An Empirical Evaluation of the Hypothesis Linking DNA Breathing Dynamics to Transcription Factor Binding Specificity

Section 1: Hypothesis Deconstruction and Operational Definitions

The analysis of any scientific hypothesis requires, first, a precise and operational definition of its components and claims, derived exclusively from observable phenomena. The central hypothesis under evaluation posits that the local, dynamic, physical fluctuations of the DNA double helix—termed "DNA breathing"—constitute a form of information "encoding" that is biologically relevant for processes such as transcription factor (TF) binding. This section will deconstruct this hypothesis into its testable, evidence-based components.

1.1 Foundational Observation: Defining "DNA Breathing" as a Physical Phenomenon

The physical basis for the hypothesis is a known property of the DNA double helix. In living cells, the two strands of the DNA molecule are not static. Due to inherent thermal motion, the strands locally and spontaneously separate and then recombine.[1] This stochastic dynamic results in transient, localized openings in the double helix.[1] This phenomenon is also referred to as "fraying".[2] These transient openings can range from the breaking of a single base pair to the formation of larger "bubbles".[1] The hypothesis under evaluation is not, therefore, based on the discovery of a new phenomenon, but rather on a new interpretation of the *informational content* and *biological function* of this existing, fundamental physical property.

1.2 Operationalizing the "Encoding" Hypothesis: From Sequence to Quantifiable Biophysical Profiles

The term "encoding" must be operationally defined in a measurable, quantifiable manner. The primary sequence of DNA—the order of Adenine (A), Thymine (T), Cytosine (C), and Guanine (G)—is the established digital, or primary, layer of encoding. The hypothesis suggests a secondary, analog layer of information that is *derived* from this primary sequence.

This secondary "encoding" is operationally defined by the quantifiable outputs of biophysical models, such as the Extended Peyrard-Bishop-Dauxois (EPBD) nonlinear DNA model.[1] These computational models, which utilize algorithms like Markov Chain Monte Carlo (MCMC) [3], translate the primary ATCG sequence into "genomic scale profiles" of specific, local, dynamic properties.[1] The key quantifiable metrics—the "encoding"—include:

- **Average Base-Pair Openings:** The average transverse displacement, or opening, of a base pair, averaged over thermal fluctuations.[1]
- **Base Flipping Probability:** The probability of a specific dynamic in which at least one base in a pair flips out of the helical stack.[1]
- **DNA Bubble Probability:** The probability of extended regions where the helix unwinds and the strands temporarily separate due to thermal motion.[1]

This "encoding" is explicitly *sequence-dependent*. The models themselves are extensions of prior non-linear models that have been modified "to include a sequence-dependent stacking term".[5] To achieve accuracy, these models are parameterized using experimentally derived force constants for the 10 unique dinucleotide steps (e.g., CG, CA, GC, AT, etc.).[5] This demonstrates that the primary ATCG sequence (Layer 1) is translated via sequence-dependent physics (e.g., local stacking energies) into a secondary, analog profile of dynamic properties (Layer 2: bubble probability, flipping probability, etc.).

Furthermore, this encoding is both dynamic and *non-local*. The biophysical models can calculate a "dynamic length".[1] This metric quantifies "how many and precisely which base-pairs experience statistically significant changes" in their dynamic profile as a result of a *single point mutation* (SNP) elsewhere.[1] This implies that the "encoding" is not a 1:1 map. A single base change (a SNP) can alter the biophysical "breathing" profile of *many* adjacent base pairs, creating a complex, non-local information landscape.

The central claim of the hypothesis is that biological machinery, such as transcription factors, "reads" *both* the primary digital sequence and this secondary analog biophysical profile.

1.3 The Measurement Challenge: Transient Dynamics and the Limits of Ensemble Averaging

A primary obstacle to the empirical validation of this hypothesis is the intrinsic nature of the phenomenon itself. DNA breathing events are, by definition, transient, stochastic, and often rare. As a result, "experimental observation of DNA breathing in real time is difficult in ensemble measurements due to the low frequency and short duration of base pair opening".[2] Traditional ensemble-based biophysical methods, such as 1H Nuclear Magnetic Resonance (NMR), are used to measure *average* properties, such as "base-pair lifetimes" and "imino

proton exchange" rates, across a large population of molecules.[6] While these methods can confirm that dynamics *exist* and are sequence-dependent (e.g., showing destabilization at mismatches [6]), they inherently average out the specific, rare, transient events that constitute the "breathing" signal.[2]

This measurement challenge necessitates the use of two distinct, advanced approaches, which form the basis of the evidence presented in this report:

1. **Computational Biophysical Modeling:** As described in 1.2, non-linear models like the pyDNA-EPBD [1] use MCMC simulations to *computationally generate* the equilibrium state and produce the probabilistic profiles (e.g., flipping probability) that cannot be easily captured in an ensemble experiment.[1]
2. **Single-Molecule Experimental Techniques:** Methods such as single-molecule Förster resonance energy transfer (smFRET) and single-molecule fluorescence linear dichroism (smFLD) *can* observe individual molecules in real-time.[2] These techniques can "report on conformational dynamics that the FRET signal is blind to" [2] and are capable of monitoring the "significant breathing at the fork junction" of a single DNA molecule.[7]

The hypothesis is, therefore, inherently a product of high-resolution methodologies. It reframes what was previously considered experimental "difficulty" or stochastic "noise" (i.e., transient, low-frequency events averaged out in ensemble measurements [2]) as the "signal" or "encoding" itself. The validation of this hypothesis depends entirely on evidence from these computational and single-molecule approaches, which are capable of resolving this dynamic, probabilistic information.

Section 2: Ten Verifiable Observations and Direct Logical Inferences

This section presents ten discrete, verifiable facts extracted from the provided experimental and computational data. Each fact is accompanied by its source and a single-sentence logical inference establishing its direct, objective relevance to the components of the hypothesis.

- **Fact 1:** *In vitro* analysis using genomic-context protein-binding microarray (gcPBM) data for basic helix-loop-helix (bHLH) transcription factor complexes (*Max-Max* and *c-Myc-Max*) showed "highly significant correlations" between TF binding and the computationally-derived "DNA flipping probability" feature.[9] This observation provides direct, controlled evidence of a quantitative relationship between a specific DNA breathing metric and protein-binding affinity.
- **Fact 2:** *In vivo* analysis of 44 transcription factors (TFs) using ChIP-seq (chromatin immunoprecipitation sequencing) binding data found that for a "large proportion of TFs," their associated "breathing features in or near core motifs are associated with binding".[9] This finding demonstrates that the correlation observed *in vitro* is not an isolated artifact and is generalizable to a wide range of TF families in a cellular context.
- **Fact 3:** A multi-modal deep learning model (EPBDxDNABERT-2) that integrates

biophysical DNA breathing features (from the EPBD model) with a genomic large language model (DNABERT-2) "significantly improves the prediction of over 660 TF-DNA" binding sites, registering an "increase in the area under the receiver operating characteristic (AUROC) metric of up to 9.6%".[12] This quantitative improvement demonstrates that DNA breathing dynamics provide non-redundant, predictive, and biologically relevant information for TF-binding specificity that is not fully captured by the primary DNA sequence alone.

- Fact 4: Experimental measurements of duplex DNA bending dynamics using electron paramagnetic resonance (EPR) revealed an "approximately fourfold" range in stepwise flexibility, with pyrimidine-purine (Py-Pu) steps (read 5' to 3') being "nearly twice as flexible" as the duplex average.13 This quantifies the strong, sequence-dependent nature of DNA's fundamental mechanical properties.
- **Fact 5:** Thermodynamic measurements of A-T base pair opening demonstrated that opening enthalpy changes (ΔH_{op}) "encompass a wide range of values," specifically from 17 to 29 kcal/mol, and are correlated with opening entropy changes (ΔS_{op}) in the central part of a duplex.[14] This observation confirms that the energetic landscape governing base-pair opening (the core of "breathing") is highly variable and sequence-dependent.
- **Fact 6:** Monte Carlo simulations used to parameterize a sequence-dependent Peyrard-Bishop-Dauxois (PBD) model "confirm that the GG/CC dinucleotide stacking is remarkably unstable" compared with the stacking in GC/CG and CG/GC dinucleotide steps.[5] This identifies a specific physical cause—dinucleotide stacking instability—as a key determinant of the sequence-dependent "breathing" profile.
- **Fact 7:** Coarse-grained simulations suggest that thermally induced DNA distortions, specifically "kinks" associated with the "spontaneous formation of internal bubbles" (i.e., breathing), "can account for ~80% of the DNA curvature present in experimentally solved [protein-bound] structures".[15] This provides a direct, quantitative link between the stochastic, local "breathing" phenomenon and the specific, functional, macroscopic *shapes* (e.g., sharp bends) that proteins recognize.
- **Fact 8:** Langevin dynamics simulations of a "Breathing DNA model" found that local denaturations (bubbles within the duplex and forks at the ends) at physiological temperature "can lower the persistence length drastically for a short DNA segment".[16] This demonstrates that local, transient breathing events have significant, non-local consequences on the overall mechanical stiffness and flexibility of the DNA polymer.
- **Fact 9:** A chemically-designed "wedge" molecule, which binds the DNA minor groove to "induce perturbations at specific base steps" and "act as an allosteric effector," was "found to enhance Exd [protein] binding by 1.5 kcal mol–1," even in the absence of direct contact with the protein.[17] This is a direct, quantitative measurement of DNA-mediated allostery, proving that a local structural perturbation in the DNA can energetically affect protein binding at a distance.
- **Fact 10:** Experimental studies on the cooperative binding of cAMP to CAPN provided a "definitive example" of "purely dynamics-driven allostery," wherein allosteric

interactions are "exclusively mediated by changes in protein motions" (i.e., conformational entropy), not by static structural changes.[18] This establishes an experimentally-verified biophysical precedent for a "breathing" or "dynamic" signal to be transmitted allosterically, affecting function *purely* through changes in motion and entropy.

Section 3: Analysis of Correlative Evidence: Transcription Factor Binding

This section provides an in-depth examination of the evidence directly linking DNA breathing features to the binding of transcription factors. The analysis proceeds from controlled *in vitro* environments to complex *in vivo* systems and, finally, to large-scale *in silico* validation.

3.1 In Vitro Validation: High-Significance Correlations in bHLH Transcription Factors

The most direct, controlled evidence for the hypothesis comes from *in vitro* studies. One such study analyzed the contribution of DNA breathing to the physical interactions of specific TFs using genomic-context protein-binding microarray (gcPBM) data.[9] This analysis focused on three basic helix-loop-helix (bHLH) TF complexes: *Max-Max homodimer (MAX)*, *Mad2-Max heterodimer (MAD)*, and *c-Myc-Max heterodimer (MYC)*.[9]

The biophysical features computed from the DNA sequence—specifically "DNA flipping probability" and "bubble propensity"—were correlated with the measured binding affinity.[9] The results for the *MAX* and *MYC* complexes showed "highly significant correlations" between the DNA flipping probability feature and the TF binding affinity.[9] This finding led to the direct inference that these specific TFs "may prefer locally and temporally melted DNA formed through breathing".[9]

Significantly, the *MAD* complex was excluded from this main analysis because it "did not show a similarly well-defined motif".[9] This exclusion acts as an implicit internal control, suggesting the observed correlations for *MAX* and *MYC* are not spurious artifacts of the model or method, but are tied to specific TF-motif interactions. This *in vitro* data provides a foundational, quantitative link between a computed breathing metric ("flipping probability") and a measured biological outcome (binding affinity).

3.2 In Vivo Generalization: Associations Across Diverse TF Families

While *in vitro* data provides controlled validation, its relevance hinges on its generalization to

the complex, *in vivo* cellular environment. The analysis was therefore extended from the three bHLH TFs to a broad set of 44 TFs using *in vivo* ChIP-seq binding data.[9]
This *in vivo* investigation found that for a "large proportion of TFs," their "breathing features in or near core motifs are associated with binding".[9] This is a critical finding. It demonstrates that the phenomenon is not an isolated artifact of the bHLH family or the artificial *in vitro* (gcPBM) conditions. The association between DNA breathing dynamics and TF binding appears to be a general principle that holds true across diverse TF families and within the native chromatin context of the cell. This generalization is a prerequisite for the hypothesis that breathing dynamics function as a widespread "encoding" mechanism.

3.3 Critical Evidentiary Limitation: Variability in Association

Objective analysis requires acknowledging data that refines or limits the hypothesis. The *in vivo* study that generalized the finding to 44 TFs also reported a crucial caveat: the "sign and magnitude of these associations vary substantially across TF families".[9]
This is a non-trivial observation. It means that DNA breathing is not a simple, universal "on/off" switch for binding. High "bubble probability," for example, does not universally *promote* binding for all TFs; for some, it might *inhibit* it. The *way* in which the biophysical "encoding" is read—whether a flexible, "breathing" site is preferred or avoided—is *dependent on the protein* (i.e., the TF family).
This variability, while a limitation on any simple, monolithic theory, is simultaneously consistent with a more complex "encoding" hypothesis. An "encoding" implies specificity. The fact that different TF families *interpret* the *same* biophysical code *differently* (some preferring locally melted DNA, others perhaps preferring rigid DNA) reinforces the idea that it is a rich, analog layer of information to be *interpreted*, not a simple physical barrier. This variability also provides a physical explanation for *why* the traditional "core motif" view is limited [9]—because the primary sequence of the motif fails to capture this secondary, protein-specific physical context.

3.4 Quantitative Validation via Multi-Modal Machine Learning

The most powerful and quantitative validation of the hypothesis comes from a large-scale *in silico* experiment. This study directly tested whether the biophysical "breathing" features provide *new, predictive information* that is not already present in the primary DNA sequence. The study designed a multi-modal deep learning model, EPBDxDNABERT-2.[12] This model integrates two streams of information:

1. **Sequence Data:** Processed by a genomic pretrained large language model, DNABERT-2, which learns complex patterns directly from the primary ATCG sequence.[12]
2. **Biophysical Data:** A profile of DNA breathing features (e.g., average coordinates, flipping features) generated by the pyDNA-EPBD model.[12]

This multi-modal model was compared to a baseline model (DNABERT-2 alone) on the task of predicting TF binding sites. The test dataset was massive, comprising 690 ChIP-seq experiments covering 161 distinct TFs.[12]

The results were definitive. The multi-modal model (EPBDxDNABERT-2) that *included* the biophysical breathing features "significantly improves the prediction of over 660 TF-DNA" binding sites.[12] The improvement was quantified as "an increase in the area under the receiver operating characteristic (AUROC) metric of up to 9.6%".[12] This "greatly enhanced" predictive power was also confirmed on *in vitro* HT-SELEX datasets for 215 TFs.[12]

The logical implication of this finding is profound. The baseline DNABERT-2 model is a state-of-the-art architecture designed to extract the maximum possible information from the *primary sequence*. The *only way* the multi-modal model could achieve a 9.6% improvement in AUROC is if the biophysical "breathing" features provided *biologically relevant, predictive information that was not redundant*—i.e., information that was *not* fully captured by the primary ATCG sequence model alone. This experiment, therefore, provides a direct, quantitative measurement of the *information content* of the "breathing" profile. It demonstrates that the biophysical dynamics are, by this objective measure, a form of "encoding" that contributes demonstrably to TF-DNA binding specificity.

Table 1: Summary of Evidence Correlating DNA Dynamics with TF Binding

Study / Method	Data Source	Subjects	Breathing Feature Analyzed	Reported Finding
In Vitro (gcPBM)	[9]	*Max* & *c-Myc* (bHLH)	DNA flipping probability	"Highly significant correlations" with binding affinity
In Vivo (ChIP-seq)	[9]	44 TFs	Breathing features near motifs	"Large proportion" of TFs show association with binding
In Vivo (ChIP-seq)	[9]	44 TFs	Breathing features near motifs	"Sign and magnitude... vary substantially across TF families"
In Silico (ML)	[12]	161 TFs (in vivo)	EPBD-derived features	"increase in the (AUROC) metric of up to 9.6%"
In Silico (ML)	[12]	215 TFs (in vitro)	EPBD-derived features	"greatly enhanced TF-binding

				predictions"

Section 4: Analysis of Biophysical Plausibility: Energetics, Mechanics, and Protein Interactions

The correlational and predictive evidence established in Section 3 requires a plausible physical mechanism. If "breathing" correlates with binding, *how* do these nanometer-scale, sub-millisecond dynamic events physically influence the binding of a protein? This section analyzes the evidence for the biophysical plausibility of such a mechanism.

4.1 Sequence-Dependent Energetics of Base Pair Opening

The "encoding" must originate in the sequence-dependent thermodynamics of the DNA helix. Experimental data confirms that the energy required to open a base pair—the fundamental event in "breathing"—is not uniform, but is highly dependent on the local sequence. Thermodynamic measurements of A-T base pair opening demonstrate that the opening enthalpy (ΔH_{op}) "encompass[es] a wide range of values," specifically from 17 to 29 kcal/mol.[14] This wide energetic range is managed, in part, by "enthalpy-entropy compensation".[14] For the central six base pairs of a measured duplex, changes in opening enthalpy were correlated to changes in opening entropy (ΔS_{op}), a mechanism that "minimizes the variations in the opening free energies (ΔG_{op})" among them.[14] However, this compensation *breaks down* for base pairs located "close to the ends of the duplex structure," which "deviate from the enthalpy-entropy compensation pattern," suggesting a "different mode of opening".[14]

A primary determinant of this energetic landscape is the dinucleotide stacking interaction. Computational PBD models, parameterized with experimental melting data, "confirm that the GG/CC dinucleotide stacking is remarkably unstable" when compared to other dinucleotide steps like GC/CG and CG/GC.[5] This inherent instability, rooted in the quantum-chemical stacking interactions, means a "GG" sequence will have a *different* intrinsic breathing profile than a "GC" sequence, even if both are part of a G/C-rich region.

Furthermore, this opening is a *cooperative* process. The free energy cost to open a single thymine base was calculated to be "almost 12 kcal mol−1" when its adjacent base pairs are unperturbed.[21] However, this cost plummets to "4 kcal mol−1" if the adjacent GC pair is *already* perturbed.[21]

Taken together, this evidence demonstrates that the primary ATCG sequence (e.g., GG/CC vs. GC/CG [5]) dictates a highly specific, cooperative [21], and non-uniform [14] *energetic landscape*. This energetic landscape, in turn, *deterministically* governs the *probability* of a transient opening—that is, it *is* the "breathing" profile.

4.2 Local Dynamics and Macroscopic Mechanical Consequences

The sequence-dependent energetic landscape (4.1) translates directly into sequence-dependent *mechanical* properties. Electron paramagnetic resonance (EPR) studies, which measure the short-time (submicrosecond) bending dynamics of duplex DNA, provide quantitative confirmation.[13] These experiments found that the "bending dynamics at a single site are a function of the sequence".[13] The data revealed an "approximately fourfold" range in stepwise flexibility.[13] This flexibility is strongly dependent on the dinucleotide-step type: pyrimidine-purine (Py-Pu) steps (when read 5' to 3') were found to be "nearly twice as flexible" as average, while purine-pyrimidine (Pu-Py) steps were "more than half as flexible as average".[13]

These highly localized, sequence-dependent dynamics (breathing and flexing) have direct, non-local consequences on the *macroscopic* mechanical properties of the DNA polymer. Langevin dynamics simulations of a "Breathing DNA model" (which explicitly includes the formation of bubbles and forks) found that these "local denaturations at a physiological temperature, despite their rare and transient presence," are sufficient to "lower the persistence length drastically for a short DNA segment".[16]

This provides a critical mechanistic link. The local, sequence-dependent *energetics* (4.1) determine the *probability* of a local breathing event (1.2), which in turn dictates the *mechanical flexibility* of that step (4.2). These local events, in aggregate, "drastically" alter the *overall stiffness* (persistence length) of the entire DNA segment.[16] A protein interacting with a 50-bp segment of DNA is therefore not interacting with a uniform rod, but with a polymer whose local and global flexibility is precisely "encoded" by its sequence-dependent breathing profile.

4.3 Mechanism of Recognition: Thermal Kinks and Conformational Selection

The evidence thus far provides a plausible *how*: a sequence-dependent breathing profile creates a sequence-dependent mechanical profile. But how does a TF *use* this information? Coarse-grained simulations provide a specific, testable mechanism that directly challenges the traditional "induced fit" model of protein-DNA binding.

The simulations propose that the DNA double helix, driven by thermal energy, *spontaneously* experiences "thermally induced kinks" (i.e., sharp bends).[15] These kinks are "associated with the spontaneous formation of internal bubbles" [15]—they are a direct, physical manifestation of "breathing."

The most striking finding from this study was a quantitative comparison between these *thermally-sampled* "kinked" states and the *experimentally-solved* structures of protein-DNA complexes from the Protein Data Bank (PDB). This comparison revealed that the "thermally

induced distortions can account for ~80% of the DNA curvature present in experimentally solved structures".[15]

This finding suggests a paradigm shift in the model of protein-DNA recognition:

1. The **Traditional (Induced Fit) Model** (implied in [22]) views the protein as the active element. The protein binds to relatively straight DNA (a passive element) and *forces* it to bend, paying a significant energetic penalty to deform the DNA into its required shape.
2. The [15] (Conformational Selection) Model reframes this. It suggests the DNA is *already active*, spontaneously "breathing" and "kinking" into these highly bent conformations due to its intrinsic, sequence-dependent thermal dynamics. The protein (now the selective element) simply "waits" for the DNA to *spontaneously adopt* the correct bent conformation and then *binds to and stabilizes* that pre-existing state.

This "conformational selection" model provides the direct, causal, and physically-grounded link from "breathing" to "binding." The "encoding" (the sequence-dependent probability of breathing/kinking) *directly* modulates the free energy of binding. A DNA sequence that has a *higher probability* of spontaneously "breathing" into the correct kinked shape will have a *higher measured binding affinity* for the protein, as the protein pays a much lower entropic and enthalpic cost to bind. The 80% figure [15] suggests this is not a minor effect, but the *dominant* mechanism for DNA bending by proteins.

4.4 Reciprocal Mechanisms: Protein-Mediated Augmentation of Dynamics

The relationship between the protein and the DNA's dynamic profile is not unidirectional. Evidence from single-molecule studies shows that proteins can, in turn, *write* to the "encoding" by altering the DNA's local breathing dynamics, creating a reciprocal feedback loop.

This was observed in single-molecule studies using simultaneous smFRET and smFLD to monitor a DNA replication fork junction.[2] The results showed three distinct phases:

1. **Intrinsic Dynamics:** First, "significant breathing" (local motions on the ~100-μs timescale) was observed at the fork junction *before* any protein was present.[7] This confirms the existence of the intrinsic, spontaneous dynamics predicted in 4.3.
2. **Weak Binding Augmentation:** Second, this "significant breathing... was greatly augmented by the presence of weakly bound helicase".[7] The initial, non-processive binding of the helicase complex "significantly perturbed" the magnitudes and relaxation times of these backbone fluctuations.[7]
3. **Processive Binding and Unwinding:** Third, these fluctuations became "still larger" and were "followed by strand separation" only *after* the *complete, tightly bound, and processive helicase complex* was assembled.[7]

This reveals a multi-step, dynamic information-processing system. (A) A protein (helicase) *reads* the intrinsic breathing profile of the DNA at the fork. (B) A *weak* initial binding event

occurs, possibly via conformational selection (4.3). (C) This weak binding *writes* to the DNA, *amplifying* its local breathing dynamics ("greatly augmented"). (D) This *newly amplified dynamic state* (a modified "encoding") is then, plausibly, the necessary signal or substrate for the assembly of the full, functional complex and its ultimate function (unwinding).
This implies the "encoding" is not a static map read by the protein, but a dynamic, reciprocal interface that is *read*, *edited*, and *re-read* by the molecular machinery.

Section 5: Analysis of DNA-Mediated Allosteric Effects

The hypothesis must also account for action at a distance, a common feature of gene regulation (e.g., an enhancer protein influencing a promoter). If a "breathing" event is local, how can it influence a binding event hundreds of base pairs away? The evidence points to DNA itself acting as an allosteric medium, transmitting information along its backbone via both structural and purely dynamic mechanisms.

5.1 Quantifying Structurally-Mediated Allostery: The "Wedge" Experiment

An allosteric effect is one in which a perturbation at one site on a molecule (e.g., DNA) influences the state or binding affinity at a distant second site, without direct contact. DNA is demonstrably capable of this. An experiment utilized a "designed 'wedge' molecule"—a sequence-targeted polyamide—to bind the DNA minor groove and "induce perturbations at specific base steps".[17] This "wedge" was targeted to the binding site of the protein Exd, but *did not* make any contact with the protein itself.[17]
The "wedge" molecule, acting as an "allosteric effector," was "found to enhance Exd binding by 1.5 kcal mol–1".[17] This is a direct, quantitative *proof* that DNA is an allosteric medium. A local structural perturbation (a "wedge," or plausibly, a "DNA bubble" or "kink" from breathing) *can* transmit energy and information *mechanically* along the helix to affect protein binding affinity at a distance. Other mechanical models and experimental data confirm this phenomenon, showing that the interaction energy "oscillates with the periodicity of the double helix" (~10 base pairs) and "decays exponentially" with distance.[23]

5.2 Evidence for Dynamics-Driven Allostery: The Entropy-Mediated Mechanism

A more subtle, and perhaps more powerful, allosteric mechanism exists, one that is "purely dynamics-driven".[18] This mechanism is particularly relevant to the "breathing" hypothesis, as it is mediated *not* by static structural changes, but by changes in *motion* and *entropy*.

Experimental studies of the cooperative binding of cAMP to the CAPN protein provided a "definitive example" of this phenomenon.[18] The data showed that the binding of the *first* cAMP molecule to its site "has a minimal effect on the conformation" (i.e., the static 3D structure) of the *second* binding site.[18] Instead, the allosteric communication occurred because the first binding event caused a "distinct alteration of protein motions," resulting in a "dramatic difference in conformational entropy" at the second site.[18] This change in *entropy* alone was sufficient to drive the observed negative cooperativity.

Another study, on the BAMHI–DNA–GRDBD complex, identified a similar "entropy-mediated mechanism" for allostery that "does not occur through any traditional models" (such as direct protein-protein contact or DNA reorganization).[25]

This evidence is crucial. It provides direct experimental proof that allostery *can be purely dynamic* (i.e., entropic). This provides a powerful, physically-grounded mechanism for the *non-local* action of DNA breathing. A TF binding at an enhancer might not *mechanically* bend the promoter (structural allostery, 5.1). Instead, its binding could *allosterically modulate the breathing dynamics* (the "dynamic length" [1]) at the distant promoter. This would change the *conformational entropy* of that promoter site, thereby changing the binding free energy and probability of RNA polymerase assembly, all without any direct physical contact. This "dynamics-driven allostery" [18] is the plausible macroscopic, functional consequence of the local, physical "breathing" phenomenon.

Section 6: Evidence-Based Conclusions

This analysis, based *exclusively* on the logical and empirical evaluation of the provided data, leads to the following conclusions regarding the hypothesis that local DNA breathing dynamics function as a form of "encoding":

1. **Correlation and Predictive Value:** The provided evidence establishes that statistically significant, quantitative correlations exist between computationally-derived DNA breathing features (e.g., "flipping probability") and experimentally-measured transcription factor binding, both *in vitro* [9] and *in vivo*.[10] Furthermore, the integration of these biophysical features provides *non-redundant, predictive information*, quantifiably improving the accuracy of state-of-the-art TF binding prediction models by up to 9.6% AUROC.[12]
2. **Sequence-Dependent Physical Basis:** The physical properties that govern DNA breathing (e.g., opening enthalpy, stacking stability, flexibility) are *demonstrably* and *quantifiably* dependent on the local dinucleotide sequence.[5] This sequence-dependence of the physical layer, which is distinct from the primary sequence itself, is the *fundamental prerequisite* for the "encoding" hypothesis.
3. **Plausible Causal Mechanisms:** The data supports *multiple*, non-exclusive causal mechanisms by which these local dynamics can physically influence protein binding:
    - **Mechanical Perturbation:** Local breathing (bubbles, kinks) "drastically" alters

macroscopic mechanical properties like persistence length [16] and accounts for the *majority* (~80%) of the DNA curvature observed in functional, protein-bound structures.[15]
    - **Conformational Selection:** This suggests a "conformational selection" model, where proteins bind to and "select" favorable, thermally-induced dynamic conformations (kinks), rather than exclusively "inducing" them. The "encoding" is thus the sequence-dependent *probability* of sampling these required conformations.[15]
    - **Allosteric Transmission:** DNA is confirmed to be an allosteric medium, capable of transmitting information via *structural* perturbations (quantified at 1.5 kcal/mol) [17] and *purely dynamic/entropic* perturbations.[18]
4. **Reciprocal Feedback and Complexity:** The relationship is not unidirectional. Evidence shows that protein binding can, in turn, *augment* local DNA breathing dynamics [7], indicating a complex, reciprocal feedback system where the "encoding" is both *read* and *edited* by the cellular machinery.
5. **Known Limitations:** The "sign and magnitude" of the observed *in vivo* correlations "vary substantially across TF families".[9] This indicates that the "encoding" is not a universal simple code (e.g., "high breathing = high binding") but is, rather, a complex physical signal that is *interpreted* differently by different protein families, precluding a simple, monolithic model.

Final Statement: Based *only* on the provided evidence, the hypothesis that local DNA breathing dynamics function as a "form of encoding" is empirically supported. The dynamics are (a) demonstrably sequence-dependent [5], (b) quantitatively correlated with transcription factor binding *in vitro* and *in vivo* [9], (c) provide non-redundant predictive power in computational models [12], and (d) are linked to TF binding via plausible, experimentally-verified physical mechanisms, including conformational selection [15] and dynamics-driven allostery.[18]

Works cited

1. Examining DNA Breathing with pyDNA-EPBD - PMC, accessed November 5, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10515784/
2. Watching DNA breath one molecule at a time - PNAS, accessed November 5, 2025, https://www.pnas.org/doi/10.1073/pnas.1316493110
3. Examining DNA breathing with pyDNA-EPBD | Bioinformatics - Oxford Academic, accessed November 5, 2025, https://academic.oup.com/bioinformatics/article/39/11/btad699/7441499
4. Examining DNA Breathing with pyDNA-EPBD - bioRxiv, accessed November 5, 2025, https://www.biorxiv.org/content/10.1101/2023.09.09.557010.full.pdf
5. nonlinear dynamic model of DNA with a sequence-dependent stacking term | Nucleic Acids Research | Oxford Academic, accessed November 5, 2025, https://academic.oup.com/nar/article/37/7/2405/1017133
6. 1H NMR determination of base-pair lifetimes in oligonucleotides ..., accessed November 5, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC135820/

7. Single-molecule FRET and linear dichroism studies of DNA ... - PNAS, accessed November 5, 2025, https://www.pnas.org/doi/10.1073/pnas.1314862110
8. Multicolor single-molecule FRET for DNA and RNA processes - PubMed - NIH, accessed November 5, 2025, https://pubmed.ncbi.nlm.nih.gov/33894656/
9. Contribution of DNA breathing to physical interactions with ..., accessed November 5, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11785057/
10. Contribution of DNA breathing to physical interactions with transcription factors - PubMed, accessed November 5, 2025, https://pubmed.ncbi.nlm.nih.gov/39896490/
11. Contribution of DNA breathing to physical interactions with transcription factors - bioRxiv, accessed November 5, 2025, https://www.biorxiv.org/content/10.1101/2025.01.20.633840v1.full.pdf
12. DNA breathing integration with deep learning foundational model advances genome-wide binding prediction of human transcription factors - PubMed Central, accessed November 5, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC11514457/
13. Sequence-dependent dynamics of duplex DNA: the applicability of a ..., accessed November 5, 2025, https://pubmed.ncbi.nlm.nih.gov/12496111/
14. Sequence-dependence of the energetics of opening of at basepairs ..., accessed November 5, 2025, https://pubmed.ncbi.nlm.nih.gov/15454449/
15. Breathing, bubbling, and bending: DNA flexibility from multimicrosecond simulations | Phys. Rev. E - Physical Review Link Manager, accessed November 5, 2025, https://link.aps.org/doi/10.1103/PhysRevE.86.021903
16. How double-stranded DNA breathing enhances its flexibility and ..., accessed November 5, 2025, https://link.aps.org/doi/10.1103/PhysRevE.81.021906
17. Allostery: DNA Does It, Too | ACS Chemical Biology, accessed November 5, 2025, https://pubs.acs.org/doi/10.1021/cb800070s
18. Dynamically driven protein allostery - PMC - NIH, accessed November 5, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC2757644/
19. Contribution of DNA breathing to physical interactions with transcription factors - bioRxiv, accessed November 5, 2025, https://www.biorxiv.org/content/10.1101/2025.01.20.633840v1
20. Advancing Transcription Factor Binding Site Prediction Using DNA Breathing Dynamics and Sequence Transformers via Cross Attention - PMC - PubMed Central, accessed November 5, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10827174/
21. Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations - PMC - PubMed Central, accessed November 5, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC149832/
22. DNA Dynamics and Single-Molecule Biology | Chemical Reviews - ACS Publications, accessed November 5, 2025, https://pubs.acs.org/doi/10.1021/cr4004117
23. Allosteric interactions in a birod model of DNA | Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences - Journals, accessed November 5, 2025, https://royalsocietypublishing.org/doi/10.1098/rspa.2018.0136

24. On the Use of Molecular Dynamics Simulations for Probing Allostery through DNA - PMC, accessed November 5, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC4776040/
25. Allosterism and signal transfer in DNA | Nucleic Acids Research - Oxford Academic, accessed November 5, 2025, https://academic.oup.com/nar/article/46/15/7554/5037728