Molecular dynamics simulation of supercoiled DNA minicircles

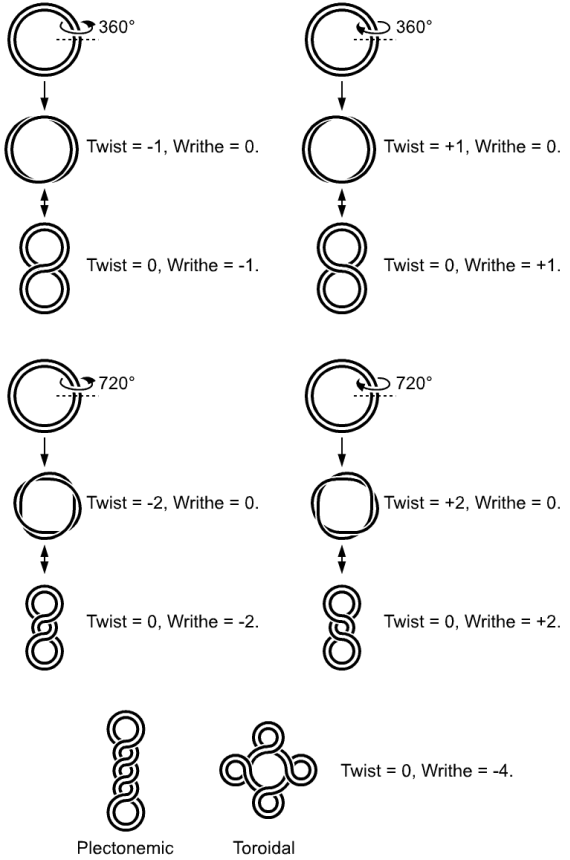
George D. Watson

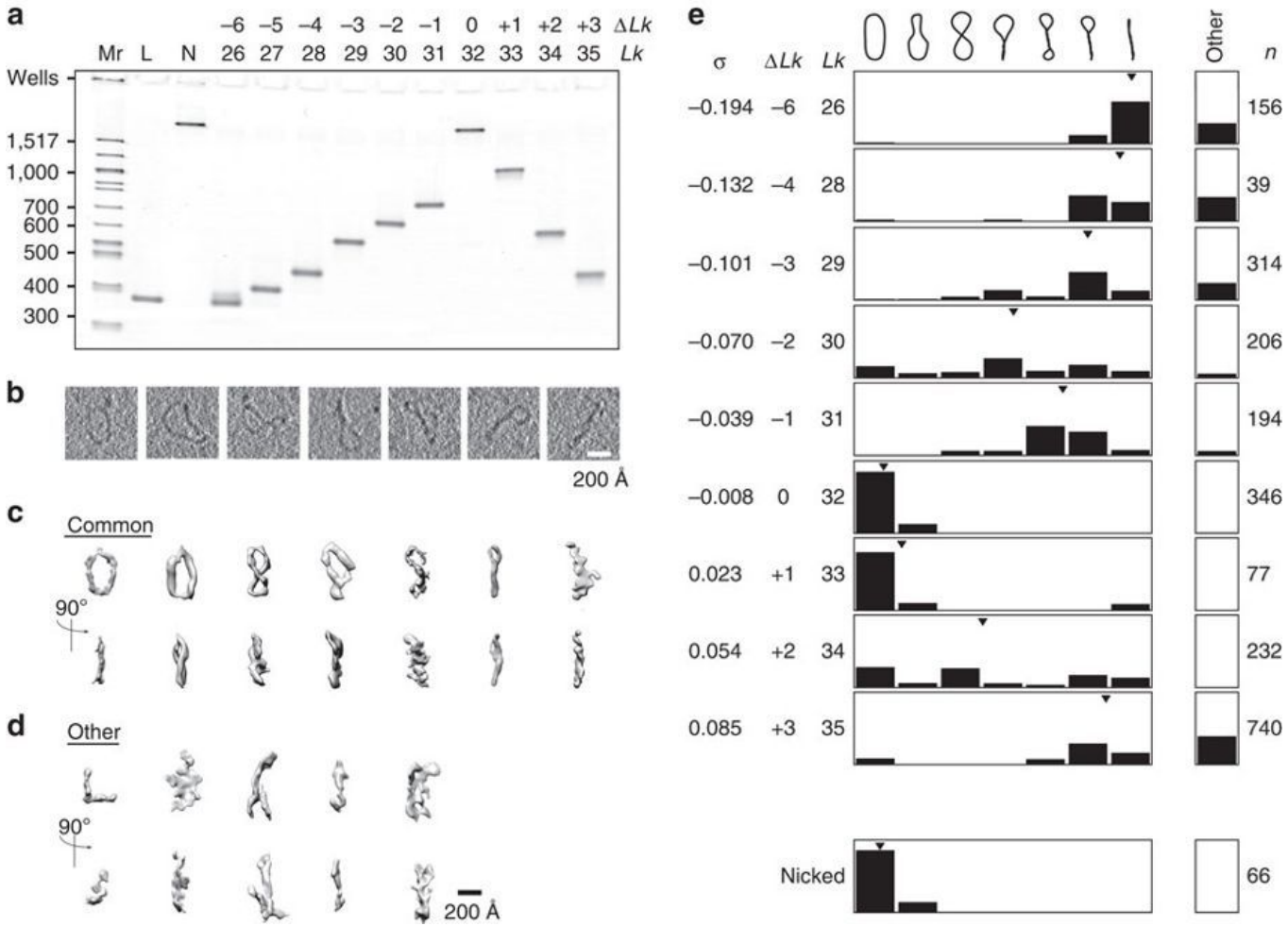
Introduction

- Gene regulation depends on protein binding & DNA structure
- Supercoiling is a key structural influence — a small change in topology can lead to large conformational changes that affect protein binding & bring distal sites closer together
- Want to be able to predict topology of arbitrary genomes & understand interplay between bound proteins and DNA topology
- Experimental methods are valuable, but struggle to resolve fine detail in dynamic processes; complementary molecular dynamics simulations can provide atomic resolution

DNA supercoiling

- Supercoiling is a deviation in the number of helical turns from the value for torsionally relaxed DNA
- Even a small deviation in either direction can have profound effects on structure & topology
- Quantified by:
  - Twist, Tw (number of coils around helix axis)
  - Writhe, Wr (number of times helix axis crosses itself)
  - Linking number, $Lk = Tw + Wr$
  - Superhelical density, $\sigma = \Delta Lk \,/\, N$ (for N bp)
- For any two intertwined closed circles in 3D space (like a DNA minicircle), Lk is a time-invariant integer but Tw & Wr may vary
- Non-zero writhe leads to all sorts of shapes...





Irobalieva R N *et al.* 2015 *Nat. Commun.* **6** 8440 doi:10.1038/ncomms9440

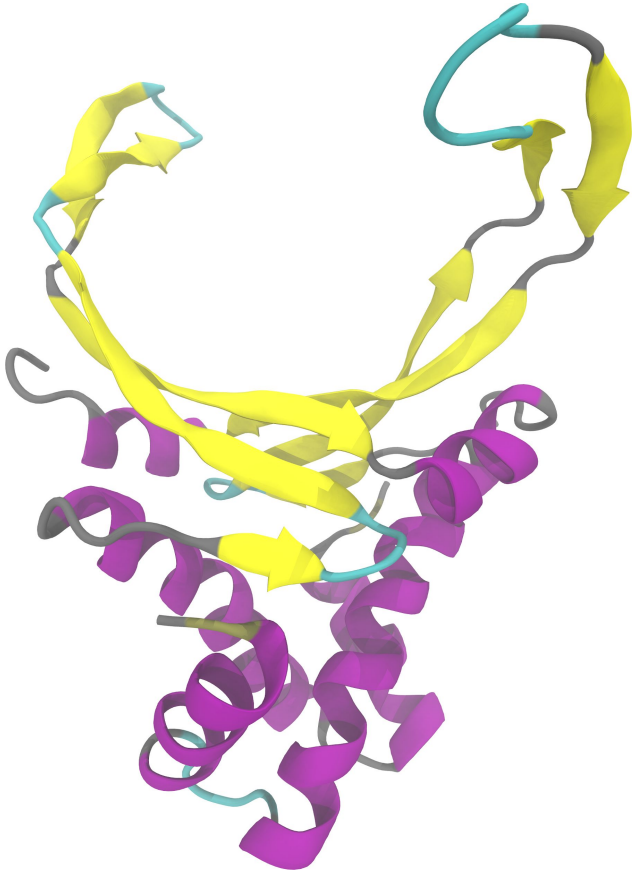
Supercoiling *in vivo*

- Prokaryotic & eukaryotic genomes are persistently negatively supercoiled
- Metabolic processes including transcription introduce dynamic changes
- Supercoiling is implicated in gene regulation [1] & the function of an epigenetic switch [2]
- Negatively supercoiled regions are associated with transcription start sites [3]
- Supercoiling-induced writhe can lead to interactions between proteins bound to distal sites [4]

[1] Baranello L *et al.* 2012 *Biochim. Biophys. Acta* **1819** 632–8 doi:10.1016/j.bbagrm.2011.12.007
[2] Norregaard K *et al.* 2013 *Proc. Natl. Acad. Sci. USA* **110** 17386–91 doi:10.1073/pnas.1215907110
[3] Kim S H *et al.* 2017 *preprint:* DNA sequence encodes the position of DNA supercoils doi:10.1101/180414
[4] Noy A *et al.* 2017 *Biophys. J.* **112** 523–31 doi:10.1016/j.bpj.2016.12.034

Nucleoid-associated proteins

- NAPs often moderate DNA topology
- IHF & HU are DNA-bending NAPs with very similar structures but little sequence similarity
- IHF binds specifically
  (to the consensus sequence WATCARNNNNTTR)
- HU binds *non*specifically to existing distortions
  (e.g. nicks, gaps, loops)
- Both bend DNA (HU 70–140°; IHF up to 160°)
- Implicated in DNA looping & gene regulation, CRISPR, biofilms, & supercoiling
- Other interesting NAPs include:
  - H-NS
  - Fis



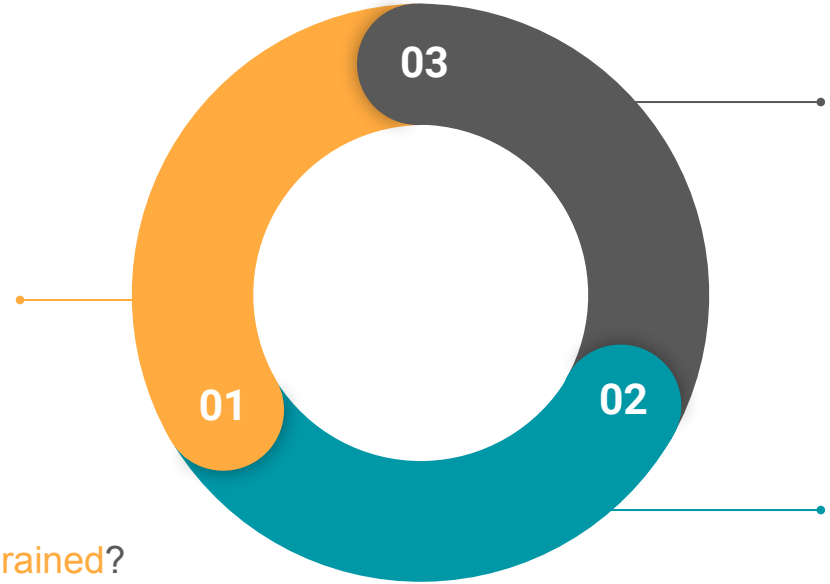
IHF structure [pdb:5j0n]

Molecular dynamics

- MD can provide dynamic, atomistic insight unavailable through experiment

Construct potential
based on the position & properties of every unit (atom or residue) in the system

Usually based on known properties of different types of atom (AMBER), but *ab initio* methods are possible for small systems



Apply force
to every unit over a very small time step; adjust velocities to ensure thermodynamic properties are stable
(e.g. Langevin thermostat)

Repeat with new positions & velocities

Integrate
at the position of every unit, in order to determine the force it will experience

- Atomistic or coarse-grained?
- Implicit or explicit solvent?
- Trade-off between speed & accuracy

General form of AMBER potential

$$V(r^N) = \sum_{\text{bonds}} k_\text{b}(l - l_0)^2 + \sum_{\text{angles}} k_\text{a}(\theta - \theta_0)^2$$

Bond lengths & angles
Hookeian potential

$$+ \sum_{\text{torsions}} \sum_{n} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)]$$

Torsions
Fourier series
(from energy associated with twisting of bonds)

$$+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}$$

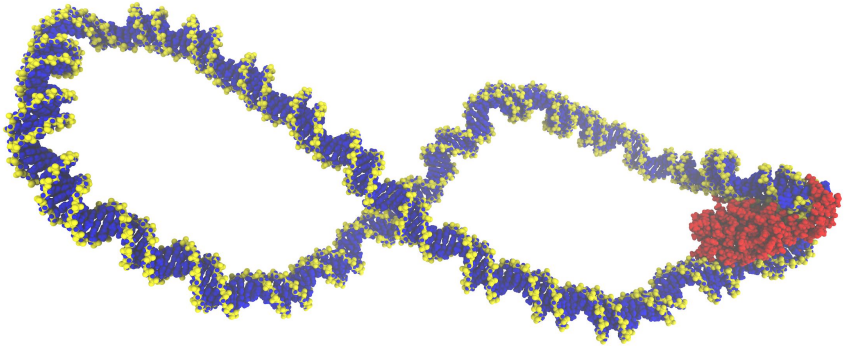
Electrostatics & van der Waals
Lennard-Jones + Coulomb potentials

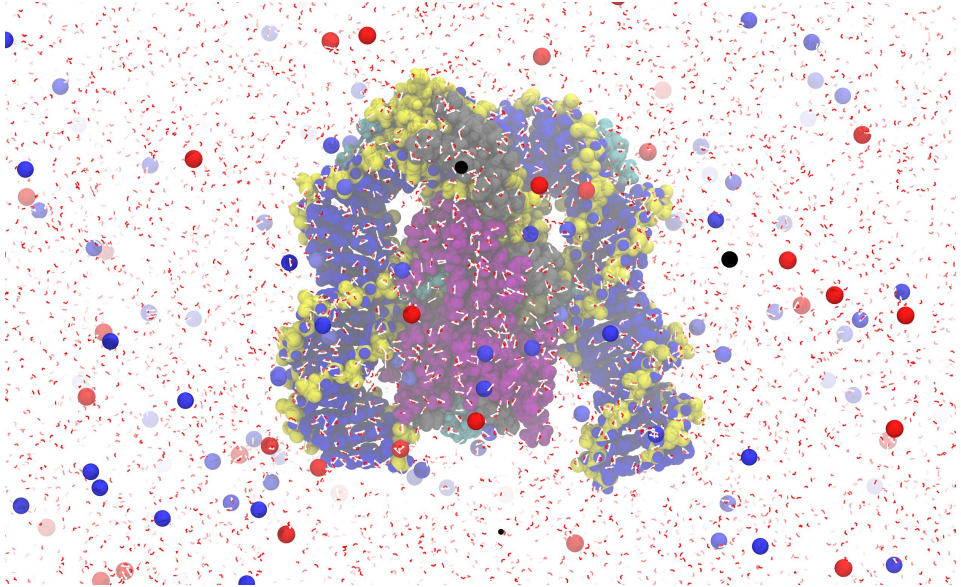
Aims & motivation

- Ultimately want to predict plectoneme formation in arbitrary genomes
- Aim to understand some predictors of plectoneme formation & how supercoiling can be moderated to regulate gene expression
- Thus, simulate DNA–protein binding to observe effect on DNA topology & emergent interactions (protein–protein, DNA–protein, or DNA–DNA)
- Make predictions testable by complementary single-molecule experiments
- Minicircle topology feeds into synthetic biology & gene therapy; understanding IHF links to biofilms & CRISPR

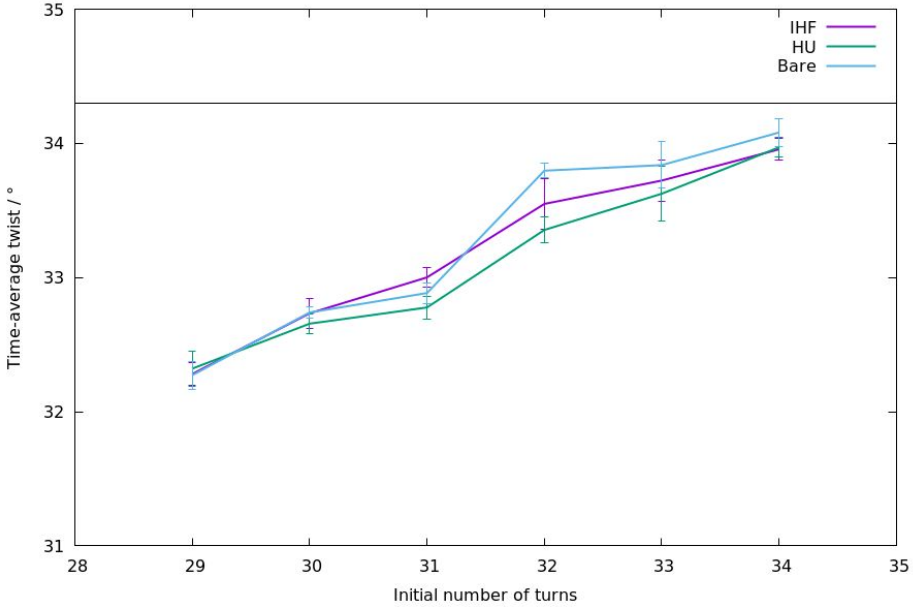
Simulations so far

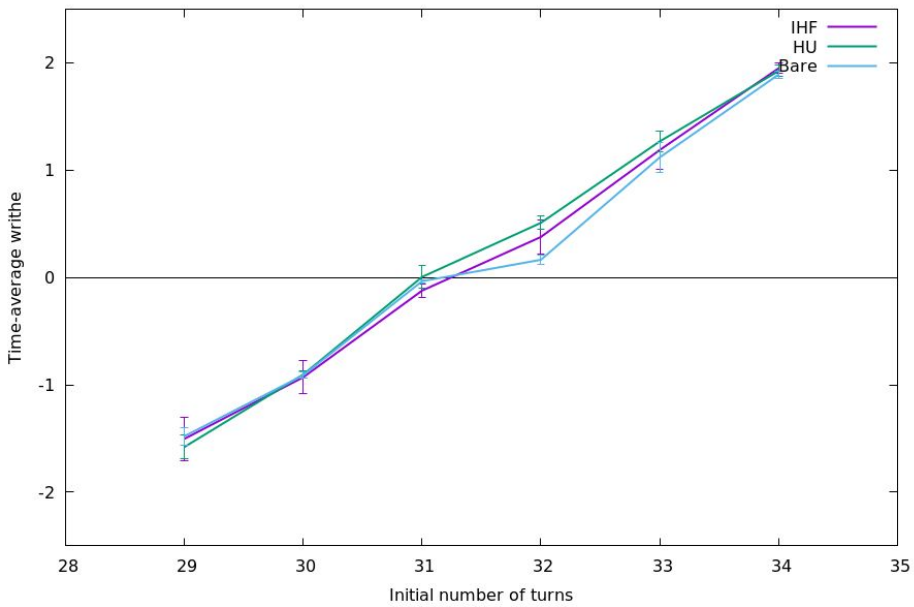
- Minicircles (336 bp)
  29 ≤ Lk ≤ 34
  Implicit solvent
  - Bare
  - + IHF
  - + HU
- Linear DNA (41 bp) + IHF
  Torsionally relaxed
  Implicit & explicit solvent





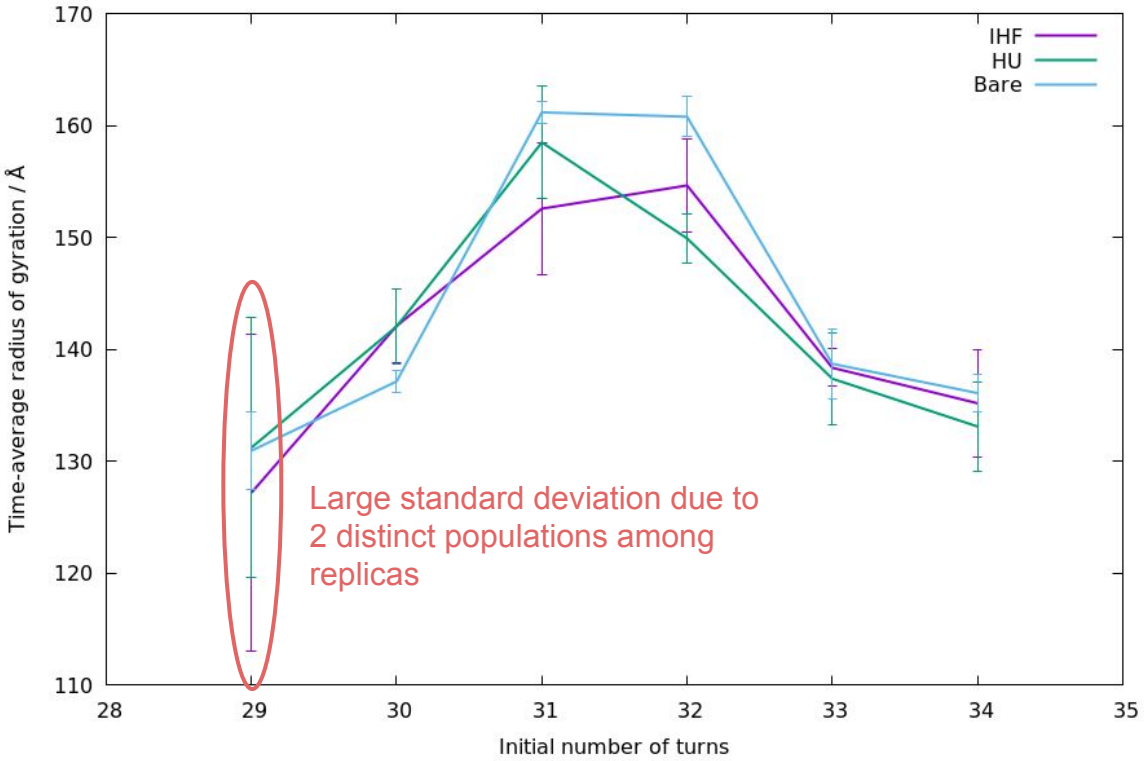
Twist & writhe



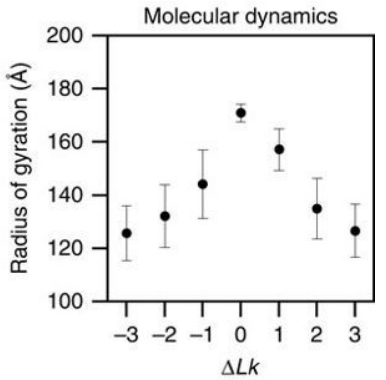


- Tw & Wr vary with Lk roughly linearly, with 0 < gradient < 1, meaning ΔLk is partitioned between the two
- Note that Lk = 31 is the most relaxed system — not Lk = 32
- IHF & HU don't seem to have much of an effect in most cases

Supercoiling enhances compaction



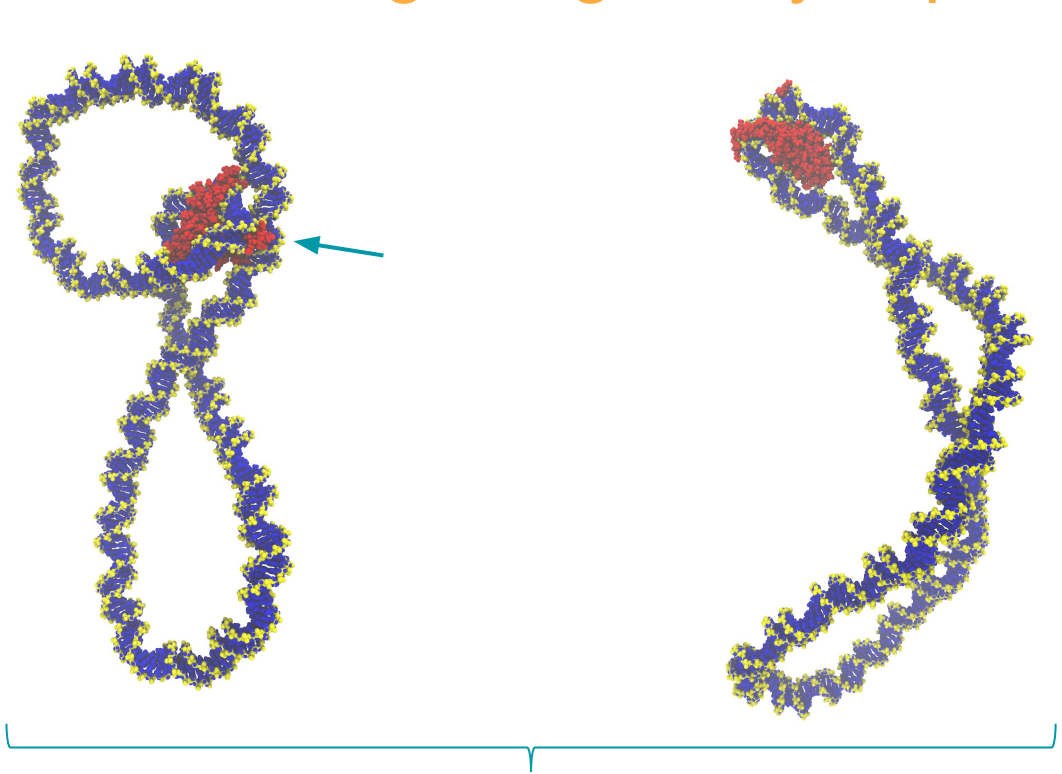
- Radius of gyration decreases with increasing $|\Delta Lk|$
- IHF & HU seem to *generally* promote compaction, but $p > 0.05$ for *most* systems



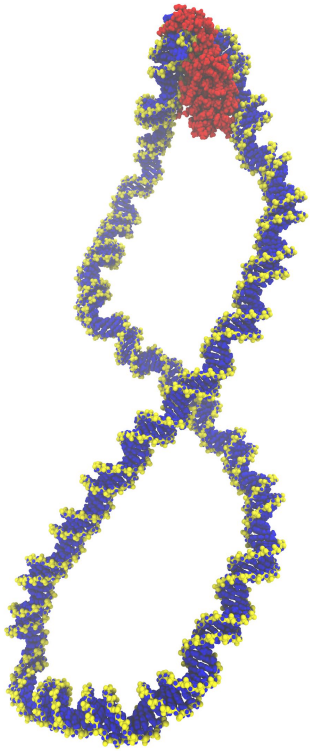
Irobalieva R N *et al.* 2015 *Nat. Commun.* **6** 8440 doi:10.1038/ncomms9440

Lk = 29

IHF can bridge negatively supercoiled DNA

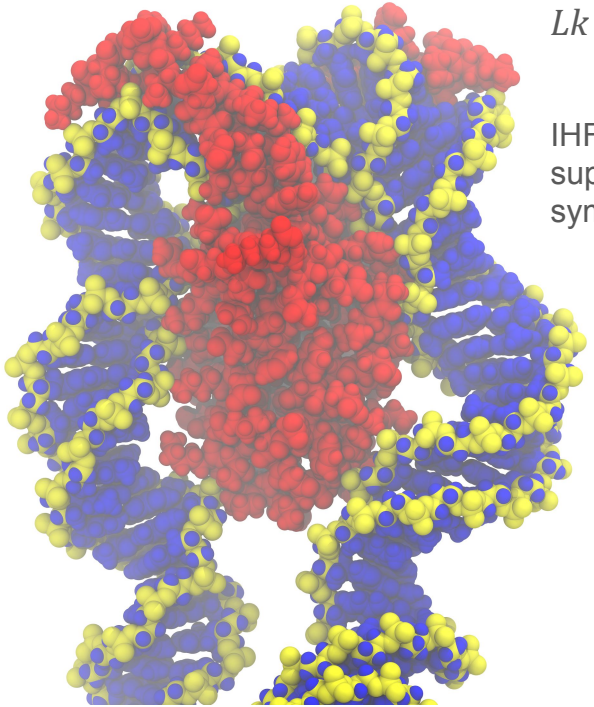


$\langle R_g \rangle$ = (112 ± 2) Å



$\langle R_g \rangle \approx$ 137 Å

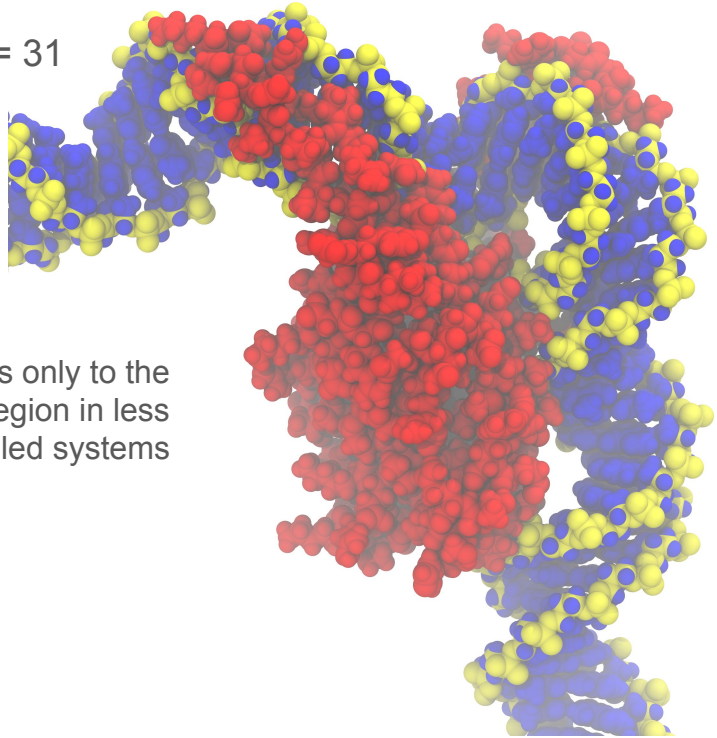
Binding mode of IHF depends on DNA topology



Lk = 29

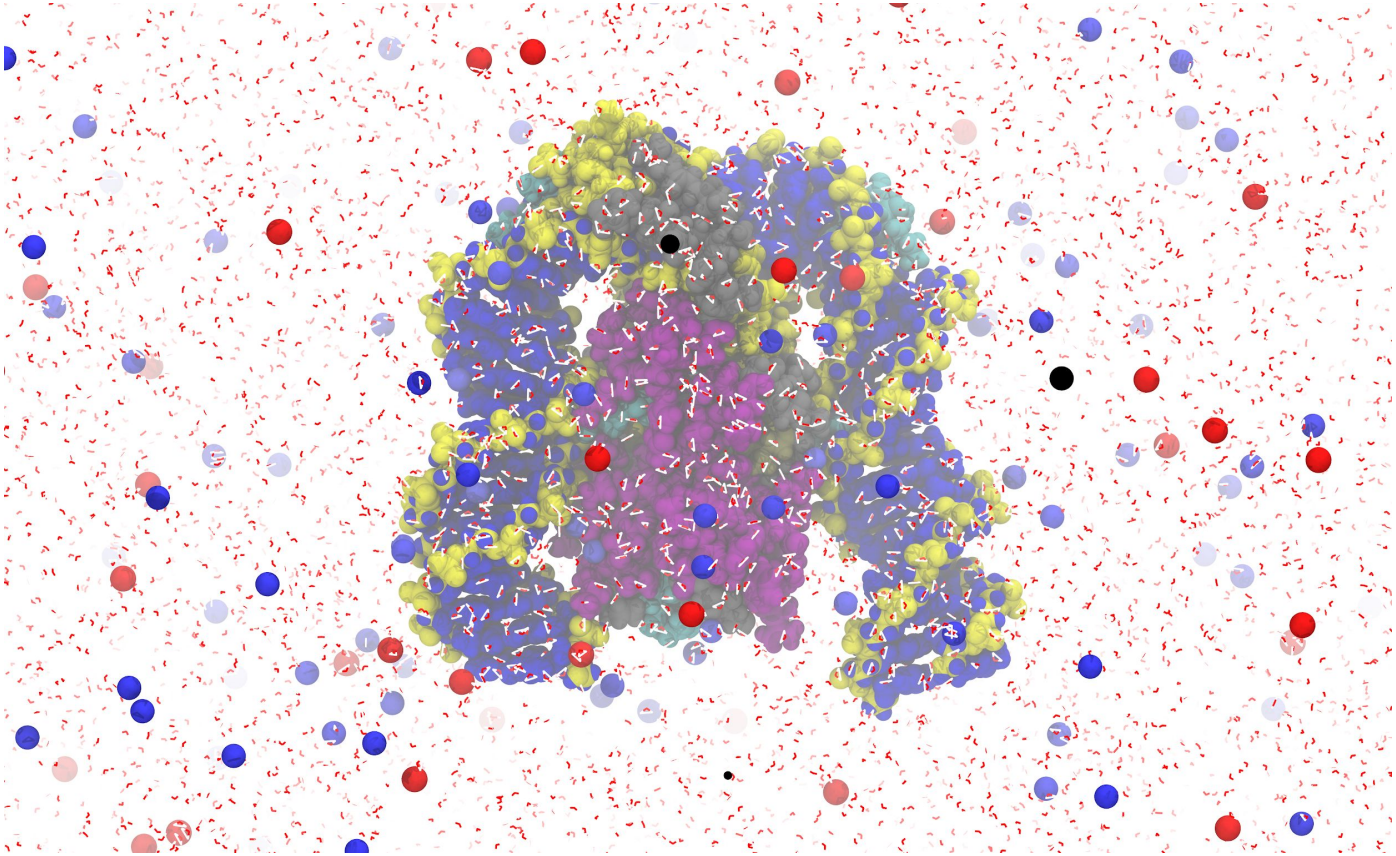
IHF binds highly supercoiled minicircles symmetrically...

Lk = 31



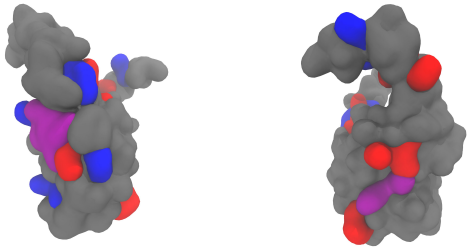
...but binds only to the AT-rich region in less supercoiled systems

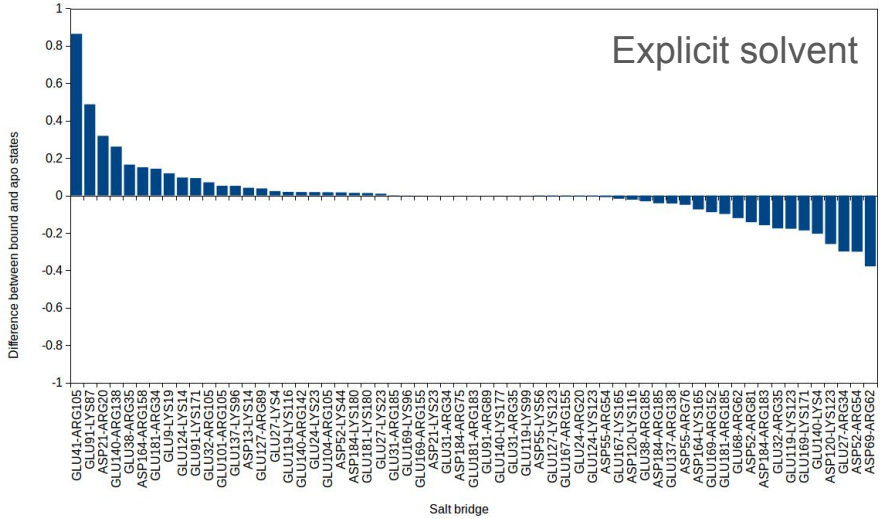
Explicit solvent

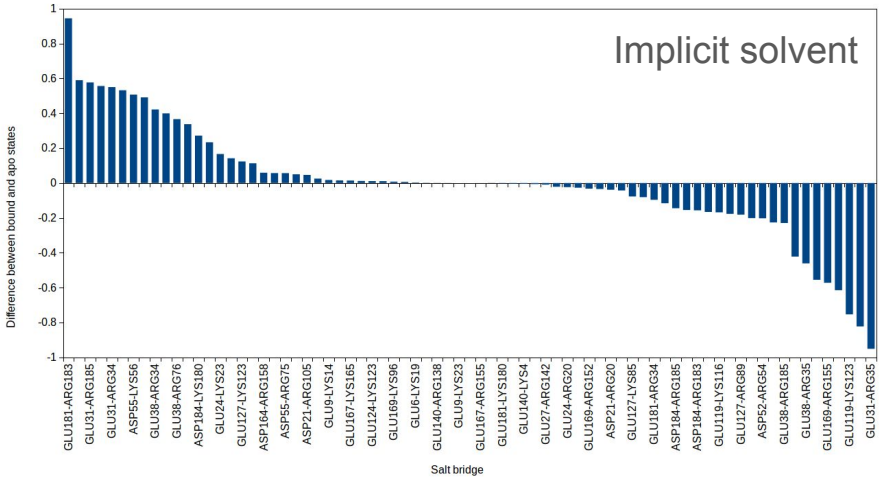


Surface salt bridges

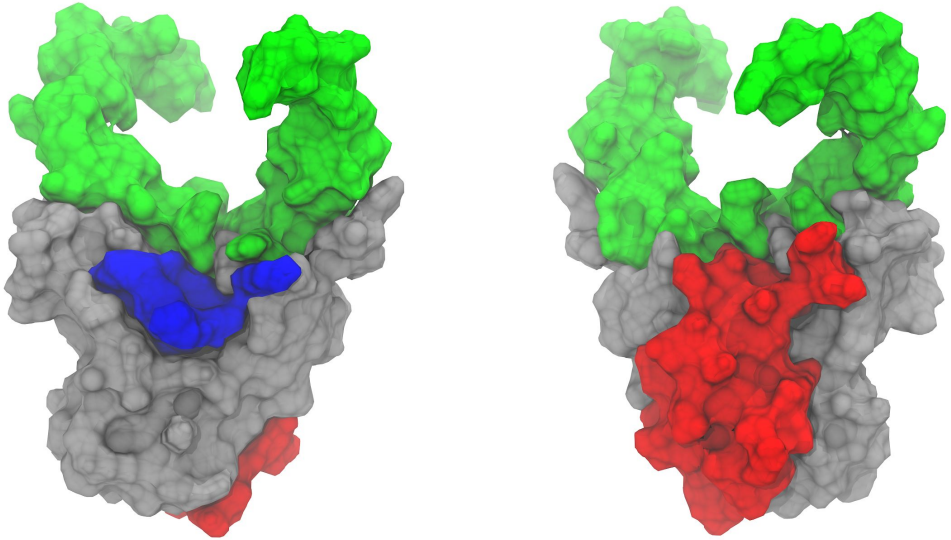
- IHF surface features many salt bridges (arginine/lysine → aspartic/glutamic acid)
- These bridges are known to differ between the DNA-bound and apo states
- Observed a significant difference between the states — but is this the same in implicit & explicit solvent?
- Important test of validity of implicit solvent approach
- No conclusions yet — needs more work



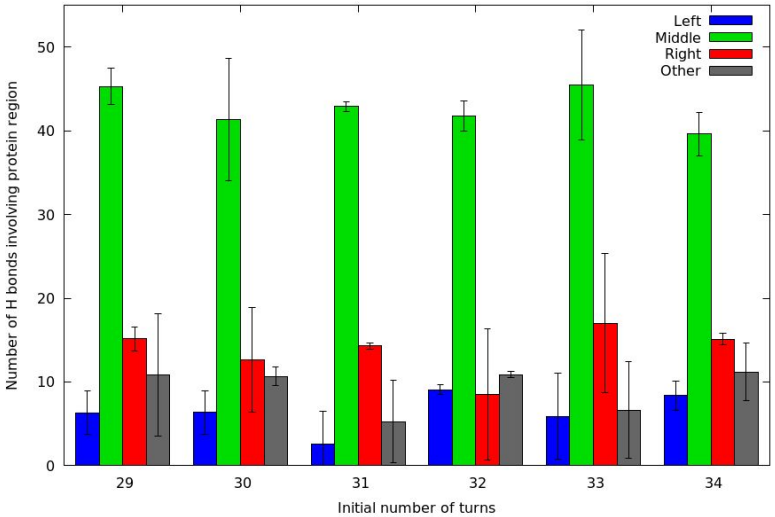




Interaction can be divided into distinct regions



Nonspecific binding site
Binds to AT-rich “right” region
Binds to other “left” region
No interaction with DNA in explicit solvent



- Differences are observed with changing Lk, but difficult to quantify due to variation between replicas
- Try defining a larger “left” region

Future work

- Improve understanding of DNA bridging by IHF
- Further explore & quantify IHF binding modes — is HU similar?
- Explore interactions between multiple proteins bound to distal sites
- Develop model (based on MD + bioinformatics + polymers) to predict plectoneme formation “hotspots”
- Converge with experiment (single-molecule, tweezers, AFM...)
    - Scale up to approach experimental lengths (coarse-grained)
    - Make predictions of plectoneme formation, protein positions, & other experimentally testable properties