Electronic Research Archive

http://www.aimspress.com/journal/era

ERA, 34(12): 9485–9518.
DOI: 10.3934/era.2026398
Received: 17 July 2026
Revised: 18 October 2026
Accepted: 03 November 2026
Published: 19 November 2026

Research article

From mutation operator geometry to Ising Hamiltonians: A variational framework for representative mutation selection

Álvaro Michelena[1], Roberto Casado[2,*], Esteban Jove[1], J. M. Corchado[2,3] and José-Luis Calvo-Rolle[1]

[1] Department of Industrial Engineering, University of A Coruña, Ciencia y Técnica Cibernética (CTC), Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC), Ferrol, A Coruña 15071, Spain

[2] Departamento de Informática y Automática, Universidad de Salamanca, Patio de Escuelas 1, Salamanca 37008, Spain

[3] Osaka Institute of Technology, 5 Chome-16-1 Omiya, Asahi-ku, Osaka 535-8585, Japan

* **Correspondence:** Email: rober@usal.es.

Abstract: Selecting representative protein mutations is an important problem in computational structural biology, where the objective is to identify a small subset of variants that captures relevant variation within a larger mutation set. Existing approaches frequently rely on individual mutation scores, predefined similarity measures, or application-specific optimization objectives, leaving open the question of whether representative selection can instead be constructed directly from a mathematical representation of modeled mutation-induced perturbations. In this work, we introduced a geometry-induced variational framework based on an operator representation of modeled mutation-induced perturbations. Each mutation was represented as a perturbation of the graph Laplacian of a common wild-type residue interaction network, embedding all candidates into a symmetric operator space endowed with the Hilbert–Schmidt structure. For a prescribed subset cardinality, the resulting perturbation magnitudes and pairwise operator distances defined a variational objective that balances deviation from the wild-type reference with geometric diversity. This formulation admitted exact binary quadratic, quadratic unconstrained binary optimization (QUBO), and Ising representations, whose data-dependent coefficients were inherited directly from the operator-derived quantities once the underlying modeling choices had been fixed. The framework was evaluated on fourteen KRAS protein missense mutations. The resulting operator configuration exhibited a structured geometry, and the variational criterion produced representative subsets with improved coverage of the induced mutation space relative to perturbation-magnitude ranking alone. Robustness was examined across different subset cardinalities, diversity weights, connectivity normalizations, and alternative residue-interaction-network constructions. The exact variational, QUBO, and Ising formulations were numerically verified to recover consistent solutions, while fixed-cardinality

simulated annealing and a cardinality-preserving quantum approximate optimization algorithm with an XY mixer (XY-QAOA) were evaluated as alternative optimization backends for the same induced objective.

Keywords: mutation operator space; representative mutation selection; graph Laplacian; variational optimization; Ising Hamiltonian; quantum optimization

1. Introduction

Mutations constitute one of the primary sources of structural and functional diversity in proteins, playing a fundamental role in evolution, human disease, protein engineering, and therapeutic response [1, 2]. The widespread adoption of high-throughput sequencing technologies has led to the identification of an enormous number of naturally occurring and clinically relevant variants, creating an increasing demand for computational methods capable of prioritizing mutations for detailed structural and functional characterization [3]. Because experimental validation remains expensive and time-consuming, selecting a small yet informative subset of representative mutations has become an essential task in computational structural biology [4, 5]. Traditionally, mutation prioritization has been approached through the analysis of individual mutation properties. Existing methods commonly rank variants according to evolutionary conservation [6], predicted pathogenicity [7], energetic stability [8], structural disruption [9], or network centrality [10]. These approaches are well suited to identifying individual mutations with potentially large biological effects. However, a ranking based exclusively on individual scores does not explicitly account for how candidate mutations relate to one another. Consequently, a subset formed by the highest-ranked variants may contain several mutations producing similar modeled perturbations while leaving other regions of the mutation set unrepresented. From the perspective of representative selection, the objective is therefore not only to retain mutations associated with large perturbations, but also to obtain a subset that covers the diversity present in the complete set of modeled mutation-induced changes. This suggests viewing representative selection as a problem defined on the organization of the mutation representation itself rather than solely as a ranking of its individual elements.

Several computational paradigms have attempted to address this challenge from different perspectives [11]. Clustering-based approaches organize mutations according to their spatial or functional relationships [12], graph-based methods exploit residue interaction networks to identify relevant regions of protein structure [10, 13], and combinatorial optimization approaches have been used to construct non-redundant representative subsets of protein sequences [14]. More recently, advances in quantum optimization have increased interest in QUBO and Ising formulations of combinatorial optimization problems [15, 16], enabling the application of quantum annealing [17, 18] and variational quantum optimization algorithms [19]. Despite their different computational implementations, these approaches typically require an objective function to be specified before the optimization stage. The coefficients of that objective are commonly determined from heuristic criteria, domain knowledge, predefined similarity measures, or application-specific scores. This separation between representation and optimization raises a more fundamental question. Rather than first designing an optimization objective and then selecting a solver, can a representative-selection problem be constructed directly from a mathematical representation of the modeled mutation-induced perturbations? More specifically, if mutations are represented as elements of a common operator space, can quantities derived from the geometry of that

representation provide both the individual and pairwise terms required to define the selection problem? Addressing this question shifts the emphasis from the optimization algorithm to the mathematical representation from which the optimization model is constructed. To investigate this idea, we introduce an operator-based representation of modeled mutation-induced perturbations. A residue interaction network of the wild-type protein is first represented through its weighted graph Laplacian. Each mutation is then associated with a mutation-dependent local reweighting of the same reference network, and the corresponding perturbation operator is defined as the difference between the mutation-dependent and wild-type Laplacians. This construction places all candidate mutations within a common space of symmetric operators. Importantly, these operators represent controlled numerical perturbations of the wild-type weighted network; they are not intended to reconstruct mutant-resolved three-dimensional protein structures.

The Hilbert–Schmidt structure of this operator space [20] provides two quantities that are used throughout the framework. The squared norm of each perturbation operator measures its magnitude relative to the wild-type reference, while the squared Hilbert–Schmidt distance between two perturbation operators measures their geometric separation. For a prescribed subset cardinality, these quantities define a variational objective that balances total perturbation magnitude against pairwise geometric diversity. Representative selection is therefore formulated as the search for a fixed-size subset containing sufficiently large perturbations while avoiding geometrically redundant choices. Once the perturbation operators have been constructed, the selection model does not require additional mutation-specific scores or externally defined pairwise similarity descriptors. The resulting variational problem possesses an exact pairwise quadratic structure. Introducing binary variables for mutation selection yields an exact constrained binary formulation, and the prescribed cardinality can be incorporated through a standard quadratic penalty to obtain a QUBO model. The subsequent QUBO-to-Ising transformation is likewise exact [15]. These reformulations do not eliminate the modeling choices involved in defining the Residue Interaction Network (RIN), the perturbation model, the Hilbert–Schmidt geometry, the subset cardinality, or the relative weight assigned to geometric diversity. Rather, their significance is that, once these modeling choices have been fixed, the QUBO and Ising representations introduce no additional mutation-specific scoring rules. In particular, the two-body Ising couplings are affine functions of the squared Hilbert–Schmidt distances between mutation perturbation operators, while the local fields also depend on the individual perturbation magnitudes. Thus, the data-dependent coefficients of the optimization model are inherited directly from the operator-derived quantities. This construction also separates the mathematical model from the algorithm used to search it. Exact enumeration, classical heuristics, quantum annealing, or gate-based variational algorithms can in principle operate on the same induced objective without changing the mutation representation or the representative-selection criterion [15, 18, 19]. In the present work, exact enumeration is used to establish reference solutions, fixed-cardinality simulated annealing provides a classical heuristic backend [21], and a cardinality-preserving XY-QAOA ansatz is used to study compatibility with gate-based variational quantum optimization [19, 22]. The optimization backends are therefore evaluated as alternative procedures for searching the same model rather than as competing definitions of representative mutation selection. The framework is evaluated on a panel of KRAS missense mutations. The numerical analysis first characterizes the geometry generated by the mutation perturbation operators and examines its sensitivity to alternative residue-interaction-network constructions. Representative subsets are then compared with a perturbation-magnitude baseline, and robustness is studied across different prescribed cardinalities

and a broad range of the diversity parameter. The analysis further examines whether the selected subsets are dominated by local network connectivity and evaluates how changes in the RIN construction affect both the global operator geometry and the discrete representative set. Finally, the exact variational, QUBO, and Ising formulations are compared numerically as consistency checks of the analytical reformulations, and classical simulated annealing and fixed-cardinality XY-QAOA are evaluated as optimization backends. The reconstruction of pairwise operator distances from the Ising couplings is therefore interpreted as a numerical verification of the analytical encoding, rather than as independent empirical evidence that the Hamiltonian discovers or preserves the geometry.

The principal contributions of this work can therefore be summarized as follows. First, we introduce an operator representation in which modeled mutation-induced perturbations are expressed relative to a common wild-type Laplacian and compared through the Hilbert–Schmidt structure. Second, we construct a fixed-cardinality representative-selection criterion directly from the resulting perturbation magnitudes and pairwise operator distances, combining individual deviation from the reference with geometric diversity. Third, we derive exact QUBO and Ising reformulations of this variational problem and make explicit the dependence of the Ising coefficients on the operator-derived quantities, thereby separating the geometry-dependent contribution from the cardinality-enforcement term. Finally, we evaluate the framework on KRAS through geometric characterization, representative-subset, and parameter robustness analyses, alternative RIN constructions, exact reformulation consistency checks, and classical and gate-based optimization backends. The remainder of this paper is organized as follows. Section 2 introduces the residue interaction network, the mutation-dependent Laplacian perturbations, and the resulting mutation operator configuration. Section 3 develops the Hilbert–Schmidt geometry, the fixed-cardinality variational formulation, its exact QUBO and Ising reformulations, and the optimization backends considered in this work. Section 4 presents the numerical evaluation, including the organization and robustness of the mutation operator geometry, representative-subset selection, numerical consistency of the exact reformulations, and the behavior of classical and gate-based optimization backends. Finally, Section 5 summarizes the conclusions, limitations, and directions for future work.

2. Operator representation of modeled mutation-induced perturbations

The variational framework developed in the following section is built upon an operator representation of modeled mutation-induced perturbations. We represent each mutation as a perturbation of the graph Laplacian associated with the wild-type residue interaction network, so that all mutations are described relative to a common reference operator. This construction embeds modeled mutation-induced changes within a common operator representation and provides the mathematical basis for comparing them systematically. In this section, we first introduce the residue interaction network and its graph Laplacian, then define the perturbation operator associated with each mutation, and finally construct the corresponding mutation operator space that will serve as the starting point for the geometric and variational formulation developed in the next section.

2.1. Residue interaction network representation

The proposed framework starts from a RIN, in which amino acid residues are represented as vertices and structural relationships between residues are encoded as weighted edges [10, 13]. We model the RIN as an undirected weighted graph $G = (V, E, W)$, where each vertex $v_i \in V$ represents an amino

acid residue, each edge $(v_i, v_j) \in E$ represents a residue pair considered to be structurally connected under the chosen RIN construction, and the weight function $W : E \to \mathbb{R}^+$ assigns a positive numerical weight to that edge. The mathematical construction developed below only requires a weighted RIN to be specified and applied consistently to the wild-type reference and its mutation-induced perturbations. The graph is represented by its weighted adjacency matrix $A_G = (a_{ij})_{i,j=1}^n$, with entries

$$a_{ij} = \begin{cases} W(v_i, v_j), & (v_i, v_j) \in E, \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

Because the interaction network is assumed to be undirected, $A_G = A_G^T$. The corresponding degree matrix is $D_G = \mathrm{diag}(d_1, \ldots, d_n)$, where $d_i = \sum_{j=1}^n a_{ij}$. From these quantities, we define the weighted graph Laplacian matrix $L = D_G - A_G$.

For the numerical implementation considered in this work, the RIN is constructed from the C_α coordinates of the protein structure. Two residues v_i and v_j are connected whenever their Euclidean C_α–C_α distance satisfies $\left\|\mathbf{r}_i - \mathbf{r}_j\right\|_2 \le r_c$, with $r_c = 7$ Å, and the corresponding edge weight is defined as $W(v_i, v_j) = \left\|\mathbf{r}_i - \mathbf{r}_j\right\|_2$. Accordingly, the edge weights used in the present implementation are geometric distance weights rather than interaction-strength or energetic weights. The cutoff determines the topology of the RIN, whereas the weighting rule determines the numerical entries of the adjacency matrix and therefore of the graph Laplacian. Alternative contact, distance-based, energetic, or physicochemical weighting schemes could be incorporated into the same mathematical construction without modifying the subsequent formalism. However, changing the RIN construction would in general modify L and, consequently, the mutation perturbation operators and the geometry derived from them. The operator geometry considered in this work should therefore be understood as conditional on the specified RIN representation. The Laplacian provides an operator representation of the residue interaction network that combines local connectivity with the global organization of the graph [23]. Since the RIN is undirected and positively weighted, L is a real symmetric positive semidefinite matrix [24, 25]. These properties make the Laplacian a convenient common representation for comparing weighted network representations associated with the same protein. In the proposed framework, the Laplacian of the wild-type protein is taken as the reference operator. The modeled perturbation associated with each mutation will therefore be represented through the change it induces with respect to this common reference. This relative operator representation provides the starting point for the mutation perturbation operators introduced in the next subsection.

2.2. Mutation-induced Laplacian perturbations

The graph Laplacian introduced in the previous subsection provides the common reference for describing modeled mutation-induced perturbations. Let L denote the Laplacian associated with the wild-type RIN. For each mutation i, a mutation-dependent weighted network is constructed by locally modifying the edge weights of the wild-type graph while preserving its vertex set and edge topology. The resulting weighted Laplacian is denoted by L_i. In the numerical model used in this work, the magnitude of this local reweighting depends on the physicochemical difference between the wild-type amino acid and its substituted residue. For a substitution $a \to b$, we define the normalized physicochemical

dissimilarity

$$s(a,b) = \sqrt{\frac{1}{3}\left[\left(\frac{h_a - h_b}{9}\right)^2 + \left(\frac{\mathrm{Vol}(a) - \mathrm{Vol}(b)}{170}\right)^2 + \left(\frac{q_a - q_b}{2}\right)^2\right]}, \tag{2.2}$$

where h, Vol, and q denote hydropathy, approximate residue volume, and side-chain charge, respectively. Hydropathy values are taken from the Kyte–Doolittle scale [26], while the residue-volume and side-chain-charge descriptors follow the physicochemical property values compiled in [27]. The three descriptors provide complementary information on hydrophobicity, size, and electrostatic changes. For mutation i, we denote the corresponding physicochemical dissimilarity by s_i. The resulting dissimilarity is mapped to a mutation-specific reweighting factor $\alpha_i = \mathrm{clip}\,[\alpha_{\max} - s_i(\alpha_{\max} - \alpha_{\min}), \alpha_{\min}, \alpha_{\max}]$, with $\alpha_{\min} = 0.20$ and $\alpha_{\max} = 0.85$. Thus, substitutions with larger physicochemical dissimilarity generate smaller values of α_i and therefore larger numerical perturbations of the local weighted network. Let m_i denote the residue affected by mutation i. For every existing edge (m_i, v_j) incident to the mutated residue, the wild-type weight is rescaled according to $W_i(m_i, v_j) = \alpha_i W(m_i, v_j)$. A weaker second-neighborhood perturbation is also introduced for edges connecting two first neighbors of the mutated residue. For such an edge (v_j, v_k), $W_i(v_j, v_k) = \beta_i W(v_j, v_k)$, where $\beta_i = 1 - \eta(1 - \alpha_i)$, with $\eta = 0.15$. All remaining edge weights are left unchanged. No vertices or edges are added or removed during this procedure; therefore, the wild-type and mutation-dependent networks share the same graph topology and differ only through their weights.

It is important to emphasize that this reweighting procedure is a numerical perturbation model rather than a reconstruction of a mutant-resolved three-dimensional protein structure. Since the original RIN weights are derived from wild-type C_α–C_α distances, the reweighted coefficients used to construct L_i should not be interpreted as literal post-mutation inter-residue distances. Their role is to provide a controlled and mutation-specific perturbation of the common weighted reference graph, allowing different amino-acid substitutions at the same residue position to generate distinct operators. Once the mutation-dependent weighted graph has been constructed, its Laplacian L_i is computed using the same weighted-Laplacian definition as for the wild-type network. Rather than treating L_i as an independent representation, we characterize mutation i through its deviation from the common wild-type reference.

Definition 1. The mutation perturbation operator associated with mutation i is defined as $\Delta L_i = L_i - L$, so that $L_i = L + \Delta L_i$.

The operator ΔL_i therefore captures the change induced by the mutation-specific local reweighting relative to the same wild-type Laplacian. Because both L and L_i are real symmetric Laplacian matrices, ΔL_i is itself real and symmetric. Consequently, all mutation perturbation operators belong to the same space of symmetric matrices and can be compared within a common operator representation, as formalized in the following subsection.

2.3. Mutation operator space

The perturbation operators introduced in the previous subsection place all modeled mutation-induced perturbations within a common mathematical representation. Since each ΔL_i is a real symmetric matrix, all mutation perturbations belong to the vector space $\mathcal{S}_n = \left\{X \in \mathbb{R}^{n\times n} : X = X^T\right\}$, consisting of all real symmetric $n \times n$ matrices. This common space allows the mutations associated with a given protein to be considered jointly rather than as unrelated perturbations.

Definition 2. Given a set of N mutations, we define the corresponding mutation operator space as the finite operator configuration $\mathcal{M} = \{\Delta L_1, \Delta L_2, \ldots, \Delta L_N\} \subseteq \mathcal{S}_n$, where each element represents the modeled perturbation associated with one mutation relative to the same wild-type reference.

The term *mutation operator space* will be used throughout this work to refer to this finite configuration embedded in the vector space $\mathcal{S}_n$. Once the perturbation operators have been constructed, this representation places all mutations in a common mathematical setting and allows their subsequent comparison to depend directly on the operators themselves, without introducing additional mutation-specific scalar scores or heterogeneous comparison descriptors [28]. To provide the structure required for such comparisons, we equip $\mathcal{S}_n$ with the Hilbert–Schmidt inner product $\langle X, Y\rangle_{HS} = \mathrm{Tr}\left(X^T Y\right)$, with $X, Y \in \mathcal{S}_n$ [20]. Since the operators in $\mathcal{S}_n$ are symmetric, this reduces to $\langle X, Y\rangle_{HS} = \mathrm{Tr}(XY)$, with the associated norm $\|X\|_{HS} = \sqrt{\mathrm{Tr}(X^2)}$. The construction of the mutation representation is therefore complete at this stage: each mutation is encoded by a perturbation operator ΔL_i, and the full mutation set is embedded in $\mathcal{S}_n$ endowed with the Hilbert–Schmidt structure.

3. Variational formulation and quantum optimization framework

The operator representation introduced in the previous section provides the starting point for constructing a geometry-based representative mutation selection problem. Once all modeled mutation-induced perturbations are expressed within the same operator space, their relative magnitude and pairwise relationships can be quantified directly from the Hilbert–Schmidt structure. These quantities are then used to define a variational objective that balances perturbation magnitude and geometric diversity. We subsequently show that this objective admits an exact binary quadratic formulation and can therefore be mapped, through the standard QUBO–Ising correspondence, onto an Ising cost Hamiltonian. The purpose of this section is thus to establish the complete mathematical chain from mutation-operator geometry to the induced optimization model, while keeping the subsequent choice of optimization algorithm conceptually separate from the modeling stage.

3.1. Geometry-induced variational formulation

The operator representation introduced in Section 2.3 allows relationships between mutations to be defined directly in terms of their perturbation operators. For two mutations i and j, we define their Hilbert–Schmidt distance as $d_{ij} = \left\|\Delta L_i - \Delta L_j\right\|_{HS}$. Since the Hilbert–Schmidt norm induces a metric on $\mathcal{S}_n$ [20, 29], this expression defines the corresponding metric on the finite mutation configuration $\mathcal{M}$. Once the perturbation operators have been constructed, no additional similarity function is required to compare them. Two quantities derived from this operator geometry are used to construct the representative-selection criterion. First, the magnitude of the perturbation induced by mutation i is measured by $V_i = \|\Delta L_i\|_{HS}^2 = \mathrm{Tr}\left(\Delta L_i^2\right)$. Second, the squared geometric separation between mutations i and j is defined as $D_{ij} = d_{ij}^2 = \left\|\Delta L_i - \Delta L_j\right\|_{HS}^2$. For a subset $R \subseteq \mathcal{M}$, these quantities induce the perturbation-magnitude functional

$$\mathcal{V}(R) = \sum_{\Delta L_i \in R} V_i \tag{3.1}$$

and the geometric-diversity functional

$$\mathcal{D}(R) = \sum_{\substack{\Delta L_i, \Delta L_j \in R \\ i<j}} D_{ij}. \tag{3.2}$$

The first term favors subsets containing mutations that induce large deviations from the wild-type reference, whereas the second favors subsets whose perturbation operators are mutually separated in the induced geometry, thereby discouraging redundant operator perturbations. We combine both contributions through the variational functional $\mathcal{J}(R) = -\mathcal{V}(R) - \mu\mathcal{D}(R)$, with $\mu > 0$, where μ is a modeling parameter controlling the trade-off between perturbation magnitude and geometric diversity. Its numerical selection does not alter the mathematical construction of the framework and is examined through sensitivity analysis in the experimental evaluation.

For a prescribed subset cardinality k, representative mutation selection is therefore formulated as

$$R^\star = \arg\min_{\substack{R \subseteq \mathcal{M} \\ |R|=k}} \mathcal{J}(R). \tag{3.3}$$

Here, k is treated as an externally specified parameter defining the desired size of the representative subset. The present formulation therefore addresses representative selection for a prescribed cardinality and does not attempt to determine the subset size adaptively. Extending the framework to the joint optimization of subset composition and cardinality constitutes a separate model-selection problem and lies outside the scope of the present work. This formulation is constructed entirely from quantities derived from the mutation perturbation operators: their magnitudes relative to the wild-type reference and their pairwise Hilbert–Schmidt distances. Once these perturbation operators have been constructed, no additional mutation-specific score or external pairwise similarity descriptor is introduced in the representative-selection model. Since the admissible family contains only $\binom{N}{k}$ subsets, a global minimizer exists for every $1 \le k \le N$. More importantly for the subsequent development, the variational functional has a pairwise quadratic structure: perturbation magnitudes contribute individual terms, whereas geometric diversity contributes pairwise interactions. This structure enables the representative-selection problem to be reformulated exactly as a binary quadratic optimization problem, providing the direct connection between the mutation-operator geometry and the QUBO formulation developed in the following subsection.

3.2. Binary quadratic reformulation of the variational problem

The variational problem introduced in the previous subsection selects a subset of fixed cardinality k from the mutation configuration $\mathcal{M}$. To express this problem in binary form, we associate each perturbation operator $\Delta L_i \in \mathcal{M}$ with a decision variable

$$x_i = \begin{cases} 1, & \text{if } \Delta L_i \in R, \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

A representative subset is therefore encoded by a binary vector $x \in \{0,1\}^N$, and the cardinality constraint becomes

$$\sum_{i=1}^{N} x_i = k. \tag{3.5}$$

Using the quantities V_i and D_{ij} introduced in Section 3.1, the variational functional can be written exactly as

$$\mathcal{J}(x) = -\sum_{i=1}^{N} V_i x_i - \mu \sum_{1\le i<j\le N} D_{ij} x_i x_j, \tag{3.6}$$

subject to the cardinality constraint above. The linear terms encode the magnitude of the individual mutation perturbations, whereas the quadratic terms encode their pairwise geometric separation. Thus, the quadratic structure is already determined by the mutation-operator geometry before any QUBO or Ising representation is introduced.

To obtain an unconstrained binary formulation, we incorporate the cardinality condition through the standard quadratic penalty [30]

$$Q_\lambda(x) = \mathcal{J}(x) + \lambda\left(\sum_{i=1}^{N} x_i - k\right)^2, \qquad \lambda > 0. \tag{3.7}$$

For binary optimization, a sufficiently large penalty parameter guarantees equivalence between the constrained and penalized problems. In particular, define

$$\Delta_{\mathcal{J}} = \max_{x,y\in\{0,1\}^N} |\mathcal{J}(x) - \mathcal{J}(y)|. \tag{3.8}$$

For every infeasible binary vector x, the cardinality violation is integer-valued and therefore satisfies

$$\left(\sum_{i=1}^{N} x_i - k\right)^2 \ge 1. \tag{3.9}$$

Let $x^\star$ denote a minimizer of $\mathcal{J}$ over the feasible set. Since $\left|\mathcal{J}(x) - \mathcal{J}(x^\star)\right| \le \Delta_{\mathcal{J}}$, every infeasible configuration satisfies

$$Q_\lambda(x) \ge \mathcal{J}(x^\star) - \Delta_{\mathcal{J}} + \lambda. \tag{3.10}$$

Consequently, if $\lambda > \Delta_{\mathcal{J}}$, then $Q_\lambda(x) > \mathcal{J}(x^\star) = Q_\lambda(x^\star)$ for every infeasible x. Hence, no infeasible configuration can be a global minimizer of Q_λ, and every global minimizer of the penalized problem satisfies the prescribed cardinality and also minimizes the original constrained problem. A directly computable sufficient bound follows from $V_i \ge 0$, $D_{ij} \ge 0$, and $\mu > 0$. Since $-B \le \mathcal{J}(x) \le 0$, with

$$B = \sum_{i=1}^{N} V_i + \mu \sum_{1\le i<j\le N} D_{ij}, \tag{3.11}$$

we have $\Delta_{\mathcal{J}} \le B$. Consequently, $\lambda > B$ provides a conservative sufficient condition for exact penalization. Expanding the penalty and using $x_i^2 = x_i$ gives

$$Q_\lambda(x) = \sum_{i=1}^{N} \left[-V_i + \lambda(1-2k)\right] x_i + \sum_{1\le i<j\le N} \left[-\mu D_{ij} + 2\lambda\right] x_i x_j + \lambda k^2. \tag{3.12}$$

Under the symmetric QUBO convention [31], $Q_\lambda(x) = x^T A x + \lambda k^2$, the corresponding matrix is therefore

$$A = -\operatorname{diag}(V) - \frac{\mu}{2} D + \lambda \mathbf{1}\mathbf{1}^T - 2\lambda k I_N. \tag{3.13}$$

This expression makes explicit the two different contributions to the binary model, $A = A_{\text{geom}} + A_{\text{card}}$, where $A_{\text{geom}} = -\operatorname{diag}(V) - \frac{\mu}{2}D$ contains the information derived from the mutation-operator representation, whereas $A_{\text{card}} = \lambda \mathbf{1}\mathbf{1}^T - 2\lambda k I_N$ arises exclusively from enforcing the prescribed subset cardinality. Hence, once the mutation perturbation operators have been constructed, the QUBO reformulation introduces no additional mutation-specific scores or pairwise similarity measures: its data-dependent coefficients are inherited directly from V_i and D_{ij}, while the penalty term serves only to enforce feasibility. This separation becomes particularly relevant in the next subsection, where the QUBO model is mapped to an Ising Hamiltonian and the dependence of its pairwise couplings on the operator-space geometry becomes explicit.

3.3. Geometry-induced Ising Hamiltonian

The binary quadratic formulation derived in the previous subsection provides an exact QUBO representation of the representative mutation problem. We now map this formulation onto an Ising model using the standard correspondence between binary and spin variables [32], $x_i = \frac{1-z_i}{2}$, and $z_i \in \{-1, +1\}$. Under the symmetric QUBO convention $Q_\lambda(x) = x^T A x + \lambda k^2$, this transformation yields the Ising energy function

$$H(z) = E_0 + \sum_{i=1}^{N} h_i z_i + \sum_{1 \le i < j \le N} J_{ij} z_i z_j, \tag{3.14}$$

where $E_0 = \lambda k^2 + \frac{1}{4}\left(\mathbf{1}^T A \mathbf{1} + \operatorname{Tr}(A)\right)$, $h_i = -\frac{1}{2}\sum_{j=1}^{N} A_{ij}$, $J_{ij} = \frac{1}{2}A_{ij}$, and $i \neq j$. This is the standard QUBO–Ising transformation [32]. Since the binary-to-spin map is bijective, the QUBO and Ising formulations possess exactly the same minimizing configurations [32]. The associated quantum cost Hamiltonian is obtained by replacing each classical spin variable z_i with the Pauli-Z operator Z_i [33, 34]:

$$H_C = E_0 I + \sum_{i=1}^{N} h_i Z_i + \sum_{1 \le i < j \le N} J_{ij} Z_i Z_j. \tag{3.15}$$

The additive constant E_0 produces only a uniform spectral shift and therefore does not affect the minimizing configuration. The relevant point for the proposed framework is not the standard binary-to-spin transformation itself, but the structure acquired by the Ising coefficients when the QUBO matrix derived from the mutation-operator geometry is substituted into these expressions. This relation is formalized in the following proposition.

Proposition 1. *Let $A = -\operatorname{diag}(V) - \frac{\mu}{2}D + \lambda \mathbf{1}\mathbf{1}^T - 2\lambda k I_N$ be the QUBO matrix obtained from the geometry-induced variational formulation. Then the one-body and two-body coefficients of the corresponding Ising Hamiltonian satisfy*

$$h_i = \frac{1}{2}V_i + \frac{\mu}{4}\sum_{j=1}^{N} D_{ij} + \lambda\left(k - \frac{N}{2}\right), \tag{3.16}$$

and, for $i \neq j$, $J_{ij} = \frac{\lambda}{2} - \frac{\mu}{4}D_{ij}$. Equivalently,

$$J_{ij} = \frac{\lambda}{2} - \frac{\mu}{4}\left\|\Delta L_i - \Delta L_j\right\|_{HS}^2. \tag{3.17}$$

Proof. From the standard QUBO–Ising correspondence,

$$h_i = -\frac{1}{2}\sum_{j=1}^{N} A_{ij}, \qquad J_{ij} = \frac{1}{2}A_{ij}. \tag{3.18}$$

For $i \neq j$, the QUBO matrix satisfies $A_{ij} = \lambda - \frac{\mu}{2}D_{ij}$, which immediately gives $J_{ij} = \frac{\lambda}{2} - \frac{\mu}{4}D_{ij}$. Similarly, summing the i-th row of A gives

$$\sum_{j=1}^{N} A_{ij} = -V_i - \frac{\mu}{2}\sum_{j=1}^{N} D_{ij} + \lambda N - 2\lambda k, \tag{3.19}$$

from which the stated expression for h_i follows. □

The proposition makes explicit how the mutation representation propagates into the Hamiltonian. The local fields depend on the perturbation magnitude V_i, the total geometric separation of mutation i from the remaining operators, and the cardinality penalty. More importantly, every pairwise Ising coupling is an affine function of the squared Hilbert–Schmidt distance between the corresponding mutation perturbation operators. The interaction coefficients are therefore not independently assigned similarity parameters: their data-dependent component is inherited directly from the mutation-operator geometry. This relation also implies that the pairwise geometry can be recovered from the Ising interaction coefficients.

Corollary 1. *Assume that $\mu > 0$. Then the squared Hilbert–Schmidt distances are uniquely determined by the Ising couplings through*

$$D_{ij} = \frac{4}{\mu}\left(\frac{\lambda}{2} - J_{ij}\right). \tag{3.20}$$

Furthermore, for any two mutation pairs (i, j) and (k, ℓ), $D_{ij} < D_{k\ell} \iff J_{ij} > J_{k\ell}$.

Proof. The result follows directly from the affine relation

$$J_{ij} = \frac{\lambda}{2} - \frac{\mu}{4}D_{ij}, \tag{3.21}$$

whose slope is non-zero and negative for $\mu > 0$. □

Thus, the numerical values of the Ising couplings differ from the original squared operator distances, but the complete pairwise distance information remains recoverable. In particular, once μ and λ are fixed, the coupling matrix provides a one-to-one encoding of the squared-distance matrix D. Together with the local fields, which also depend on the perturbation magnitudes V_i, this shows that, for fixed k, μ, and λ, the nonconstant part of the Hamiltonian is completely determined by the geometric quantities V and D. The geometric and cardinality contributions identified at the QUBO level are also preserved by the Ising transformation.

Proposition 2. *Let $A = A_{\text{geom}} + A_{\text{card}}$, with $A_{\text{geom}} = -\operatorname{diag}(V) - \frac{\mu}{2}D$ and $A_{\text{card}} = \lambda\mathbf{1}\mathbf{1}^T - 2\lambda k I_N$. Then the corresponding cost Hamiltonian can be written as $H_C = H_{\text{geom}} + H_{\text{card}} + E_0 I$, where H_{geom} contains the contributions induced by the perturbation magnitudes and pairwise operator distances, whereas H_{card} contains the contribution arising from the quadratic penalty enforcing the prescribed subset cardinality.*

Proof. The Ising coefficients h_i and J_{ij} depend linearly on the entries of the QUBO matrix A. Therefore, the decomposition $A = A_{\text{geom}} + A_{\text{card}}$ induces the corresponding decomposition of the one-body and two-body Ising coefficients, and hence of the cost Hamiltonian. □

This decomposition preserves throughout the modeling pipeline the distinction between the information extracted from the mutation-operator representation and the algebraic contribution required to enforce feasibility. The geometry-induced component determines the data-dependent structure of the optimization model, whereas the cardinality component only enforces the prescribed number of selected mutations.

3.4. Optimization backends for the induced Hamiltonian

The construction developed in the previous subsections is complete once the Ising cost Hamiltonian H_C has been obtained. At this stage, the representative mutation problem is fully specified independently of the computational method used to solve it. The remaining task is therefore to determine a minimum-energy configuration of the induced Hamiltonian and map that configuration back to the corresponding subset of mutation perturbation operators. At the classical spin level, the penalized optimization problem is

$$z^\star = \arg\min_{z\in\{-1,+1\}^N} H(z). \tag{3.22}$$

The resulting spin configuration is transformed back into binary variables through $x_i^\star = \frac{1-z_i^\star}{2}$, from which the representative mutation subset is recovered as $R^\star = \left\{\Delta L_i \in \mathcal{M} : x_i^\star = 1\right\}$. Consequently, different optimization procedures can operate on the same objective without modifying the mutation representation, the variational criterion, or the QUBO and Ising formulations derived in the preceding subsections. For small problem instances, exhaustive enumeration provides the exact minimum and therefore serves as a reference solution. The same optimization problem can also be addressed using classical heuristic procedures such as simulated annealing, which explores the discrete energy landscape through a temperature-controlled stochastic process [21]. When the prescribed cardinality k is enforced directly by the optimization backend, the search may instead be restricted to the feasible subspace

$$\mathcal{F}_k = \left\{ x \in \{0,1\}^N : \sum_{i=1}^N x_i = k \right\}. \tag{3.23}$$

Within this subspace, the quadratic cardinality penalty vanishes identically, and minimizing the penalized QUBO or Ising objective is therefore equivalent to minimizing the original variational objective over the admissible configurations. In the numerical experiments, simulated annealing exploits this structure through cardinality-preserving swap moves, so every intermediate configuration remains feasible. The fixed-cardinality structure can likewise be incorporated directly into gate-based variational quantum optimization. Rather than using the standard QAOA initialization $|+\rangle^{\otimes N}$ together with a transverse-field mixer, we consider a cardinality-preserving QAOA ansatz defined on the Hamming-weight-k subspace. The initial state is the uniform Dicke state

$$|D_k^N\rangle = \binom{N}{k}^{-1/2} \sum_{|x|=k} |x\rangle, \tag{3.24}$$

which has support only on feasible binary configurations [22]. Mixing is generated by local XY interactions on a ring,

$$H_M^{XY} = \frac{1}{2} \sum_{(i,j)\in E_{\text{ring}}} \left(X_i X_j + Y_i Y_j \right), \tag{3.25}$$

where E_{ring} denotes the set of nearest-neighbor edges of the ring. Since H_M^{XY} preserves the total Hamming weight, the variational evolution remains entirely within the prescribed-cardinality subspace. The corresponding depth-p state is

$$|\psi_p(\boldsymbol{\beta}, \boldsymbol{\gamma})\rangle = \prod_{\ell=1}^{p} U_M^{XY}(\beta_\ell) e^{-i\gamma_\ell H_C} |D_k^N\rangle, \tag{3.26}$$

where $U_M^{XY}(\beta_\ell)$ denotes the cardinality-preserving mixer evolution. The parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ are obtained by minimizing $\left\langle \psi_p(\boldsymbol{\beta}, \boldsymbol{\gamma}) \middle| H_C \middle| \psi_p(\boldsymbol{\beta}, \boldsymbol{\gamma}) \right\rangle$. Because every basis state reached by this ansatz satisfies the cardinality constraint, the penalty contribution is constant and vanishes on the feasible subspace. Therefore, the quantum variational procedure optimizes the same geometry-induced representative-selection objective as the constrained classical formulation.

In the numerical implementation used in Section 4, the XY mixer is applied through a symmetric even–odd–even decomposition of local two-qubit XY rotations on the ring. The resulting mixer graph is connected on the fixed-cardinality configuration space considered in the experiments, allowing transitions between all feasible configurations. The Dicke state is initialized exactly in the state-vector simulation; the circuit cost associated with preparing this state on physical hardware is not included in the present analysis. More generally, the same cost Hamiltonian may also be used as the problem Hamiltonian in quantum annealing, where the system evolves from an initial Hamiltonian toward H_C according to an adiabatic or annealing schedule [35, 36]. Quantum annealing is not evaluated numerically in the present work, but its compatibility illustrates that the geometry-induced Hamiltonian is independent of the particular optimization mechanism. Exact enumeration, classical Ising heuristics, constrained gate-based variational optimization, and quantum annealing therefore differ in how they search the energy landscape, while operating on the same underlying optimization model. The complete computational workflow of the proposed framework can be summarized in Algorithm 1. The computational cost can be separated into model construction and optimization. Let n denote the number of residues and N the number of candidate mutations. Starting from the wild-type weighted RIN, constructing the N mutation-dependent weighted networks and their corresponding Laplacian perturbations requires at most $O(Nn^2)$ operations under a direct dense-matrix implementation. Evaluating the N perturbation magnitudes likewise requires $O(Nn^2)$ operations, whereas constructing the complete pairwise distance matrix D requires $O(N^2n^2)$ operations. Once V and D have been obtained, construction of the QUBO and Ising coefficients requires $O(N^2)$ operations and $O(N^2)$ storage in the general dense case. The pairwise-distance calculation is therefore the dominant polynomial cost of the model-construction stage. The subsequent optimization stage is combinatorial. For a prescribed cardinality k, exhaustive enumeration requires consideration of $\binom{N}{k}$ candidate subsets and therefore becomes impractical as N grows. Simulated annealing and other heuristic Ising solvers avoid exhaustive enumeration, although their runtime and convergence behavior depend on the particular algorithm and problem instance. In direct gate-based quantum encoding, one binary decision variable is associated with each mutation and therefore N qubits are required. The geometry-induced cost Hamiltonian is generally dense and may

Algorithm 1 Geometry-induced representative mutation selection

Require: Wild-type weighted RIN $G = (V, E, W)$, set of N mutations, prescribed subset size k, diversity parameter μ, and penalty parameter λ

Ensure: Representative mutation subset $R^\star$

Compute the wild-type Laplacian L.
for $i = 1, \ldots, N$ **do**
    Compute the physicochemical dissimilarity s_i associated with mutation i.
    Compute the corresponding reweighting factors α_i and β_i.
    Construct the mutation-dependent weighted RIN by applying the local reweighting defined in Section 2.2.
    Compute the mutation-dependent Laplacian L_i.
    Compute the perturbation operator $\Delta L_i = L_i - L$.
end for
for $i = 1, \ldots, N$ **do**
    Compute the perturbation magnitude $V_i = \|\Delta L_i\|_{HS}^2$.
end for
for all pairs $1 \le i < j \le N$ **do**
    Compute the pairwise squared distance $D_{ij} = \|\Delta L_i - \Delta L_j\|_{HS}^2$.
end for
Construct the geometry-induced variational objective $\mathcal{J}$.
Construct the penalized QUBO matrix $A = -\operatorname{diag}(V) - \frac{\mu}{2}D + \lambda \mathbf{1}\mathbf{1}^T - 2\lambda k I_N$.
Map the QUBO formulation to the Ising cost Hamiltonian H_C.
Minimize the resulting objective using a selected optimization backend, either over the penalized full binary space or directly within the feasible fixed-cardinality subspace.
Convert the minimizing spin configuration into binary variables.
Recover $R^\star = \{\Delta L_i \in \mathcal{M} : x_i^\star = 1\}$.

contain up to $O(N^2)$ pairwise Z_iZ_j terms. By contrast, the cardinality-preserving XY mixer employed in the numerical experiments contains $O(N)$ local two-qubit interactions on a ring. The resulting circuit depth depends on the decomposition of the dense cost layer, the mixer implementation, and the connectivity of the target hardware. The preparation of the initial Dicke state introduces an additional hardware-dependent cost that is not included in the present state-vector study. Analogously, the physical embedding cost for quantum annealing is hardware-dependent. Hence, the proposed formulation provides a polynomial-time construction of the optimization model, while the difficulty of obtaining its global minimum remains governed by the combinatorial optimization stage. This separation between modeling and optimization is a central feature of the proposed framework. The mutation perturbation operators determine the geometric quantities (V, D), these quantities define the variational functional, and the resulting objective determines the QUBO and Ising formulations through exact reformulations. The choice of optimization backend therefore occurs only after the mathematical model has been fixed and does not alter the operator-derived information encoded in the induced Hamiltonian.

4. Numerical results

This section evaluates the mathematical framework developed in Section 3 through a sequence of numerical experiments. Rather than focusing on a particular quantum optimization algorithm, the experiments examine each stage of the proposed modeling pipeline: the geometry induced by the mutation perturbation operators, the resulting representative-selection problem, the robustness of the selected subsets, the numerical consistency of the exact QUBO and Ising reformulations, and the behavior of classical and gate-based optimization backends applied to the same induced objective.

4.1. Experimental setup

All numerical experiments were performed on human KRAS using a panel of fourteen missense mutations located at recurrent mutational sites. The analyzed panel comprised six substitutions at residue G12 (G12A, G12C, G12D, G12R, G12S, and G12V), three substitutions at G13 (G13C, G13D, and G13V), three substitutions at Q61 (Q61H, Q61L, and Q61R), together with K117N and A146T. Table 1 summarizes the fixed numerical configuration used throughout the experiments, while the methodological choices requiring further interpretation are described below.

Table 1. Summary of the experimental configuration adopted throughout the numerical evaluation.

Component	Configuration
Structural reference	Human KRAS, PDB 4OBE, chain A
Mutation panel	14 missense mutations
Main RIN	169 C_α nodes, 666 edges, 7 Å cutoff, Euclidean-distance weights
Mutation reweighting	$\alpha_{\min} = 0.20$, $\alpha_{\max} = 0.85$, $\eta = 0.15$; fixed topology
Operator representation	$\Delta L_i = L_i - L$
Operator geometry	$V_i = \|\|\Delta L_i\|\|_{\mathrm{HS}}^2$, $D_{ij} = \|\|\Delta L_i - \Delta L_j\|\|_{\mathrm{HS}}^2$
Main cardinality	$k = 5$
Cardinality robustness	$k = 4, 6$
Main diversity parameter	$\mu = 1$
Sensitivity analyses	μ; distance-weighted RINs at 6, 7, 8 Å; binary RIN at 7 Å
Baseline	Top-V_i ranking
Exact reference	Exhaustive feasible-subset search
QUBO penalty	$\lambda = 1.05B$
Classical backend	Fixed-k simulated annealing; 100 restarts, 1000 steps
Quantum backend	Dicke-state XY-QAOA; $p = 1, \ldots, 5$, 20 multistart runs, maximum 800 iterations, 3000 sampling shots
Geometric diagnostics	Hierarchical clustering, silhouette analysis, multidimensional scaling (MDS), cluster coverage, perturbation-diversity analysis
Exact consistency checks	Variational/QUBO/Ising minima and reconstruction of D_{ij} from J_{ij}
Software	Python, NumPy, SciPy, NetworkX, Biopython, scikit-learn, joblib, and Matplotlib
Reproducibility	Public GitHub repository

The wild-type residue interaction network was constructed from the guanosine diphosphate (GDP)-bound human KRAS crystal structure PDB 4OBE, chain A [37], as reported in Table 1, by representing each standard residue containing a C_α atom as a node. Two residues were connected whenever their C_α–C_α Euclidean distance was below the prescribed cutoff, and each retained edge in the main representation was weighted directly by that Euclidean distance. The alternative RIN constructions listed in Table 1 were used only for the sensitivity analysis described later. For each mutation i, the mutation-dependent network was generated from the same wild-type topology by applying the local numerical reweighting model described in Section 2.2. No vertices or edges were added or removed. Accordingly, the resulting coefficients are interpreted as model-defined perturbations of the weighted network rather than as reconstructed post-mutation inter-residue distances. The Laplacian L_i of each mutation-dependent network was compared with the wild-type Laplacian L through the perturbation operator $\Delta L_i = L_i - L$. The resulting operator geometry was characterized by the perturbation magnitudes $V_i = \|\Delta L_i\|_{\mathrm{HS}}^2$ and the pairwise squared distances $D_{ij} = \|\Delta L_i - \Delta L_j\|_{\mathrm{HS}}^2$. These quantities define the variational objective and, through the exact reformulations derived in Section 3, determine the corresponding QUBO and Ising representations.

The representative-selection experiments used $k = 5$ as the main cardinality, with $k = 4$ and $k = 6$ included as robustness cases. The diversity parameter was set to the nominal value $\mu = 1$ in the main experiments and subsequently varied in a dedicated sensitivity analysis. For $k = 5$, the selected subset remains unchanged throughout the tested interval from $\mu = 0.1$ to $\mu = 5$. The value $\mu = 1$ is therefore used as a nominal reference within this stable region rather than as a separately optimized hyperparameter. The representative-selection model was evaluated against a perturbation-magnitude baseline that selects the k mutations with the largest values of V_i. Exact exhaustive search over the feasible subsets was used to determine the reference global optimum. The geometry of the resulting selections was further examined through hierarchical clustering, multidimensional scaling, cluster coverage, and perturbation-diversity analyses. In particular, the number of hierarchical clusters was selected from the data using the silhouette score rather than prescribed a priori. Optimization backends were evaluated separately from the representative-selection analysis. The penalized QUBO/Ising formulation used the value of λ reported in Table 1 to verify the exact equivalence with the constrained variational problem. By contrast, simulated annealing and XY-QAOA were implemented directly in the fixed-cardinality subspace, where every admissible state satisfies $|x| = k$ and the cardinality penalty therefore vanishes identically. The classical backend used cardinality-preserving swap moves. The quantum backend was evaluated through a state-vector simulation of a cardinality-preserving XY-QAOA ansatz initialized in the uniform Dicke state over all Hamming-weight-k configurations. The XY mixer was defined on a ring and preserves the feasible subspace by construction. The numerical settings used for both optimization backends are summarized in Table 1. Finally, the exact QUBO-to-Ising reformulation was numerically verified by comparing the global minima of the variational, QUBO, and Ising formulations. The affine relation between the Ising couplings and the squared operator distances was independently checked at numerical precision by reconstructing D_{ij} from J_{ij}. These calculations are treated as numerical consistency checks of the analytical identities derived in Section 3 rather than as independent empirical evidence for geometry preservation. All computations were implemented using the scientific-computing libraries listed in Table 1. The complete computational pipeline, including the input data, executable notebook, generated results, and scripts required to reproduce the numerical analysis, is publicly available at the project GitHub repository.

4.2. *Geometry of the mutation operator space*

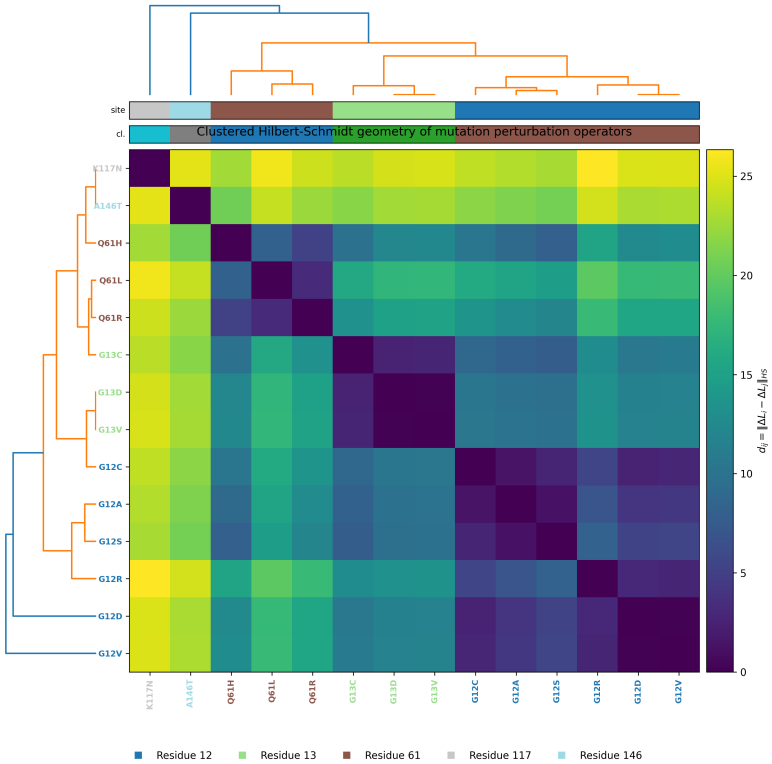


Figure 1. Pairwise Hilbert-Schmidt distance matrix between mutation perturbation operators together with hierarchical clustering. The distance matrix exhibits coherent geometric neighborhoods. A silhouette-based analysis selects five clusters, corresponding in the present dataset to the G12, G13, and Q61 mutation families together with separate K117N and A146T groups.

Before analyzing representative subsets, we first characterize the geometry induced by the mutation-operator representation introduced in Section 2. Each mutation is represented by a perturbation operator ΔL_i, and pairwise relationships are quantified through the Hilbert–Schmidt distance $d_{ij} = \|\Delta L_i - \Delta L_j\|_{\mathrm{HS}}$. The purpose of this subsection is therefore not to optimize representative subsets, but to examine the organization of the finite operator configuration generated by the perturbation model and to assess how strongly this organization depends on the adopted RIN construction. Figure 1 presents the complete pairwise Hilbert–Schmidt distance matrix together with the corresponding hierarchical clustering. The matrix exhibits a clear block structure rather than an approximately uniform distribution of distances, showing that the modeled perturbation operators form distinct geometric neighborhoods. To avoid prescribing the number of groups a priori, the number of hierarchical clusters was selected using the silhouette score over candidate partitions containing between two and eight clusters. The maximum silhouette value was obtained for five clusters, with a score of 0.583. The resulting partition separates the G12, G13, and Q61 substitutions into three groups, while K117N and A146T form two additional clusters. This site-aligned organization should be interpreted in the context of the perturbation construction itself. Mutations affecting the same residue modify overlapping local edge neighborhoods of the wild-type RIN and therefore naturally generate related Laplacian perturbations, while the mutation-

specific physicochemical reweighting distinguishes different substitutions within each site. Conversely, mutations at different residue positions act on different local network neighborhoods, which can produce greater separation in operator space. Thus, the observed clustering provides evidence that the operator representation retains the local structure encoded by the perturbation model; it should not be interpreted as an independent recovery of biological mutation classes.

While hierarchical clustering characterizes local relationships, a two-dimensional MDS embedding provides a complementary global visualization of the operator configuration. Figure 2 shows this embedding, where each point represents a perturbation operator, colors identify the mutated residue, and marker size is proportional to the perturbation magnitude $V_i = \|\Delta L_i\|_{\mathrm{HS}}^2 = \mathrm{Tr}(\Delta L_i^2)$. The embedding is consistent with the organization identified by hierarchical clustering. Mutations associated with the same residue occupy nearby regions, whereas K117N and A146T remain well-separated from the G12, G13 and Q61 families. At the same time, substantial differences in V_i are observed among operators that belong to the same geometric neighborhood. Pairwise separation and individual perturbation magnitude therefore provide complementary information: one describes the position of a mutation relative to the remaining operator configuration, whereas the other quantifies its deviation from the wild-type reference.

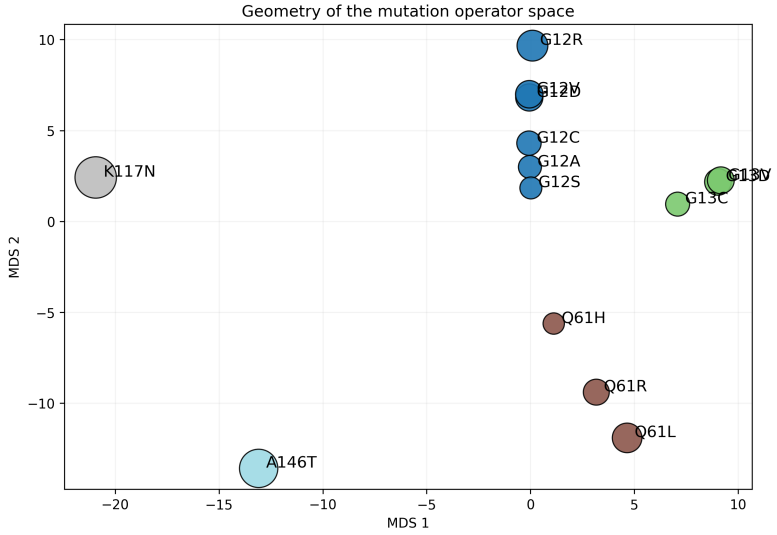


Figure 2. Two-dimensional MDS visualization of the mutation operator space constructed from Hilbert–Schmidt distances. Colors denote mutation sites and marker sizes are proportional to the perturbation magnitude $V_i = \mathrm{Tr}(\Delta L_i^2)$.

The distribution of pairwise Hilbert–Schmidt distances is shown in Figure 3. The observed distances range from approximately 0.17 to 26.34, with a mean of 13.66, a median of 12.81, and a standard deviation of 7.45. The broad distribution indicates that the pairwise separations are not concentrated around a single scale. This variability is relevant for representative selection because the diversity contribution $\mathcal{D}(R) = \sum_{i<j} D_{ij}$ can discriminate among subsets containing operators with substantially different degrees of geometric separation rather than behaving as an approximately constant correction.

The second geometric ingredient of the variational formulation is the perturbation magnitude. Figure 4 ranks the fourteen mutations according to V_i. K117N and A146T exhibit the largest magnitudes in the analyzed panel, followed by G12R and Q61L, while the remaining G12, G13, and Q61 substitutions span a broad range of values. Importantly, this ranking does not reproduce the information contained in

the pairwise-distance geometry. Mutations with comparatively large V_i may still lie close to another high-magnitude perturbation in operator space. This distinction motivates the use of both V_i and D_{ij} in the variational formulation: the former favors large deviations from the wild-type reference, whereas the latter penalizes geometrically redundant selections.

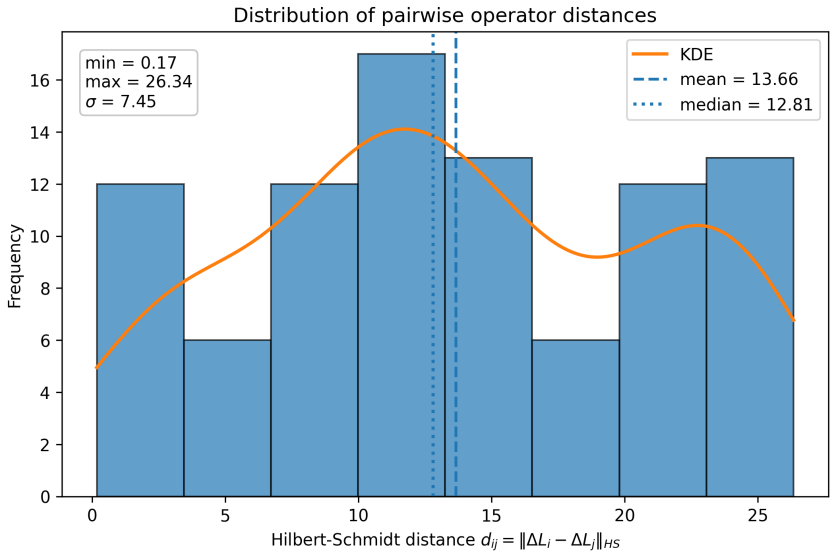


Figure 3. Distribution of pairwise Hilbert-Schmidt distances between mutation perturbation operators. The broad observed range shows that the induced operator geometry contains substantial variation in pairwise separation.

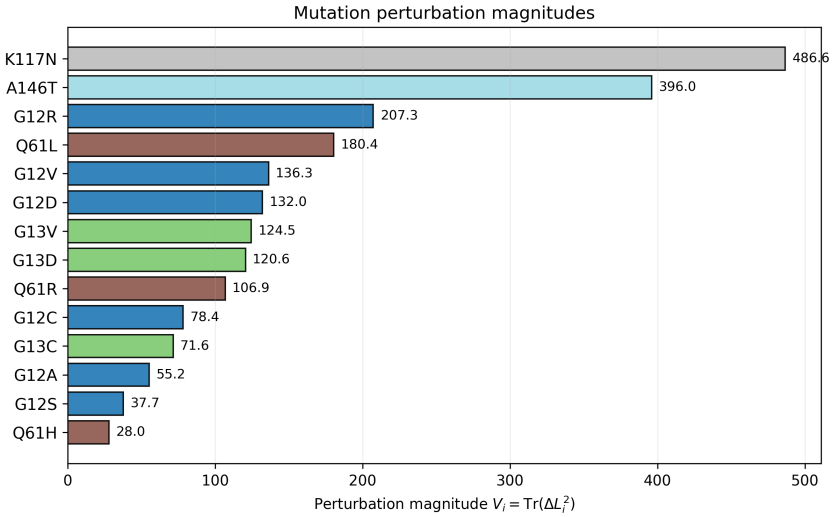


Figure 4. Ranking of mutation perturbation magnitudes $V_i = \mathrm{Tr}(\Delta L_i^2)$. The substantial variability across mutations provides the individual-magnitude component of the variational criterion, complementary to the pairwise operator distances.

Because the operator geometry depends on the underlying RIN representation, we additionally evaluated its sensitivity to alternative network constructions. The 7 Å distance-weighted RIN used

in the main analysis contains 666 edges. Reducing the cutoff to 6 Å produces 454 edges, whereas increasing it to 8 Å produces 809 edges. A binary-contact representation using the same 7 Å topology was also considered. Despite these changes, the complete pairwise operator-distance geometry remained strongly correlated with the main representation. Relative to the 7 Å distance-weighted baseline, the Pearson/Spearman correlations of the pairwise Hilbert–Schmidt distances were 0.976/0.980 for the 6 Å network, 0.969/0.984 for the 8 Å network, and 0.998/0.998 for the binary 7 Å representation. Figure 5 summarizes these correlations across the four tested RIN constructions and makes explicit that the alternative representations remain close to the baseline geometry despite substantial changes in network density and edge weighting. These results indicate that the global organization of the operator configuration remains similar across the tested cutoff and edge-weighting choices, although it is not strictly invariant under changes in the RIN construction.

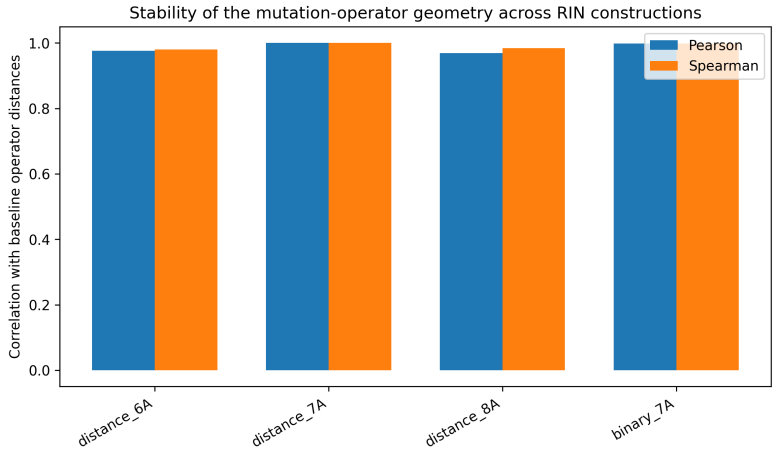


Figure 5. Sensitivity of the mutation-operator geometry to alternative RIN constructions. Pearson and Spearman correlations are computed between the complete set of pairwise Hilbert–Schmidt distances obtained for each RIN construction and those of the main 7 Å distance-weighted representation. The high correlations observed for the 6 Å, 8 Å, and binary 7 Å alternatives indicate that the global operator-space organization is stable across the tested network definitions, while not implying exact invariance.

The representative subsets obtained under these alternative geometries are examined in the following subsection, where the robustness of the actual selection is considered separately from the robustness of the underlying pairwise geometry.

4.3. Variational representative subset selection

Having characterized the geometry induced by the mutation perturbation operators, we next evaluate whether this information changes the composition of representative subsets relative to selection based only on individual perturbation magnitude. For a prescribed cardinality k, the representative subset is obtained by minimizing the variational functional introduced in Section 3.1, $\mathcal{J}(R) = -\mathcal{V}(R) - \mu\mathcal{D}(R)$, where $\mathcal{V}(R)$ is the total perturbation magnitude of the selected operators and $\mathcal{D}(R)$ quantifies their pairwise geometric diversity. The primary analysis uses $k = 5$ and $\mu = 1$, as specified in Section 4.1. Exact enumeration is used here to identify the model-defined optimum. The comparison at the model

level is therefore made against the perturbation-magnitude baseline that selects the k largest values of V_i, while randomly sampled subsets provide context for the distribution of feasible solutions. Optimization backends are evaluated separately in Section 4.5.

Figure 6 shows the resulting perturbation-magnitude–diversity plane. For $k = 5$, the exact variational optimum is $R_{\text{var}} = \{\text{G12R}, \text{G13V}, \text{Q61L}, \text{K117N}, \text{A146T}\}$, with $\mathcal{V}(R_{\text{var}}) = 1394.667$, $\mathcal{D}(R_{\text{var}}) = 5180.325$, and $\mathcal{J}(R_{\text{var}}) = -6574.992$. By contrast, selecting the five largest individual perturbation magnitudes gives $R_{\text{top-}V} = \{\text{G12V}, \text{G12R}, \text{Q61L}, \text{K117N}, \text{A146T}\}$, with $\mathcal{V}(R_{\text{top-}V}) = 1406.470$, $\mathcal{D}(R_{\text{top-}V}) = 5042.881$, and $\mathcal{J}(R_{\text{top-}V}) = -6449.351$. Thus, relative to the magnitude-only subset, the variational solution decreases total perturbation magnitude by only approximately 0.84%, while increasing geometric diversity by approximately 2.73%. The difference between the two subsets consists of a single replacement: G12V is replaced by G13V.

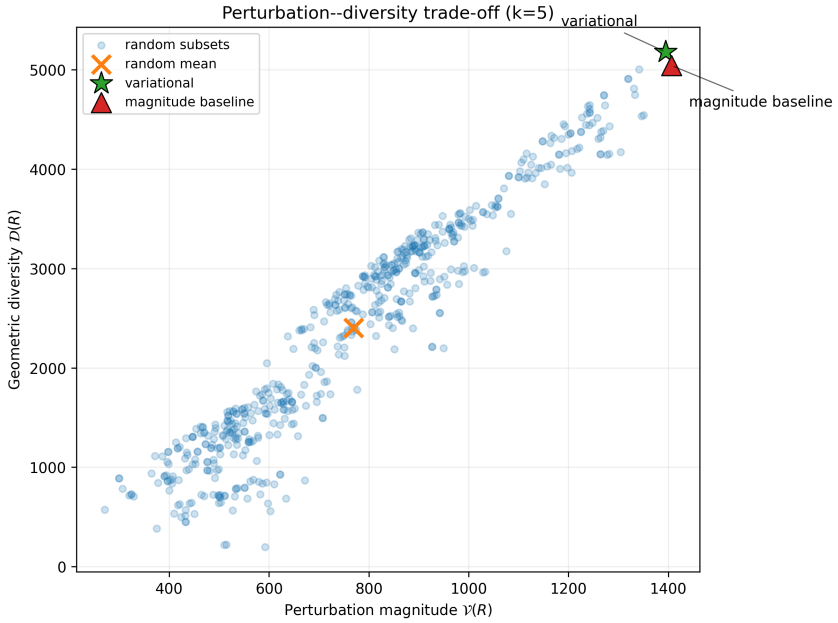


Figure 6. Perturbation magnitude $\mathcal{V}(R)$ versus geometric diversity $\mathcal{D}(R)$ for representative subsets of cardinality $k = 5$. Randomly sampled subsets provide the background distribution of feasible solutions, while the perturbation-magnitude baseline and the exact variational optimum are highlighted.

The geometric effect of this replacement is more apparent in the MDS representation shown in Figure 7. The magnitude-only criterion selects both G12V and G12R from the same local operator-space neighborhood and contains no representative from the G13 group. The variational criterion retains G12R but replaces G12V with G13V, thereby extending the selected subset into a previously uncovered region of the operator configuration. This change is also reflected quantitatively in direct coverage measures. The minimum Hilbert–Schmidt distance between any pair of selected representatives increases from 2.723 for the magnitude-only subset to 13.426 for the variational subset, an approximately 4.9-fold increase. At the same time, the mean distance from each mutation operator to its nearest selected representative decreases from 4.152 to 2.893, corresponding to an improvement of approximately 30.3% in this coverage measure. These results show that the diversity term does not simply increase the

aggregate pairwise distance: in this instance, it specifically removes a nearly redundant selection while improving representation of the complete operator configuration. The same effect is visible when the five data-driven clusters identified in Section 4.2 are used as a discrete measure of coverage. The magnitude-only subset covers four of the five clusters because both G12V and G12R belong to the G12 cluster, whereas no G13 mutation is selected. The variational subset contains G12R, G13V, Q61L, K117N, and A146T, thereby placing one representative in each of the five clusters. Cluster coverage consequently increases from $4/5$ to $5/5$. Since these clusters arise from the modeled operator geometry, this result should be interpreted as improved coverage of the induced perturbation space rather than as an independent recovery of biological mutation classes.

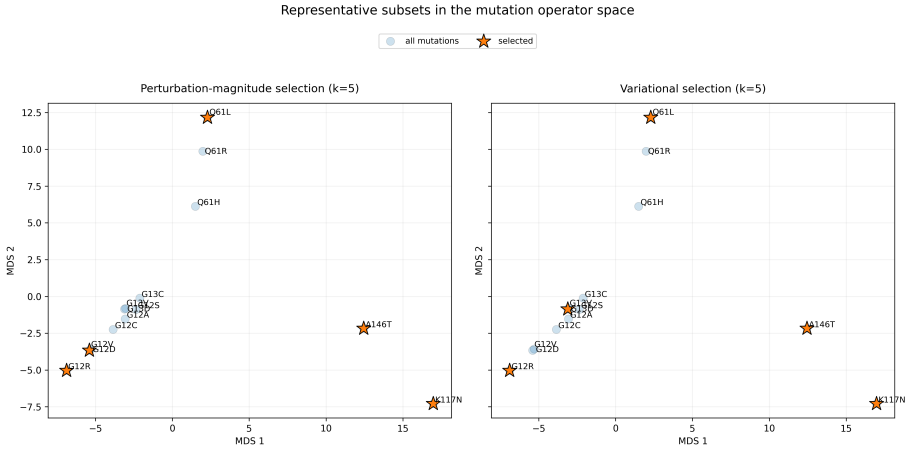


Figure 7. Comparison of the representative subsets obtained from the perturbation-magnitude baseline and the geometry-induced variational criterion for $k = 5$. Selected mutations are indicated on the common MDS representation of the operator configuration. The variational solution replaces G12V with G13V, reducing local redundancy and extending coverage to the G13 operator-space neighborhood.

Table 2. Coverage of the five data-driven operator-space clusters by the perturbation-magnitude baseline and the proposed variational selection for $k = 5$.

Method	Selected mutations	Cluster coverage
Perturbation magnitude	G12V, G12R, Q61L, K117N, A146T	4 / 5
Variational optimization	G12R, G13V, Q61L, K117N, A146T	5 / 5

The site composition of the variational subset also admits a useful biological interpretation within the KRAS case study. G12 and G13 lie in the P-loop involved in nucleotide binding, whereas Q61 is located in the switch-II region and participates in the GTP-hydrolysis mechanism. K117 and A146 are located in conserved nucleotide-binding regions involved in interactions with the guanine nucleotide and in the regulation of nucleotide exchange [38]. The five mutational sites considered here therefore span distinct aspects of KRAS nucleotide handling, and the variational subset retains representation from each site rather than concentrating two representatives at G12. This observation provides biological context for the geometric coverage result, but the specific preference for G12R over G12V or for G13V over other G13 substitutions remains a consequence of the present perturbation model and should not

be interpreted as a ranking of their clinical or biological importance.

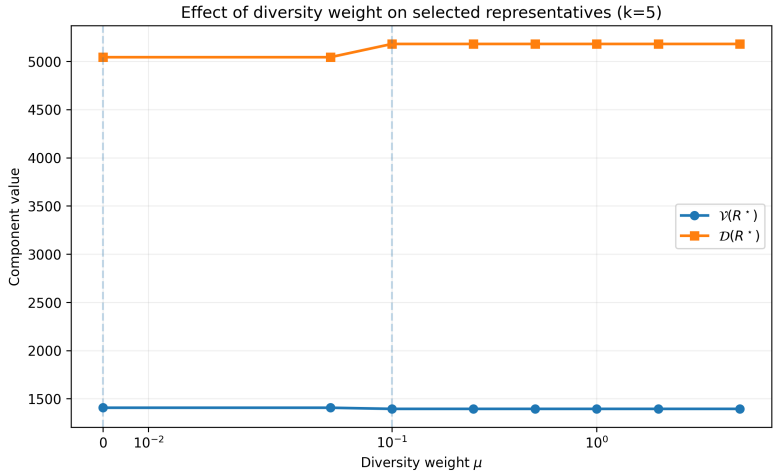


Figure 8. Evolution of the perturbation-magnitude and geometric-diversity components of the optimal $k = 5$ subset as a function of the diversity weight μ. A change in the optimal composition occurs between $\mu = 0.05$ and $\mu = 0.10$, after which both components remain constant over the tested range.

The influence of the diversity parameter μ is shown in Figures 8 and 9. At $\mu = 0$ and $\mu = 0.05$, the optimum coincides with the perturbation-magnitude subset, {G12V, G12R, Q61L, K117N, A146T}. At $\mu = 0.10$, the solution changes to {G12R, G13V, Q61L, K117N, A146T}, and this subset remains unchanged for all tested values up to $\mu = 5$. The parameter sweep therefore identifies a discrete change in the optimum between $\mu = 0.05$ and $\mu = 0.10$, followed by a broad region of selection stability. This behavior supports the use of $\mu = 1$ as a nominal value within a stable regime rather than as a parameter tuned to obtain a particular subset.

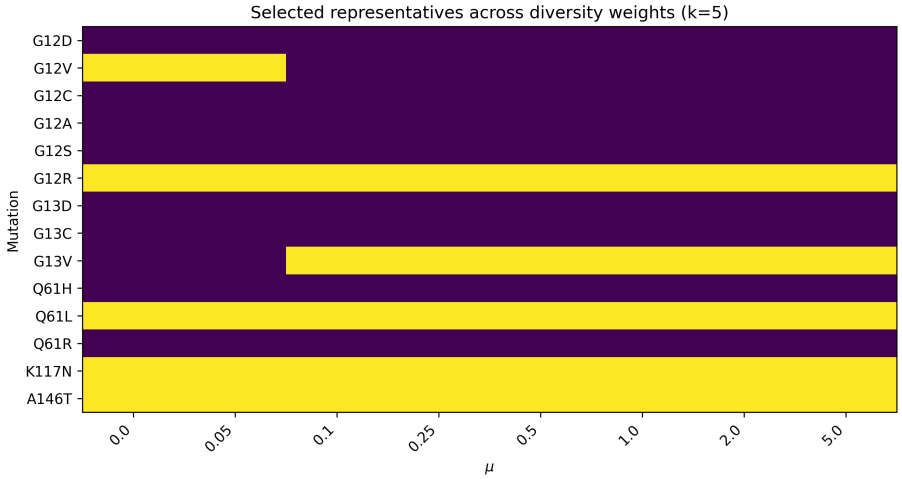


Figure 9. Representative mutations selected for $k = 5$ across the tested values of the diversity parameter μ. The magnitude-only composition is retained for $\mu \leq 0.05$, whereas the geometry-aware subset remains unchanged from $\mu = 0.10$ to 5.

To assess whether the observed behavior is specific to the main cardinality, the exact analysis was repeated for $k = 4$ and $k = 6$. For $k = 4$, the variational and perturbation-magnitude criteria coincide and select {G12R, Q61L, K117N, A146T}. For $k = 5$, the diversity term introduces G13V in place of G12V, as discussed above. For $k = 6$, the variational optimum is {G12R, G13V, Q61L, Q61R, K117N, A146T}, whereas the perturbation-magnitude criterion selects {G12D, G12V, G12R, Q61L, K117N, A146T}. The $k = 4$ result is particularly informative because it shows that the diversity term does not necessarily alter the subset when the magnitude-based solution is already compatible with the joint criterion. At larger cardinalities, however, the geometric contribution reduces the concentration of representatives within the G12 family and redistributes part of the subset toward other operator-space neighborhoods.

As an additional robustness check, we examined whether the representative subsets were driven primarily by differences in the local connectivity of the mutated residues. We therefore repeated the exact selection after normalizing the perturbation magnitude V_i by both the residue degree and the weighted degree of the corresponding site in the wild-type RIN. Under both normalization schemes, the variational optimum remained unchanged for $k = 4$, $k = 5$, and $k = 6$. In particular, the $k = 5$ subset remained {G12R, G13V, Q61L, K117N, A146T}. This diagnostic indicates that the selected composition is not solely a consequence of larger raw Hilbert–Schmidt magnitudes associated with more highly connected local neighborhoods. Finally, we examined whether the representative subset remained stable under the alternative RIN constructions introduced in Section 4.2. For $k = 5$ and $\mu = 1$, the 6 Å distance-weighted network and the binary 7 Å network recover the same subset as the main 7 Å distance-weighted representation. The denser 8 Å network instead selects G12V in place of G13V, yielding a Jaccard similarity of 0.667 with the main subset. Nevertheless, G12R, Q61L, K117N, and A146T are retained under all four tested RIN constructions. Figure 10 provides a direct visualization of this behavior, showing the selected mutations under each network definition and making the stable four-mutation core explicit. Thus, the representative selection exhibits a stable core while the marginal fifth representative is sensitive to the denser 8 Å construction.

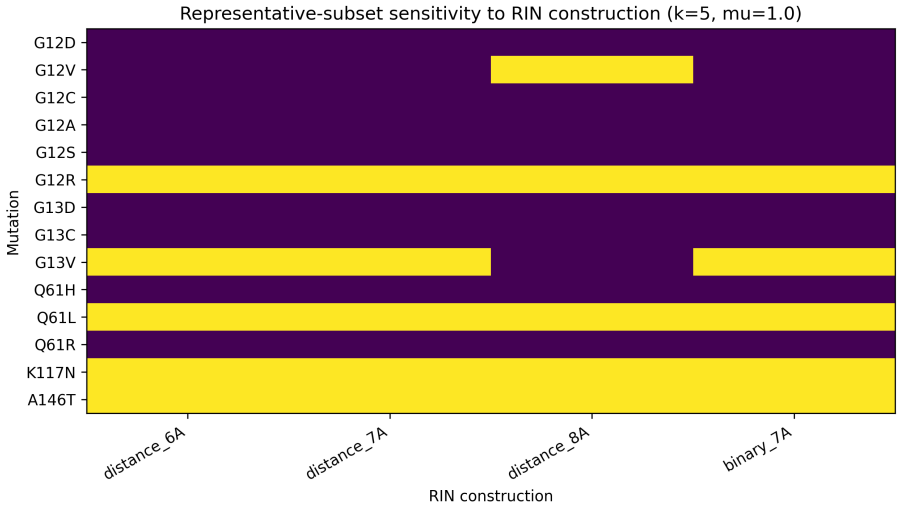


Figure 10. Sensitivity of the representative subset to the RIN construction for $k = 5$ and $\mu = 1$. Highlighted cells indicate selected mutations under each network definition. G12R, Q61L, K117N, and A146T are retained across all tested constructions, whereas the fifth representative changes from G13V to G12V for the denser 8 Å distance-weighted network.

This result complements the geometric sensitivity analysis of Section 4.2: the global operator geometry remains strongly correlated across RIN definitions, but the discrete optimum need not be strictly invariant when two candidate representatives compete near the selection boundary.

4.4. Numerical consistency of the geometry-induced Ising encoding

Sections 4.2 and 4.3 established the representative-selection problem directly from the Hilbert–Schmidt geometry of the mutation perturbation operators. Sections 3.2 and 3.3 subsequently showed that this problem admits exact QUBO and Ising reformulations. The purpose of the present subsection is not to provide independent empirical evidence that the Ising Hamiltonian preserves the operator geometry, since this property already follows analytically from the derived coupling relation

$$J_{ij} = \frac{\lambda}{2} - \frac{\mu}{4} D_{ij}, \qquad i \neq j, \tag{4.1}$$

but rather to verify numerically that the complete computational implementation reproduces the analytical identities up to floating-point precision. For fixed λ and $\mu > 0$, the squared Hilbert–Schmidt distances can be recovered directly from the Ising couplings as

$$\widehat{D}_{ij} = \frac{4}{\mu}\left(\frac{\lambda}{2} - J_{ij}\right). \tag{4.2}$$

We therefore perform two complementary consistency checks: first, we compare the exact minimizers of the variational, QUBO, and Ising formulations; second, we reconstruct the complete operator-distance geometry from the Ising coupling matrix and compare it numerically with the original geometry.

The first check was performed for all three cardinalities considered in the robustness analysis, $k = 4$, 5, and 6. In every case, exact enumeration of the variational problem, the penalized QUBO, and the corresponding Ising Hamiltonian recovered the same representative subset and the same minimum objective value up to numerical precision. The maximum QUBO–Ising energy discrepancy observed across these tests was 1.86×10^{-9}. This agreement provides a direct numerical verification of the exact reformulations derived in Section 3.

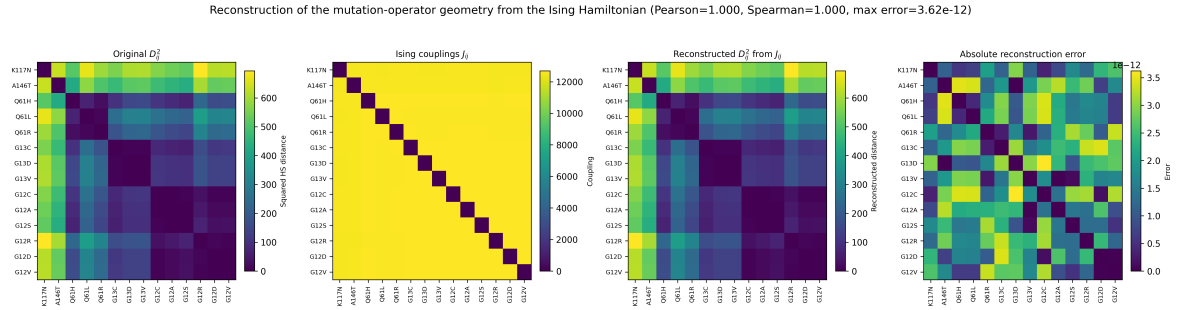


Figure 11. Numerical consistency check of the analytical reconstruction of the mutation-operator geometry from the Ising Hamiltonian. Left: Squared Hilbert–Schmidt distance matrix D_{ij} computed from the mutation perturbation operators. Middle-left: Ising coupling matrix obtained from the QUBO formulation. Middle-right: Squared-distance matrix $\widehat{D}_{ij}$ reconstructed from the Ising couplings using the analytical affine relation. Right: Absolute reconstruction error. The reconstructed matrix agrees with the original one up to numerical precision, with Pearson and Spearman correlations equal to 1.000 and a maximum absolute error of 3.62×10^{-12}.

Figure 11 shows the second consistency check. The original squared-distance matrix D and the reconstructed matrix $\widehat{D}$ are visually indistinguishable, while their difference is restricted to floating-point errors of order 10^{-12}. The Pearson and Spearman correlations between the two matrices are both 1.000, and the maximum absolute reconstruction error is 3.62×10^{-12}. These numerical results are therefore consistent with the exact affine relation between J_{ij} and D_{ij} derived in Section 3.3. They should be interpreted as validation of the implementation of that relation rather than as an independent empirical demonstration of geometry preservation.

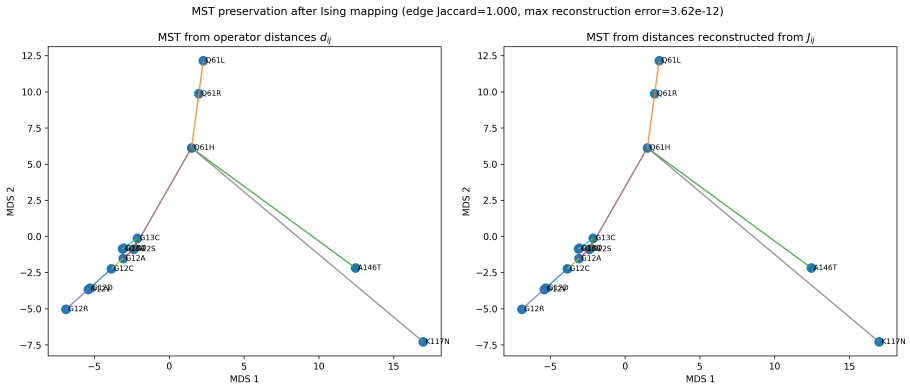


Figure 12. Minimum spanning trees obtained from the original Hilbert–Schmidt distances d_{ij} (left) and from the distances reconstructed from the Ising couplings (right). The two trees contain the same 13 edges, yielding an edge Jaccard similarity of 1.000.

As a derived diagnostic of the reconstructed distances, we also compared the corresponding minimum spanning trees. Figure 12 shows that both trees contain exactly the same 13 edges, giving an edge Jaccard similarity of 1.000. This result is expected once the complete pairwise distance matrix has been reconstructed to numerical precision, but it provides an intuitive check that the implementation also reproduces the neighborhood relations induced by the original operator-space distances.

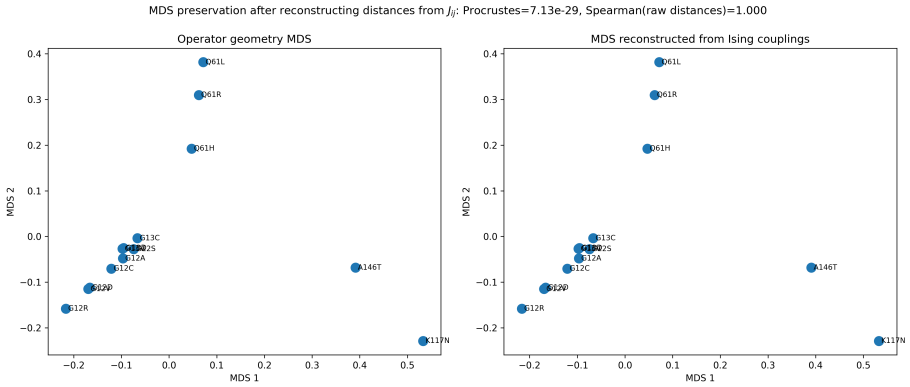


Figure 13. Comparison between the two-dimensional MDS embedding obtained from the original Hilbert–Schmidt distances (left) and the corresponding embedding obtained from distances reconstructed from the Ising couplings (right). The embeddings agree up to numerical precision, with a Procrustes disparity of 7.13×10^{-29} and Pearson and Spearman correlations of the embedding distances equal to 1.000.

A complementary global visualization is provided by the MDS embeddings in Figure 13. Applying the same embedding procedure to the original and reconstructed Hilbert–Schmidt distance matrices produces geometrically equivalent configurations up to numerical precision. The Procrustes disparity is 7.13×10^{-29}, while both the Pearson and Spearman correlations between the corresponding embedding distances are 1.000. As for the MST comparison, this agreement is a consequence of the accurate reconstruction of the underlying pairwise distances and is reported as an additional numerical consistency diagnostic rather than as independent evidence for the analytical mapping. The principal quantitative results are summarized in Table 3.

Table 3. Numerical consistency checks of the variational, QUBO, and Ising formulations and of the reconstruction of the operator-space geometry from the Ising couplings.

Metric	Value
Variational/QUBO/Ising minimizer agreement ($k = 4, 5, 6$)	Yes
Maximum QUBO–Ising energy mismatch	1.86×10^{-9}
Pearson correlation (D_{ij}, reconstructed $\widehat{D}_{ij}$)	1.000
Spearman correlation (D_{ij}, reconstructed $\widehat{D}_{ij}$)	1.000
Pearson correlation (d_{ij}, reconstructed $\widehat{d}_{ij}$)	1.000
Spearman correlation (d_{ij}, reconstructed $\widehat{d}_{ij}$)	1.000
Maximum reconstruction error in D_{ij}	3.62×10^{-12}
MST edge Jaccard similarity	1.000
Procrustes disparity	7.13×10^{-29}

Taken together, these checks verify the numerical implementation of the complete reformulation chain. The constrained variational problem, its penalized QUBO representation, and the corresponding Ising Hamiltonian recover the same exact solutions for the tested cardinalities, while the Ising couplings reconstruct the squared operator distances according to the analytical relation derived in Section 3.3. The remaining numerical discrepancies are at floating-point precision. Consequently, the role of these experiments is to confirm computational consistency between the theoretical objects introduced in Section 3, while the evaluation of classical and quantum optimization backends for the resulting objective is considered separately in the following subsection.

4.5. Performance of classical and gate-based optimization backends

After verifying the exact variational–QUBO–Ising reformulation in the previous subsection, we finally evaluate alternative computational backends for searching the resulting objective. The purpose of this experiment is not to define alternative representative-selection models, since the objective and target cardinality have already been fixed by the operator geometry, but to characterize how a classical heuristic and a gate-based variational quantum ansatz behave on the same optimization problem. Exact enumeration of the penalized objective over all binary configurations is used to establish the reference global optimum, which is verified to coincide with the minimum in the fixed-cardinality $k = 5$ subspace. Simulated annealing is then restricted to this feasible subspace through cardinality-preserving swap moves, while the quantum experiment uses the Dicke-state, cardinality-preserving XY-QAOA construction described in Section 3.4. Thus, simulated annealing and XY-QAOA operate on the same fixed-cardinality objective whose exact optimum is known independently.

Table 4. Summary of the optimization-backend evaluation for the $k = 5$ representative-selection problem. Exact enumeration establishes the reference optimum of the penalized objective. Simulated annealing statistics are computed over 100 independent restarts, whereas XY-QAOA is evaluated over circuit depths $p = 1, \ldots, 5$.

Backend	Diagnostic	Result
Exact enumeration	Reference optimum	−6574.992
Fixed-k simulated annealing	Successful restarts	100/100
Fixed-k simulated annealing	Median / mean gap	0/0
XY-QAOA	Relative expected-energy gap, $p = 1$	38.05%
XY-QAOA	Relative expected-energy gap, $p = 5$	22.45%
XY ring mixer	Reachable feasible states	2002/2002

The exact global minimum of the penalized objective is −6574.992, and it coincides with the minimum obtained after restricting the search to the $k = 5$ feasible subspace. The corresponding configuration is therefore the exact representative subset identified in Section 4.3. Fixed-cardinality simulated annealing recovered this optimum in all 100 independent restarts using 1000 swap steps per restart. Both the median and mean optimization gaps are consequently zero for this instance. This result shows that the classical heuristic robustly recovers the known optimum for the present small problem under the adopted numerical settings. It should not, however, be interpreted as evidence of polynomial-time performance or of comparable behavior for substantially larger mutation panels. The gate-based experiment was carried out through state-vector simulation of the cardinality-preserving XY-QAOA ansatz for depths $p = 1, \ldots, 5$. The initial state is the uniform Dicke state over the $\binom{14}{5} = 2002$ feasible bit strings, and the local XY ring mixer preserves the Hamming weight throughout the evolution. As an implementation check, the mixer graph was verified to be connected over the complete fixed-cardinality basis, with all 2002 feasible configurations reachable. The Dicke state is initialized exactly in the state-vector simulation; the additional circuit cost required to prepare this state on quantum hardware is not included in the present experiment.

Figure 14 reports the expected-energy gap relative to the exact optimum. This quantity decreases from 2502.056 at $p = 1$ to 1475.848 at $p = 5$. In relative terms, the gap evolves as 38.05%, 31.71%, 30.62%, 25.83%, and 22.45% for $p = 1, \ldots, 5$, respectively, corresponding to an approximately 41% reduction between the shallowest and deepest circuits considered. At each depth, the variational parameters were optimized from 20 random initializations. For $p > 1$, an additional initialization was obtained by embedding the optimized parameters of the preceding depth and setting the new layer to the identity. Consequently, each deeper ansatz can reproduce the preceding optimized state. The monotonic decrease observed in the expected-energy gap should therefore be interpreted within this nested warm-start protocol rather than as evidence that QAOA performance generally improves monotonically with circuit depth.

Expected energy alone does not characterize the sampling behavior of the optimized variational state. We therefore additionally evaluate the single-shot probability P_{opt} assigned to the exact optimum and the corresponding number of independent samples required to obtain the optimum with at least 95% probability. These diagnostics are shown in Figure 15. The optimal-state probabilities for $p = 1, \ldots, 5$ are 0.006689, 0.008260, 0.010464, 0.008764, and 0.002868, respectively. Under independent sampling from the corresponding optimized states, these probabilities imply requirements of 447, 362, 285, 341,

and 1044 shots, respectively, to reach a 95% probability of observing the optimum at least once. Unlike the expected-energy gap, neither sampling diagnostic improves monotonically with circuit depth. In particular, the lowest expected-energy gap is obtained at $p = 5$, whereas the largest probability assigned to the exact optimum occurs at $p = 3$.

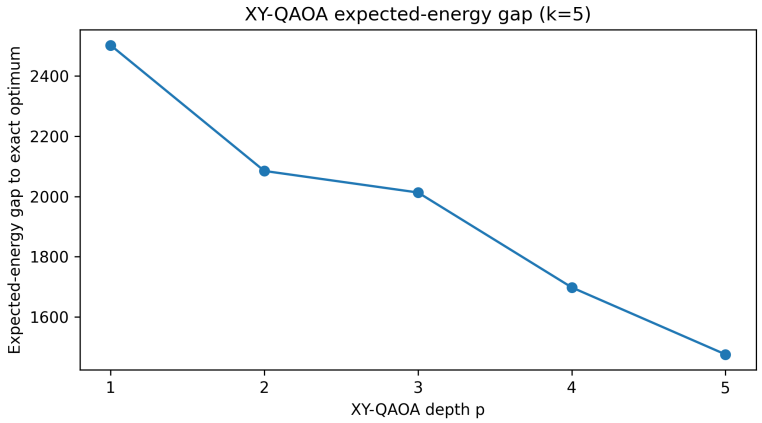


Figure 14. Expected-energy gap of the fixed-cardinality XY-QAOA state relative to the exact optimum for circuit depths $p = 1, \ldots, 5$. Under the nested warm-start optimization protocol, the gap decreases from 2502.056 at $p = 1$ to 1475.848 at $p = 5$, corresponding to an approximately 41% reduction.

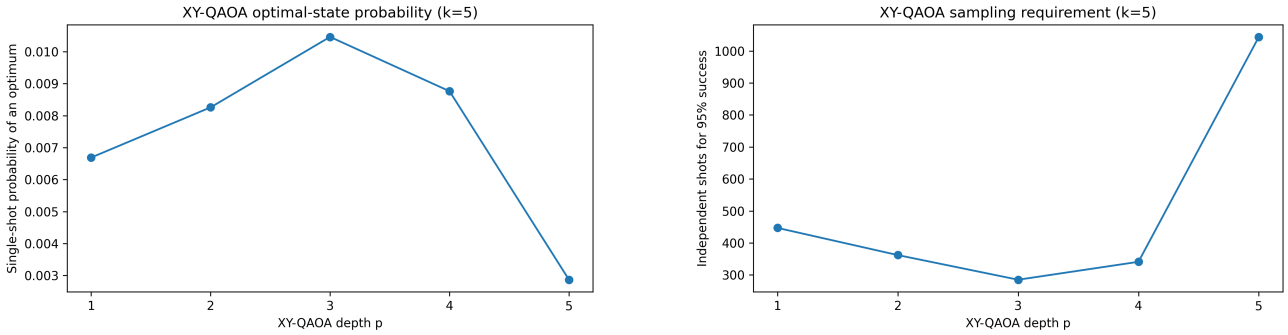


Figure 15. Sampling diagnostics for fixed-cardinality XY-QAOA as a function of circuit depth. Left: Single-shot probability P_{opt} assigned to the exact optimum. Right: The number of independent shots required to obtain the optimum with at least 95% probability. In contrast to the expected-energy gap, these sampling diagnostics are non-monotonic with depth, showing that improvement of the expectation value does not necessarily imply increased probability mass on the exact optimum.

The expectation-value and sampling diagnostics therefore capture complementary aspects of the variational state. Within the adopted nested optimization protocol, increasing depth improves the optimized expected objective value over the tested range, but it does not produce a corresponding monotonic improvement in exact-optimum sampling probability. Consequently, the present state-vector results should be interpreted as a backend-compatibility and depth-behavior study rather than as evidence of quantum advantage or of superior optimization performance relative to the classical

heuristic. For this small instance, fixed-cardinality simulated annealing already recovers the exact optimum in every tested restart. The relevance of the XY-QAOA experiment is instead that the same geometry-induced objective can be implemented within a gate-based, constraint-preserving variational architecture while maintaining feasibility throughout the quantum evolution. Assessing whether such an approach becomes competitive for larger instances, under realistic state-preparation, noise, sampling, and hardware constraints, remains outside the scope of the present study.

5. Conclusions

This work introduced a geometry-induced framework for representative mutation selection based on an operator representation of modeled mutation-induced perturbations. Starting from a common wild-type residue interaction network, each mutation is represented through the corresponding Laplacian perturbation operator, placing all candidates within a common symmetric operator space endowed with the Hilbert–Schmidt structure. The resulting perturbation magnitudes and pairwise operator distances provide the quantities required to define a fixed-cardinality variational selection problem that balances deviation from the wild-type reference with geometric diversity.

A central feature of the framework is the separation between model construction and optimization. Once the RIN representation, perturbation model, operator geometry, subset cardinality, and diversity weight have been specified, the representative-selection objective admits exact binary quadratic, QUBO, and Ising reformulations. The data-dependent terms of these formulations are inherited directly from the operator-derived quantities, while the cardinality contribution is introduced only to enforce feasibility in the unconstrained representation. In particular, the two-body Ising couplings are affine functions of the squared Hilbert–Schmidt distances between perturbation operators. The numerical reconstruction of these distances from the Ising couplings and the agreement of the variational, QUBO, and Ising minima should therefore be interpreted as consistency checks of the analytical reformulation rather than as independent empirical evidence of geometry preservation. The KRAS case study illustrates how this construction behaves in practice. The mutation perturbation operators form a structured finite geometry, and the variational criterion selects subsets that improve coverage of the induced operator configuration relative to perturbation-magnitude ranking alone. For the main $k = 5$ analysis, the variational subset replaces a geometrically redundant G12 representative with a G13 mutation, increasing the minimum separation between selected representatives and covering all five data-driven operator-space clusters while incurring only a small reduction in total perturbation magnitude. The selected composition remains stable over a broad interval of the diversity parameter and exhibits a stable core across alternative RIN constructions, although the fifth representative changes for the denser 8 Å network. Additional analyses across $k = 4, 5, 6$ and degree-normalized perturbation magnitudes further indicate that the observed representative sets are not determined solely by a single cardinality choice or by local network connectivity. The optimization experiments reinforce the conceptual separation between the representative-selection model and the procedure used to search it. Exact enumeration provides the reference optimum, fixed-cardinality simulated annealing recovers that optimum in all tested restarts for the present instance, and cardinality-preserving XY-QAOA demonstrates that the same induced objective can be implemented within a gate-based variational architecture while remaining entirely in the feasible subspace. The QAOA results should be interpreted as a backend-compatibility and depth-behavior study rather than as evidence of quantum advantage: deeper circuits reduce the

optimized expected-energy gap under the adopted nested warm-start protocol, but the probability of sampling the exact optimum does not improve monotonically with depth.

The contribution of the present work is therefore not a new mutation-impact score or a claim that a unique optimization problem is determined independently of modeling choices. Rather, it provides a systematic route from a chosen operator representation of modeled mutation-induced perturbations to a representative-selection objective whose QUBO and Ising coefficients are determined exactly by the resulting geometric quantities. This separation between representation, model construction, and optimization may be useful beyond the specific KRAS example considered here and may extend to other graph-based representative-selection problems in which candidate objects admit a common operator description. Several limitations remain. The numerical study is restricted to a single protein system and to model-defined perturbations of a static wild-type RIN rather than mutant-resolved three-dimensional structures. The subset cardinality is prescribed externally, and the selected representatives retain some dependence on the adopted RIN construction. Moreover, the gate-based quantum experiments are performed through state-vector simulation and do not include the hardware cost of Dicke-state preparation, device noise, connectivity constraints, or finite-resource execution. Future work should therefore examine cross-protein generalization, richer and possibly ensemble-based interaction representations, adaptive treatment of subset cardinality, alternative operator geometries, and larger-scale classical and quantum optimization under realistic computational and hardware constraints.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The CITIC, as a center accredited for excellence within the Galician University System and a member of the CIGUS Network, receives subsidies from the Department of Education, Science, Universities, and Vocational Training of the Xunta de Galicia. Additionally, it is co-financed by the European Union through the FEDER Galicia 2021-27 operational program (Ref. ED431G 2023/01).

Conflict of interest

Álvaro Michelena, Esteban Jove and Jose Luis Calvo-Rolle are guest editors for the special issue of the ERA and was not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

References

1. M. Soskine, D. S. Tawfik, Mutational effects and the evolution of new protein functions, *Nat. Rev. Genet.*, **11** (2010), 572–582. https://doi.org/10.1038/nrg2808
2. N. Tokuriki, D. S. Tawfik, Protein dynamism and evolvability, *Science*, **324** (2009), 203–207. https://doi.org/10.1126/science.1169375

3. K. J. Karczewski, L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, Q. Wang, et al., The mutational constraint spectrum quantified from variation in 141,456 humans, *Nature*, **581** (2020), 434–443. https://doi.org/10.1038/s41586-020-2308-7
4. W. Xu, A. Li, Y. Zhao, Y. Peng, Decoding the effects of mutation on protein interactions using machine learning, *Biophys. Rev.*, **6** (2025), 011307. https://doi.org/10.1063/5.0249920
5. Q. Pan, T. B. Nguyen, D. B. Ascher, D. E. V. Pires, Systematic evaluation of computational tools to predict the effects of mutations on protein stability in the absence of experimental structures, *Briefings Bioinf.*, **23** (2022), bbac025. https://doi.org/10.1093/bib/bbac025
6. P. C. Ng, S. Henikoff, SIFT: Predicting amino acid changes that affect protein function, *Nucleic Acids Res.*, **31** (2003), 3812–3814. https://doi.org/10.1093/nar/gkg509
7. I. Adzhubei, D. M. Jordan, S. R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2, *Curr. Protoc. Hum. Genet.*, **76** (2013), 7.20.1–7.20.41. https://doi.org/10.1002/0471142905.hg0720s76
8. J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, The FoldX web server: an online force field, *Nucleic Acids Res.*, **33** (2005), W382–W388. https://doi.org/10.1093/nar/gki387
9. T. Preeprem, G. Gibson, SDS, a structural disruption score for assessment of missense variant deleteriousness, *Front. Genet.*, **5** (2014), 82. https://doi.org/10.3389/fgene.2014.00082
10. L. Di Paola, M. De Ruvo, P. Paci, D. Santoni, A. Giuliani, Protein contact networks: an emerging paradigm in chemistry, *Chem. Rev.*, **113** (2013), 1598–1613. https://doi.org/10.1021/cr3002356
11. Y. Liu, W. S. B. Yeung, P. C. N. Chiu, D. Cao, Computational approaches for predicting variant impact: An overview from resources, principles to applications, *Front. Genet.*, **13** (2022), 981005. https://doi.org/10.3389/fgene.2022.981005
12. M. J. Meyer, R. Lapcevic, A. E. Romero, M. Yoon, J. Das, J. F. Beltrán, et al., mutation3D: cancer gene prediction through atomic clustering of coding variants in the structural proteome, *Hum. Mutat.*, **37** (2016), 447–456. https://doi.org/10.1002/humu.22963
13. D. Yehorova, B. Di Geronimo, M. Robinson, P. M. Kasson, S. C. L. Kamerlin, Using residue interaction networks to understand protein function and evolution and to engineer new proteins, *Curr. Opin. Struct. Biol.*, **89** (2024), 102922. https://doi.org/10.1016/j.sbi.2024.102922
14. M. W. Libbrecht, J. A. Bilmes, W. S. Noble, Choosing non-redundant representative subsets of protein sequence data sets using submodular optimization, *Proteins Struct. Funct. Bioinf.*, **86** (2018), 454–466. https://doi.org/10.1002/prot.25461
15. D. De Santis, S. Tirone, S. Marmi, V. Giovannetti, Optimized QUBO formulation methods for quantum computing, *Quantum Sci. Technol.*, **11** (2026), 015056. https://doi.org/10.1088/2058-9565/ae3b70
16. A. Abbas, A. Ambainis, B. Augustino, A. Bärtschi, H. Buhrman, C. Coffrin, et al., Challenges and opportunities in quantum optimization, *Nat. Rev. Phys.*, **6** (2024), 718–735. https://doi.org/10.1038/s42254-024-00770-9
17. F. A. Quinton, P. A. S. Myhr, M. Barani, P. C. del Granado, H. Zhang, Quantum annealing applications, challenges and limitations for optimisation problems compared to classical solvers, *Sci. Rep.*, **15** (2025), 12733. https://doi.org/10.1038/s41598-025-96220-2

18. S. Yarkoni, E. Raponi, T. Bäck, S. Schmitt, Quantum annealing for industry applications: introduction and review, *Rep. Prog. Phys.*, **85** (2022), 104001. https://doi.org/10.1088/1361-6633/ac8c54
19. K. Blekos, D. Brand, A. Ceschini, C. H. Chou, R. H. Li, K. Pandya, et al., A review on quantum approximate optimization algorithm and its variants, *Phys. Rep.*, **1068** (2024), 1–66. https://doi.org/10.1016/j.physrep.2024.03.002
20. B. Simon, *Operator Theory: A Comprehensive Course in Analysis, Part 4*, American Mathematical Society, 2015. https://doi.org/10.1090/simon/004
21. S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi, Optimization by simulated annealing, *Science*, **220** (1983), 671–680. https://doi.org/10.1126/science.220.4598.671
22. W. Liu, J. Cheng, Y. Ma, X. Li, X. Li, Quantum alternating operator ansatz with block-ring mixer and hot-start strategy for graph edit distance problem, *Quantum Mach. Intell.*, **8** (2026), 91. https://doi.org/10.1007/s42484-026-00436-y
23. E. Maiorino, A. Rizzi, A. Sadeghian, A. Giuliani, Spectral reconstruction of protein contact networks, *Physica A*, **471** (2017), 804–817. https://doi.org/10.1016/j.physa.2016.12.046
24. D. A. Spielman, Spectral graph theory, in *Combinatorial Scientific Computing*, CRC Press, (2012), 495–524.
25. F. R. K. Chung, *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics, American Mathematical Society, **92** (1997). https://doi.org/10.1090/cbms/092
26. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.*, **157** (1982), 105–132. https://doi.org/10.1016/0022-2836(82)90515-0
27. T. S. Lee, S. J. Potts, M. J. McGinniss, C. M. Strom, Multiple property tolerance analysis for the evaluation of missense mutations, *Evol. Bioinf.*, **2** (2006). https://doi.org/10.1177/117693430600200019
28. S. S. Li, Z. M. Liu, J. Li, Y. B. Ma, Z. Y. Dong, J. W. Hou, et al., Prediction of mutation-induced protein stability changes based on the geometric representations learned by a self-supervised method, *BMC Bioinf.*, **25** (2024), 282. https://doi.org/10.1186/s12859-024-05876-6
29. R. A. Horn, C. R. Johnson, *Matrix Analysis*, 2^{nd} edition, Cambridge University Press, Cambridge, 2012. https://doi.org/10.1017/CBO9781139020411
30. G. Yuan, B. Ghanem, An exact penalty method for binary optimization based on MPEC formulation, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **31** (2017), 2867–2875. https://doi.org/10.1609/aaai.v31i1.10795
31. F. Glover, G. Kochenberger, R. Hennig, Y. Du, Quantum bridge analytics I: a tutorial on formulating and using QUBO models, *Ann. Oper. Res.*, **314** (2022), 141–183. https://doi.org/10.1007/s10479-022-04634-2
32. A. Lucas, Ising formulations of many NP problems, *Front. Phys.*, **2** (2014), 5. https://doi.org/10.3389/fphy.2014.00005
33. L. Nagpal, A. Kumar, S. R. Hassan, Mapping of fermionic lattice models for Ising solvers, *Sci. Rep.*, **16** (2026), 16249. https://doi.org/10.1038/s41598-026-44886-7
34. A. Glos, A. Krawiec, Z. Zimborás, Space-efficient binary optimization for variational quantum computing, *npj Quantum Inf.*, **8** (2022), 39. https://doi.org/10.1038/s41534-022-00546-y

35. C. C. McGeoch, *Adiabatic Quantum Computation and Quantum Annealing: Theory and Practice*, Springer Nature, Cham, 2014. https://doi.org/10.1007/978-3-031-02518-1

36. S. Morita, H. Nishimori, Mathematical foundation of quantum annealing, *J. Math. Phys.*, **49** (2008), 125210. https://doi.org/10.1063/1.2995837

37. J. C. Hunter, D. Gurbani, S. B. Ficarro, M. A. Carrasco, S. M. Lim, H. G. Choi, et al., In situ selectivity profiling and crystal structure of SML-8-73-1, an active site inhibitor of oncogenic K-Ras G12C, *Proc. Natl. Acad. Sci. U.S.A.*, **111** (2014), 8895–8900. https://doi.org/10.1073/pnas.1404639111

38. K. M. Haigis, KRAS alleles: the devil is in the detail, *Trends Cancer*, **3** (2017), 686–697. https://doi.org/10.1016/j.trecan.2017.08.006



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)