Cell. Mol. Life Sci. 64 (2007) 3–16
1420-682X/07/010003-14
DOI 10.1007/s00018-006-6296-z


Cellular and Molecular Life Sciences

Review

# The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding

**C. J. Wilson [a, b, c] H. Zhan [a, d] and L. Swint-Kruse [d] and K. S. Matthews [a, b, *]**

[a] Department of Biochemistry and Cell Biology, Rice University, 6100 Main Street, Houston, Texas 77005 (USA), Fax: +1 713 348 6149, e-mail: ksm@rice.edu
[b] Keck Center for Interdisciplinary Bioscience Training, Rice University, 6100 Main Street, Houston, Texas 77005 (USA)
[c] Division of Biology, California Institute of Technology, Mail Code 147-75, Pasadena, California 91125 (USA)
[d] Department of Biochemistry and Molecular Biology, MS 3030, The University of Kansas Medical Center, Kansas City, Kansas 66160 (USA)

Received 27 June 2006; received after revision 1 August 2006; accepted 28 September 2006
Online First 13 November 2006

**Abstract.** In 1961, Jacob and Monod proposed the operon model for gene regulation based on metabolism of lactose in *Escherichia coli* [1]. This proposal was followed by an explication of allosteric behavior by Monod and colleagues [2]. The operon model rationally depicted how genetic mechanisms can control metabolic events in response to environmental stimuli via coordinated transcription of a set of genes with related function (e.g. metabolism of lactose). The allosteric response found in the lactose repressor and many other proteins has been extended to a variety of cellular signaling pathways in all organisms. These two models have shaped our view of modern molecular biology and captivated the attention of a surprisingly broad range of scientists. More recently, the lactose repressor monomer was used as a model system for experimental and theoretical explorations of protein folding mechanisms. Thus, the *lac* system continues to advance our molecular understanding of genetic control and the relationship between sequence, structure and function.

**Keywords.** Lactose repressor, allostery, genetic regulation, transcription regulation, inducer, operator.

## Transcription regulation

Efficient sensing and utilization of changing food supplies is key to survival. Bacteria share with many organisms the strategy of shifting concentrations of metabolic catalysts in response to varying nutrient availabilities. When glucose is plentiful, *Escherichia coli* utilizes this carbon source exclusively, even when other sugars are present. Only when glucose is depleted does the bacterium upregulate expression of proteins that transport and metabolize other carbon sources, such as lactose. Understanding the molecular mechanisms by which an organism selects the most energy-efficient metabolites from a constantly changing environment has been a fundamental problem in biology over the last 5 decades.

## The Jacob-Monod repression model

The seminal work of Jacob and Monod [1] generated a theoretical framework based on their studies of lactose utilization in *E. coli* that delineated how bacteria might switch from one carbon source to another. At the heart of their remarkable hypothesis is the *lac* operon. This unit of DNA encodes a group of 'structural genes' that are transcribed as a single messenger RNA (mRNA) in response to glucose/lactose ratios. Jacob and Monod postulated that

* Corresponding author.

the lactose structural genes are coordinately regulated by a molecular toggle switch ('repressor'), which represses transcription until it interacts with a chemical 'inducer'. Repression is mediated by reversible interaction between the repressor and a regulatory element, called an 'operator.' Induction of the repressor results in production of mRNA, the encoded proteins and consequently enzymatic activity of these structural genes. (Note that at the time of Jacob and Monod's observations [1], the chemical identities of two of the three regulatory components were unknown.) Empirical evidence soon identified (i) a soluble regulatory agent – lactose repressor protein – that controlled expression of the structural genes responsible for lactose metabolism [3, 4], (ii) a specific interaction between the repressor and the DNA regulatory operator [5, 6], and (iii) modulation of this binding event by inducer sugars [7, 8].

## The *lac* operon

A variety of reviews cover many elements of the *lac* operon and particularly its repressor [9–13]; we will touch only on a subset of the salient elements in this review. In the absence of lactose, the protein product of the *lacI* gene (LacI) binds with high affinity to the *lac* operon at a specific operator DNA sequence [14, 15]. In turn, RNA polymerase binding, elongation and/or initiation have each been reported to be compromised [16–18]. The outcome is a dramatic reduction, but not elimination, of *lac* mRNA. Specifically, the regulated genes are *lac*Z, *lac*Y and *lac*A, which encode three proteins involved in lactose metabolism – $\beta$-galactosidase, *lac* permease and thiogalactoside transacetylase, respectively (Fig. 1) [1, 9, 19, 20]. The transmembrane protein, *lac* permease, facilitates lactose transport [21]. $\beta$-Galactosidase cleaves lactose into galactose and glucose, the first step in lactose metabolism [8, 22]. Thiogalactoside transacetylase transfers an acetyl group from coenzyme A (CoA) to the hydroxyl group of galactosides. Although not essential for lactose metabolism, transacetylase is physiologically important for maintaining cellular viability [20].

The permease and galactosidase activities are also essential for the process of 'turning on' *lac* enzyme production (Fig. 1). When lactose is available, basal levels of *lac* permease and $\beta$-galactosidase transport this sugar

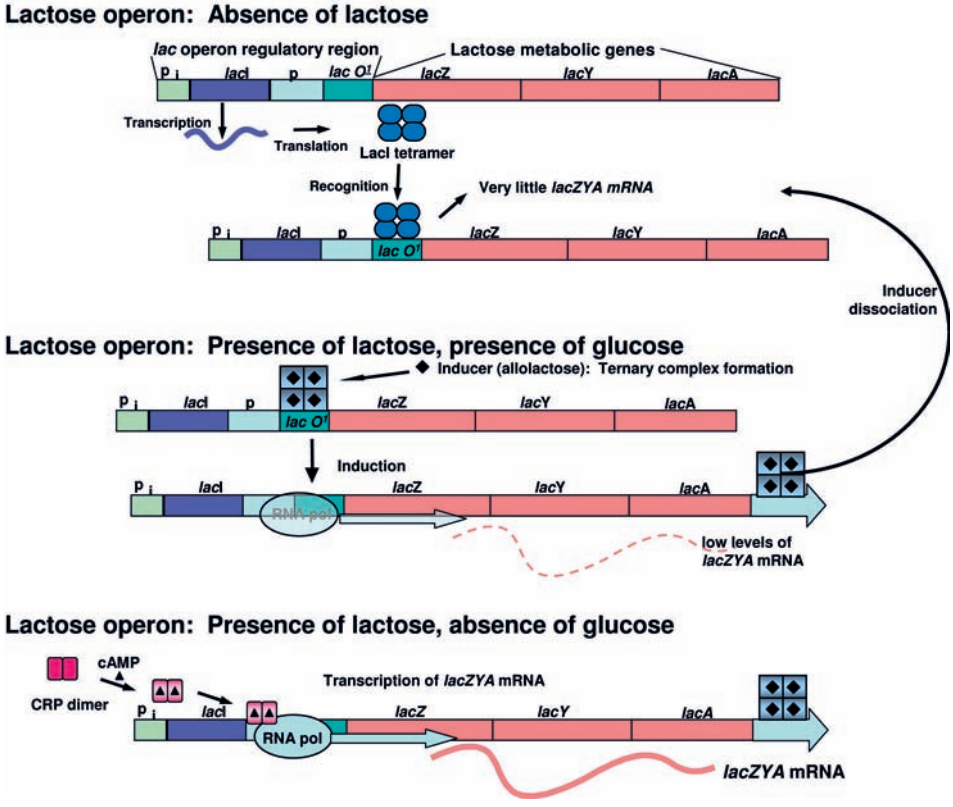


**Figure 1.** Schematic of the lactose operon. In the absence of lactose, *lac*I mRNA is transcribed from the *lac*I promoter ($p_i$) and translated into LacI protein, which binds with high affinity to the lactose operator ($O^1$). Since the LacI binding site overlaps the *lac* promoter (p), transcription of the *lac* operon structural genes encoding $\beta$-galactosidase (*lac*Z), lactose permease (*lac*Y) and thiogalactoside transacetylase (*lac*A) is inhibited. The structural genes encode the metabolic enzymatic activity required for lactose utilization. In the presence of lactose, a secondary metabolite, 1,6-allolactose, is produced that binds to LacI and diminishes its affinity for the operator site, resulting in a small increase in the amounts of *lac* metabolic enzymes produced. If glucose is absent, the CAP-cAMP complex forms and facilitates transcription of the *lac* structural proteins, resulting in a large increase in the amount of *lac* metabolic enzymes generated.

and convert a small fraction of lactose into 1,6-allolactose [8]. 1,6-Allolactose binds with high affinity to LacI and 'induces' the repressor protein by lowering its affinity for the operator sequence, thereby allowing transcription of the polycistronic *lac*ZYA mRNA [8]. This feature is now commonly exploited in molecular biology and biotechnology by substituting *lac*ZYA DNA with other desired genes. Allolactose can also be substituted with a variety of gratuitous inducers (e.g. isopropyl-*β*,D-thiogalactoside, IPTG) that are not metabolized but elicit the induction response [1, 23].

Interestingly, when *E. coli* are provided both glucose and lactose, the cells preferentially metabolize glucose [24, 25]. Even when the repressor does not occupy the operator site, the operon is transcribed infrequently and remains largely inactive until glucose is depleted. The mechanism for selective utilization of available metabolites was not identified until the discovery of the catabolite activator/repressor protein and its regulator – cyclic AMP (reviewed in [24]). When glucose levels are depleted, adenylyl cyclase produces high levels of cAMP from ATP. The substrate cAMP subsequently binds to cyclic AMP-dependent catabolite activator/repressor protein (CAP). In turn, the CAP-cAMP complex binds a DNA sequence in the promoter region located just upstream of the *lac* promoter (Fig. 1). This DNA binding event enhances the affinity of RNA polymerase for the promoter and thereby increases the expression of the *lac* structural genes nearly 50-fold [24, 25]. Together, repression and activation account for the selective utilization of available metabolites and thereby the growth behavior originally observed by Jacob and Monod.

## The operators and DNA looping

The primary operator site ($O^1$) for the *lac* operon was sequenced by Gilbert and Maxam [14] nearly a decade after Jacob and Monod published their model [1]. The *lac* repressor protein binds to this largely palindromic operator, and the LacI•$O^1$ complex is disrupted by the presence of the inducer IPTG [7]. In addition to $O^1$, two auxiliary pseudo-operators ($O^2$ and $O^3$), were identified with sequences similar to the primary operator [26, 27]. *In vivo*, these pseudo-operators in the Z- and I-genes are required for maximal repression. We now know that the tetrameric *lac* repressor is ideally suited to bind two operators simultaneously, creating so-called 'repression loops' [26–32]. DNA looping enhances LacI affinity for multi-operator sequences, and supercoiling these DNAs yields complexes with remarkable stability [33–38]. A number of synthetic operator variants have also been constructed (e.g. [39]) and have proved very useful for understanding molecular mechanisms of repression. The most notable artificial operator is the perfectly symmetric sequence $O^{sym}$ [40, 41]; use of this sequence was critical for obtaining a high-resolution crystal structure of the repressor protein (see below).

## Molecular mechanism of *lac* operon repression

The molecular mechanism of repressing the *lac* operon requires that LacI protein be capable of binding both operator DNA and inducing sugar. The possibility of competitive binding by these ligands was eliminated by the demonstration that protease digestion selectively cleaves LacI to two fragments: (i) a tetrameric 'core' (residues 60–360 of the monomer) that retains inducer binding properties and (ii) a monomeric N-terminal headpiece (amino acids 1–59) capable of binding DNA, albeit with substantially reduced affinity [42, 43]. For induction to occur, the binding sites on the different functional domains must have a means of communication – an allosteric relationship. Additional deletion studies later identified a small C-terminal tetramerization domain and showed that the dimeric unit of LacI retained DNA- and inducer-binding properties, as well as allosteric regulation [44, 45].

In 1963, Monod, Changeux and Jacob used the concept of allostery to describe concerted structural isomerizations that result when an effector ligand binds to a protein and influences its ability to perform other function(s) [2]. Monod, Wyman and Changeux later formalized this idea mathematically, resulting in the 'MWC' model [46]. A competing, sequential model of 'induced fit' was postulated by Koshland and colleagues [47]. Both models invoked conformational changes in the protein. These ideas of how an effector ligand induces a reversible change in the conformation and function of the protein are still commonly accepted interpretations of the word 'allosteric'. However, the past 5 years have seen an explosion in the molecular understanding of allostery and its prevalence in a wide range of proteins (e.g. [48–50]).

## LacI function: ligand binding and allosteric regulation

Once the biological activity of LacI was defined, attention turned to biochemical and biophysical analysis of its function. For operator DNA and sugar binding, equilibrium analysis with purified protein gave valuable insight into thermodynamic features and general mechanistic aspects of LacI function (reviewed in [9–12]). Three classes of sugar effectors were identified. Most sugars induced diminished affinity of LacI for DNA (inducers), but a few made DNA binding even tighter (anti-inducers); a third group bound specifically to the protein but had no functional effect (neutral) [23]. Allosteric regulation can be quantitated as the difference in affinities for one ligand in the absence and presence of saturating levels of the other ligand, and kinetics studies provide insight into the

time scale of binding events and allosteric response (e.g. [51–54]).

**LacI structure and allosteric conformational change**

The first indications of global conformational change associated with the allosteric response to inducer binding were provided by spectroscopic and hydrodynamic studies [55, 56]. Ultraviolet (UV)-difference spectra of LacI modified with N-bromosuccinimide and unmodified protein indicated that aromatic residues in the trypsin-resistant core domain are affected by conformational isomerization due to inducer binding [57]. Inducer binding also resulted in a measurable difference in the sedimentation coefficient of the repressor [55]. Moreover, when small crystals of LacI were exposed to the chemical inducer, they instantaneously 'cracked,' indicating a significant structural shift when this sugar bound to the protein [58]. These experiments were part of a huge body of biochemical exploration during the 3 decades that passed before X-ray structures of LacI were acquired. Even in the absence of structural information, biochemical and biophysical work was extremely powerful at 'seeing' features of LacI. However, in order to place residue-specific results in the context of the structure, we will depart from chronological order to first describe the atomic-level structures.

Despite Herculean efforts by several groups, high-resolution determination of any portion of this key gene regulator remained elusive until the isolated N-terminal domain bound to DNA was solved using nuclear magnetic resonance (NMR) [59] and the tetrameric tryptic core fragment bound to IPTG yielded to crystallization [60]. Subsequently, three-dimensional structures were determined for multiple forms of LacI in various liganded states [13, 61–64]. From the crystallographic structures, LacI is a tetrameric protein comprised of identical, ~37.5 kDa, monomers, as shown in Figure 2 [60, 61]. The protein is best described as a 'dimer of dimers', consistent with functional results of deleting the tetramerization domain [44]. Monomers are separated structurally into three domains: (i) a helix-turn-helix motif and hinge region that bind operator DNA, (ii) a core domain that binds to LacI sugar ligands in a cleft between two subdomains (N- and C-) and encompasses the monomer-monomer interface, and (iii) a C-terminal domain that comprises the tetramerization domain (see Fig. 2a–c [60, 61]). These structural regions correspond almost perfectly with the domains identified by functional deletion described above.

**DNA binding domain and hinge region**

The DNA binding domain (residues 1–50) of the *lac* repressor folds into a classical helix-turn-helix motif [65] that fits into the major groove of operator DNA and forms electrostatic interactions with the phosphate backbone, burying more than 3300 Å$^2$ of solvent accessible surface area [59, 61]. The hinge region (residues 50–60) connects the DNA binding domain to the core of the repressor [61, 62]. This region appears to be devoid of secondary structure in the absence of DNA [61], providing the helix-turn-helix DNA binding domain a wide range of motion [66]. When the repressor interacts with operator DNA, a coil-to-helix transition occurs in this hinge region, coupled with interactions in the DNA minor groove that generate an ~45° kink in the operator and thereby distort the canonical B-form of the DNA [61, 62, 67–69]. In addition, the hinge region forms extensive interactions with the core N-subdomain of the adjacent monomer [62, 70].

**The regulatory core domain**

The structure of the central core domain (residues 60–340) belongs to the very diverse periplasmic binding protein superfamily [71], sometimes designated the Venus flytrap motif [72]. The core domain consists of two structurally similar subdomains (N- and C-), each with three structural layers ($\alpha\beta\alpha$) comprising a six-stranded parallel $\beta$-sheet flanked by $\alpha$-helices [60, 61]. These two subdomains are intertwined such that one $\alpha$-helix and one $\beta$-strand (helix 13 and strand K, Fig. 2c) in the N-subdomain are contributed by C-terminal sequences, resulting in three polypeptide-crossover points between the two subdomains. The central portion of this three-strand interconnection demarcates a cleft between the two subdomains that constitutes the effector ligand binding site.

The functional DNA binding unit of LacI is the dimer [44, 73, 74], which is held together via an extensive monomer•monomer interface on the core domain. Structural studies show that this interface consists of two sub-interfaces, composed by each of the N-subdomain and C-subdomains (Fig. 2a). The interface between monomers buries ~2200 Å$^2$ of solvent accessible surface area, nearly equally distributed between the interfaces of the two subdomains [60, 61]. Upon binding inducer, residues in the N-subdomain show significant movement with respect to their locations in the DNA-bound complex, a property that is crucial for allosteric regulation [61, 75–77]. In contrast, the C-subdomains remain essentially unchanged in their orientation in all liganded states (Fig. 2b).

**The tetramerization domain**

Unlike many homotetramers, LacI does not have a point group symmetry and is in essence a V-shaped dimer of dimers [60, 61]. Only a small ~300 Å$^2$ surface area stabilizes this arrangement [10, 60, 61], and the relationship between the dimers appears flexible [78, 79]. In contrast, the C-terminal region that mediates tetramer assembly forms a highly stable interface comprised of a four-helix, anti-parallel bundle [45, 61]. Deletion of 5–32 amino acids from the C-terminus of LacI and alteration of this region limit oligomerization to the dimeric state [44, 45]. Inducer binding affinity and cooperativity of the dele-

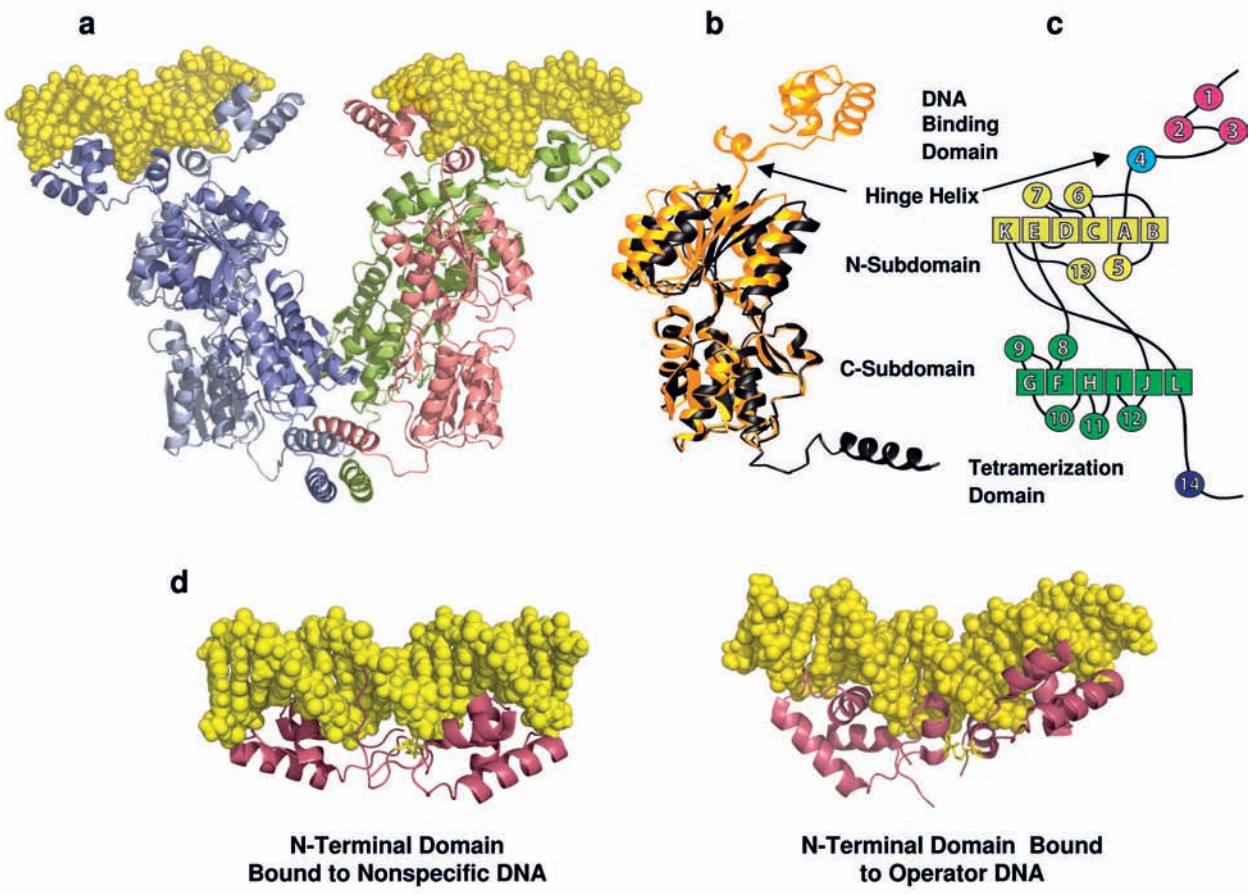


**Figure 2.** LacI structures. (*a*) Tetrameric LacI•O$^{sym}$ complex (PDB file 1lbg). Note that DNA is bound to each dimer within the tetramer. The left dimer is blue/grey, the right green/pink. (*b*) Comparison of monomeric LacI structure in the presence (gold) and absence of DNA (black) (from PDB files 1efa and 1lbh) [61]. The domains within LacI are labeled to orient all structures. The C-subdomain undergoes minimal alteration between the IPTG- and DNA-bound states, whereas the N-subdomain is displaced in a motion akin to screwing a jar top [61]. The DNA binding domain is not resolved in the crystal structure of the IPTG-bound protein; this region appears highly mobile in the apoprotein [66]. (*c*) LacI monomer topology and domain structure. Note that the topology of folding involves three crossovers between the N- and C-subdomains. (*d*) Nonspecific and specific DNA complexes of the N-terminal DNA binding domain (amino acids 1–62 cross-linked at V52 by a disulfide bond) (PDB files 1OSL and 1L1M) [68, 69, 94]. Note that the nonspecific complex of the disulfide-linked N-terminal DNA binding domain does not alter DNA topology, whereas the complex with operator DNA distorts the double helix in a manner similar to the full-length LacI [61].

tion mutants and wild-type protein are very similar; thus, subunit communication with respect to inducer binding involves the dimer unit, even within the tetrameric protein [44, 73]. Although the dimeric protein exhibits diminished operator affinity [44, 45], the decrease is not due to a change in intrinsic DNA binding affinity but to increased dissociation of dimer to monomer and thermodynamic coupling of monomer•monomer assembly and operator binding [73].

## Residue-specific analysis of LacI function

To provide a more complete picture of the function of this protein, we will now return to a brief summary the data amassed over several decades. A highly resolved understanding of amino acid side-chain participation in oligomer assembly, ligand binding and allosteric response has been generated using genetic analyses, spectroscopy, chemical modification and mutational variants.

### Genetic analysis

Miller and colleagues (e.g. [80–82]) used 14 different suppressors of nonsense mutations to generate more than 4000 single amino acid substitutions throughout the protein. Gordon et al. [83] also generated a significant set of random mutations in the LacI gene that abolished repression. Phenotypic analyses of these mutations generated a broad-based, if low-resolution, functional description of each amino acid position. In general, LacI variants could be segregated into two distinct populations: (i) those incapable of repressing, presumably due to the absence of requisite operator DNA interaction ($I^-$); and (ii) repressor molecules that do not respond to IPTG, presumably due either to inability to bind inducer or to undergo the re-

quisite allosteric changes ($I^s$). Some positions could have either phenotype, depending upon the substitution.
A summary of this massive genetic analysis revealed that functional variants were not uniformly scattered throughout the sequence: large stretches of the primary sequence are quite tolerant to substitutions without phenotypic effect, whereas other regions appear critical for active repressor [80–83]. As expected, the N-terminal DNA binding domain had a number of $I^-$ residues, whereas the core domain had more scattered $I^-$ regions that often map onto the residues required for folding and assembly from the structure [80–83]. In contrast, few $I^s$ variants were found in the N-terminal DNA binding domain of the protein, but a larger number were scattered throughout the core domain, often in groups. Variations in the primary sequence that affect function can occur for a range of reasons, but the patterns suggested a role for a subset of these residues in binding inducer [80–83]. When mutational effects of substitutions were later mapped onto the LacI structure, a clear rationale emerged for many [82].

**Spectral analysis**

Spectroscopic characterization employed both intrinsic and extrinsic probes. LacI contains two tryptophan residues (at positions 201 and 220) and eight tyrosine residues, five of which are situated in the core. Consistent with later structural data, studies of the wild-type protein and variants at the two tryptophan sites showed that W220 is partially exposed to solvent in the absence of inducer, whereas W201 is buried and makes a smaller contribution to the fluorescence spectrum [84]. In addition, W220 is protected from oxidation by UV irradiation or N-bromosuccinimide in the presence of the inducer IPTG [57, 85]. The LacI-inducer interaction results in an emission blue shift, apparently due to direct interaction between the indole side chain of W220 and the sugar. The direct correlation of fluorescence shift with IPTG binding emerged as an important tool in many thermodynamic and kinetic studies of LacI and LacI-variant ligand association and allostery (e.g. [44, 53, 54, 86, 87]).

**Chemical modification**

Chemical modification of reactive side chains provided information on LacI recognition, binding and allosteric response (reviewed in [9]). Many of these probes targeted cysteine residues and successfully reported protein interactions upon binding operator, nonspecific DNA, and inducer. The three cysteines in LacI are unreactive with reagents that target solvent-exposed side chains (e.g. iodoacetamide), but modification with apolar probes revealed distinct reactivities and local environments. C281 was predicted to be located at the monomer•monomer subunit interface, C107 was in a relatively apolar environment in the core domain, and the relatively protected location of C140 in the core domain was identified to be in close proximity to the DNA binding domain. Consistent with the allosteric nature of LacI, the environments for C107 and C140 were found to be altered upon inducer binding.
In addition to cysteine modification, chemically modified lysine, histidine, arginine and tyrosine residues generated substantial, detailed functional data for the *lac* repressor protein (reviewed in [9]). Chemical modification of lysine residues using trinitrobenzenesulfonate or dansyl chloride showed that reaction diminished both operator and nonspecific DNA binding; K33, K37 and K108 were protected from chemical modification in the presence of DNA. Patterns of activity lost upon treatment or protection from modification illuminated the roles of other critical residues in DNA binding and function: arginine side chains were targeted using 2,3-butanedione and phenylglyoxal, histidine with diethyl pyrocarbonate, and tyrosine residues via tetranitromethane or iodination. Chemical studies clearly demonstrated structural shifts in the presence of inducer. $^1$H NMR studies using selectively deuterated tyrosine [88] and $^{19}$F NMR spectra of 3-fluorotyrosine-substituted LacI [89] provided further evidence for an allosteric response that engaged significant portions of the protein.

**Biophysical analysis of LacI mutational variants**

A number of LacI mutations and their functional effects were identified in early studies, but the nature of the amino acid changes was unknown until the advent of facile molecular biological methodology. Once site-directed mutagenesis became routinely available, more than 100 variants were created, the resulting proteins were purified, and functional change was assessed by biochemical/biophysical methods (described in [90]). Prior to the available structures, these data were important for mapping ligand binding sites and protein-protein interfaces and frequently provided satisfying explanations for observed phenotypes [9, 82]. With the published structures, site-directed mutagenesis has been used to test a number of structure/function theories (e.g. [91]), sometimes with outcomes that contradict structural predictions (e.g. [92]). The latter observation illustrates the limitations of utilizing protein structural ‘snapshots’ and led to the application of computational approaches (see below). A residue-specific summary of results from chemical modification and site-specific mutagenesis can be found at http://www.bioc.rice.edu/~ksm/ or http://www.kumc.edu/biochemistry/swintkruse.html.

**Targeted molecular dynamics simulations: the allosteric response in motion**

The observed crystallographic ‘endpoints’ for the LacI functional cycle – the repressed and induced states – cannot provide direct information regarding the dynamic al-

losteric mechanism during the transition between these states. However, relevant motions can be interpolated *in silico* using targeted molecular dynamics simulation (TMD) to predict the most probable atomic-level allosteric routes in LacI [75] (Fig. 3a). The advantage of TMD is that it uses two experimentally determined structures to constrain the simulation, dramatically reducing the number of requisite calculations while retaining the spatial and temporal resolution necessary to reveal relevant motions (e.g. backbone dynamics, side-chain motion, aromatic ring flipping).

The key results predicted by the simulation are shown in Figure 3a. Movements originate asymmetrically in the inducer-binding site near D149 of one monomer and propagate to the adjacent monomer via a network of noncovalent interactions of three interconnected routes [75]. The first changes begin in the inducer binding site of one monomer and propagate through the central $\beta$-sheet of the N-subdomain core to the N-terminal monomer•monomer interface that includes residue K84. At this point, K84 moves out of the otherwise hydrophobic interface, with consequent rearrangements in the second pathway that result in transient $\pi$-stacking between H74 and H74′. This interaction is also part of a third pathway that occurs from rear of the inducer-binding pocket proximal to residue F161 through the bottom of this pocket. The intermediate structures display significant asymmetry within this homodimer. All of these changes reposition the N-subdomains, which in turn alters the requisite contacts to the DNA binding domain and thereby reduces binding affinity. The results from the *in silico* allosteric trajectories are in excellent agreement with the large body of data from biochemical and genetic experiments and also provide hypotheses for new explorations.

## Integration of genetic, biochemical, structural and dynamic data

X-ray crystallographic and biochemical studies were essential to a broader understanding of the general topology and function of the *lac* repressor. As indicated previously, the crystal structures corresponding to the functional modes of induced and repressed states of LacI reveal two distinct protein conformations [60, 61], but not the pathway between them. Comparison shows that inducer binding to LacI changes the orientation of the N-subdomains relative to each other and to the C-subdomain (Fig. 2b), and the C-subdomain serves as a rigid scaffold that maintains the functional oligomeric state. This isomerization alters both the intrasubunit orientations between the subdomains of a given monomer and intersubunit interactions between the N-subdomains at each monomer•monomer interface. Ultimately, the N-subdomain change is propagated to the hinge helices, presumably resulting in their unfolding. Both the hinge and DNA binding domain are predicted by simulation to acquire increased mobility [70, 93]. Neither the DNA binding domain nor the hinge is resolved in the x-ray structures [61], and NMR studies demonstrate significant flexibility [66]. Direct observation of hinge helix unfolding in full-length LacI has proved elusive.

In answer to this deficiency, Kalodimos and colleagues utilized NMR analysis of a truncated LacI variant to decipher the structural consequences of DNA binding on the N-terminal DNA binding domain [68, 69, 94–96]. These studies used a short N-terminal segment that was based on a full-length V52 variant [97, 98]. Formation of a disulfide bond at position 52 allowed the truncated protein to maintain the dimeric form needed for high DNA binding affinity. Experiments generated a detailed, time-resolved account of critical association, recognition and dissociation events. Under high-salt conditions, these NMR studies show that the helix-turn-helix motif is folded, whereas the hinge region remains unstructured in the free state or when bound to nonspecific DNA (Fig. 2d). Furthermore, the considerable flexibility of the DNA binding domain in the presence of non-operator DNA would allow the repressor to scan the DNA *in vivo* by translocation in a single dimension, facilitating rapid target location, which may be a critical component of its biological function. Once the specific DNA site is recognized, the hinge helix simultaneously forms and interacts with the minor groove of the nearly palindromic operator ($O^1$) [68, 69].

A subset of residues in the hinge helix directly interact with the core domain of the same monomer, and others make significant interactions with the partner core domain [61]. Inserting glycine residues between the hinge helix and the core domain progressively reduced affinity for the operator and simultaneously diminished allosteric response to the inducer [99, 100]. Disruption of DNA binding may arise from altered spacing of the DNA binding domains within a given dimer, increased entropic cost of binding, distortion of the DNA binding domain by core interaction or decreased structural capacity for the hinge helix formation required for binding in the minor groove.

Specific interactions between the hinge helices and between the N-subdomain and the hinge helix are also important for allosteric signaling. The site of closest approach between the hinge helices is V52-V52′. Variants at this position have a variety of effects on function [101], and V52C formed the basis for creating a disulfide-linked DNA binding domain for NMR studies (see above) [97]. Interestingly, this covalent interaction impedes the allosteric response of the full-length protein for both wild-type operator and its symmetric variant, $O^{sym}$. However, V52C-oxidized and other V52 variants are responsive to inducer for other operator variants, indicating that DNA sequence plays an active role in determining LacI allosteric properties [98, 101].

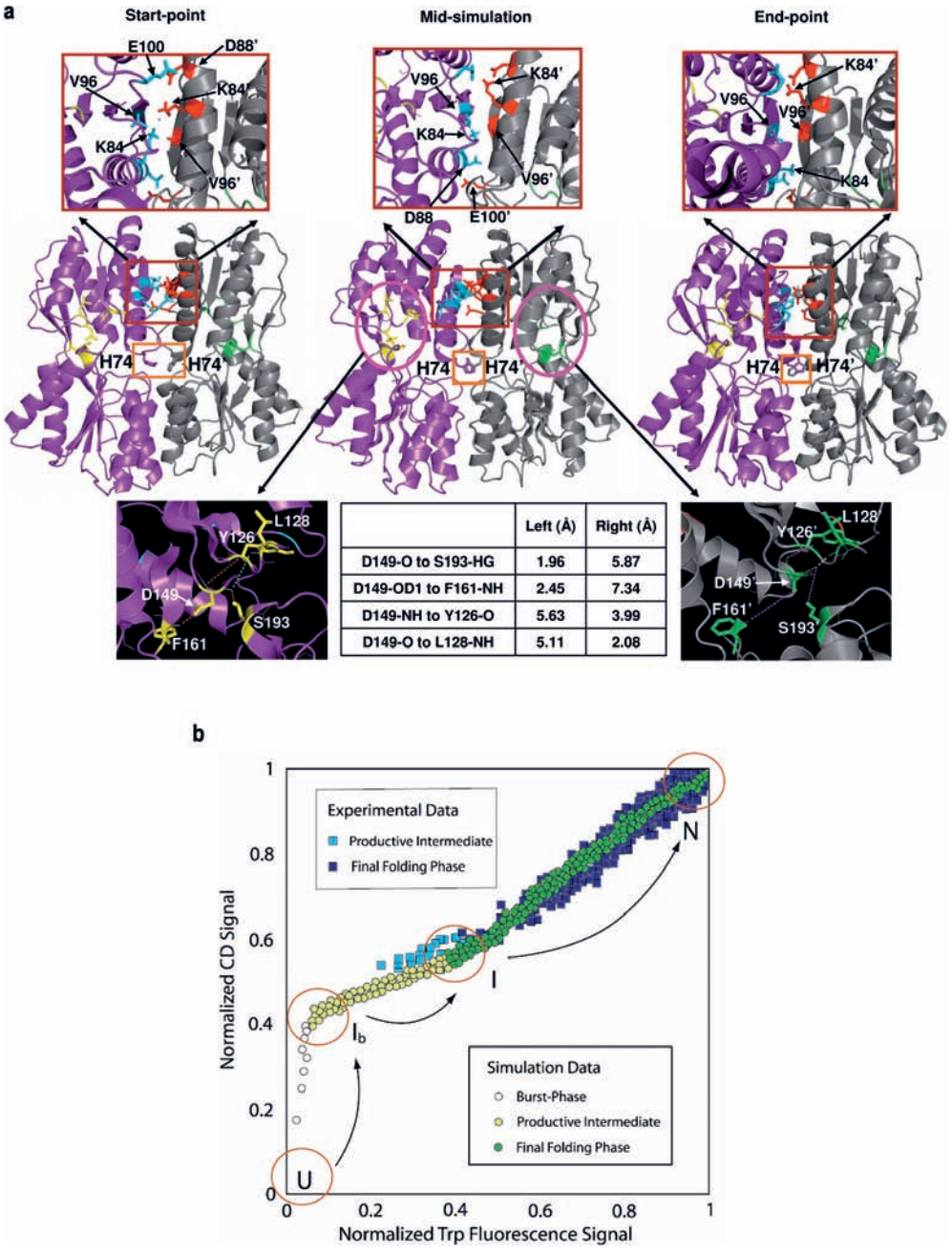


**Figure 3.** Kinetic features of allostery and folding. (*a*) Representative structures from the allosteric trajectories predicted by targeted molecular dynamics simulations [75]: The core domain dimer structure is shown for the starting structure, a mid-point in the simulation, and the target end-point structure. Three areas are highlighted. (i) The region near the inducer site that encompasses D149, S193 and F161 is circled in pink in the middle structure and is enlarged for each monomer below this mid-simulation structure. Note the differences in interatomic distances for D149–S193, D149–F161, D149–Y126 and D149–L128 between the two monomers within a dimer, demonstrating the asymmetry within this homodimer. (ii) The N-subdomain monomer • monomer interface encompasses K84 from both subunits (red box). The detailed, close-up structures above the dimer structures show how K84 positions change during the allosteric transition. Note that at the start-point both K84 side chains are buried in the interface; at the end-point, both are solvent-exposed and interact with negatively charged side chains. In mid-simulation, the K84′ from the 'trigger' monomer (right) has moved to the exposed position, whereas K84 in the 'response' monomer (left) remains buried. (iii) His 74 is located at the monomer • monomer interface near of the inducer binding site (orange box). These residues are shifted and alter interactions with other amino acids between the start- and end-point structures. At mid-simulation the imidazole moieties of these residues are $\pi$-stacked. (*b*) Kinetic features of monomer folding. Normalized CD signals are plotted as a function of fluorescence signals determined under the same experimental conditions. Open circles correspond to the signal theoretically observed in the burst phase, which is within the deadtime of the experimental determinations and therefore not detected; the yellow circles and light blue squares correspond to the theoretical and experimental accumulation of the productive intermediate, respectively; finally, the green circles and dark blue squares represent the final folding phase from theoretical and experimental results, respectively. U, unfolded protein; Ib, burst phase, non-productive intermediate; I, productive intermediate; N, native, folded protein.

Shifting to the core domain, one of the salient allosteric residues is K84. The importance of this residue was first flagged in mutagenesis studies [87]; TMD simulations provided a new context for reinterpreting the data and formulating new hypotheses [75]. In the repressed state, the 4-aminobutyl side chain of this charged residue is buried in the N-subdomain's relatively hydrophobic interface, forming contacts with the carbonyl moiety of V94 and V96′ of the adjacent monomer (Fig. 3a). Upon induction, this basic residue is partially solvent-exposed outside the interface and forms extensive electrostatic interactions with residues E100′ and D88 [61]. Substitutions at position 84 have a profound effect on allosteric induction (phenotype $I^s$) or DNA binding (phenotype $I^-$) [82]. K84L and K84A substitutions diminish allosteric response and affect inducer binding kinetics. The latter are biphasic and several orders of magnitude slower than those for wild-type protein [86, 87]. The slowed rates allowed operator release to be correlated with binding of the second inducer molecule to the dimeric form of the protein [86]. This result is consistent with the TMD prediction that the allosteric conformational change is asymmetric: the biphasic inducer binding kinetics for the K84 variants may provide a direct observation of a fleeting allosteric intermediate state detected by the targeted molecular dynamics simulation [75].

## LacI as a model system for protein folding and function

### The folding landscape of LacI

*In vitro* and *in silico* studies show that small single-domain proteins (<100 residues), often fold by two-state equilibrium and kinetic mechanisms (reviewed in [102, 103]). Recent experimental results and theoretical studies suggest that the mechanism and rate for most small, fast-folding proteins is strongly determined by the native state topology, as opposed to fine details of the polypeptide sequence (reviewed in [102, 103]). Many larger proteins (>100 residues) do not follow two-state mechanisms, but populate intermediate structures in their kinetic folding reactions [102, 103]. Moreover, the largest proteins (>200 residues) are typically composed of multiple domains or subdomains, as is the case for LacI (e.g. [104]). A comprehensive assessment of complex folding behaviors associated with large, topologically intricate proteins like LacI can likely be achieved through the combination of experimental and theoretical/computational approaches [105, 106].

### Thermodynamic analysis of tetrameric LacI folding

The first experimental studies of LacI protein folding used fluorescence and circular dichroism spectroscopy, analytical ultracentrifugation and functional activities at varied urea concentration to show that folding occurs in a single cooperative transition with no evidence of populated intermediates [107]. The absence of the concentration dependence anticipated for tetramer dissociation/unfolding indicated that disruption of the monomer•monomer interface and monomer unfolding are concerted reactions that occur prior to the dissociation of the dimer•dimer interface. The extremely high stability of K84L and K84A allowed dissection of the binding energy for the tetramerization domain from the energy of monomer folding and for the formation of the monomer•monomer interface [107, 108]. An overall free energy change of ~250 kJ $mol^{-1}$ was calculated for dissociation of all interfaces and unfolding of the tetrameric *lac* repressor, reflecting the exceptional stability of this protein. However, a detailed understanding of the folding mechanism/rate was not feasible for tetrameric LacI.

### Comparison of theory and experiment for folding of monomeric LacI

To secure both theoretical [105] and experimental [106] analysis of the folding landscape for a large protein with multiple subdomains, a monomeric variant of LacI was developed through the elimination of eleven C-terminal residues of the tetramerization domain [44] in conjunction with the L251A point mutation that destabilizes the monomer•monomer interface [109]. Monitoring fluorescence and circular dichroism signals (Fig. 3b), the experimentally determined folding mechanism for this LacI monomer involves an off-path burst-phase engaging the transient formation of structures that must unfold. The latter is presumed to reflect a misfolded state that must unfold in order to form a productive intermediate prior to formation of the fully folded protein. The conformations generated in the theoretical pathway were analyzed to estimate circular dichroism and fluorescence signals for comparison with experimental data. In an unprecedented correlation, the key features of the probable experimental route, including the mechanistic complexity and relative spectroscopic signal changes, are nearly identical to those captured by a simple topology-based $C_\alpha$ model [110] (Fig. 3b). The correspondence of simulation and experimental results confirms the folding mechanism of monomeric LacI with a level of confidence that could not be achieved by either approach alone. The combination allows structural interpretation of intermediates observed, shows the presence of long-range interactions in the unproductive intermediate and confirms experimentally the states predicted in the simulation.

### VFT structural comparisons – principles for protein design?

A number of bacterial repressor proteins, known as the LacI/GalR family, have moderate sequence identity (usu-

ally between 15 and 35%), have high structural similarity and exhibit homology in their target DNA sites [111]. LacI/GalR family members bind and are responsive to a variety of chemical effectors, including purines and a wide array of sugars. The monomer core domain of this family shares common ancestry with the periplasmic sugar binding proteins (PBPs), a widely distributed bacterial protein superfamily that mediates chemotaxis and the uptake of nutrients [71, 72, 112]. In both families, ligand binding facilitates closure of the two subdomains, with a motion that resembles a Venus flytrap plant (see the LacI core domain structure in Fig. 2). The structural and functional diversity of the PBP superfamily has allowed the engineering of systems that make use of their distinctive conformational changes, effector site adaptability, and enzymatic activity or biological properties (e.g. [113–116]). The VFT fold has also been identified in domains of several eukaryotic receptors that contain extracellular ligand binding domains – including glutamate-gated ion channels, G-protein-coupled receptors and atrial natriuretic peptide guanylate cyclase receptors – potentially generated by fusions with membrane-bound proteins during the process of evolution [72, 117].

Comparison of these protein families can yield insight into structural differences key to their unique functions. For example, a salient difference between the repressor proteins/receptor proteins and the PBPs is oligomeric state: all of the PBPs are monomeric. The ability of the LacI/GalR family to form dimeric or higher-order oligomers is crucial for high specificity in DNA binding. To elucidate the evolutionary determinants of assembly, a structure-based sequence alignment of several repressors and periplasmic binding proteins and a series of contact maps in network representation were generated for the repressor interfaces [76, 77]. These analyses revealed that the primary sequences corresponding to the interfaces of LacI/GalR family C-subdomains differ significantly from the monomeric PBPs in the region around LacI residues 281 and 282, sites in LacI known to be important in assembly. The secondary and tertiary structures of this region are remarkably conserved, but in the repressor family the side chains are largely hydrophobic, whereas the PBPs have charged residues. The ability to move easily between dimer and monomer was demonstrated by T41 monomeric LacI, demonstrated to be due to the charged point mutation Y282D [118, 119]. *In vitro* characterization of this monomer variant demonstrated that the protein was folded with inducer binding properties comparable to the wild-type repressor [120]. Interestingly, dimerization of the Y282D variant could be compensated by a second-site mutation (K84L or K84A) in the N-subdomain [108], and 22 additional compensatory second-site mutations were identified by phenotypic screening for variants with restored repression function [76]. The large number and diversity of second site revertants suggests that oligomerization of the PBP-like protein (LacI monomer) is robust, with many different scenarios for assembly. In a larger context, this work demonstrates that proteins have evolved to have sequences that are robust – plastic – and achieve their native and properly assembled state with few traps arising from discordant energetic signals [121].

One other feature that must have evolved is the ability of the Venus flytrap domain to interact with other domains. In the LacI/GalR family, the interface between the DNA binding domain and the core domain is mediated by the hinge region. Bioinformatics analysis coupled with structural analysis and mutational effects on function indicate that this interface is unique for each family member [70]. Thus, domain boundaries/interfaces provide another site for modulating protein function ([122], L. Swint-Kruse et al., unpublished).

## Future thoughts

Since the discovery of LacI, this genetic regulatory protein has been routinely utilized for negative, inducible control of gene expression, as a module in regulatory circuit design (e.g. [123]), as a mechanism to control expression in eukaryotic systems *in vivo* (e.g. [124–126]), as a means of purifying selected DNAs (e.g. [127]), as a means to link DNA-containing structures via operator sequences (e.g. [128, 129]), and as a way to generate conditional gene silencing in complex systems [130]. Despite this diversity, only the DNA binding capacity of LacI has been utilized. The next generation of applications will build on recent insights into the structure, function, folding, assembly and allosteric properties that can be utilized for de novo design, for redesign of existing proteins for prescribed properties or for extrapolation to other systems that bear significant structural similarity. Indeed, the information amassed for LacI provides an excellent basis for the development of novel biological functions (e.g. a protein that binds DNA and responds to a novel ligand) that exploit our intricate knowledge of this prototypic genetic regulatory protein for functional adaptation and protein engineering of VFT and LacI/GalR family members.

*Acknowledgements.* Work performed in the laboratories of the authors was supported by the NIH (GM22441 to K. S. M.; L. S. K. is supported by P20 RR 17708) and the Robert A. Welch Foundation (C-576 to K. S. M.); C. J. W. was a trainee of the Houston Area Molecular Biophysics Training Program (NIH GM08280) in the Keck Center for Interdisciplinary Bioscience Training.

*Note.* Where necessary reviews are cited to provide access to the primary literature.

1 Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356.
2 Monod, J., Changeux J.-P. and Jacob, F. (1963) Allosteric proteins and cellular control systems. J. Mol. Biol. 6, 306–329.

3 Gilbert, W. and Müller-Hill, B. (1966) Isolation of the Lac repressor. Proc. Natl. Acad. Sci. USA 56, 1891–1898.
4 Riggs, A. D. and Bourgeois, S. (1968) On the assay, isolation, and characterization of the *lac* repressor. J. Mol. Biol. 34, 361–364.
5 Gilbert, W. and Müller-Hill, B. (1967) The Lac operator is DNA. Proc. Natl. Acad. Sci. USA 58, 2415–2421.
6 Riggs, A. D., Bourgeois, S., Newby, R. F. and Cohn, M. (1968) DNA binding of the *lac* repressor. J. Mol. Biol. 34, 365–368.
7 Riggs, A. D., Newby, R. F. and Bourgeois, S. (1970) *lac* repressor-operator interaction. II. Effect of galactosides and other ligands. J. Mol. Biol. 51, 303–314.
8 Jobe, A. and Bourgeois, S. (1972) *lac* repressor-operator interaction. VI. The natural inducer of the *lac* operon. J. Mol. Biol. 69, 397–408.
9 Matthews, K. S. and Nichols, J. C. (1998) Lactose repressor protein: functional properties and structure. Prog. Nucl. Acid Res. Mol. Biol. 58, 127–164.
10 Lewis, M. (2005) The *lac* repressor. C. R. Biol. 328, 521–548.
11 Miller, J. H. and Reznikoff, W. S. (1980) The Operon, 2nd Edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
12 Müller-Hill, B. (1996) The Lac operon: A Short History of a Genetic Paradigm, Walter de Gruyter, Berlin.
13 Bell, C. E. and Lewis, M. (2001) The Lac repressor: a second generation of structural and functional studies. Curr. Opin. Struct. Biol. 11, 19–25.
14 Gilbert, W. and Maxam, A. (1973) The nucleotide sequence of the *lac* operator. Proc. Natl. Acad. Sci. USA 70, 3581–3584.
15 Riggs, A. D., Suzuki, H. and Bourgeois, S. (1970) *lac* repressor-operator interaction. I. Equilibrium studies. J. Mol. Biol. 48, 67–83.
16 Schlax, P. J., Capp, M. W. and Record, M. T. Jr (1995) Inhibition of transcription initiation by *lac* repressor. J. Mol. Biol. 245, 331–350.
17 Lee, J. and Goldfarb, A. (1991) *lac* repressor acts by modifying the initial transcribing complex so that it cannot leave the promoter. Cell 66, 793–798.
18 Straney, S. B. and Crothers, D. M. (1987) Lac repressor is a transient gene-activating protein. Cell 51, 699–707.
19 Miller, J. H., Calos, M. P., Galas, D., Hofer, M., Büchel, D. E. and Müller-Hill, B. (1980) Genetic analysis of transpositions in the *lac* region of *Escherichia coli*. J. Mol. Biol. 144, 1–18.
20 Roderick, S. L. (2005) The *lac* operon galactoside acetyltransferase. C. R. Biol. 328, 568–575.
21 Guan, L. and Kaback, H. R. (2006) Lessons from lactose permease. Annu. Rev. Biophys. Biomol. Struct. 35, 67–91.
22 Juers, D. H., Heightman, T. D., Vasella, A., McCarter, J. D., Mackenzie, L., Withers, S. G. and Matthews, B. W. (2001) A structural view of the action of *Escherichia coli* (*lac*Z) beta-galactosidase. Biochemistry 40, 14781–14794.
23 Barkley, M. D., Riggs, A. D., Jobe, A. and Bourgeois, S. (1975) Interaction of effecting ligands with *lac* repressor and repressor-operator complex. Biochemistry 14, 1700–1712.
24 Kolb, A., Busby, S., Buc, H., Garges, S. and Adhya, S. (1993) Transcriptional regulation by cAMP and its receptor protein. Annu. Rev. Biochem. 62, 749–795.
25 Emmer, M., deCrombrugghe, B., Pastan, I. and Perlman, R. (1970) Cyclic AMP receptor protein of *E. coli*: its role in the synthesis of inducible enzymes. Proc. Natl. Acad. Sci. USA 66, 480–487.
26 Reznikoff, W. S., Winter, R. B. and Hurley, C. K. (1974) The location of the repressor binding sites in the *lac* operon. Proc. Natl. Acad. Sci. USA 71, 2314–2318.
27 Pfahl, M., Gulde, V. and Bourgeois, S. (1979) 'Second' and 'third operator' of the *lac* operon: an investigation of their role in the regulatory mechanism. J. Mol. Biol. 127, 339–344.
28 Mossing, M. C. and Record, M. T. Jr (1986) Upstream operators enhance repression of the *lac* promoter. Science 233, 889–892.
29 Besse, M., von Wilcken-Bergmann, B. and Müller-Hill, B. (1986) Synthetic *lac* operator mediates repression through *lac* repressor when introduced upstream and downstream from *lac* promoter. EMBO J. 5, 1377–1381.
30 Brenowitz, M., Pickar, A. and Jamison, E. (1991) Stability of a Lac repressor mediated 'looped complex'. Biochemistry 30, 5986–5998.
31 Borowiec, J. A., Zhang, L., Sasse-Dwight, S. and Gralla, J. D. (1987) DNA supercoiling promotes formation of a bent repression loop in *lac* DNA. J. Mol. Biol. 196, 101–111.
32 Becker, N. A., Kahn, J. D. and Maher, L. J., 3rd (2005) Bacterial repression loops require enhanced DNA flexibility. J. Mol. Biol. 349, 716–730.
33 Whitson, P. A., Hsieh, W.-T., Wells, R. D. and Matthews, K. S. (1987) Influence of supercoiling and sequence context on operator DNA binding with *lac* repressor. J. Biol. Chem. 262, 14592–14599.
34 Hsieh, W.-T., Whitson, P. A., Matthews, K. S. and Wells, R. D. (1987) Influence of sequence and distance between two operators on interaction with the *lac* repressor. J. Biol. Chem. 262, 14583–14591.
35 Krämer, H., Amouyal, M., Nordheim, A. and Müller-Hill, B. (1988) DNA supercoiling changes the spacing requirement of two *lac* operators for DNA loop formation with *lac* repressor. EMBO J. 7, 547–556.
36 Krämer, H., Niemöller, M., Amouyal, M., Revet, B., von Wilcken-Bergmann, B. and Müller-Hill, B. (1987) *lac* repressor forms loops with linear DNA carrying two suitably spaced *lac* operators. EMBO J. 6, 1481–1491.
37 Levandoski, M. M., Tsodikov, O. V., Frank, D. E., Melcher, S. E., Saecker, R. M. and Record, M. T. Jr (1996) Cooperative and anticooperative effects in binding of the first and second plasmid $O^{sym}$ operators to a LacI tetramer: evidence for contributions of non-operator DNA binding by wrapping and looping. J. Mol. Biol. 260, 697–717.
38 Eismann, E. R. and Müller-Hill, B. (1990) *lac* repressor forms stable loops *in vitro* with supercoiled wild-type *lac* DNA containing all three natural *lac* operators. J. Mol. Biol. 213, 763–765.
39 Betz, J. L., Sasmor, H. M., Buck, F., Insley, M. Y. and Caruthers, M. H. (1986) Base substitution mutants of the *lac* operator: *in vivo* and *in vitro* affinities for *lac* repressor. Gene 50, 123–132.
40 Sadler, J. R., Sasmor, H. and Betz, J. L. (1983) A perfectly symmetric *lac* operator binds the *lac* repressor very tightly. Proc. Natl. Acad. Sci. USA 80, 6785–6789.
41 Simons, A., Tils, D., von Wilcken-Bergmann, B. and Müller-Hill, B. (1984) Possible ideal *lac* operator: *Escherichia coli lac* operator-like sequences from eukaryotic genomes lack the central G-C pair. Proc. Natl. Acad. Sci. USA 81, 1624–1628.
42 Platt, T., Files, J. G. and Weber, K. (1973) *Lac* repressor: specific proteolytic destruction of the $NH_2$-terminal region and loss of the deoxyribonucleic acid binding activity. J. Biol. Chem. 248, 110–121.
43 Ogata, R. T. and Gilbert, W. (1978) An amino-terminal fragment of *lac* repressor binds specifically to *lac* operator. Proc. Natl. Acad. Sci. USA 75, 5851–5854.
44 Chen, J. and Matthews, K. S. (1992) Deletion of lactose repressor carboxyl-terminal domain affects tetramer formation. J. Biol. Chem. 267, 13843–13850.
45 Alberti, S., Oehler, S., von Wilcken-Bergmann, B. and Müller-Hill, B. (1993) Genetic analysis of the leucine heptad repeats of Lac repressor: evidence for a 4-helical bundle. EMBO J. 12, 3227–3236.
46 Monod, J., Wyman, J. and Changeux, J. P. (1965) On the nature of allosteric transitions: a plausible model. J. Mol. Biol. 12, 88–118.
47 Koshland, D. E. Jr, Nemethy, G. and Filmer, D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5, 365–385.

48 Swain, J. F. and Gierasch, L. M. (2006) The changing landscape of protein allostery. Curr. Opin. Struct. Biol. 16, 102–108.
49 Kern, D. and Zuiderweg, E. R. P. (2003) The role of dynamics in allosteric regulation. Curr. Opin. Struct. Biol. 13, 748–757.
50 Gunasekaran, K., Ma, B. and Nussinov, R. (2004) Is allostery an intrinsic property of *all* dynamic proteins? Proteins 57, 433–443.
51 Daly, T. J. and Matthews, K. S. (1986) Allosteric regulation of inducer and operator binding to the lactose repressor. Biochemistry 25, 5479–5484.
52 Dunaway, M., Olson, J. S., Rosenberg, J. M., Kallai, O. B., Dickerson, R. E. and Matthews, K. S. (1980) Kinetic studies of inducer binding to *lac* repressor • operator complex. J. Biol. Chem. 255, 10115–10119.
53 Friedman, B. E., Olson, J. S. and Matthews, K. S. (1977) Interaction of *lac* repressor with inducer: kinetic and equilibrium measurements. J. Mol. Biol. 111, 27–39.
54 Laiken, S. L., Gross, C. A. and von Hippel, P. H. (1972) Equilibrium and kinetic studies of *Escherichia coli lac* repressor-inducer interactions. J. Mol. Biol. 66, 143–155.
55 Ohshima, Y., Matsuura, M. and Horiuchi, T. (1972) Conformational change of the *lac* repressor induced with the inducer. Biochem. Biophys. Res. Commun. 47, 1444–1450.
56 Matthews, K. S. (1974) Ultraviolet difference spectra of the lactose repressor protein. II. Trypsin core protein. Biochim. Biophys. Acta 359, 334–340.
57 O'Gorman, R. B. and Matthews, K. S. (1977) Fluorescence and ultraviolet spectral studies of *lac* repressor modified with N-bromosuccinimide. J. Biol. Chem. 252, 3572–3577.
58 Pace, H. C., Lu, P. and Lewis, M. (1990) *lac* repressor: crystallization of intact tetramer and its complexes with inducer and operator DNA. Proc. Natl. Acad. Sci. USA 87, 1870–1873.
59 Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. and van Gunsteren, W. F. (1985) A protein structure from nuclear magnetic resonance data: *lac* repressor headpiece. J. Mol. Biol. 182, 179–182.
60 Friedman, A. M., Fischmann, T. O. and Steitz, T. A. (1995) Crystal structure of *lac* repressor core tetramer and its implications for DNA looping. Science 268, 1721–1727.
61 Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. and Lu, P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science 271, 1247–1254.
62 Bell, C. E. and Lewis, M. (2000) A closer view of the conformation of the Lac repressor bound to operator. Nat. Struct. Biol. 7, 209–214.
63 Bell, C. E. and Lewis, M. (2001) Crystallographic analysis of Lac repressor bound to natural operator *O1*. J. Mol. Biol. 312, 921–926.
64 Bell, C. E., Barry, J., Matthews, K. S. and Lewis, M. (2001) Structure of a variant of *lac* repressor with increased thermostability and decreased affinity for operator. J. Mol. Biol. 313, 99–109.
65 Slijper, M., Bonvin, A. M. J. J., Boelens, R. and Kaptein, R. (1996) Refined structure of *lac* repressor headpiece (1–56) determined by relaxation matrix calculations from 2D and 3D NOE data: change of tertiary structure upon binding to the *lac* operator. J. Mol. Biol. 259, 761–773.
66 Wade-Jardetzky, N., Bray, R. P., Conover, W. W., Jardetzky, O., Geisler, N. and Weber, K. (1979) Differential mobility of the N-terminal headpiece in the *lac*-repressor protein. J. Mol. Biol. 128, 259–264.
67 Spronk, C. A. E. M., Slijper, M., van Boom, J. H., Kaptein, R. and Boelens, R. (1996) Formation of the hinge helix in the *lac* repressor is induced upon binding to the *lac* operator. Nat. Struct. Biol. 3, 916–919.
68 Kalodimos, C. G., Boelens, R. and Kaptein, R. (2002) A residue-specific view of the association and dissociation pathway in protein-DNA recognition. Nat. Struct. Biol. 9, 193–197.
69 Kalodimos, C. G., Biris, N., Bonvin, A. M. J. J., Levandoski, M. M., Guennuegues, M., Boelens, R. and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 305, 386–389.
70 Swint-Kruse, L., Larson, C., Pettitt, B. M. and Matthews, K. S. (2002) Fine-tuning function: correlation of hinge domain interactions with functional distinctions between LacI and PurR. Protein Sci. 11, 778–794.
71 Tam, R. and Saier, M. H. Jr (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiol. Rev. 57, 320–346.
72 Felder, C. B., Graul, R. C., Lee, A. Y., Merkle, H. P. and Sadee, W. (1999) The Venus flytrap of periplasmic binding proteins: an ancient protein module present in multiple drug receptors. AAPS PharmSci. 1: E2.
73 Chen, J. and Matthews, K. S. (1994) Subunit dissociation affects DNA binding in a dimeric *lac* repressor produced by C-terminal deletion. Biochemistry 33, 8728–8735.
74 Chen, J., Alberti, S. and Matthews, K. S. (1994) Wild-type operator binding and altered cooperativity for inducer binding of *lac* repressor dimer mutant R3. J. Biol. Chem. 269, 12482–12487.
75 Flynn, T. C., Swint-Kruse, L., Kong, Y., Booth, C., Matthews, K. S. and Ma, J. (2003) Allosteric transition pathways in the lactose repressor protein core domains: asymmetric motions in a homodimer. Protein Sci. 12, 2523–2541.
76 Swint-Kruse, L., Elam, C. R., Lin, J. W., Wycuff, D. R. and Matthews, K. S. (2001) Plasticity of quaternary structure: twenty-two ways to form a LacI dimer. Protein Sci. 10, 262–276.
77 Swint-Kruse, L. (2004) Using networks to identify fine structural differences between functionally distinct protein states. Biochemistry 43, 10886–10895.
78 Villa, E., Balaeff, A. and Schulten, K. (2005) Structural dynamics of the *lac* repressor-DNA complex revealed by a multiscale simulation. Proc. Natl. Acad. Sci. USA 102, 6783–6788.
79 Swigon, D., Coleman, B. D. and Olson, W. K. (2006) Modeling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation. Proc. Natl. Acad. Sci. USA 103, 9879–9884.
80 Miller, J. H. (1984) Genetic studies of *lac* repressor. XII. Amino acid replacements in the DNA binding domain of the *Escherichia coli lac* repressor. J. Mol. Biol. 180, 205–212.
81 Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S. and Miller, J. H. (1994) Genetic studies of the *lac* repressor. XIV. Analysis of 4000 altered *Escherichia coli lac* repressors reveals essential and non-essential residues, as well as 'spacers' which do not require a specific sequence. J. Mol. Biol. 240, 421–433.
82 Suckow, J., Markiewicz, P., Kleina, L. G., Miller, J., Kisters-Woike, B. and Müller-Hill, B. (1996) Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. J. Mol. Biol. 261, 509–523.
83 Gordon, A. J. E., Burns, P. A., Fix, D. F., Yatagai, F., Allen, F. L., Horsfall, M. J., Halliday, J. A., Gray, J., Bernelot-Moens, C. and Glickman, B. W. (1988) Missense mutation in the *lacI* gene of *Escherichia coli*: inferences on the structure of the repressor protein. J. Mol. Biol. 200, 239–251.
84 Royer, C. A., Gardner, J. A., Beechem, J. M., Brochon J.-C. and Matthews, K. S. (1990) Resolution of the fluorescence decay of the two tryptophan residues of *lac* repressor using single tryptophan mutants. Biophys. J. 58, 363–378.
85 Spodheim-Maurizot, M., Charlier, M. and Helene, C. (1985) Photochemical modifications of *lac* repressor: effect of effectors binding on tryptophan photooxidation. Photochem. Photobiol. 42, 353–359.
86 Swint-Kruse, L., Zhan, H. and Matthews, K. S. (2005) Integrated insights from simulation, experiment, and mutational

analysis yield new details of LacI function. Biochemistry 44, 11201–11213.
87 Chang W.-I., Olson, J. S. and Matthews, K. S. (1993) Lysine 84 is at the subunit interface of *lac* repressor protein. J. Biol. Chem. 268, 17613–17622.
88 Matthews, K. S., Wade-Jardetzky, N. G., Graber, M., Conover, W. W. and Jardetzky, O. (1977) High resolution 1H NMR of a selectively deuterated analog of the *lac* repressor. Biochim. Biophys. Acta 490, 534–538.
89 Jarema, M. A. C., Lu, P. and Miller, J. H. (1981) Genetic assignment of resonances in the NMR spectrum of a protein: *lac* repressor. Proc. Natl. Acad. Sci. USA 78, 2707–2711.
90 Swint-Kruse, L. and Matthews, K. S. (2004) Thermodynamics, protein modification, and molecular dynamics in characterizing lactose repressor protein: strategies for complex analyses of protein structure-function. Methods Enzymol. 379, 188–209.
91 Swint-Kruse, L., Zhan, H., Fairbanks, B. M., Maheshwari, A. and Matthews, K. S. (2003) Perturbation from a distance: mutations that alter LacI function through long-range effects. Biochemistry 42, 14004–14016.
92 Barry, J. K. and Matthews, K. S. (1999) Substitutions at histidine 74 and aspartate 278 alter ligand binding and allostery in lactose repressor protein. Biochemistry 38, 3579–3590.
93 Swint-Kruse, L., Matthews, K. S., Smith, P. E. and Pettitt, B. M. (1998) Comparison of simulated and experimentally determined dynamics for a variant of the LacI DNA binding domain, Nlac-P. Biophys. J. 74, 413–421.
94 Kalodimos, C. G., Boelens, R. and Kaptein, R. (2004) Toward an integrated model of protein-DNA recognition as inferred from NMR studies on the *lac* repressor system. Chem. Rev. 104, 3567–3586.
95 Kalodimos, C. G., Bonvin, A. M. J. J., Salinas, R. K., Wechselberger, R., Boelens, R. and Kaptein, R. (2002) Plasticity in protein-DNA recognition: *lac* repressor interacts with its natural operator *O1* through alternative conformations of its DNA binding domain. EMBO J. 21, 2866–2876.
96 Kalodimos, C. G., Folkers, G. E., Boelens, R. and Kaptein, R. (2001) Strong DNA binding by covalently linked dimeric Lac headpiece: evidence for the crucial role of the hinge helices. Proc. Natl. Acad. Sci. USA 98, 6039–6044.
97 Falcon, C. M., Swint-Kruse, L. and Matthews, K. S. (1997) Designed disulfide between N-terminal domains of lactose repressor disrupts allosteric linkage. J. Biol. Chem. 272, 26818–26821.
98 Falcon, C. M. and Matthews, K. S. (2001) Engineered disulfide linking the hinge regions within lactose repressor dimer increases operator affinity, decreases sequence selectivity, and alters allostery. Biochemistry 40, 15650–15659.
99 Falcon, C. M. and Matthews, K. S. (1999) Glycine insertion in the hinge region of lactose repressor protein alters DNA binding. J. Biol. Chem. 274, 30849–30857.
100 Falcon, C. M. and Matthews, K. S. (2000) Operator DNA sequence variation enhances high affinity binding by hinge helix mutants of lactose repressor protein. Biochemistry 39, 11074–11083.
101 Zhan, H., Swint-Kruse, L. and Matthews, K. S. (2006) Extrinsic interactions dominate helical propensity in coupled binding and folding of the lactose repressor protein hinge helix. Biochemistry 45, 5896–5906.
102 Gillespie, B. and Plaxco, B. W. (2004) Using protein folding rates to test protein folding theories. Annu. Rev. Biochem. 73, 837–859.
103 Matthews, C. R. (1993) Pathways of protein folding. Annu. Rev. Biochem. 62, 653–683.
104 Xu, D. and Nussinov, R. (1997) Favorable domain size in proteins. Fold. Design 3, 11–17.
105 Das, P., Wilson, C. J., Fossati, G., Wittung-Stafshede, P., Matthews, K. S. and Clementi, C. (2005) Characterization of the folding landscape of monomeric lactose repressor: Quantitative comparison of theory and experiment. Proc. Natl. Acad. Sci. USA 102, 14569–14574.
106 Wilson, C. J., Das, P., Clementi, C., Matthews, K. S. and Wittung-Stafshede, P. (2005) The experimental folding landscape of monomeric lactose repressor, a large two-domain protein, involves two kinetic intermediates. Proc. Natl. Acad. Sci. USA 102, 14563–14568.
107 Barry, J. K. and Matthews, K. S. (1999) Thermodynamic analysis of unfolding and dissociation in lactose repressor protein. Biochemistry 38, 6520–6528.
108 Nichols, J. C. and Matthews, K. S. (1997) Combinatorial mutations of *lac* repressor. Stability of monomer-monomer interface is increased by apolar substitution at position 84. J. Biol. Chem. 272, 18550–18557.
109 Dong, F., Spott, S., Zimmermann, O., Kisters-Woike, B., Müller-Hill, B. and Barker, A. (1999) Dimerisation mutants of Lac repressor. I. A monomeric mutant, L251A, that binds *lac* operator DNA as a dimer. J. Mol. Biol. 290, 653–666.
110 Gō, N. (1983) Theoretical studies of protein folding. Annu. Rev. Biophys. Bioeng. 12, 183–210.
111 Weickert, M. J. and Adhya, S. (1992) A family of bacterial regulators homologous to Gal and Lac repressors. J. Biol. Chem. 267, 15869–15874.
112 Mowbray, S. L. and Björkman, A. J. (1999) Conformational changes of ribose-binding protein and two related repressors are tailored to fit the functional need. J. Mol. Biol. 294, 487–499.
113 Hellinga, H. W. and Marvin, J. S. (1998) Protein engineering and the development of genetic biosensors. Trends Biotechnol. 16, 183–189.
114 Marvin, J. S. and Hellinga, H. W. (2001) Conversion of a maltose receptor into a zinc biosensor by computational design. Proc. Natl. Acad. Sci. USA 98, 4955–4960.
115 Bolon, D. N., Voigt, C. A. and Mayo, S. L. (2002) De novo design of biocatalysts. Curr. Opin. Chem. Biol. 6, 125–129.
116 Dwyer, M. A. and Hellinga, H. W. (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering. Curr. Opin. Struct. Biol. 14, 495–504.
117 Baumgartner, J. W., Kim, C., Brissette, R. E., Inouye, M., Park, C. and Hazelbauer, G. L. (1994) Transmembrane signalling by a hybrid protein: communication from the domain of chemoreceptor Trg that recognizes sugar-binding proteins to the kinase/phosphatase domain of osmosensor EnvZ. J. Bacteriol. 176, 1157–1163.
118 Schmitz, A., Schmeissner, U., Miller, J. H. and Lu, P. (1976) Mutations affecting the quaternary structure of the *lac* repressor. J. Biol. Chem. 251, 3359–3366.
119 Chen, J. and Matthews, K. S. (1992) T41 mutation in *lac* repressor is Tyr282 to Asp. Gene 111, 145–146.
120 Daly, T. J. and Matthews, K. S. (1986) Characterization and modification of a monomeric mutant of the lactose repressor protein. Biochemistry 25, 5474–5478.
121 Bryngelson, J. D. and Wolynes, P. G. (1987) Spin glasses and the statistical mechanics of protein folding. Proc. Natl. Acad. Sci. USA 84, 7524–7528.
122 Swint-Kruse, L., Zhan, H., Meinhardt, S. and Tungtur, S. (2006) Functional consequences of domain swapping between the LacI/GalR transcription regulatory proteins. Biophys. J. 90, 346A.
123 Atsumi, S. and Little, J. W. (2004) Regulatory circuit design and evolution using phage lambda. Genes Dev. 18, 2086–2094.
124 Hu, M. C.-T. and Davidson, N. (1987) The inducible *lac* operator-repressor system is functional in mammalian cells. Cell 48, 555–566.
125 Fieck, A., Wyborski, D. L. and Short, J. M. (1992) Modifications of the *E. coli* Lac repressor for expression in eukaryotic cells: effects of nuclear signal sequences on protein activity

and nuclear accumulation. Nucleic Acids Res. 20, 1785–1791.

126 Cronin, C. A., Gluba, W. and Scrable, H. (2001) The *lac* operator-repressor system is functional in the mouse. Genes Dev. 15, 1506–1517.

127 Hasche, A. and Voss, C. (2005) Immobilisation of a repressor protein for binding of plasmid DNA. J. Chromtogr. A 1080, 76–82.

128 Wu H.-Y., Lau, K. and Liu, L. F. (1992) Interlocking of plasmid DNAs due to *lac* repressor-operator interaction. J. Mol. Biol. 228, 1104–1114.

129 Bystricky, K., Laroche, T., van Houwe, G., Blaszczyk, M. and Gasser, S. M. (2005) Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. J. Cell Biol. 168, 375–387.

130 Higuchi, M., Tsutsumi, R., Higashi, H. and Hatakeyama, M. (2004) Conditional gene silencing utilizing the *lac* repressor reveals a role of SHP-2 in *cagA*-positive *Helicobacter pylori* pathogenicity. Cancer Sci. 95, 442–447.