AlphaFold's Bayesian Roots in Probability Kinematics

Thomas Hamelryck
University of Copenhagen, Denmark

Kanti V. Mardia
University of Leeds, UK

Abstract

The seminal breakthrough of AlphaFold in protein structure prediction relied on a learned potential energy function parameterized by deep models, in contrast to its successors AlphaFold2 and AlphaFold3, which lack an explicit probabilistic interpretation. While AlphaFold's potential was originally justified by heuristic analogy to physical *potentials of mean force*, we show that it can instead be understood as a principled instance of *probability kinematics* (PK), also known as *Jeffrey conditioning*, a generalization of Bayesian updating. This reinterpretation reveals that AlphaFold is a generalized Bayesian model that explicitly defines a posterior distribution over structures, providing a deeper explanation of its success and a foundation for future model design. To demonstrate this framework with precision, we introduce a tractable synthetic model in which an angular random walk prior is updated with distance-based evidence via PK, directly mirroring AlphaFold's mechanism. This setting allows us to explore the probabilistic foundations of AlphaFold in a clear and interpretable way. Our work connects a landmark in protein structure prediction to a broader class of compositional deep generative models and points to new opportunities for principled probabilistic approaches.

1 INTRODUCTION

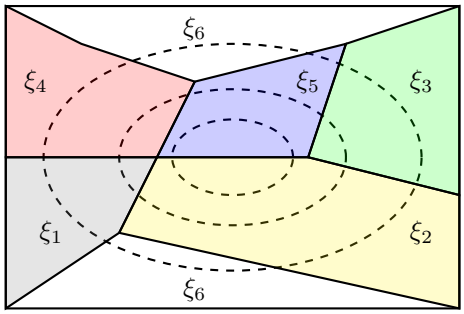
The protein structure prediction problem – predicting the three-dimensional (3D) structure of a protein from its amino acid sequence – has been a central unsolved challenge for more than 50 years (Dill and MacCallum, 2012). In 2024, the shared Nobel Prize in Chemistry was awarded to John Jumper and Demis Hassabis for AlphaFold's solution of this problem (Callaway, 2024). Here, we focus on the initial formulation of AlphaFold, which we will refer to as AlphaFold1 (AF1) (Senior et al., 2019, 2020). AF1 makes use of a potential function parameterized by deep learning and (as we show here) rooted in generalized Bayesian updating, which differs considerably from the end-to-end deep learning pipeline approach of its two successors that despite increased pragmatic accuracy lack an explicit, principled, SE(3)-invariant probabilistic model over the conformational space of proteins (Jumper et al., 2021; Abramson et al., 2024; Jing et al., 2025). Predictions by AF1 are obtained by minimizing this potential energy function with respect to the protein's dihedral angles. We argue that AF1's probabilistic nature is well worth reconsidering as a basis for further development, as we discuss below.

AF1's potential function was justified heuristically by appealing to physical *potentials of mean force* (PMFs) (Senior et al., 2019, 2020) as used in the physics of liquids, first formulated by Kirkwood in 1935 (Kirkwood, 1935). Here, we instead interpret AF1 as an instance of an under-recognized generalization of Bayesian updating called *Jeffrey conditioning* or *probability kinematics* (PK). PK was introduced by the philosopher of logic and probability Richard C. Jeffrey in 1957 (Jeffrey, 1957, 1990, 2004), and specifies how to incorporate *soft evidence*, i.e., updated probabilities on a partition of a random variable's sample space (see Figure 1), rather than conditioning on a concrete event. AF1 updates a prior distribution over a protein's dihedral angles with soft evidence on pairwise distances between amino acids, both obtained from deep models.

This theoretical paper offers the following contributions:

- Building on a definition of PK for probability density functions (PDFs) and infinite partitions tailored to the case of AF1, we propose a generalized Bayesian reinterpretation of AF1´s potential function, and make explicit its functional form as a

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).



$$(A) \quad \pi(\boldsymbol{\omega}) = \sum_i \pi(\boldsymbol{\omega} \mid \xi_i)\pi(\xi_i),$$

$$(B) \quad p(\boldsymbol{\omega} \mid \xi_i) = \pi(\boldsymbol{\omega} \mid \xi_i)$$

$$(C) \quad \Rightarrow p(\boldsymbol{\omega}) = \sum_i \pi(\boldsymbol{\omega} \mid \xi_i)p(\xi_i)$$

$$(D) \quad \Rightarrow p(\boldsymbol{\omega}) = \frac{p\left(\xi(\boldsymbol{\omega})\right)}{\pi\left(\xi(\boldsymbol{\omega})\right)}\pi(\boldsymbol{\omega})$$

Figure 1: **Schematic example of generalized Bayesian updating by PK.** The dotted lines denote the equiprobability contours of the probability density function (PDF) of a prior distribution, $\pi(\boldsymbol{\omega})$, in this example defined on a rectangle. The solid lines represent a finite partition $\{\xi_i\}$ of the sample space of $\pi(\boldsymbol{\omega})$, with probabilities $\{\pi(\xi_i)\}$. $\pi(\boldsymbol{\omega})$ can thus be written as in (A). New evidence is then obtained in the form of updated probabilities for the partition,$\{p(\xi_i)\}$. PK updates the prior $\pi(\boldsymbol{\omega})$ to the posterior $p(\boldsymbol{\omega})$ by keeping the conditional following (B), which is called the *Jeffrey-condition* or *J-condition*, and replacing the marginal as in (C). An equivalent expression (D), where $\xi(\boldsymbol{\omega})$ is the partition element that contains $\boldsymbol{\omega}$, is obtained by applying Bayes' formula to (C). The latter form, called the *reference ratio* update, is of interest when the conditional in (C) is not readily available or tractable, as is the case for AF1. PK readily generalizes to infinite partitions (see Section 3).

prior density adjusted by a ratio of PDFs.

- We contrast our framework with AF1´s heuristic explanation based on PMFs and link its deep model to other statistical models of protein structure. This bridge is timely, as there is a growing need for modeling epistemic and aleatoric uncertainty (Bowman, 2024; Ramasamy et al., 2025; Chakravarty et al., 2024; Agarwal and McShan, 2024; Gavalda-Garcia et al., 2025; Chakravarty et al., 2025; López-Sagaseta and Urdiciain, 2025).

- We introduce a tractable two-dimensional synthetic model that is analogous to AF1's potential. The model updates an angular prior with evidence in the form of a PDF over a distance. This setup mirrors the probabilistic framework behind AF1's potential and allows us to demonstrate the update's precision via hypothesis testing. In the Appendix, we also provide an application to a small 3D protein model.

- We reveal PK as a previously unrecognized foundation for formulating deep generative models of biomolecular structure.

- Our results show PK's relevance and potential in formulating *compositional* deep generative models, i.e., models constructed by combining simpler probabilistic components into a coherent whole (Ghahramani, 2015; Wood et al., 2014; Bingham et al., 2019).

Our aim is not to improve or assess empirical accuracy, as AF1's historical success is well established (Senior et al., 2019, 2020), but to clarify its previously unrecognized statistical foundation and point out its potential for future work. To do this, we make use of a principled probabilistic framework, tested in a controlled synthetic setting where accuracy can be assessed.

2 BACKGROUND

2.1 Preliminaries

The protein amino acid sequence, $\boldsymbol{s}$, of length L, is represented by a $L \times 20$ tensor of one-hot encodings of the 20 amino acid symbols. $\boldsymbol{M}$ is an $M \times L \times 20$ tensor that contains the multiple sequence alignment of M protein sequences related to $\boldsymbol{s}$. Following AF1, a protein structure will be represented by its backbone atoms only, including $N, C\alpha, C, C\beta$ atoms (see Figure 2). The 3D coordinates of these four atoms for each amino acid are in the tensor $\boldsymbol{X}$, which is $L \times 4 \times 3$.

Protein structure is often parameterized using dihedral angles (Ramachandran et al., 1963),[1] which has the advantage of invariance to rotation and translation and statistical simplicity (Edgoose et al., 1998; Boomsma et al., 2008; Hamelryck et al., 2012). Let $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ be vectors of L dihedral angles in $(-\pi, \pi]$. We will write $\boldsymbol{X}(\boldsymbol{\phi}, \boldsymbol{\psi})$ for the realization of $\boldsymbol{X}$ up to an arbitrary rotation and translation for a given set of dihedral angles. Following AF1 (Senior et al., 2019, 2020), we

[1] A *dihedral angle* or *torsion angle*, defined by four consecutively bonded atoms, is the angle between the planes formed by the first three atoms and the last three. Under the common assumption that bond lengths and bond angles take fixed, ideal values, a protein's 3D structure is fully determined by its dihedral angles.

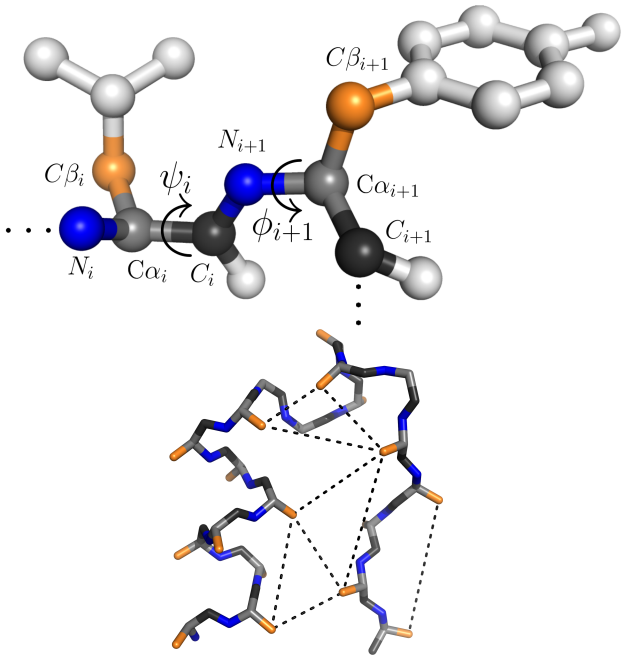


Figure 2: **(top)** Two amino acids in a protein, with relevant atoms and dihedral angles labeled. Atoms that do not play a role in the AF1 potential, including side chain and backbone oxygen atoms, are shown in white. The continuation of the protein backbone is indicated with dots. **(bottom)** A protein with some $C\beta - C\beta$ distances shown as dotted lines. Only the atoms relevant to the AF1 potential are shown. Figures made with PyMol (Schrödinger, LLC, 2015).

assume ideal bond angles and lengths, *trans* peptide bonds, and ignore amino acid side chains and backbone oxygen atoms, without any loss of generality (see Figure 2).

Let $\boldsymbol{D}$ represent the $L \times L$ distance matrix of distances between L $C\beta$ atoms, with $L(L-1)/2$ unique values. Note that although these pairwise distances only involve the $C\beta$ atoms, they are taken as representative distances between amino acids. A given set of atomic coordinates, $\boldsymbol{X}$, or dihedral angles, $(\boldsymbol{\phi}, \boldsymbol{\psi})$, determines $\boldsymbol{D}$ fully – we will write $\boldsymbol{D}(\boldsymbol{X})$ and $\boldsymbol{D}(\boldsymbol{\phi}, \boldsymbol{\psi})$ for the realization of $\boldsymbol{D}$ for given coordinates or dihedral angles.

We will use the following generic notation to emphasize features common to PK and PMFs. Let $\boldsymbol{\omega}$ be a random variable, and let $\boldsymbol{\xi}$ be a random variable obtained via a many-to-one mapping of $\boldsymbol{\omega}$. Thus, $\boldsymbol{\xi}$ is the random variable and $\boldsymbol{\xi}(\boldsymbol{\omega})$ is its concrete *realization* for a given value of $\boldsymbol{\omega}$. We will call $\boldsymbol{\omega}$ a *fine-grained* random variable, and $\boldsymbol{\xi}$ a corresponding *coarse-grained* or *collective* random variable. Depending on the context, $\boldsymbol{\omega}$ and $\boldsymbol{\xi}$ may be matrices, vectors, sets or scalars; in the latter two cases we omit the boldface. These symbols are reused and explained in their context as needed. Notably, in the previous paragraph, $\boldsymbol{D}$ is a coarse grained, collective variable, while $\boldsymbol{X}$ is a corresponding fine-grained variable, corresponding to instances of $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$, respectively. Appendix (A) gives an overview of notation and abbreviations.

2.2 AF1's potential

AF1's potential contains four terms, three of which are parameterized by deep models. A fourth, repulsive potential is derived from physical considerations and merely serves to prevent atoms from overlapping. The first three potentials are central to this article. Section 4 provides details on the PDFs mentioned below.

The first potential concerns the dihedral angles and is given by

$$V_{\mathrm{A}}(\boldsymbol{\phi}, \boldsymbol{\psi}) = -\sum_i \log p\left(\phi_i, \psi_i \mid \boldsymbol{s}, \boldsymbol{M}\right), \tag{1}$$

where the sum runs over all amino acids. The second energy term contains the pairwise distances between the $C\beta$ atoms, which serve as representative distances between amino acids i and j,

$$V_{\mathrm{D}}(\boldsymbol{D}) = -\sum_{i,j,i\neq j} \log p\left(D_{i,j} \mid \boldsymbol{s}, \boldsymbol{M}\right). \tag{2}$$

As we will show in Section 4, the formulation of a consistent joint probability distribution over angles and distances requires a third term – a so-called *reference potential*, which also concerns pairwise distances. In AF1, this potential depends only on the number of amino acids in the protein, L, and a one-hot vector that specifies the positions of Glycine residues, $\mathbb{I}_{\mathrm{Gly}}(\boldsymbol{s})$. Glycine is a small amino acid that allows exceptional conformational freedom to the protein chain. The reference potential is

$$V_{\mathrm{R}}(\boldsymbol{D}) = -\sum_{i,j,i\neq j} \log p\left(D_{i,j} \mid L, \mathbb{I}_{\mathrm{Gly}}(\boldsymbol{s})\right). \tag{3}$$

What AF1 refers to as a Kirkwood PMF is then obtained from their difference,

$$V_{\mathrm{PMF}}\left(\boldsymbol{D}\right) = V_{\mathrm{D}}\left(\boldsymbol{D}\right) - V_{\mathrm{R}}\left(\boldsymbol{D}\right). \tag{4}$$

The final potential brings the three potentials above together, and combines it with the predefined repulsive potential,

$$V(\boldsymbol{\phi}, \boldsymbol{\psi}) = V_{\mathrm{PMF}}\left(\boldsymbol{D}(\boldsymbol{\phi}, \boldsymbol{\psi})\right) + V_{\mathrm{A}}(\boldsymbol{\phi}, \boldsymbol{\psi}) + V_{\mathrm{Repulsive}}\left(\boldsymbol{X}(\boldsymbol{\phi}, \boldsymbol{\psi})\right). \tag{5}$$

Note that $\boldsymbol{D}$ and $\boldsymbol{X}$ (up to an arbitrary rotation and translation) in the above expression are fully determined by the dihedral angles, $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ (see Figure 2).

Prediction by AF1 amounts to minimizing the potential as a function of the dihedral angles using gradient descent. In the next section, we explain how AF1 justifies this potential by heuristic analogy with PMFs for liquid systems.

2.3 AF1's potential justified as a PMF

Physical potentials of mean force A PMF can be understood as an *averaged*, *mean* or *effective* potential regarding some coordinates (typically called the *collective* or *coarse-grained variables*) of a physical system, while averaging over the remaining degrees of freedom (Chandler, 1987; Zuckerman, 2010; Hartmann et al., 2011). A PMF is defined as

$$p(\boldsymbol{\xi}) \propto \exp(-\mathrm{PMF}(\boldsymbol{\xi})/kT) \tag{6}$$

$$= \int \delta\left(\boldsymbol{\xi} - \boldsymbol{\xi}(\boldsymbol{X})\right) \exp(-U(\boldsymbol{X})/kT)\mathrm{d}\boldsymbol{X}, \tag{7}$$

where k is Boltzmann's constant, T is the temperature, $\boldsymbol{X}$ are the atomic coordinates of the system, $\boldsymbol{\xi}$ is the collective random variable (e.g., distances between a subset of the atoms), $\boldsymbol{\xi}(\boldsymbol{X})$ is the realization of $\boldsymbol{\xi}$ at $\boldsymbol{X}$, $\delta(.)$ is the Dirac delta function, and the exponential factor is the Boltzmann distribution associated with the internal energy U.

PMFs were introduced by Kirkwood in his seminal paper from 1935 (Kirkwood, 1935), applied to mixtures of liquids. For this particular physical system, Kirkwood proposed expressions of the form,

$$W_{\mathrm{Kirkwood}}(r) = -kT \log\left(\frac{p(r)}{p_{\mathrm{Reference}}(r)}\right), \tag{8}$$

where r is the distance between two identified liquid particles. The numerator in the log is the probability density observed in a liquid, while the denominator refers to a so-called *reference distribution*. In the case of Kirkwood potentials, the reference distribution is well-defined, corresponding to an ideal state of non-interacting particles (Kirkwood, 1935). It serves to isolate genuine intermolecular interactions from features that are purely geometric or entropic. Technically, a Kirkwood potential is thus a *relative* PMF, corresponding to a difference between the PMFs of the interacting liquid and its corresponding non-interacting reference system. We now turn to the heuristic adoption of formulas resembling Kirkwood potentials, as done by AF1, for biomolecular prediction and simulation.

Knowledge based potentials Knowledge based potentials (KBPs) attempt to formulate empirical energy functions based on statistics derived from the database of known structures (Pohl, 1971; Tanaka and Scheraga, 1976; Miyazawa and Jernigan, 1985; Sippl, 1990). KBPs based on the pairwise distances between $C\beta$ atoms were introduced by Sippl (1990) and widely adopted (Sippl, 1995, 1993a; Sippl et al., 1996). The energy of a protein structure following such a KBP is given by

$$\Delta E_{\boldsymbol{s}}(\boldsymbol{X}) = \sum_{i<j} -\log\left(\frac{p\left(D_{i,j}(\boldsymbol{X}) \mid s_i, s_j\right)}{p_{\mathrm{R}}\left(D_{i,j}(\boldsymbol{X}) \mid s_i, s_j\right)}\right), \tag{9}$$

where s_i, s_j are amino acid labels. These heuristic potentials were loosely inspired by the ratio of PDFs arising in Kirkwood potentials (Eq. 8).

Early Rosetta work (Simons et al., 1997) used a Sippl-style potential involving discretized distances combined with Monte Carlo sampling of protein fragments. Sippl-style potentials were also used for protein structure evaluation, fold recognition and threading, and model refinement (Zhou and Zhou, 2002; Shen and Sali, 2006; Lu et al., 2008; Zhang and Zhang, 2010). In the 2010s, later versions of Rosetta and other programs instead typically relied on predicted contacts/distances as straightforward restraints, which is distinct from Sippl-style potentials and probability kinematics.

Despite pragmatic success, their theoretical justification as genuine PMFs was widely contested (Thomas and Dill, 1996; Ben-Naim, 1997; Koppensteiner and Sippl, 1998). Notably, unlike physical PMFs and Kirkwood potentials, KBPs are not derived from a Boltzmann distribution of a single physical system, but from a heterogeneous data set of different protein structures, and the reference distribution is undefined (Ben-Naim, 1997; Koppensteiner and Sippl, 1998). Attempts at *qualitative, heuristic* justification invoked properties of random polymers (Finkelstein et al., 1995) and probabilistic reasoning (Simons et al., 1997). Subsequently, KBPs were *quantitatively, formally* explained (Hamelryck et al., 2010; Valentin et al., 2014), and eventually recognized as an instance of PK (Hamelryck et al., 2015), to which we turn next.

2.4 Probability kinematics

As in the classic example of *Whitworth's horses* (Diaconis and Zabell, 1982) (see Appendix C), new information often does not arrive in the form of a deterministic event, but rather as a revision in the probability of various hypotheses. PK provides a framework for updating beliefs under such uncertain evidence. Unlike classical Bayesian conditioning, which updates a prior $\pi(\boldsymbol{\omega})$ to a posterior $p(\boldsymbol{\omega}) \propto p(\boldsymbol{y} \mid \boldsymbol{\omega})\pi(\boldsymbol{\omega})$ using data $\boldsymbol{y}$, PK updating uses revised probabilities on a partition of the sample space of $\boldsymbol{\omega}$ (Figure 1).

Given a prior distribution $\pi(\boldsymbol{\omega})$ over a sample space Ω, and a finite partition $\{\xi_i\}$ of Ω along with new marginal

probabilities $p(\xi_i)$ that differ from the prior probabilities $\pi(\xi_i)$, the PK update prescribes the posterior as

$$p(\boldsymbol{\omega}) = \sum_i \pi(\boldsymbol{\omega} \mid \xi_i) p(\xi_i), \tag{10}$$

by construction making use of the *Jeffrey- or J-condition* $p(\boldsymbol{\omega} \mid \xi_i) = \pi(\boldsymbol{\omega} \mid \xi_i)$.

Unlike conventional Bayesian inference, PK does not rely on an explicit likelihood function, and its evidence is not a concrete event. Nonetheless, PK preserves essential features of Bayesian updating. First, PK satisfies the Dutch book argument (Armendt, 1980). Second, PK reduces to conventional Bayesian updating when for one of the partition elements, $p(\xi_i) = 1$. Third, among all distributions that match the updated marginals $\{p(\xi_i)\}$, the PK update minimizes the Kullback-Leibler divergence to the prior π (Diaconis and Zabell, 1982). This links the method to a variational principle (Blei et al., 2017; Kingma and Welling, 2019) that fits the Bayesian interpretation of probability (Diaconis and Zabell, 1982). PK is usually formulated in terms of probabilities and finite partitions (Meehan and Zhang, 2020). In the next section, we outline the more general probabilistic framework needed to address AF1.

3 PROBABILISTIC FRAMEWORK

In this section, we first formulate a definition of PK that is suitable for our purposes (i.e., concerning PDFs and infinite partitions), and review some of PK's properties. As a preliminary to its application in the case of AF1, we consider the case of pairwise distance KBPs and contrast PK with physical PMFs.

3.1 PK for the case of PDFs and infinite partitions

We provide a definition tailored to the present case, based on (Diaconis and Zabell, 1982) and (Meehan and Zhang, 2020), while avoiding the formalism of measure theory and irrelevant boundary cases.

Definition 1 (Probability kinematics for infinite partitions and PDFs)**.** *Given:*

- *A random variable $\boldsymbol{\omega} \in \Omega$, where Ω is a smooth manifold (e.g., the space of m dihedral angles, the m-torus $\mathbb{T}^m$), equipped with a prior density $\pi(\boldsymbol{\omega})$.*
- *A random variable $\boldsymbol{\xi}$, defined as a many-to-one function of $\boldsymbol{\omega}$. For $\boldsymbol{\omega} \in \Omega$, the corresponding realization is written $\boldsymbol{\xi}(\boldsymbol{\omega}) \in \Xi$, where Ξ is a smooth manifold (e.g., the space of valid pairwise distances between n points in $\mathbb{R}^3$). The prior density of $\boldsymbol{\xi}$ is induced by that of $\boldsymbol{\omega}$ as*

$$\pi(\boldsymbol{\xi}) = \int \delta\big(\boldsymbol{\xi}(\boldsymbol{\omega}) - \boldsymbol{\xi}\big)\, \pi(\boldsymbol{\omega})\, \mathrm{d}\boldsymbol{\omega}. \tag{11}$$

  Each value $\boldsymbol{\xi} \in \Xi$ corresponds to the partition element $\{\boldsymbol{\omega} \in \Omega : \boldsymbol{\xi}(\boldsymbol{\omega}) = \boldsymbol{\xi}\}$.
- ***Evidence** as a new density $p(\boldsymbol{\xi})$.*
- *Densities $\pi(\boldsymbol{\xi})$ and $p(\boldsymbol{\xi})$ are positive on a common support and normalized. We will call $\boldsymbol{\omega}$ and $\boldsymbol{\xi}$ the **fine-grained** and the **coarse-grained or collective** random variables, respectively.*
- *We assume all functions and densities are sufficiently well-behaved to ensure the existence and measurability of the integrals and conditional densities involved.*

*A **probability kinematics** update defines a posterior density $p(\boldsymbol{\omega})$ by satisfying the following conditions, which together ensure the evidence is incorporated while preserving conditional distributions:*

1. ***Faithfulness**: The marginal density of $\boldsymbol{\xi}$ matches the evidence,*

$$p(\boldsymbol{\xi}) = \int p(\boldsymbol{\omega}) \delta(\boldsymbol{\xi}(\boldsymbol{\omega}) - \boldsymbol{\xi})\, d\boldsymbol{\omega}. \tag{12}$$

2. ***J-Condition**: The conditional density of $\boldsymbol{\omega}$ given $\boldsymbol{\xi}$ is preserved,*

$$p(\boldsymbol{\omega}|\boldsymbol{\xi}) = \pi(\boldsymbol{\omega}|\boldsymbol{\xi}). \tag{13}$$

3. ***Probability kinematics update rule**: The posterior $p(\boldsymbol{\omega})$ is obtained as*

$$p(\boldsymbol{\omega}) = \int \pi(\boldsymbol{\omega} \mid \boldsymbol{\xi}) p(\boldsymbol{\xi})\, d\boldsymbol{\xi}. \tag{14}$$

4. ***Reference ratio update rule**: Equivalently, after applying Bayes' formula to (14),*

$$p(\boldsymbol{\omega}) = \int \frac{\delta(\boldsymbol{\xi} - \boldsymbol{\xi}(\boldsymbol{\omega}))\pi(\boldsymbol{\omega})}{\pi(\boldsymbol{\xi})} p(\boldsymbol{\xi})\, d\boldsymbol{\xi} \tag{15}$$

$$= \frac{p(\boldsymbol{\xi}(\boldsymbol{\omega}))}{\pi(\boldsymbol{\xi}(\boldsymbol{\omega}))} \pi(\boldsymbol{\omega}). \tag{16}$$

   The Dirac delta function arises from the fact that $p(\boldsymbol{\xi} \mid \boldsymbol{\omega}) = \delta(\boldsymbol{\xi} - \boldsymbol{\xi}(\boldsymbol{\omega}))$. Note that the integrand is zero when $\boldsymbol{\xi} \neq \boldsymbol{\xi}(\boldsymbol{\omega})$, and that (16) follows directly from the J-condition given by (13). This formulation of PK is useful when the conditional in (14) is unavailable.

***Consequence**: The posterior $p(\boldsymbol{\omega})$ has marginal $p(\boldsymbol{\xi})$ and preserves the conditional $\pi(\boldsymbol{\omega} \mid \boldsymbol{\xi})$.*

3.2 KBPs versus PMFs

KBPs can be understood as an (approximate) application of PK (Hamelryck et al., 2010; Mardia et al., 2011; Valentin et al., 2014; Hamelryck et al., 2015). Consider a prior distribution over protein structure, $\pi(\boldsymbol{X})$. Without any loss of generality, we leave out the conditioning on sequence. We are now given evidence $p(\boldsymbol{D})$ about pairwise distances between a subset of the coordinates, e.g., the $C\beta$ coordinates. Thus, the coarse-grained variable $\boldsymbol{D}$ induces a partition on the sample space of the fine-grained variable $\boldsymbol{X}$. These variables correspond to $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ in Definition 1, respectively. The PK update following (16) is then given by

$$p(\boldsymbol{X}) = \int \frac{\delta(\boldsymbol{D}(\boldsymbol{X}) - \boldsymbol{D})\pi(\boldsymbol{X})}{\pi(\boldsymbol{D})} p(\boldsymbol{D}) \mathrm{d}\boldsymbol{D} \tag{17}$$

$$= \frac{p(\boldsymbol{D}(\boldsymbol{X}))}{\pi(\boldsymbol{D}(\boldsymbol{X}))} \pi(\boldsymbol{X}). \tag{18}$$

Note that $\boldsymbol{D}$ is a random variable, while $\boldsymbol{D}(\boldsymbol{X})$ is its realization at $\boldsymbol{X}$. A minus-log transform recovers the KBP given by (9),

$$\Delta E(\boldsymbol{X}) = -\log\left(\frac{p(\boldsymbol{D}(\boldsymbol{X}))}{\pi(\boldsymbol{D}(\boldsymbol{X}))}\right) - \log\left(\pi(\boldsymbol{X})\right), \tag{19}$$

if we assume conditional independence of pairwise distances (see also Appendix E). The second term does not appear in standard KBP expressions, as it is typically implicitly accounted for by Monte Carlo sampling (Simons et al., 1997) or scoring of existing structures (Sippl, 1993b).

The difference between PK/KBPs and PMFs is now clear: while PMFs amount to a projection and marginalization of a given Boltzmann distribution of a single physical system as in (7), often formulated relative to a corresponding non-interacting system by subtracting a reference energy as in (8), PK/KBPs correspond to a generalized Bayesian update of a prior as in (14-16). Moreover, PK/KBPs and PMFs differ in producing distributions over fine- ($\boldsymbol{\omega}$) and coarse-grained ($\boldsymbol{\xi}$) variables, respectively, i.e.,

$$\text{PMFs: } p(\boldsymbol{\xi}) = \int \delta(\boldsymbol{\xi} - \boldsymbol{\xi}(\boldsymbol{\omega})) p(\boldsymbol{\omega}) \mathrm{d}\boldsymbol{\omega} \tag{20}$$

$$\text{PK/KBPs: } p(\boldsymbol{\omega}) = \int \frac{\delta(\boldsymbol{\xi} - \boldsymbol{\xi}(\boldsymbol{\omega}))\pi(\boldsymbol{\omega})}{\pi(\boldsymbol{\xi})} p(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} \tag{21}$$

We now show that AF1's potential can be understood from the perspective of PK.

4 AF1's POTENTIAL RE-EVALUATED

4.1 AF1's potential as an instance of PK

In this section, we present a reformulation of the AF1 potential in probabilistic terms. First, we reinterpret the potential over the dihedral angles as an empirical prior that is updated with soft evidence on pairwise distances. Building on previous work on probabilistic models of dihedral angles (Edgoose et al., 1998; Hamelryck et al., 2006; Boomsma et al., 2008; Lennox et al., 2009; Zhao et al., 2010; Lennox et al., 2010) based on directional statistics (Mardia and Jupp, 2009; Ley and Verdebout, 2017; Pewsey and García-Portugués, 2021; Mardia, 2025a,b), this prior is formulated as

$$\pi(\boldsymbol{\phi}, \boldsymbol{\psi}; \boldsymbol{\Theta}) = \prod_i \mathrm{vM}\left(\phi_i; \boldsymbol{\Theta}\right) \mathrm{vM}\left(\psi_i; \boldsymbol{\Theta}\right), \tag{22}$$

where the parameters are obtained from a deep neural network (NN), $\boldsymbol{\Theta} = \mathrm{NN}_{\mathrm{Dihedral}}(\boldsymbol{s}, \boldsymbol{M})$, and vM is the von Mises distribution. This circular distribution resembles a Gaussian distribution for large κ and is suitable for modeling dihedral angles in $(-\pi, \pi]$ (Mardia and Jupp, 2009). The prior over dihedral angles implies a matching prior over distances,[2]

$$\pi(\boldsymbol{D}; \boldsymbol{\Gamma}) \propto \prod_{i,j,i\neq j} p\left(D_{i,j}; \boldsymbol{\Gamma}_{i,j}\right), \tag{23}$$

where $\boldsymbol{\Gamma} = \mathrm{NN}_{\mathrm{Reference}}(L, \mathbb{I}_{\mathrm{Gly}}(\boldsymbol{s}))$.

The local prior will be updated with soft evidence on distances following the generalized Bayesian approach of PK. The updated probability distribution over the distances is,

$$p(\boldsymbol{D}; \boldsymbol{\Lambda}) \propto \prod_{i,j,i\neq j} p\left(D_{i,j}; \boldsymbol{\Lambda}_{i,j}\right), \tag{24}$$

where $\boldsymbol{\Lambda} = \mathrm{NN}_{\mathrm{Distance}}(\boldsymbol{s}, \boldsymbol{M})$. We now have all the elements in place to invoke the machinery of PK. Following PK's reference ratio update given by (16), which avoids the use of the unavailable conditional used in (14), the posterior is obtained as

$$p(\boldsymbol{\phi}, \boldsymbol{\psi}; \boldsymbol{\Theta}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}) \propto \frac{p(\boldsymbol{D}(\boldsymbol{\phi}, \boldsymbol{\psi}); \boldsymbol{\Lambda})}{\pi(\boldsymbol{D}(\boldsymbol{\phi}, \boldsymbol{\psi}); \boldsymbol{\Gamma})} \pi(\boldsymbol{\phi}, \boldsymbol{\psi}; \boldsymbol{\Theta}). \tag{25}$$

By construction, AF1 adopts the J-condition induced by a prior over dihedral angles (Boomsma et al., 2008; Valentin et al., 2014). Why is the PK update needed, if dihedral angles are in principle sufficient to determine the structure? The answer is well known: due

[2] We discuss the incorrect assumption of independent pairwise distances in Appendix E.

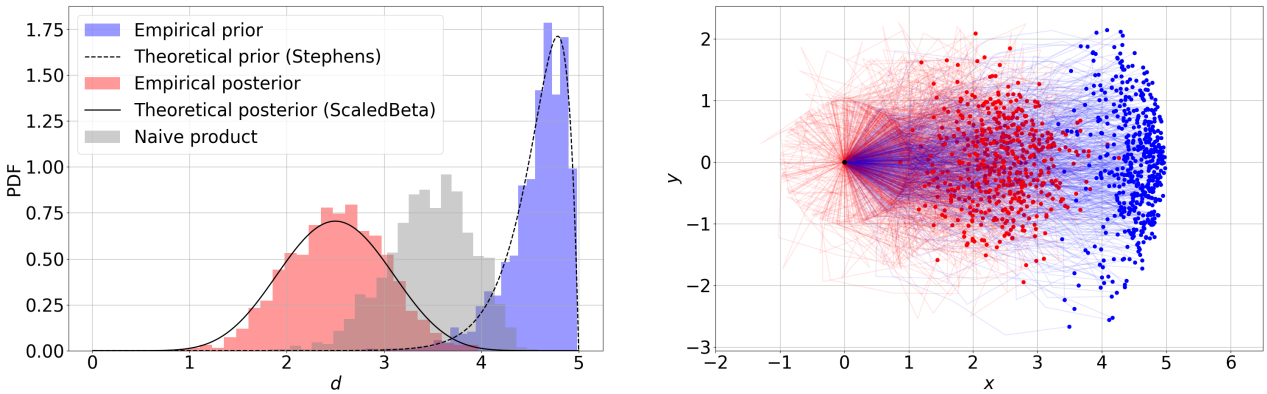


Figure 3: **(left)** Empirical PDFs of the origin-to-end distance, d, for the unmodified VRW prior (with $\mu = 0, \kappa = 10$, $N = 5$; blue histogram), the PK-updated VRW (red histogram), and a naive product of the ScaledBeta($\alpha = 10, \beta = 10, N = 5$) target distribution with the VRW prior, i.e., without dividing by the reference distribution (gray histogram). The latter does not fit the target distribution, clearly showing that the reference distribution is required. Also shown are the PDFs of the target ScaledBeta distribution (black line) and of the theoretical distribution for d (Stephens distribution; dashed line), showing an excellent fit in both cases. **(right)** 500 2D trajectories, $\boldsymbol{\chi}(\boldsymbol{\theta})$, sampled from the VRW prior (blue) and the PK-updated distribution (red). A dot denotes the end point of a random walk. As expected, the endpoints of the PK-updated distribution lie closer to the origin.

to an elbow effect, the dihedral prior is accurate for amino acids that are close together in the protein sequence, but becomes increasingly unreliable for amino acids that are distant (Holmes and Tsai, 2004; Gong et al., 2005). This *locality* is a well-known limitation of probabilistic models of proteins based on dihedral angles, and is due to the accumulation of small errors and improper covariance representation (Arts et al., 2024). In AF1, PK provides a minimal and coherent update of the prior, incorporating soft evidence on Euclidean distances while harnessing the prior's useful representation of local structure. The J-condition adopted by AF1 is further discussed in Appendix D. Some practical, approximate aspects of AF1's application of PK, including modeling the pairwise distances as conditionally independent and approximating the reference distribution, are discussed in Appendix E.

AF1 predictions are obtained by gradient descent minimization in the space of the dihedral angles, $\boldsymbol{\phi}, \boldsymbol{\psi}$. This approximates a maximum a posteriori (MAP) estimate obtained from the compositional empirical Bayes model constructed using PK. However, PK allows the formulation of well-defined posterior distributions involving angles, coordinates and distances, as we demonstrate in the next section using a tractable synthetic model.

5 EXPERIMENTS

Motivation To demonstrate PK in a way relevant to protein structure prediction, we formulate a synthetic model that by design has many crucial properties in common with the AF1 case (summarized in Appendix B). For both AF1 and the synthetic model, a generative model that specifies Euclidean coordinates based on an angular distribution is updated using distance evidence.

Our synthetic model is *deliberately minimal* to demonstrate the feasibility and precision of PK updates involving angles, coordinates and distances in a tractable model[3]. Our probabilistic framework and synthetic model are thus meant to provide a bedrock for applications to more complex systems in future work.

As prior, we use a von Mises random walk (VRW) in Euclidean 2D space. As coarse grained variable, we use the length of the *resultant vector*, that is, the distance between the origin and the end of the VRW, which has the advantage that its distribution is known (Mardia and Jupp, 2009; Stephens, 1969).

The similarities with AF1 include the following elements: (a) both models make use of the von Mises distribution for a prior over angles, (b) the angles sampled from the von Mises distributions specify Euclidean coordinates that can be interpreted as atom positions, (c) subsequent atomic positions are at a given, fixed distance from each other, (d) the prior is updated with evidence concerning distances between a subset of the atoms, (e) the conditional PDF of angles given distances is not available, requiring the use of PK's reference ratio update given by (16), and (f) the updated

[3]We provide an application to a small 3D peptide model, which requires estimation of the reference distribution, in Appendix F.

atomic positions are less spatially extended than the prior (see also Appendix D and Figure 5). Specifically, as we detail below, angle vector $\boldsymbol{\theta}$, 2D coordinates $\boldsymbol{\chi}$ and distance d of the synthetic model correspond to $(\boldsymbol{\phi}, \boldsymbol{\psi}), \boldsymbol{X}$ and $\boldsymbol{D}$ of AF1, respectively. The correspondences between AF1 and the synthetic model are summarized in Appendix B.

Setup Let $\boldsymbol{\chi}$ be an $(N+1)\times 2$ matrix of 2-dimensional coordinates, resulting from a VRW with N steps. The VRW with mean μ and concentration (i.e., inverse variance) κ is,

$$\begin{cases} \boldsymbol{\chi}_i = (0,0), & i = 0 \\ \boldsymbol{\chi}_i = \boldsymbol{\chi}_{i-1} + \boldsymbol{v}(\theta_i), & 0 < i \leq N, \end{cases} \tag{26}$$

where $\boldsymbol{\theta} = \{\theta_i \sim \mathrm{vM}(\mu, \kappa)\}_{i=1}^{N}$ are angles in $(-\pi, \pi]$ and $\boldsymbol{v}(\theta_i)$ is the corresponding 2D unit vector. As in AF1, the coordinates are fully determined by the angles; as before, we will write $\boldsymbol{\chi}(\boldsymbol{\theta})$ for the realization of $\boldsymbol{\chi}$ at $\boldsymbol{\theta}$. For this case, an approximate PDF of the Euclidean distance between start and end of the VRW, i.e., the length of the resultant vector, $d(\boldsymbol{\chi}) = \|\boldsymbol{\chi}_0 - \boldsymbol{\chi}_N\|$, is given by Stephens (1969),

$$2N\gamma\left(1 - \frac{d}{N}\right) \sim \mathrm{Chi}^2_{N-1}, \tag{27}$$

where $\frac{1}{\gamma} = \frac{1}{\kappa} + \frac{3}{8\kappa^2}$. For $\kappa \geqslant 4$, the approximation is excellent. We will write $d \sim \mathrm{Stephens}(d; \kappa, N)$. Note that the maximum value of d is N, since each step has unit length. We define the prior distribution $\pi_{\mathrm{VRW}}(\boldsymbol{\theta}; \mu, \kappa)$ as the generative model induced by the VRW. The marginal probability over d for this prior is

$$\begin{aligned} \pi_{\mathrm{VRW}}(d; \kappa, N) &= \mathrm{Stephens}(d; \kappa, N) \\ &= \int \delta\left(d - d(\boldsymbol{\theta})\right) \pi_{\mathrm{VRW}}(\boldsymbol{\theta}; \mu, \kappa, N) \mathrm{d}\boldsymbol{\theta}. \end{aligned} \tag{28}$$

Note that d is a random variable, while $d(\boldsymbol{\theta})$ is the realization of d for a specific VRW with angles $\boldsymbol{\theta}$. Now, we will demonstrate how PK allows us to update the prior using soft information on the distribution of d. We wish to modify the distribution over trajectories such that the marginal distribution of d/N follows a $\mathrm{Beta}(\alpha, \beta)$ distribution[4] with given parameters (α, β) instead of $\mathrm{Stephens}(d; \kappa)$. The *target distribution* over d thus becomes after applying the change of variable formula,

$$p(d; \alpha, \beta, N) = \frac{1}{N} \frac{1}{B(\alpha, \beta)} \left(\frac{d}{N}\right)^{\alpha-1} \left(1 - \frac{d}{N}\right)^{\beta-1}, \tag{29}$$

[4]Note that as $d \in [0, N]$,we divide by N to be within the support of the Beta distribution.

where $B(\alpha, \beta)$ is the standard normalization constant of the Beta distribution and $d \in [0, N]$. We will write $d \sim \mathrm{ScaledBeta}(d; \alpha, \beta, N)$.

We can apply PK to update the prior, following $p(\boldsymbol{\theta}) = \int \pi_{\mathrm{VRW}}(\boldsymbol{\theta} \mid d) p(d) \mathrm{d}\boldsymbol{\theta}$, but the marginal $\pi_{\mathrm{VRW}}(\boldsymbol{\theta} \mid d)$ is not available in a tractable form for a VRW. Therefore, as is the case for AF1, we use PK's reference ratio update given by (16), making use of the fact that the marginal over d according to the prior is given by (27). The PK update has the following form,

$$p(\boldsymbol{\theta}) = \frac{p\left(d(\boldsymbol{\theta})\right)}{\pi_{\mathrm{VRW}}\left(d(\boldsymbol{\theta})\right)} \pi_{\mathrm{VRW}}(\boldsymbol{\theta}), \tag{30}$$

which, after filling in the concrete PDFs, results in the final model,

$$\begin{aligned} &p(\boldsymbol{\theta}; \mu, \kappa, \alpha, \beta, N) = \\ &\quad \frac{\mathrm{ScaledBeta}\left(d(\boldsymbol{\theta}); \alpha, \beta, N\right)}{\mathrm{Stephens}\left(d(\boldsymbol{\theta}); \kappa, N\right)} \pi_{\mathrm{VRW}}(\boldsymbol{\theta}; \mu, \kappa, N). \end{aligned} \tag{31}$$

We sample from this model using the No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014), as implemented in the deep probabilistic programming language Pyro (Bingham et al., 2019), with default parameters. We use 1000 warm-up and 1000 posterior samples, taking under one minute on a standard laptop with an Intel i7 CPU. For code availability, see Appendix 8.

Results and evaluation We set $N = 5$ and $\mu = 0, \kappa = 10$ for the VRW, and $\mathrm{ScaledBeta}(\alpha = 10, \beta = 10, N = 5)$ as target. Figure 3 visually confirms that the correct posterior is attained. To further evaluate the model, we compared the empirical distributions of d to the target distribution using the one-sample Kolmogorov–Smirnov (KS) test (Virtanen et al., 2020), including every fifth NUTS sample to reduce autocorrelation. Across 10 NUTS simulations, the KS statistic ranges from 0.04 to 0.08, with a corresponding median p-value of 0.41, indicating a good match.

Ablation experiment In contrast and as expected, a naive reweighting approach that omits the reference distribution produced a much poorer fit, with the KS statistic ranging from 0.62 to 0.70, and $p < 10^{-70}$ in all cases. The deviation is also very obvious visually (Figure 3, gray histogram). This ablation experiment confirms the importance of including the reference distribution.

6 DISCUSSION AND RELATED WORK

KBPs (Sippl, 1990) were quantitatively explained by appealing to (16) by Hamelryck *et al.* (Hamelryck

et al., 2010; Mardia et al., 2011; Mardia and Hamelryck, 2012; Valentin et al., 2014), and later recognized as an instance of PK (Hamelryck et al., 2015). However, these references predate AF1, and are unconnected to deep learning. Moreover, the connection between KBPs and PK, while formally sound, has remained virtually unexplored within the deep learning paradigm applied to biomolecular structure. To our knowledge, our work is the first to analyze AF1 as an instance of PK.

Although PK is a well-defined concept in probability theory (Jeffrey, 1990; Diaconis and Zabell, 1982; Jeffrey, 2004; Chan and Darwiche, 2005; Meehan and Zhang, 2020), as far as we know, it has seen almost no explicit adoption in deep models: Yu (2020) applies PK to training Bayesian neural networks with soft evidence. Munk et al. (2023) use PK to incorporate soft evidence in probabilistic models and stochastic simulators.

This work connects statistical mechanics, probabilistic modeling, and deep learning, which are increasingly converging in the field of protein structure prediction (Durumeric et al., 2023; Jing et al., 2025; Olsson, 2026). The success of AF1 and its roots in PK hint at a general approach to formulate deep models that are compositional (Ghahramani, 2015; Bingham et al., 2019; Wood et al., 2014). This approach consists of updating an empirical Bayes prior concerning a fine-grained variable with soft evidence on a coarse-grained variable, both obtained from deep models. We point out that PK is agnostic regarding the nature of its components, bringing a wide range of deep generative models potentially within its scope (Tomczak, 2024).

Finally, probability kinematics has been connected to *credal set theory*, where uncertainty is represented by closed, convex sets of probability distributions rather than a single posterior (Marchetti and Antonucci, 2018; Caprio and Gong, 2023; Caprio and Mukherjee, 2023), providing opportunities for formulating deep generative models that can explicitly accommodate imprecise, partially conflicting or even inconsistent information.

7 CONCLUSIONS

We interpret AF1 as an instance of PK and demonstrated the probabilistic framework using a synthetic model with qualitative evaluation. Beyond clarifying the foundations of a paradigm-shifting method, our results imply that PK offers a principled foundation for building compositional deep probabilistic models based on generalized Bayesian updating (Ghahramani, 2015; Bingham et al., 2019; Wood et al., 2014).

8 CODE AVAILABILITY

Code for reproducing the synthetic von Mises random walk experiment in Section 5 is available at: `https://github.com/thamelry/AlphaFoldPK`.

9 ACKNOWLEDGMENTS

We thank Wouter Boomsma, Jesper Ferkinghoff-Borg, Jes Frellsen, Carsten Hartmann, John T. Kent, Ola Rønning, Manfred Sippl and Sandy Zabell for discussions and comments, and an anonymous reviewer for pointing out the connection to credal sets. We acknowledge funding from the VILLUM Experiment Programme (grant 50240) and the Independent Research Fund Denmark (grant 10.46540/5244-00031B).

References

J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold3. *Nature*, 630:493–500, 2024.

V. Agarwal and A. C. McShan. The power and pitfalls of AlphaFold2 for structure prediction beyond rigid globular proteins. *Nature Chemical Biology*, 20:950–959, 2024.

M. AlQuraishi. Parallelized natural extension reference frame: Parallelized conversion from internal to Cartesian coordinates. *Journal of Computational Chemistry*, 40:885–892, 2019.

B. Armendt. Is there a Dutch book argument for probability kinematics? *Philosophy of Science*, 47: 583–588, 1980.

M. Arts, J. Frellsen, and W. Boomsma. Internal-coordinate density modelling of protein structure: Covariance matters. *Transactions on Machine Learning Research*, 2024.

A. Ben-Naim. Statistical potentials extracted from protein structures: Are these meaningful potentials? *Journal of Chemical Physics*, 107:3698–3706, 1997.

E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall, and N. D. Goodman. Pyro: Deep universal probabilistic programming. *Journal of Machine Learning Research*, 20:1–6, 2019.

D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112:859–877, 2017.

W. Boomsma, K. Mardia, C. Taylor, J. Ferkinghoff-Borg, A. Krogh, and T. Hamelryck. A generative, probabilistic model of local protein structure. *Proceedings of the National Academy USA*, 105:8932–8937, 2008.

G. R. Bowman. AlphaFold and protein folding: Not dead yet! The frontier is conformational ensembles. *Annual Review of Biomedical Data Science*, 7:51–57, 2024.

E. Callaway. Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures. *Nature*, 634:525–526, 2024.

M. Caprio and R. Gong. Dynamic precise and imprecise probability kinematics. In *International Symposium on Imprecise Probability: Theories and Applications (ISIPTA)*, pages 72–83, 2023.

M. Caprio and S. Mukherjee. Ergodic theorems for dynamic imprecise probability kinematics. *International Journal of Approximate Reasoning*, 152:325–343, 2023.

D. Chakravarty, J. W. Schafer, E. A. Chen, J. F. Thole, L. A. Ronish, M. Lee, and L. L. Porter. AlphaFold predictions of fold-switched conformations are driven by structure memorization. *Nature Communications*, 15:7296, 2024.

D. Chakravarty, M. Lee, and L. L. Porter. Proteins with alternative folds reveal blind spots in AlphaFold-based protein structure prediction. *Current Opinion in Structural Biology*, 90:102973, 2025.

H. Chan and A. Darwiche. On the revision of probabilistic beliefs using uncertain evidence. *Artificial Intelligence*, 163:67–90, 2005.

D. Chandler. *Introduction to modern statistical mechanics*. Oxford University Press, 1987.

P. Clifford and N. Green. Distances in Gaussian point sets. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 97, pages 515–524. Cambridge University Press, 1985.

P. Diaconis and S. L. Zabell. Updating subjective probability. *Journal of the American Statistical Association*, 77:822–830, 1982.

K. Dill and J. MacCallum. The protein-folding problem, 50 years on. *Science*, 338:1042–1046, 2012.

I. L. Dryden and K. V. Mardia. *Statistical shape analysis with applications in R, Second Edition*. Wiley, 2016.

A. E. Durumeric, N. E. Charron, C. Templeton, F. Musil, K. Bonneau, A. S. Pasos-Trejo, Y. Chen, A. Kelkar, F. Noé, and C. Clementi. Machine learned coarse-grained protein force-fields: Are we there yet? *Current Opinion in Structural Biology*, 79:102533, 2023.

T. Edgoose, L. Allison, and D. Dowe. An MML classification of protein structure that knows about angles and sequence. In *Pacific Symposium on Biocomputing*, pages 585–596, 1998.

A. V. Finkelstein, A. Y. Badretdinov, and A. M. Gutin. Why do protein architectures have Boltzmann-like statistics? *Proteins: Structure, Function, and Bioinformatics*, 23:142–150, 1995.

J. Gavalda-Garcia, B. Dixit, A. Díaz, A. Ghysels, and W. Vranken. Gradations in protein dynamics captured by experimental NMR are not well represented by AlphaFold2 models and other computational metrics. *Journal of Molecular Biology*, 437:168900, 2025.

Z. Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 521:452–459, 2015.

H. Gong, P. J. Fleming, and G. D. Rose. Building native protein conformation from highly approximate backbone torsion angles. *Proceedings of the National Academy of Sciences*, 102:16227–16232, 2005.

T. Hamelryck, J. Kent, and A. Krogh. Sampling realistic protein conformations using local structural bias. *PLoS Computational Biology*, 2(9):e131, 2006.

T. Hamelryck, M. Borg, M. Paluszewski, J. Paulsen, J. Frellsen, C. Andreetta, W. Boomsma, S. Bottaro, and J. Ferkinghoff-Borg. Potentials of mean force for protein structure prediction vindicated, formalized and generalized. *PLoS ONE*, 5:e13714, 2010.

T. Hamelryck, K. V. Mardia, and J. Ferkinghoff-Borg, editors. *Bayesian methods in structural bioinformatics*. Springer, 2012.

T. Hamelryck, W. Boomsma, J. Ferkinghoff-Borg, J. Foldager, J. Frellsen, J. Haslett, and D. Theobald. Proteins, physics and probability kinematics: A Bayesian formulation of the protein folding problem. In I. Dryden and J. Kent, editors, *Geometry Driven Statistics*, pages 356–376. Wiley, 2015.

C. Hartmann, J. C. Latorre, and G. Ciccotti. On two possible definitions of the free energy for collective variables. *European Physical Journal Special Topics*, 200:73–89, 2011.

M. D. Hoffman and A. Gelman. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:1593–1623, 2014.

J. B. Holmes and J. Tsai. Some fundamental aspects of building protein structures from fragment libraries. *Protein Science*, 13:1636–1650, 2004.

R. C. Jeffrey. *Contributions to the theory of inductive probability*. PhD thesis, Princeton University, 1957.

R. C. Jeffrey. *The logic of decision*. The University of Chicago Press, 2nd edition, 1990. Originally published in 1965 by McGraw-Hill.

R. C. Jeffrey. *Subjective probability: The real thing*. Cambridge University Press, 2004.

B. Jing, B. Berger, and T. Jaakkola. AI-based methods for simulating, sampling, and predicting protein ensembles. *arXiv:2509.17224*, 2025.

J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596:583–589, 2021.

W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica A*, 32: 922–923, 1976.

D. P. Kingma and M. Welling. An introduction to variational autoencoders. *Foundations and Trends in Machine Learning*, 12:307–392, 2019.

J. G. Kirkwood. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics*, 3:300–313, 1935.

W. A. Koppensteiner and M. J. Sippl. Knowledge-based potentials – Back to the roots. *Biochemistry (Moscow)*, 63:247–252, 1998.

K. P. Lennox, D. B. Dahl, M. Vannucci, and J. W. Tsai. Density estimation for protein conformation angles using a bivariate von Mises distribution and Bayesian nonparametrics. *Journal of the American Statistical Association*, 104:586–596, 2009.

K. P. Lennox, D. B. Dahl, M. Vannucci, R. Day, and J. W. Tsai. A Dirichlet process mixture of hidden Markov models for protein structure prediction. *Annals of Applied Statistics*, 4:916, 2010.

C. Ley and T. Verdebout. *Modern directional statistics*. Chapman and Hall/CRC, 2017.

J. López-Sagaseta and A. Urdiciain. Severe deviation in protein fold prediction by advanced AI: A case study. *Scientific Reports*, 15:4778, 2025.

M. Lu, A. D. Dousis, and J. Ma. OPUS-PSP: An orientation-dependent statistical all-atom potential derived from side-chain packing. *Journal of Molecular Biology*, 376:288–301, 2008.

S. Marchetti and A. Antonucci. Imaginary kinematics. *arXiv:1808.00329*, 2018.

K. V. Mardia. Fisher's legacy of directional statistics, and beyond to statistics on manifolds. *Journal of Multivariate Analysis*, 207:105404, 2025a.

K. V. Mardia. Profile: Kanti Mardia. *Significance*, 22:36–38, 2025b.

K. V. Mardia and T. Hamelryck. On 'Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation' by P. Fearnhead and D. Prangle. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74:462–463, 2012. Discussion.

K. V. Mardia and P. E. Jupp. *Directional statistics*. John Wiley & Sons, 2009.

K. V. Mardia, J. Frellsen, M. Borg, J. Ferkinghoff-Borg, and T. Hamelryck. A statistical view on the reference ratio method. In *LASR 2011 Proceedings*, pages 55–61. Leeds University Press, 2011.

A. Meehan and S. Zhang. Jeffrey meets Kolmogorov: A general theory of conditioning. *Journal of Philosophical Logic*, 49:941–979, 2020.

S. Miyazawa and R. L. Jernigan. Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules*, 18:534–552, 1985.

A. Munk, A. Mead, and F. Wood. Uncertain evidence in probabilistic models and stochastic simulators. In *International Conference on Machine Learning (ICML)*, pages 25486–25500, 2023.

S. Olsson. Generative molecular dynamics. *Current Opinion in Structural Biology*, 96:103213, 2026.

J. Parsons, J. B. Holmes, J. M. Rojas, J. Tsai, and C. E. Strauss. Practical conversion from torsion space to Cartesian space for in silico protein synthesis. *Journal of Computational Chemistry*, 26:1063–1068, 2005.

A. Pewsey and E. García-Portugués. Recent advances in directional statistics. *TEST*, 30:1–58, 2021.

F. M. Pohl. Empirical protein energy maps. *Nature New Biology*, 234:277–279, 1971.

G. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7:95–99, 1963.

P. Ramasamy, J. Zuallaert, L. Martens, and W. F. Vranken. Assessing the relation between protein phosphorylation, AlphaFold3 models and conformational variability. *preprint bioRxiv*, 2025.

Schrödinger, LLC. The PyMOL molecular graphics system, Version 3.0.0. Available at `https://pymol.org/`, 2015.

A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. Nelson, A. Bridgland, et al. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). *Proteins: Structure, Function, and Bioinformatics*, 87:1141–1148, 2019.

A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. Nelson, A. Bridgland, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 577:706–710, 2020.

M.-Y. Shen and A. Sali. Statistical potential for assessment and prediction of protein structures. *Protein Science*, 15:2507–2524, 2006.

K. T. Simons, C. Kooperberg, E. Huang, and D. Baker. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of Molecular Biology*, 268:209–225, 1997.

M. J. Sippl. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *Journal of Molecular Biology*, 213:859–883, 1990.

M. J. Sippl. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *Journal of Computer-Aided Molecular Design*, 7:473–501, 1993a.

M. J. Sippl. Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*, 17:355–362, 1993b.

M. J. Sippl. Knowledge-based potentials for proteins. *Current Opinion in Structural Biology*, 5:229–235, 1995.

M. J. Sippl, M. Ortner, M. Jaritz, P. Lackner, and H. Flöckner. Helmholtz free energies of atom pair interactions in proteins. *Folding and Design*, 1:289–298, 1996.

M. A. Stephens. Tests for the von Mises distribution. *Biometrika*, 56:149–160, 1969.

S. Tanaka and H. A. Scheraga. Medium-and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules*, 9:945–950, 1976.

P. D. Thomas and K. A. Dill. Statistical potentials extracted from protein structures: How accurate are they? *Journal of Molecular Biology*, 257:457–469, 1996.

J. M. Tomczak. *Deep generative modeling*. Springer, 2nd edition, 2024.

J. B. Valentin, C. Andreetta, W. Boomsma, S. Bottaro, J. Ferkinghoff-Borg, J. Frellsen, K. V. Mardia, P. Tian, and T. Hamelryck. Formulation of probabilistic models of protein structure in atomic detail using the reference ratio method. *Proteins: Structure, Function, and Bioinformatics*, 82:288–299, 2014.

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272, 2020.

F. Wood, J. W. Meent, and V. Mansinghka. A new approach to probabilistic programming inference. In *Artificial Intelligence and Statistics (AISTATS)*, pages 1024–1032. PMLR, 2014.

E. Yu. Bayesian neural networks with soft evidence. *preprint arXiv:2010.09570*, 2020.

J. Zhang and Y. Zhang. A novel side-chain orientation dependent potential derived from random-walk reference state for protein fold selection and structure prediction. *PLoS ONE*, 5:e15386, 2010.

F. Zhao, J. Peng, J. Debartolo, K. Freed, T. Sosnick, and J. Xu. A probabilistic and continuous model of protein conformational space for template-free modeling. *Journal of Computional Biology*, 17:783–798, 2010.

H. Zhou and Y. Zhou. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Science*, 11:2714–2726, 2002.

D. M. Zuckerman. *Statistical physics of biomolecules: An introduction*. CRC press, 2010.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator, if your work uses existing assets. [Yes]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

(a) The full text of instructions given to participants and screenshots. [Not Applicable]

(b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

APPENDIX

A NOTATION AND ABBREVIATIONS

Symbol / Abbreviation	Meaning (and shape of arrays)
AF1	AlphaFold version 1
KBP	Knowledge Based Potential
MAP	Maximum A Posteriori estimate
MSA	Multiple Sequence Alignment
NN	Neural Network
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
PK	Probability Kinematics (Jeffrey conditioning)
PMF	Potential of Mean Force
$\mathrm{vM}(\phi; \mu, \kappa)$	von Mises distribution, PDF with support $(-\pi, \pi]$
VRW	von Mises Random Walk
$\pi(\cdot)$	Prior PDF
AF1	
L	Length (number of amino acids) of a protein
$\boldsymbol{s}$	Protein sequence, one-hot encoding, shape $L \times 20$
$\boldsymbol{M}$	MSA, shape $M \times L \times 20$ for M sequences
$\mathbb{I}_{\mathrm{Gly}}(\boldsymbol{s})$	Glycines in protein sequence, one-hot encoding, shape L
$\boldsymbol{X}$	Coordinates of protein backbone, shape $L \times 4 \times 3$
$\boldsymbol{X}(\boldsymbol{\phi}, \boldsymbol{\psi})$	Realization of $\boldsymbol{X}$ for given dihedral angles
$\boldsymbol{\phi}, \boldsymbol{\psi}$	Protein dihedral angles in $(-\pi, \pi]$, shape L each
$\boldsymbol{D}$	Cβ-Cβ distance matrix, shape $L \times L$
$\boldsymbol{D}(\boldsymbol{X}), \boldsymbol{D}(\boldsymbol{\phi}, \boldsymbol{\psi})$	Realization of $\boldsymbol{D}$ for given coordinates or angles
$\pi(\boldsymbol{\phi}, \boldsymbol{\psi}; \boldsymbol{\Theta})$	Prior over dihedral angles (von Mises)
$\pi(\boldsymbol{D}; \boldsymbol{\Gamma})$	Prior over pairwise distances
$p(\boldsymbol{D}; \boldsymbol{\Lambda})$	Target PDF of $\boldsymbol{D}$, used for PK update
Synthetic model (von Mises random walk, VRW)	
N	Number of random walk steps
$\boldsymbol{\chi}$	2D coordinates of VRW starting at $(0, 0)$, shape $(N+1) \times 2$
$\boldsymbol{\theta}$	Angles of a VRW, shape N
d	Origin-to-end distance of a VRW, $d \in \mathbb{R}^+$
$d(\boldsymbol{\chi}), d(\boldsymbol{\theta})$	Realization of d for given VRW 2D coordinates or angles
$\pi_{\mathrm{VRW}}(\boldsymbol{\theta}; \mu, \kappa, N)$	Prior over VRW angles (von Mises)
$\pi_{\mathrm{VRW}}(d; \kappa, N)$, $\mathrm{Stephens}(d; \kappa, N)$	Prior over VRW origin-to-end distance d
$\mathrm{ScaledBeta}(d; \alpha, \beta, N)$	Target PDF of d, used for PK update
PMF and PK summary	
$\boldsymbol{\omega}$	A *fine-grained* random variable (e.g., $\boldsymbol{X}$)
$\boldsymbol{\xi}$	A *coarse-grained* or *collective* random variable (e.g., $\boldsymbol{D}$)
$\boldsymbol{\xi}(\boldsymbol{\omega})$	Realization of $\boldsymbol{\xi}$ given $\boldsymbol{\omega}$ (e.g., $\boldsymbol{D}(\boldsymbol{X})$)
J-condition	Jeffrey condition, posterior $p(\boldsymbol{\omega} \mid \boldsymbol{\xi}) = \pi(\boldsymbol{\omega} \mid \boldsymbol{\xi})$
PMF; maps $p(\boldsymbol{\omega})$ to $p(\boldsymbol{\xi})$	$p(\boldsymbol{\xi}) = \int \delta(\boldsymbol{\xi} - \boldsymbol{\xi}(\boldsymbol{\omega})) p(\boldsymbol{\omega}) \mathrm{d}\boldsymbol{\omega}$
PK; updates $\pi(\boldsymbol{\omega})$ with $p(\boldsymbol{\xi})$	$p(\boldsymbol{\omega}) = \int \pi(\boldsymbol{\omega} \mid \boldsymbol{\xi}) p(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi}$
Reference ratio formulation of PK	$p(\boldsymbol{\omega}) = \int \frac{\delta(\boldsymbol{\xi} - \boldsymbol{\xi}(\boldsymbol{\omega})) \pi(\boldsymbol{\omega})}{\pi(\boldsymbol{\xi})} p(\boldsymbol{\xi}) \mathrm{d}\boldsymbol{\xi} = \frac{p(\boldsymbol{\xi}(\boldsymbol{\omega}))}{\pi(\boldsymbol{\xi}(\boldsymbol{\omega}))} \pi(\boldsymbol{\omega})$

B AF1 AND THE SYNTHETIC MODEL

Table 1 summarizes the conceptual correspondences between AF1 and our synthetic von Mises random walk (VRW) model. Both models share the same probabilistic structure: a prior distribution over (fine-grained) angular variables induces Euclidean coordinates, which in turn determine a (coarse-grained) distance variable. The map from angles to distances is many-to-one in both cases. In AF1, dihedral angles generate 3D coordinates, $\boldsymbol{X}$, and corresponding $C\beta - C\beta$ distances; in the synthetic model, VRW angles generate 2D coordinates and an origin-to-end distance or *resultant length*, d. This analogy ensures that the PK update plays the same role in both cases, i.e., an angular prior is updated with distance information, allowing us to study the probabilistic foundation of AF1 in a tractable and interpretable setting.

Table 1: Comparison of AF1 and the synthetic model.

AF1	**Synthetic model**	**Correspondence**
Dihedral angles, $(\boldsymbol{\phi}, \boldsymbol{\psi})$	VRW angles, $\boldsymbol{\theta}$	Prior over angles (von Mises)
3D atomic coordinates, $\boldsymbol{X}$	2D VRW coordinates, $\boldsymbol{\chi}$	Coordinates generated by angles
Pairwise distances, $\boldsymbol{D}$	Resultant length, d	PK update with distance information

C WHITWORTH'S HORSES

This simple application of PK is discussed by Diaconis and Zabell (Diaconis and Zabell, 1982). Three horses, A, B, and C, have prior winning probabilities $\pi(A) = 1/2$, $\pi(B) = 1/4$, and $\pi(C) = 1/4$, respectively. New evidence increases the probability of A winning from $\pi(A) = 1/2$ to $p(A) = 2/3$, without providing any further information on the other horses. Standard Bayesian conditioning cannot apply, as no specific event has occurred. Instead, PK prescribes an update that satisfies the new marginal, $p(A) = 2/3$ and $p(A^c) = 1/3$, while preserving the conditional probabilities, $\pi(B \mid A^c) = \pi(C \mid A^c) = 1/2$. This yields the updated posterior probabilities,

$$p(B) = \pi(B \mid A^c)p(A^c) = 1/2 \cdot 1/3 = 1/6, \tag{32}$$

$$p(C) = \pi(C \mid A^c)p(A^c) = 1/2 \cdot 1/3 = 1/6, \tag{33}$$

while previously $\pi(B) = \pi(C) = 1/4$.

D LOCAL NATURE OF AF1 PRIOR AND J-CONDITION

Motivation As the dihedral angles are sufficient to generate 3D structures, one might ask why it is necessary to add evidence on pairwise distances. It is well known that probabilistic models based on internal coordinates, such as dihedral angles, struggle to reproduce realistic distances between amino acids that are far apart in sequence position (Arts et al., 2024). This is due to the highly structured covariance induced by the chain geometry: small local perturbations in dihedral angles propagate in a correlated way through the backbone.

Experiments To illustrate this in the context of AF1, we sampled dihedral angles from the prior and reconstructed the corresponding 3D coordinates. AF1 code that can be executed for CASP13 targets is available from GitHub.[5] AF1 was applied to CASP13 target TS1019s2, corresponding to the PT-VENN domain-containing protein, Protein Data Bank Identifier 8EY4, chain B. For each amino acid, the prior is specified as probabilities over the Ramachandran plot, binned at $10^\circ \times 10^\circ$ resolution, obtained from a deep model (see Section 4) Senior et al. (2020). For each amino acid, we picked a random bin according to the specified probabilities, and set the dihedral angles to the bin's centroid. We then construct the 3D coordinates of the $N, C\alpha, C$ backbone from the sampled dihedral angles. The sampled structures were superimposed on the true structure using a standard algorithm Kabsch (1976); Dryden and Mardia (2016).

Results and discussion Averaging over 100 sampled 3D structures, we visualized the resulting $C\alpha - C\alpha$ distogram and compared it to that of the native structure (see Figure 4). The prior captures distances between residues that are close together in the sequence well (i.e., close to the diagonal in Figure 4), but fails to produce

[5] https://github.com/google-deepmind/deepmind-research/tree/master/alphafold_casp13

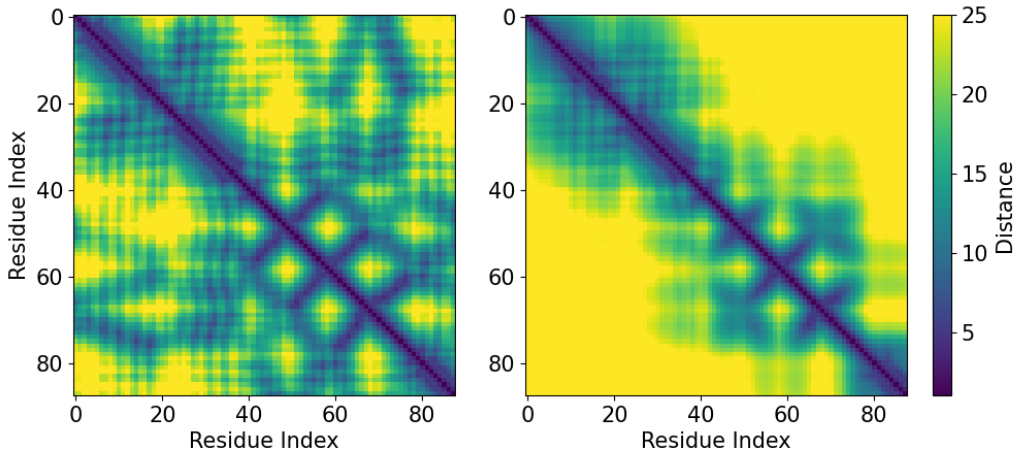


Figure 4: Distogram of the native structure (**left**) and the averaged distogram over 100 samples from the AF1 prior (**right**) for the PT-VENN domain-containing protein (Protein Data Bank Identifier 8EY4, Chain B). Distances are in Ångström. The prior accurately captures short-range distances between amino acids neighboring in sequence (i.e., distances close to the diagonal in the distogram on the right), but not long-range distances between amino acids far apart in the sequence, which necessitates updating the prior with distance information.

correct distances for the other cases, generating structures that are insufficiently compact. This is also visible in Figure 5, where the ensemble of 3D structures sampled from the AF1 prior deviates significantly from the compact, folded conformation. This result illustrates AF1's adopted J-condition, $p(\boldsymbol{\phi}, \boldsymbol{\psi} \mid \boldsymbol{D}) = \pi(\boldsymbol{\phi}, \boldsymbol{\psi} \mid \boldsymbol{D})$, and the corresponding PK update, which performs the minimal perturbation of the locally valid prior necessary to match the marginal distribution over the distances, $p(\boldsymbol{D})$.

E PRACTICAL ASPECTS OF THE AF1 POTENTIAL

Our statistical framework explains AF1's potential as the update of a prior over dihedral angles with information on pairwise distances. For practical reasons, the application of PK by AF1 is heuristic in a few respects, which we clarify here.

First, the reference distribution over the coarse grained variable, $\pi(\boldsymbol{D}(\boldsymbol{\phi}, \boldsymbol{\psi}); \boldsymbol{\Gamma})$, is not directly obtained from the prior over the fine-grained variable, $\pi(\boldsymbol{\phi}, \boldsymbol{\psi}; \boldsymbol{\Theta})$. Rather, it is approximated by training a model that uses as input only the sequence length, L, and the positions of the Glycine residues, $\mathbb{I}_{\text{Gly}}(\boldsymbol{s})$.

Second, AF1 does not use proper PDFs over the distance matrix, $\boldsymbol{D}$. For n points in $\mathbb{R}^d$, the $n(n-1)/2$ pairwise distances cannot vary independently: they must satisfy geometric constraints to correspond to a valid 3D configuration. As discussed by Clifford and Green (1985), valid distances lie on a smaller manifold of dimension $nd - d(d+1)/2$. By modeling probability distribution over $\boldsymbol{D}$ as a product of independent marginals, AF1 defines an approximate distribution that is *singular*, i.e., it may place nonzero density on geometrically invalid configurations.

These choices reflect pragmatic compromises that ultimately preserve the core principle of PK: updating a prior distribution concerning a fine-grained variable using soft evidence on a coarse-grained variable. In addition, minimization of the AF1 potential as a function of the dihedral angles ensures realizable 3D coordinates and a corresponding valid distance matrix.

F PROTEIN EXPERIMENT

Motivation To demonstrate that PK readily extends to realistic three–dimensional systems of relevance to protein structure prediction, we applied the framework to a model that represents a small protein consisting of eight amino acids. The experimental design closely parallels the two-dimensional von Mises random walk (VRW) of Section 5, but now uses dihedral angles describing the 3D coordinates of the protein backbone. As in AF1,

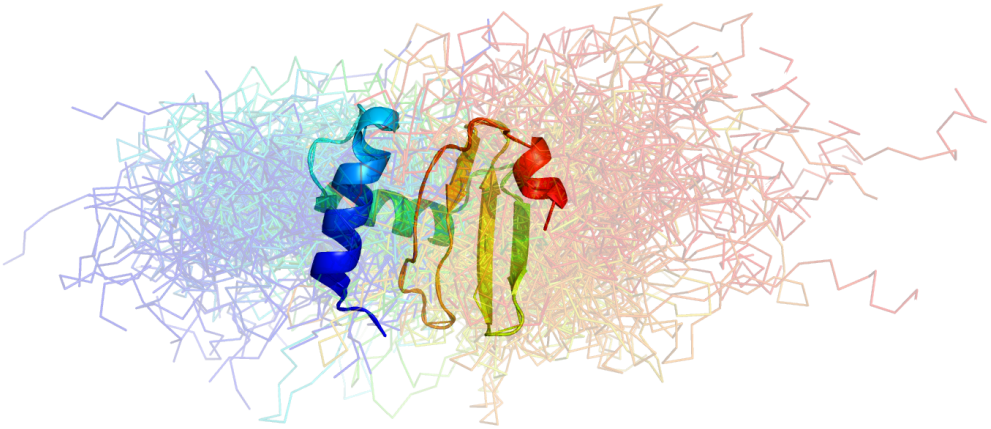


Figure 5: 100 samples from the AF1 prior (transparent, thin ribbons) superimposed on the cartoon representation of the native structure of the PT-VENN domain-containing protein (Protein Data Bank Identifier 8EY4, Chain B). In all cases, the N-terminus is shown in blue and C-terminus in red. Note how the samples from the prior are spatially extended compared to the compact native structure, which is representative for probabilistic models based on dihedral angles (Arts et al., 2024). Figure made with PyMol Schrödinger, LLC (2015).

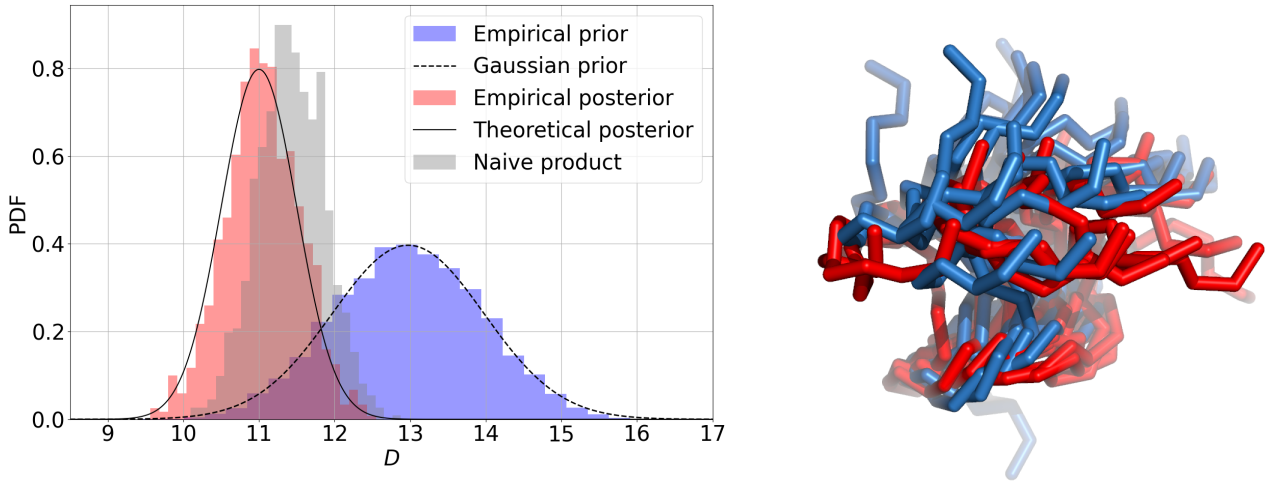


Figure 6: PK update applied to a 3D protein model (von Mises angular prior, eight amino acids). **(left)** Empirical PDFs of the $C\alpha - C\alpha$ start–to–end distance, D, for the unmodified angular prior (blue histogram), the PK–updated posterior (red histogram), and a naive product of the Gaussian target with the empirical prior, i.e., without dividing by the reference distribution (gray histogram). Also shown are the Gaussian distribution that approximates $\pi(D)$ (dashed black), estimated from prior samples, and the theoretical posterior (i.e., the desired target Gaussian distribution, $p(D)$; solid black), showing good agreement with the actual PK posterior. **(right)** 3D conformations from the prior (blue) and the PK–updated distribution (red), generated by the PNERF algorithm from the dihedral angles. Figure made with PyMol (Schrödinger, LLC, 2015).

a generative model that generates 3D Euclidean coordinates from (von Mises) angular priors is updated with evidence on inter-atomic distances. Recall that the 2D VRW synthetic model was chosen because the distance prior $\pi(d)$ associated with the angular prior is analytically available (see Section 5, Setup). This is not the case for the protein model. Thus, the distance prior must be estimated empirically from samples generated from the angular prior. Remarkably, a simple Gaussian fit to these samples provides a sufficient approximation to recover the correct posterior, highlighting the robustness of the PK formulation. In the next paragraph, we describe the protein model in detail.

Setup As before, let the $L \times 4 \times 3$ tensor $\boldsymbol{X}$ contain the atomic coordinates $(N, C\alpha, C, C\beta)$ of a protein backbone, and let $\boldsymbol{X}(\boldsymbol{\phi}, \boldsymbol{\psi})$ be the realization of $\boldsymbol{X}$ at $(\boldsymbol{\phi}, \boldsymbol{\psi})$. In this section, we will consider a protein consisting of eight amino acids (i.e., $L = 8$). Following AF1, bond lengths and bond angles are fixed to ideal values, and all peptide bonds were set to the *trans* conformation, without any loss of generality. Again following AF1, the prior over each dihedral angle follows a von Mises distribution. For our protein model, the prior over the dihedral angles of the i-th amino acid is

$$\phi_i \sim \mathrm{vM}(\mu_\phi, \kappa_\phi), \qquad \psi_i \sim \mathrm{vM}(\mu_\psi, \kappa_\psi), \tag{34}$$

where $\mu_\phi = -60°$, $\mu_\psi = -40°$, which corresponds to a canonical right-handed α-helix, and $\kappa_\phi = \kappa_\psi = 20$. This prior distribution corresponds to a probability distribution over the space of α-helix-like 3D conformations. As is the case for AF1, we will update this angular prior with pairwise distance information using PK. Concretely, we use PK to alter the PDF of the distance[6] D between the first and the last $C\alpha$ atoms,

$$D(\boldsymbol{\phi}, \boldsymbol{\psi}) = \|\mathbf{X}(\boldsymbol{\phi}, \boldsymbol{\psi})_{0,a} - \mathbf{X}(\boldsymbol{\phi}, \boldsymbol{\psi})_{-1,a}\|, \tag{35}$$

where $D(\boldsymbol{\phi}, \boldsymbol{\psi})$ is the realization of D at $(\boldsymbol{\phi}, \boldsymbol{\psi})$ and a is the index of the $C\alpha$ atom. In contrast to the tractable synthetic model based on the VRW, the induced prior distribution over the distance, $\pi(D)$, is not available for our protein model. Therefore, $\pi(D)$ was approximated by a Gaussian,

$$\pi(D) \approx \mathcal{N}(\mu_\pi, \sigma_\pi^2), \qquad \mu_\pi = 12.97 \text{ Å}, \ \sigma_\pi^2 = 1.02 \text{ Å}, \tag{36}$$

where (μ_π, σ_π) were robustly estimated (using the median and median absolute deviation, respectively) from 10^4 samples from the prior. Figure 6 (left) shows the histogram of the samples from the prior (blue) and the fitted Gaussian (dashed black line). As target distribution, we chose a Gaussian distribution that imposes a smaller mean distance and a sharper distribution,

$$p(D) = \mathcal{N}(\mu_D, \sigma_D^2), \qquad \mu_D = 11 \text{ Å}, \ \sigma_D^2 = 0.25 \text{ Å}. \tag{37}$$

As is the case for AF1 and the VRW, the conditional distribution, i.e., $p(\boldsymbol{\phi}, \boldsymbol{\psi} \mid D)$, is not available. Therefore, we again use the reference formulation of PK, which does not require the conditional, but uses a ratio of marginals. Following Eq. (16), the PK update is given by

$$p(\boldsymbol{\phi}, \boldsymbol{\psi}) = \frac{p(D(\boldsymbol{\phi}, \boldsymbol{\psi}))}{\pi(D(\boldsymbol{\phi}, \boldsymbol{\psi}))} \pi(\boldsymbol{\phi}, \boldsymbol{\psi}), \tag{38}$$

which after filling in the concrete PDFs (34, 36, 37) results in

$$p(\boldsymbol{\phi}, \boldsymbol{\psi}) = \frac{\mathcal{N}(D(\boldsymbol{\phi}, \boldsymbol{\psi}) \mid \mu_D, \sigma_D)}{\mathcal{N}(D(\boldsymbol{\phi}, \boldsymbol{\psi}) \mid \mu_\pi, \sigma_\pi)} \prod_{i=0}^{L-1} \mathrm{vM}(\phi_i \mid \mu_\phi, \kappa_\phi)\mathrm{vM}(\psi_i \mid \mu_\psi, \kappa_\psi). \tag{39}$$

Similarly to the VRW model, we sample the model using the No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014) as implemented in Pyro (Bingham et al., 2019). We use 1000 warm-up and 1000 posterior samples, and set the maximum tree depth to 6. 3D coordinates were calculated from the dihedral angles using the Parallelized Natural Extension Reference Frame (PNERF) algorithm (Parsons et al., 2005; AlQuraishi, 2019),[7] i.e., $\boldsymbol{X}(\boldsymbol{\phi}, \boldsymbol{\psi})$ in (35) is $\mathrm{PNERF}(\boldsymbol{\phi}, \boldsymbol{\psi})$. Execution takes under two minutes on a standard laptop with an Intel i7 CPU.

[6] We use a single distance (D) instead of all pairwise distances ($\boldsymbol{D}$) without any loss of generality because (a) this makes analysis easier and (b) as formulating non-singular PDFs over distance matrices is somewhat non-trivial (Clifford and Green, 1985), we leave this for future work. We use a capital letter despite the scalar nature of D to emphasize its conceptual correspondence to the distance matrix $\boldsymbol{D}$.

[7] We used the PyTorch implementation available from `https://github.com/FelixOpolka/pnerf-pytorch`.

Results and evaluation Figure 6 (left) compares the empirical distributions of D obtained under the prior (blue) and the posterior obtained from PK (red). The PK posterior visually coincides well with the imposed target distribution (Gaussian given by 37, black line). As in the VRW case, we further evaluate the fit between the posterior samples and the target distribution using the KS test, including every fifth NUTS sample to reduce autocorrelation. Across ten NUTS runs, the KS statistics range from 0.03 to 0.10, with a median p-value equal to 0.41, confirming a good fit.

Ablation experiment In contrast and as expected, omitting the reference distribution and using a naive product, i.e., $p(D(\boldsymbol{\phi}, \boldsymbol{\psi}))\pi(\boldsymbol{\phi}, \boldsymbol{\psi})$, produced a markedly poorer fit (gray histogram in Figure 6, left). The KS statistics ranged from 0.33 to 0.40 with $p < 10^{-18}$ in all cases, confirming the poor fit. As in the case of the VRW, the ablation study confirms the importance of the reference distribution.

Discussion This experiment demonstrates and confirms that the PK update is applicable to a protein model, despite the need for an estimation of the reference distribution: even though the prior marginal $\pi(D)$ is unavailable in closed form, a simple Gaussian approximation estimated from prior samples suffices to reproduce a matching posterior distribution.