Enabling efficient experimental design in the context of high-dimensional generative models

Samuel Willis[1]* **Alexandru I. Stere**[4] **Dragos D. Margineantu**[5] **Henry T. Oldroyd**[2]

John A. Fozard[2] **Carl Henrik Ek**[1,3] **Henry Moss**[1, 2] **Erik Bodin**[1]*

[1]University of Cambridge [2]Lancaster University [3]Karolinska Institutet

[4]Boeing Commercial Airplanes [5]Boeing AI

Abstract

Many scientific optimisation tasks require finding designs consistent with physical or experimental constraints. Generative models such as diffusion models have proven to be powerful tools to propose designs, but remain challenging to control, often requiring significant model-specific procedures and compute to meet the constraints. We introduce surrogate latent spaces — Euclidean subspaces defined by examples — that enable standard optimisation algorithms to search for designs efficiently; including when the objective function is a black-box, non-differentiable, or expensive to evaluate, e.g. be evaluated through simulation or real-world experiments. We outline the underlying principles governing surrogate spaces, and demonstrate that the approach allows generating protein backbone designs using RFDIFFUSION with a sequence length that was previously infeasible.

1 Introduction

Generative models such as diffusion and flow matching have transformed the synthesis of high-dimensional data across domains, from images and video to molecules and protein structures [Ho et al., 2020, Song et al., 2020b, Lipman et al., 2022, Watson et al., 2023].

Yet, in many scientific applications, the goal is not merely to sample from a data distribution but to *optimise* for specific, measurable outcomes: a molecule's stability, a material's conductivity, or an aerodynamic coefficient. While these objectives often can be evaluated through simulation or experiment, they are typically non-differentiable, noisy, or expensive to compute. This creates a fundamental incompatibility between the differentiable conditioning mechanisms used in modern generative models and the black-box nature of many scientific optimisation tasks.

Recent work has explored steering generative models through prompt engineering or gradient-based guidance [Dhariwal and Nichol, 2021, Song et al., 2024, Fan et al., 2023], but these approaches presuppose access to well-behaved conditioning signals and are unsuitable when objectives arise from physical processes or empirical evaluation. Alternatively, *Latent Space Optimisation* (LSO) methods [Gómez-Bombarelli et al., 2018, Kusner et al., 2017, Moss et al., 2025] propose searching directly within the latent representation of a generative model. However, when applied to modern sample-based models such as diffusion or flow matching, direct optimisation in latent coordinates is impeded by the high dimensionality of the latent variable and the absence of guarantees that arbitrary perturbations remain within the model's support [Bodin et al., 2024].

*Equal contribution

European Conference on Neural Information Processing Systems (EurIPS) Workshop: SimBioChem 2025.

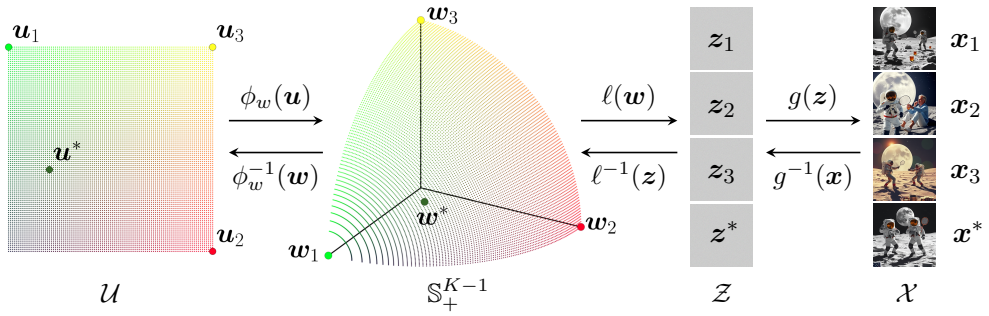


Figure 1: **Illustration of a surrogate latent space.** *(left)* A low-dimensional surrogate space $\mathcal{U}$ is mapped via the surrogate chart ϕ_w onto the positive orthant of the unit sphere *(mid left)*, which is then transported via ℓ to valid latent realisations $\boldsymbol{z} \sim p$ *(mid right)*, before finally being decoded by the generative model g into objects $\boldsymbol{x}$ *(right)*. All mappings are bijective, and enable optimisation algorithms to operate in $\mathcal{U}$ while guaranteeing validity, uniqueness, and approximate stationarity.

To address this, we propose to construct controllable, low-dimensional *surrogate latent spaces*, defined using examples. These spaces, which are Euclidean and bounded, are specified by a freely chosen number of examples and will contain a smoothly indexed set of similar solutions. As a result, the formed search spaces are easy to navigate using standard optimisation algorithms such as Bayesian Optimisation [Shahriari et al., 2015], CMA-ES [Hansen, 2016] and others, without retraining, fine-tuning, or gradient propagation. We demonstrate using an experiment on protein design that searching within a surrogate space is substantially more efficient than searching within the original latent space.

2 Surrogate latent spaces

We consider a deterministic generative model $g : \mathcal{Z} \to \mathcal{X}$ that maps latent variables $\boldsymbol{z} \sim p(\boldsymbol{z})$ to observable samples $\boldsymbol{x} \in \mathcal{X}$. Such deterministic mappings can for example be specified by a diffusion model — the probability-flow ODE [Song et al., 2020b] for continuous-time or DDIM [Song et al., 2020a] for discrete-time — or using by a flow matching model [Lipman et al., 2022]. The latent space $\mathcal{Z}$ is typically high-dimensional ($D \gg 10^4$), making direct optimisation intractable and unsafe: arbitrary perturbations of z can leave the model's support, yielding invalid or nonsensical generations [Bodin et al., 2024].

Our goal is to construct a *low-dimensional Euclidean space* $U = [0, 1]^{K-1}$ (where $K \ll D$) which can be optimised over efficiently, while guaranteeing that every point $\boldsymbol{u} \in U$ corresponds to a valid latent $\boldsymbol{z} \sim p(\boldsymbol{z})$, in turn meeting the requirements for valid generation $\boldsymbol{x} = g(\boldsymbol{z})$. The space is defined by K examples, specifically by their corresponding latent realisations $\{\boldsymbol{z}_k\}_{k=1}^K$.

2.1 Choosing the seeds

Let $\{\boldsymbol{z}_k\}_{k=1}^K$ be latent vectors corresponding to K designs; which we will refer to as *seeds*. Seeds can be chosen in a task-informed manner, for example, be obtained from inversions [Song et al., 2020b,a] of 'good' known designs (existing data from prior experiments scoring highly on some task), or be (filtered) high scoring solutions originally sampled from the latent distribution.

High scoring seeds tend to define spaces containing high scoring solutions, but importantly, *the solutions found within a surrogate space can be substantially higher scoring than the seeds defining it*; an example of this is shown in Figure 3. Therefore, even defining a surrogate space with *random* seeds is worthwhile if high scoring seeds are not available a-priori.

The number of seeds determines the dimensionality of the search space, and should be set with consideration to the optimisation algorithm used (which are designed for different dimensionalities) and the total evaluation budget; larger spaces typically contain a better *best* solution than smaller spaces, but can be more difficult to search within.

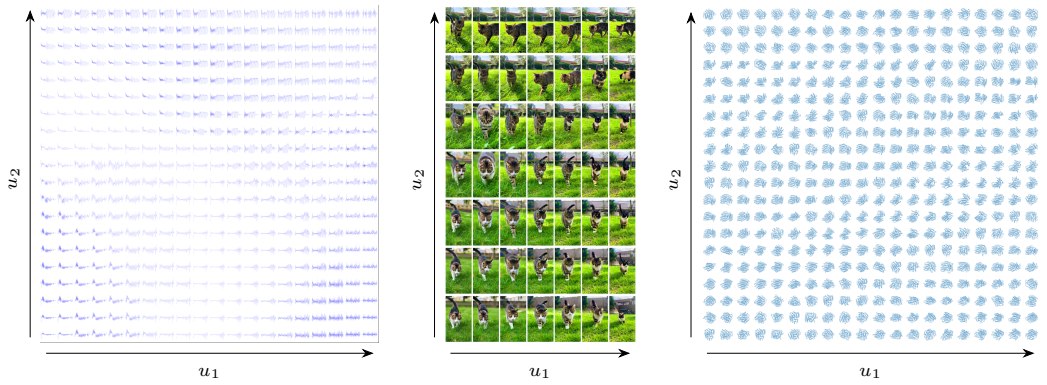


Figure 2: **We can build smooth surrogate spaces for any generative model.** (*left*) Waveform generations over a grid of a 2D slice of a 7D surrogate space formed from 8 seed latents and the 8256-dimensional StableAudio2.0 text-to-audio generation model [Evans et al., 2025]. (*middle*) The first frames of a similarly constructed grid of videos from the 4,308,480-dimensional HunyuanVideo text-to-video generation model [Kong et al., 2024]. (*right*) A grid of proteins over a 2D surrogate space formed from 3 seed latents corresponding to 3 proteins using RFDiffusion [Watson et al., 2023].

2.2 Surrogate space construction

We define a smooth bijection

$$\phi(\boldsymbol{u}; \{\boldsymbol{z}_k\}) = \ell(\phi_w(\boldsymbol{u}); \{\boldsymbol{z}_k\}), \qquad \boldsymbol{u} \in U = [0,1]^{K-1},$$

referred to as the *surrogate chart*, which is composed of two maps internally; a *weight chart* $\phi_w : U \to \mathbb{S}_+^{K-1}$ which maps Euclidean coordinates to the positive orthant of the unit hypersphere, and $\ell : \mathbb{S}_+^{K-1} \to \mathcal{Z}$ which combines the seeds according to the seed weights $\boldsymbol{w} = \phi_w(\boldsymbol{u})$. Following Bodin et al. [2024], ℓ is defined using a latent-optimal linear (LOL) transport map ensuring that linear combinations of the seeds remain distributed as $p(\boldsymbol{z})$:

$$\boldsymbol{z} = \ell(\boldsymbol{w}; \{\boldsymbol{z}_k\}) = T_{\leftarrow}(\boldsymbol{\xi}\boldsymbol{w}), \quad \boldsymbol{\xi} = [T_{\rightarrow}(\boldsymbol{z}_1), \ldots, T_{\rightarrow}(\boldsymbol{z}_K)]^\top,$$

where $T_{\rightarrow}$ and $T_{\leftarrow}$ are transport maps between $p(\boldsymbol{z})$ and an auxiliary distribution p_ϵ that is rotationally invariant and closed under aggregation (e.g., a zero-mean Gaussian or uniform hyperspherical law). See Figure 1 for an illustration, and Section A for details, including for transport maps for various latent distributions and for the inverse of the surrogate chart.

2.3 Principles for the optimisation landscape

The surrogate chart is designed to obtain surrogate spaces according to three key principles for compatibility with standard optimisation algorithms:

- P1 **Validity:** All coordinates in U map to valid latent samples supported by the generative model: $\boldsymbol{u} \mapsto \boldsymbol{z} = \phi(\boldsymbol{u}; \{\boldsymbol{z}_k\}) \Rightarrow \boldsymbol{z} \sim p(\boldsymbol{z})$. This prevents degeneracy and ensures meaningful generations and objective evaluations during optimisation.
- P2 **Uniqueness:** All locations must encode unique objects given the seeds, i.e. the mapping $\boldsymbol{u} \mapsto \boldsymbol{z}$ is bijective. This avoids redundant representations and ensures that distinct optimisation steps correspond to distinct generated objects.
- P3 **Stationarity:** The relationship between objects' similarity as a function of their Euclidean distance in the surrogate space should be approximately maintained for any pair of objects throughout the space. Euclidean distances in U approximately preserve cosine similarity in the latent space: $\mathrm{sim}(z_i, z_j) \approx v(\|u_i - u_j\|_2)$, where v is a monotone kernel induced by the geometry of the chart.

Together, these properties make surrogate latent spaces well suited for standard black-box optimisation algorithms — including Bayesian Optimisation [Shahriari et al., 2015] and CMA-ES [Hansen, 2016] — while remaining agnostic to the underlying generative model. In Appendix A we show that these principles hold.

2.4 Practical Use in Scientific Optimisation

In scientific and engineering contexts, the objective function $f(\boldsymbol{x})$ — scoring a design — often corresponds to a simulation or experiment (e.g., aerodynamic drag, protein stability, or material energy). Direct gradient propagation through f is typically impractical or impossible, and conditioning generative models on f would require task-specific fine-tuning [Krishnamoorthy et al., 2023, Fan et al., 2023]. By introducing surrogate spaces, allowing us to define informed low-dimensional search spaces for the task at hand, we can search over valid generations as:

$$\boldsymbol{u}^* = \arg\max_{\boldsymbol{u} \in U} f(g(\phi(\boldsymbol{u}))) \qquad \boldsymbol{z} = \phi(\boldsymbol{u}) \qquad \boldsymbol{x} = g(\boldsymbol{z}) \tag{1}$$

This approach generalises across modalities and objective functions, from text-to-image to protein backbones, while requiring no model retraining and only a small number of evaluations of f.

3 Illustrative Experiment: Protein Optimisation

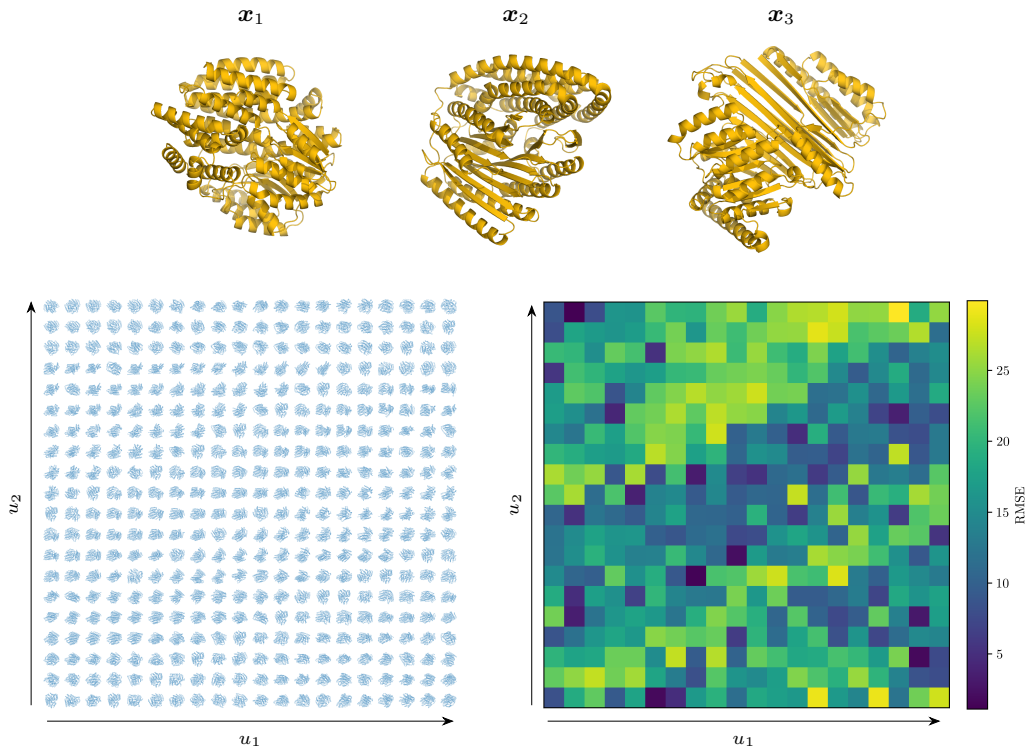


Figure 3: **2D surrogate space for proteins.** A surrogate space $\mathcal{U}^2$ defined by $K = 3$ seed latents (top) yields a structured objective landscape. *Left:* grid of generated backbones across $\mathcal{U}^2$. *Right:* corresponding evaluation scores (C_α–frame RMSE). Notably, significantly lower RMSE proteins exist in the evaluated grid than the evaluations corresponding the seeds (top left, top right, and bottom right corners).

Our methodology focuses on the latent variable and is agnostic to the generative model itself. This allows it to be applied to various data modalities as long as a generative model is available, which we illustrate in Figure 2. We will now demonstrate on a protein backbone design task that surrogate spaces allows us to search for designs efficiently.

Setup. We employ RFDIFFUSION [Watson et al., 2023] as a deterministic generative model of protein backbones, and define a surrogate space using latent vectors corresponding to a small set of seed structures. Each coordinate $\boldsymbol{u} \in U$ maps to a latent $\boldsymbol{z} = \phi(\boldsymbol{u})$ and hence to a backbone $\boldsymbol{x} = g(\boldsymbol{z})$. The objective function $f(\boldsymbol{x})$ evaluates the reconstruction error of each design by computing the C_α–frame RMSE between the generated backbone and its structure predicted by ALPHAFOLD2 [Pak

et al., 2023]. The objective function $f(\boldsymbol{x})$ is non-differentiable[2] and expensive, but we are yet able to deploy a standard black-box optimisation algorithm (CMA–ES [Hansen, 2016]) directly in U.

Results. Surrogate spaces consistently yield RMSEs, higher rates of successful recovery (RMSE$<$ 2.0 Å), and a greater number of structurally diverse successful proteins. Random sampling failed in most trials, consistent with prior results. In contrast, optimisation in $\mathcal{U}$ consistently produced recoverable proteins; random seeds already improved success rates at no additional cost, while filtered seeds (taken as the 24 with lowest RMSE out of 100 random designs, as no a-priori solutions were known) further reduced RMSE and increased yield over ten-fold.

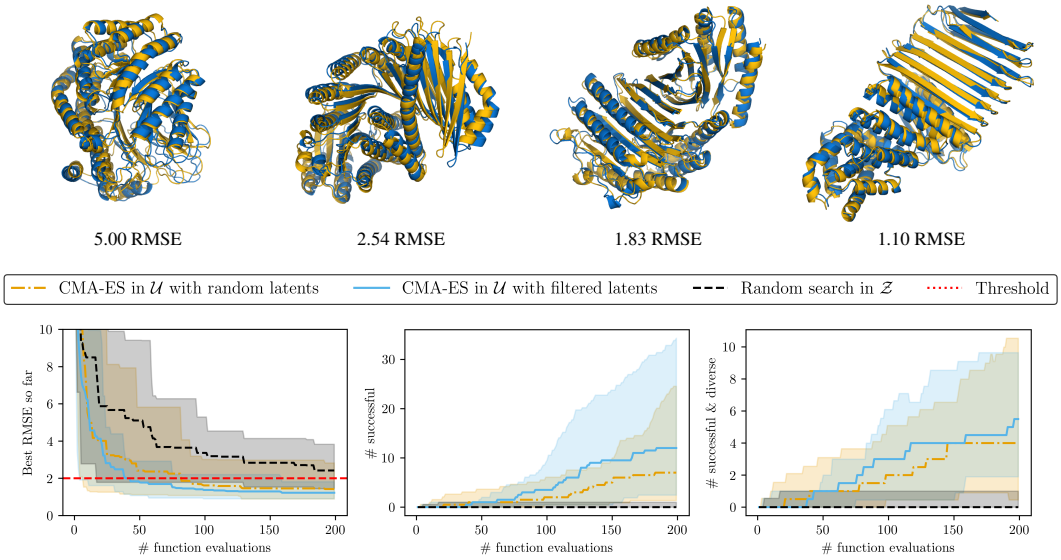


Figure 4: **Protein design with surrogate latent spaces.** *Top:* Representative generations showing the RMSE discrepancy between the RFDIFFUSION backbone *(yellow)* and their ALPHAFOLD2 regeneration *(blue)*. *Bottom:* comparison of standard sampling from the model versus optimisation in our surrogate spaces using CMA–ES. Plots report the median and 90% confidence interval of the best RMSE per step as well as the number of successful and diverse designs. We define a successful generation as meeting the RMSE target from Watson et al. [2023]. Diversity is based on the TM-score [Zhang and Skolnick, 2004] and detailed in Section G.

4 Conclusion

Surrogate latent spaces provides an interface between generative models and scientific optimisation. By constructing low-dimensional surrogate spaces mapping to a model's latent space, we can search the latent space efficiently and guarantee that all explored points remain valid under the generative model. This method transforms pre-trained diffusion or flow models into controllable spaces of candidate solutions that can be explored based on prior knowledge and experiments.

References

Erik Bodin, Alexandru Stere, Dragos D Margineantu, Carl Henrik Ek, and Henry Moss. Linear combinations of latents in generative models: subspaces and beyond. *arXiv preprint arXiv:2408.08558*, 2024.

[2]While the objective we use (reconstruction error from RFDIFFUSION → PROTEINMPNN → ALPHAFOLD2) is nominally differentiable, in practice gradients are statistically weak and prohibitively costly under the small evaluation budgets typical for structure evaluation; moreover, many biologically meaningful objectives (e.g., wet-lab viability, binding or stability assays) are non-differentiable.

J Dauparas, I Anishchenko, N Bennett, H Bai, R J Ragotte, L F Milles, B I M Wicky, A Courbet, R J de Haas, N Bethel, P J Y Leung, T F Huddy, S Pellock, D Tischer, F Chan, B Koepnick, H Nguyen, A Kang, B Sankaran, A K Bera, N P King, and D Baker. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, October 2022.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Zach Evans, Julian D Parker, CJ Carr, Zack Zukowski, Josiah Taylor, and Jordi Pons. Stable audio open. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025.

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models. *Advances in Neural Information Processing Systems*, 36:79858–79885, 2023.

Rafael Gómez-Bombarelli, Jennifer N Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, Ryan P Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS central science*, 4(2):268–276, 2018.

Kailash A Hambarde and Hugo Proenca. Information retrieval: recent advances and beyond. *IEEE Access*, 11:76581–76604, 2023.

Nikolaus Hansen. The cma evolution strategy: A tutorial. *arXiv preprint arXiv:1604.00772*, 2016.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Weijie Kong, Qi Tian, Zijian Zhang, Rox Min, Zuozhuo Dai, Jin Zhou, Jiangfeng Xiong, Xin Li, Bo Wu, Jianwei Zhang, et al. Hunyuanvideo: A systematic framework for large video generative models. *arXiv preprint arXiv:2412.03603*, 2024.

Siddarth Krishnamoorthy, Satvik Mehul Mashkaria, and Aditya Grover. Diffusion models for black-box optimization. In *International Conference on Machine Learning*, pages 17842–17857. PMLR, 2023.

Matt J Kusner, Brooks Paige, and José Miguel Hernández-Lobato. Grammar variational autoencoder. In *International conference on machine learning*, pages 1945–1954. PMLR, 2017.

Black Forest Labs. Flux. `https://github.com/black-forest-labs/flux`, 2024.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

Henry B Moss, Sebastian W Ober, and Tom Diethe. Return of the latent space cowboys: Re-thinking the use of vaes for bayesian optimisation of structured spaces. In *International Conference on Machine Learning*, 2025.

Masahiro Nomura and Masashi Shibata. cmaes: A simple yet practical python library for cma-es. *arXiv preprint arXiv:2402.01373*, 2024.

Marina A Pak, Karina A Markhieva, Mariia S Novikova, Dmitry S Petrov, Ilya S Vorobyev, Ekaterina S Maksimova, Fyodor A Kondrashov, and Dmitry N Ivankov. Using AlphaFold to predict the impact of single mutations on protein stability and function. *PLoS One*, 18(3):e0282689, March 2023.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1): 148–175, 2015.

John Parr Snyder. *Map projections–A working manual*, volume 1395. US Government Printing Office, 1987.

Bowen Song, Soo Min Kwon, Zecheng Zhang, Xinyu Hu, Qing Qu, and Liyue Shen. Solving inverse problems with latent diffusion models via hard data consistency. In *ICLR*, 2024.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

Harald Steck, Chaitanya Ekanadham, and Nathan Kallus. Is cosine-similarity of embeddings really about similarity? In *Companion Proceedings of the ACM Web Conference 2024*, pages 887–890, 2024.

Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vaacute;zquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with rfdiffusion. *Nature*, 620 (7976):1089ndash;1100, Jul 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06415-8. URL `http://dx.doi.org/10.1038/s41586-023-06415-8`.

Yang Zhang and Jeffrey Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins*, 57(4):702–710, December 2004.

A Defining the surrogate chart

The key idea behind our proposed method is to build a search space $\mathcal{U}$ using a collection of K so-called *seeds* that leads to a coordinate system able to generate objects with similar properties to the seeds. The surrogate and the latent space are related through a smooth bijective mapping $\phi : \mathcal{U}, \cdot \to \mathcal{Z}$, which we refer to as the *surrogate chart*. The surrogate chart is associated with the seed latents $\{\boldsymbol{z}_k\}_{k=1}^K$, $\boldsymbol{z}_k \in \mathcal{Z}$ and defined as,

$$\phi(\mathbf{u}, \{\mathbf{z}_k\}_{k=1}^K) = \mathbf{z}, \qquad \mathbf{z} := l\big(\mathbf{w}, \{\mathbf{z}_k\}_{k=1}^K\big), \qquad \mathbf{w} := \phi_w(\mathbf{u}), \tag{2}$$

where $\phi_w : \mathcal{U} \to \mathbb{S}_+^{K-1}$ and $l : \mathbb{S}_+^{K-1}, \cdot \to \mathcal{Z}$ are invertible functions. Here, $\mathbb{S}_+^{K-1}$ denotes the positive orthant of the unit $(K-1)$-sphere,

$$\mathbb{S}_+^{K-1} := \big\{\, \mathbf{w} \in \mathbb{R}^K \;\big|\; \|\mathbf{w}\|_2 = 1,\ w_i \geq 0 \ \ \forall i \,\big\}.$$

The surrogate chart ϕ thus defines a coordinate system for a subset of $\mathcal{Z}$, with inverse mapping

$$\phi^{-1}(\mathbf{z}, \{\mathbf{z}_k\}_{k=1}^K) = \mathbf{u}, \qquad \mathbf{u} := \phi_w^{-1}(\mathbf{w}), \qquad \mathbf{w} := l^{-1}\big(\mathbf{z}, \{\mathbf{z}_k\}_{k=1}^K\big). \tag{3}$$

A.1 Ensuring model support for all coordinates in $\mathcal{U}$

To ensure validity (P1), each coordinate $\boldsymbol{u} \in \mathcal{U}$ maps to a latent realisation $\boldsymbol{z} \in \mathcal{Z}$ via a linear combination weight vector $\boldsymbol{w} \in \mathbb{S}_+^{K-1}$. To guarantee that all coordinates $\boldsymbol{u} \in \mathcal{U}$ are valid, the weight vector $\boldsymbol{w}$ must map to a $\boldsymbol{z} \sim p$, where p is the latent distribution of the generative model. To ensure this, the function l (Equation 2) is formed via a *Latent Optimal Linear combinations (LOL)* transport map [Bodin et al., 2024] which via an 'inner' latent variable $\boldsymbol{\epsilon} \sim p_\epsilon$ guarantees that a linear combination of the seed latents follows the latent distribution p

$$\boldsymbol{z} = l(\boldsymbol{w}, \{\boldsymbol{z}_k\}_{k=1}^K) := \mathcal{T}_{\leftarrow}(\boldsymbol{\epsilon}) \qquad \boldsymbol{\epsilon} := \boldsymbol{\xi}\boldsymbol{w} \qquad \boldsymbol{\xi} := [\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_K]^T \qquad \boldsymbol{\epsilon}_k := \mathcal{T}_{\rightarrow}(\boldsymbol{z}_k), \tag{4}$$

where $\mathcal{T}_{\rightarrow}$ and $\mathcal{T}_{\leftarrow}$ maps from p to p_ϵ and back, respectively, and $l^{-1}(\boldsymbol{z}, \{\boldsymbol{z}_k\}_{k=1}^K) = \boldsymbol{\xi}^+\mathcal{T}_{\rightarrow}(\boldsymbol{z})/||\boldsymbol{\xi}^+\mathcal{T}_{\rightarrow}(\boldsymbol{z})||_2$ where $\boldsymbol{\xi}^+$ is the Moore–Penrose inverse of $\boldsymbol{\xi}$, see Section B.

Amenable inner latents An inner latent distribution p_ϵ is amenable to our methodology if it is zero-mean, rotationally invariant, and closed under aggregation with unit–ℓ_2 weights. Formally, if $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_K \overset{\text{i.i.d.}}{\sim} p_\epsilon$ and $\boldsymbol{w} \in \mathbb{S}^{K-1}$, then an inner latent distribution p_ϵ is amenable to our methodology if it is zero-mean, rotationally invariant, and closed under aggregation with unit–ℓ_2 weights. Formally, if $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_K \overset{\text{i.i.d.}}{\sim} p_\epsilon$ and $\boldsymbol{w} \in \mathbb{S}^{K-1}$, then

$$\boldsymbol{z} \;=\; \mathcal{T}_{\leftarrow}(\boldsymbol{\xi}\boldsymbol{w}), \qquad \boldsymbol{\xi} = [\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_K]^T,$$

again satisfies $\boldsymbol{z} \sim p$. Specifying the maps $T_{\rightarrow}$ and $\mathcal{T}_{\leftarrow}$ to map between the distributions p and p_ϵ is the framework for applying our method to a model at hand:

- **Gaussian latents.** If $p = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, then closure holds directly under $\|\boldsymbol{w}\|_2 = 1$, and we may set $\mathcal{T}_{\rightarrow} = \mathcal{T}_{\leftarrow} = \mathrm{id}$. If $p = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} \neq \mathbf{0}$, then $\mathcal{T}_{\rightarrow}(\boldsymbol{z}) = \boldsymbol{z} - \boldsymbol{\mu}$ and $\mathcal{T}_{\leftarrow}(\boldsymbol{\epsilon}) = \boldsymbol{\epsilon} + \boldsymbol{\mu}$, which centres the distribution before aggregation and restores the mean afterwards. This case was treated in Bodin et al. [2024].
- **Hyperspherical latents.** If $p = \mathrm{Unif}(\mathbb{S}^{D-1})$, where $\boldsymbol{z} \in \mathbb{R}^D$, then $\mathcal{T}_{\rightarrow} = \mathrm{id}$, while $\mathcal{T}_{\leftarrow}$ normalises any linear combination back onto the sphere, $\mathcal{T}_{\leftarrow}(\boldsymbol{\epsilon}) = \frac{\boldsymbol{\epsilon}}{\|\boldsymbol{\epsilon}\|}$. Because this construction is rotation–equivariant, it preserves the uniform law on the sphere.
- **Composite latents**. If the latent variable $\boldsymbol{z}$ can be decomposed into M statistically independent components such that $\boldsymbol{z} = \{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(M)}\}$ where $p(\boldsymbol{z}) \propto p(\boldsymbol{z}^{(1)}) \ldots p(\boldsymbol{z}^{(M)})$, then each component $\boldsymbol{z}^{(m)}, m \in [1, \ldots, M]$ can be mapped separately. This allows, for example, for a model having two or more latent variables, to map these independently, and concatenate their inner latents when computing linear combinations.
- **General scalar distributions** If a latent variable $\boldsymbol{z}$ has individual dimensions z_i that are independent of the others (see above), those elements follow scalar distributions which can be transported optimally to e.g. $\mathcal{N}(0, 1)$ — which is amenable — using the respective cumulative distribution function as proposed in Bodin et al. [2024].

A.2 Ensuring uniqueness for generations from $\mathcal{U}$

We now address principle (P2), that every point in $\mathcal{U}$ specifies an unique realisation of $\boldsymbol{z}$. In Bodin et al. [2024] linear combinations of the seeds are defined for the entire hyperplane, however, following the transformation, such combinations can all be indexed by a bounded set; the weights $\boldsymbol{w} \in \mathbb{S}^{K-1}$ is a sufficient such set to index all possible Latent Optimal Linear combinations, as we show in Section C. As each coordinate $\boldsymbol{u} \in \mathcal{U}$ specifies a unique $\boldsymbol{w} \in \mathbb{S}_+^{K-1}$, which in turn specifies a unique $\boldsymbol{z}$, it follows that each coordinate in $\mathcal{U}$ maps to an unique latent realisation.

The focus on linear combinations weights residing on the positive orthant $\mathbb{S}_+^{K-1} \subset \mathbb{S}^{K-1}$ rather than the whole hypersphere will be motivated by the principle addressed in the next section. But we point out that, to represent positive associations to the seeds we only need positive weights $\mathbb{S}_+^{K-1}$, as only those induce (positive cosine) similarity to the seeds. Note that negative associations (to encourage dissimilarity) to any particular seed could still be represented, by negating the corresponding seed latent. Moreover, as $\mathbb{S}_+^{K-1}$ is a subset of $\mathbb{S}^{K-1}$, uniqueness still holds.

A.3 Ensuring approximate object similarity stationarity

We will now address our final principle of stationarity (P3), motivated by a well-established observation in generative modelling that similarity between generated objects is captured by the cosine similarity of latent vectors. This principle underlies embedding models [Devlin et al., 2019, Mikolov et al., 2013, Kingma and Welling, 2013], widely in e.g. information retrieval [Hambarde and Proenca, 2023] and model alignment [Radford et al., 2021]. We adopt the same assumption in this work.

A common implicit assumption when computing cosine similarities is that latent vectors are centred and isotropic, e.g., unit Gaussian [Steck et al., 2024]. While this may not hold for general latent distributions, it does hold for the *inner latent* representation $\boldsymbol{\epsilon}$ (Equation 4), to which any latent $\boldsymbol{z}$ indexed by $\boldsymbol{w} \in \mathbb{S}^{K-1}$ admits a mapping. Hence, $\boldsymbol{\epsilon}$ serves as the central latent variable in this section.

We adopt $\mathrm{sim}_z(\boldsymbol{z}_i, \boldsymbol{z}_j) := \mathrm{sim}_e(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j)$ where sim_e is the cosine similarity, which is

$$\mathrm{sim}_z(\boldsymbol{z}_i, \boldsymbol{z}_j) = \mathrm{sim}_e(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \frac{\boldsymbol{\epsilon}_i^\top \boldsymbol{\epsilon}_j}{\|\boldsymbol{\epsilon}_i\| \, \|\boldsymbol{\epsilon}_j\|} = \frac{(\boldsymbol{\xi}\boldsymbol{w}_i)^\top \boldsymbol{\xi}\boldsymbol{w}_j}{\|\boldsymbol{\xi}\boldsymbol{w}_i\| \, \|\boldsymbol{\xi}\boldsymbol{w}_j\|} \tag{5}$$

for $\boldsymbol{z}_i = \phi(\boldsymbol{u}_i)$ and $\boldsymbol{z}_j = \phi(\boldsymbol{u}_j)$, where $\boldsymbol{u}_i, \boldsymbol{u}_j \in \mathcal{U}$. For large latent dimensionality D, this reduces to

$$\frac{(\boldsymbol{\xi}\boldsymbol{w}_i)^\top \boldsymbol{\xi}\boldsymbol{w}_j}{\sqrt{(\boldsymbol{w}_i^\top \boldsymbol{\xi}^\top \boldsymbol{\xi}\, \boldsymbol{w}_i)(\boldsymbol{w}_j^\top \boldsymbol{\xi}^\top \boldsymbol{\xi}\, \boldsymbol{w}_j)}} \xrightarrow[D\to\infty]{\text{a.s.}} \boldsymbol{w}_i^\top \boldsymbol{w}_j$$

since $\|\boldsymbol{w}_i\| = \|\boldsymbol{w}_j\| = 1$, $\mathbb{E}[\epsilon_{k,d}] = 0$, and $\boldsymbol{\xi}^\top \boldsymbol{\xi} \xrightarrow[D\to\infty]{\text{a.s.}} D\sigma^2 \boldsymbol{I}$ for independent $\{\boldsymbol{\epsilon}_k\}$. In Section E we show that this effect dominates already at practical dimensionalities, allowing us to control the similarity of objects across $\mathcal{U}$ through the design of ϕ_w.

To preserve similarity as a function of Euclidean distance, we require, for some function v

$$v(\|\mathbf{u}_i - \mathbf{u}_j\|_2) = \phi_w(\mathbf{u}_i)^\top \phi_w(\mathbf{u}_j), \quad \forall\, \mathbf{u}_i, \mathbf{u}_j \in \mathcal{U}, \tag{6}$$

the form of a stationary kernel. This condition can only hold approximately, since $\phi_w : [0,1]^{K-1} \to \mathbb{S}_+^{K-1}$ maps flat to curved space, and Gaussian curvature is preserved under local isometries.

This is analogous to the impossibility of constructing a flat map of the globe that preserves all distances — the classic cartographic problem Snyder [1987]. Restricting to an orthant reduces curvature and thus approximation error, but exact preservation is unattainable for any ϕ_w. In Appendix D, we specify two variants, including one based on the Knothe–Rosenblatt (KR) map, which we found performs well empirically and which we adopt in experiments where not otherwise specified.

B Inverse of the l map

In this section we will derive the inverse of the map l introduced in Section A.

Let $\{\boldsymbol{z}_k\}_{k=1}^K \subset \mathcal{Z}$ be seeds, and define their inner latents $\boldsymbol{\epsilon}_k = \mathcal{T}_\to(\boldsymbol{z}_k)$. Let $\boldsymbol{\xi} = [\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_K] \in \mathbb{R}^{D\times K}$. Assume:

(A1) $\boldsymbol{\xi}$ has full column rank, so that $\boldsymbol{\xi}^+\boldsymbol{\xi} = I_K$;

(A2) For all $\boldsymbol{\epsilon} \in \mathbb{R}^D$,

$$\mathcal{T}_{\rightarrow}\big(\mathcal{T}_{\leftarrow}(\boldsymbol{\epsilon})\big) = \alpha(\boldsymbol{\epsilon})\,\boldsymbol{\epsilon}, \qquad \alpha(\boldsymbol{\epsilon}) > 0.$$

Define the l map

$$l(\boldsymbol{w}, \{\boldsymbol{z}_k\}) = \mathcal{T}_{\leftarrow}(\boldsymbol{\xi}\boldsymbol{w}), \qquad \boldsymbol{w} \in \mathbb{S}_+^{K-1}.$$

Then l is invertible with

$$l^{-1}(\boldsymbol{z}, \{\boldsymbol{z}_k\}) = \frac{\boldsymbol{\xi}^+\,\mathcal{T}_{\rightarrow}(\boldsymbol{z})}{\|\boldsymbol{\xi}^+\,\mathcal{T}_{\rightarrow}(\boldsymbol{z})\|}.$$

Proof. Let $\boldsymbol{z} = l(\boldsymbol{w}, \{\boldsymbol{z}_k\}) = \mathcal{T}_{\leftarrow}(\boldsymbol{\xi}\boldsymbol{w})$. Applying $\mathcal{T}_{\rightarrow}$ and using (A2) gives

$$\mathcal{T}_{\rightarrow}(\boldsymbol{z}) = \mathcal{T}_{\rightarrow}\big(\mathcal{T}_{\leftarrow}(\boldsymbol{\xi}\boldsymbol{w})\big) = \alpha(\boldsymbol{\xi}\boldsymbol{w})\,\boldsymbol{\xi}\boldsymbol{w}.$$

Multiplying by $\boldsymbol{\xi}^+$ and using (A1),

$$\boldsymbol{\xi}^+\mathcal{T}_{\rightarrow}(\boldsymbol{z}) = \alpha(\boldsymbol{\xi}\boldsymbol{w})\,(\boldsymbol{\xi}^+\boldsymbol{\xi})\boldsymbol{w} = \alpha(\boldsymbol{\xi}\boldsymbol{w})\,\boldsymbol{w}.$$

Thus $\boldsymbol{\xi}^+\mathcal{T}_{\rightarrow}(\boldsymbol{z})$ is a positive scalar multiple of $\boldsymbol{w}$. Normalising cancels the unknown factor,

$$\frac{\boldsymbol{\xi}^+\mathcal{T}_{\rightarrow}(\boldsymbol{z})}{\|\boldsymbol{\xi}^+\mathcal{T}_{\rightarrow}(\boldsymbol{z})\|} = \boldsymbol{w},$$

which establishes the result. □

[When normalisation is redundant] Normalisation in the inverse formula is redundant if and only if

$$\mathcal{T}_{\rightarrow} \circ \mathcal{T}_{\leftarrow} = \mathrm{id},$$

that is, when $\alpha(\boldsymbol{\epsilon}) \equiv 1$.

- **Gaussian latents.** $\mathcal{T}_{\rightarrow}$ and $\mathcal{T}_{\leftarrow}$ are exact inverses, so $\alpha = 1$. No normalisation needed.
- **Hyperspherical latents.** With $\mathcal{T}_{\leftarrow}(\boldsymbol{\epsilon}) = \boldsymbol{\epsilon}/\|\boldsymbol{\epsilon}\|$ and $\mathcal{T}_{\rightarrow} = \mathrm{id}$, one has $\alpha(\boldsymbol{\epsilon}) = 1/\|\boldsymbol{\epsilon}\|$. Normalisation is essential.
- **Independent scalar latents mapped via CDF to Gaussian.** Exact inverses, so $\alpha = 1$. Normalisation is redundant.

C $\mathbb{S}^N$ is a sufficient index for Latent Optimal Linear combinations

In this section we will show that linear combination weights on the unit hypersphere is sufficient to index all Latent Optimal Linear combinations [Bodin et al., 2024]. We will first address the Gaussian case and then the general case.

Gaussian latents A linear combination

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{w}, \tag{7}$$

where $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K]$, $\boldsymbol{w} \in \mathbb{R}^K$, $\boldsymbol{z}_k \in \mathbb{R}^D$, $\boldsymbol{z}_k \sim p$ and $p = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ has distribution

$$\boldsymbol{y} \sim \mathcal{N}(\alpha\boldsymbol{\mu}, \beta\boldsymbol{\Sigma}) \tag{8}$$

where $\alpha = \sum_1^K w_i$ and $\beta = \sum_1^K w_i^2$. The variable $\boldsymbol{y}$ does not follow the same distribution p as $\boldsymbol{z}_k$ at weights yielding $\alpha \neq 1$ and $\beta \neq 1$. In Bodin et al. [2024] the following map was proposed for the linear combinations $\boldsymbol{y}$ in the Gaussian case

$$\mathcal{T}(\boldsymbol{y}) = (1 - \frac{\alpha}{\beta})\boldsymbol{\mu} + \frac{\boldsymbol{y}}{\sqrt{\beta}} \tag{9}$$

which is the Monge optimal map between $\mathcal{N}(\alpha\boldsymbol{\mu}, \beta\boldsymbol{\Sigma})$ and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

We can rewrite Equation 9 as

$$\mathcal{T}(\boldsymbol{y}) = (1 - \frac{\alpha}{||\boldsymbol{w}||})\boldsymbol{\mu} + \frac{\boldsymbol{y}}{||\boldsymbol{w}||}, \tag{10}$$

and note that if $\boldsymbol{\mu} = \boldsymbol{0}$, then the transformed variable is invariant to the norm of the weights $\boldsymbol{w}$.

As we can treat the requirement of $\boldsymbol{\mu} = \boldsymbol{0}$ by centring the distribution for a known mean vector, it follows that $\boldsymbol{w} \in \mathbb{S}^{K-1}$ is sufficient to index all such transformed variables.

General case In the non-Gaussian setting, the same principle applies once we introduce an amenable inner latent distribution p_ϵ (Section A). For any latent distribution p, we construct transport maps $\mathcal{T}_\rightarrow$ and $\mathcal{T}_\leftarrow$ such that $\boldsymbol{\epsilon} = \mathcal{T}_\rightarrow(\boldsymbol{z}) \sim p_\epsilon$ and $\boldsymbol{z} = \mathcal{T}_\leftarrow(\boldsymbol{\epsilon}) \sim p$. Because p_ϵ is rotationally invariant and closed under aggregation with unit–ℓ_2 weights, any linear combination

$$\boldsymbol{\epsilon} = \boldsymbol{\xi}\boldsymbol{w}, \qquad \boldsymbol{\xi} = [\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_K]^T, \qquad \boldsymbol{\epsilon}_k \sim p_\epsilon, \tag{11}$$

with $\boldsymbol{w} \in \mathbb{S}^{K-1}$ again satisfies $\boldsymbol{\epsilon} \sim p_\epsilon$. Applying the inverse transport then yields

$$\boldsymbol{z} = \mathcal{T}_\leftarrow(\boldsymbol{\epsilon}) \sim p, \tag{12}$$

showing that the weights $\boldsymbol{w}$ on the unit hypersphere are sufficient to index all latent-optimal linear combinations, regardless of the underlying distribution p.

Concretely, the Gaussian case corresponds to the choice $p_\epsilon = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, where closure holds directly under $\|\boldsymbol{w}\| = 1$. For other distributions, p_ϵ and the associated transport maps adapt accordingly: hyperspherical latents are closed under normalisation, composite latents can be mapped component-wise, and scalar independent latents can be treated dimension-wise via their cumulative distribution functions. In all cases, the invariance of p_ϵ under unit–ℓ_2 aggregation ensures that $\boldsymbol{w} \in \mathbb{S}^{K-1}$ is a sufficient index.

Summary Both the Gaussian case and the general case rely on the same underlying mechanism: linear aggregation in a latent space that is invariant under unit–ℓ_2 weighting, together with a suitable transport map back to the target distribution p. This establishes that restricting to $\boldsymbol{w} \in \mathbb{S}^{K-1}$ is always sufficient to represent all Latent Optimal Linear combinations, independent of the specific form of p.

D Weight charts ϕ_w

Let K be the number of seeds. The *weight chart* is a map

$$\phi_w : [0,1]^{K-1} \;\to\; \mathbb{S}_+^{K-1},$$

where $\mathbb{S}_+^{K-1} = \{w \in \mathbb{R}^K : \|w\|_2 = 1,\ w_i \geq 0\}$ is the positive orthant of the unit hypersphere.

Angular coordinates chart (spherical angles). Set $\theta_i = \frac{\pi}{2} u_i \in (0, \frac{\pi}{2})$ for $i = 1, \ldots, K-1$ and define

$$w_1 = \cos\theta_1, \qquad w_k = \Big(\prod_{i=1}^{k-1} \sin\theta_i\Big)\cos\theta_k \;\; (k = 2, \ldots, K-1), \qquad w_K = \prod_{i=1}^{K-1} \sin\theta_i. \tag{13}$$

Inverse: recover angles by $\theta_1 = \arccos(w_1)$ and $\theta_k = \arccos\big(w_k / \prod_{i=1}^{k-1} \sin\theta_i\big)$ for $k \geq 2$, then $u_i = \frac{2}{\pi}\theta_i$. **Notes**: smooth, *not* equal-area.

Knothe–Rosenblatt (KR) chart. Let $U \in (0,1)^{K-1}$ and define independent stick-breaks

$$v_k \;=\; I_{u_k}^{-1}\Big(\tfrac{1}{2}, \tfrac{K-k}{2}\Big), \quad k = 1, \ldots, K-1,$$

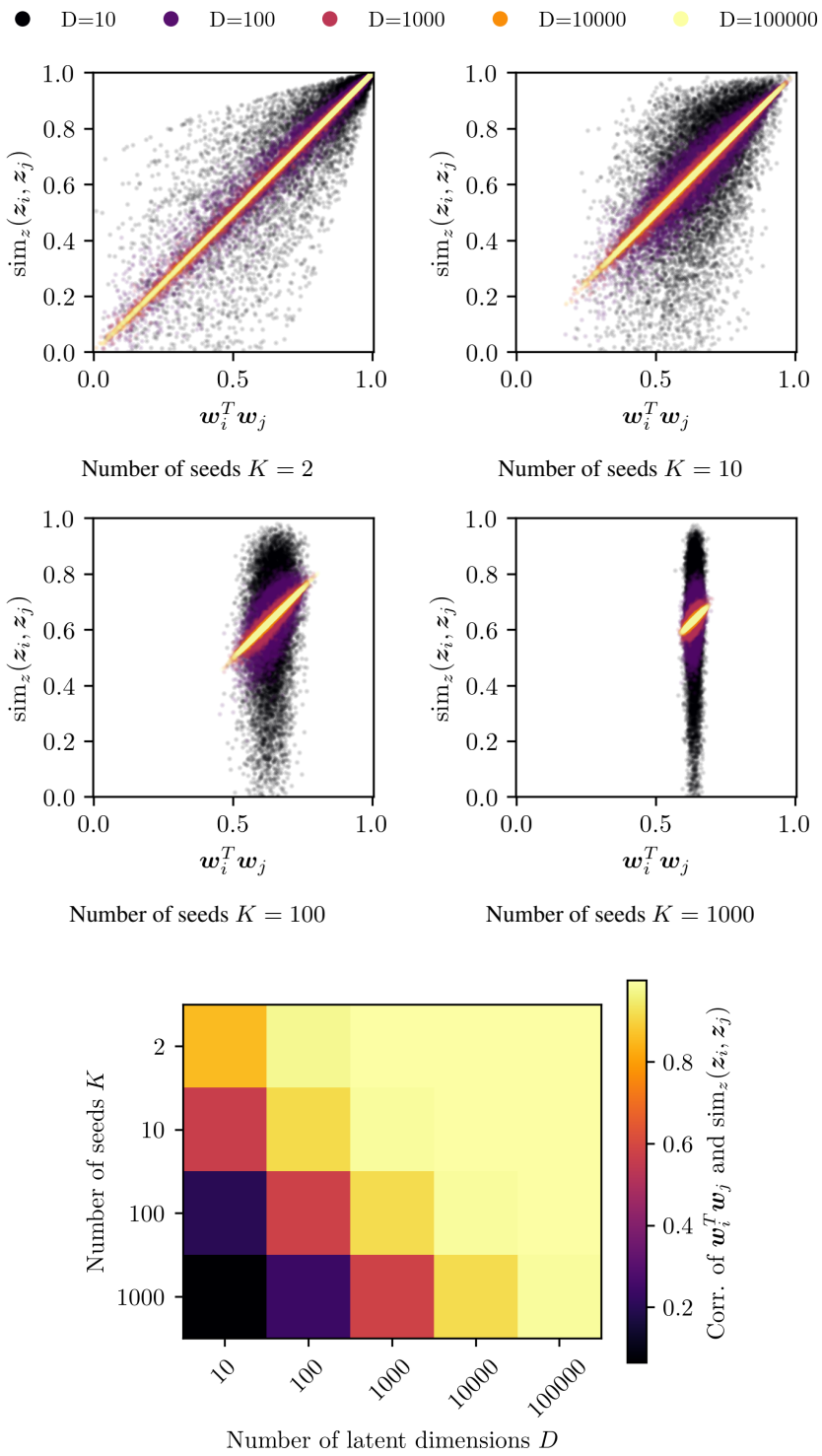
where $I_\cdot^{-1}(a,b)$ is the inverse regularised incomplete beta. Set (Dirichlet stick-breaking)

$$z_1 = v_1, \qquad z_k = v_k \prod_{i=1}^{k-1}(1 - v_i) \;\; (k = 2, \ldots, K-1), \qquad z_K = \prod_{i=1}^{K-1}(1 - v_i), \qquad w_i = \sqrt{z_i}. \tag{14}$$

Inverse: with $z = w \odot w$ and $s_k = \sum_{j=k}^{K} z_j$,

$$v_k = \frac{z_k}{s_k}, \qquad u_k = I_{v_k}\Big(\tfrac{1}{2}, \tfrac{K-k}{2}\Big), \quad k = 1, \ldots, K-1.$$

Notes: smooth, pushes $\mathrm{Unif}([0,1]^{K-1})$ to the uniform surface measure on $\mathbb{S}_+^{K-1}$ (equal-area); use stable `betainc`/`betaincinv`.



Top image

Figure 5: (*top*) Dot products of the weights $\boldsymbol{w}_i^T \boldsymbol{w}_j$ and the cosine similarity $\mathrm{sim}_\epsilon(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j)$ for uniformly drawn samples in $\mathcal{U}$ for $K = 2, 10, 100, 1000$, respectively for various dimensions D, where $\boldsymbol{\epsilon} \in \mathbb{R}^D$. The number of samples per setting is $10,000$, with 100 realisations of the seeds drawn from $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and 100 uniformly sampled $\boldsymbol{u}$ per sampled seeds realisation. (*bottom*) Estimated correlations between $\boldsymbol{w}_i^T \boldsymbol{w}_j$ and ϕ_ϵ uses all the samples per setting.

E The weight chart ϕ_w sets the similarity structure

In Figure 5 we demonstrate numerical evidence for the claim in Section A.2 that the dot product $\boldsymbol{w}_i^T \boldsymbol{w}_j$ is the dominant factor in determining the cosine similarity between two latent variables $\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j \in \mathbb{R}^D$ indexed by a surrogate latent space $\mathcal{U}$. We see that already $D \approx 100, K \leq 10$ yields an dominating $\boldsymbol{w}_i^T \boldsymbol{w}_j$, as shown by Pearson correlations of more than 0.95, and correlations very close to 1 for higher dimensionalities of D (at a rate dependent on K). For reference, typical diffusion and flow matching models [Rombach et al., 2022, Labs, 2024, Lipman et al., 2022] have a D of *tens of thousands* to *hundreds of thousands*, and Kong et al. [2024] has a dimensionality of several *million*.

F RFdiffusion

We adopt the pipeline of Watson et al. [2023], consisting of: (1) backbone generation with RFDIFFUSION, (2) sequence design with PROTEINMPNN [Dauparas et al., 2022], (3) structure reconstruction with ALPHAFOLD2 [Pak et al., 2023], and (4) evaluation by C_α–frame RMSE. Lower RMSE indicates closer agreement between the generated and reconstructed backbones.

In step 1, candidate backbones are sampled from the original RFDIFFUSION model using DDIM. A backbone is defined as the set of C_α coordinates and residue-wise rotations, but does not include categorical amino acid identities. In step 2, each backbone is completed with $M = 8$ amino acid sequences predicted by PROTEINMPNN. This introduces the missing categorical information; however, the predictions are noisy, motivating multiple samples. In step 3, the sequences are passed to ALPHAFOLD2, which reconstructs 3D structures from sequence alone, testing whether the backbone proposed by RFDIFFUSION is compatible with realistic sequences. In step 4, reconstructed proteins are aligned to the original backbones, and C_α RMSE is computed. For each backbone we report the best sequence (minimum RMSE over $M = 8$), following the evaluation protocol of Watson et al. [2023]. For optimiser setups, see H

As in Watson et al. [2023] we adopt a threshold of $T = 2.0$ Å RMSE to define successful recovery, however we drop their secondary filtering metric of designs having PAE < 5.0 to focus on proof of principle, although in future this could naturally be supported by considering multi-objective optimisation. For fairness, all baselines were recomputed under our evaluation. Each optimisation run used 200 iterations, twice the 100 generations of the original paper.

RFDIFFUSION parametrises a backbone of length N by residue-wise frames $(\boldsymbol{x}_{pos}^{(t)}, \boldsymbol{x}_{rot}^{(t)}) \in \mathbb{R}^{3N} \times \mathrm{SO}(3)^N$, where $\boldsymbol{x}_{pos}^{(t)}$ are C_α coordinates and $\boldsymbol{x}_{rot}^{(t)}$ are orientations derived from N–C_α–C triplets, measured from a reference frame. The forward diffusion process applies Gaussian noise to $\boldsymbol{x}_{pos}$ and Brownian motion on $\boldsymbol{x}_{rot}$; generation is by reverse integration of the probability–flow ODE. The resulting latent is

$$\boldsymbol{z} = (\boldsymbol{z}_{pos}, \boldsymbol{z}_{rot}) = (\boldsymbol{x}_{pos}^{(T)}, \boldsymbol{x}_{rot}^{(T)}), \quad \boldsymbol{z}_{pos} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}^{\mathbf{3N}}),\ \boldsymbol{z}_{rot} \sim \mathrm{Unif}(\mathrm{SO}(3)^N).$$

Because $\boldsymbol{z}_{pos}$ follows a Gaussian distribution and we parametrise $\boldsymbol{z}_{rot}$ as quaternions which are uniformly distributed on $\boldsymbol{z}_{rot} \sim \mathrm{Unif}(\mathbb{S}^{3N})$, we can directly apply composite latents from Section A.1 to construct surrogate latent spaces $\mathcal{U}$.

Figure 3 shows a N=2 dimensional latent space formed from K=3 seed latents, over which a grid of protein structures have been generated and evaluated according to the target objective. Clear structure in shown in the objective space which makes this objective amenable to optimisation.

For the protein optimisation experiments we set the number of seed latents to $K = 24$. Optimisation is performed in $\mathcal{U}$ via CMA–ES, with candidates mapped back into $\mathcal{Z}$ for decoding and evaluation. Two seed selection strategies were used. *Random seeds:* sampled directly from the prior distributions, incurring no additional cost. *Filtered seeds:* obtained by first generating 100 backbones from the base model, ranking them by RMSE, and selecting the top $K = 24$ latents as seeds. None of these passed the $T = 2.0$ threshold, but they provided a stronger starting point than random seeds. The extra cost relative to random seeds is generating and evaluating the pipeline 100 times.

Method	Best RMSE ↓	SUCC ↑	SUCCDIV ↑
Random in $\mathcal{Z}$	2.33	0	0
CMA–ES in $\mathcal{U}^{K-1}$ (random seeds)	1.19	7	4
CMA–ES in $\mathcal{U}^{K-1}$ (filtered seeds)	1.08	12	5.5

Table 1: Protein design at $N = 600$ after 200 iterations. SUCC: successful recoveries (RMSE$< T$); SUCCDIV: distinct clusters of successful backbones. Medians over 10 runs.

G Template Modelling Score (TM-score)

The Template Modelling score (TM-score) is a widely used measure of structural similarity between two protein backbones. Unlike RMSE, which is sensitive to local deviations and scales poorly with chain length, the TM-score is normalised to the length of the target protein and therefore more suitable for comparing proteins of different sizes [Zhang and Skolnick, 2004].

Given a target structure of length L and a comparison structure, the TM-score is defined as

$$\mathrm{TM{-}score} \;=\; \max_{\text{alignments}} \frac{1}{L} \sum_{i=1}^{L} \frac{1}{1 + \left(\frac{d_i}{d_0(L)}\right)^2}, \tag{15}$$

where d_i is the distance between the ith pair of aligned C_α atoms under a given alignment, and $d_0(L) = 1.24\sqrt[3]{L - 15} - 1.8$ is a normalisation factor that accounts for protein length. The score lies in $[0, 1]$, with higher values indicating greater structural similarity.

As a rule of thumb, TM-score > 0.5 indicates that two structures share the same fold, while TM-score < 0.17 corresponds to similarity expected by chance.

In our case, all generations are of equal length, so RMSE remains valid; however, using TM-score not only allows us to apply established interpretative thresholds, but also lets us follow Watson et al. [2023] in treating two designs as *non-diverse* if their pairwise TM-score exceeds 0.6.

Diversity counting. To compute the number of diverse generations reported in Section 3, we apply the following greedy procedure: 1. Sort generated proteins by reconstruction accuracy (lowest RMSE first). 2. Initialise the diverse set with the best structure. 3. For each subsequent protein, compute its TM-score against all members of the current diverse set. 4. Add it to the diverse set if its TM-score is ≤ 0.6 with respect to all previously accepted members; otherwise, discard it.

This ensures that each counted generation is both accurate (passes the RMSE threshold) and structurally distinct under TM-score. *Note:* because a newly generated protein may achieve lower RMSE than existing members of the diverse set while simultaneously being non-diverse with respect to several of them, the overall count of diverse structures can decrease across iterations.

H Optimiser setups

Optimiser setups:

- **CMA-ES.**. We use the implementation from Nomura and Shibata [2024] with population size 4 and $\sigma = 0.2$.
- **Random search in $\mathcal{Z}$.** Standard random (and independent) sampling from the latent distribution.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly describe the contribution—constructing surrogate latent spaces that permit black-box optimisation over the outputs of generative models. The stated claims match both the methodological framework and the empirical demonstrations across images, audio, video, and protein design.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper highlights several practical considerations throughout, such as the requirement for deterministic sampling from the generative model, the role of seeds in defining the surrogate subspace, and the way optimisation budgets affect results. These are discussed contextually rather than as an explicit limitations section.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Formal derivations and proofs are provided in the appendices, and all assumptions required for the constructions are stated or referenced in the corresponding sections.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper describes seed selection, surrogate dimensionalities, optimisation procedures (e.g., CMA-ES), evaluation metrics, and generative model configurations. These details are sufficient for reproduction with the same publicly available pretrained models.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: Code is not yet released because a full-version submission is still in preparation. The experiments rely on publicly available pretrained generative models, and the procedures described in the paper are sufficient for independent reproduction.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies optimisation budgets, seed counts, generative models used, scoring metrics, and protocol details; further information is provided in the appendices.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Protein-design experiments report median trajectories with 90% confidence intervals and success/diversity statistics capturing optimisation variability.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper specifies use of pretrained diffusion/flow-matching models, single-GPU inference for generation, and evaluation compute (e.g., AlphaFold2). Hardware requirements and computational scale are apparent from the experimental description.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work uses public pretrained generative models, no human subjects or private data, and presents no ethical or safety concerns beyond those already associated with standard generative-model usage.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification: The paper discusses scientific design-optimisation impacts. The approach does not introduce new misuse vectors beyond existing generative models; societal risk is low.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper does not release new high-risk models or datasets; it constructs surrogate search spaces on top of existing public generative models.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All generative models and toolchains used (e.g., Stable Diffusion, StableAudio2, HunyuanVideo, RFDiffusion, AlphaFold2, ProteinMPNN) are cited and come with known public licenses.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not introduce or release new datasets, pretrained models, or assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing or human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: No human-subjects research is conducted.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLMs are not part of the method development or experimental pipeline.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.