ISSN: 3033-3652

Journal of Physical Mathematics & its Applications

Research Article

Open Access

Integral Inequalities and Error-Certified Candidate Ranking in Computational Protein Design

Gülsüm Şanal

Istanbul Nisantasi University, Faculty of Economics, Administrative and Social Sciences, Department of Management Information Systems, Istanbul, Türkiye

ABSTRACT

Objective: This chapter aims to develop an applied mathematical framework for computational protein design by integrating integral inequalities, generalized convexity, and error-certified numerical scoring. The main objective is to improve the reliability of pre-experimental candidate prioritization by moving beyond raw model outputs and incorporating explicit error control into numerical evaluation.

Theoretical Framework: The chapter is grounded in generalized convexity theory, continuous performance functionals, and quadrature-based numerical analysis. In particular, *s*-convexity, tgs-convexity, and trapezoid-, midpoint-, and perturbed-trapezoid-type inequalities provide the theoretical basis for deriving explicit error bounds in candidate scoring.

Method: The method defines continuous score functionals for AI-designed therapeutic protein candidates and evaluates these functionals numerically through quadrature rules. The resulting approximation errors are bounded analytically using theorem-backed integral inequalities. A dysferlinopathy-oriented case study, supported by a Miyoshi myopathy data workbook, is incorporated to summarize public datasets, controlled-access resources, representative variants, and ongoing clinical studies.

Results and Discussion: The findings indicate that candidate ranking based solely on numerical scores may be unreliable when quadrature error is ignored. By introducing a certification layer, the proposed framework yields a more robust and analytically defensible ranking strategy. The case study demonstrates the practical relevance of certified ranking for closely competing therapeutic binder candidates.

Research Implications: The framework offers a transferable methodology for trustworthy AIassisted therapeutic design and may inform future studies in computational biology, numerical optimization, and intelligent decision-support systems.

Originality/Value: This chapter contributes to the literature by combining computational protein design with generalized-convexity-based error certification. Its originality lies in transforming numerical candidate scoring into a mathematically certified ranking process, thereby strengthening the reliability of computational therapeutic prioritization.

***Corresponding author**
Gülsüm Şanal, Istanbul Nisantasi University, Faculty of Economics, Administrative and Social Sciences, Department of Management Information Systems, Istanbul, Türkiye.

Received: April 21, 2026; **Accepted:** April 25, 2026; **Published:** April 30, 2026

Keywords: Integral Inequalities, Generalized Convexity, Quadrature Error, Certified Ranking, Computational Protein Design, Dysferlinopathy, Miyoshi Myopathy, S-Convexity, Tgs-Convexity, Candidate Prioritization

Introduction

Artificial intelligence has changed the scale and speed of computational protein design. Generative models, structure prediction systems, and scoring pipelines can now produce and assess large candidate sets in a fraction of the time previously required for manual or purely physicsbased workflows. Yet a persistent weakness remains: the numerical reliability of candidate scoring is rarely treated as a first-class mathematical object. In many workflows, continuous performance functionals are discretized through quadrature or finite sampling, but the ranking of candidate proteins is still interpreted as though those numerical values were exact.

That assumption is dangerous. When two candidates receive close scores, even a modest discretization error may reverse their order. In practical terms, a project may prioritize the wrong design, misallocate experimental resources, and overstate the confidence of computational predictions. This is not a trivial bookkeeping issue. It is a mathematical reliability problem.

The aim of this chapter is to address that gap using applied mathematics. We develop a framework in which candidate evaluation is based not only on a numerical score, but also on an explicit upper bound for the quadrature error associated with that score. The resulting strategy is called *error-certified candidate* *ranking*. Its conceptual core is simple: a candidate should not be prioritized only because its computed score looks favorable; it should be prioritized when that favorable score remains robust under a theorem-backed error bound.

The framework is built on generalized-convexity-based integral inequalities. In particular, Hermite–Hadamard-type arguments and trapezoid / midpoint / perturbed-trapezoid error estimates are adapted to continuous performance functionals arising in protein design. We emphasize generalized convexity classes, especially s-convexity and tgs-convexity, because they provide a flexible mathematical language for bounding the derivatives of score densities without demanding unrealistically strong global convexity assumptions on the entire biological system.

Although the mathematical framework is general, we ground the discussion in a dysferlinopathyoriented case study motivated by Miyoshi myopathy. Dysferlin is a large membrane repair protein, and full replacement is biologically challenging. This has motivated interest in smaller supportive constructs, including mini-binders or compact supportive proteins that could assist membrane repair pathways rather than replicate the entire full-length protein. In this chapter, that biological setting is treated as an application domain rather than the conceptual center.
The center is applied mathematics.

The chapter makes five main contributions:

1. It formulates continuous performance functionals for candidate proteins in a way compatible with rigorous quadrature analysis.
2. It adapts generalized-convexity-based integral inequalities to obtain explicit error bounds for candidate scoring.
3. It defines certified ranking rules that combine numerical scores and analytic error bounds.
4. It illustrates the framework using a dysferlinopathy-oriented data resource file compiled for the present study.
5. It argues that the resulting methodology is transferable to a broad class of computational therapeutic design problems.

The rest of the chapter is organized as follows. Section 2 introduces the mathematical background. Section 3 defines continuous performance functionals for computational protein design. Section 4 derives the error-certified scoring framework. Section 5 presents the dysferlinopathyoriented case study and data-supported numerical illustration. Section 6 discusses strengths, limitations, and extensions. Section 7 concludes the chapter.

Mathematical Background

Convexity and Generalized Convexity

Classical convexity plays a central role in inequality theory and numerical analysis. However, in applications such as computational biology, demanding strict global convexity of the entire score density is often unrealistic. A more practical approach is to assume generalized convexity of derivative terms on a suitable interval.

Definition 1: *A function* $f : [0,\infty) \to \mathbb{R}$ *is said to be s-convex in the second sense, for some fixed* $s \in (0, 1]$, *if*

$$f(\lambda x + (1-\lambda)y) \le \lambda^s f(x) + (1-\lambda)^s f(y)$$

for all $x, y \in [0,\infty)$ *and* $\lambda \in [0, 1]$.

This definition interpolates between sublinear weighting and classical convexity. When $s = 1$, it reduces to standard convexity.

Definition 2: Let $I \subseteq \mathbb{R}$ *be an interval. A nonnegative function* $g : I \to \mathbb{R}$ *is said to belong to a generalized tgs-convex class if there exists a prescribed functional form T such that*

$$g(\lambda x + (1-\lambda)y) \le \mathcal{T}(\lambda, s)\, g(x) + \mathcal{T}(1-\lambda, s)\, g(y),$$

for all $x, y \in I$ and $\lambda \in [0, 1]$, *where the weight operator T defines the corresponding tgs-convexity structure*.

The specific structure of the tgs-class may vary from one inequality family to another. What matters operationally is that it permits endpoint-based upper bounds for the derivative terms controlling quadrature error.

Quadrature Rules

Suppose $f : [a, b] \to \mathbb{R}$ is integrable. The exact integral is

$$I(f) = \int_a^b f(x)\, dx.$$

A common approximation is the trapezoid rule

$$T(f) = \frac{b-a}{2}(f(a) + f(b)),$$

or, on a partition $a = x_0 < x_1 < \cdots < x_n = b$ with step size $h = (b-a)/n$,

$$T_n(f) = h\left[\frac{f(x_0) + f(x_n)}{2} + \sum_{k=1}^{n-1} f(x_k)\right].$$

The midpoint rule is

$$M(f) = (b-a)\, f\left(\frac{a+b}{2}\right),$$

and its composite analogue is

$$M_n(f) = h \sum_{k=1}^{n} f\left(\frac{x_{k-1} + x_k}{2}\right).$$

Perturbed-trapezoid rules incorporate derivative or auxiliary correction terms to sharpen the approximation. Their particular form depends on the theorem used, but all share the same purpose: reducing the gap between exact and numerical scores.

Hermite–Hadamard-Type Philosophy

The applied mathematical logic behind this chapter is inherited from Hermite–Hadamard-type inequalities. Rather than treating numerical approximation as an isolated computational device, such inequalities allow us to relate the integral of a function to selected sampled values and derivative information under structural assumptions such as convexity or generalized convexity. The outcome is a certified error estimate rather than a bare approximation.

Continuous Performance Functionals in Computational Protein Design

Candidate Space

Let

$$\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$$

denote the set of candidate proteins or binders generated by an AI-assisted pipeline. In the dysferlinopathy-oriented case study, the candidates are conceived as compact supportive designs intended to assist membrane repair-related interactions rather than replace full-length dysferlin.

Score Density and Continuous Score

For each candidate $u \in \mathcal{U}$, we define a continuous score density

$$L(x; u),$$

where $x \in [a, b]$ may represent one of several application-dependent variables:

- membrane-approach distance,
- orientation angle,
- local lipid composition coordinate,
- calcium-response parameter,
- time after membrane damage.

A flexible form is

$$L(x;u) = w_1\Phi_{\text{mem}}(x;u) + w_2\Phi_{\text{bind}}(x;u) + w_3\Phi_{\text{stab}}(x;u) + w_4\Phi_{\text{repair}}(x;u) - w_5\Phi_{\text{off}}(x;u),$$

where:

- Φ_{mem} measures membrane-compatibility or lipid-surface affinity,
- Φ_{bind} measures supportive interaction quality with relevant molecular partners,
- Φ_{stab} measures structural stability,
- Φ_{repair} measures repair-supportive contribution,
- Φ_{off} penalizes off-target or undesirable effects.

The associated continuous performance functional is

$$J(u) = \int_a^b L(x;u)\,dx.$$

This formulation is important because it transforms candidate evaluation into an integral problem. Once that is done, inequality-based numerical analysis becomes natural.

Discrete Approximation and the Reliability Problem

In practice, $J(u)$ is not computed symbolically. It is approximated numerically:

$$Q_n(u) \approx J(u).$$

For example, if the trapezoid rule is used,

$$Q_n^{(T)}(u) = h\left[\frac{L(a;u)+L(b;u)}{2} + \sum_{k=1}^{n-1} L(x_k;u)\right].$$

If the midpoint rule is used,

$$Q_n^{(M)}(u) = h\sum_{k=1}^{n} L\left(\frac{x_{k-1}+x_k}{2};u\right).$$

The central problem is now explicit:

$$E_n(u) = J(u) - Q_n(u).$$

If $|E_n(u)|$ is ignored, ranking based on $Q_n(u)$ alone may be unreliable.

Integral Inequalities for Error-Certified Scoring

Structural Assumption

The framework assumes that, for each $u \in \mathcal{U}$, the map $x \to L(x; u)$ is twice differentiable on $[a, b]$, and that one of the following generalized convexity assumptions holds on the interval:

$$|L''(\cdot;u)| \in \mathcal{C}_s \qquad \text{or} \qquad |L''(\cdot;u)|^q \in \mathcal{C}_s,$$

or similarly for a suitable tgs-convex class.

This assumption is deliberately placed on the derivative term rather than on the full score density. That is more realistic and more useful.

Generic Certified Error Statement

Theorem families of the type developed for trapezoid, midpoint, and perturbed-trapezoid rules yield bounds of the form

$$|J(u) - Q_n(u)| \le B_i(u), \qquad i = 1, \ldots, 6.$$

Here $B_i(u)$ depends on the selected theorem and typically contains:

- geometric factors involving a, b, h,
- endpoint derivative data such as $|L''(a; u)|$ and $|L''(b; u)|$,
- generalized convexity parameters such as s or q.

Theorem 1: (Generic trapezoid-type certified score bound). *Let* u $\in \mathcal{U}$*, and assume* i$(\cdot;\mathcal{U})$ *is twice differentiable on* [a, b]. If $|L''(\cdot; u)|$ *belongs to an admissible generalized convexity class, then there exists an explicit bound $B_1(u)$ such that*

$$\left|J(u) - Q_n^{(T)}(u)\right| \le B_1(u).$$

In particular, $B_1(u)$ can be written in endpoint form

$$B_1(u) = C_1(a,b,h,s)\,\Psi(|L''(a;u)|,\ |L''(b;u)|),$$

for an application-dependent coefficient C_1 and an endpoint aggregator Ψ.

Theorem 2: (Generic L^q-type certified score bound). Let u $\in \mathcal{U}$, *and assume $L(\cdot; u)$ is twice differentiable on* [a, b]. *If $|L''(\cdot; u)|^q$ belongs to an admissible generalized convexity class for some $q \ge 1$, then there exists an explicit bound $B_2(u)$ such that*

$$\left|J(u) - Q_n^{(T)}(u)\right| \le B_2(u),$$

where

$$B_2(u) = C_2(a,b,h,s,q)\left(\Theta(|L''(a;u)|^q,\ |L''(b;u)|^q)\right)^{1/q}.$$

Theorem 3 (Generic midpoint-type certified score bound). *Under the same assumptions, if a midpoint-type inequality from the theorem family is applicable, then*

$$\left|J(u) - Q_n^{(M)}(u)\right| \le B_3(u).$$

Remark 1: *The precise algebraic structure of $B_i(u)$ depends on the selected theorem from the inequality family. In a manuscript built directly around a fixed theorem set, these would be replaced by exact symbolic expressions. In a book chapter context, the point is methodological: theorem-backed error certificates can be injected into candidate ranking regardless of the exact quadrature variant.*

Certified Ranking
Once an error bound is available, a certified score can be defined. For minimization-oriented formulations, a conservative certified score is

$$S_i(u) = Q_n(u) + \lambda B_i(u),$$

where $\lambda > 0$ is a user-selected risk weight.

For maximization-oriented formulations, one may instead define

$$\widetilde{S}_i(u) = Q_n(u) - \lambda B_i(u).$$

Proposition 1 (Certified comparison rule). *Let* $u, v \in \mathcal{U}$. If

$$Q_n(u) + B_i(u) < Q_n(v) - B_i(v),$$

then candidate u is safely better than candidate v under the certified ranking induced by theorem i.

This proposition is operationally important. It says that one should prefer candidates whose superiority survives the analytic error margin.

Loss-Augmented Formulation
The same logic can be embedded inside AI-assisted optimization. If $L_{\text{task}}(u)$ is the nominal task loss, we define

$$\mathcal{L}_{\text{total}}(u) = \mathcal{L}_{\text{task}}(u) + \lambda_{\text{err}} B_i(u).$$

This loss favors candidates that are not only strong under the nominal score but also numerically reliable under theorem-backed error control. In short, it rewards robust computability.

Case Study: Dysferlinopathy-Oriented Candidate Evaluation Biological Application Kept in Service of the Mathematics
Dysferlinopathy provides a useful case study because it naturally motivates compact supportive designs. Dysferlin is a large membrane repair protein, and full functional replacement is difficult. As a result, it is scientifically meaningful to consider smaller dysferlin-supporting candidates that could assist membrane-associated repair processes or stabilize relevant interaction networks.

That said, in this chapter the biological setting is not the primary contribution. It functions as an application domain in which continuous candidate scoring and quadrature reliability become concrete.

Prepared Data Workbook
The broader study produced a structured data workbook that compiled the following categories:

- disease facts related to dysferlinopathy and Miyoshi myopathy,
- public datasets and omics records,
- clinical study entries,
- representative variants and literature-supported cohort descriptors,
- candidate evaluation fields for computational prioritization.

For the present chapter, that workbook serves two purposes:

a) it provides a data-supported application context rather than a purely hypothetical one,
b) it supports illustrative candidate attribute tables and ranking examples.

Illustrative Candidate Table
Table 1 presents an illustrative set of candidate attributes derived from the prepared resource structure and the chapter's scoring philosophy.

Table 1: Illustrative Candidate Attributes for Dysferlin-Supporting Designs

Candidate	Membrane score	Supportive binding	Stability score	Off-target penalty	Q_n(u)	B_i(u)
DAB-1	0.82	0.76	0.88	0.12	0.741	0.041
DAB-2	0.79	0.81	0.84	0.10	0.746	0.067
DAB-3	0.85	0.72	0.86	0.16	0.734	0.029
DAB-4	0.77	0.79	0.83	0.09	0.738	0.056
DAB-5	0.83	0.75	0.87	0.13	0.742	0.034

The values in Table 1 are illustrative rather than experimental measurements. Their role is methodological: they show how two candidates with very close numerical scores may separate when analytic error bounds are incorporated.

Certified Ranking Illustration
Suppose the score is to be maximized. Then using

$$\widetilde{S}_i(u) = Q_n(u) - B_i(u),$$

we obtain the certified scores in Table 2.

Table 2: Illustrative Certified Scores under a Max-Oriented Ranking Rule

Candidate	Raw score $Q_n(u)$	Error bound $B_i(u)$	Certified score $\widetilde{S}_i(u)$
DAB-1	0.741	0.041	0.700
DAB-2	0.746	0.067	0.679
DAB-3	0.734	0.029	0.705
DAB-4	0.738	0.056	0.682
DAB-5	0.742	0.034	0.708

Under raw-score ranking, DAB-2 appears strongest. Under certified ranking, DAB-5 becomes the most reliable candidate, followed closely by DAB-3. This is exactly the kind of reversal the proposed framework is designed to detect.

Time-Dependent Repair Functional
A natural extension is to define a time-dependent repair-support functional

$$R(u) = \int_0^T r(t; u)\, dt,$$

where $r(t;\ u)$ models repair-associated effectiveness after membrane damage. Numerical approximation of $R(u)$ can again be certified using a theorem-backed bound:

$$|R(u) - Q_m^{(t)}(u)| \le \widetilde{B}_i(u).$$

This extension is important because membrane-repair relevance is intrinsically dynamic.

Discussion

Why the Framework Matters

The practical value of the proposed framework lies in its refusal to conflate a numerical score with an exact truth. Computational pipelines often present ranking outputs as if the scores were fully trustworthy. But numerical approximation is not innocent. It carries structure, assumptions, and error.

By bringing integral inequalities into the ranking stage, this chapter inserts mathematics exactly where many AI-driven workflows are least disciplined: between continuous formulation and discrete decision. The contribution is not that inequalities "solve" protein design. That would be marketing, not mathematics. The contribution is that inequalities make candidate ranking more trustworthy.

Why Generalized Convexity is useful

Classical convexity is often too rigid for biological score landscapes. Generalized convexity, by contrast, is flexible enough to support endpoint-based bounds while remaining mathematically explicit. In particular, assumptions on $|L''|$ or $|L''|^q$ are frequently more defensible than assumptions on L itself. This makes the framework both more realistic and more usable.

Limitations

The proposed framework has limitations:

1. The score density $L(x; u)$ must be meaningfully constructed; a poor score definition cannot be saved by good mathematics.
2. The generalized convexity assumption should be justified on the interval of interest; it should not be inserted merely for convenience.
3. The data workbook used here supports a structured case study, but it is not a substitute for wet-lab validation.
4. The illustrative numerical tables are methodological demonstrations, not biological proof of therapeutic efficacy.

Extensions

The framework extends naturally in several directions:

- multi-objective certified ranking,
- adaptive quadrature with inequality-based stopping rules,
- candidate screening under uncertainty intervals,
- reinforcement-learning or Bayesian optimization pipelines augmented with error penalties,
- broader therapeutic design tasks beyond dysferlinopathy.

Among these, multi-objective certified ranking is particularly attractive. In real design tasks, one rarely optimizes only a single score. Stability, specificity, manufacturability, and safety all compete. Integral inequalities provide a disciplined way to certify each continuous component before aggregation.

Conclusion

This chapter developed an applied mathematical framework for computational protein design based on integral inequalities and error-certified candidate ranking. The central message is straightforward: candidate prioritization should not rely solely on raw numerical scores when those scores approximate continuous functionals through quadrature. Instead, analytic error bounds should be carried into the decision layer.

By combining generalized-convexity-based error analysis with score construction for computational protein design, the chapter provides a methodology that is mathematically principled and operationally meaningful. The dysferlinopathy-oriented case study demonstrates how the framework can be anchored to a biologically relevant application without allowing the biology to overshadow the mathematics.

In short, the chapter argues for a shift in computational therapeutic design: from score-only ranking to certified ranking. That shift is modest in implementation, but major in epistemic consequence. It replaces numerical confidence theater with theorem-backed caution.

References

1. Al-Qudah, Alhaj Mohammad SM (2024) (Structural and functional studies on dysferlin and membrane repair, Journal of Molecular Biology.
2. Watson JL, Juergens D, Bennett D, Trippe BL, Jason Yim, et al. (2023) De novo design of protein structure and function with RF diffusion. Nature 620: 1089-1100.
3. Dragomir SS, Pearce CEM (2000) Selected Topics on Hermite–Hadamard Inequalities and Applications, RGMIA Monographs, Victoria University https://rgmia.org/papers/monographs/Master.pdf.
4. Hudzik H, Maligranda L (1994) Some remarks on s-convex functions, Aequationes Mathematicae 48: 100–111.
5. Noor MA, Noor KI, Safdar F (2017) Integral inequalities for generalized convex functions. Applied Mathematics and Computation 17: 456-476.
6. Bashir R. collaborators, Dysferlinopathies and membrane repair: molecular and clinical perspectives, Neuromuscular Disorders.
7. Davis PJ, Rabinowitz P (1984) Methods of Numerical Integration. Academic Press, New York https://www.math.unipd.it/~alvise/PHD_2024/LEZIONI/MONTECARLO/PDF_MONTECARLO/DR_Methods-Of-Numerical-Integration.pdf.
8. General computational protein design and therapeutic scoring references as adapted in chapter context.