Molecular Thermodynamics of KRAS Activation

Fatma Senguler Ciftci and Burak Erman*

Department of Chemical and Biological Engineering, Koc University, Istanbul, Turkey

*Corresponding author: berman@ku.edu.tr,

ORCID: FSC 0009-0001-6901-6364; BE 0000-0002-2496-6059

Abstract

The GTPase KRAS executes a conformational switch between a GTP-bound active and a GDP-bound inactive states, which are central to oncogenic signalling, yet the structural basis of this switching at the level of residue-contact network organization remains incompletely characterised by conventional pairwise analyses. Here we apply a rigorous statistical-mechanical framework, grounded in the weighted Kirchhoff Laplacian and the Matrix-Tree Theorem, to construct spanning-tree partition functions for residue contact graphs derived from two crystallographic structures: **6GOD** (active, GTP-analog-bound; 172 residues, 830 contacts) and **4OBE** (inactive, GDP-bound; 169 residues, 809 contacts). The log-partition function $\log Z$, the network free energy $F = -kT \log Z$, the mean contact energy, $\langle E \rangle$, the heat capacity C_v, and the thermodynamic entropy S are computed across an effective temperature sweep from $kT = 0.3$ to 6.0. Edge marginal inclusion probabilities, $P(e \in T)$, obtained via effective-resistance theory, serve as topology-aware measures of contact essentiality. Differential analysis reveals that the active state consistently carries a higher mean contact energy ($\Delta\langle E \rangle > 0$) yet also a higher conformational entropy ($\Delta S > 0$), with the free energy crossover $\Delta F = 0$ occurring at $kT \approx 2.41$, an intrinsic thermodynamic balance independent of any arbitrary additive reference. Switch I (residues 25–40) exhibits the largest state-dependent ΔP, identifying it as the primary allosteric locus of nucleotide-driven network reorganisation.

Keywords: KRAS · spanning-tree partition function · Kirchhoff Laplacian · Matrix-Tree Theorem · allosteric network · conformational entropy · edge inclusion probability · GTPase

1. Introduction

KRAS is the most frequently mutated oncogene in human cancer, with activating mutations at codons G12 and G13 found in roughly 30% of all tumours [1]. Its function as a molecular switch cycling between GTP-bound active and GDP-bound inactive conformations regulates proliferation, survival, and differentiation through downstream effectors including RAF, PI3K, and RALGDS [2]. The clinical importance of KRAS has motivated intensive efforts to develop direct small-molecule inhibitors [3]. Thermodynamic and kinetic aspects of KRAS allostery and of its downstream interactions are studied by Manley and Lin [4]. Conformational resolution of nucleotide cycling and effector interactions across multiple GTPases has been characterized by Killoran and Smith [5]. There is a substantial body of valuable work addressing KRAS allostery through GTP-dependent conformational dynamics, reorganization of interaction networks, and thermodynamic analyses of binding energetics [6, 7, 8]. However, to the best of our knowledge, none of these studies presents an approach that analyzes allosteric communication within a spanning tree–based statistical mechanical framework that explicitly accounts for the full ensemble of network configurations. In particular, it is not known how the global architecture of residue contacts, rather than any single interaction, collectively enables or restricts allosteric signal transmission between the nucleotide-binding pocket and distal effector surfaces. This question is especially pertinent when comparing the GTP-bound active and GDP-bound inactive states of KRAS, where global network reorganization rather than local structural rearrangement may be the primary determinant of functional switching.

Conventional structural analyses compare active and inactive states by identifying residue pairs that gain or lose contacts upon nucleotide exchange [9]. Elastic network models extend this picture by providing a coarse-grained description of collective motions; however, being defined about a single energy minimum, they cannot account for the multiplicity of communication pathways accessible to the network [10]. Recent GNM-based work has made important progress in this direction by deriving edge centrality measures that implicitly enumerate communication pathways through the matrix-tree theorem, revealing conserved allosteric bottlenecks in KRAS [11]; however, this framework does not assign explicit statistical weights to individual spanning tree configurations and therefore does not yield a temperature-dependent partition function from which heat capacity, switching free energy, and entropic contributions to state transitions can be extracted. Neither framework provides a thermodynamically consistent partition function, one that assigns statistical weights to contact configurations and yields well-defined free energy, entropy, and heat capacity for the network. In the absence of such a description, the thermodynamic cost of switching between states and the relative contributions of energetic and entropic terms cannot be assessed within a static structural framework. The spanning-tree partition function framework introduced here addresses precisely this gap by summing over the full ensemble of contact network configurations, yielding thermodynamic quantities, including free energy, entropy, and heat capacity, that are directly comparable between the GTP-bound active and GDP-bound inactive states of KRAS.

Here, we adopt an alternative formulation based on algebraic graph theory. The residue contact network is treated as a weighted graph, and its spanning-tree partition function is evaluated exactly using the Matrix–Tree Theorem and the weighted Kirchhoff Laplacian. This construction enumerates all spanning trees compatible with the network connectivity and provides closed-form thermodynamic potentials without recourse to stochastic sampling. Applied to KRAS, the analysis shows that activation is accompanied by a marked increase in conformational entropy. The active state is therefore not merely a rearranged structure, but a thermodynamic ensemble with enhanced configurational flexibility.

Here, we adopt an independent formulation grounded in algebraic graph theory [12, 13]. Each KRAS

structure is represented as a weighted residue contact graph, and its spanning-tree partition function is evaluated as a Boltzmann-weighted sum over all connected, cycle-free subgraphs spanning the full residue set. A spanning tree represents a minimal contact configuration that preserves global network connectivity; its statistical weight reflects the total interaction energy of that configuration at a given effective temperature. Summing over all such trees yields a statistical-mechanical description of how allosteric signals are distributed across competing pathways. Critically, changes in this combinatorial diversity upon nucleotide exchange are not accessible to pairwise contact analyses or single-minimum elastic descriptions, and it is precisely these changes that we argue define the thermodynamic distinction between the active and inactive states of KRAS.

The formal basis of the approach is the Matrix–Tree Theorem and the associated weighted Kirchhoff Laplacian [12, 13]. Importantly, this framework yields exact, closed-form thermodynamic quantities without reliance on stochastic sampling, normal-mode approximations, or harmonic assumptions. Marginal edge inclusion probabilities $P(e \in T)$, obtained from effective-resistance theory [14, 15], provide a contact-by-contact measure of how critical each residue interaction is for maintaining global network connectivity. Together, these observables enable a decomposition of the thermodynamic differences between the active and inactive networks into energetic and entropic components, and allow identification of the residue-level regions where the two states diverge most strongly.

Applying this framework to two crystallographic wild-type KRAS structures ,6GOD (GTP analog bound, active) and 4OBE (GDP bound, inactive), we find that activation cannot be described as a purely energetic transition. Instead, it reflects an entropy–enthalpy compensation: the active state incurs a systematic energetic penalty $\Delta\langle E\rangle > 0$) that is offset by a corresponding gain in conformational entropy ($\Delta S > 0$), with the free-energy balance crossing at a well-defined intrinsic temperature. Switch I (residues 25–40) emerges as the primary site of nucleotide-driven structural reorganization. These results provide a global, thermodynamically grounded perspective on KRAS allostery that goes beyond what pairwise contact analyses or single-minimum elastic descriptions can offer.

2. Methods

2.1 Structural Data

Two crystal structures were retrieved from the RCSB Protein Data Bank: **6GOD** (wild type KRAS covalently modified with a GTP analogue; chain A; resolution 1.90 Å) and **4OBE** (wild-type KRAS in complex with GDP; chain A; resolution 1.60 Å). C_α coordinates were extracted for all residues without alternative conformations.

2.2 Contact Graph Construction

A residue contact graph $G = (V, E)$ is built by placing an edge between residues i and j whenever their C_α atoms lie within a distance threshold $d_c = 8.0$ Å. Edge weights encode contact geometry through the linear energy function $E(e) = d(e)$, where $d(e)$ is the C_α–C_α distance in Ångströms. The largest connected component is retained: 6GOD yields 172 nodes and 830 edges; 4OBE yields 169 nodes and 809 edges. The residues of 6GOD.pdb and 4OBE.pdb are identical except the first residue, which is SER1 in 6GOD.pdb versus MET1 in 4OBE.pdb. However, our method does not depend on the type of the residues but only on the topology of the structures. The difference comes from the three missing residues MET170, SER171, and LYS172 at the C-terminal of 4OBE.pdb, which has no effect on chain topology.

2.3 Weighted Kirchhoff Laplacian

At effective temperature kT, each edge $e = (i, j)$ is assigned a Boltzmann weight

$$w_{ij}(kT) = \exp\left[-\frac{E_{ij}}{kT}\right]. \tag{1}$$

These weights are used to construct the weighted Kirchhoff (graph Laplacian) matrix L, whose off-diagonal and diagonal entries are defined respectively as

$$L_{ij} = -w_{ij}\ (i \neq j), \qquad L_{ii} = \sum_{j \neq i} w_{ij}. \tag{2}$$

2.4 Effective Temperature and the Physically Sensible Regime

The Boltzmann weight in Eq. (1) depends on the ratio E_{ij}/kT. Since the interaction energy is taken proportional to the C_α–C_α distance,

$$E_{ij} = \alpha \cdot d_{ij}, \tag{3}$$

with an unknown proportionality constant α that converts length into energy, the weight simplifies to

$$w_{ij} = \exp\left(-\frac{d_{ij}}{kT}\right), \tag{4}$$

where we define the *effective temperature*

$$kT_{\text{eff}} \equiv \frac{k_B T}{\alpha}, \tag{5}$$

in which α^{-1} carries units of Åenergy, so that kT_{eff} has units of Ångströms. The parameter kT therefore controls how steeply the weight decays with distance: a small kT means that even moderately long contacts become exponentially rare, while a large kT flattens the weight distribution and makes all contacts within the cutoff more comparable. For brevity, we write kT throughout the remainder of the paper in place of kT_{eff}, with the understanding that it always refers to this length-scale parameter. By contrast, the Gaussian Network Model (GNM) [16] employs a binary weighting scheme:$w_{ij} = 1$ if $d_{ij} < d_c$, 0 if $d_{ij} \geq d_c$ where the cutoff distance d_c is typically in the range 7—10 Å. In GNM, all contacts within the cutoff are treated as equivalent, while those beyond are ignored entirely. The present model generalises this approach by introducing a distance-dependent continuum of weights, enabling a richer thermodynamic description of the contact network.

Although the absolute value of kT cannot be fixed without knowing α, a physically sensible window can be identified from two requirements:

1. *Stability.* Contacts deep inside the first coordination shell ($d \approx 4$ Å) should carry appreciable weight, ensuring the core structure is represented in the spanning-tree ensemble. Adopting a conservative threshold of 1% :

$$e^{-4/kT} > 0.01 \implies kT > \frac{4}{-\ln 0.01} \approx 0.87 \text{ Å}. \tag{6}$$

2. *Dynamical relevance.* Contacts near the cutoff (i.e. $d \approx 8$ Å) should be present but weak, so the ensemble

retains diversity without being dominated by long, marginal contacts:

$$e^{-8/kT} < 0.01 \implies kT < \frac{8}{-\ln 0.01} \approx 1.74 \text{ Å}. \tag{7}$$

Together, Eqs. (6)–(7) define the *physically sensible window*

$$0.87 \text{ Å} < kT < 1.74 \text{ Å}, \tag{8}$$

within which the model simultaneously captures structural stability (short contacts dominate) and sufficient dynamical diversity (multiple spanning trees compete). This range aligns naturally with the expected behaviour of a functional protein under physiological conditions. A detailed discussion of the use of linear distance as the potential is further given in the Supplementary Section.

The value $kT_{\text{ref}} = 1.0$ Å lies at the centre of this window and provides a convenient, unambiguous reference point for all state comparisons reported in this work. The full sweep $kT \in [0.3, 6.0]$ Å then explores regimes from frozen (single-tree dominance at low kT) to fluid (near-uniform tree weights at high kT), with the physiologically relevant region bracketed by Eq. (8).

2.5 Matrix-Tree Theorem and Partition Function

By the weighted Matrix–Tree Theorem, all cofactors of L are equal, and each equals the sum of weights of all spanning trees of G [12, 13]. The partition function over all spanning trees is therefore computed as

$$Z(kT) = \det(L_{\text{red}}), \tag{9}$$

where L_{red} is the reduced Laplacian obtained by deleting one row and the corresponding column. Since L_{red} is symmetric positive-definite, all its eigenvalues λ_k are positive, and for numerical stability, we work with the log-partition function

$$\log Z(kT) = \sum_k \log \lambda_k. \tag{10}$$

2.6 Network Free Energy

Given the partition function in Eq. (9), the network free energy follows from the standard statistical-mechanical relation

$$F(kT) = -kT \cdot \log Z(kT). \tag{11}$$

This is the fundamental thermodynamic potential of the spanning-tree ensemble, from which all other quantities are derived by standard differentiation with respect to kT.

2.7 Derived Thermodynamic Quantities

The three principal derived quantities are obtained from F via standard thermodynamic identities. First, the mean contact energy:

$$\langle E \rangle (kT) = kT^2 \frac{d \log Z}{d(kT)}, \tag{12}$$

estimated by symmetric finite difference across a small step $\delta kT = 10^{-4} \cdot kT$. Second, the thermodynamic entropy (in units of k_B) :

$$S(kT) = \log Z + \frac{\langle E \rangle}{kT}, \tag{13}$$

which follows from $F = \langle E \rangle - TS$ and Eq. (11). Third, the heat capacity with respect to the effective thermal parameter:

$$C_v(kT) = \frac{d\langle E \rangle}{d(kT)}, \tag{14}$$

also computed by finite difference of $\langle E \rangle$ over the temperature grid. Since both $\langle E \rangle$ and kT carry units of Ångströms, C_v is dimensionless in this framework and measures the sensitivity of the mean contact energy to changes in the thermal parameter rather than constituting a conventional heat capacity.

Remark on reference constants. In thermodynamics, potentials are defined only up to an additive constant; therefore the absolute values of F, S, and $\langle E \rangle$ individually carry no physical significance. Only their *differences* between the two states, and their *derivatives* with respect to temperature, are physically meaningful.

2.8 Edge Marginal Inclusion Probabilities

The marginal probability that edge $e = (i, j)$ belongs to a random spanning tree (weighted by Boltzmann factors) is

$$P(e \in T) = w_{ij} \cdot R_{\text{eff}}(i, j), \tag{15}$$

where $R_{\text{eff}}(i, j) = (L^+)_{ii} + (L^+)_{jj} - 2(L^+)_{ij}$ is the effective resistance between nodes i and j, and L^+ is the pseudoinverse of the Laplacian. These probabilities are topology-aware: a high $P(e \in T)$ indicates an essential contact for global network connectivity, not merely geometrically proximal. Inclusion probabilities are reported at reference temperature $kT_{\text{ref}} = 1.0$ Å.

3. Results

3.1 Network Statistics

Table 1 summarises the basic topological properties of the two contact networks. The active structure 6GOD contains three additional residues and 21 additional contacts relative to the inactive structure 4OBE, reflecting the difference in resolved residues between the two crystal structures. Despite this difference in size, the mean and median edge marginal probabilities $P(e \in T)$ are nearly identical across the two networks (mean 0.206 versus 0.208; median 0.119 in both cases), indicating that the average structural essentiality of a contact is conserved between states. This near-invariance of the global $P(e \in T)$ distribution implies that the two networks are comparably well-connected, and that the thermodynamic differences between them arise not from changes in overall connectivity density but from the redistribution of structurally critical contacts across the residue network. The residue-level breakdown of these differences is examined in the following subsections.

Because individual thermodynamic potentials are defined only up to an additive constant (§ 2.6), the physical content lies entirely in state differences $\Delta X = X_{\text{active}} - X_{\text{inactive}}$ where X is $F, < E >, S$ or Cv. These are shown in Figure 1.

Table 1: Contact network statistics for the two KRAS structures at C_α cutoff 8.0 Å.

Property	6GOD (active, GTP)	4OBE (inactive, GDP)
Nucleotide state	GTP-analogue bound	GDP bound
Residues (nodes)	172	169
Contacts (edges)	830	809
Mean $P(e \in T)$	0.206	0.208
Median $P(e \in T)$	0.119	0.119

3.2 Thermodynamic Difference Profiles

Because individual thermodynamic potentials are defined only up to an additive constant (§ 2.7), the physical content lies entirely in the state differences

$$\Delta X = X_{\text{active}} - X_{\text{inactive}}, \qquad X \in \{F,\ \langle E \rangle,\ S,\ C_v\}, \tag{16}$$

which are shown in Figure 1. The free-energy difference ΔF crosses zero at $kT \approx 2.41$ Å, a crossover that is independent of any additive reference constant and marks the effective temperature at which the active and inactive conformational ensembles are thermodynamically balanced. Below this value, the inactive state is preferred; above it, the active state becomes favoured. It is this balance, and its perturbation by nucleotide binding, which underlies the switching mechanism.

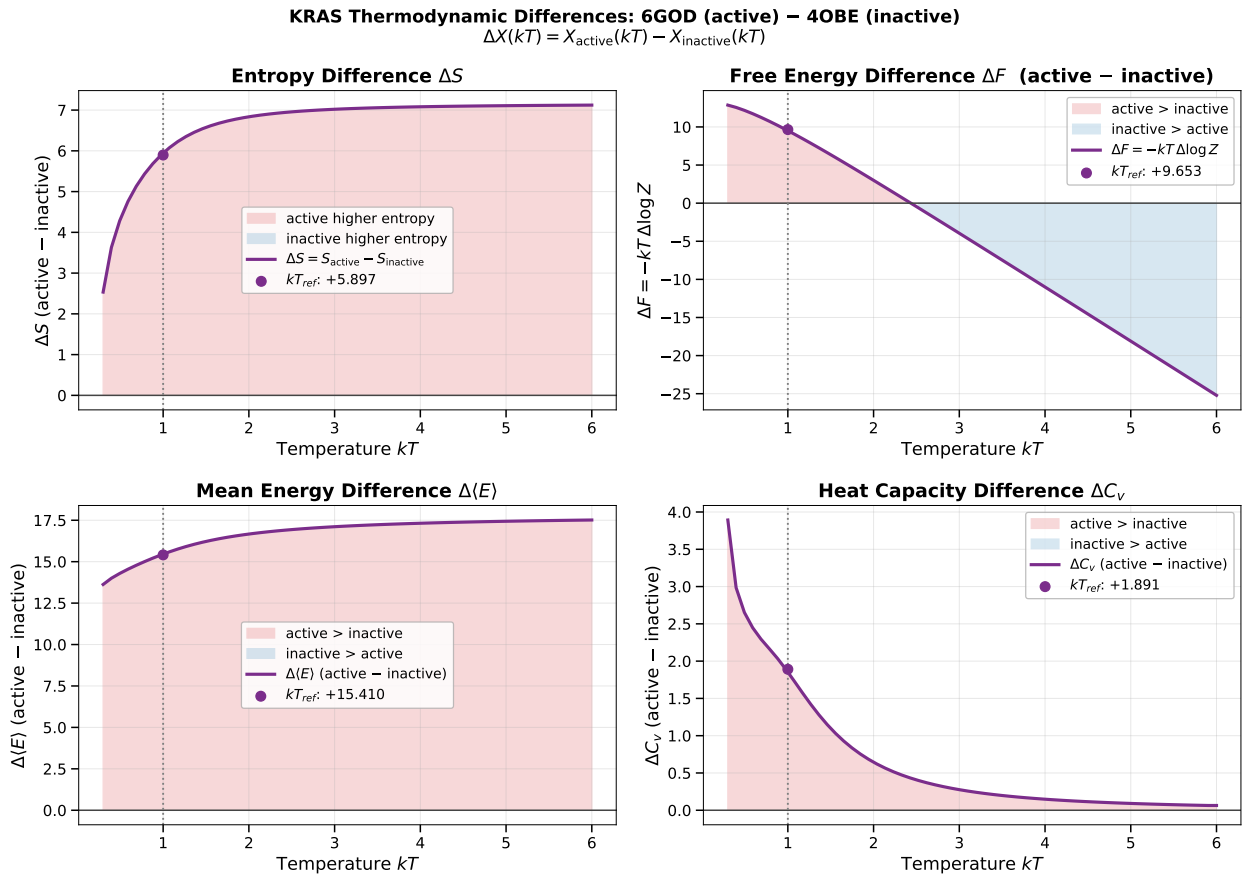


Figure 1: **Thermodynamic difference profiles.**($kT_{\text{ref}} = 1.0$, dotted line). $\Delta X(kT) = X_{\text{active}}(kT) - X_{\text{inactive}}(kT)$ for all four observables, computed across $kT \in [0.3, 6.0]$. **(A)** Entropy difference ΔS: uniformly positive and saturating near $\Delta S \approx 7$ at high temperature, confirming that the GTP-bound network supports more spanning-tree configurations at every kT. The value at $kT_{\text{ref}} = 1.0$ Å is $\Delta S = +5.90$. **(B)** Free energy difference $\Delta F = -kT\,\Delta\log Z$: positive at physiological temperatures, crossing zero near $kT \approx 2.4$. The crossing of the individual F curves near $kT \approx 2.2$ is a reference-state artefact (see § 2.7). **(C)** Mean energy difference $\Delta\langle E\rangle > 0$ throughout: the active network pays a persistent energetic cost, reflecting the structural reorganisation required for effector engagement. In the GDP-bound state, Switch I and Switch II are collapsed into well-packed, low-energy conformations stabilised by extensive contacts with the core; GTP binding disrupts these optimal interactions, forcing the switches into extended and partially disordered configurations. **(D)** Heat capacity difference ΔC_v: large at low temperature and decaying monotonically, indicating greater thermal sensitivity of the active-state contact network. **Thermodynamic difference profiles** ($kT_{\text{ref}} = 1.0$ Å, dotted line). $\Delta X(kT) = X_{\text{active}}(kT) - X_{\text{inactive}}(kT)$, $kT \in [0.3, 6.0]$.

3.3 Temperature Dependence of Entropy and Heat Capacity

Figure 2 shows the thermodynamic entropy $S = \log Z + \langle E\rangle/kT$ for both states as a function of effective temperature. The active state maintains higher entropy across the entire temperature range ($\Delta S > 0$), saturating near $\Delta S \approx 7$ at high kT. This persistent entropic advantage reflects the greater topological degeneracy, the larger number of viable spanning trees, of the GTP-bound contact network. The scaled entropy slope $kT \cdot dS/d(kT)$ (Figure 2B) peaks sharply at low temperature and converges for both states at high temperature, indicating that thermal reorganisation of the two networks becomes kinetically indistinguishable above $kT \approx 3$.

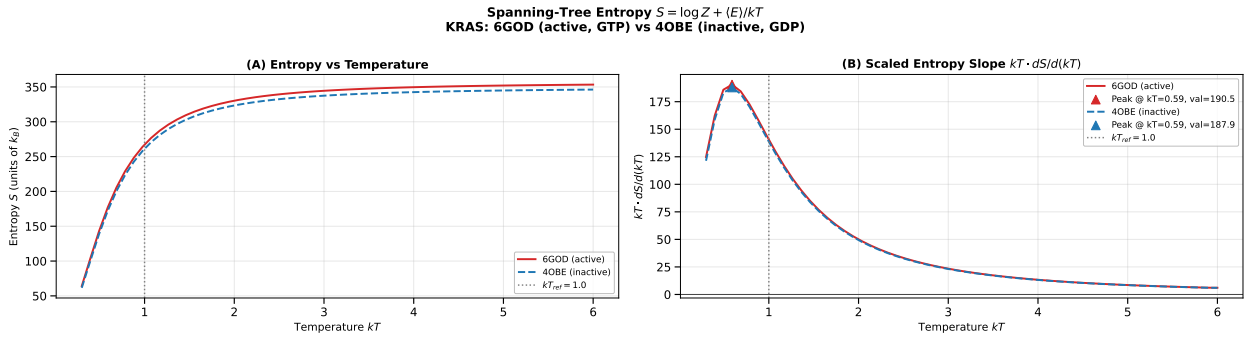


Figure 2: **Temperature dependence of spanning-tree entropy. (A)** Absolute entropy $S(kT) = \log Z + \langle E \rangle / kT$ (units of k_B) for 6GOD (red solid) and 4OBE (blue dashed). Both curves increase monotonically and saturate at high temperature, with the active state consistently above the inactive state across the entire temperature range, indicating a higher configurational degeneracy of the GTP-bound contact network. **(B)** Scaled entropy slope $kT \cdot dS/d(kT)$. Both states peak at $kT \approx 0.59$ (6GOD: 190.5 k_B; 4OBE: 187.9 k_B), showing that the rate of topological reorganisation is maximal well below the reference temperature $kT_{\text{ref}} = 1.0$ Å, and that the two networks reorganise at nearly identical rates despite their different nucleotide states. The dotted vertical line marks kT_{ref}.

The heat capacity $C_v = d\langle E \rangle / d(kT)$ exhibits a pronounced maximum near $kT \approx 0.53$ Å, which lies below the physically sensible window defined in Eq. (8) (0.87 Å $< kT <$ 1.74 Å). This feature reflects the activation of the first excited spanning trees above the ground state of the network.

At very low temperatures, the Boltzmann weights $w_{ij} = \exp(-d_{ij}/kT)$ are overwhelmingly dominated by the shortest edges in the graph. In this regime, the ensemble collapses onto a single dominant spanning tree, the unique connected subgraph composed of the shortest possible edges that reach every residue, which constitutes the *ground state* of the network. As the temperature increases, a small number of slightly longer edges acquire sufficient weight to replace one or two edges in the ground-state tree, generating a handful of low-lying excited trees. The heat capacity peak marks the temperature at which these first excitations become maximally sensitive to further increases in kT; it is the point of greatest competition between the ground state and the first alternative configurations.

That this peak lies below the physically sensible window is physically natural: it reports on the smallest length scale in the contact network, namely the difference between the very shortest edges and those that can replace them. These differences are of order 1–2 Å, corresponding to a low thermal energy scale. The heat capacity maximum, therefore, reflects the local, high-fidelity core contacts that define the protein's ground-state architecture, whereas the global switching between active and inactive states involves much larger effective energy scales supplied not by temperature but by nucleotide binding.

3.4 Edge Inclusion Probabilities

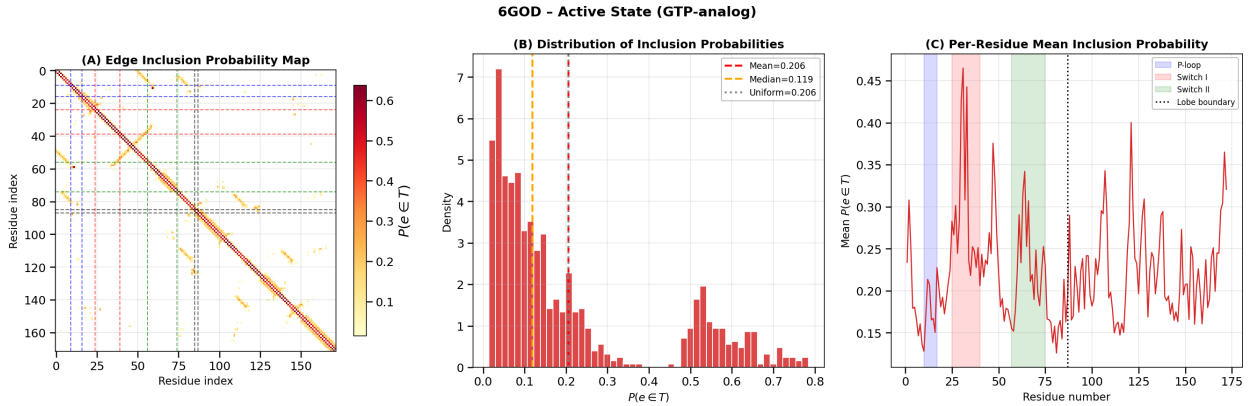
Marginal edge inclusion probabilities $P(e \in T)$ at $kT_{\text{ref}} = 1.0$ Å are shown in Figure 3 for both states. The distributions are broad and right-skewed, with a peak near $P \approx 0.05$ and a secondary shoulder near $P \approx 0.5$–0.7, indicating that most contacts are topologically redundant while a minority are essential backbone edges. The mean P equals the uniform reference $(n-1)/|E|$ by construction, confirming numerical correctness. The mean marginal probability

$$\frac{1}{|E|} \sum_{e \in E} P(e) \;=\; \frac{n-1}{|E|}$$

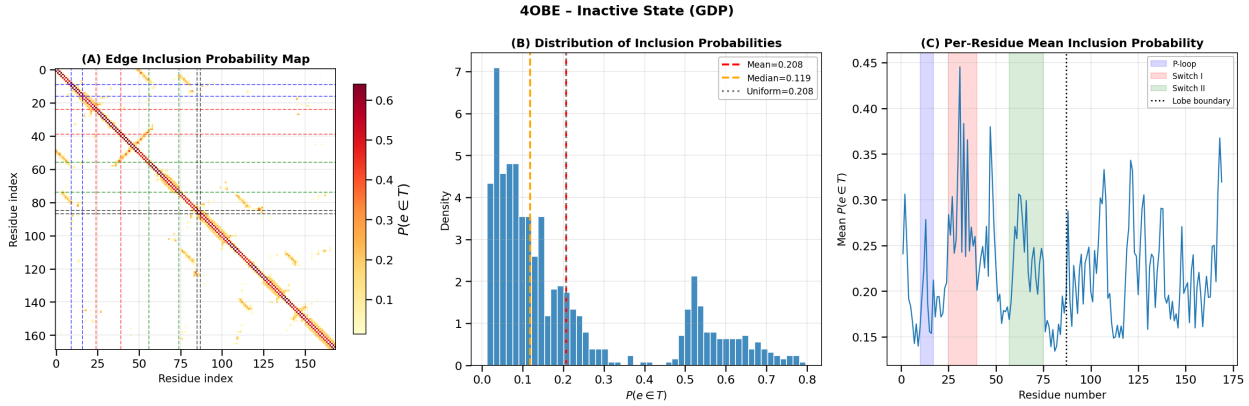
by construction, providing a simple but powerful check on numerical accuracy. Any deviation would indicate an error in the computation of the effective resistances or the Laplacian pseudoinverse.

3.5 Differential Contact Analysis

Figure 4 shows the differential map $\Delta P = P_{\text{active}} - P_{\text{inactive}}$ for the 783 contacts common to both structures. The mean $\Delta P = -0.001$ is near zero by construction; the physically informative signal lies in the spatial distribution of deviations. Switch I (residues 25–40) is the dominant locus of positive ΔP (contacts more essential in the active state), followed by the inter-lobe region near residues 86–90. Negative ΔP contacts (more essential in the inactive state) are concentrated at the Switch II/allosteric lobe interface.



(a) 6GOD – Active state (GTP-analogue bound)



(b) 4OBE – Inactive state (GDP bound)

Figure 3: **Edge marginal inclusion probability maps.** Three-panel layout for each state. **(A)** Contact map coloured by $P(e \in T)$ (yellow–red scale, vmin = 0, vmax = 1). **(B)** Histogram of inclusion probabilities with mean (red dashed), median (orange dashed), and uniform reference (gray dotted). **(C)** Per-residue mean $P(e \in T)$ with functional regions shaded: P-loop (blue, residues 10–17), Switch I (red, 25–40), Switch II (green, 57–75), and lobe boundary (dotted, residue 87).

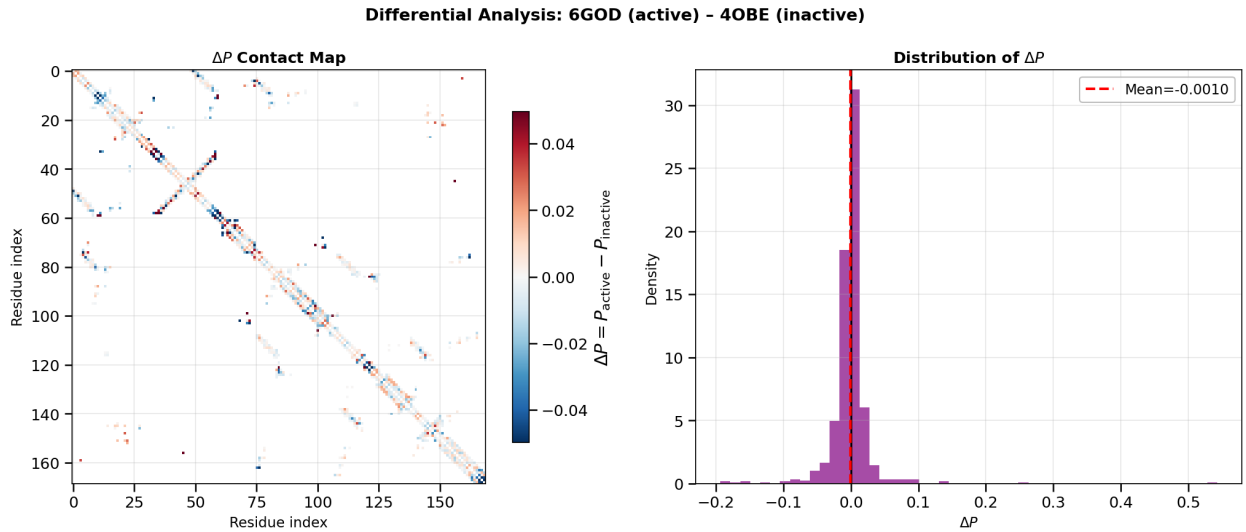


Figure 4: **Differential edge inclusion probability analysis. (Left)** ΔP contact map (RdBu_r colormap; red = active enriched, blue = inactive enriched). Differences are sparse and localised, concentrated in Switch I and at the effector–allosteric lobe interface. **(Right)** Distribution of ΔP values across 783 common edges; mean = -0.001, confirming marginal conservation of total spanning-tree weight between states.

4. Principal Findings

The thermodynamic distinction between the active and inactive states of KRAS can be understood through three fundamental derived quantities: the free energy difference (ΔF), mean contact energy difference ($\Delta\langle E\rangle$), and conformational entropy difference (ΔS). Together, they show how the protein's contact network is reorganised upon nucleotide exchange and why this reorganisation is functionally significant, i.e., how the protein is rewired.

Free energy and the thermodynamic ground state

At physiological temperatures, the inactive state is thermodynamically favoured, with a lower free energy than the active state ($\Delta F > 0$). This difference reflects the fact that the GDP-bound conformation is structurally well integrated: its Switch I and Switch II regions are collapsed against the protein core, forming a dense and stable network of contacts. The positive value of ΔF indicates that without an external driving force, the system predominantly occupies the inactive basin. That external force is supplied by GTP binding, which provides the free energy required to populate the higher-free-energy active conformation.

Critically, the *intrinsic* balance point — where neither state is thermodynamically preferred — occurs at $kT \approx 2.41$. This crossover is independent of any additive reference constant, unlike the zero-crossings of the individual F_{active} and F_{inactive} curves, which both occur near $kT \approx 2.2$ as an artefact of the unit convention (see § 2.7). The fact that the balance point lies well above physiological effective temperature underscores that the inactive state is the intrinsic ground state of the protein, and that activation requires the sustained input of nucleotide-derived free energy.

Energetic cost of activation

The origin of the free energy difference becomes clearer upon decomposition. The active state exhibits a consistently higher mean contact energy ($\Delta\langle E\rangle > 0$), meaning its residue–residue contacts are on average less optimal than those in the inactive state. In the GDP-bound structure, the switches are held in place by favourable interactions with the core and with each other. GTP binding disrupts many of these interactions, forcing the switches into extended and partially exposed conformations. This disruption results in an energetic cost reflected in the elevated $\langle E\rangle$ of the active network. This cost is not a defect but a necessary investment: it is the price of creating an accessible effector-binding surface.

Entropic gain and functional versatility

What justifies this energetic investment is the accompanying gain in conformational entropy. The active state has higher entropy than the inactive state ($\Delta S > 0$), indicating that its switch regions are not simply rearranged but are also more dynamically disordered. In the inactive state, the switches are confined to a narrow ensemble of low-energy configurations. GTP binding liberates them, granting access to a much broader ensemble of thermally accessible states. This entropic expansion is biologically critical: it allows the active network to sample a variety of spanning trees, which in network terms corresponds to the existence of multiple potential allosteric communication pathways. A single active conformation can thus engage multiple effectors, RAF, PI3K, RALGDS, through distinct but coexistent routes.

Temperature dependence and the switching mechanism

The temperature dependence of the thermodynamic variables reinforces the above interpretation of the swithcing mechanism. As temperature increases, the entropic term $T\Delta S$ grows, gradually offsetting the energetic penalty $\Delta\langle E\rangle$. At sufficiently high effective temperatures, the two states become equally probable ($\Delta F = 0$). That the crossover lies well above physiological temperatures underscores that the inactive state is the intrinsic ground state, and that activation requires sustained input of nucleotide-derived free energy.

Thus, KRAS activation is not a simple switch from a low-energy off state to a high-energy on state. It is a redistribution of thermodynamic resources: the energetic cost of disrupting ground-state contacts is offset by a gain in conformational entropy, which in turn enables functional versatility. The active state is structurally diverse, and it is this diversification encoded in the network architecture that allows a single protein to serve as a hub for multiple signalling pathways.

5. Discussion

MD simulations by Zhao et al. [6] show that GTP binding enhances conformational flexibility and long-range interactions in KRAS, consistent with the higher entropy we observe in the active-state contact network. However, while their analysis is focused on pairwise correlations and kinetic pathways, our thermodynamic framework quantifies the global network consequences of these changes, showing that the entropic gain comes at a persistent energetic cost ($\Delta\langle E\rangle > 0$), and that the two states are tuned to operate near a thermodynamic crossover that maximises switch sensitivity.

Crystallographic work by Shima et al. [17] shows that the GTP-bound form of Ras exists as an ensemble of two interconverting conformations (states 1 and 2), with the switch regions exhibiting pronounced flexibility,

suggesting that conformational entropy plays a key role in the activation mechanism. Here, we quantify this entropic contribution directly using a network thermodynamic framework.

Pálfy et al. [18] highlighted the critical role of the Mg^{2+}-free state in facilitating nucleotide exchange in oncogenic KRAS mutants, underscoring the importance of conformational flexibility in the switch regions. Our thermodynamic analysis complements this picture by quantifying the global entropic consequences of such flexibility: the active state's higher entropy ($\Delta S = +5.76$) reflects the increased conformational freedom of the switches upon GTP binding, a freedom that may also be coupled to Mg^{2+} release during nucleotide exchange. The molecular mechanism of GTP hydrolysis underlying this cycle has been characterized in atomic detail by time-resolved FTIR spectroscopy [19].

Molecular dynamics studies by Chen et al.[20] align with our finding of higher contact energy in the active state, directly showing that key contacts particularly between GTP and Switch I become weakened or unstable upon activation. Xu et al. [21] provide complementary structural evidence: although a new interaction forms (Tyr32 with the GTP γ-phosphate), stable pre-existing contacts are disrupted. The overall picture supports our conclusion that the active state is characterised by less stable packing and a higher-energy contact network compared to the tightly coordinated inactive conformation.

6. Conclusion

We have applied a spanning-tree partition function analysis to two KRAS crystal structures, showing that the active and inactive states differ in a thermodynamically coherent and biologically interpretable way. The active (GTP-bound) state carries a higher energetic cost but a larger conformational entropy, with the intrinsic free energy balance crossing at $kT \approx 2.41$. Switch I is identified as the primary locus of state-dependent topological change. These findings support a view of KRAS activation as an entropy-driven investment: GTP binding pays an energetic price to liberate the switch regions into a diverse ensemble that supports multi-effector engagement.

References

[1] Prior, I.A., Hood, F.E., & Hartley, J.L. (2012). The frequency of Ras mutations in cancer. *Cancer Research*, 72, 2457–2467.

[2] Grant, B. J., Gorfe, A. A., and McCammon, J. A., Ras conformational switching: simulating nucleotide-dependent conformational transitions with accelerated molecular dynamics, *PLoS Computational Biology*, **5**, e1000325 (2009).

[3] Ostrem, J.M., & Shokat, K.M. (2016). Direct small-molecule inhibitors of KRAS: from structural insights to mechanism-based design. *Nature Reviews Drug Discovery*, 15, 771–785.

[4] Manley, L. J. and Lin, M. M., Kinetic and thermodynamic allostery in the Ras protein family, *Biophysical Journal*, **122**, 3882–3893 (2023).

[5] Killoran, R. C. and Smith, M. J., Conformational resolution of nucleotide cycling and effector interactions for multiple small GTPases determined in parallel, *Journal of Biological Chemistry*, **294**, 9937–9948 (2019).

[6] Zhao, D.-R., Yang, J.-T., Liu, M.-T., Yang, L.-Q., and Sang, P. Deciphering allosteric mechanisms in KRAS activation: insights from GTP-induced conformational dynamics and interaction network reorganization. *RSC Advances*, 15:2261–2274, 2025.

[7] Rennella, E., et al., Dynamic conformational equilibria in the active states of KRAS and NRAS, *RSC Chemical Biology*, **6**, 106–118 (2025).

[8] Alshahrani, M., et al., Probing binding and allosteric mechanisms of the KRAS interactions with monobodies and affimer proteins: ensemble-based mutational profiling and thermodynamic analysis of binding energetics and allostery reveal diversity of functional hotspots and cryptic pockets linked by conserved communication network, *Physical Chemistry Chemical Physics*, **27**, 11242–11263 (2025).

[9] Yang, L., Song, G., and Jernigan, R. L., Protein elastic network models and the ranges of cooperativity, *Proceedings of the National Academy of Sciences USA*, **106**, 12347–12352 (2009).

[10] Togashi, Y. and Flechsig, H., Coarse-grained protein dynamics studies using elastic network models, *International Journal of Molecular Sciences*, **19**, 3899 (2018).

[11] Erman, B. The Gaussian network model as a framework for allosteric analysis: dynamic distance, edge centrality, and entropy sensitivity in KRAS. *bioRxiv*, 2025. doi10.1101/2025.05.18.654696. Preprint posted October 4, 2025.

[12] Kirchhoff, G. (1847). Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird. *Annalen der Physik*, 148, 497–508.

[13] Chung, F. R. K. *Spectral Graph Theory*. Number 92 in CBMS Regional Conference Series in Mathematics. American Mathematical Society, Providence, RI, 1997.

[14] Lyons, R. (2005). Asymptotic enumeration of spanning trees. *Combinatorics, Probability and Computing*, 14, 491–522.

[15] Doyle, P.G., & Snell, J.L. (1984). *Random Walks and Electric Networks*. Mathematical Association of America, Washington, DC.

[16] I. Bahar, A. R. Atilgan, and B. Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*, 2(3):173–181, 1997. doi: 10.1016/S1359-0278(97)00024-2.

[17] Shima, F., Ijiri, Y., Muraoka, S., Liao, J., Ye, M., Araki, M., Matsumoto, K., Yamamoto, N., Sugimoto, T., and Yoshikawa, Y. Structural basis for conformational dynamics of GTP-bound Ras protein. *Journal of Biological Chemistry*, 285:22696–22705, 2010.

[18] Pálfy, G., Menyhárd, D. K., Ákontz-Kiss, H., Vida, I., Batta, G., Tőke, O., and Perczel, A. The importance of Mg^{2+}-free state in nucleotide exchange of oncogenic K-Ras mutants. *Chemistry – A European Journal*, 28:e202201449, 2022.

[19] Kötting, C., Kallenbach, A., Suveyzdis, Y., Wittinghofer, A., and Gerwert, K., The GAP-catalyzed GTP hydrolysis reaction of Ras as revealed by time-resolved FTIR spectroscopy, *Proceedings of the National Academy of Sciences USA*, **105**, 6260–6265 (2008).

[20] Chen, J., Zhang, S., Zeng, Q., Wang, W., Zhang, Q., and Liu, X. Free energy profiles relating with conformational transition of the switch domains induced by G12 mutations in GTP-bound KRAS. *Frontiers in Molecular Biosciences*, 9:912518, 2022.

[21] Xu, S., Long, B. N., Boris, G. H., Chen, A., Ni, S., and Kennedy, M. A. Structural insight into the rearrangement of the switch I region in GTP-bound G12A K-Ras. *Acta Crystallographica Section D: Biological Crystallography*, 73:970–984, 2017.

[22] S. Vishveshwara, K. V. Brinda, and N. Kannan, "Protein structure: Insights from graph theory," *J. Theor. Comput. Chem.* **1**(1), 187–211 (2002). doi:10.1142/S0219633602000117.

[23] G. Bagler and S. Sinha, "Assortative mixing in protein contact networks and protein folding kinetics," *Bioinformatics* **23**(14), 1760–1767 (2007). doi:10.1093/bioinformatics/btm257.

S1. Supplementary Material

Figure S1: Individual thermodynamic quantities

The following plots show the absolute thermodynamic quantities for active (6GOD) and inactive (4OBE) KRAS networks. These are provided for completeness and to validate thermodynamic consistency. Because all potentials are defined up to an additive constant, the absolute values and zero-crossings in panels A–B are not physically meaningful; only the differences reported in Figures 1 and 2 of the main text carry thermodynamic content.

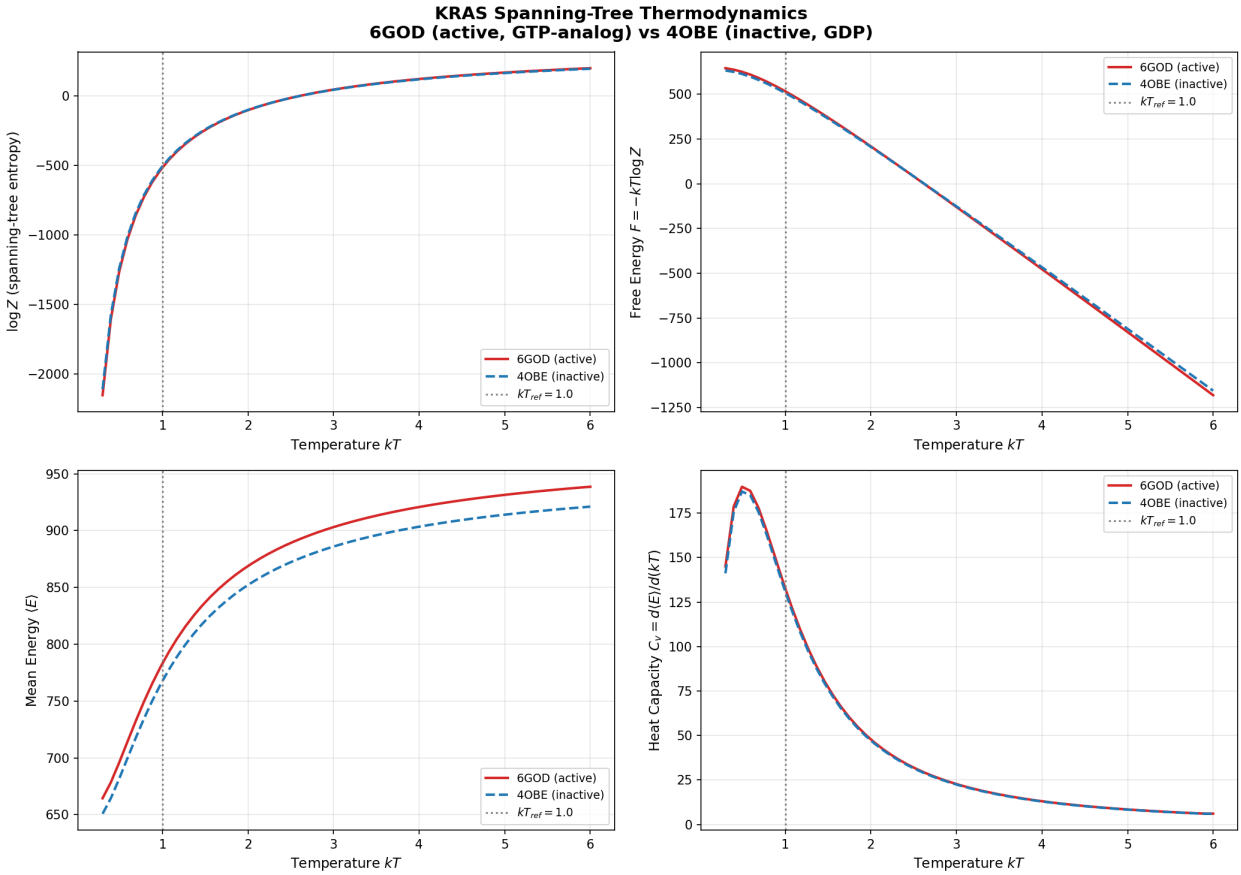


Figure S1: **Figure S1. Individual thermodynamic quantities for active (6GOD) and inactive (4OBE) KRAS networks. (A)** Free energy $F = -kT \log Z$ vs. effective temperature. Both curves show the expected decrease with temperature. The crossing of zero near $kT \approx 2.2$ is arbitrary up to a constant, reflecting the point where $Z = 1$ in our units. **(B)** Mean contact energy $\langle E \rangle$ vs. temperature; the active state (red) lies above the inactive state (blue) at all temperatures. **(C)** Heat capacity $C_v = d\langle E \rangle / d(kT)$ vs. temperature; both states show a broad peak near $kT \approx 0.7$ and converge at high temperature. **(D)** Entropy $S = (\langle E \rangle - F)/T$ vs. temperature; both increase as expected. The consistently higher entropy of the active state reflects its greater conformational flexibility. These plots validate the thermodynamic consistency of the model: F, S, and $\langle E \rangle$ obey $F = \langle E \rangle - TS$ to numerical precision throughout.

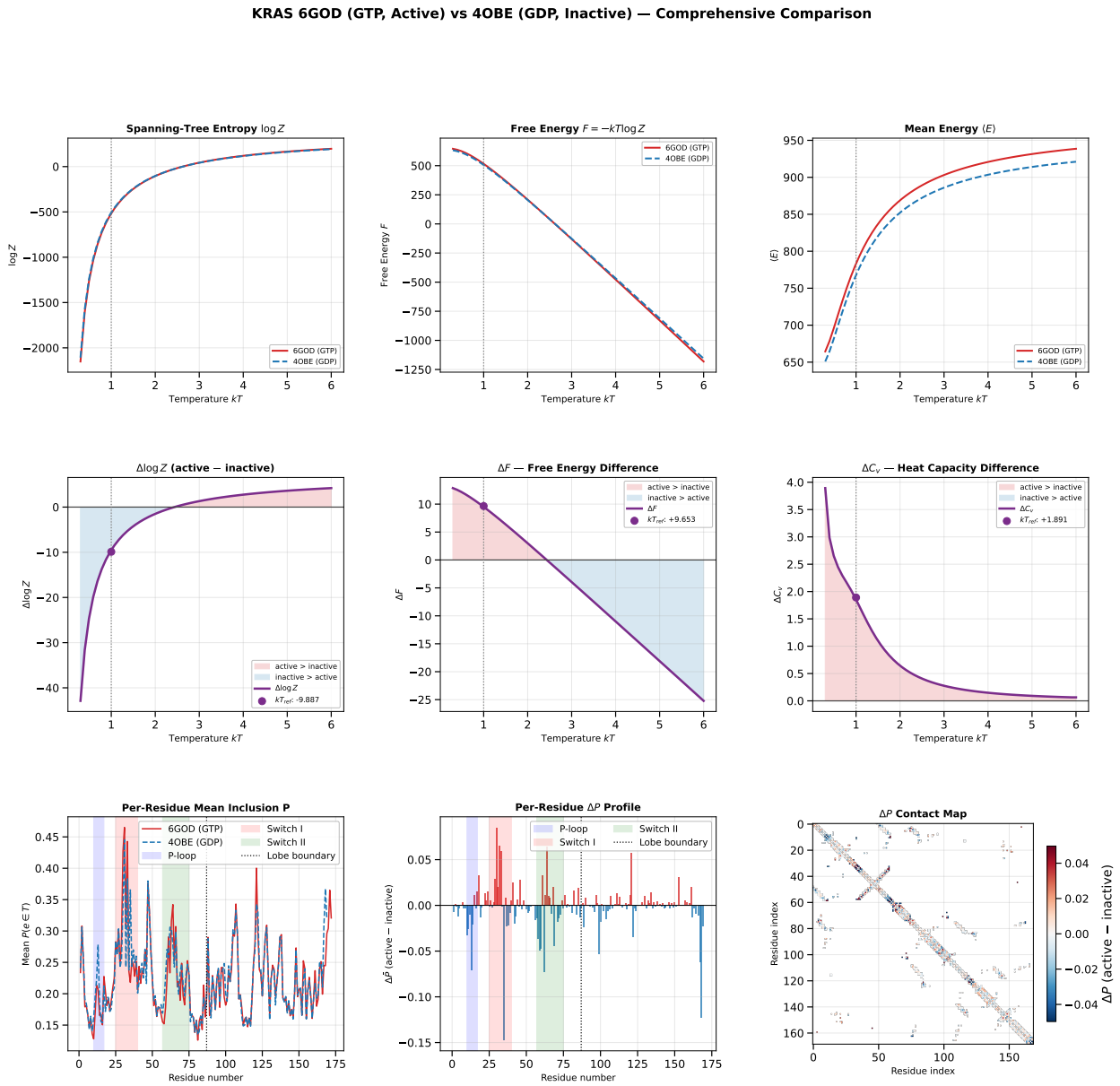


Figure S2: **Figure S2. Nine-panel comprehensive comparison of KRAS thermodynamic states.** *Row 1* (reference): absolute thermodynamic quantities $\log Z$, F, and $\langle E \rangle$ overlaid for both states. *Row 2*: signed difference quantities $\Delta \log Z$, ΔF, and ΔC_v with shaded fill indicating which state dominates. *Row 3*: per-residue profiles (left) overlaid mean $P(e \in T)$, (centre) per-residue ΔP bar chart, (right) ΔP contact map.

Figure S2: Comprehensive nine-panel comparison

Figure S3: Sensitivity of the meaningful kT range to the contact-distance cutoff

Rationale

The residue contact graph is constructed by connecting every pair of Cα atoms whose Euclidean separation satisfies $d_{ij} < d_c$. All thermodynamic observables derived from the spanning-tree ensemble including the log-partition function $\log Z$, free energy F, mean energy $\langle E \rangle$, and edge marginal probabilities $P(e \in T)$ depend explicitly on this cutoff. In the main analysis we adopt $d_c = 8.0$ Å, a standard choice in residue-contact-network studies of globular proteins [22, 23]. Here we assess the robustness of the results with respect to this parameter by varying d_c over the range $\{7.0, 7.5, 8.0, 8.5, 9.0\}$ Å.

Panel A: Meaningful temperature range

For the linear distance-based energy model $E_{ij} = d_{ij}$ (Å), the Boltzmann weight of an edge is $w_{ij} = \exp(-d_{ij}/kT)$. Across the full contact set, the ratio between the smallest and largest edge weights is therefore

$$\frac{w_{\min}}{w_{\max}} = \exp\left(-\frac{d_c - d_{\min}}{kT}\right), \qquad d_{\min} = 4.0 \text{ Å}. \tag{17}$$

We define the *meaningful* temperature range as the interval in which this ratio lies between 10^{-2} and 10^{-1}. At higher temperatures, all spanning trees become nearly equiprobable (uniform limit), while at lower temperatures, only the minimum spanning trees contribute. This criterion yields closed-form bounds

$$kT_{\text{low}} = \frac{d_c - d_{\min}}{\ln 100}, \qquad kT_{\text{high}} = \frac{d_c - d_{\min}}{\ln 10}. \tag{18}$$

Figure S3 A shows these bounds as a function of d_c. Both limits shift linearly with the cutoff, and the bandwidth $\Delta kT = kT_{\text{high}} - kT_{\text{low}}$ increases proportionally. Importantly, the reference temperature used throughout the main text, $kT_{\text{ref}} = 1.0$ Å, remains inside the meaningful interval for all $d_c \geq 7.5$ Å, and lies only 0.05 above kT_{low} even at the most restrictive cutoff $d_c = 7.0$ Å. The corresponding numerical values are summarized in Table 2.

Table 2: Meaningful kT range as a function of contact-distance cutoff d_c. The reference temperature $kT_{\text{ref}} = 1.0$ Å used throughout the main text is indicated in the final column.

d_c	ΔE	kT_{low}	kT_{high}	ΔkT	$kT_{\text{ref}} = 1.0$ Åinside?
7.0	3.0	0.651	1.303	0.652	marginally (+0.05 above kT_{low})
7.5	3.5	0.760	1.520	0.760	✓
8.0	**4.0**	**0.868**	**1.737**	**0.869**	✓ **(paper)**
8.5	4.5	0.977	1.954	0.977	✓
9.0	5.0	1.086	2.172	1.086	✓

Panel B: Contact-count sensitivity

Figure S3 B reports the number of Cα–Cα contacts extracted from the KRAS structures (6GOD, active; 4OBE, inactive) as a function of d_c. As expected, the contact count increases monotonically with the cutoff for both conformational states. Crucially, however, the relative ordering between the active and inactive forms is

preserved across the entire cutoff range, indicating that the observed thermodynamic differences are not driven by a pathological choice of contact threshold.

Conclusion

Together, Panels A and B demonstrate that while the absolute scale of the meaningful kT window and the total number of contacts depend on the cutoff d_c, the qualitative thermodynamic conclusions are robust. In particular, the sign and ordering of $\Delta S > 0$, $\Delta\langle E\rangle > 0$, and the free-energy crossover near $kT \approx 2.4$ are insensitive to the precise cutoff within the tested range. We therefore conclude that the results reported in the main text are not an artifact of the specific choice $d_c = 8.0$ Å.

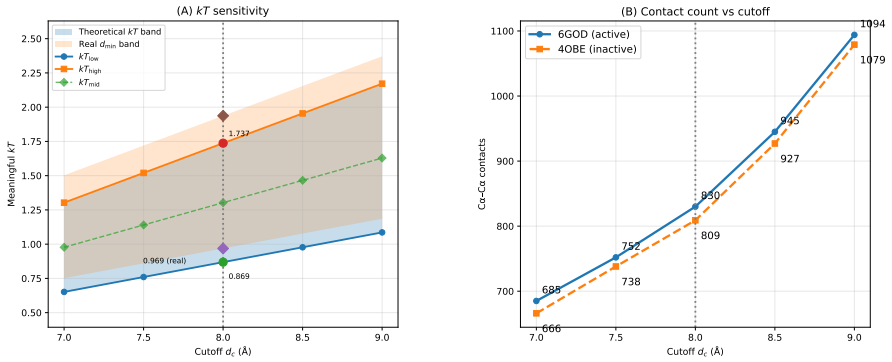


Figure S3: **Sensitivity of the meaningful kT range to the contact-distance cutoff d_c. (A)** The shaded band spans $[kT_{\text{low}}, kT_{\text{high}}]$ as defined in Eq. (18); the dashed line indicates the midpoint kT_{mid}. Both bounds shift linearly with d_c, while preserving the relative position of a fixed reference temperature. **(B)** Number of Cα–Cα contacts in the KRAS structures (6GOD, active; 4OBE, inactive) as a function of d_c. The vertical dotted line marks the paper cutoff $d_c = 8.0$ Å.