Regulation of Transcription by DNA Bending, Twisting, Looping, and Topology Part II: In vivo approaches

Jason D. Kahn
Department of Chemistry and Biochemistry
University of Maryland, College Park

Experiments in Physical Biology

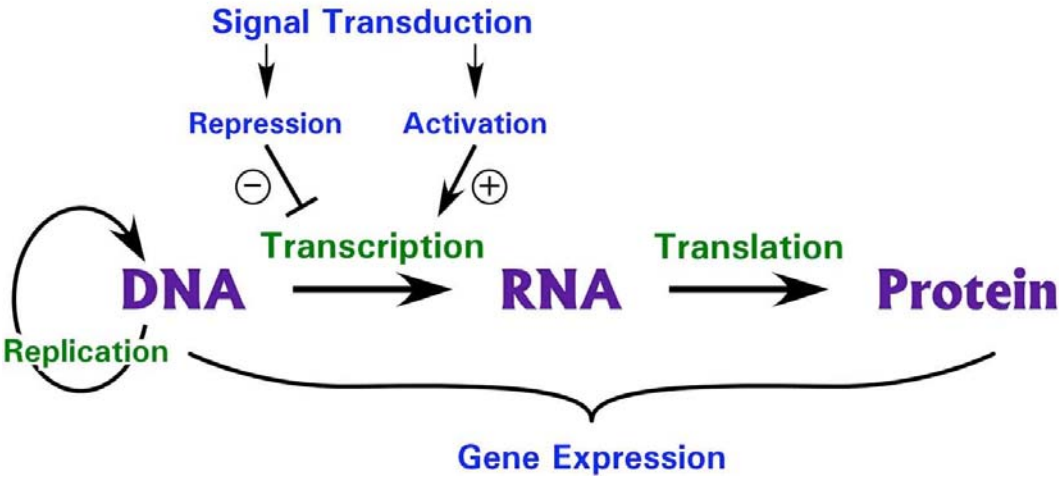
IMA, University of Minnesota

May 2-3, 2005

Kahn Lab Research

- **Design and construct interesting DNAs.**
- **Characterize their interactions with proteins: bending, twisting, writhing, looping, topology**
- **Goals:**
  - **Better methods for large-scale structure.**
  - **What are some of the structures?**
  - **Functional consequences of changes in structure?**
  - **Design new systems.**
- **Techniques:**
  - **Design and construction of reporter strains**
  - **Quantitation of reporter gene activity**
- **Systems:**
  - **Lac repressor looping**
  - **Transcriptional activation by Nitrogen Regulatory Protein C (NtrC)**

Central Dogma of Molecular Biology



J. Kahn

Review of in vitro techniques

- Design of interesting DNA:
  - Model, synthesize and/or clone short DNAs containing sequence-directed bends, protein binding sites, internal loops, fluorophores, controlled length, circularity
- Characterize DNA shape
  - Measure J factor for cyclization ± ligands using kinetics of T4 DNA ligase mediated cyclization vs. bimolecular reactions
  - Assess product topology (linking number difference w.r.t. relaxed), typically by electrophoresis in the presence of intercalators
  - Bulk and SM FRET between fluorophores measures relative distances
- Characterize Protein binding
  - Electrophoretic mobility shift analysis can provide relative binding constants, absolute binding constants with some care and luck, information re bending and looping
  - Footprinting: observe protection against cleavage agents due to occlusion by bound protein
- Guiding Principle: Any DNA that is preorganized so as to minimize the free energy needed to deform into the geometry needed to bind protein will form a more stable complex with the protein. And vice versa. And the converse goes for protein deformation.
- Examples: TBP binding to supercoiled DNA, LacI to pre-bent loops

J. Kahn

In vivo Techniques

- Design of interesting DNA:
  - Model, synthesize and clone DNA fragments containing sequence-directed bends, protein binding sites. Typically we are interested in promoter sequences that drive transcription.
  - Systematic deletion of upstream control regions in order to identify the proteins that act at a particular gene is called “promoter bashing.”
- Characterize DNA function
  - Measure the activity of a “reporter gene” linked to the promoter sequence of interest
  - Readout is typically through a chromogenic reaction like X-gal hydrolysis to give blue bacterial colonies, or *o*-Nitrophenylgalactoside hydrolysis to give a yellow product
  - Other common reporters include chloramphenicol acetyltransferase, luciferase, green fluorescent protein, or an essential gene that allows the cell to survive.
- Characterize protein effect on DNA shape and function
  - Introduce or delete proteins that are expected to bind to the DNA. Use “expression vectors” to make the protein in vivo. Often express mutant or truncated proteins.
  - Some proteins directly affect RNA polymerase or “recruit” coactivators and corepressors that control chromatin structure or RNA polymerase
  - “Architectural” factors act on DNA structure to allow other proteins to make functional interactions. Examples include bacterial HU and IHF proteins, eukaryotic HMG proteins.
  - Introduce or remove ligands that control the activity of binding proteins. Examples: IPTG is an “inducer” of the lac operon, acts by reducing the affinity of Lac repressor for operator DNA. Removal of ammonia induced activity of NtrC.
- Ideally, should control amounts of protein and DNA binding site in the cell. This can be difficult and is often ignored, which is problematic if stoichiometry is important.

J. Kahn

Complex example

BAMBED, Vol. 30, No. 6, pp. 353–362, 2002

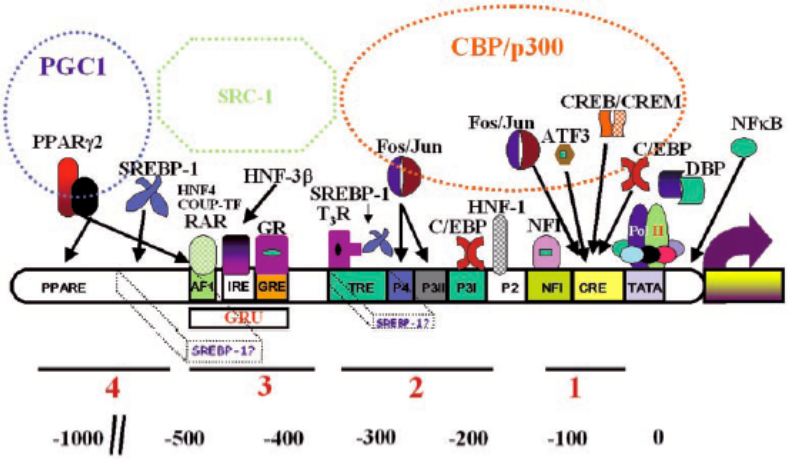


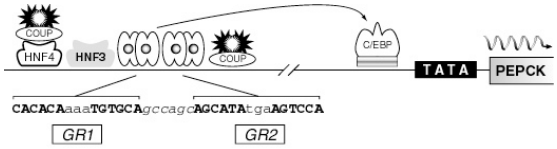
FIG. 1. **The PEPCK-C gene promoter showing the transcription factors and coactivators that regulate the transcription of the gene.** *T_3R*, thyroid hormone receptor; *DBP*, D-binding protein; *CREM*, cAMP regulatory element modifier; *RAR*, retinoic acid receptor; *TRE*, thyroid hormone regulatory element; *Pol II*, RNA polymerase II; *GRE*, glucocorticoid regulatory element; *P2*, *P3(I)*, *P3(II)*, and *P4*, are protein binding sites identified by DNase I footprinting (40); *PPARE*, peroxisome proliferator-activated receptor element; *IRE*, insulin response element.

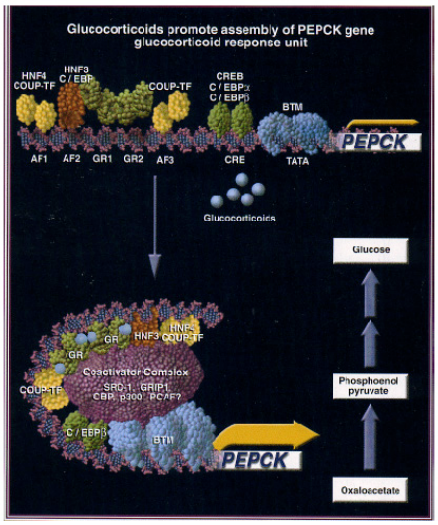
- Review from R. Hanson’s group on PEPCK (phosphoenolpyruvate carboxy kinase) regulation.

J. Kahn

Connection to Shape

- Structural models are inferred from known or imagined structures of individual proteins bound to DNA.
- This kind of complex is called an “enhanceosome.” Not clear whether or all components really need to be present simultaneously.
- We study simpler systems!





J. Kahn

R. Miesfeld, www.biochem.arizona.edu/classes/bioc460/spring/rlm/RLM40.3.html

Exploring the Role of DNA Flexibility and Architectural Proteins HU and HMGB2 in Lac Repressor Looping

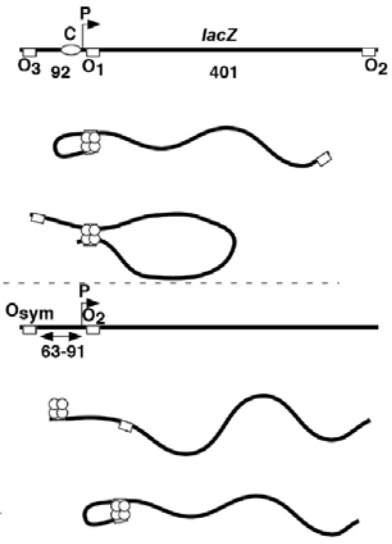
Concept, Design and Experiments by Nicole Becker and Jim Maher (Mayo Clinic)
Statistical Weights/DNA Mechanics Model by J. Kahn
Becker, Kahn, and Maher, JMB, in press 2005

J. Kahn

Lac Repressor Looping in vivo

- Related systems were pioneered by the groups of Benno Müller-Hill and Tom Record: demonstrated Lac looping, torsion and distance dependence.
- Here it is applied to the role of architectural proteins.

Wild Type Lac Promoter:
The strong O_1 operator blocks transcription. The weak O_2 and O_3 operators upstream and downstream loop. The CAP activator binds at C.

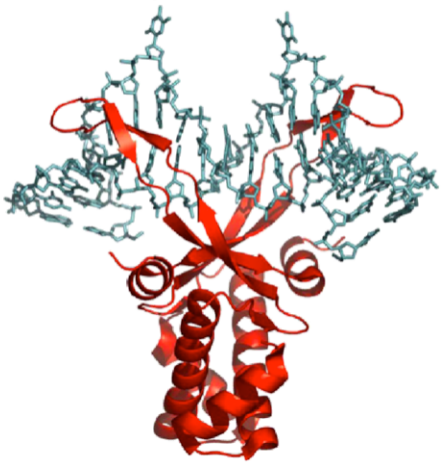


Simplified Test System:
Make a collection of promoter variants with different operator spacings. The weak O_2 operator is the one that blocks transcription. The strong O_{sym} operator is upstream. The strong *lac* UV5 promoter removes the need for CAP. This maximizes sensitivity to looping.

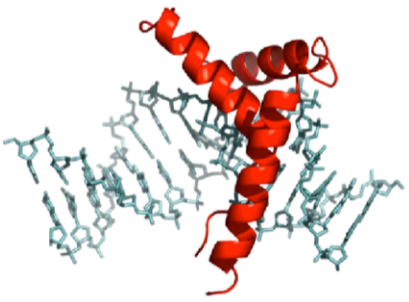
J. Kahn

Probe Role of Architectural Proteins

- Introduce promoter constructs into cells that contain one, both, or neither of the architectural proteins shown.
- Measure β-galactosidase (LacZ) activity ± inducer IPTG
- How do these proteins affect efficiency of repression?



Semi/non-specific *E. coli* HU DNA bending protein



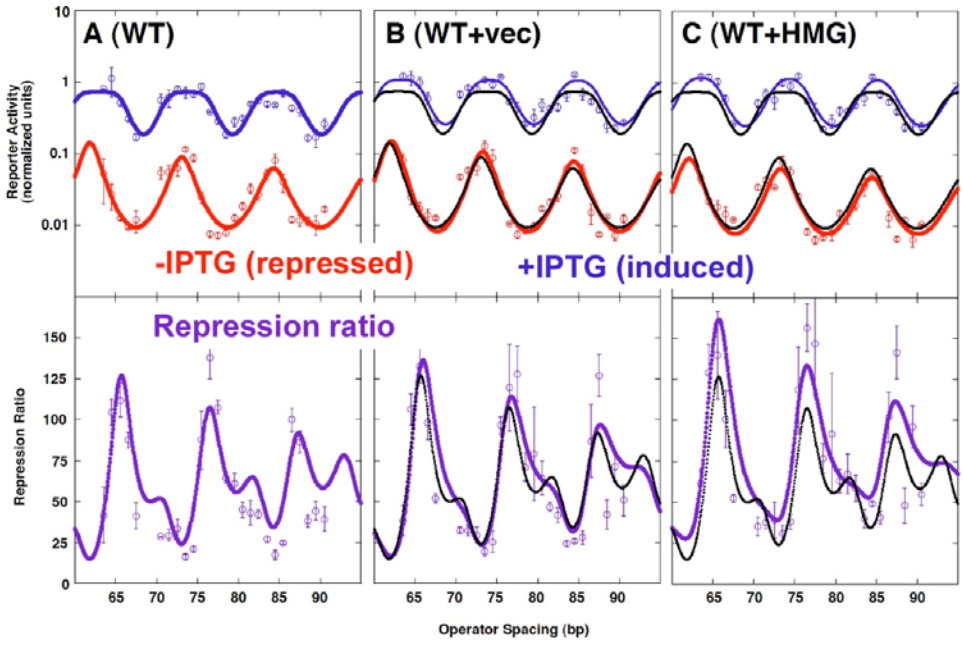
Semi/non-specific human HMGB1 DNA bending protein

Can it substitute?

J. Kahn

Experimental Data

- See obvious torsional oscillations



J. Kahn

Where's the Applied Mathematics?

- The partition function for the system is the sum of the possible states of the O_2 operator:

$$[Free]+[Specific\ Loop]+[NS\ Loop]+[Single\ bound]=[O2]$$

- This is expressed in terms of the equilibrium constants for different states as follows:

$$[Free]\left(1+K_{SL}+K_{NSL}+K_{O2}\right)=[O2]$$

$$\text{where } K_{SL}=\frac{[Specific\ Loop]}{[Free]},\ K_{NSL}=\frac{[NS\ Loop]}{[Free]},\ \text{and } K_{O2}=\frac{[Single\ bound]}{[Free]}.$$

where we have absorbed the constant cellular concentration of lac repressor into each of the three equilibrium constants.

- How do we measure and model the equilibrium constants?

J. Kahn

Analysis of Repression Data

- The experimental fraction bound is given by

$$f_{bound} = \frac{\text{max induced activity - observed activity}}{\text{max induced activity}}$$

- The theoretical fraction bound is given by

$$\begin{aligned} f_{bound} &= ([O2]-[Free])/[O2] \\ &= ([Specific\ Loop]+[NS\ Loop]+[Single\ bound])/[O2] \\ &= \left(K_{SL}+K_{NSL}+K_{O2}\right)/\left(1+K_{SL}+K_{NSL}+K_{O2}\right) \end{aligned}$$

- Only K_{SL} depends on inter-operator distance. The other K's are amplitudes and offsets.

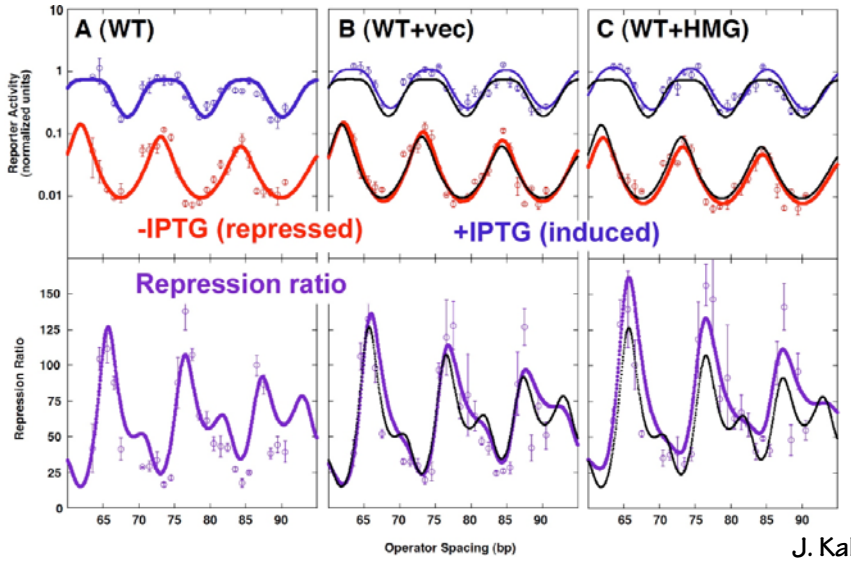
$$K_{SL} = \sum_{i=-5}^{5} K_{\max} e^{-(sp-sp_{optimal}+i\cdot hr)^2/2\sigma_{Tw}^2}$$

- We can estimate an apparent in vivo twist flexibility for the protein-DNA loop system. In this case the data did not show distance dependence, so the apparent persistence length is too short to measure.

J. Kahn

Analysis of Experimental Data

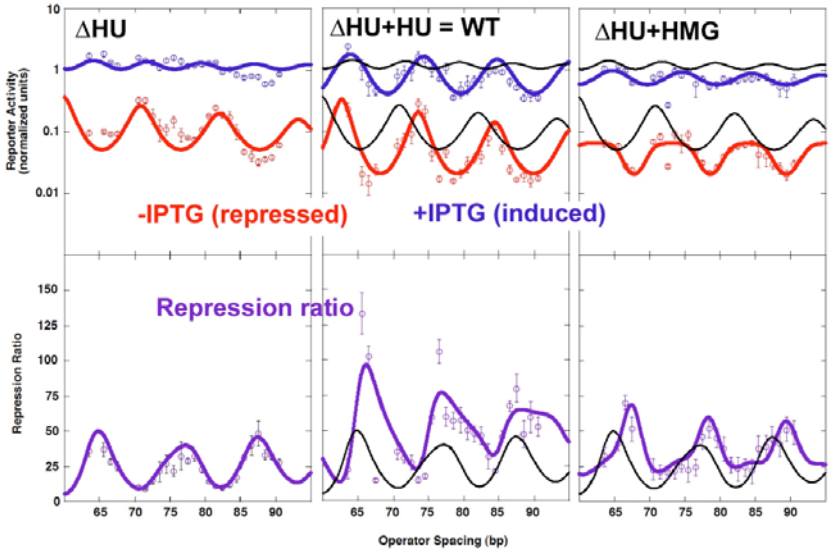
- Top curves are upside-down sums of Gaussians
- Note even induced promoter is still partially repressed via looping!
- Spacing/helical repeat changes ±IPTG give shoulders in repression ratio. May be due to altered loop geometry or dynamic supercoiling effects?



J. Kahn

Can HMGB1 substitute?

- HU deletion strain shows little or repression when induced: bending flexibility does matter.
- HMGB1 enhances loop-dependent repression, but also enhances non-phase dependent repression.



J. Kahn

Lac Loop Conclusions

- As we have seen, DNA in vivo is surprisingly flexible. Evolution has had a long time to work out how to shape steel.
- In vivo loops can be modeled, but we need to consider both specific and non-specific effects of architectural proteins. They are not completely interachangeable.
- Induction is only relative: loops are stable enough to persist.

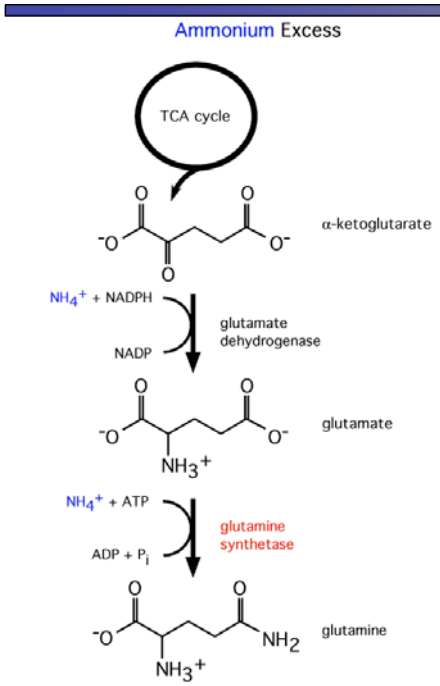
J. Kahn

Exploring the Role of DNA Geometry in Transcriptional Activation by the *E. coli* NtrC Protein, Using a Semi-random DNA Shape Library and Designed Variants

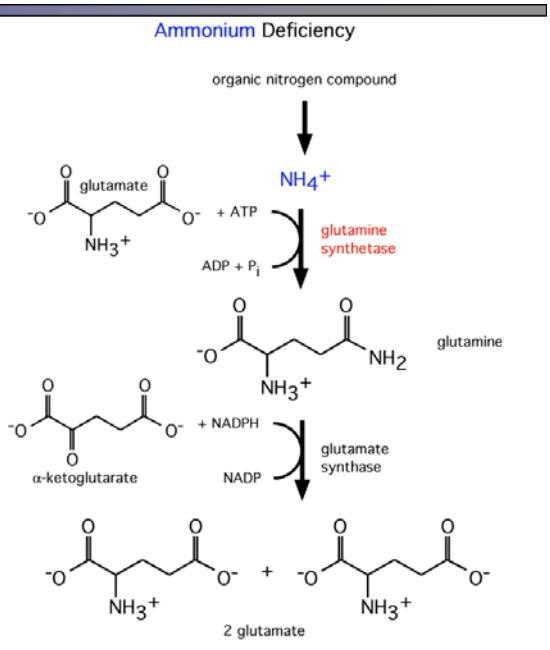
Lilja, Jenssen, and Kahn, JMB 2004

J. Kahn

Nitrogen Metabolism in *E. coli*



Net: α-KG+NH_4^++NADPH -> Glu

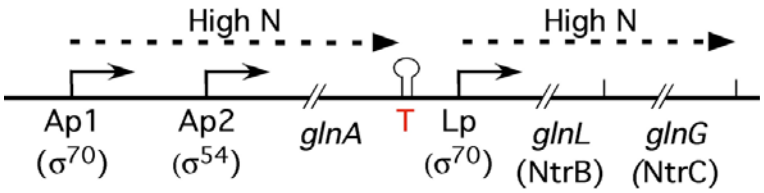


Net: α-KG+NH_4^++NADPH+ATP -> Glu

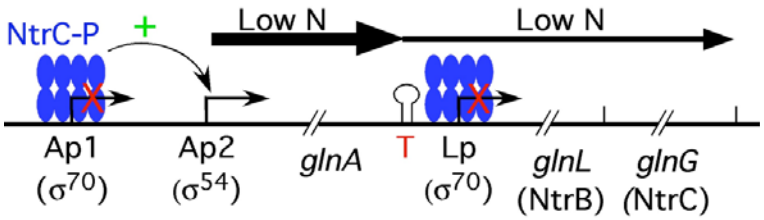
J. Kahn

The *E. coli glnALG* Operon

Under normal conditions: Enough GlnA to make Gln

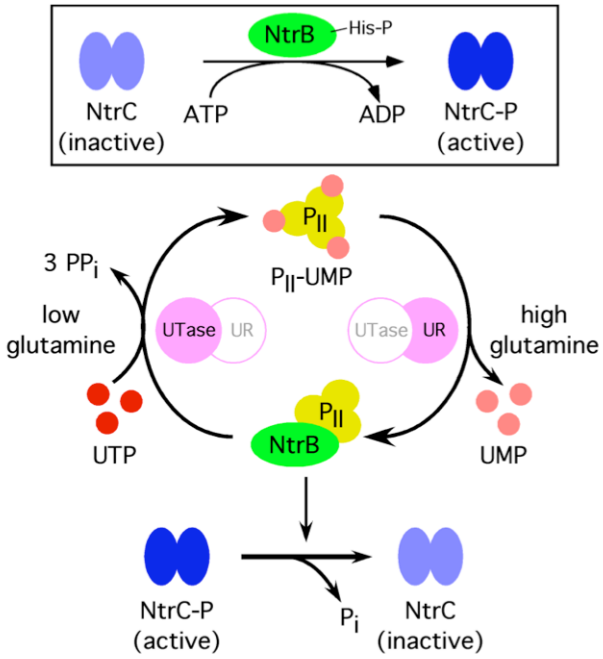


During nitrogen starvation: GlnA needed to take up NH_4^+



J. Kahn

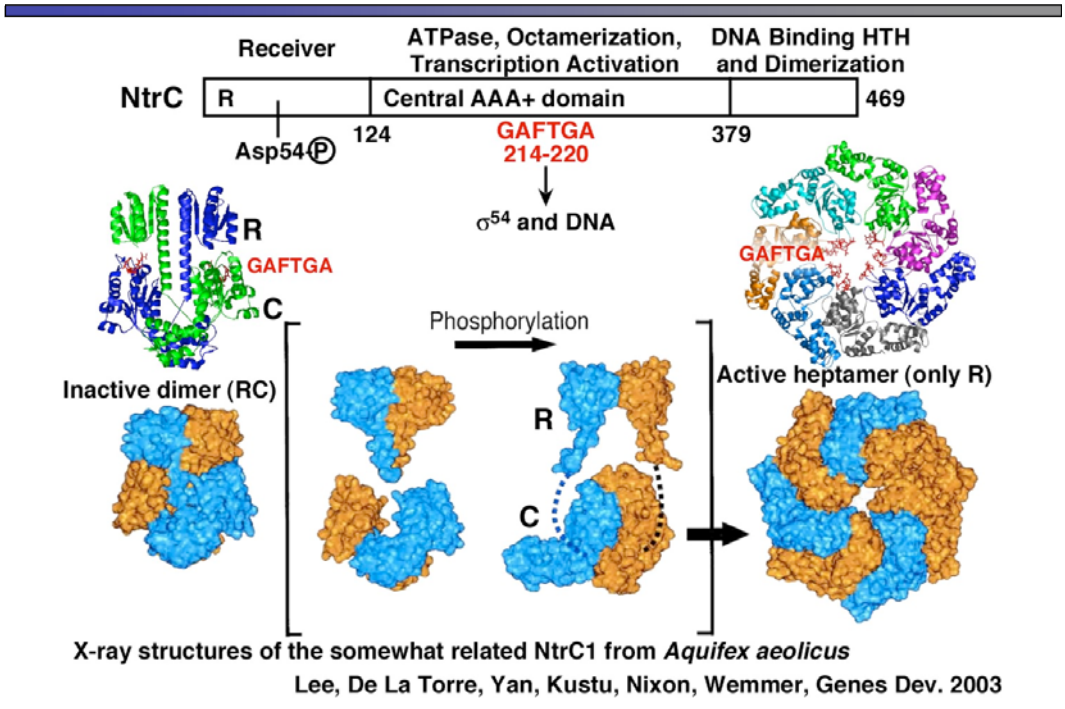
NtrC is Part of a Two-Component Regulatory System



Adapted from: Magasanik and Neidhardt (1987), E. coli and S. typhimurum, pp1321

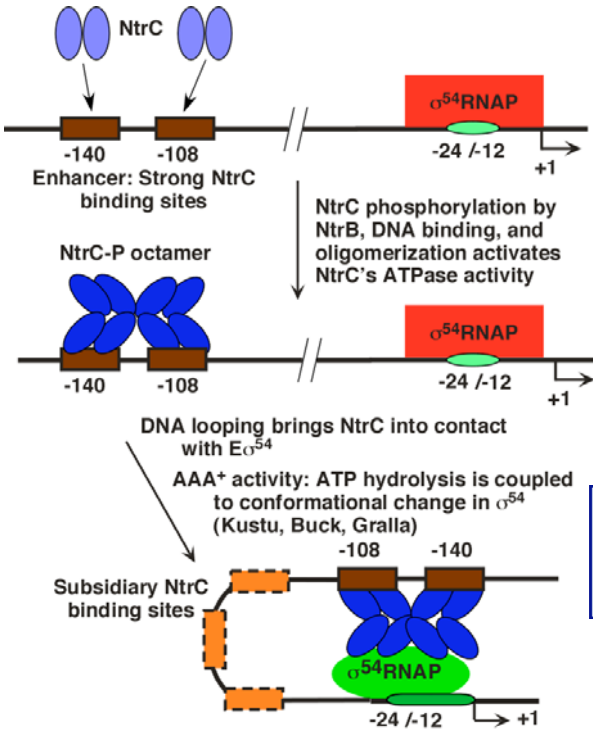
J. Kahn

Domain Organization of NtrC



J. Kahn

Transcription Initiation at the *glnAp2* Promoter

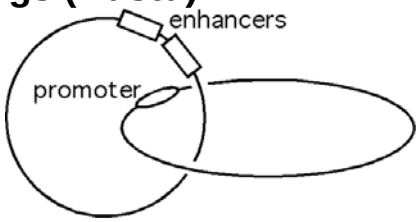


What is the role of DNA structure in looping and transcription activation?

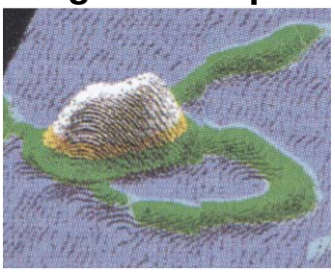
J. Kahn

DNA Looping at *glnAp2*

- **NtrC sites can function from >500 bp upstream or downstream of the promoter (Magasanik)**
- **Promoter and enhancers can interact when located on separate catenated rings (Kustu)**

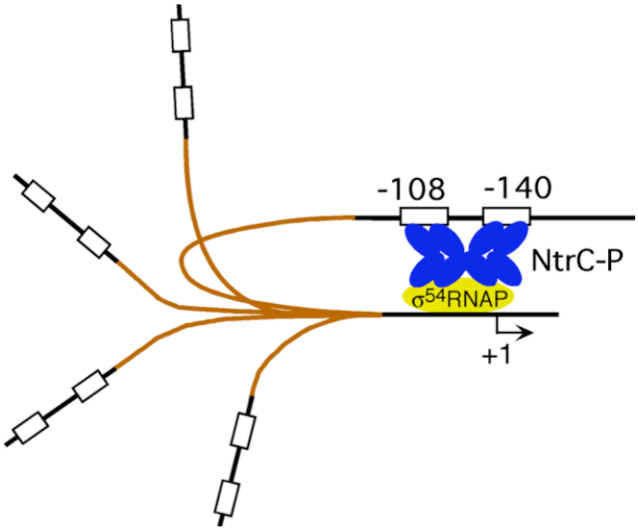


- **EM and AFM images of loops (Rippe, Bustamante)**



J. Kahn

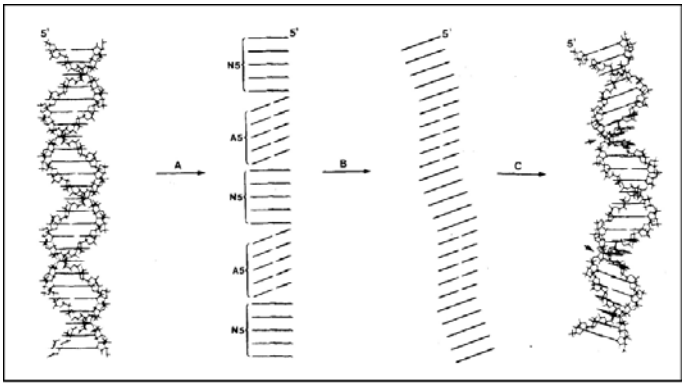
Experimental Approach to Optimum Geometry



- Make a library of bent DNA looping regions
- Isolate optimal transcription templates in vivo

J. Kahn

A-Tracts: Intrinsically Bent DNA



Crothers et.al. (1990). *J. Biol. Chem.*

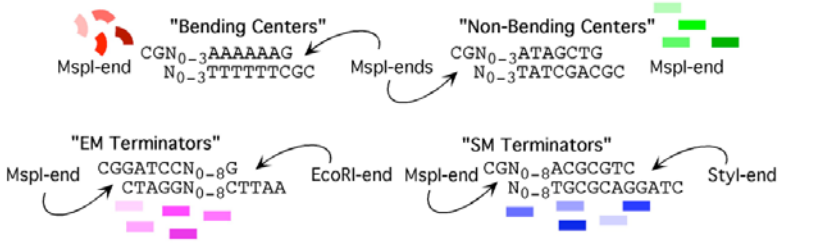
Steps A-C are equivalent to compressing the minor groove at the centers of the A-tracts.

A-tracts repeated in phase with the helical repeat add up to give substantial DNA bends

J. Kahn

Construction of DNA Fragments with Different Geometries

We used double-stranded oligonucleotides with different lengths, shapes, and overhangs...

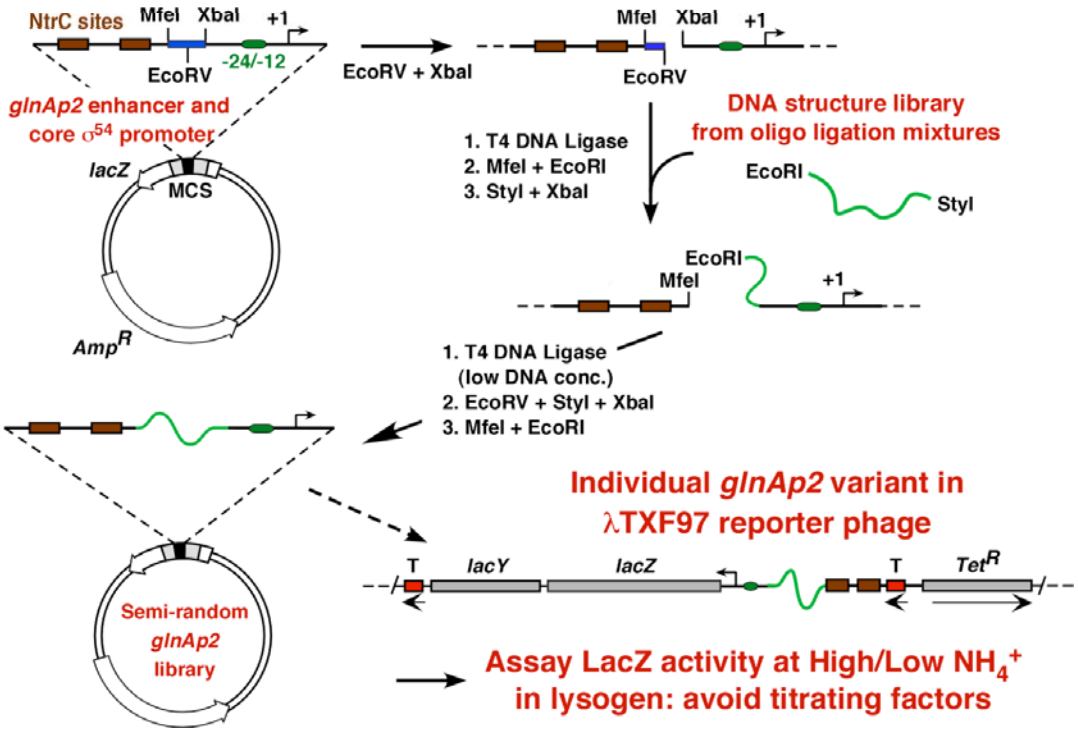


...and polymerized them...

...to form a library of DNA molecules with random length and structure

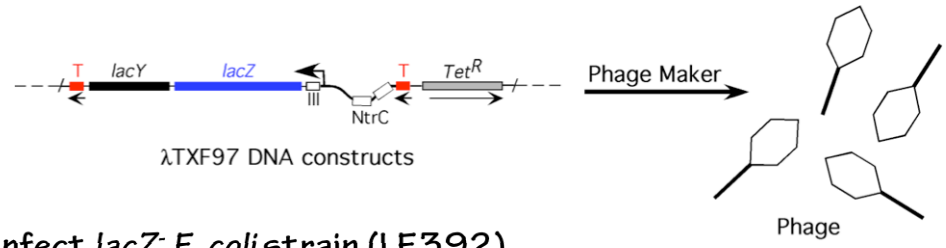
J. Kahn

Forced Ligation Approach to Library Cloning



J. Kahn

Characterization of Promoter Constructs



- Infect *lacZ*⁻ *E. coli* strain (LE392)
- Select lysogens on Tetracycline
- Grow lysogens in LB or medium with limited nitrogen to induce the NtrC, present at wild-type levels
- Assay β-galactosidase activity under both growth conditions
- Map transcription start sites with S1 Nuclease; increased transcription is due to initiation from *glnAp2*, not *glnAp1*
- Sequence phage DNA
- Model the shapes of all promoter constructs

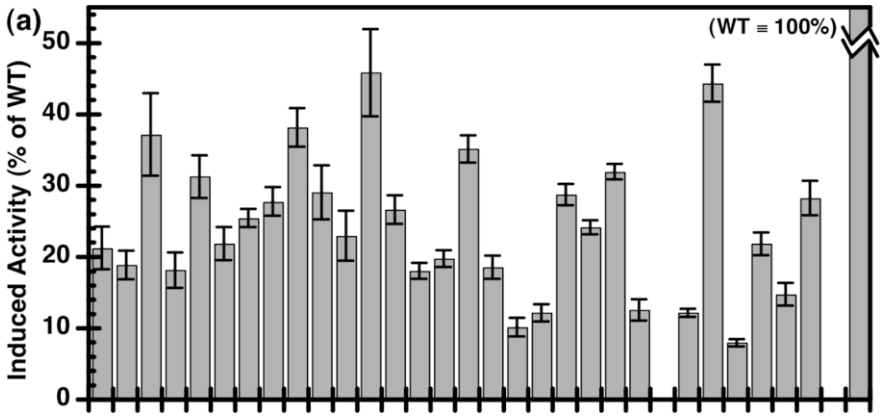
J. Kahn

Examples of Collected Data

Construct	Limiting Ammonium (Miller Units)	Abundant Ammonium (Miller Units)
Wild Type	2143 ± 437	589 ± 110
AL74B	259 ± 16	86 ± 5
AL76Q1	596 ± 46	201± 6
AL78C	450 ± 40	134 ± 19
AL81Q2	575 ± 65	159 ± 11
AL82Z	450 ± 23	100 ± 5
AL84Y	637 ± 43	176 ± 15
AL85U	497 ± 13	128 ± 4
AL93	542 ± 31	128 ± 3
AL94	745 ± 41	196 ± 3
AL101	353 ± 9	120 ± 11
AL195A	279 ± 22	123 ± 15
AL197A	556 ± 22	164 ± 7
AL198E	638 ± 34	205 ± 10
AL199E	433 ± 13	175 ±6
AL200E	473 ± 12	187 ± 9
AL201E	840 ± 7	180 ± 4
AL202P	445 ± 31	180 ± 11

J. Kahn

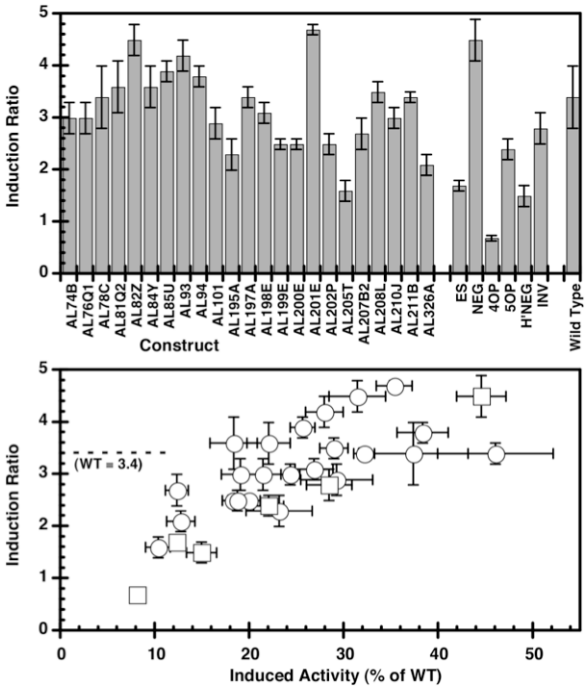
Transcriptional Activity



- Four-fold range of induced activity
- No random construct has >50% of wild type activity
- Does activity correlate with geometry?

J. Kahn

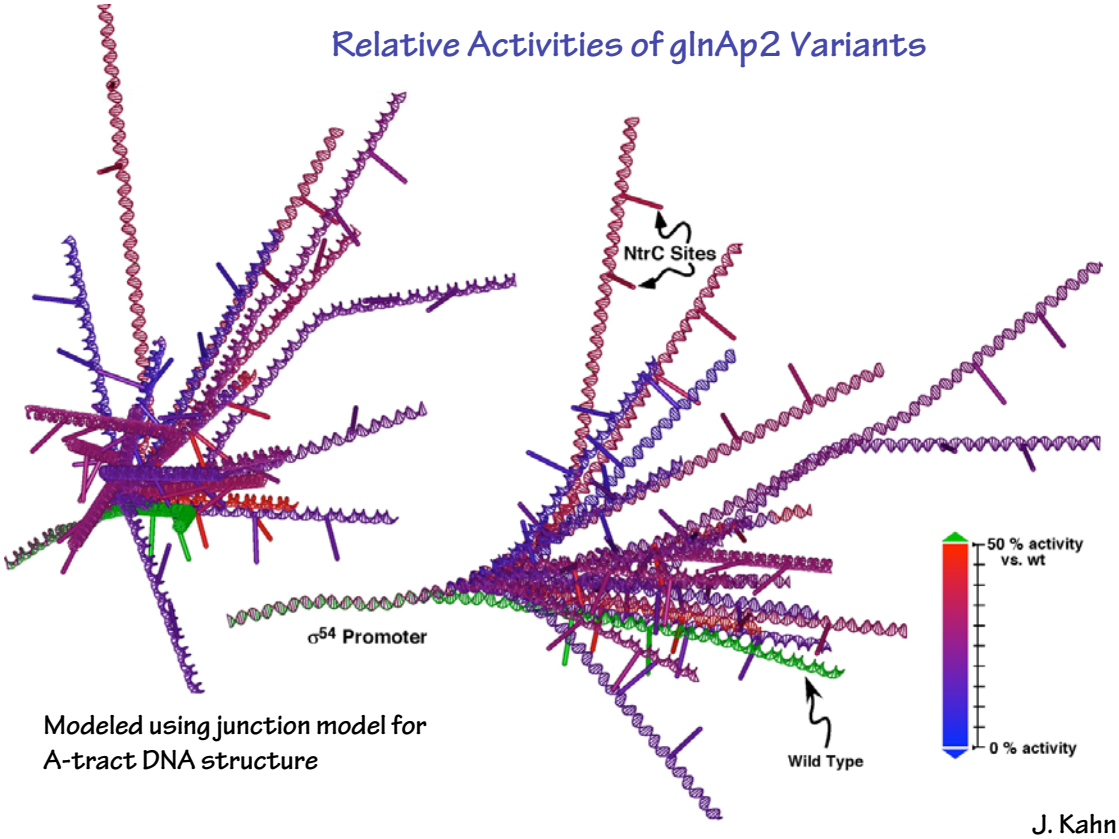
Induction ratios



- Except for the most inactive promoters, induction ratio is ~constant and ~wt.
- S1 mapping shows that ~70% of basal transcription is from *glnAp1*, >90% of induced transcription is from *glnAp2*
- Suggests mechanistic connection between *glnAp1* and *glnAp2*?
- For example, closed complex at *glnAp2* could block polymerases from upstream. Then occasional activation of *glnAp2* would allow *glnAp1* transcription to proceed.

J. Kahn

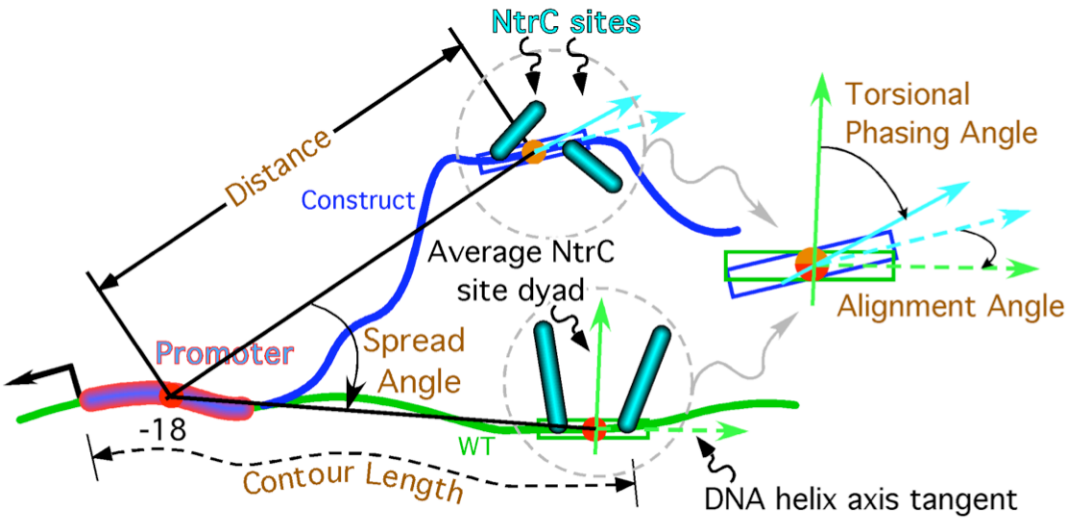
Relative Activities of glnAp2 Variants



Modeled using junction model for A-tract DNA structure

J. Kahn

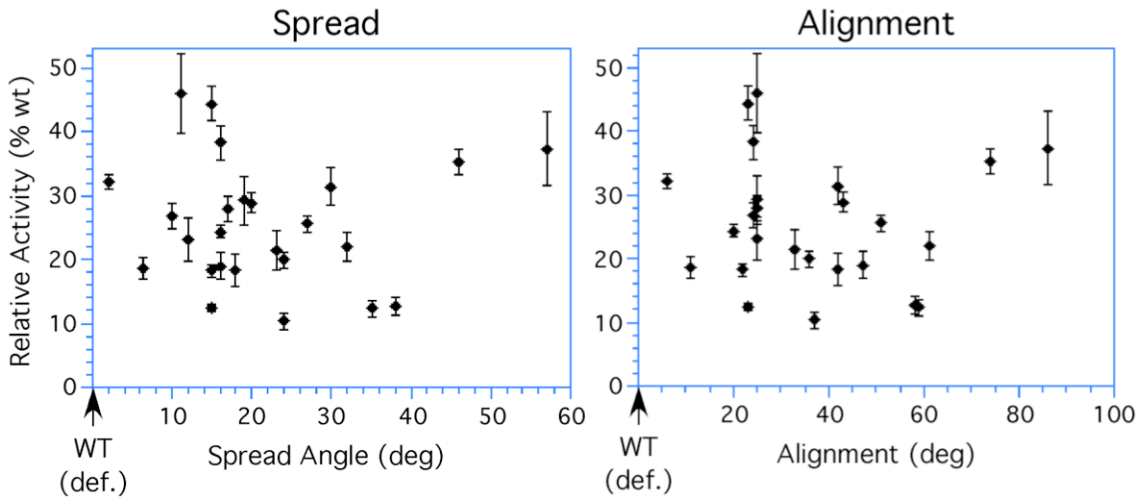
Geometric Definitions



Ignores some minor issues:
Supercoiling, likely presence of DNA packaging proteins, macromolecular crowding.

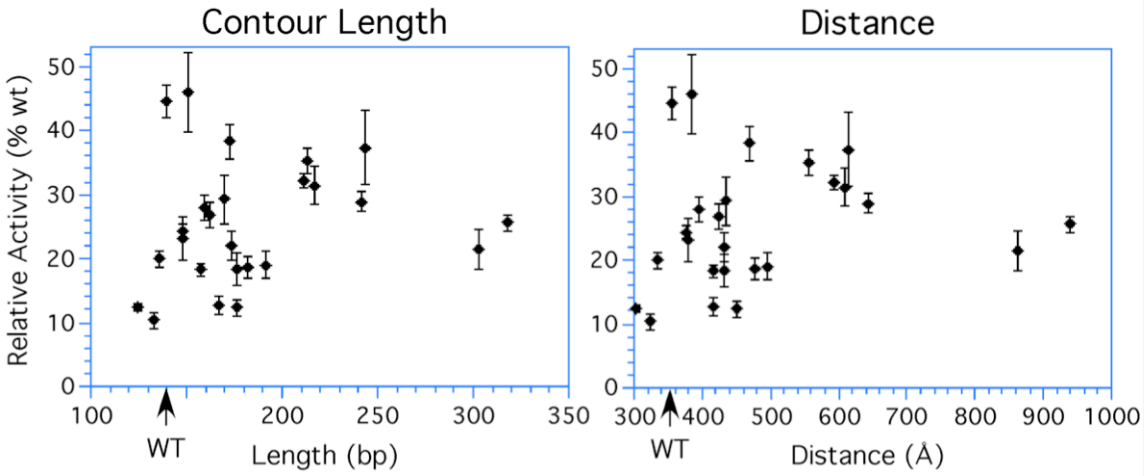
J. Kahn

Activity vs. Positioning or Alignment of the Enhancers



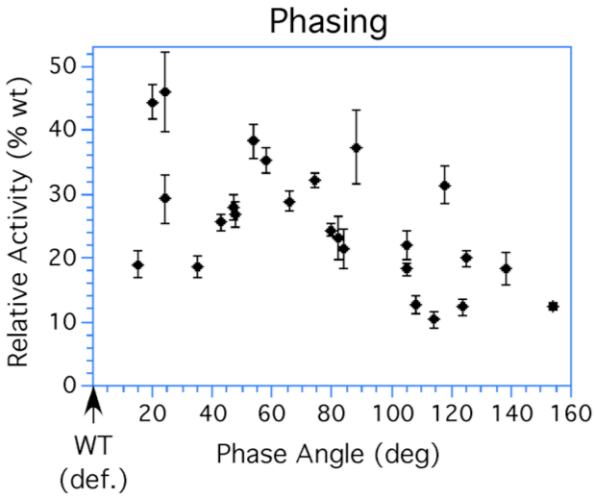
J. Kahn

Activity vs. Promoter-Enhancer Separation



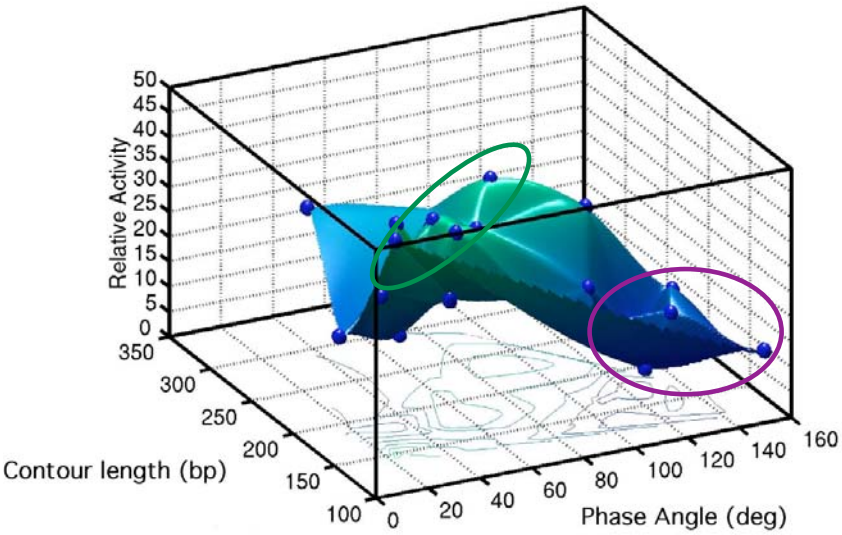
J. Kahn

Activity vs. Relative Torsional Phasing of Enhancers



J. Kahn

Preferred Geometry of Active Promoters?

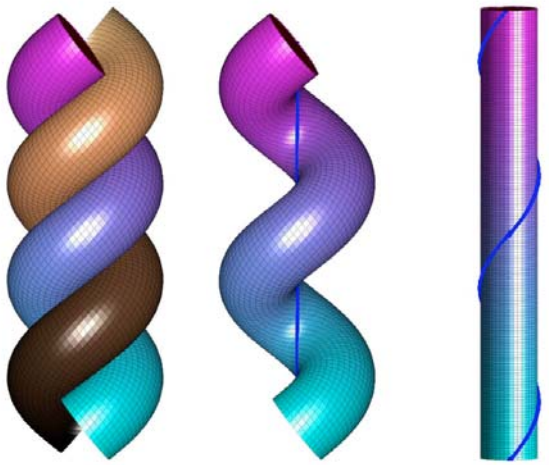


- Short, torsionally mismatched promoters are inactive
- Otherwise there is only subtle dependence on geometry
- Increase in optimum phase angle with length is consistent with apparent helical repeat increase in supercoiled DNA

J. Kahn

Apparent Helical Repeat in Supercoiled DNA

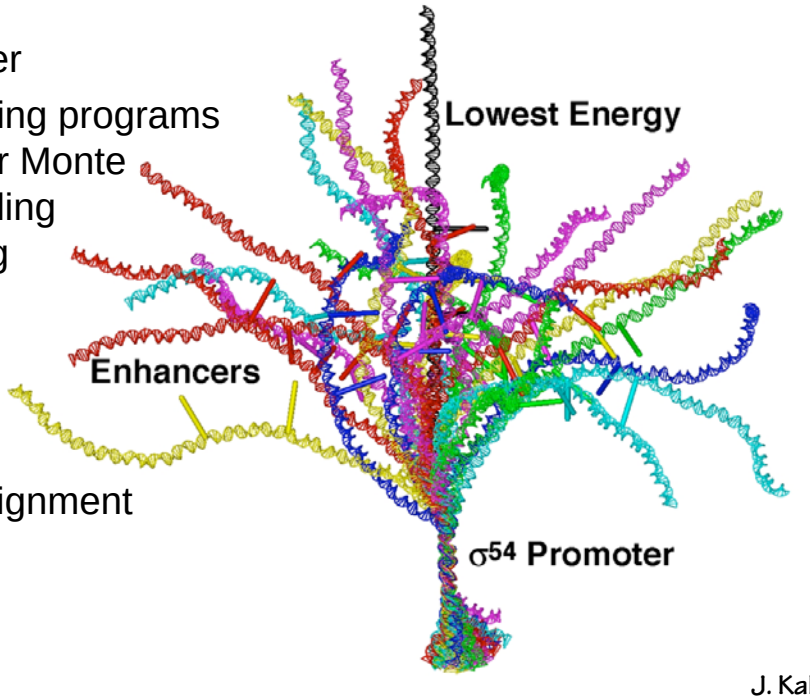
- DNA duplexes in a plectoneme trace out helical paths on each other.
- Therefore must cover more distance to reach an apposed location.
- Leads to apparent twist decrease (Record, Schleif), i.e. increased helical repeat.



J. Kahn

Thermally Accessible Ensemble

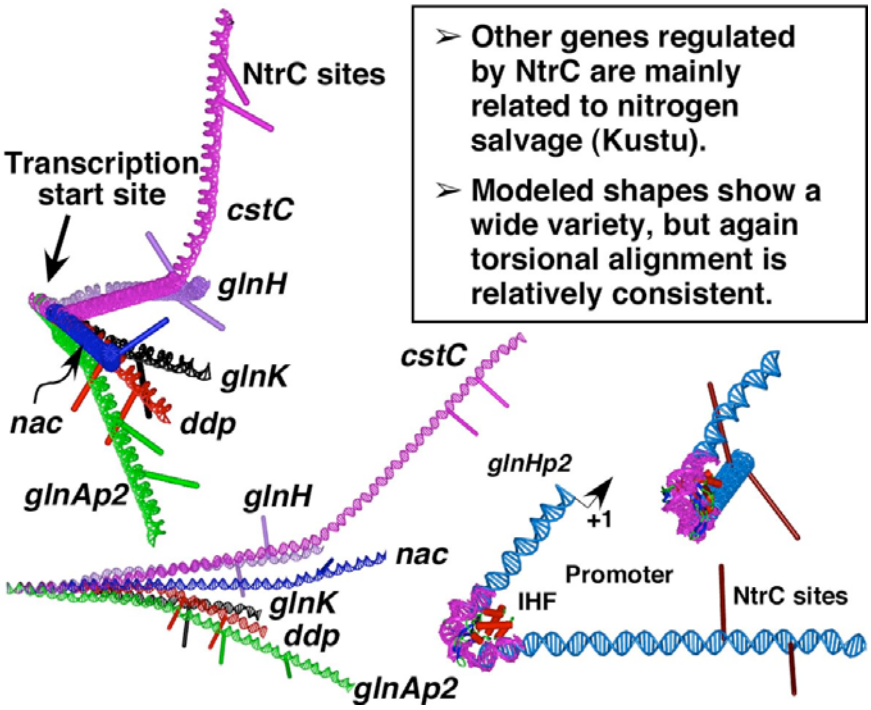
- WT promoter
- Modeled using programs designed for Monte Carlo modeling for DNA ring closure
- Wide variety of shapes
- Torsional alignment maintained



J. Kahn

Geometry Generality?

- Other genes regulated by NtrC are mainly related to nitrogen salvage (Kustu).
- Modeled shapes show a wide variety, but again torsional alignment is relatively consistent.



J. Kahn

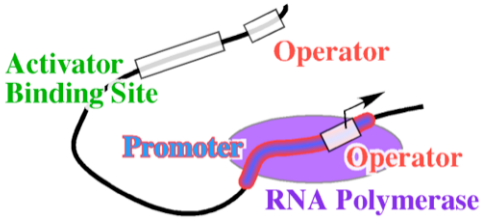
Connections

- Rippe's laboratory has found no systematic effects of DNA structure on activation *in vitro*, upon inserting curved sequences into the *glnAp2* promoter region
- In contrast, other model systems for DNA looping, such as repression of transcription by AraC or LacI have been shown to be strongly dependent on DNA helical phasing, weakly dependent on distance, and sensitive to DNA bending

J. Kahn

Activation vs. Repression

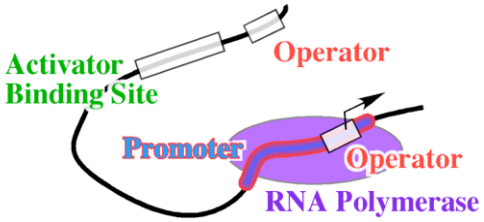
A promoter that is initially "firing" 1 % of the time…



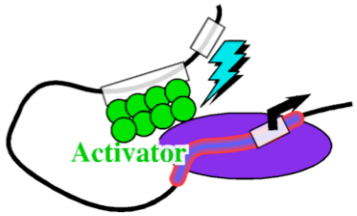
J. Kahn

Activation vs. Repression

A promoter that is initially "firing" 1 % of the time…



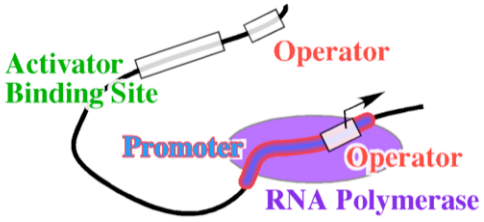
can be activated 10-fold by an activator that contacts the target RNA polymerase only 9% of the time, assuming that it activates whenever it is in contact:



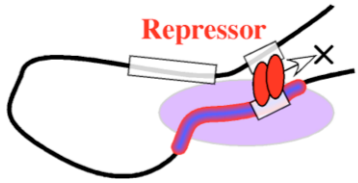
J. Kahn

Activation vs. Repression

A promoter that is initially "firing" 1 % of the time…



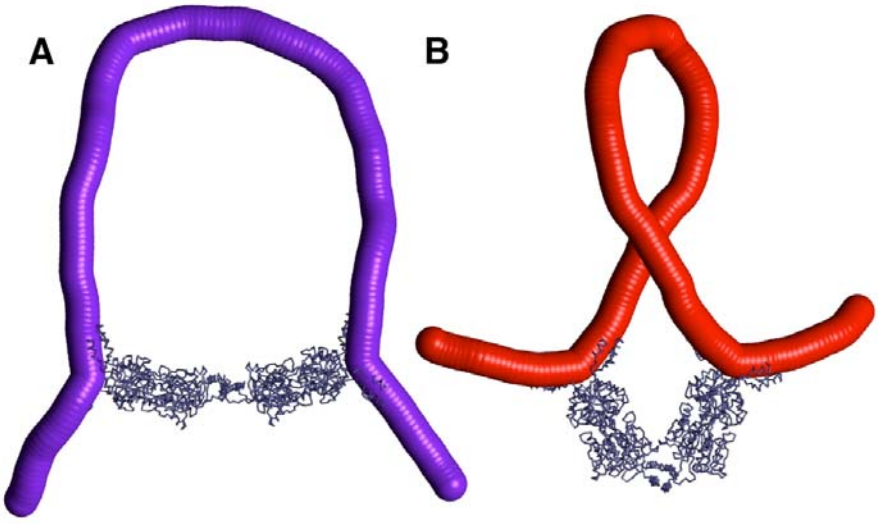
must be contacted by a repressor 90% of the time for 10-fold repression, again assuming efficient repression by bound protein:



J. Kahn

Repression loops are flexible also

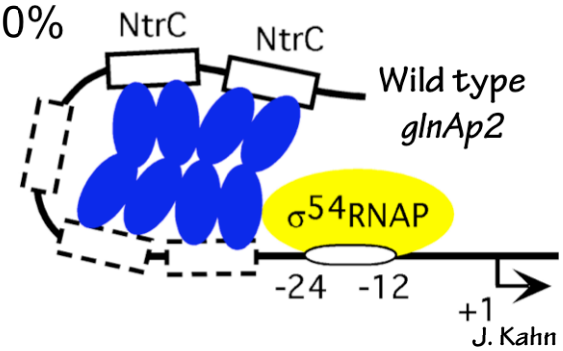
- Proposed geometries for Lac repressor loops (yesterday)
- Repressors are smaller than activation partners, and only bivalent: more sensitive to geometry



J. Kahn

Summary: Effects of Geometry

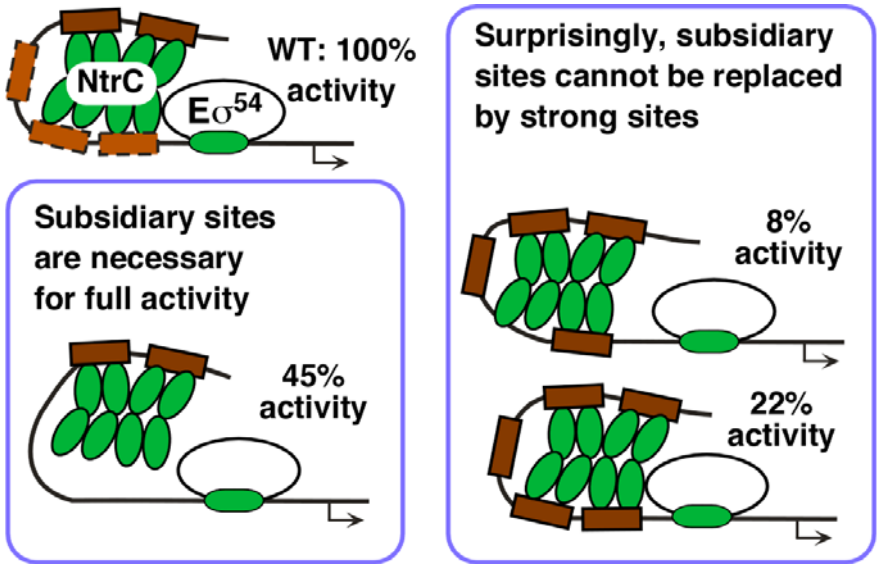
- NtrC tolerates a wide variety of DNA shapes. Only short molecules with the enhancer on the wrong side of the DNA are inactive. Little or no effect of bending. Biology has evolved to obviate effects of DNA physical chemistry.
- This is in accord with results from Rippe and Maher groups on activation.
- No variant is more than 50% as active as wild type. Due to subsidiary sites? Rippe proposed NtrC wrapping:



J. Kahn

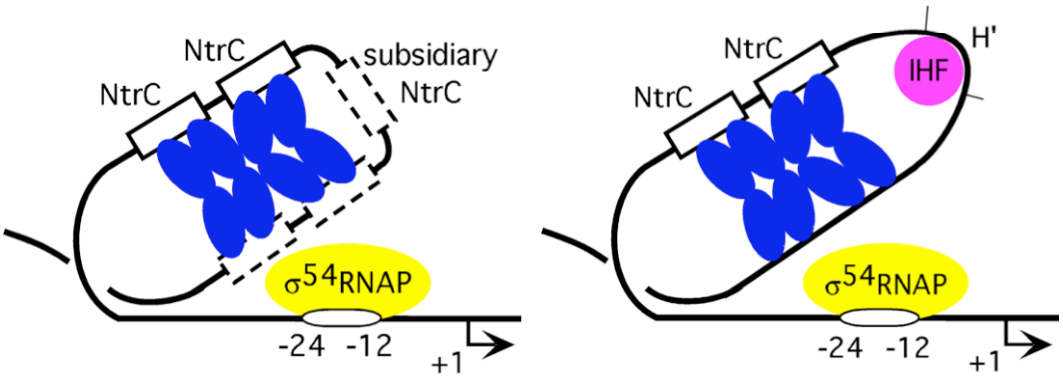
An NtrC Enhanceosome?

- Subsidiary sites must have intermediate affinity for full function:



J. Kahn

Cannot Simply Reposition Enhanceosome

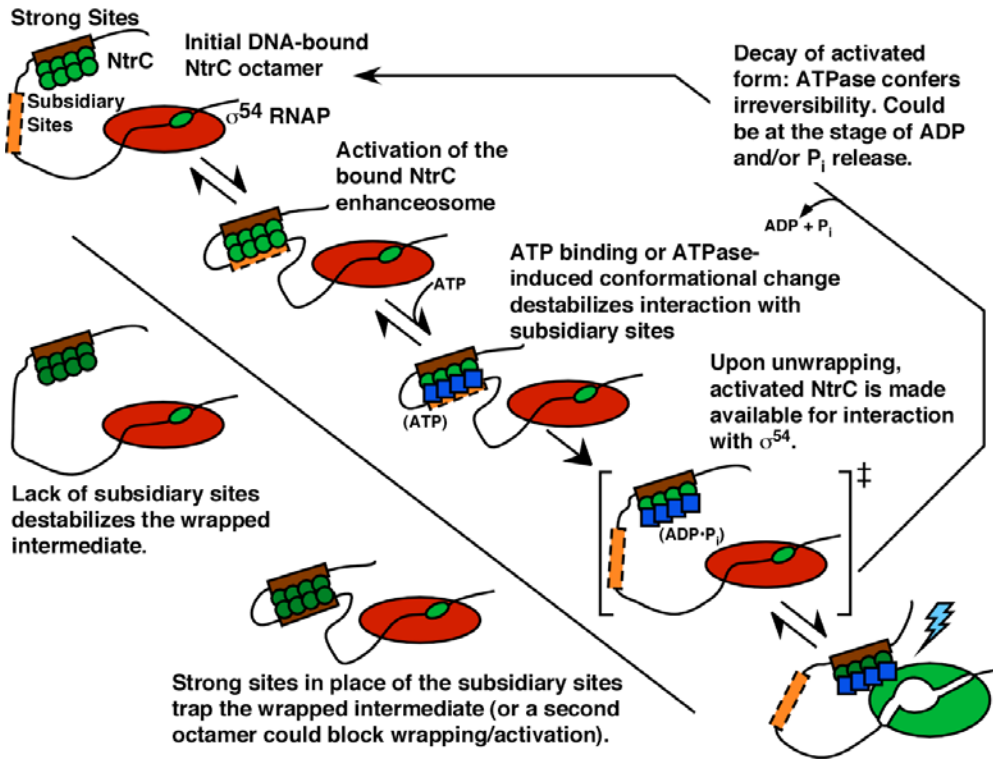


28 %

15 %

J. Kahn

Proposed Dynamic Wrapping Mechanism for Activation



J. Kahn

Conclusions

- The semi-random DNA structure library approach should be generally applicable to other systems in which DNA structure is believed to be important.
- For classical looping, the NtrC octamer only needs to be brought into the general vicinity of the polymerase. Flexibility and/or multivalency of the DNA and enhanceosome make the looping interaction insensitive to precise orientation.
- Subsidiary NtrC sites are important for full activation of transcription from wild type *glnAp2*, and the enhanceosome is not easily repositioned.
- Dynamic DNA wrapping in the enhanceosome may be needed for activation. Analogous to other AAA^+ protein molecular machines.
- You can't always get what you want.

J. Kahn

NtrC Acknowledgements

Dr. Anders Lilja

- Dr. James Jenssen
- Raymond Cheong, Dr. Ruchi Mehta
- Dr. Thomas Linn (University of Western Ontario)

  Dr. Alexander Ninfa (University of Michigan)
- Dr. Robert Weisberg (NIH)

What Would We Like to See Modeled?

- Multi-scale modeling to combine all-atom and reduced representations: realistic protein-DNA flexibility.
- Multi-protein DNA complexes with appropriate on/off rates. Consideration of intracellular architectural proteins.
- Effects of supercoiling on shape.
- Dynamics due to transcription or other tracking processes.
- Chromatin!
- Suggestions for non-obvious but possible experiments.

Thank You!

J. Kahn