's life. **Huifang You**, since I've known you, you have been a lovely fairy who brings the best energy to every gathering. I hope you achieve your goals and build a great life in Hangzhou, Hangzhou is such a dreamy place to settle down. **Yupei Li**, my first impression of you at your birthday party was of a lovely, humorous girl with a bright laugh. I am glad you got your PR in Sweden, and I wish the best for you and Florian. **David Holmberg & Madeleine Drabe**, it was a pleasure to share our foodie enthusiasm with you. You inspired us with your cooking, and I hope we inspired you a little, too. **Dennis Leenheer,** thank you for the tasty home-brewed beers. I wish you the best in either your beer career or science; I know you will shine either way. **Letian Bao**, thank you for the scientific discussions and co-working time when I was in Tony's lab. Your insights were invaluable. **Zhen Chen & Sicong Wang**, it was such a coincidence to discover that we are not only from the same hometown and also from the same high school yet met here in Uppsala! **Haiyan Liu**, thank you for being my first professor in academia. You showed me a great example of female leadership in science, and I admired the harmonious lab environment you created. **Lixiang Zhao**, thank you for always being helpful and willing to assist during my Bachelor's thesis. **Yu Mei**, thank you for teaching me experiments on the side while you were working as a PhD student. **Ziqi Jin**, thank you for being a mentor figure for me during my bachelor thesis time. I learned a lot following you around for DNA extraction and other tasks. **Peifen Gao & Yejun Qiao**, I am glad we were roommates for four years. I deeply appreciate that we have kept in touch until today. **Jinsong Xu,** thank you for the company during our first and second college years, hanging out for student organizations and making our college lives fun. **Wanye Wei,** thank you for encouraging me to study abroad back then. **Yupeng Gu**, who would have thought we'd still be in contact since middle school? I wish you a happy marriage and the best for your family. **Yamin Lu**, though we talk on and off, I know we share a similar vision and strength for life. Let's keep solving life problems one by one—we will figure it all out. 最后也是最重要的，感谢我爸爸妈妈，永远支持我的决定也对我很放心，感谢爸爸妈妈理解我一生不羁放纵爱自由的心。希望爸爸妈妈长命百岁，永远健健康康，好好锻炼身体，我们每年一起国内国外旅游。

12. References

1. Hoopes, L. Introduction to the gene expression and regulation topic room. *Nature Education* **1**, (2008).
2. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).
3. Browning, D. F. & Busby, S. J. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* **2**, 57–65 (2004).
4. Alberts, B. *et al. Molecular Biology of the Cell: Seventh International Student Edition with Registration Card*. (W. W. Norton & Company, 2022).
5. Riggs, A. D., Bourgeois, S. & Cohn, M. The lac represser-operator interaction. *J. Mol. Biol.* **53**, 401–417 (1970).
6. Smoluchowski, M. v. Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen. *Z. Phys. Chem. (N F)* **92U**, 129–168 (1918).
7. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry* **20**, 6929–6948 (1981).
8. Richter, P. H. & Eigen, M. Diffusion controlled reaction rates in spheroidal geometry. *Biophys. Chem.* **2**, 255–263 (1974).
9. Berg, O. G. & Blomberg, C. Association kinetics with coupled diffusional flows. Special application to the lac repressor--operator system. *Biophys. Chem.* **4**, 367–381 (1976).
10. von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
11. Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L. & Xie, X. S. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5752–5757 (2006).
12. McKinney, K., Mattia, M., Gottifredi, V. & Prives, C. p53 linear diffusion along DNA requires its C terminus. *Mol. Cell* **16**, 413–424 (2004).

13. Gorman, J. *et al.* Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2-Msh6. *Mol. Cell* **28**, 359–370 (2007).
14. Elf, J., Li, G.-W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).
15. Hammar, P. *et al.* The lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598 (2012).
16. Marklund, E. *et al.* DNA surface exploration and operator bypassing during target search. *Nature* **583**, 858–861 (2020).
17. Milo, R. & Phillips, R. *Cell Biology by the Numbers*. (Garland Science, 2015).
18. Grönlund, A., Lötstedt, P. & Elf, J. Transcription factor binding kinetics constrain noise suppression via negative feedback. *Nat. Commun.* **4**, 1864 (2013).
19. Jones, D. L., Brewster, R. C. & Phillips, R. Promoter architecture dictates cell-to-cell variability in gene expression. *Science* **346**, 1533–1536 (2014).
20. Ali, M. Z., Parisutham, V., Choubey, S. & Brewster, R. C. Inherent regulatory asymmetry emanating from network architecture in a prevalent autoregulatory motif. *Elife* **9**, (2020).
21. Morrison, M., Razo-Mejia, M. & Phillips, R. Reconciling kinetic and thermodynamic models of bacterial transcription. *PLoS Comput. Biol.* **17**, e1008572 (2021).
22. Gilbert, W. & Müller-Hill, B. Isolation of the lac repressor. *Proc. Natl. Acad. Sci. U. S. A.* **56**, 1891–1898 (1966).
23. Griffiths, A., Doebley, J., Peichel, C. & Wassarman, D. A. *Introduction to Genetic Analysis*. (W.H. Freeman, New York, NY, 2019).
24. Slutsky, M. & Mirny, L. A. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035 (2004).
25. Kaback, H. R., Bibi, E. & Roepe, P. D. β-GALACTOSIDE TRANSPORT IN E. COLI: A FUNCTIONAL DISSECTION OF LAC PERMEASE. in *Proteins: Form and Function* 59–69 (Elsevier, 1990). doi:10.1016/b978-1-85166-512-9.50010-8.
26. Juers, D. H., Matthews, B. W. & Huber, R. E. LacZ β-galactosidase: structure and function of an enzyme of historical and molecular biological importance: LacZβ-Galactosidase. *Protein Sci.* **21**, 1792–

1807 (2012).
27. Lewis, M. *et al.* Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* **271**, 1247–1254 (1996).
28. Chen, J., Alberti, S. & Matthews, K. S. Wild-type operator binding and altered cooperativity for inducer binding of lac repressor dimer mutant R3. *J. Biol. Chem.* **269**, 12482–12487 (1994).
29. Lewis, M. The lac repressor. *C. R. Biol.* **328**, 521–548 (2005).
30. Friedman, A. M., Fischmann, T. O. & Steitz, T. A. Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science* **268**, 1721–1727 (1995).
31. Gilbert, W. & Maxam, A. The nucleotide sequence of the lac operator. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3581–3584 (1973).
32. Krämer, H. *et al.* lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. *EMBO J.* **6**, 1481–1491 (1987).
33. Whitson, P. A., Hsieh, W. T., Wells, R. D. & Matthews, K. S. Influence of supercoiling and sequence context on operator DNA binding with lac repressor. *J. Biol. Chem.* **262**, 14592–14599 (1987).
34. Dunaway, M. & Matthews, K. S. Hybrid tetramers of native and core lactose repressor protein. Assessment of operator and nonspecific DNA binding parameters and their relationship. *J. Biol. Chem.* **255**, 10120–10127 (1980).
35. Mandal, N., Su, W., Haber, R., Adhya, S. & Echols, H. DNA looping in cellular repression of transcription of the galactose operon. *Genes Dev.* **4**, 410–418 (1990).
36. Sadler, J. R., Sasmor, H. & Betz, J. L. A perfectly symmetric lac operator binds the lac repressor very tightly. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 6785–6789 (1983).
37. Oehler, S., Eismann, E. R., Krämer, H. & Müller-Hill, B. The three operators of the lac operon cooperate in repression. *EMBO J.* **9**, 973–979 (1990).
38. Berger, M. F. & Bulyk, M. L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat. Protoc.* **4**, 393–411 (2009).
39. Winter, R. B., Berg, O. G. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions. *Biochemistry* **20**, 6961–6977 (1981).
40. Kandavalli, V., Zikrin, S., Elf, J. & Jones, D. Anti-correlation of LacI

association and dissociation rates observed in living cells. *Nat. Commun.* **16**, 764 (2025).
41. Kalodimos, C. G. *et al.* Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386–389 (2004).
42. Bell, C. E. & Lewis, M. Crystallographic analysis of Lac repressor bound to natural operator O1. *J. Mol. Biol.* **312**, 921–926 (2001).
43. Bell, C. E. & Lewis, M. A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol.* **7**, 209–214 (2000).
44. Swint-Kruse, L. & Matthews, K. S. Allostery in the LacI/GalR family: variations on a theme. *Curr. Opin. Microbiol.* **12**, 129–137 (2009).
45. Blackwood, E. M. & Eisenman, R. N. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* **251**, 1211–1217 (1991).
46. Schütz, S. *et al.* Intrinsically disordered regions in the transcription factor MYC:MAX modulate DNA binding via intramolecular interactions. *Biochemistry* (2024) doi:10.1021/acs.biochem.3c00608.
47. Lüking, M., Elf, J. & Levy, Y. Conformational Change of Transcription Factors from Search to Specific Binding: A lac Repressor Case Study. *J. Phys. Chem. B* **126**, 9971–9984 (2022).
48. Sengupta, R., Capp, M. W., Shkel, I. A. & Record, M. T., Jr. The mechanism and high-free-energy transition state of lac repressor-lac operator interaction. *Nucleic Acids Res.* **45**, 12671–12680 (2017).
49. Zhan, H., Swint-Kruse, L. & Matthews, K. S. Extrinsic interactions dominate helical propensity in coupled binding and folding of the lactose repressor protein hinge helix. *Biochemistry* **45**, 5896–5906 (2006).
50. Daber, R., Sochor, M. A. & Lewis, M. Thermodynamic analysis of mutant lac repressors. *J. Mol. Biol.* **409**, 76–87 (2011).
51. Gatti-Lafranconi, P., Dijkman, W. P., Devenish, S. R. A. & Hollfelder, F. A single mutation in the core domain of the lac repressor reduces leakiness. *Microb. Cell Fact.* **12**, 67 (2013).
52. Soares, R. R. G. *et al.* Pooled optical screening in bacteria using chromosomally expressed barcodes. *Commun. Biol.* **8**, 851 (2025).
53. Mukherjee, S. *et al.* Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat. Genet.* **36**, 1331–1339 (2004).
54. Bulyk, M. L., Gentalen, E., Lockhart, D. J. & Church, G. M.

Quantifying DNA-protein interactions by double-stranded DNA arrays. *Nat. Biotechnol.* **17**, 573–577 (1999).
55. Bulyk, M. L., Huang, X., Choo, Y. & Church, G. M. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 7158–7163 (2001).
56. Berger, M. F. *et al.* Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* **24**, 1429–1435 (2006).
57. Badis, G. *et al.* Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720–1723 (2009).
58. Siggers, T., Duyzend, M. H., Reddy, J., Khan, S. & Bulyk, M. L. Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex. *Mol. Syst. Biol.* **7**, 555 (2011).
59. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10**, 354–360 (2013).
60. Revzin, A. & von Hippel, P. H. Direct measurement of association constants for the binding of Escherichia coli lac repressor to non-operator DNA. *Biochemistry* **16**, 4769–4776 (1977).
61. Beta-Galactosidase Assay (A better Miller). https://openwetware.org/mediawiki/index.php?title=Beta-Galactosidase_Assay_(A_better_Miller)&oldid=620416.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 2614

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-571679

ACTA UNIVERSITATIS
UPSALIENSIS
2025