Goodness-of-fit tests based on Wasserstein distance to detect changes on local protein conformations

Javier González-Delgado

Institut de Mathématiques de Toulouse, LAAS-CNRS

Journée toulousaine “Statistique pour la Biologie”

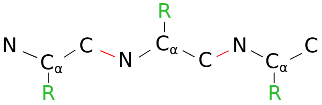
November 15, 2022

Joint work with

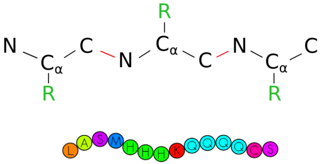
Pierre Neuvial[1] Juan Cortés[2] Pau Bernadó[3] Alberto González-Sanz[1,4]

1. Institut de Mathématiques de Toulouse, Université de Toulouse, CNRS, Toulouse, France.
2. LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France.
3. Centre de Biologie Structurale, Université de Montpellier, INSERM, CNRS, France.
4. ImUva, Universidad de Valladolid.

Proteins : sequence and conformation

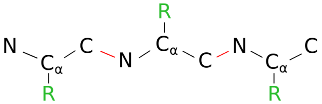


Proteins : sequence and conformation



Primary structure

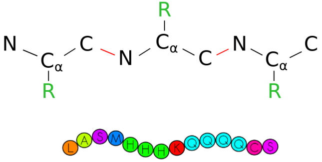
Proteins : sequence and conformation



Primary structure

Secondary structure (conformation)

Proteins : sequence and conformation

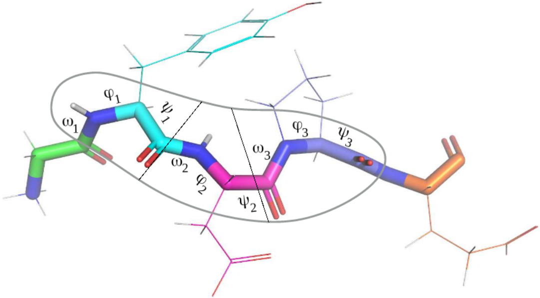


Primary structure

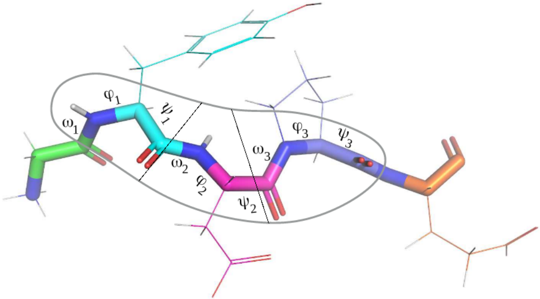
Secondary structure (conformation)

Sequence $\overset{?}{\Leftrightarrow}$ 3D structure $\Leftrightarrow$ Function

Local protein conformation



Local protein conformation

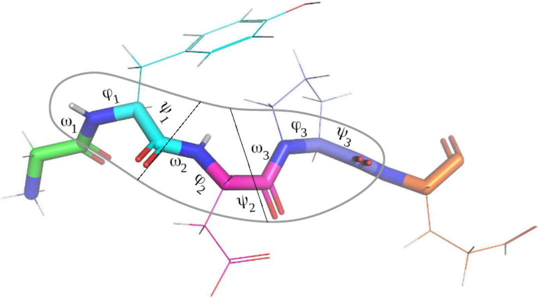


At each amino-acid, local conformation is given by the torsion dihedral angles [1]

- $\omega_i \in \{0, \pi\}$ is fixed,
- $(\varphi_i, \psi_i) \in [-\pi, \pi] \times [-\pi, \pi]$ determine local conformation.

1. G.N. Ramachandran et al. “Stereochemistry of polypeptide chain configurations”. In : *Journal of Molecular Biology* 7.1 (1963), p. 95-99

Local protein conformation

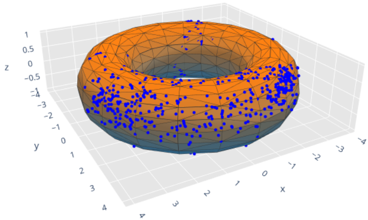


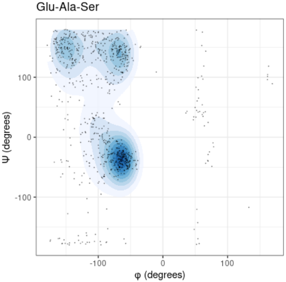
(φ, ψ) dihedral angles

- are physically restricted for globular (structured) proteins,
- are **random** for **Intrinsically Disordered Proteins** (IDP).

Local protein conformation

Local conformation is given by a **probability measure supported on $\mathbb{T}^2$**.





Goal

Detecting changes on **local protein conformations**

Goal

Detecting changes on **local protein conformations**

$\Updownarrow$

Detecting changes on **probability measures supported on** $\mathbb{T}^2$

Goal

Detecting changes on **local protein conformations**

$\Updownarrow$

Detecting changes on **probability measures supported on** $\mathbb{T}^2$

$\Updownarrow$

Goodness-of-fit testing for measures supported on $\mathbb{T}^2$

Goal

Detecting changes on **local protein conformations**

$$\Updownarrow$$

Detecting changes on **probability measures supported on** $\mathbb{T}^2$

$$\Updownarrow$$

Goodness-of-fit testing for measures supported on $\mathbb{T}^2$

First question : choice of a **metric** between distributions (statistic).

Optimal Transport Theory
Wasserstein distance

- C. VILLANI. *Topics in Optimal Transportation*. Providence, Rhode Island : American mathematical society, 2003
- C. VILLANI. *Optimal Transport : Old and New*. Springer-Verlag Berlin Heidelberg, 2008

Optimal Transport Theory

Wasserstein distance

- C. VILLANI. *Topics in Optimal Transportation*. Providence, Rhode Island : American mathematical society, 2003
- C. VILLANI. *Optimal Transport : Old and New*. Springer-Verlag Berlin Heidelberg, 2008

p-Wasserstein distance between two arbitrary measures

$$\mathcal{W}_p^p(\mu,\nu) = \min_{\pi\in\mathcal{U}(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)^p \mathrm{d}\pi(x,y) = \min_{(X,Y)} \left\{\mathbb{E}_{(X,Y)}(c(X,Y)^p) \,:\, X\sim\mu \;\; Y\sim\nu\right\}.$$

Two-sample goodness-of-fit test based on Wasserstein distance for measures on $\mathbb{R}^2/\mathbb{Z}^2$

Two-sample goodness-of-fit test based on Wasserstein distance for measures on $\mathbb{R}^2/\mathbb{Z}^2$

Goodness-of-fit test

Let $P, Q \in \mathcal{P}(\mathbb{R}^2/\mathbb{Z}^2)$ and P_n, Q_m the corresponding empirical probability measures.
The goal is to assess the null hypothesis :

$$H_0 \, : \, P = Q \tag{1}$$

by considering the p-value

$$\mathbb{P}_{H_0}(\mathcal{W}_p^p(P_n, Q_m) \geq w_{nm}), \tag{2}$$

where w_{nm} is the statistic realization.

Two-sample goodness-of-fit test based on Wasserstein distance for measures on $\mathbb{R}^2/\mathbb{Z}^2$

Goodness-of-fit test

Let $P, Q \in \mathcal{P}(\mathbb{R}^2/\mathbb{Z}^2)$ and P_n, Q_m the corresponding empirical probability measures. The goal is to assess the null hypothesis :

$$H_0 \, : \, P = Q \tag{1}$$

by considering the p-value

$$\mathbb{P}_{H_0}(\mathcal{W}_p^p(P_n, Q_m) \geq w_{nm}), \tag{2}$$

where w_{nm} is the statistic realization.

Problem

We need to know the null distribution of $\mathcal{W}_p^p(P_n, Q_m)$.

Two-sample goodness-of-fit test based on Wasserstein distance for measures on $\mathbb{R}^2/\mathbb{Z}^2$

Goodness-of-fit test

Let $P, Q \in \mathcal{P}(\mathbb{R}^2/\mathbb{Z}^2)$ and P_n, Q_m the corresponding empirical probability measures. The goal is to assess the null hypothesis :

$$H_0 \;:\; P = Q \tag{1}$$

by considering the p-value

$$\mathbb{P}_{H_0}(\mathcal{W}_p^p(P_n, Q_m) \geq w_{nm}), \tag{2}$$

where w_{nm} is the statistic realization.

Problem

We need to know the null distribution of $\mathcal{W}_p^p(P_n, Q_m)$.

Exact tests are unfeasible when dimension is higher than one

Distribution of $\mathcal{W}_p^p(P_n, Q_m)$ when $P = Q$ is unknown.

Testing the equality of N_g projections to closed geodesics[2]

2. J. González-Delgado et al. *Two-sample goodness-of-fit tests on the flat torus based on Wasserstein distance and their relevance to structural biology*. arXiv :2108.00165. 2021

Testing the equality of N_g projections to closed geodesics[2]

Strategy

For each $i \in \{1, \ldots, N_g\}$:

1. Project both samples to the i-th closed geodesic,
2. Test the equality of both projected samples : get the i-th p-value.

Finally : aggregate the N_g p-values.

2. J. González-Delgado et al. *Two-sample goodness-of-fit tests on the flat torus based on Wasserstein distance and their relevance to structural biology*. arXiv :2108.00165. 2021

Testing the equality of N_g projections to closed geodesics[2]

Strategy

For each $i \in \{1, \ldots, N_g\}$:

1. Project both samples to the i-th closed geodesic,
2. Test the equality of both projected samples : get the i-th p-value.

Finally : aggregate the N_g p-values.

Relevant points

- How to choose geodesics : deterministic/random (uniform ?) sampling,
- How to project samples to a given geodesic.

2. J. González-Delgado et al. *Two-sample goodness-of-fit tests on the flat torus based on Wasserstein distance and their relevance to structural biology*. arXiv :2108.00165. 2021

Geodesics on $\mathbb{T}^2$

Geodesics[3] on $\mathbb{T}^2$ are the images by the canonical projection of straight lines on $\mathbb{R}^2$. Lines with irrational slope map to geodesics which are dense on $\mathbb{T}^2$, and only lines with rational slope map to *closed* geodesics on the torus, which are **closed spirals isomorphic to $\mathbb{R}/\mathbb{Z}$**.

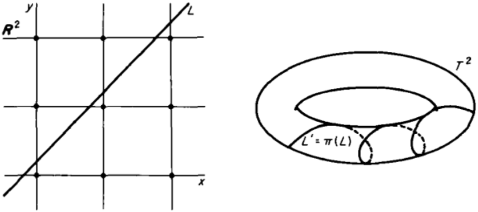


Figure : (Boothby, 1975)[3]

3. William M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Pure and Applied Mathematics. Academic Press, London, 1975

Two-sample goodness-of-fit test for measures on $\mathbb{R}/\mathbb{Z}$

Using Wasserstein distance as test statistic

Let $P^c, Q^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$ and P_n^c, Q_m^c be their corresponding empirical probability measures. We aim to test

$$H_0 : P^c = Q^c \qquad \text{against} \qquad H_1 : P^c \neq Q^c.$$

4. Julie Delon et al. "Fast transport optimization for Monge cost on the circle". In : *SIAM Journal on Applied Mathematics* 70.7/8 (2010), p. 2239-2258

Two-sample goodness-of-fit test for measures on $\mathbb{R}/\mathbb{Z}$

Using Wasserstein distance as test statistic

Let $P^c, Q^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$ and P_n^c, Q_m^c be their corresponding empirical probability measures. We aim to test

$$H_0 : P^c = Q^c \qquad \text{against} \qquad H_1 : P^c \neq Q^c.$$

If we denote by F, G the cumulative distribution functions of P^c, Q^c respectively, defined as $F(t) = P^c([0,t])$. Then, we have[4]

$$\mathcal{W}_2^2(P^c, Q^c) = \inf_{\alpha \in \mathbb{R}} \int_0^1 \left(F^{-1}(t) - (G - \alpha)^{-1}(t)\right)^2 \, dt, \tag{3}$$

where the pseudo-inverse is defined as $F^{-1}(s) = \inf\{t \, : \, F(t) > s\}$.

4. Julie Delon et al. "Fast transport optimization for Monge cost on the circle". In : *SIAM Journal on Applied Mathematics* 70.7/8 (2010), p. 2239-2258

Two-sample goodness-of-fit test for measures on $\mathbb{R}/\mathbb{Z}$

Using Wasserstein distance as test statistic

Let $P^c, Q^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$ and P_n^c, Q_m^c be their corresponding empirical probability measures. We aim to test

$$H_0 : P^c = Q^c \qquad \text{against} \qquad H_1 : P^c \neq Q^c.$$

If we denote by F, G the cumulative distribution functions of P^c, Q^c respectively, defined as $F(t) = P^c([0, t])$. Then, we have[4]

$$\mathcal{W}_2^2(P^c, Q^c) = \inf_{\alpha \in \mathbb{R}} \int_0^1 \left(F^{-1}(t) - (G - \alpha)^{-1}(t)\right)^2 \, dt, \tag{3}$$

where the pseudo-inverse is defined as $F^{-1}(s) = \inf\{t \, : \, F(t) > s\}$.

Consequence

The Optimal Transport problem on $\mathbb{R}/\mathbb{Z}$ reduces to the same problem on $[0, 1)$ if both measures are **relocated on the real line choosing as origin** the minimizing element α.

4. Julie Delon et al. "Fast transport optimization for Monge cost on the circle". In : *SIAM Journal on Applied Mathematics* 70.7/8 (2010), p. 2239-2258

Two-sample goodness-of-fit test for measures on $\mathbb{R}/\mathbb{Z}$

Using Wasserstein distance as test statistic

Let $P^c, Q^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$ and P^c_n, Q^c_m be their corresponding empirical probability measures. We aim to test

$$H_0 : P^c = Q^c \qquad \text{against} \qquad H_1 : P^c \neq Q^c.$$

If we denote by F, G the cumulative distribution functions of P^c, Q^c respectively, defined as $F(t) = P^c([0,t])$. Then, we have[4]

$$\mathcal{W}_2^2(P^c, Q^c) = \inf_{\alpha \in \mathbb{R}} \int_0^1 \left(F^{-1}(t) - (G - \alpha)^{-1}(t)\right)^2 \, dt, \tag{3}$$

where the pseudo-inverse is defined as $F^{-1}(s) = \inf\{t \,:\, F(t) > s\}$.

Consequence

The Optimal Transport problem on $\mathbb{R}/\mathbb{Z}$ reduces to the same problem on $[0,1)$ if both measures are **relocated on the real line choosing as origin** the minimizing element α.

But... $\mathcal{W}_2^2(P^c, Q^c)$ is not distribution-free under H_0.

4. Julie Delon et al. "Fast transport optimization for Monge cost on the circle". In : *SIAM Journal on Applied Mathematics* 70.7/8 (2010), p. 2239-2258

Two-sample goodness-of-fit test for measures on $\mathbb{R}/\mathbb{Z}$

Finding a distribution-free statistic (I)

Adapting the idea from Ramdas et. al[5] to $\mathbb{R}/\mathbb{Z}$

Instead of comparing F to G, compare $G(F^{-1})$ to a uniform distribution.

5. Aaditya Ramdas et al. "On Wasserstein Two Sample Testing and Related Families of Nonparametric Tests". In : *Entropy* 19 (sept. 2015). DOI : 10.3390/e19020047

Two-sample goodness-of-fit test for measures on $\mathbb{R}/\mathbb{Z}$

Finding a distribution-free statistic (I)

Adapting the idea from Ramdas et. al[5] to $\mathbb{R}/\mathbb{Z}$

Instead of comparing F to G, compare $G(F^{-1})$ to a uniform distribution.

Two advantages :

- Explicit form of the minimizer α,
- Distribution-free statistic under H_0.

5. Aaditya Ramdas et al. "On Wasserstein Two Sample Testing and Related Families of Nonparametric Tests". In : *Entropy* 19 (sept. 2015). DOI : 10.3390/e19020047

Two-sample goodness-of-fit test for measures on $\mathbb{R}/\mathbb{Z}$

Finding a distribution-free statistic (I)

Adapting the idea from Ramdas et. al[5] to $\mathbb{R}/\mathbb{Z}$

Instead of comparing F to G, compare $G(F^{-1})$ to a uniform distribution.

Two advantages :

- Explicit form of the minimizer α,
- Distribution-free statistic under H_0.

Lemma (Explicit form of α)

Let $P^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$, and F be its cumulative distribution function. Let U be the uniform distribution on $\mathbb{R}/\mathbb{Z}$. Then,

$$\mathcal{W}_2^2(P^c, U) = \int_0^1 \left(F^{-1}(t) - t - \alpha_0(F)\right)^2 \, dt,$$

where the optimal origin is given by $\alpha_0(F) = \int_0^1 (F^{-1}(t) - t)\, dt$.

5. Aaditya Ramdas et al. "On Wasserstein Two Sample Testing and Related Families of Nonparametric Tests". In : *Entropy* 19 (sept. 2015). DOI : 10.3390/e19020047

Two-sample goodness-of-fit test for measures on $\mathbb{R}/\mathbb{Z}$

Finding a distribution-free statistic (II)

We consider the statistic

$$T^c_{nm} = \frac{nm}{n+m}\mathcal{W}_2^2(G_m \# P^c_n, U) = \frac{nm}{n+m}\int_0^1 \left(G_m(F_n^{-1}(t)) - t - \alpha_0(F_n^{-1}(G_m))\right)^2 dt,$$

where $G_m \# P^c_n$ is the *push-forward* measure of P^c_n through G_m, and whose CDF is $G_m(F_n^{-1})$.

Two-sample goodness-of-fit test for measures on $\mathbb{R}/\mathbb{Z}$

Finding a distribution-free statistic (II)

We consider the statistic

$$T^c_{nm} = \frac{nm}{n+m}\mathcal{W}_2^2(G_m\#P_n^c, U) = \frac{nm}{n+m}\int_0^1 \left(G_m(F_n^{-1}(t)) - t - \alpha_0(F_n^{-1}(G_m))\right)^2 dt,$$

where $G_m\#P_n^c$ is the *push-forward* measure of P_n^c through G_m, and whose CDF is $G_m(F_n^{-1})$.

Proposition (Distribution free under H_0)

Let $P^c, Q^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$, P_n^c, Q_m^c be their corresponding empirical probability measures, and F_n, G_m be their empirical cumulative distribution functions. If $\frac{n}{m} \to \lambda$ when $n, m \to \infty$ for some $\lambda \in [0, \infty)$ then, under $P^c = Q^c$, it holds that

$$T^c_{nm} = \frac{nm}{n+m}\mathcal{W}_2^2(G_m\#P_n^c, U) \xrightarrow[n,m]{w} \int_0^1 \mathbb{B}(t)^2\, dt - \left(\int_0^1 \mathbb{B}(t)\, dt\right)^2,$$

where $\mathbb{B}$ is a standard Brownian bridge, and the weak convergence is understood as convergence of probability measures on the space of right-continuous functions with left limits.

Two-sample goodness-of-fit test for measures on $\mathbb{R}/\mathbb{Z}$

Test definition

Two-sample test for measures on $\mathbb{R}/\mathbb{Z}$

$$\pi^c_{nm} = \begin{cases} 1 & \text{if} \qquad T^c_{nm} \geq c^c_{nm}(\alpha) \\ 0 & \text{otherwise} \end{cases}$$

where the critical value $c^c_{nm}(\alpha)$ is given by

$$c^c_{nm}(\alpha) = \inf \{ t > 0 \, : \, F^c_{nm}(t) \geq 1 - \alpha \} \ ,$$

with F^c_{nm} denoting the distribution function of T^c_{nm} under H_0. Equivalently, a p-value for this test is $p^c_{nm} = 1 - F^c_{nm}(T^c_{nm})$.

Two-sample goodness-of-fit test for measures on $\mathbb{R}/\mathbb{Z}$

Test definition

Two-sample test for measures on $\mathbb{R}/\mathbb{Z}$

$$\pi^c_{nm} = \begin{cases} 1 & \text{if} \qquad T^c_{nm} \geq c^c_{nm}(\alpha) \\ 0 & \text{otherwise} \end{cases}$$

where the critical value $c^c_{nm}(\alpha)$ is given by

$$c^c_{nm}(\alpha) = \inf \left\{ t > 0 \,:\, F^c_{nm}(t) \geq 1 - \alpha \right\} ,$$

with F^c_{nm} denoting the distribution function of T^c_{nm} under H_0. Equivalently, a p-value for this test is $p^c_{nm} = 1 - F^c_{nm}(T^c_{nm})$.

Proposition (Consistency)

Let $P^c, Q^c \in \mathcal{P}(\mathbb{R}/\mathbb{Z})$. If $P^c \neq Q^c$, it holds

$$\lim_{n,m\to\infty} \mathbb{P}\left(\pi^c_{nm} = 1\right) = 1 \quad \text{for any } \alpha > 0.$$

Aggregating N_g p-values

Let p_i be the i-th p-value, for $i = 1, \ldots, N_g$. Then, under H_0 and assuming that the p-values are independent, the statistic

$$T^g_{nm,N_g} = \min_{i=1}^{N_g} p_i$$

follows a $\beta(1, N_g)$ distribution.

Aggregating N_g p-values

Let p_i be the i-th p-value, for $i = 1, \ldots, N_g$. Then, under H_0 and assuming that the p-values are independent, the statistic

$$T^g_{nm,N_g} = \min_{i=1}^{N_g} p_i$$

follows a $\beta(1, N_g)$ distribution. Consequently, the random variable

$$F_{\beta(1,N_g)}\left(T^g_{nm,N_g}\right),$$

where $F_{\beta(1,N_g)}$ denotes the $\beta(1, N_g)$ CDF, follows a uniform distribution on $[0, 1]$.

Aggregating N_g p-values

Let p_i be the i-th p-value, for $i = 1, \ldots, N_g$. Then, under H_0 and assuming that the p-values are independent, the statistic

$$T^g_{nm,N_g} = \min_{i=1}^{N_g} p_i$$

follows a $\beta(1, N_g)$ distribution. Consequently, the random variable

$$F_{\beta(1,N_g)}\left(T^g_{nm,N_g}\right),$$

where $F_{\beta(1,N_g)}$ denotes the $\beta(1, N_g)$ CDF, follows a uniform distribution on $[0, 1]$.

Two-sample test for measures on $\mathbb{T}^2$

$$\pi^g_{nm,N_g} = \begin{cases} 1 & \text{if} \quad F_{\beta(1,N_g)}\left(T^g_{nm,N_g}\right) \leq \alpha \\ 0 & \text{otherwise} \end{cases} \qquad (N_g\text{-geod})$$

Aggregating N_g p-values

Let p_i be the i-th p-value, for $i = 1, \dots, N_g$. Then, under H_0 and assuming that the p-values are independent, the statistic

$$T^g_{nm,N_g} = \min_{i=1}^{N_g} p_i$$

follows a $\beta(1, N_g)$ distribution. Consequently, the random variable

$$F_{\beta(1,N_g)}\left(T^g_{nm,N_g}\right),$$

where $F_{\beta(1,N_g)}$ denotes the $\beta(1, N_g)$ CDF, follows a uniform distribution on $[0, 1]$.

Two-sample test for measures on $\mathbb{T}^2$

$$\pi^g_{nm,N_g} = \begin{cases} 1 & \text{if} \quad F_{\beta(1,N_g)}\left(T^g_{nm,N_g}\right) \leq \alpha \\ 0 & \text{otherwise} \end{cases} \qquad (N_g\text{-geod})$$

Proposition (Consistency)

Let $P, Q \in \mathcal{P}(\mathbb{T}^2)$ and P^c_i (resp. Q^c_i), $i = 1, \dots, N_g$, be the circular projected distributions of P (resp. Q) to N_g closed geodesics of $\mathbb{T}^2$. If $P^c_i \neq Q^c_i$ for at least one $i \in \{1, \dots, N_g\}$, it holds

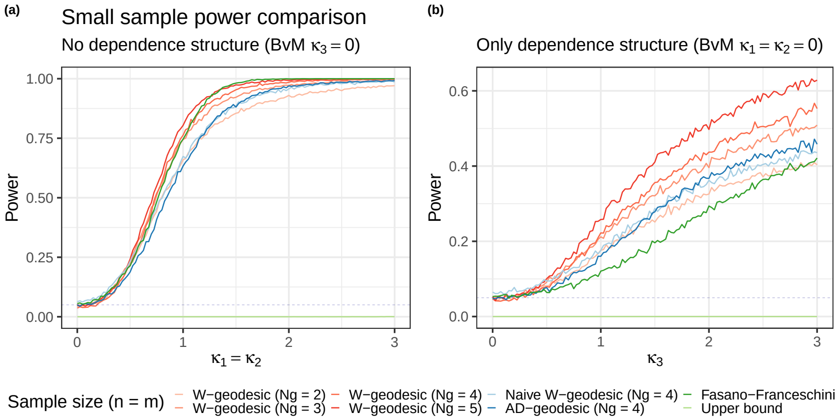
$$\lim_{n,m\to\infty} \mathbb{P}\left(\pi^g_{nm,N_g} = 1\right) \quad \text{for any } \alpha > 0.$$

Power analysis

Alternatives converging to the null (Uniform on $\mathbb{T}^2$ vs. bivariate von Mises converging to uniform)

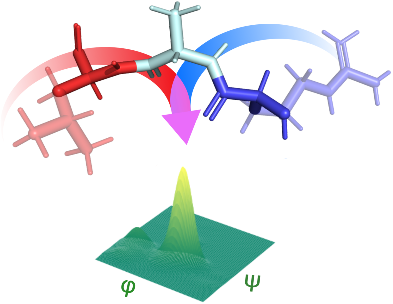
Power analysis

Alternatives converging to the null (Uniform on $\mathbb{T}^2$ vs. bivariate von Mises converging to uniform)



Applications (I)

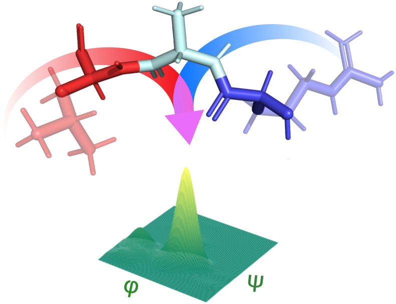
Effect of neighboring amino-acids on (ϕ, ψ) distribution[6]



6. Javier González-Delgado et al. "Statistical proofs of the interdependence between nearest neighbor effects on polypeptide backbone conformations". In : *Journal of Structural Biology* 214.4 (2022), p. 107907

Applications (I)

Effect of neighboring amino-acids on (ϕ, ψ) distribution[6]



H_0 : Flory's Isolated Pair hypothesis (IPH)[7]

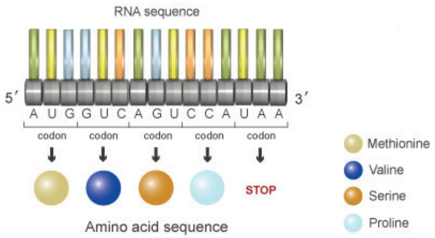
The identities of left and right amino-acids do not have an effect on (ϕ, ψ) distribution

6. Javier González-Delgado et al. "Statistical proofs of the interdependence between nearest neighbor effects on polypeptide backbone conformations". In : *Journal of Structural Biology* 214.4 (2022), p. 107907

7. Paul. J. Flory et al. "Statistical mechanics of chain molecules". In : *Biopolymers* 8.5 (1969), p. 699-700

Applications (II)

Effect of translated codon on (ϕ, ψ) distribution[8]

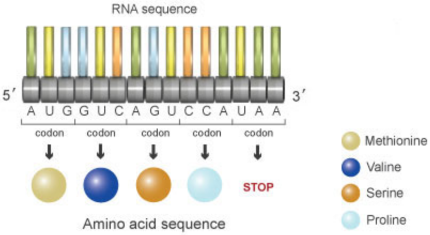


8. J. González-Delgado et al. *Statistical tests to detect differences between codon-specific Ramachandran plots*. Submitted. 2022

9. Image : https ://www.nature.com/scitable/topicpage/the-information-in-dna-determines-cellular-function-6523228/

Applications (II)

Effect of translated codon on (ϕ, ψ) distribution[8]



Question

Does the identity of the translated codon have an effect on (ϕ, ψ) distribution ?

8. J. González-Delgado et al. *Statistical tests to detect differences between codon-specific Ramachandran plots*. Submitted. 2022

9. Image : https ://www.nature.com/scitable/topicpage/the-information-in-dna-determines-cellular-function-6523228/

Applications (II)

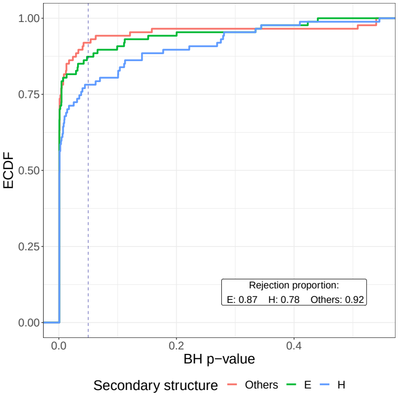
The identity of the translated codon has an effect on (ϕ, ψ) distribution

- Comparison of synonymous **codon-specific** (ϕ, ψ) distributions, considering **three** different types of **secondary structures** (α-helixes, β-sheets, others).

Applications (II)

The identity of the translated codon has an effect on (ϕ, ψ) distribution

- Comparison of synonymous **codon-specific** (ϕ, ψ) distributions, considering **three** different types of **secondary structures** (α-helixes, β-sheets, others).



Some other applications

- Detect the effect of a **sequence mutation** on secondary structure (and thus in function)

Some other applications

- Detect the effect of a **sequence mutation** on secondary structure (and thus in function)
- Assess the quality/convergence of **Molecular Dynamics simulations**

Some other applications

- Detect the effect of a **sequence mutation** on secondary structure (and thus in function)
- Assess the quality/convergence of **Molecular Dynamics simulations**
- Compare **entire sequences** and detect outlier amino-acids

Some other applications

- Detect the effect of a **sequence mutation** on secondary structure (and thus in function)
- Assess the quality/convergence of **Molecular Dynamics simulations**
- Compare **entire sequences** and detect outlier amino-acids
- Understand how to **design** a protein performing a specific **function**

Some other applications

- Detect the effect of a **sequence mutation** on secondary structure (and thus in function)
- Assess the quality/convergence of **Molecular Dynamics simulations**
- Compare **entire sequences** and detect outlier amino-acids
- Understand how to **design** a protein performing a specific **function**

Thank you for your attention !