Maximum Entropy in the Gaussian Network Model: A Thermodynamic Reference State for Protein Dynamics

Burak Erman

Chemical and Biological Engineering, Koc University, Istanbul, Turkey

email: berman@ku.edu.tr

Keywords: Gaussian Network Model, Maximum Entropy, Protein Dynamics, Thermodynamic Reference State , Soft Matter, Kirchhoff Matrix , NMR Correlations , Allosteric Mechanisms , Helmholtz Free Energy, Computational Biophysics

Abstract

This paper presents a rigorous proof that the Gaussian Network Model (GNM) with uniform spring constants maximizes entropy subject only to constraints imposed by the native contact topology. We demonstrate that this entropic optimality establishes the GNM as a fundamental thermodynamic reference state for protein dynamics, rather than merely a computational convenience. Our analysis shows that proteins, as soft matter systems, are predominantly governed by entropic considerations, with thermal fluctuations distributed among their degrees of freedom in a manner that maximizes configurational entropy. The uniform-spring constant GNM provides the least biased representation of protein dynamics, explaining its remarkable success in predicting experimental observables including NMR-measured correlations. We discuss how applying additional constraints to this maximum-entropy reference state, such as non-uniform spring constants or specific covariance requirements, inevitably reduces entropy while increasing Helmholtz free energy. This reference state framework offers new insights into allosteric mechanisms, drug binding effects, and the evolutionary balance between thermodynamic optimality and functional specialization in proteins. Previous approaches that constrained GNM fluctuations while maximizing entropy are contextualized within this theoretical foundation, highlighting pathways for model refinement while maintaining entropic principles.

1. Introduction

Proteins are flexible, entropy-driven systems whose thermally induced and coordinated fluctuations contribute significantly to their biological functions. The Gaussian Network Model (GNM), a minimalist framework employing uniform harmonic springs and native contact topology, assumes Gaussian fluctuations, justified by the Central Limit Theorem, resulting in linear, spring-like interactions that are

represented using a Kirchhoff Laplacian matrix, which incorporates interactions within the first coordination shell of each residue[1, 2].

In this paper, we prove that the entropy of the protein described by the GNM is maximum subject to the constraint encoded in this Laplacian matrix. The entropic properties of coarse-grained protein models have been extensively studied [3-13], though the specific maximum entropy property of the uniform spring constant GNM has not been previously demonstrated. Molecular dynamics (MD) simulations have consistently shown that proteins sample conformations in a manner that maximizes configurational entropy under native state constraints [14], providing computational evidence that complements our theoretical analysis. Since maximum entropy represents a fundamental thermodynamic imperative that governs all physical systems under given constraints [15], our proof establishes the GNM as not merely simple but optimal for faithfully representing native protein dynamics. This optimality is particularly relevant given observations that proteins exist in states of marginal stability, making their entropic properties highly responsive to external perturbations [15-17].

The Gaussian Network Model's maximum entropy property, yielding the least biased correlations for a given topology, provides a theoretical foundation for its accurate prediction of experimentally measurable cross-correlations [18, 19] and explains its utility in understanding diverse biological phenomena, including allosteric information transfer, molecular transport mechanisms, and structure-based drug design [20].

2. The GNM Maximum Entropy Theorem

Theorem: Given a connected graph $G=(V,E)$, V representing vertices, E edge set, with $|V|=N$ and number of edges $|E|$ in a Gaussian Network Model, the entropy is maximized when all spring constants $k_{ij}=k$ for $\{i,j\}\in E$ under the constraint $\sum_{\{i,j\}\in E} k_{ij}=|E|k$.

The Kirchhoff matrix Γ is defined as:

$$\Gamma_{ij}=-k_{ij} \quad if \quad \{i,j\}\in E$$

$$\Gamma_{ii}=\sum_{j\neq i} k_{ij} \tag{1}$$

$$\Gamma_{ij}=0 \quad otherwise.$$

The eigenvalues of Γ are $\lambda_1=0,\lambda_2,...,\lambda_N>0$. The entropy is:

$$S=\frac{1}{2}\left[N\left(1+\log\left(2\pi T\right)\right)-\log\det\Gamma\right] \tag{2}$$

where $\det\Gamma$ is the product of non-zero eigenvalues and log indicates natural logarithm. Maximizing S requires minimizing

$$\log\det\Gamma=\sum_{i=2}^{N}\log\lambda_i \tag{3}$$

***Lemma: Matrix-Tree Theorem*[21]**

For a connected graph G, the product of non-zero eigenvalues is:

$$\prod_{i=2}^{N} \lambda_i = \sum_{T \in \Upsilon} \prod_{\{i,j\} \in T} k_{ij} \tag{4}$$

where Υ is the set of all spanning trees, and each tree T has N-1 edges. For uniform spring constants $k_{ij} = k$:

$$\prod_{i=2}^{N} \lambda_i = |\Upsilon| k^{N-1} \tag{5}$$

where $|\Upsilon|$ is the number of spanning trees.

Main proof:

Consider two Kirchoff matrices:

Γ_1 : Uniform spring constants $k_{ij} = k$ for all $\{i, j\} \in E$.

Γ_2 : Variable spring constants k_{ij} with $\sum_{\{i,j\} \in E} k_{ij} = |E| k$.

For Γ_1:

$$\det(\Gamma_1) = |\Upsilon| k^{N-1} \text{ , } \log\det(\Gamma_1) = \log|\Upsilon| + (N-1)\log k \tag{6}$$

For Γ_2:

$$\det(\Gamma_2) = \sum_{T \in \Upsilon} \prod_{\{i,j\} \in T} k_{ij} \text{ , } \log\det(\Gamma_2) = \log\left(\sum_{T \in \Upsilon} \prod_{\{i,j\} \in T} k_{ij} \right) \tag{7}$$

Define $p_T = \prod_{\{i,j\} \in T} k_{ij}$, the product of spring constants over the $N-1$ edges in spanning tree T. We need to show:

$$\sum_{T \in \Upsilon} p_T \geq |\Upsilon| k^{N-1} \tag{8}$$

with equality when $k_{ij} = k$. By the arithmetic-geometric mean (AM-GM) inequality[22]:

$$\frac{\sum_{T \in \Upsilon} p_T}{|\Upsilon|} \geq \left(\prod_{T \in \Upsilon} p_T \right)^{1/|\Upsilon|} \tag{9}$$

so:

$$\sum_{T\in\Upsilon} p_T \geq |\Upsilon| \left(\prod_{T\in\Upsilon} p_T \right)^{1/|\Upsilon|} \tag{10}$$

Equality holds when $p_T = p_{T'}$ for all $T, T' \in \Upsilon$, i.e., the product of spring constants is the same for every spanning tree. Since p_T depends on the specific edges $\{i, j\} \in T$, equal p_T across all trees implies $k_{ij} = k$ for all $\{i, j\} \in E$, as non-uniform k_{ij} would produce varying products depending on the edges in each tree. The constraint $\sum k_{ij} = |E| k$ ensures that $k_{ij} = k$ is the only solution satisfying $p_T = p_{T'}$ for all T, T'. In the uniform case, $k_{ij} = k$ gives $p_T = k^{N-1}$, satisfying the constraint. Thus:

$$\log\det(\Gamma_2) \geq \log\det(\Gamma_1) \tag{11}$$

implying $S_2 \leq S_1$ with equality when $k_{ij} = k$.

The function given by

$$f\left(\{k_{ij}\}\right) = \log\left(\sum_{T\in\Upsilon} \prod_{\{i,j\}\in T} k_{ij} \right) \tag{12}$$

is convex because it is the logarithm of a sum of monomials $p_T = \prod_{\{i,j\}\in T} k_{ij}$. Functions of the form $\log\left(\sum_i \prod_j x_j^{a_{ij}} \right)$ are convex , as the sum of monomials is convex, and the logarithm preserves convexity[23]. Convexity guarantees that any local minimum is global. Since $k_{ij} = k$ satisfies the first-order optimality condition (via AM-GM), and the constraint $\sum k_{ij} = |E| k$, it is the **unique global minimizer** of f .

Conclusion

The entropy is maximized when $k_{ij} = k$, as this minimizes $\log\det(\Gamma)$. The AM-GM inequality, Lagrange multiplier analysis, and convexity of f confirm that any deviation from uniform spring constants increases $\log\det(\Gamma)$, reducing entropy.

Since the energy, $E = \frac{1}{2}(N-1)$ in $k_B T$ units, of the GNM is independent of k_{ij} , maximizing entropy minimizes the Helmholtz free energy, $A = U - TS$.

3. Entropy Maximization and Its Implications for Protein Dynamics

3.1 Entropic Dominance in Protein Systems

The predominant entropic nature of proteins is a direct consequence of their soft matter characteristics [17]. When constraints are applied to a protein system, it naturally evolves toward states that maximize entropy subject to those constraints. This maximization process is particularly critical for understanding protein function, as it leads not only to uncorrelated fluctuations but, more importantly, to coordinated motions essential for biological activity.

3.2 The GNM as a Thermodynamic Reference State

Our theorem demonstrates that when contact potentials are uniform (with identical off-diagonal elements in the Kirchhoff matrix), the GNM achieves maximum entropy under the sole constraint of maintaining a given contact topology. This establishes the uniform-spring constant GNM as a fundamental thermodynamic reference state against which modified models can be compared.

The concept of a reference state is central to thermodynamics and statistical mechanics. Just as the ideal gas serves as a reference state for real gases—allowing deviations to be quantified and understood in terms of molecular interactions—the uniform-spring constant GNM provides a reference state for protein dynamics. This reference state represents the least biased probability distribution of conformational fluctuations consistent with the topological constraints of the protein structure.

The significance of this reference state extends beyond computational convenience. It provides a baseline against which we can measure and interpret the effects of additional constraints or perturbations to the system. Any deviation from the uniform-spring constant model necessarily reduces entropy, indicating a departure from this reference state toward more specialized dynamic behavior.

Implications of the Reference State Framework:

1. **Quantifiable Thermodynamic Costs**: The difference in entropy (or equivalently, in Helmholtz free energy) between the reference GNM and a constrained model provides a direct measure of the thermodynamic "cost" of imposing additional constraints.
2. **Evolutionary Insights**: The existence of a maximum entropy reference state suggests that evolution may have optimized proteins to balance between the thermodynamic drive toward maximum entropy and the functional requirements that necessitate specific dynamic patterns.
3. **Model Interpretation**: Any refinement to the GNM (such as non-uniform spring constants) can now be understood not merely as a computational adjustment but as a thermodynamically meaningful deviation from the reference state.
4. **Perturbation Analysis**: Changes in protein structure, such as those induced by mutations, ligand binding, or environmental factors, can be evaluated based on how they alter the system's proximity to this reference state.

3.3 Specific Implications of the GNM Maximum Entropy Theorem

Specifically, our findings show that:

1. Any additional constraints applied to the GNM, whether imposing pair correlations or modifying entries of the Kirchhoff matrix while preserving contact topology, will produce a model with entropy lower than that of the original GNM.

2. Once constraints are applied, entropy maximization principles dictate that the system should reorganize to maximize its entropy within these new boundaries.
3. Since the energy of the GNM is independent of spring constants [24], the Helmholtz free energy of the entropy-maximized system will be minimum. Conversely, any constrained system will inevitably have higher Helmholtz free energy than the uniform spring constant GNM.

3.3 Applications of Constrained GNM Models

The application of constraints to the GNM opens avenues for analyzing specific biological processes. For instance, constraining fluctuations at a drug binding site and subsequently maximizing the entropy of the protein-ligand complex can reveal information transfer pathways critical for allosteric drug action.

Two previous examples exist where GNM fluctuations have been constrained followed by entropy maximization:

1. Razon and Bahar [25] maximized entropy subject to specific covariance constraints (C_ij = c_ij, derived from experimental data such as NMR), which yielded non-uniform spring constants.
2. Erman [26] applied information theory to incorporate many-body effects into the spring constant matrix, followed by entropy maximization.

Both approaches resulted in modified GNM formulations with higher Helmholtz free energy compared to the uniform spring constant model. The difference between these constrained models and the uniform GNM effectively quantifies the potential of the latter as a reference state.

4. Theoretical Foundation and Predictive Power

The Gaussian Network Model's maximum entropy property, yielding the least biased fluctuation distribution for a given topology, underpins its robustness and predictive power. By maximizing configurational entropy with uniform harmonic springs, the GNM ensures universal predictions, applicable across proteins from small domains to large complexes, unlike models requiring protein-specific parameters.

Its Gaussian fluctuations, justified by the Central Limit Theorem and validated by numerous molecular dynamics simulations showing near-Gaussian distributions [27], accurately predict NMR observables, including B-factors, $\left(\Gamma^{-1}\right)_{ii}$, and cross-correlations $\left(\Gamma^{-1}\right)_{ij}$ measurable via NOE and residual dipolar couplings. These correlations show allosteric pathways, enabling structure-based drug design by identifying ligand-binding effects. The GNM's entropy maximization at minimum Helmholtz free energy ensures thermodynamic consistency, formalizing its success beyond the 1997 framework's implicit entropic role [1].

5. Model Extensions and Limitations

5.1 Adaptability to Perturbed States

The GNM's adaptability to perturbed states is a key strength. For ligand-bound proteins, an updated Kirchhoff Laplacian matrix Γ reflecting altered contacts re-maximizes entropy, predicting shifted correlations critical for drug targeting. However, the harmonic assumption limits the GNM to Gaussian fluctuations, potentially missing non-Gaussian dynamics observed in MD, such as subtle skewness in

residue displacements. This raises the question of improving the GNM to capture such effects while retaining its entropic optimality.

5.2 Model Refinements

One improvement involves non-uniform spring constants ($k_{ij} \neq k$), adjusting Gamma to prioritize specific correlations in functionally important regions. While maintaining Gaussian fluctuations, this requires optimizing k_{ij} to maximize entropy under new constraints (e.g., fixed covariances), as arbitrary k_{ij} introduce bias, reducing entropy. Explicit entropy maximization, via constrained optimization, ensures optimality but sacrifices the GNM's simplicity [25, 26].

Alternatively, geometric nonlinearity, nonlinear strain-displacement relations preserving quadratic potentials, offers a promising extension. By modifying Γ to account for nonlinear contact geometry, the GNM can model large deformations in perturbed protein structures while maximizing entropy and retaining Gaussianity, as suggested by MD's near-Gaussian distributions. This aligns with the GNM's minimalist principles, avoiding the complexity of non-harmonic potentials, which demand re-deriving non-Gaussian distributions.

Non-harmonic models, capturing conformational switches, further depart from the GNM's framework, requiring new entropy maximization strategies and computational intensity akin to MD. Such improvements trade universality for specificity, highlighting the GNM's strength: its entropy-maximized simplicity excels for collective dynamics.

6. Implications for Protein Function: An Entropic Perspective

6.1 Entropy Reduction as a Functional Mechanism

These findings suggest that the uniform spring constant GNM represents an idealized state where entropy is maximized under minimal constraints. Real proteins operate under additional constraints imposed by their specific functional requirements. The entropy maximization principle provides a theoretical framework for understanding how proteins balance the need for specific functional motions against the thermodynamic drive toward maximum entropy.

A compelling analogy exists between protein dynamics and the elasticity of cross-linked polymeric rubbers [28]. When rubber is stretched, polymer chains are deformed into elongated conformations, decreasing the system's entropy. This entropy reduction is converted into mechanical work, defining the essential function of rubber elasticity. Similarly, in proteins, when constraints are applied (such as ligand binding, post-translational modifications, or conformational changes), the system's entropy decreases from its maximum value in the unconstrained GNM. This entropy reduction can be viewed as being converted into "functional work", the specific, coordinated motions required for biological activity.

In this framework, allosteric regulation and signal transduction can be understood as processes where local constraints (e.g., binding events) reduce local entropy, creating an "entropy gradient" that drives information transfer through the protein structure. The system reorganizes to maximize entropy under these new constraints, generating specific motion patterns that constitute the protein's functional response. Thus, protein function emerges from the interplay between entropy maximization and targeted constraint application.

6.2 Energetic Cost of Functional Specialization

The quantification of free energy differences between constrained and uniform GNM models offers a measure of the energetic cost associated with functional specialization in proteins. This framework potentially allows us to evaluate the evolutionary trade-offs between functional efficiency and thermodynamic optimality in protein systems.

Just as engineered materials balance elasticity with specific mechanical properties, evolution appears to have optimized proteins to balance entropic freedom with functional constraints. Highly conserved, functionally critical regions often experience the largest entropy reduction upon constraint application, suggesting that the "entropic cost" of functional specialization has been minimized through evolutionary selection.

7. Applications and Future Directions

For drug design, the GNM's correlations guide initial allosteric targeting, while extended models refine predictions for specific perturbations. Future NMR experiments validating cross-correlations in various protein systems, combined with MD insights, can guide these improvements, ensuring the GNM's entropic framework bridges theory and application in protein dynamics.

8. Conclusion

The entropy is maximized when spring constants are uniform, as this minimizes the determinant of the Kirchhoff matrix. The AM-GM inequality and convexity confirm that any deviation from uniform spring constants increases this determinant, reducing entropy.

Since the energy of the GNM (in k_BT units) is independent of spring constants, maximizing entropy minimizes the Helmholtz free energy. This establishes the GNM not merely as a computational convenience but as a thermodynamically optimal reference state for protein dynamics under topological constraints.

The reference state concept allows us to understand protein dynamics through a thermodynamic perspective, where functional specialization and evolutionary adaptation can be quantified as departures from the maximum entropy state. This theoretical foundation enhances our ability to interpret experimental measurements of protein dynamics and may guide the development of more refined models that balance universal principles with system-specific details.

By establishing the uniform spring constant GNM as the maximum entropy reference state, we provide a unifying framework that connects the empirical success of the GNM with fundamental principles of statistical thermodynamics. This connection not only validates the widespread use of the GNM in computational biology but also opens new avenues for understanding protein function as an interplay between entropic maximization and functional constraint.

References

[1] Bahar I, Atilgan A R and Erman B 1997 Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential *Fold Des* **2** 173-81

[2] Haliloglu T, Bahar I and Erman B 1997 Gaussian dynamics of folded proteins *Physical Review Letters* **79** 3090-3

[3] Das P, Matysiak S and Clementi C 2005 Balancing energy and entropy: A minimalist model for the characterization of protein folding landscapes *Proceedings of the National Academy of Sciences* **102** 10141-6

[4] Clementi C 2008 Coarse-grained models of protein folding: toy models or predictive tools? *Current opinion in structural biology* **18** 10-5

[5] Foley T T, Shell M S and Noid W G 2015 The impact of resolution upon entropy and information in coarse-grained models *The Journal of chemical physics* **143**

[6] Jin J, Pak A J, Durumeric A E, Loose T D and Voth G A 2022 Bottom-up coarse-graining: Principles and perspectives *J Chem Theory Comput* **18** 5759-91

[7] Souza P C, Alessandri R, Barnoud J, Thallmair S, Faustino I, Grünewald F, Patmanidis I, Abdizadeh H, Bruininks B M and Wassenaar T A 2021 Martini 3: a general purpose force field for coarse-grained molecular dynamics *Nature methods* **18** 382-8

[8] Tozzini V 2010 Minimalist models for proteins: a comparative analysis *Quarterly reviews of biophysics* **43** 333-71

[9] Noid W G 2023 Perspective: Advances, challenges, and insight for predictive coarse-grained models *The Journal of Physical Chemistry B* **127** 4174-207

[10] Heath A P, Kavraki L E and Clementi C 2007 From coarse-grain to all-atom: toward multiscale analysis of protein landscapes *Proteins: Structure, Function, and Bioinformatics* **68** 646-61

[11] Ganguly D and Chen J 2011 Topology-based modeling of intrinsically disordered proteins: balancing intrinsic folding and intermolecular interactions *Proteins: Structure, Function, and Bioinformatics* **79** 1251-66

[12] Tozzini V 2010 Multiscale modeling of proteins *Accounts of chemical research* **43** 220-30

[13] Hills Jr R D, Lu L and Voth G A 2010 Multiscale coarse-graining of the protein energy landscape *PLoS computational biology* **6** e1000827

[14] Cavalli A, Camilloni C and Vendruscolo M 2013 Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle *The Journal of chemical physics* **138**

[15] Dill K and Bromberg S 2010 *Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience*: Garland Science)

[16] Dill K and Alonso D 1988 *Protein Structure and Protein Engineering*: Springer) pp 51-8

[17] Whitford P C, Sanbonmatsu K Y and Onuchic J N 2012 Biomolecular dynamics: order–disorder transitions and energy landscapes *Reports on Progress in Physics* **75** 076601

[18] Xu D, Meisburger S P and Ando N 2021 Correlated motions in structural biology *Biochemistry* **60** 2331-40

[19] Meisburger S P, Thomas W C, Watkins M B and Ando N 2017 X-ray scattering studies of protein structural dynamics *Chemical reviews* **117** 7615-72

[20] Erman B 2025 Fluctuation-driven synergy, redundancy, signal to noise ratio and error correction in protein allostery *Physical Biology* **22** 026003

[21] Bollobás B 1998 *Modern graph theory* vol 184: Springer Science & Business Media)

[22] Steele J M 2004 *The Cauchy-Schwarz master class: an introduction to the art of mathematical inequalities*: Cambridge University Press)

[23] Boyd S P and Vandenberghe L 2004 *Convex optimization*: Cambridge university press)

[24] Bahar I, Atilgan A R, Demirel M C and Erman B 1998 Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability *Physical Review Letters* **80** 2733
[25] Lezon T R and Bahar I 2010 Using entropy maximization to understand the determinants of structural dynamics beyond native contact topology *PLoS computational biology* **6** e1000816
[26] Erman B 2025 Extending the Gaussian Network Model: Integrating Local, Allosteric, and Structural Factors for Improved Residue-Residue Correlation Analysis *BioRxiv*
[27] Micheletti C, Carloni P and Maritan A 2004 Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models *Proteins: Structure, Function, and Bioinformatics* **55** 635-45
[28] Mark J E and Erman B 2007 *Rubberlike elasticity: a molecular primer*: Cambridge University Press)