Molecular Flip-Flops Formed by Overlapping Fis Sites

Paul N. Hengen [a,b,e], Ilya G. Lyakhov [a,e],

Lisa E. Stewart [c], and Thomas D. Schneider [d,f,*]

[a] *Intramural Research Support Program, SAIC, NCI Frederick, Frederick, MD, USA*

[b] *Current address: Applied Biosystems, 3833 North First Street, San Jose, CA, 95134-1701, USA*

[c] *Volunteer to the National Cancer Institute at Frederick. Current address: Alkami Biosystems, Inc. P.O. Box 11216, Berkeley, CA 94712-2216, USA*

[d] *Laboratory of Experimental and Computational Biology, National Cancer Institute at Frederick, P. O. Box B, Building 469, Room 144, Frederick, MD 21702-1201, USA, 301) 846-5581 (-5532 for messages), fax: (301) 846-5598.*

[e] *These authors contributed equally.*

[f] *The publisher or recipient acknowledges the right of the U.S. Government to retain a nonexclusive, royalty-free license in and to any copyright covering the article.*

Abstract

The DNA binding protein Fis frequently uses pairs of sites 7 or 11 base pairs apart. Two overlapping Fis sites separated by 11 base pairs are found in the *E. coli* origin of chromosomal replication. Only one of these sites is bound by Fis at a time, so the structure is a molecular flip-flop that could direct alternative firing of replication complexes in opposite directions. Alternatively, the flip-flop could represent part of an on-off switch for replication. Because they can be used to create precise switched states, molecular flip-flops could be used as the basis of a novel molecular computer.

Key words: Fis DNA binding, *oriC*, flip-flop, information theory, sequence walker

version = 2.15 of fff.tex 2003 October 16

running title: Fis Flip-Flops

Introduction

Fis is a well characterized site-specific DNA binding protein which is known to bend DNA and is involved in many site-specific recombination systems [1]. In addition, it autoregulates its own promoter and activates other promoters [2,3]. When *Escherichia coli* encounters a rich nutritional medium, the number of Fis molecules increases from nearly zero to 25,000-50,000 dimers per cell [4]. Estimates of the number of Fis sites in the *E. coli* genome based on the average information in Fis sites give a similar number, indicating that most Fis molecules are controlling genetic systems throughout the genome [5–8].

Information analysis of Fis binding sites and their surrounding sequences has revealed previously unidentified sites adjacent to known ones [5]. We observed that pairs of Fis sites are often separated by 7 or 11 bases in many genetic systems [5,9]. These Fis sites often overlap the binding sites of other proteins in biologically significant places such as the Xis site of λ *att* [10], *dif*, *nrd*, *ndh* and the *fis* promoter (data not shown). To understand the significance of these pairs we sought to determine whether Fis binds cooperatively or antagonistically at the adjacent sites. In this study we show that in artificial DNA constructs, overlapping Fis sites 7 or 11 base pairs apart cannot be bound simultaneously and therefore act as a molecular flip-flop.

DNA replication starts at 84.6 minutes on the circular *E. coli* K-12 chromosome at a locus called *oriC* [11,12]. Bidirectional replication starting at *oriC* [13,14] is completed in the

* Corresponding author.
Email addresses: ilyakhov@ncifcrf.gov (Ilya G. Lyakhov), stewartl@alkami.com (Lisa E. Stewart), toms@al.m.mit.e.. (Thomas D. Schneider).
URL: http://www.lecb.ncifcrf.gov/~toms/ (Thomas D. Schneider).

terminus region half way around the chromosome [15]. Replication is dependent on the DnaA protein, which binds at 5 sites in *oriC*. Using sequence walkers [10], we observed that there are likely to be two Fis sites 11 base pairs apart wedged precisely between two of the DnaA sites. We show that these sites are not bound simultaneously.

Materials and Methods

Sequence analysis programs

Delila system programs were used for handling sequences and information calculations [16–21,10]. Figures were generated automatically from raw GenBank data using Delila and UNIX script programs. Further information is available at http://www.lecb.ncifcrf.gov/~toms/.

Design of Fis binding experiments

Synthetic DNAs containing strong Fis sites separated by 11 and 7 base pairs were designed by selecting from the most frequent bases at each position in the Fis sequence logo [5]. These were then merged with the same sequence shifted by 11 or 7 base pairs by comparing the $R_{iw}(b,l)$ values for various choices. (Note: the consensus sequence of the early model we used was TTTG(G/C)TCAAAATTTGA(G/C)CAAA which differs from that of the logo [22]. The 0 and strongly conserved ± 7 positions are underlined.) Five extra bases were added to the ends based on the natural sequences around the *hin* proximal and medial sites for the overlap 11 oligo, and the sequences around *cin* external and proximal sites were used for the overlap 7 oligo [5]. The DNAs were made self complementary (Fig. 4a, b). Sites separated by 23 bases were created starting with the 11 base separated DNA and duplicating the central overlap region. A *Bam*HI site was also inserted and the DNA was flanked by *Eco*RI sites (Fig. 4c). Oligos were synthesized with biotin on the $5'$ end and gel purified (Oligos Etc., Wilsonville, OR, USA). To ensure thorough annealing, they were heated to 90°C for 10 minutes, and slowly cooled to room temperature. The annealed products were

electrophoresed through an 8% (w/v) polyacrylamide gel, and the bands corresponding to the linear duplex DNA of the correct size were sliced from the gel. DNA was recovered by electroelution and extracted with isoamyl alcohol to remove ethidium bromide. A non-specific control DNA was composed of the two 66 bp *Hin*FI fragments from bacteriophage ϕX174 (Life Technologies, Inc.). This DNA did not shift even at the highest concentration of Fis used. Gel mobility shift experiments were performed as described previously, using chemiluminescent detection [5]. Fis used in the experiment shown in Fig. 5 was a gift from Reid Johnson.

Hairpin DNA oligos containing the *oriC* region (Fig. 7) were synthesized with a 5′-tetramethylrhodamine modification and gel purified (Oligos Etc.). By using a hairpin, an annealing step is not required and the oligos are exactly equimolar [5]. Horizontal 8% PAGE was used for the gel mobility shift assay and band positions were visualized with a FMBIO II fluorescent scanner (Hitachi) with an excitation wavelength of 532 nm and detection at 605 nm.

Fis protein isolation

For the experiment shown in Fig. 7, Epicurian Coli® BL21-Gold(DE3) cells (Stratagene) [23], transformed with pRJ1077 plasmid (a gift from Reid Johnson) were used for Fis protein isolation by modification of a previous method [24]. A 10 ml overnight culture was added to 1 L of SOB media, shaken at 37°C for 2 hours 30 minutes, induced by 1 μM IPTG and shaken again for 1 hour. The culture was centrifuged and the cell pellet was resuspended in 50 ml FLBC buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA), containing 200 mM NaCl and 0.1 mM PMSF, sonicated with a Cole-Parmer Ultrasonic Processor model CP 70T 3 times for 10 minutes (40W) on ice and centrifuged at 2000 g for 20 minutes. The supernatant was loaded on a 10 ml Q-Sepharose® Fast Flow (Amersham Pharmacia Biotech) column equilibrated with the same buffer and washed with 10 ml of FLBC buffer containing 300

mM NaCl. Flow-through and wash fractions were combined and the mix was loaded on a 10 mL Heparin Sepharose® CL-6B (Amersham Pharmacia Biotech), washed with 50 ml of FLBC buffer, containing 300 mM NaCl and eluted with a 100-ml 300-2000 mM NaCl gradient of FLBC buffer. Fractions were analyzed by SDS-electrophoresis.

Molecular Models of Fis Binding to DNA

In order to understand the sequence logo for Fis binding sites [5], we constructed three dimensional models of Fis-DNA binding. The models of Fis interacting with DNA were built using Insight II software from Biosym Technologies, Inc., on an IRIS computer (Silicon Graphics, Inc.), and displayed with RasMol 2.5, available at
http://molbiol.soton.ac.uk/rasmol.html
or
ftp://ftp.dcs.ed.ac.uk/pub/rasmol/ .
The Fis protein coordinates are those of the Protein Data Bank
(http://www.rcsb.org/pdb/) entry 1fia .

In Fis-DNA models built by earlier workers either B-form DNA with no particular base sequence was smoothly bent and placed next to the X-ray crystal model of Fis [25,26], or the single distal *hin* Fis binding site (V01370, 180, 8.9 bits) was used and kinks added to the DNA at positions based on the DNase I hypersensitive pattern [27]. Since the sequence logo gives a detailed model for the sequence conservation at Fis sites, we used this information to construct three models:

- a single Fis homodimer binding to the consensus sequence as determined by the logo (Fig. 3A).
- two Fis homodimers binding to two Fis sites spaced 11 base pairs apart on DNA (Fig. 3B).
- two Fis homodimers binding to two Fis sites spaced 7 base pairs apart on DNA (Fig. 3C).

We used several criteria for making the first model. As noted by previous workers, Arg85 and Lys91 in the D helix are likely to be involved in DNA binding since mutations at these amino acids interfere with DNA binding [25,26]. The logo shows highly conserved G-C base pairs at -7 and $+7$ in major groove positions, as is common when contacts are made by proteins with helix-turn-helix motifs [27,28]. In addition, methylation of these bases blocks Fis from binding [29] and Fis protects these bases against methylation [30]. However, as mentioned in previous studies, with B-form DNA the G's at ± 7 would be further apart than the distance between the two DNA-contacting D helices of Fis. Fis can deflect DNA from 40 degrees to 90 degrees [2,25,27,31], so we can attain a shorter distance between the two DNA contact points at ± 7 either by 'bending' (a gradual curvature of the helix axis over a sequence of more than two bases) or by 'kinking' (a sharp change in the DNA strand's axial direction between two adjacent bases) at one or more locations between the two contacts [32].

If specific contacts are not formed, a smoothly bent DNA might be expected to produce smooth sequence conservation over the binding region, but the logo shows strong variable conservation of base sequences at positions -5 through +5, so it is likely that the deflections are not smooth. Furthermore, both CA and TG, which are conserved at positions -4 to -3 and $+3$ to $+4$ [26] (*i.e.* positions ± 3.5, have been shown to display unusually large roll (rotation about the long axis between base pairs [33]. The CRP-DNA complex displays a 90 degrees bend that results from CA and TG kinking [34,35]. (See [28] for the CRP sequence logo.)

Additionally, the other two pyrimidine-purine (Py-Pu) base sequences conserved in Fis sites, TA [36] and CG [37–40] show this high-roll characteristic. These four kinking dinucleotides appear frequently at ± 3.5 and ± 7.5. They account for the base frequencies in the logo and they identify kink locations which could result in the observed DNA deflection [41]. Examination of the individual Fis binding sites confirms this observation. With the exception of seven Fis sites in our set: *fis* X62399 292, *nrd* K02672 3266, *aldB* L40742 153, *hin* V01370

68, *cin* X01828 289, *gin* M10193 389, and *hns* X07688 655, sites which do not have a Py-Pu step at ±3.5 do have one displaced just one position left or right [34]. In addition, these Py-Pu kinks form during energetic minimization of a Fis-DNA complex [42]. We therefore constructed our model using two 42 degree kinks at ±3.5. More kinking would prevent Fis from fitting into the two successive major grooves. Two 50 degree kinks were also added at positions ±7.5 because this was the largest observed kink found for CAP protein binding [35].

The midsection of the logo is AT rich and this should create a higher twist, leading to a compression in the minor groove [32,43,26]. We incorporated this compression into our model by making the helical repeat through the central A-T tract (bases −2 to +2) 10 base pairs per turn and setting the remainder of the model to 10.6 base pairs per turn for B-form DNA [44–46].

We tried to maximize the overall DNA bend angle of Fis bound to DNA and found that the best we could do, while avoiding stereochemical collisions, was 60 degrees. This model of a single Fis homodimer contacting DNA (Fig. 3A) has two kinks and one compression that allow the four nitrogens of Arg85 to be within contact distance (< 3.5 angstroms) of either the O6 or N7 acceptor of each G at ±7 and Lys91 to be within 3.5 angstroms of the phosphate backbone at ±1.5 with no major structural conflicts.

To investigate the consequences of two Fis molecules binding to nearby sites, we constructed 3 dimensional models with Fis sites separated by 11 or 7 base pairs. For the two overlapping Fis models, we extended the general scheme of modeling a Fis homodimer to the synthetic DNA sequences used in our gel shift experiment.

For the overlap 11 model (Fig. 3B), different kinks from the two adjacent sites would be in the same place. Because 50 degree kinks would prevent Fis from fitting into either site, we made both 42 degrees. We incorporated 50 degree kinks on the outside two positions, as in

the single site model. We also increased the twist to −36 degrees at all A-T pairs within the central region [36].

For the overlap 7 model (Fig. 3C), the two outside TG and CA ends were kinked 50 degrees to conform to the single Fis model. However, two pairs which would normally be kinked were not Py-Pu pairs in this sequence. We excluded kinking of the left site at ±7.5 due to a Py-Py pair (CT) and we excluded kinking of the right site at −7.5 due to a Pu-Pu pair (AG). Therefore we only added 42 degree kinks at three positions: at two Py-Pu pairs (left −3.5 CA and right +3.5 TG) which correspond to the single Fis model, and at one in the center (TA). As in the overlap 11 model, twists of −36 degrees were made throughout any A-T tract to be consistent with the single homodimer model.

We found that two Fis proteins separated by 11 base pairs would strongly interpenetrate. On the other hand, two Fis proteins separated by 7 base pairs have a minimal van der Waals force conflict between the two central D helices, but this might be accommodated for by flexibility of the DNA-protein complex. Given that the D helices cannot fit directly into B-form DNA, there is some uncertainty as to how Fis binds DNA. We thought that 11-base separated Fis molecules (Fig. 3B) would compete for binding but that a 7-base separation (Fig. 3C) might allow simultaneous binding.

Results and Discussion

Fis site structures suggest self-competition

In a previous study we used 60 experimentally defined Fis binding sites to construct a highly detailed and reliable model of Fis binding to DNA [5]. When we searched DNA sequences for Fis sites using this information-theory based weight matrix model, we observed Fis sites spaced 11 base pairs apart in the DNA inversion systems *hin*, *gin*, and *min* (see Fig. 5 in reference [5]). A search of the entire *E. coli* genome shows that Fis sites are frequently

separated by 11 bases (Fig. 1). Although the same number of sites are found in a genome-sized equiprobable random sequence, this does not mean that the sites are not functional. Many binding sites have just the amount of information ($R_{sequence}$) needed for them to be found in the genome ($R_{frequency}$) [18]. According to Shannon's principles [47], a well-coded communications system looks like random noise from the outside, and Fis site pairs may follow this principle. So the spike of sites coming in pairs is caused by the pattern of Fis itself; it is intrinsic to the Fis model. We cannot tell if the pattern evolved because Fis sites were required to overlap or whether Fis sites overlap because the pattern evolved. The peaks of Fis sites every three bases in the genomic scan come from coding effects. The peak of spacing at 8 bases and excess sites at spacings of 1 and 3 bases were only noticed during revisions of this paper; we do not understand their significance.

⇐Fig 1

Since B-form DNA twists every 10.6 base pairs, the sites should be on the same side of the DNA. While it is possible for two adjacent proteins to bind simultaneously by a subtle interleaving of their DNA contacts, as in the case of RNA polymerases [48,49], it seems more likely that in this case they will compete for binding in the major groove since after an 11 base shift the sequence logo shows that the predominant G at −7 corresponds to the G at +4 and the C at −4 corresponds to the C at +7 (red arrows in Fig. 2). Competition between these internally redundant patterns [9] would allow Fis to change the point at which the DNA is bent. Perhaps this is important for inversion.

⇐Fig 2

In contrast to the 11 base spacing described above, in bacteriophage P1 *cin*, bacteriophage *cin*, *E. coli e14 pin* sites, and in the genome scan (Fig. 1), pairs of Fis sites are observed at spacings of 7 bases. This would place the Fis dimers 122° apart on B-DNA (360° − 360° per turn × 7 bases/10.6 bases per turn = 122°). After a 7 base shift, the sequence logo shows that the predominant G at −7 would match the A/T/g/c of the minor groove on the opposite face of the DNA at coordinate 0, while the C/T at −4 would match the T/C at +3 and the A/G at −3 would match the G/A at +4 (green arrows in Fig. 2). This allows for

the possibility that the two proteins bind at the same time, which might also be important to the function of these regions.

To investigate the consequences of two Fis molecules binding to nearby sites, we constructed 3 dimensional models (Fig. 3). We found that two Fis proteins bound to sites separated by 11 base pairs might strongly interpenetrate. In contrast, a 7 base pair separation might only have a minimal van der Waals force conflict between the two central D helices. This might be accommodated for by flexibility of the DNA-protein complex, given that there is some uncertainty as to how Fis binds DNA. The models suggest that 11-base separated Fis molecules would compete for binding but that a 7-base separation might allow simultaneous binding. ⇐Fig 3

These ideas are supported by the preliminary observation that synthetic DNA containing either the *hin* proximal or medial Fis sites are bound by Fis in electrophoretic mobility gel shift assays [5]. When these overlapping sites were together on the same fragment with a spacing of 11 bases only one band shift was observed, suggesting that only one of the sites can be bound at a time. To test whether this is the case requires using high concentrations of Fis and strong Fis sites to ensure that both sites would be bound if that were possible.

Test of the self-competition model

To determine whether overlapping Fis sites can be simultaneously bound by Fis, we synthesized self-complementary DNA oligos with strong Fis sites that overlap by either 7 or 11 base pairs (Fig. 4a,b) or that were separated by 23 base pairs (Fig. 4c) and tested their properties by gel shift. Neither the 11 nor the 7 overlapping sites showed a doubly-shifted band, even at an extremely high Fis/DNA ratio and with exceptionally strong ($>$ 12 bits) Fis binding sites (Fig. 5), suggesting that only one Fis molecule could bind to each DNA fragment. The DNA fragment with two Fis sites separated by 23 base pairs did double shift (Fig. 5), demonstrat- ⇐Fig 4 ⇐Fig 5

ing that two well separated Fis sites can simultaneously bind two Fis molecules. However, Fis can create a ladder of complexes on non-specific sequences [50], and this might account for the double-shifts. Under our conditions with short DNAs, a non-specific (all positions < 1 bit) 66 bp DNA fragment barely shifted at high Fis concentration (data not shown) so the secondary shifts were not from non-specific binding. These results demonstrate that Fis sites separated by 7 or 11 bases cannot be bound simultaneously.

Genetic implications of Fis flip-flops

The *Escherichia coli tyrT* promoter has three Fis sites separated by 20 and 31 base pairs. This separation corresponds to our 23 base pair separated control experiment (Fig. 4 and Fig. 5), in which two Fis molecules bind independently. The separation in *tyrT* is sufficient for three Fis dimers to simultaneously position themselves on the same face of the DNA to cooperatively bind an αCTD subunit and activate transcription of stable RNA promoters [51]. In addition to this activation mechanism, which is based on separated sites, Fis may also have evolved another control mechanism that uses overlapping sites.

When we scanned our Fis individual information model across various sequences, we discovered 7 and 11 spacings at DNA inversion regions, the *fis*, *nrd*, and *ndh* promoters, and at *dif*, *E. coli oriC* and λ *att* [10,5]. In the latter three systems, sequence walkers for Fis sites overlap binding sites of other proteins in biologically significant places, so we do not think that Fis sites appear at this spacing merely because of the internal redundancy of the site. Scanning with the Fis weight matrix also reveals two strong Fis sites previously identified in *oriC* at coordinates 202 and 213 although only one was thought to be bound [52–54]. Footprinting data from three different groups show protection covering one, the other and both sites (Fig. 6). ⇐Fig 6

We confirmed both *oriC* Fis sites individually by a gel mobility shift experiment (Fig. 7). ⇐Fig 7

Because of our previous work on Fis promoters [5], we knew that we could use sequence walkers [10] to engineer the sites. This way we could know that we had destroyed one site without affecting the other and without creating new sites [48]. When both sites were mutated (oligo **nn**), no shift was observed until a high Fis concentration (1000 nM) was used. We take this concentration to represent non-specific binding for all four oligos used in the experiment. When the left (oligo **no**) or right (oligo **on**) Fis sites were mutated, only a single shifted band was observed. This confirms our prediction that both sites can bind Fis. In the **on** control a high band appeared at the highest Fis concentration. From the lower concentration lanes, we know that there is one specific binding site and, apparently, at this extremely high concentration a second Fis molecule binds nonspecifically on this particular DNA sequence. This additional band may be explained by the existence of a number of weak (< 0 bit) sites that can bind DNA at high Fis concentration. In any case this band is absent in the experimental lane (**oo**).

The experiment also shows that the left Fis site, which is closer to DnaA R2 (oligo **on**, 8.4 bits) binds slightly less strongly than the one closer to DnaA R3 (oligo **no**, 9.1 bits) confirming the respective individual information contents, which differ by about 1 bit. When both sites were wild-type (oligo **oo**) Fis binding was also observed but no supershift was visible. We conclude that only one Fis molecule can bind at a time between DnaA R2 and DnaA R3 of *oriC*.

The **oo** oligo can be bound in two distinct ways, so its association constant to Fis should be the sum of the two individual site association constants. This effect may have practical applications since creating overlapping binding sites will double the sensitivity of a biological detection system [55].

The two Fis sites at *oriC* fit exactly between the R2 and R3 DnaA sites and have similar individual information contents, suggesting that their binding energies are similar [56], so

in the absence of other effects Fis could occupy them for nearly equal fractions of the time as a molecular flip-flop. The two states have not been recognized before because DNA footprinting only shows one predominant state or shows both states simultaneously, and such footprints have hitherto been interpreted as representing single sites. That is, the macroscopic experiments did not reveal that there are two distinct binding modes at dual Fis sites.

Our results resolve two previously conflicting reports. Gille *et al* found that Fis binding and DnaA binding at R3 are mutually exclusive [52] but Margulies and Kaguni found that they could bind concurrently [57]. The controversy may be resolved by noting that different experimental techniques were used and that, because consensus sequences were being used, it was not clear that there are two Fis sites [22]. The experiments by Gille *et al* were DNase I footprints, which show protection of the entire R2-R3 region when Fis is bound prior to DnaA, as would be expected from a mixture of two states. The experiments by Margulies and Kaguni were footprints and gel shifts. If DnaA binds to R3, Fis might be blocked at position 213 but Fis could still bind at the other site at position 202. It is possible that both experiments produced valid data but for different states of the flip-flop.

If binding by DnaA and Fis are mutually exclusive [52], then the position of a Fis-induced DNA bend could be controlled by DnaA and the binding of DnaA could be controlled by Fis. During nutritional upshifts when there is a high Fis concentration [4], occupancy of one Fis site should ensure that only one DnaA site is available at a time. DnaA directs the loading of the DnaB helicase which, in turn, determines the orientation of the DNA polymerase or DnaG primase [58] and therefore the direction of replication [59–61,12,62]. Since the Fis flip-flop probably controls which of the two oppositely oriented DnaA sites can be bound, it may control the alternative firing of replication complexes in opposite directions [63]. This appears to be consistent with the divergent directions of DNA initiation observed in this region, as shown in Fig. 6 [64]. Indeed, the absence of Fis leads to asynchronous replication

[65], and at high temperatures *fis* null mutants have been shown to form filamentous cells and have aberrant nucleoid segregation [66]. Although Fis is pleiotropic, these and other observations [67,68] are consistent with Fis being required for proper replication initiation. However, initiation using purified components in the absence of Fis is bidirectional [59,69]. One possibility is that in the absence of Fis, loading orientation is random [69] and that initiation *in vivo* fails unless the complexes are oppositely oriented. This should occur ~25% of the time, and indeed when initiating molecules were counted, only 36% formed SSB bubbles [59]. Since there may be other explanations for these data, further experiments will be needed to determine how the Fis flip-flop is involved in initiation.

A general model for how DnaA is involved in origin replication has been proposed [70,71] in which the R1 DnaA site is thought to be involved in opening the adjacent AT rich 13 mer region. That model does not include the two competing Fis sites demonstrated in this paper. To combine the models, one possibility is that the genetic structure at R2-R3 is only functional at nutritional upshifts when there is a high concentration of Fis in the cell [4]. Under these circumstances coordination of replication fork firing may be critical to start the first or subsequent rounds of replication correctly.

Are these proposals supported by mutations in the Fis site(s)? An experiment by Weigel *et al* was intended to destroy 'the' Fis site between R2 and R3 [72], however analysis by sequence walkers shows that the *oriC*131 AACTCAA to ATGTGTA mutation decreased the left Fis site (at coordinate 202) from 8.4 to 3.1 bits, while leaving the right site at 213 unchanged (analysis not shown). When placed in the *E. coli* chromosome, the mutant shows a moderate change of asynchrony of initiation by flow cytometry. Unfortunately, the mutation also created an 8.9 bit DnaA site which makes the experimental results difficult to interpret. (This 8.9 bit site also was created in earlier experiments [53].) In another experiment, replacing six bases between R2 and R3 by a *Bam*HI site decreased *oriC* dependent plasmid transformation by 57 fold [73]. Sequence walker analysis (not shown) indicates that this mutation

destroys both Fis sites (< 2 bits) while leaving the R2 and R3 DnaA sites intact. A 10 base pair insertion at coordinate 203 (presumably between bases 203 and 204) destroyed the MPE (Methidiumpropyl-EDTA-Fe^{++}) footprint of the Fis site at 202 but the site at 213 still showed an MPE footprint [54]. This same insert reduced transformation frequency of an *oriC* plasmid into a *polA* strain. These results suggest that the Fis flip-flop is important for replication from *oriC*.

When does the flip-flop state change? *In vivo* footprinting shows that during the cell cycle DnaA sites R1, R2 and R4 are bound, but R3 is not occupied [74]. R3 becomes occupied at initiation of DNA replication [75]. R3 is also bound by DnaA more weakly than R2 *in vitro* [76,77]. This is consistent with the information measures which suggest that R3 is $13.5 - 12.1 = 1.4$ bits or at least $2^{1.4} = 2.6$ fold weaker (Fig. 6). In the absence of other processes, R3 should be bound less frequently. We suggest that for the majority of the cell cycle DnaA bound to R2 blocks Fis at position 202, which allows Fis to bind at 213 and which, in turn, entirely blocks R3 from being bound by DnaA. Replication initiation may temporarily alter the flip-flop state, exposing R3. These data suggest the alternative hypothesis that the flip-flop is part of an on-off switch controlling initiation in the presence of Fis, especially during nutritional upshift [4].

Closely spaced sites are often bound cooperatively, as in the classical example of T4 gene 32 autogenous regulation [78], and even at overlapping LexA sites on opposite faces of the DNA [79,80]. In contrast, Fis represents the unusual situation where a protein competes with itself by binding at overlapping positions. Self-occlusion has been observed in artificial constructs, where one ribosome is apparently blocked by the presence of another ribosome bound nearby [81], by polymerases at promoters [48,82,49,83], and in enzyme complexes [84]. An interplay of factors may be typical of complex flip-flop mechanisms. For example, as many as five Fis sites are likely to be in λ *att*. Two of these are spaced 11 base pairs apart, with one of them overlapping an Xis site [10]. Likewise, at the *E. coli dif* locus, where the XerC and XerD

site-specific recombination proteins bind [85], one finds an overlapping set of 7 weak Fis sites; three of these are separated by 11 bases (data not shown).

The positioning of Fis binding sites relative to one another and to the binding sites of other proteins therefore appears to be key for the ability of Fis to perform many diverse functions. Fis has evolved a transcriptional activation mode in which sites are on the same face of the DNA and are sufficiently far apart to be bound simultaneously [51]. Fis may also have specifically evolved to allow for two competitive binding modes. When the sites are on the same face of DNA (11 bp apart), a single Fis molecule could disengage and rebind to move the bend location between two possible places without changing the overall direction of the DNA. When sites are on nearly opposite faces (7 bp apart), shifting a Fis dimer molecule would cause the bend direction to change by 122°. How these cogs fit into the larger picture of pleiotropic Fis functions remains to be determined. However, an 11 base pair shift of a bend would have dramatic effects on the *oriC* DNA initiation complex structure such as the ones suggested by Messer *et al* [54,86].

Technological Uses of Molecular Flip-Flops

Molecular flip-flops such as the Fis flip-flop have at least two technological uses. First, binding equations for the flip-flop show that, compared to a single binding site, a dual-competitive site (in which both parts have the same binding energy) should have a doubled association constant [82]. This curious property may be useful in biological assays, because it increases the sensitivity without increasing the amount of bound protein [55].

Second, the flip-flop and other exclusionary binding site clusters may be used to construct a molecular computer since overlapping binding sites can provide Boolean logic [55]. Previous molecular-based computers have used DNA hybridization and PCR [87], and carbon monoxide on copper surfaces [88] among other methods. Steps have been made to evolve

genetic circuits [89] and it has been demonstrated that the stability of gene networks can be increased by autoregulation [90] to provide sharp logical responses for 'digital genetics'. Genetic networks have also been shown to provide distinct toggle switching [91,92] and Boolean logic [93]. In conjunction with these cell-sized chemical circuits, molecular flip-flops could be used to build molecular-scale circuits with sharp Boolean responses. A NOR gate can be constructed from a set of three binding sites, **r1**, **a** and **r2**, in which the middle site **a** is the binding site for a transcriptional activator **A** [55] (Fig. 8). When bound, **A** causes transcription of a downstream operon that can contain a gene for a signal such as GFP (Green Fluorescence Protein [94]) or a DNA binding protein for another part of the circuit. In the gate, **r1** overlaps **a** and **r2** overlaps **a** from the other side. As shown in Fig. 7, design of such constructs is facilitated by sequence walkers [10]. If either the **R1** or the **R2** binding protein exists in the cell or solution, then the **A** binding protein is excluded and the transcript is not expressed. Because this forms a NOR gate with amplification and fan-out at every step, any digital genetic circuit can be constructed, including complete computers [95–98]. This method can take advantage of the inherent precision of protein/DNA recognition to provide low error rates [99,100]. Unlike the gene-sized logic circuits previously described [93], which require about 3000 bases of DNA (1μm), the functional component of these gates can be about 30 bases (10 nm) in length.

⇐Fig 8

Acknowledgments

We thank Reid Johnson for generously supplying Fis protein and pRJ1077; Nagarajan Pattabiraman, Connor McGrath, George McGregor, Rick Gussio and Mike Ma for advice and help with molecular graphics; the Frederick Biomedical Supercomputing Center for access to computer resources; Tom Hunter for enlightening discussions; Peter Rogan, Denise Rubens, R. M. Stephens, Keith Robison, Stacy L. Bartram, Dhruba Chattoraj, Ryan Shultzaberger, and Danielle Needle for comments on the manuscript. The content of this publication does

not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract numbers N01-C0-56000 and N01-C0-12400.

References

[1] Johnson, R. C. and Simon, M. I. (1987) Enhancers of site-specific recombination in bacteria. *Trends in Genetics,* **3**, 262–267.

[2] Finkel, S. E. and Johnson, R. C. (1992) The Fis protein: it's not just for DNA inversion anymore. *Mol. Microbiol.,* **6**, 3257–3265.

[3] Finkel, S. E. and Johnson, R. C. (1992) The Fis protein: it's not just for DNA inversion anymore (erratum). *Mol. Microbiol.,* **6**, 1023.

[4] Ball, C. A., Osuna, R., Ferguson, K. C., and Johnson, R. C. (1992) Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. *J. Bacteriol.,* **174**, 8043–8056.

[5] Hengen, P. N., Bartram, S. L., Stewart, L. E., and Schneider, T. D. (1997) Information analysis of Fis binding sites. *Nucleic Acids Res.,* **25**(24), 4994–5002 http://www.lecb.ncifcrf.gov/~toms/paper/fisinfo/.

[6] Travers, A., Schneider, R., and Muskhelishvili, G. (2001) DNA supercoiling and transcription in Escherichia coli: The FIS connection. *Biochimie,* **83**, 213–217.

[7] Ussery, D., Larsen, T. S., Wilkes, K. T., Friis, C., Worning, P., Krogh, A., and Brunak, S. (2001) Genome organisation and chromatin structure in *Escherichia coli*. *Biochimie,* **83**, 201–212.

[8] Dorman, C. J. and Deighan, P. (2003) Regulation of gene expression by histone-like proteins in bacteria. *Curr Opin Genet Dev,* **13**, 179–184.

[9] Schneider, T. D. and Mastronarde, D. (1996) Fast multiple alignment of ungapped DNA sequences using information theory and a relaxation method. *Discrete Applied Mathematics,* **71**, 259–268 http://www.lecb.ncifcrf.gov/~toms/paper/malign.

[10] Schneider, T. D. (1997) Sequence walkers: a graphical method to display how binding proteins interact with DNA or RNA sequences. *Nucleic Acids Res.,* **25**, 4408–4415 http://www.lecb.ncifcrf.gov/~toms/paper/walker/, erratum: NAR 26(4): 1135, 1998.

[11] Messer, W. and Weigel, C. (1996) Initiation of chromosome replication. In Neidhardt, F. C., Curtiss III, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., (eds.), Escherichia coli *and* Salmonella*: Cellular and Molecular Biology*, Washington, D.C.: American Society for Microbiology Press Vol. 2, pp. 1579–1601.

[12] Baker, T. A. and Bell, S. P. (1998) Polymerases and the replisome: machines within machines. *Cell,* **92**, 295–305.

[13] Prescott, D. M. and Kuempel, P. L. (1972) Bidirectional replication of the chromosome in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA,* **69**, 2842–2845.

[14] Bird, R. E., Louarn, J., Martuscelli, J., and Caro, L. (1972) Origin and sequence of chromosome replication in *Escherichia coli*. *J. Mol. Biol.,* **70**, 549–566.

[15] Hill, T. M. (1996) Features of the chromosomal terminus region. In Neidhardt, F. C., Curtiss III, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., (eds.), Escherichia coli *and* Salmonella*: Cellular and Molecular Biology*, Washington, D.C.: American Society for Microbiology Press Vol. 2, pp. 1602–1614.

[16] Schneider, T. D., Stormo, G. D., Haemer, J. S., and Gold, L. (1982) A design for computer nucleic-acid sequence storage, retrieval and manipulation. *Nucleic Acids Res.,* **10**, 3013–3024.

[17] Schneider, T. D., Stormo, G. D., Yarus, M. A., and Gold, L. (1984) Delila system tools. *Nucleic Acids Res.,* **12**, 129–140.

[18] Schneider, T. D., Stormo, G. D., Gold, L., and Ehrenfeucht, A. (1986) Information content of binding sites on nucleotide sequences. *J. Mol. Biol.,* **188**, 415–431 http://www.lecb.ncifcrf.gov/~toms/paper/schneider1986/.

[19] Schneider, T. D. and Stephens, R. M. (1990) Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.,* **18**, 6097–6100 http://www.lecb.ncifcrf.gov/~toms/paper/logopaper/.

[20] Stephens, R. M. and Schneider, T. D. (1992) Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J. Mol. Biol.,* **228**, 1124–1136 http://www.lecb.ncifcrf.gov/~toms/paper/splice/.

[21] Schneider, T. D. (1997) Information content of individual genetic sequences. *J. Theor. Biol.,* **189**(4), 427–441 http://www.lecb.ncifcrf.gov/~toms/paper/ri/.

[22] Schneider, T. D. (2002) Consensus Sequence Zen. *Applied Bioinformatics,* **1**(3), 111–119 http://www.lecb.ncifcrf.gov/~toms/papers/zen/.

[23] Studier, F. W. and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.,* **189**, 113–130.

[24] Pan, C. Q., Finkel, S. E., Cramton, S. E., Feng, J. A., Sigman, D. S., and Johnson, R. C. (1996) Variable structures of Fis-DNA complexes determined by flanking DNA- protein contacts. *J. Mol. Biol.,* **264**, 675–695.

[25] Kostrewa, D., Granzin, J., Koch, C., Choe, H.-W., Raghunathan, S., Wolf, W., Labahn, J., Kahmann, R., and Saenger, W. (1991) Three-dimensional structure of the *E. coli* DNA-binding protein FIS. *Nature,* **349**, 178–180.

[26] Kostrewa, D., Granzin, J., Stock, D., Choe, H.-W., Labahn, J., and Saenger, W. (1992) Crystal structure of the factor for inversion stimulation FIS at 2·0 Å resolution. *J. Mol. Biol.,* **226**, 209–226.

[27] Yuan, H. S., Finkel, S. E., Feng, J.-A., Kaczor-Grzeskowiak, M., Johnson, R. C., and Dickerson, R. E. (1991) The molecular structure of wild-type and a mutant Fis protein: Relationship between mutational changes and recombinational enhancer function or DNA binding. *Proc. Natl. Acad. Sci. USA,* **88**, 9558–9562.

[28] Papp, P. P., Chattoraj, D. K., and Schneider, T. D. (1993) Information analysis of sequences that bind the replication initiator RepA. *J. Mol. Biol.,* **233**, 219–230.

[29] Bruist, M. F., Glasgow, A. C., Johnson, R. C., and Simon, M. I. (1987) Fis binding to the recombinational enhancer of the Hin DNA inversion system. *Genes Dev.,* **1**, 762–772.

[30] Thompson, J. F., deVargas, L. M., Koch, C., Kahmann, R., and Landy, A. (1987) Cellular factors couple recombination with growth phase: Characterization of a new component in the λ site-specific recombination pathway. *Cell,* **50**, 901–908.

[31] Thompson, J. F. and Landy, A. (1988) Empirical estimation of protein-induced DNA bending angles: applications to λ site-specific recombination complexes. *Nucleic Acids Res.,* **16**, 9687–9705.

[32] Wu, H.-M. and Crothers, D. M. (1984) The locus of sequence-directed and protein-induced DNA bending. *Nature,* **308**, 509–513.

[33] Dickerson, R. E. (1989) Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.,* **17**, 1797–1803.

[34] Barber, A. M. and Zhurkin, V. B. (1990) CAP binding sites reveal pyrimidine-purine pattern characteristic of DNA bending. *J. Biomol. Struct. Dyn.,* **8**, 213–232.

[35] Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) Crystal structure of a CAP-DNA complex: The DNA is bent by 90°. *Science,* **253**, 1001–1007.

[36] Goodsell, D. S., Kaczor-Grzeskowiak, M., and Dickerson, R. E. (1994) The crystal structure of C-C-A-T-T-A-A-T-G-G; implications for bending of *B*-DNA at T-A steps. *J. Mol. Biol.,* **239**, 79–96.

[37] Gorin, A. A., Zhurkin, V. B., and Olson, W. K. (1995) *B*-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.,* **247**, 34–48.

[38] Suzuki, M. and Yagi, N. (1995) Stereochemical basis of DNA bending by transcription factors. *Nucleic Acids Res.,* **23**, 2083–2091.

[39] Crothers, D. M., Haran, T. E., and Nadeau, J. G. (May 5, 1990) Intrinsically bent DNA. *J. Biol. Chem.,* **265**, 7093–7096.

[40] Cognet, J. A. H., Boulard, Y., and Fazakerley, G. V. (1995) Helical parameters, fluctuations, alternative hydrogen bonding, and bending in oligonucleotides containing a mismatched base-pair by NOESY distance restrained and distance free molecular dynamics. *J. Mol. Biol.,* **246**, 209–226.

[41] Lazarus, L. R. and Travers, A. A. (1993) The *Escherichia coli* FIS protein is not required for the activation of *tyrT* transcription on entry into exponential growth. *EMBO J.,* **12**, 2483–2494.

[42] Sandmann, C., Cordes, F., and Saenger, W. (1996) Structure model of a complex between the factor for inversion stimulation (FIS) and DNA: Modeling protein–DNA complexes with dyad symmetry and known protein structures. *Proteins: Structure, Function, and Genetics,* **25**, 486–500.

[43] Lipanov, A. A. and Chuprina, V. P. (1987) The structure of poly(dA):poly(dT) in a condensed state and in solution. *Nucleic Acids Res.,* **15**(14), 5833–5844.

[44] Peck, L. J. and Wang, J. C. (1981) Sequence dependence of the helical repeat of DNA in solution. *Nature,* **292**, 375–378.

[45] Rhodes, D. and Klug, A. (1980) Helical periodicity of DNA determined by enzyme digestion. *Nature,* **286**, 573–578.

[46] Rhodes, D. and Klug, A. (1981) Sequence-dependent helical periodicity of DNA. *Nature,* **292**, 378–380.

[47] Shannon, C. E. (1948) A mathematical theory of communication. *Bell System Tech. J.,* **27**, 379–423, 623–656 http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html.

[48] Goodrich, J. A. and McClure, W. R. (1991) Competing promoters in prokaryotic transcription. *Trends Biochem Sci,* **16**, 394–397.

[49] Strainic Jr, M. G., Sullivan, J. J., Collado-Vides, J., and deHaseth, P. L. (2000) Promoter interference in a bacteriophage lambda control region: effects of a range of interpromoter distances. *J. Bacteriol.,* **182**, 216–220.

[50] Bétermier, M., Galas, D. J., and Chandler, M. (1994) Interaction of Fis protein with DNA: Bending and binding specificity of binding. *Biochimie,* **76**, 958–967.

[51] Muskhelishvili, G., Travers, A. A., Heumann, H., and Kahmann, R. (1995) FIS and RNA polymerase holoenzyme form a specific nucleoprotein complex at a stable RNA promoter. *EMBO J.,* **14**, 1446–1452.

[52] Gille, H., Egan, J. B., Roth, A., and Messer, W. (1991) The FIS protein binds and bends the origin of chromosomal DNA replication, *oriC*, of *Escherichia coli*. *Nucleic Acids Res.,* **19**, 4167–4172.

[53] Roth, A., Urmoneit, B., and Messer, W. (1994) Functions of histone-like proteins in the initiation of DNA replication at *oriC* of *Escherichia coli*. *Biochimie,* **76**, 917–923.

[54] Messer, W., Hartmann-Kühlein, H., Langer, U., Mahlow, E., Roth, A., Schaper, S., Urmoneit, B., and Woelker, B. (1992) The complex for replication initiation of *Escherichia coli*. *Chromosoma,* **102**, S1–S6.

[55] Schneider, T. D. and Hengen, P. N. MOLECULAR COMPUTING ELEMENTS: GATES AND FLIP-FLOPS. (2003) US Patent application, patent pending, WO 99/42929, PCT/US99/03469. European Patent granted, number 1057118. http://www.lecb.ncifcrf.gov/~toms/patent/molecularcomputing/.

[56] Schneider, T. D. (1991) Theory of molecular machines. II. Energy dissipation from molecular machines. *J. Theor. Biol.,* **148**, 125–137 http://www.lecb.ncifcrf.gov/~toms/paper/edmm/.

[57] Margulies, C. and Kaguni, J. M. (1998) The FIS protein fails to block the binding of DnaA protein to *oriC*, the *Escherichia coli* chromosomal origin. *Nucleic Acids Res.,* **26**, 5170–5175.

[58] Hiasa, H. and Marians, K. J. (1999) Initiation of bidirectional replication at the chromosomal origin is directed by the interaction between helicase and primase. *J. Biol. Chem.,* **274**, 27244–27248.

[59] Baker, T. A., Funnell, B. E., and Kornberg, A. (1987) Helicase action of dnaB protein during replication from the *Escherichia coli* chromosomal origin *in vitro*. *J. Biol. Chem.,* **262**, 6877–6885.

[60] Marszalek, J. and Kaguni, J. M. (1994) DnaA protein directs the binding of DnaB protein in initiation of DNA replication in Escherichia coli. *J. Biol. Chem.,* **269**, 4883–4890.

[61] Sutton, M. D., Carr, K. M., Vicente, M., and Kaguni, J. M. (1998) *Escherichia coli* DnaA protein. The N-terminal domain and loading of DnaB helicase at the *E. coli* chromosomal origin. *J. Biol. Chem.,* **273**, 34255–34262.

[62] Carr, K. M. and Kaguni, J. M. (2001) Stoichiometry of DnaA and DnaB protein in initiation at the Escherichia coli chromosomal origin. *J. Biol. Chem.,* **276**, 44919–44925.

[63] Meijer, M. and Messer, W. (1980) Functional analysis of minichromosome replication: bidirectional and unidirectional replication from the *Escherichia coli* replication origin, *oriC*. *J. Bacteriol.,* **143**, 1049–1053.

[64] Seufert, W. and Messer, W. (1987) Start sites for bidirectional *in vitro* DNA replication inside the replication origin, *oriC*, of *Escherichia coli*. *EMBO J.,* **6**, 2469–2472.

[65] Boye, E., Lyngstadaas, A., Løbner-Olesen, A., Skarstad, K., and Wold, S. (1982) Regulation of DNA replication in *Escherichia coli*. In Fanning, E., Knippers, R., and Winnacker, E.-L., (eds.), *DNA Replication and the Cell cycle*, Berlin: Springer-Verlag Vol. 43, pp. 15–26.

[66] Filutowicz, M., Ross, W., Wild, J., and Gourse, R. L. (1992) Involvement of Fis protein in replication of the *Escherichia coli* chromosome. *J. Bacteriol.,* **174**, 398–407.

[67] Hiasa, H. and Marians, K. J. (1994) Fis cannot support *oriC* DNA replication *in vitro*. *J. Biol. Chem.,* **269**, 24999–25003.

[68] Wold, S., Crooke, E., and Skarstad, K. (1996) The *Escherichia coli* Fis protein prevents initiation of DNA replication from *oriC in vitro*. *Nucleic Acids Res.,* **24**, 3527–3532.

[69] Fang, L., Davey, M. J., and O'Donnell, M. (1999) Replisome assembly at *oriC*, the replication origin of *E. coli*, reveals an explanation for initiation sites outside an origin. *Mol Cell,* **4**, 541–553.

[70] Speck, C. and Messer, W. (2001) Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *EMBO J.,* **20**, 1469–1476.

[71] Messer, W., Blaesing, F., Jakimowicz, D., Krause, M., Majka, J., Nardmann, J., Schaper, S., Seitz, H., Speck, C., Weigel, C., Wegrzyn, G., Welzeck, M., and Zakrzewska-Czerwinska, J. (2001) Bacterial replication initiator DnaA. Rules for DnaA binding and roles of DnaA in origin unwinding and helicase loading. *Biochimie,* **83**, 5–12.

[72] Weigel, C., Messer, W., Preiss, S., Welzeck, M., Morigen, and Boye, E. (2001) The sequence requirements for a functional Escherichia coli replication origin are different for the chromosome and a minichromosome. *Mol. Microbiol.,* **40**, 498–507.

[73] Crooke, E., Thresher, R., Hwang, D. S., Griffith, J., and Kornberg, A. (1993) Replicatively active complexes of DnaA protein and the Escherichia coli chromosomal origin observed in the electron microscope. *J. Mol. Biol.,* **233**, 16–24.

[74] Samitt, C. E., Hansen, F. G., Miller, J. F., and Schaechter, M. (1989) *In vivo* studies of DnaA binding to the origin of replication of *Escherichia coli*. *EMBO J.,* **8**, 989–993.

[75] Cassler, M. R., Grimwade, J. E., and Leonard, A. C. (1995) Cell cycle-specific changes in nucleoprotein complexes at a chromosomal replication origin. *EMBO J.,* **14**, 5833–5841.

[76] Margulies, C. and Kaguni, J. M. (1996) Ordered and sequential binding of DnaA protein to *oriC*, the chromosomal origin of *Escherichia coli*. *J. Biol. Chem.,* **271**, 17035–17040.

[77] Weigel, C., Schmidt, A., Rückert, B., Lurz, R., and Messer, W. (1997) DnaA protein binding to individual DnaA boxes in the *Escherichia coli* replication origin, *oriC*. *EMBO J.,* **16**, 6574–6583.

[78] Miller, E. S., Karam, J. D., and Spicer, E. (1994) Control of translation initiation: mRNA structure and protein repressors. In J. D. Karam *et al.*, (ed.), *Molecular Biology of Bacteriophage T4*, Washington, D. C.: American Society for Microbiology pp. 193–205.

[79] Ebina, Y., Takahara, Y., Kishi, F., and Nakazawa, A. (1983) LexA protein is a repressor of the colicin E1 gene. *J. Biol. Chem.,* **258**, 13258–13261.

[80] Lloubes, R., Granger-Schnarr, M., Lazdunski, C., and Schnarr, M. (1991) Interaction of a regulatory protein with a DNA target containing two overlapping binding sites. *J. Biol. Chem.,* **266**, 2303–2312.

[81] Barrick, D., Villanueba, K., Childs, J., Kalil, R., Schneider, T. D., Lawrence, C. E., Gold, L., and Stormo, G. D. (1994) Quantitative analysis of ribosome binding sites in *E. coli.*. *Nucleic Acids Res.,* **22**, 1287–1295.

[82] Goodrich, J. A. and McClure, W. R. (1992) Regulation of open complex formation at the *Escherichia coli* galactose operon promoters. Simultaneous interaction of RNA polymerase, *gal* repressor and CAP/cAMP. *J. Mol. Biol.,* **224**, 15–29.

[83] Xu, J. and Koudelka, G. B. (2000) Mutually exclusive utilization of P_R and P_{RM} promoters in bacteriophage 434 O_R. *J. Bacteriol.,* **182**, 3165–3174.

[84] Mande, S. S., Sarfaty, S., Allen, M. D., Perham, R. N., and Hol, W. G. (1996) Protein-protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase. *Structure,* **4**, 277–286.

[85] Blakely, G. W. and Sherratt, D. J. (1994) Interactions of the site-specific recombinases XerC and XerD with the recombination site *dif*. *Nucleic Acids Res.,* **22**, 5613–5620.

[86] Woelker, B. and Messer, W. (1993) The structure of the initiation complex at the replication origin, *oriC*, of *Escherichia coli*. *Nucleic Acids Res.,* **21**, 5025–5033.

[87] Adleman, L. M. (1994) Molecular computation of solutions to combinatorial problems. *Science,* **266**, 1021–1024.

[88] Heinrich, A. J., Lutz, C. P., Gupta, J. A., and Eigler, D. M. (2002) Molecule cascades. *Science,* **298**, 1381–1387.

[89] Yokobayashi, Y., Weiss, R., and Arnold, F. H. (2002) Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. USA,* **99**, 16587–16591.

[90] Becskei, A. and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. *Nature,* **405**, 590–593.

[91] Monod, J. and Jacob, F. (1961) General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. *Cold Spring Harb. Symp. Quant. Biol.,* **26**, 389–401.

[92] Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature,* **403**, 339–342.

[93] Guet, C. C., Elowitz, M. B., Hsing, W., and Leibler, S. (2002) Combinatorial synthesis of genetic networks. *Science,* **296**, 1466–1470.

[94] Tsien, R. Y. (1998) The green fluorescent protein. *Annu Rev Biochem,* **67**, 509–544.

[95] Wait, J. V. (1967) Symbolic logic and practical applications. In Klerer, M. and Korn, G. A., (eds.), *Digital Computer User's Handbook*, N. Y.: McGraw-Hill Book Company, Inc. pp. 4–3 to 4–28 show that NAND is sufficient.

[96] Gersting, J. L. (1986) Mathematical structures for computer science, W. H. Freeman and Co., New York second edition.

[97] Schilling, D. L., Belove, C., Apelewicz, T., and Saccardi, R. J. (1989) Electronic circuits, discrete and integrated, McGraw-Hill, New York third edition.

[98] Schneider, T. D. (1994) Sequence logos, machine/channel capacity, Maxwell's demon, and molecular computers: a review of the theory of molecular machines. *Nanotechnology,* **5**, 1–18 http://www.lecb.ncifcrf.gov/~toms/paper/nano2/.

[99] Shannon, C. E. (1949) Communication in the presence of noise. *Proc. IRE,* **37**, 10–21.

[100] Schneider, T. D. (1991) Theory of molecular machines. I. Channel capacity of molecular machines. *J. Theor. Biol.,* **148**, 83–123 http://www.lecb.ncifcrf.gov/~toms/paper/ccmm/.

[101] Schneider, T. D. (1996) Reading of DNA sequence logos: Prediction of major groove binding by information theory. *Meth. Enzym.,* **274**, 445–455 http://www.lecb.ncifcrf.gov/~toms/paper/oxyr/.

[102] Schneider, T. D. (2001) Strong minor groove base conservation in sequence logos implies DNA distortion or base flipping during replication and transcription initiation. *Nucl. Acid Res.,* **29**(23), 4881–4891 http://www.lecb.ncifcrf.gov/~toms/paper/baseflip/.

[103] Messer, W., Meijer, M., Bergmans, H. E. N., Hansen, F. G., von Meyenburg, K., Beck, E., and Schaller, H. (1979) Origin of replication, *oriC*, of the *Escherichia coli* K12 chromosome: nucleotide sequence. *Cold Spring Harb. Symp. Quant. Biol.,* **43 Pt 1**, 139–145.

[104] Messer, W., Egan, B., Gille, H., Holz, A., Schaefer, C., and Woelker, B. (1991) The complex of *oriC* DNA with the DnaA initiator protein. *Res. Microbiol.,* **142**, 119–125.

[105] Cassler, M. R., Grimwade, J. E., McGarry, K. C., Mott, R. T., and Leonard, A. C. (1999) Drunken-cell footprints: nuclease treatment of ethanol-permeabilized bacteria reveals an initiation-like nucleoprotein complex in stationary phase replication origins. *Nucleic Acids Res.,* **27**, 4570–4576.

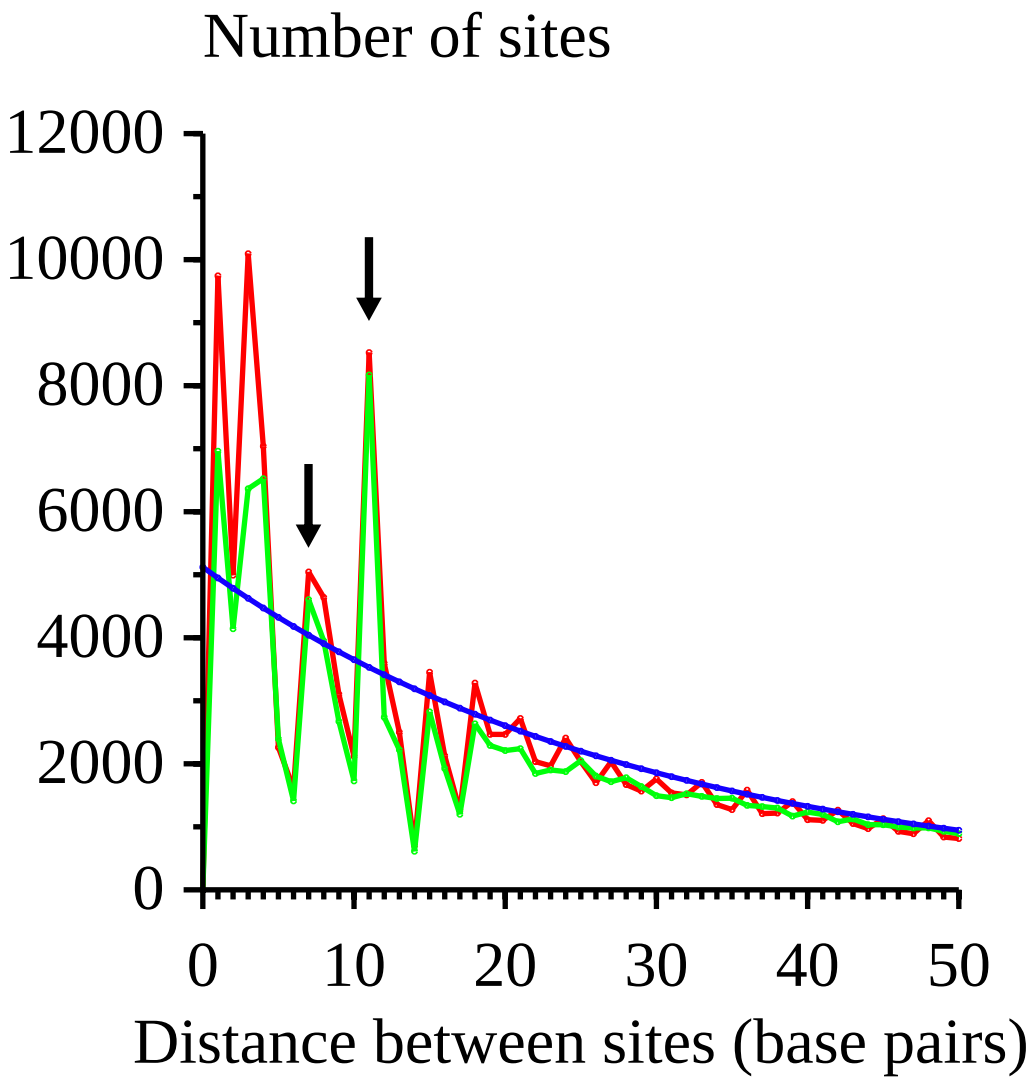


Fig. 1. Predicted Fis site spacings.

An information-theory based model of Fis protein DNA binding was searched across the *E. coli* genome. The distances between the zero coordinates of successive sites were recorded and tabulated in this graph. Red curve: search of *E. coli* genome; green curve: search over equiprobable random sequence; blue curve: mathematical model for randomly placed sites. This model was constructed by considering a genome of size $G = 4639221$ bases (the U00096 *E. coli* K-12 MG1655 genome) having $n = 154112$ sites ($R_i > 0$ bits) so that the probability of a site being at one position is $p = n/G$. Then the number of sites with separation d is $Gp^2(1-p)^d$. Similar results are obtained for a 2.5 bit cutoff, the lowest observed Fis site in our set. Arrows indicate spacings of 7 and 11 base pairs.

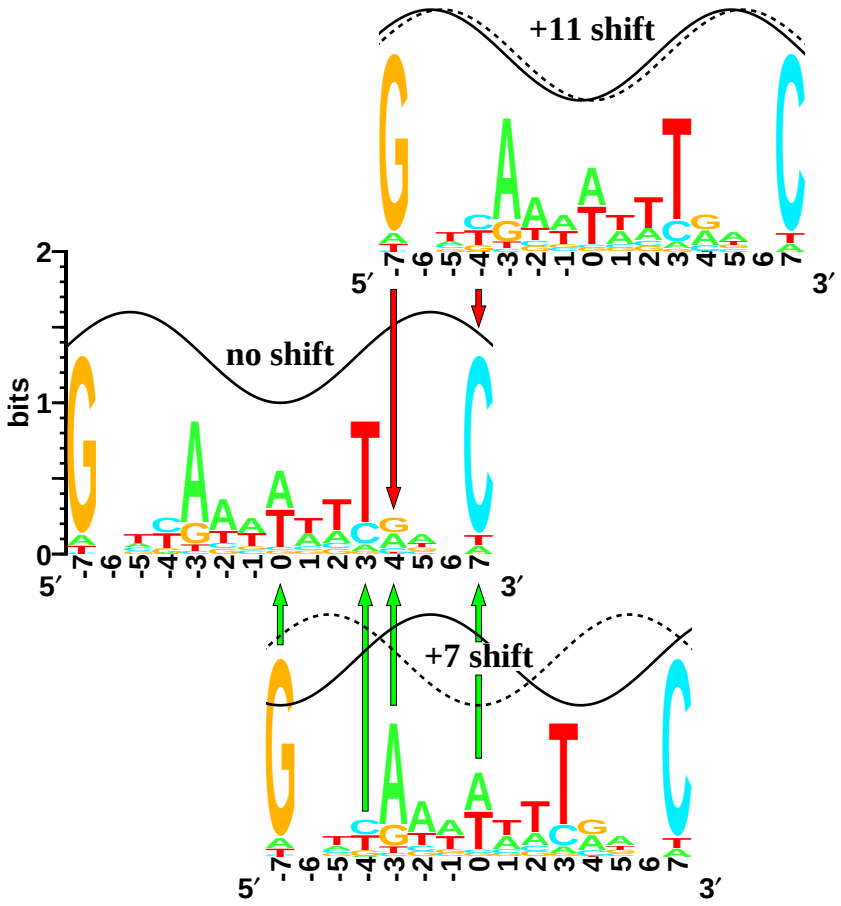


Fig. 2. Self-similarity of Fis binding sites.

The sequence logo for Fis [19,5] is shown three times. The upper and lower logos are shifted +11 and +7 bases to the right (respectively) relative to the middle logo. Dashed waves indicate the phase of the shifted site; solid waves indicate the phase of the unshifted site. The in-phase sine waves, with a wavelength of 10.6 bases, show that Fis sites shifted by 11 bases would be on the same face of the DNA [28,101,102], while the out-of-phase waves of Fis sites shifted by 7 bases indicate binding to opposite faces. Arrows are at positions where the logo is self-similar after a shift. Red arrows (pointing downwards from the +11 shift) mean that the contacts by Fis to the bases would interfere because they would be on the same face of the DNA. Green arrows (pointing upwards from the +7 shift) mean that the contacts could be simultaneous because they are on opposite faces. In a sequence logo, the height of each letter is proportional to the frequency of the corresponding base at that position in the sites, and the height of the stack of letters represents the sequence conservation in bits. For clarity, the sine waves run from 1 to 1.6 bits.

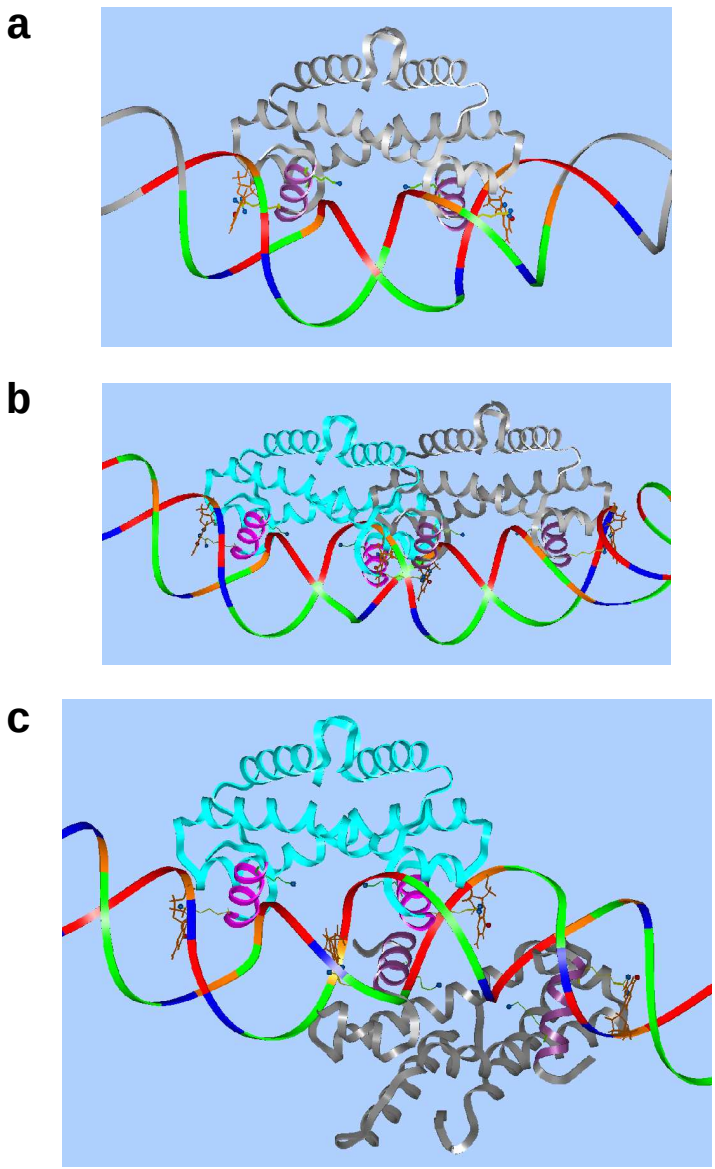


Fig. 3. Fis binding models.

(**a**) A single Fis dimer binding to DNA. (**b**) Two Fis dimers binding to Fis sites separated by 11 base pairs. (**c**) Two Fis dimers binding to Fis sites separated by 7 base pairs. The DNA backbone is color coded: A: green, C: blue, G: orange, T: red. The models of Fis interacting with DNA were built using Insight II software from Biosym Technologies, Inc., on an IRIS computer (Silicon Graphics, Inc.), and displayed with RasMol 2.5, available at http://molbiol.soton.ac.uk/rasmol.html or ftp://ftp.dcs.ed.ac.uk/pub/rasmol/. The Fis protein coordinates are those of the Protein Data Bank (http://www.rcsb.org/pdb/) entry 1fia. (See Materials and Methods for further details.)

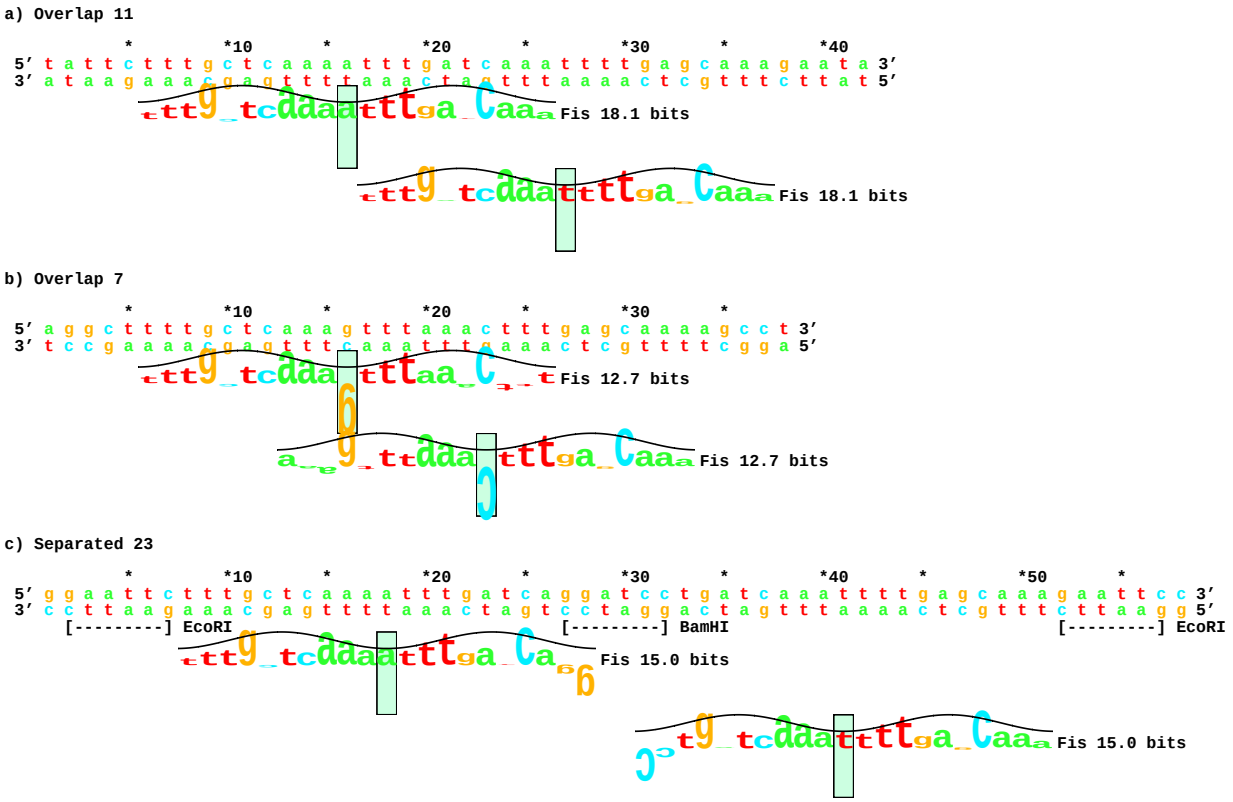


Fig. 4. Oligonucleotide design of overlapping and separated Fis binding sites.

The predicted Fis sites are shown by sequence walkers floating below each self-complementary DNA sequence [10,5]. In a walker, the vertical green box marks the zero base of the binding site. The box also shows the vertical scale, with the upper edge being at +2 bits and the lower edge being at −3 bits. The height of each letter is determined from the bit value in the individual information weight matrix [21,10,5]. Negative weights are represented by drawing the letter upside-down and placing it below the zero bit level. To indicate predicted relative orientations, the peaks of sine waves correspond to where Fis would bind into the major groove. Three DNAs were designed, each having two Fis sites spaced 11, 7 and 23 bases apart. Design details are given in Materials and Methods. The total strength of a site is the sum of the information weights for each base. The 18.1 bit Fis sites are 3.4 standard deviations higher than the average Fis site in natural sequences [5,21]. The 12.7 and 15.0 bit sites are 1.6 and 2.4 standard deviations above average (respectively).

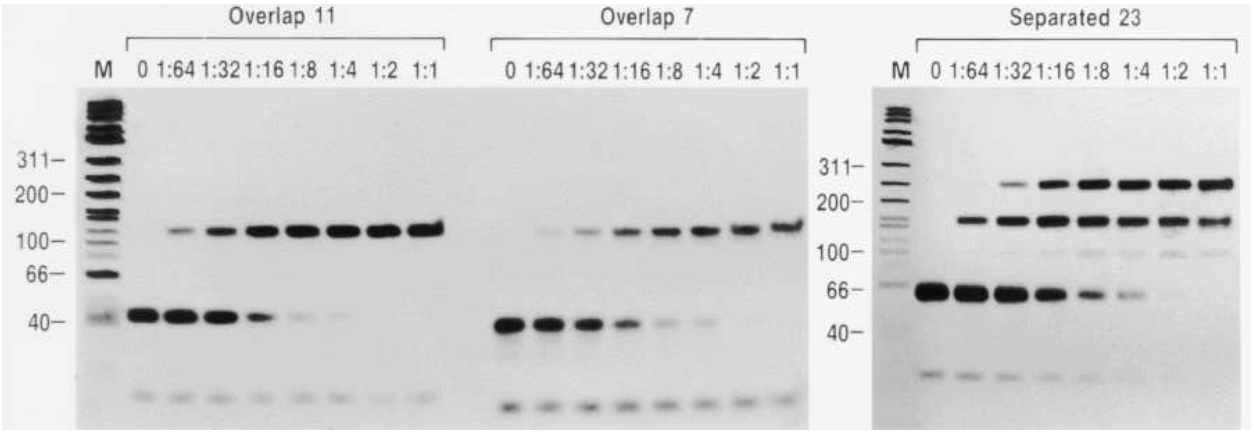


Fig. 5. Mobility shift experiments for 11 and 7 base pair overlapping and 23 base pair separated Fis sites.

Each lane contains increasing concentrations of Fis protein, beginning with no Fis, Fis diluted 1 to 64, etc. The 1:1 dilution was at 2200 nM Fis. This concentration was chosen intentionally so that with the 1 nM of DNA used in this experiment, the protein/DNA ratio was 2-fold higher than that needed to strongly shift DNA containing the 8.9 bit wild-type *hin* distal Fis site [29]. The sequences are given in Fig. 4. Marker lanes (M) contain 10 ng of biotinylated ϕX174 *Hinf* I digested DNA standards (Life Technologies, Inc.). Sizes are indicated in bp. The lowest band in most lanes of the figure is single-stranded oligonucleotide DNA. In the "Separated 23" experiment, at high concentrations, Fis proteins are apparently able to capture the single-stranded DNA when it has folded into a hairpin. This produces a faint band near the 100 bp marker.

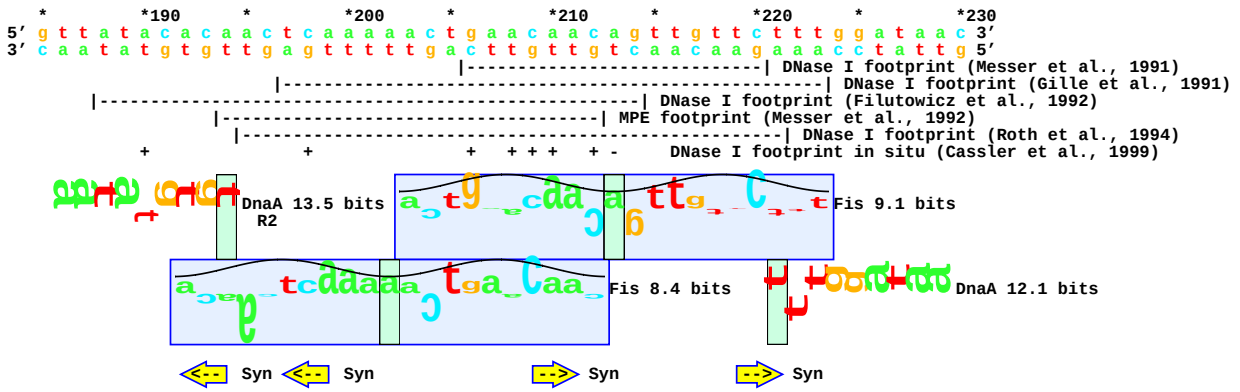


Fig. 6. Positions of Fis and DnaA sites at the *E. coli oriC* shown by sequence walkers. Sequence data are from GenBank accession K01789 [103]. The horizontal dashes below the sequence represent regions protected by Fis. Locations of DnaA sites are from [104] and Fis footprint data are from [104,52,66,54,53,105]. The asymmetric DnaA individual information matrix was created from 27 experimentally demonstrated DnaA binding sites [102]. DNA synthesis start sites are indicated by yellow arrows and 'Syn' [64], however start sites have also been mapped to the left side of *oriC* [69]. Blue boxes mark two Fis sites separated by 11 bases. DnaA site directionality is indicated by letters turned sideways in the direction that DnaA binds [10].

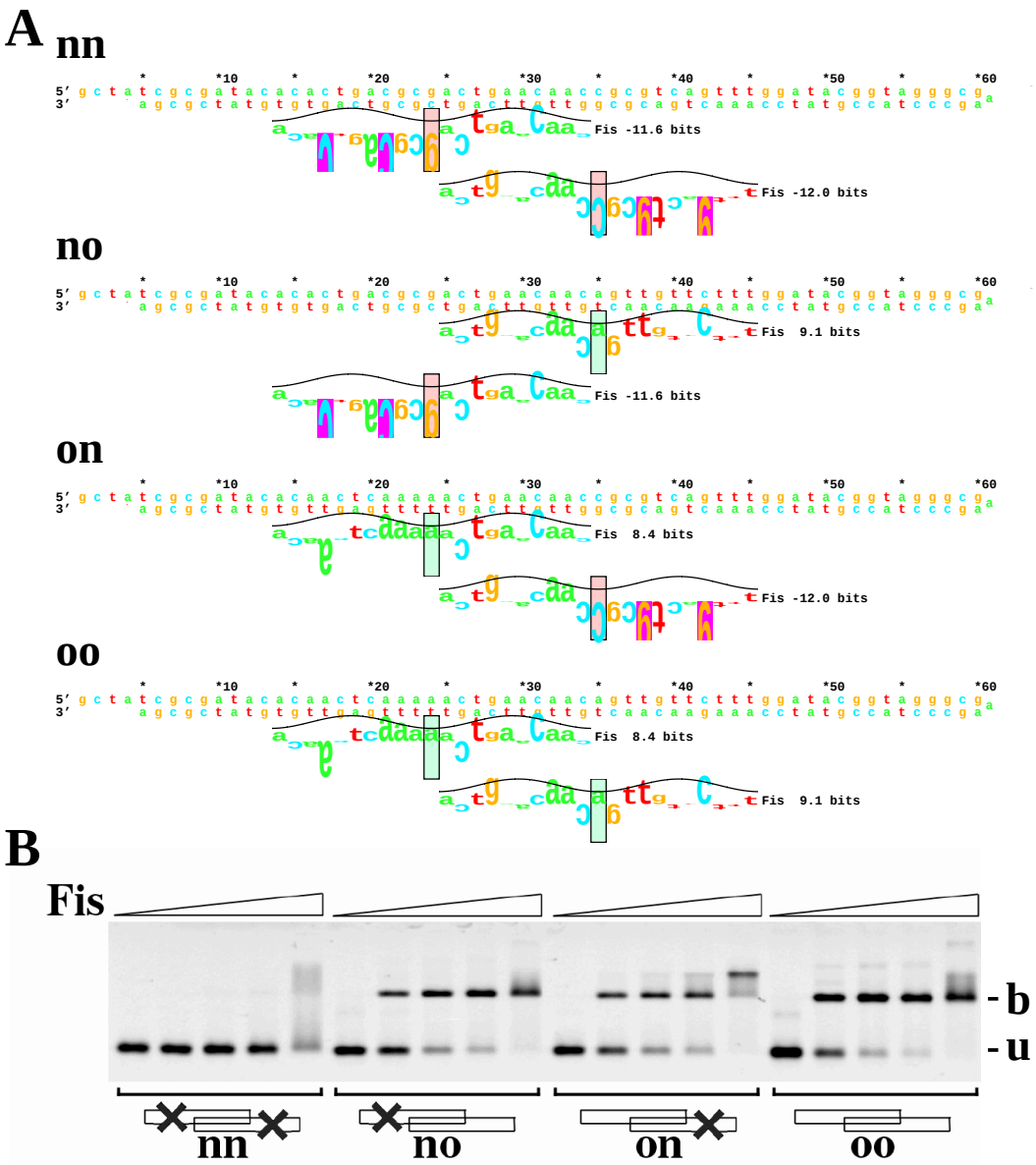


Fig. 7. *E. coli oriC* can bind only one Fis molecule at a time between DnaA sites R2 and R3.

A. Design of wild-type and mutated Fis sites from *E. coli oriC*. Four hairpin oligos were designed and designated **nn**, **no**, **on** and **oo** where **n** means no site because of engineered mutations (pink boxes, with information less than zero) and **o** means that there is a complete wild-type origin Fis site (green boxes, with positive information). For example, **no** contains only the Fis site closest to R3 on the right side.

B. Gel mobility shift assay with *oriC* sites using the oligos shown in part A at a concentration of 10 nM each. Fis concentrations were 0, 30, 100, 300 and 1000 nM. u: unbound DNA; b: Fis bound DNA.

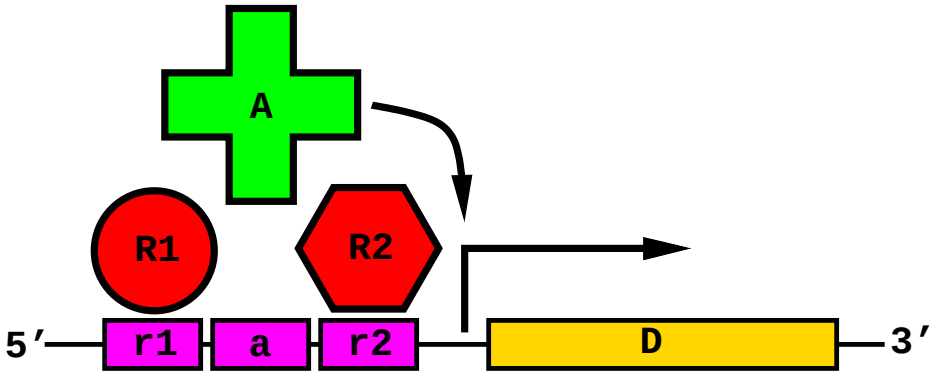
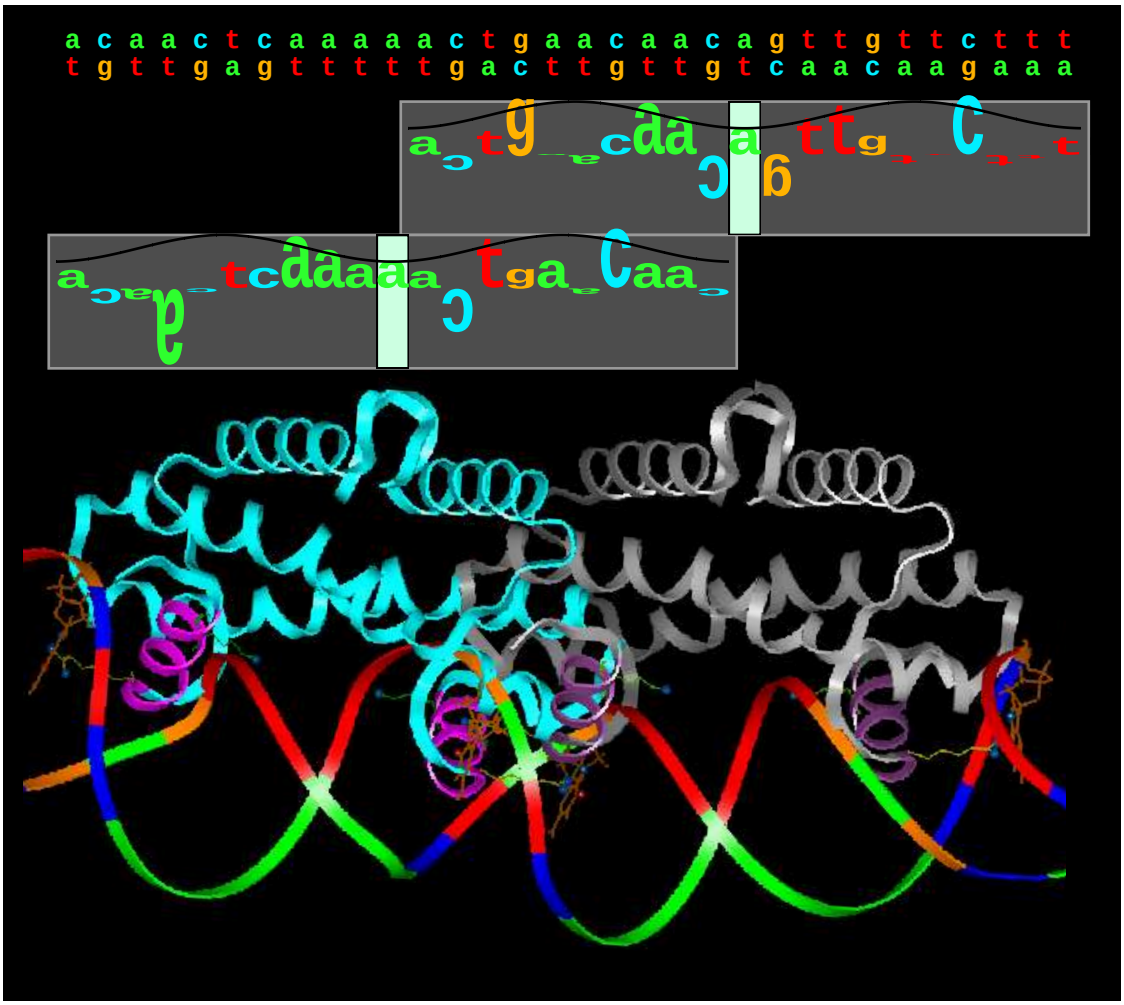


Fig. 8. NOR Gate molecular computer.

An activator protein molecule **A** (green plus) binds to a DNA molecule at position **a**. When the activator binds, it turns on the promoter for gene **D**. Two repressor protein molecules **R1** and **R2** (red circle and red hexagon respectively) bind to DNA at positions **r1** and **r2**. Binding to either **r1** or **r2** interferes with binding by **A**, so the activator can only bind when the two repressors are absent. Assigning the presence of a molecule as '1' or 'true' and the absence as '0' or 'false', then **D = R1 NOR R2**. By connecting such NOR gates together, any computer circuit can be built.



Proposed cover figure. "Fis binding sites form a bi-stable flip-flop in the *Escherichia coli* origin of replication."