Modèles génératifs pour les séquences de protéines

Generative models for protein sequences

Marion Chauveau

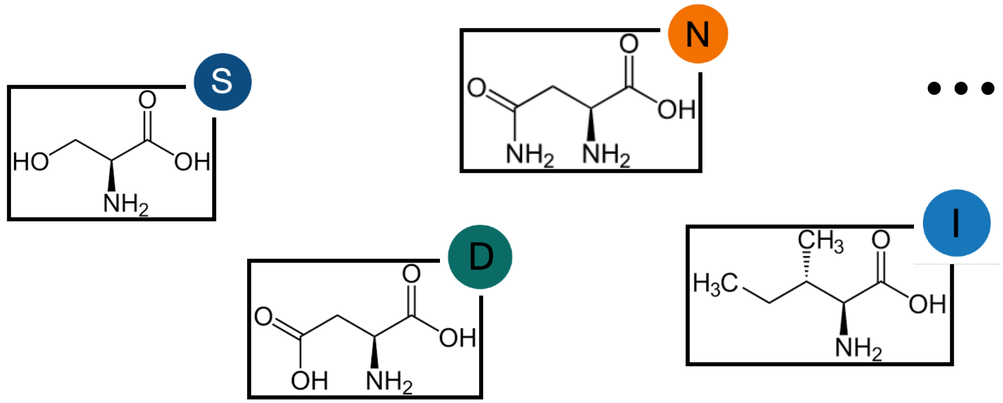
Thesis advisors: Ivan Junier and Olivier Rivoire

PhD Thesis Defense | October 1st, 2025 | ESPCI Paris

Proteins

The basics

- 20 building blocks: amino acids

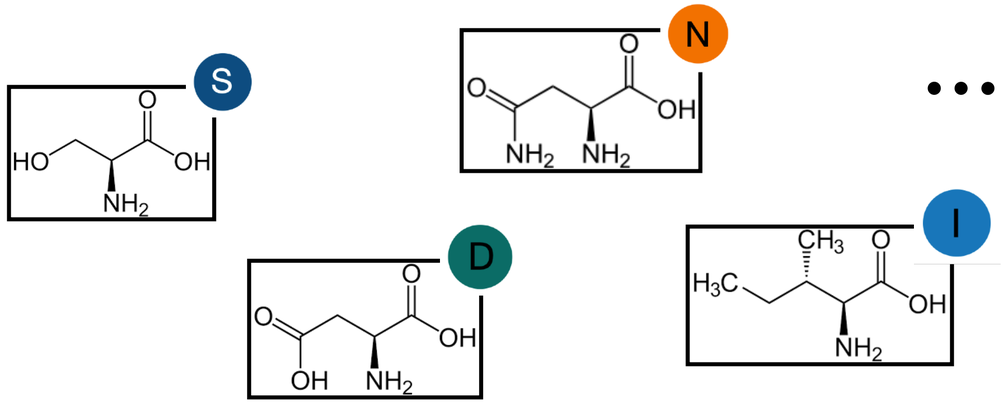


Alberts *et al.*, **2022**

Proteins

The basics

- 20 building blocks: amino acids
- Assembled into chains
- Sequence encoded in the DNA



A protein sequence

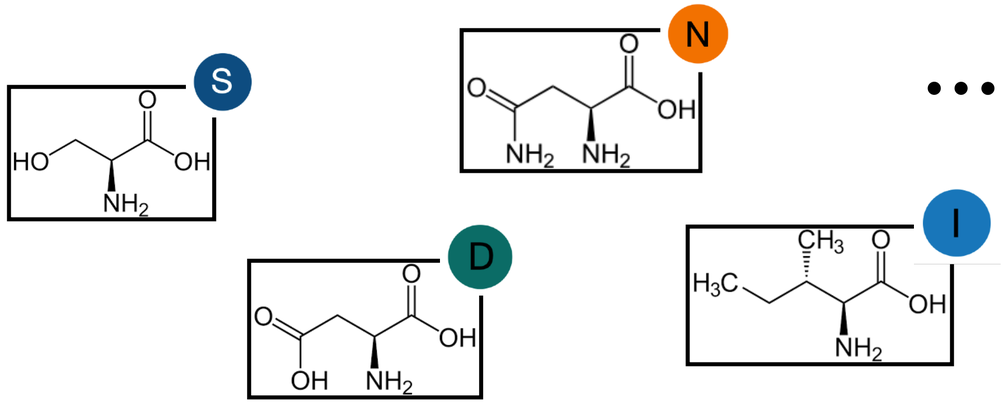
IVGGYTCQ ••• CNYVDWIQ

Alberts *et al.*, **2022**

Proteins

The basics

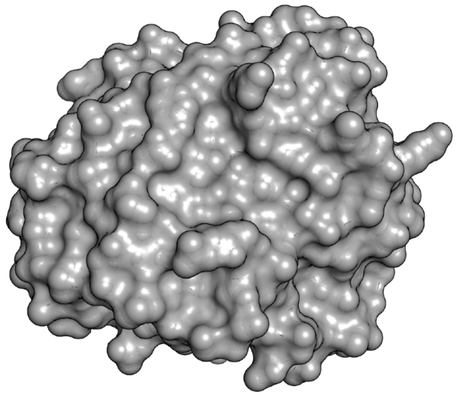
- 20 building blocks: amino acids
- Assembled into chains
- Sequence encoded in the DNA
- Non-trivially ordered matter
- Life's essential machines (catalysis, transport, signaling, immune defense, structure…)



A protein sequence

IVGGYTCQ ••• CNYVDWIQ

A protein structure representation



Alberts *et al.*, **2022**

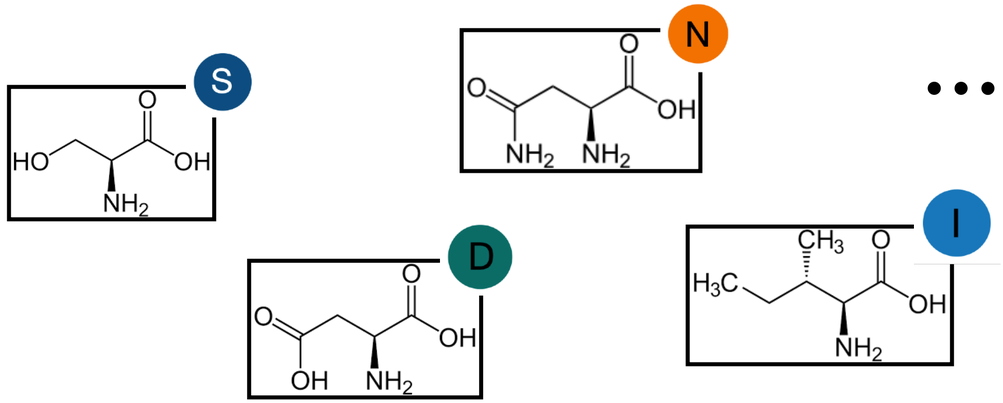
Proteins

The basics

- 20 building blocks: amino acids
- Assembled into chains
- Sequence encoded in the DNA
- Non-trivially ordered matter
- Life's essential machines (catalysis, transport, signaling, immune defense, structure…)

Performance & adaptability

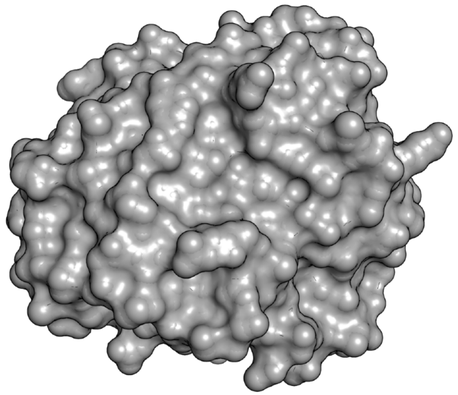
- High catalytic efficiency (enzymes)
- Specific interaction with target molecules
- Capacity to evolve new functions



A protein sequence

IVGGYTCQ ••• CNYVDWIQ

A protein structure representation



Alberts *et al.*, **2022**

Proteins

The basics

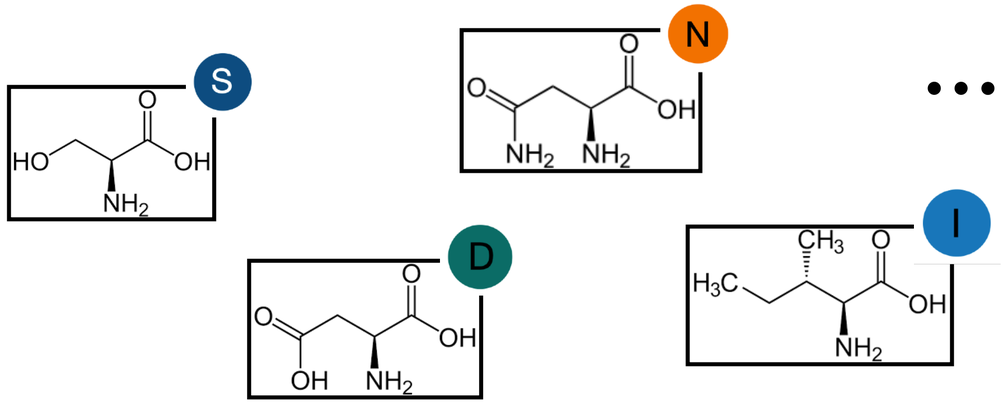
- 20 building blocks: amino acids
- Assembled into chains
- Sequence encoded in the DNA
- Non-trivially ordered matter
- Life's essential machines (catalysis, transport, signaling, immune defense, structure…)

Performance & adaptability

- High catalytic efficiency (enzymes)
- Specific interaction with target molecules
- Capacity to evolve new functions

Why are proteins worth studying?

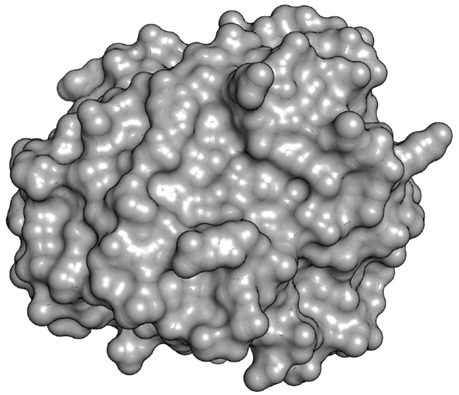
- Understand how proteins work
- Applications in medicine, industry…



A protein sequence

IVGGYTCQ ••• CNYVDWIQ

A protein structure representation

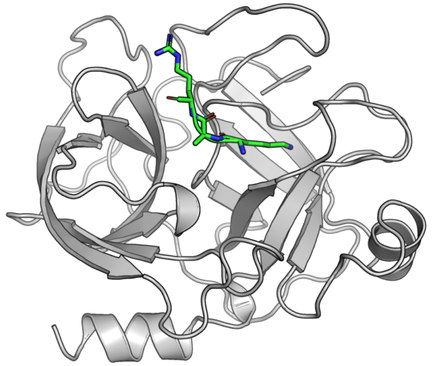


Alberts *et al.*, **2022**

Study proteins by learning from evolution

Different approaches

- Observe their structure (X-ray crystallography…)
- Simulate them using structure (Molecular Dynamics…)
- Modify them (mutational scans…)
- Learn from evolution (statistical models…) Etc…



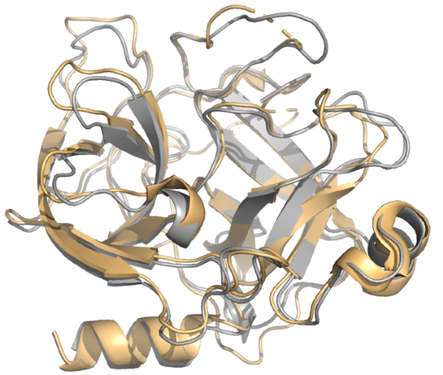
Rat Trypsin structure

Kendrew *et al.,* Nature, **1958**
C Edgar *et al., Current opinion in structural biology,* **2006**
M Fowler *et al.,* Nature methods, **2014**
Cocco *et al.,* Rep. Prog. Phys., **2018**

Study proteins by learning from evolution

Different approaches

- Observe their structure (X-ray crystallography…)
- Simulate them using structure (Molecular Dynamics…)
- Modify them (mutational scans…)
- Learn from evolution (statistical models…) Etc…




IVGGYTCQ ••• CNYVDWIQ


IVGGR--R ••• AQFVNWID

Why statistical models?

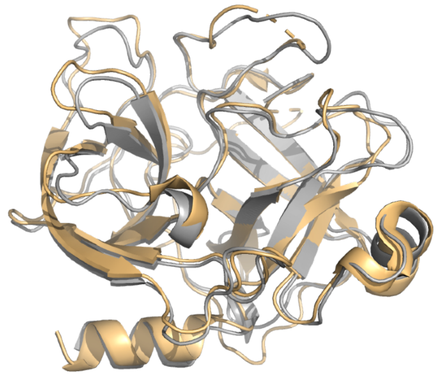
- Protein sequences are shaped by evolutionary pressure
- Many sequences can have similar structure and function: family

Kendrew *et al.,* Nature, **1958**
C Edgar *et al., Current opinion in structural biology,* **2006**
M Fowler *et al.,* Nature methods, **2014**
Cocco *et al.,* Rep. Prog. Phys., **2018**

Study proteins by learning from evolution

Different approaches

- Observe their structure (X-ray crystallography…)
- Simulate them using structure (Molecular Dynamics…)
- Modify them (mutational scans…)
- Learn from evolution (statistical models…) Etc…



Why statistical models?

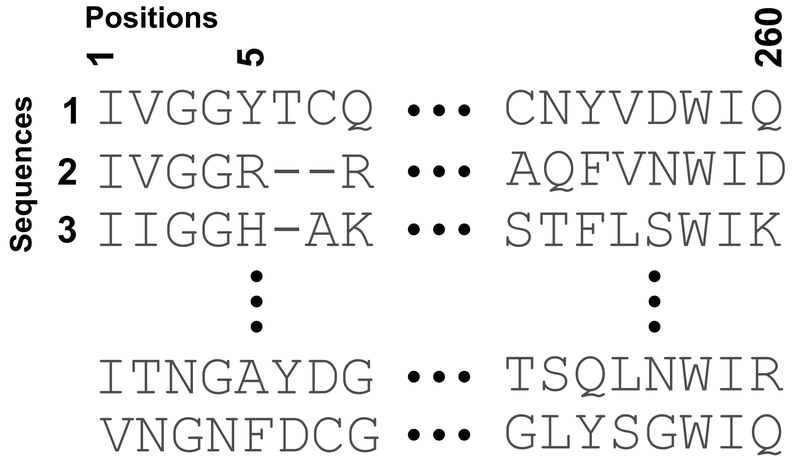
- Protein sequences are shaped by evolutionary pressure
- Many sequences can have similar structure and function: family


IVGGYTCQ ••• CNYVDWIQ
IVGGR--R ••• AQFVNWID

How?

- Collect sequences with shared structure and function
- Build a Multiple Sequence Alignment (MSA)
- Look for statistical signatures

Multiple Sequence Alignment



Kendrew *et al.,* Nature, **1958**
C Edgar *et al., Current opinion in structural biology,* **2006**
M Fowler *et al.,* Nature methods, **2014**
Cocco *et al.,* Rep. Prog. Phys., **2018**

Study proteins by learning from evolution

Different approaches

- Observe their structure (X-ray crystallography…)
- Simulate them using structure (Molecular Dynamics…)
- Modify them (mutational scans…)
- Learn from evolution (statistical models…) Etc…

Why statistical models?

- Protein sequences are shaped by evolutionary pressure
- Many sequences can have similar structure and function: family

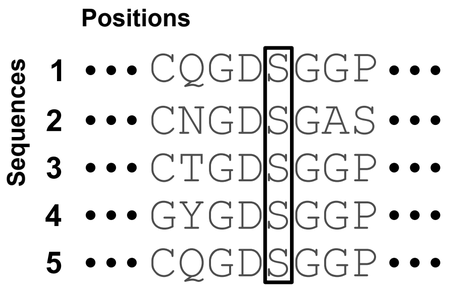
How?

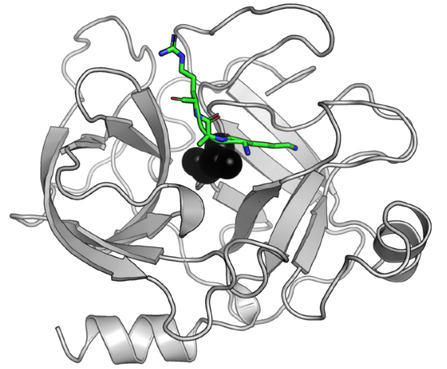
- Collect sequences with shared structure and function
- Build a Multiple Sequence Alignment (MSA)
- Look for statistical signatures

To find

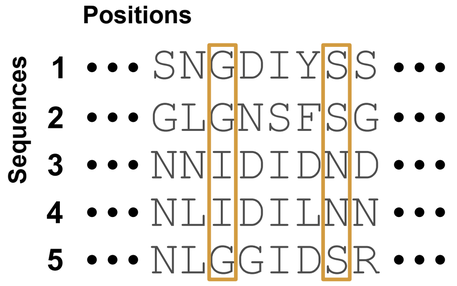
- Conservations (important amino-acids)
- Correlations (important interactions)

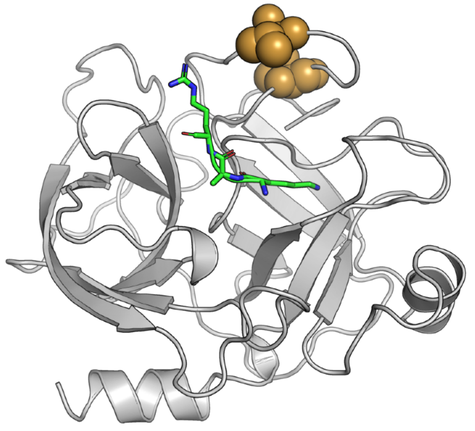
Conservation





Correlations



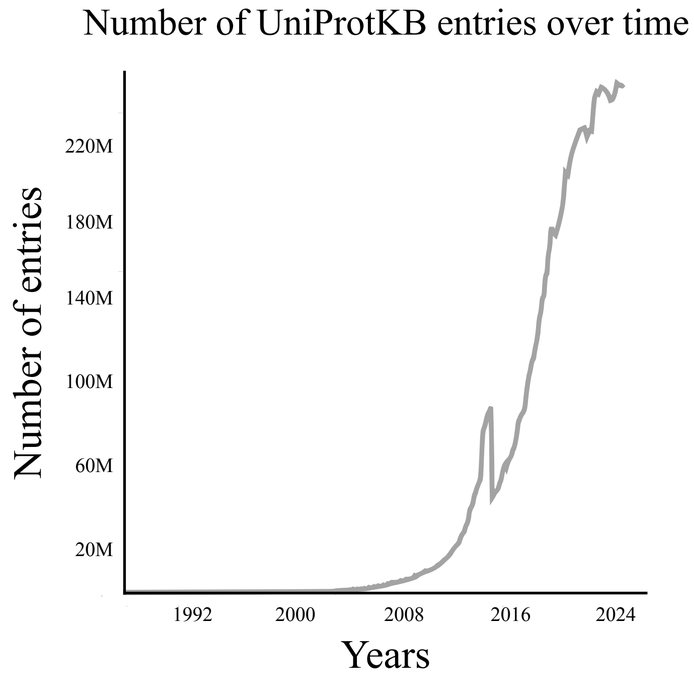


Kendrew *et al.,* Nature, **1958**
C Edgar *et al., Current opinion in structural biology,* **2006**
M Fowler *et al.,* Nature methods, **2014**
Cocco *et al.,* Rep. Prog. Phys., **2018**

Statistical learning on a protein family

Enough data for advanced methods

- Massive expansion of available sequences in the last two decades
- Approaches grounded in statistical physics
- Deep learning methods



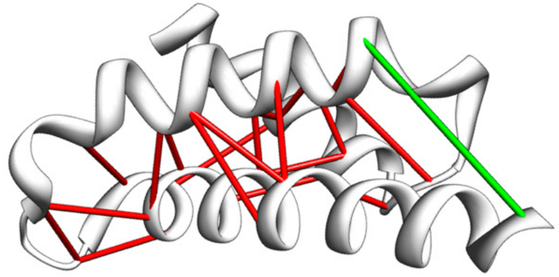
Shendure *et al., Nature biotechnology,* **2008**
Jain *et al., Genome biology*, **2016**
Weigt *et al.,* PNAS, **2009**
Rivoire *et al.,* PLoS CB, **2016**
UniProt, NAR, **2025**

Statistical learning on a protein family

Enough data for advanced methods

- Massive expansion of available sequences in the last two decades
- Approaches grounded in statistical physics
- Deep learning methods

Contact prediction

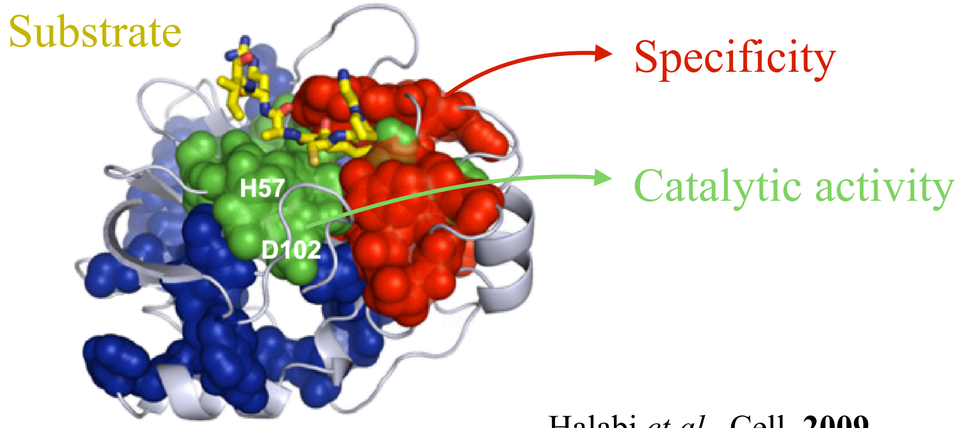


Morcos F *et al.*, PNAS, **2011**

Different methods capture different interaction scales

- From contacting pairs (ex: Direct Coupling Analysis)
- To coevolving groups (ex: Statistical Coupling Analysis)

S1A sectors mapped onto rat trypsin structure



Halabi *et al.*, Cell, **2009**

Shendure *et al., Nature biotechnology,* **2008**
Jain *et al., Genome biology*, **2016**
Weigt *et al.,* PNAS, **2009**
Rivoire *et al.,* PLoS CB, **2016**
UniProt, NAR, **2025**

Statistical learning on a protein family

Enough data for advanced methods

- Massive expansion of available sequences in the last two decades
- Approaches grounded in statistical physics
- Deep learning methods

Contact prediction

Morcos F *et al.*, PNAS, **2011**

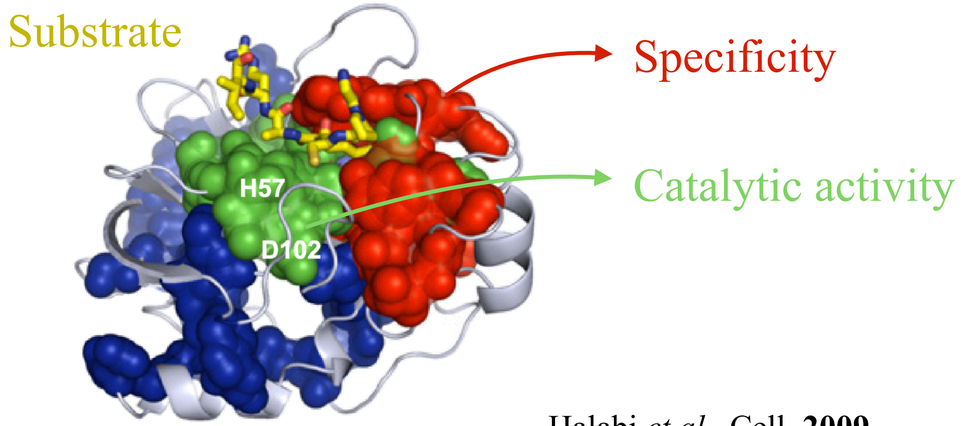
Different methods capture different interaction scales

- From contacting pairs (ex: Direct Coupling Analysis)
- To coevolving groups (ex: Statistical Coupling Analysis)

Statistical Coupling Analysis applied to S1A family

- 3 coevolving groups (sectors)
- Structurally connected
- Functionally independent (mutagenesis experiments)

S1A sectors mapped onto rat trypsin structure



Halabi *et al.*, Cell, **2009**

Shendure *et al., Nature biotechnology,* **2008**
Jain *et al., Genome biology*, **2016**
Weigt *et al.,* PNAS, **2009**
Rivoire *et al.,* PLoS CB, **2016**
UniProt, NAR, **2025**

Generative models for protein sequences

Training data

Generative models for protein sequences

Training data

Train

Generative model

Generative models for protein sequences

Training data

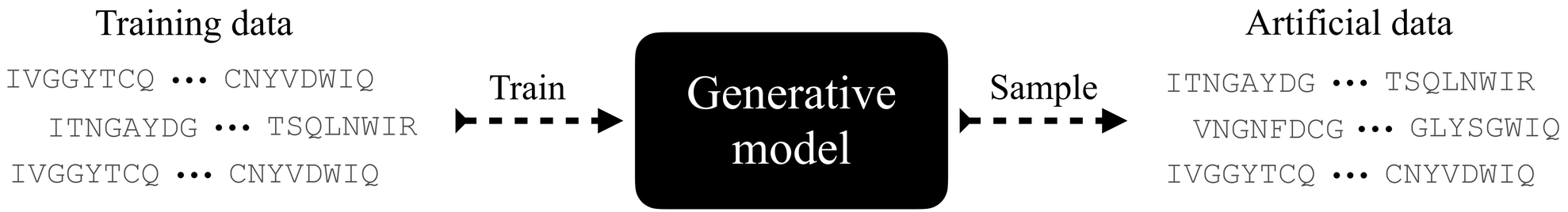
Train

Generative model

Sample

Artificial data

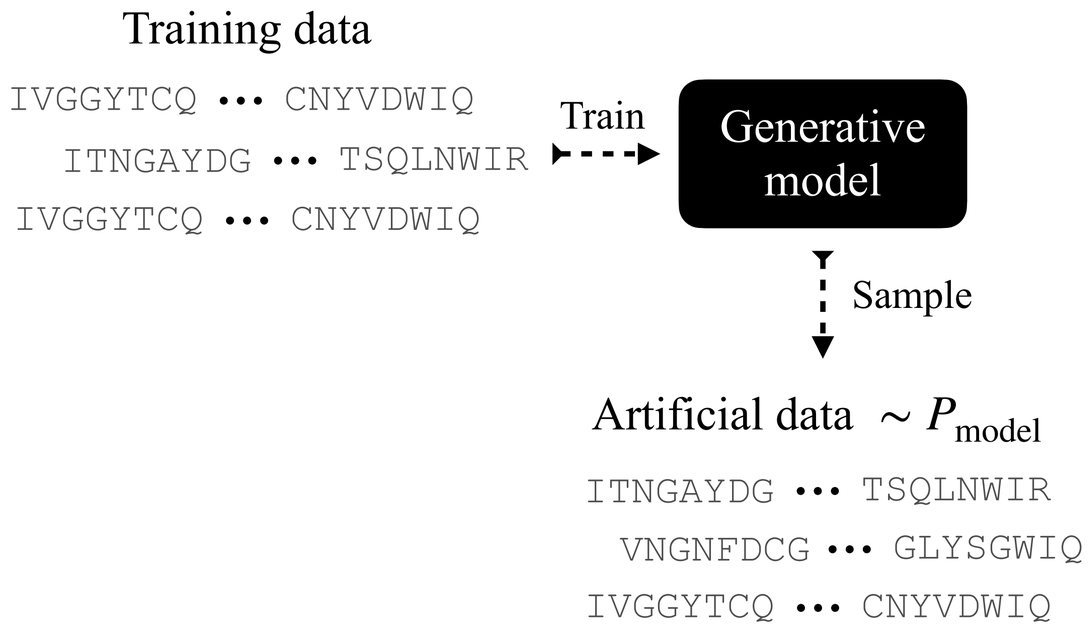
Generative models for protein sequences



Generative models for protein sequences

Principle

- Probability distribution over protein sequences



Hawkins-Hooker *et al., PLoS CB*, **2021**
Repecka *et al.,* Nature Machine Intelligence, **2021**
Sgarbossa *et al.,* Elife, **2023**
Watson L *et al.,* Nature, **2023**

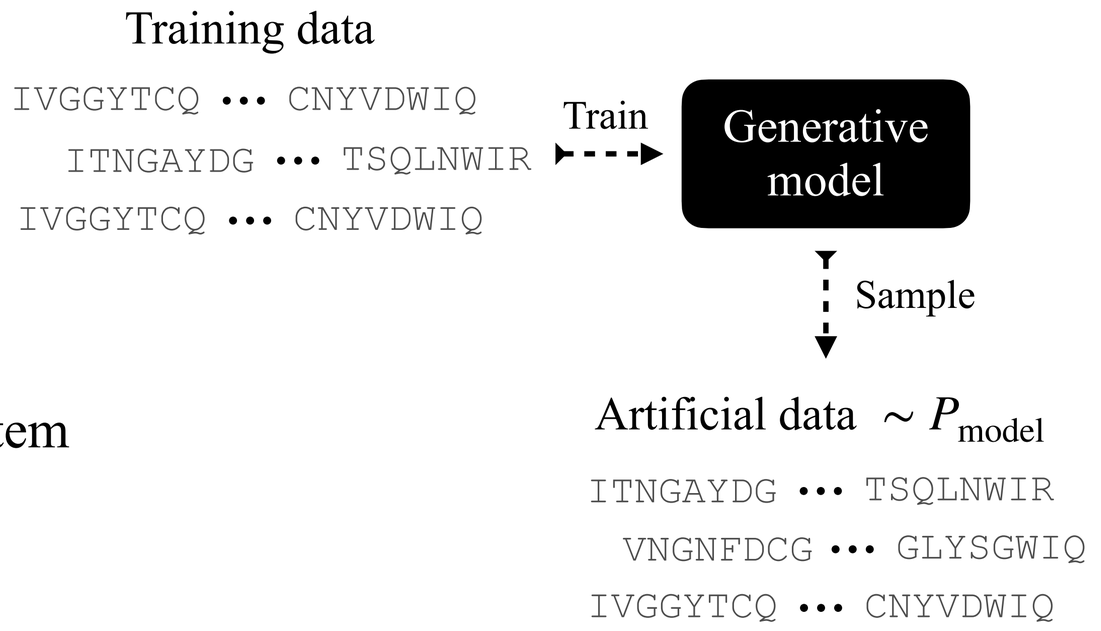
Generative models for protein sequences

Principle

- Probability distribution over protein sequences

For what?

- Design-oriented goal: make new proteins
- Framework for understanding: parametrization of the system



Hawkins-Hooker *et al., PLoS CB*, **2021**
Repecka *et al.,* Nature Machine Intelligence, **2021**
Sgarbossa *et al.,* Elife, **2023**
Watson L *et al.,* Nature, **2023**

Generative models for protein sequences

Principle

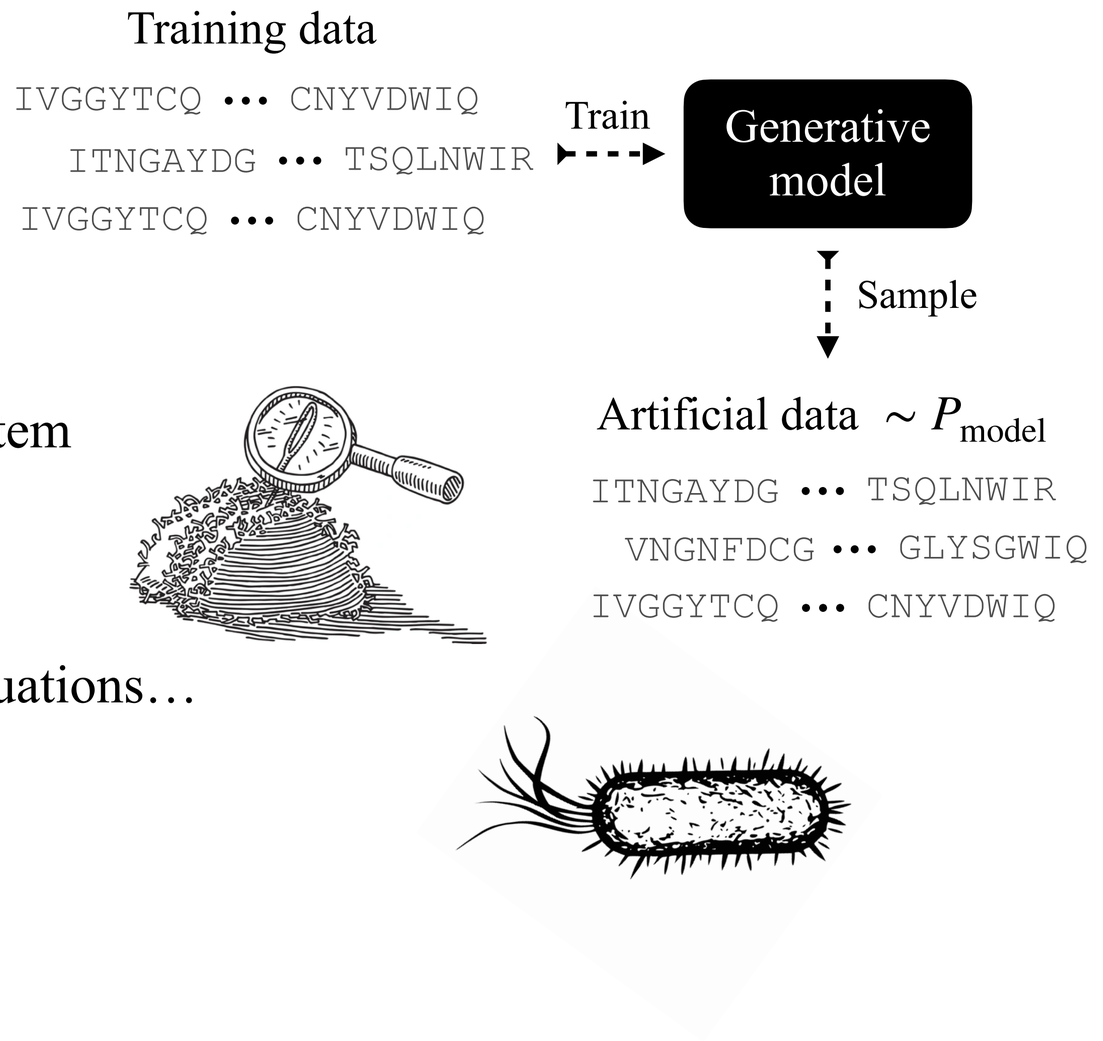
- Probability distribution over protein sequences

For what?

- Design-oriented goal: make new proteins
- Framework for understanding: parametrization of the system

Modeling a protein family

- High dimensional sequence space $\sim 10^{66} - 10^{650}$
- Model evaluation: wet-lab characterization, in silico evaluations…

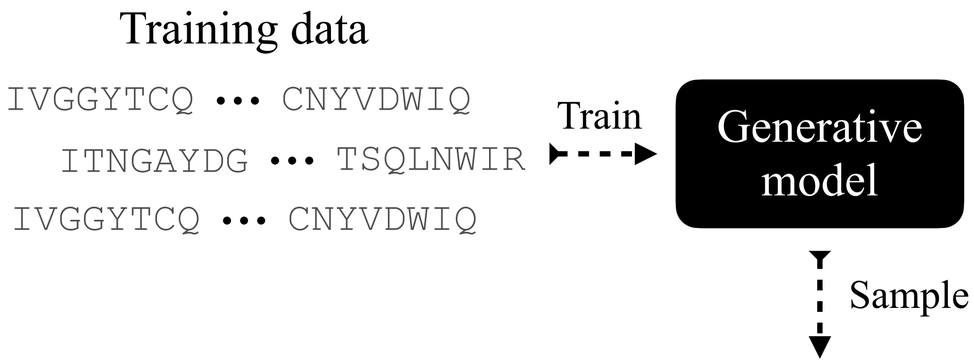


Hawkins-Hooker *et al., PLoS CB*, **2021**
Repecka *et al.,* Nature Machine Intelligence, **2021**
Sgarbossa *et al.,* Elife, **2023**
Watson L *et al.,* Nature, **2023**

Generative models for protein sequences

Principle

- Probability distribution over protein sequences

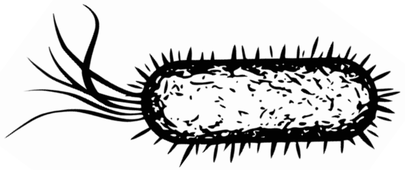


For what?

- Design-oriented goal: make new proteins
- Framework for understanding: parametrization of the system

Modeling a protein family

- High dimensional sequence space $\sim 10^{66} - 10^{650}$
- Model evaluation: wet-lab characterization, in silico evaluations…



Model

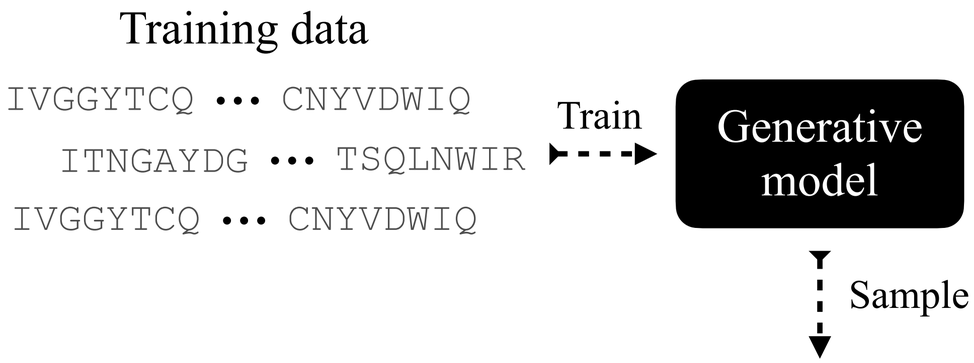
- Variational Autoencoders, Diffusion models, Transformers, Restricted Boltzmann Machine, Boltzmann Machine…

Hawkins-Hooker *et al., PLoS CB*, **2021**
Repecka *et al.,* Nature Machine Intelligence, **2021**
Sgarbossa *et al.,* Elife, **2023**
Watson L *et al.,* Nature, **2023**

Generative models for protein sequences

Principle

- Probability distribution over protein sequences

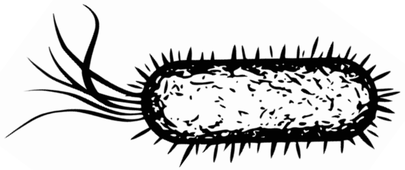


For what?

- Design-oriented goal: make new proteins
- Framework for understanding: parametrization of the system

Modeling a protein family

- High dimensional sequence space $\sim 10^{66} - 10^{650}$
- Model evaluation: wet-lab characterization, in silico evaluations…



Model

- Variational Autoencoders, Diffusion models, Transformers, Restricted Boltzmann Machine, Boltzmann Machine…

Boltzmann Machine

- Interpretability
- Mapping with other models
- Generative capacity experimentally tested (Russ *et al.* 2020)

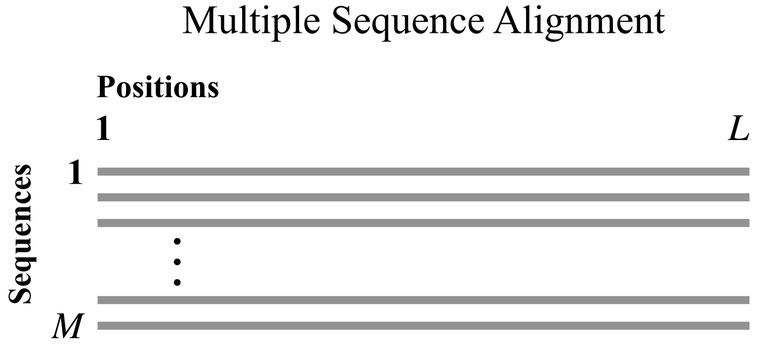
Hawkins-Hooker *et al., PLoS CB*, **2021**
Repecka *et al.,* Nature Machine Intelligence, **2021**
Sgarbossa *et al.,* Elife, **2023**
Watson L *et al.,* Nature, **2023**

Modeling a protein family with a Boltzmann Machine

Modeling a protein family with a Boltzmann Machine (BM)

Principle

- Modeling of a protein family



$$M \sim 10^2 - 10^5$$

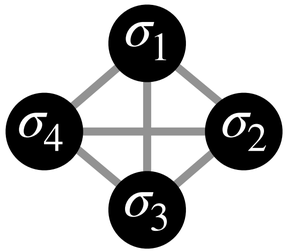
$$L \sim 50 - 500$$

Hopf *et al.,* Nature biotechnology, **2017**
Russ *et al.,* Science, **2020**
Morcos *et al.,* PNAS, **2011**
Weigt *et al.,* PNAS, **2009**
Ackley *et al.,* Cognitive science, **1985**
Cocco *et al.,* Rep. Prog. Phys., **2018**

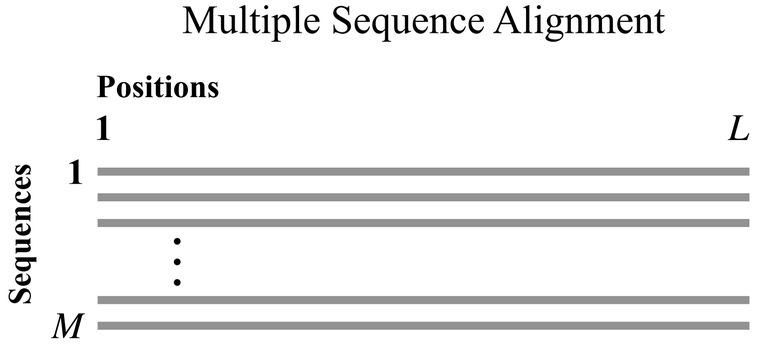
Modeling a protein family with a Boltzmann Machine (BM)

Principle

- Modeling of a protein family
- Graphical model: fully connected graph



$$P(\{\sigma_i\}_{i=1,\dots,L})$$



$$M \sim 10^2 - 10^5$$

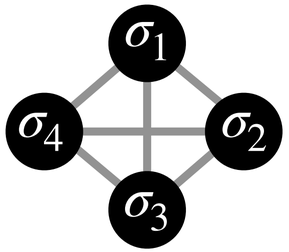
$$L \sim 50 - 500$$

Hopf *et al.,* Nature biotechnology, **2017**
Russ *et al.,* Science, **2020**
Morcos *et al.,* PNAS, **2011**
Weigt *et al.,* PNAS, **2009**
Ackley *et al.,* Cognitive science, **1985**
Cocco *et al.,* Rep. Prog. Phys., **2018**

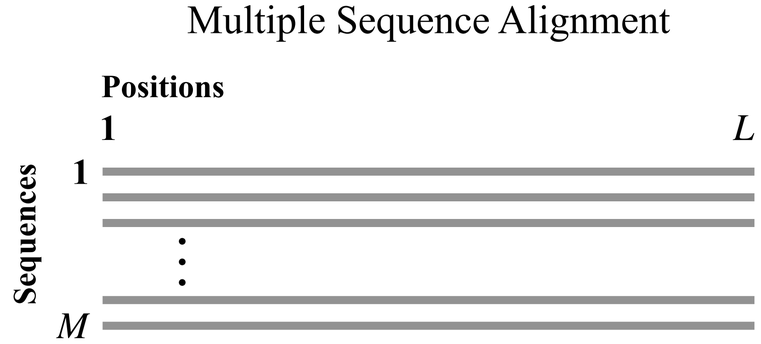
Modeling a protein family with a Boltzmann Machine (BM)

Principle

- Modeling of a protein family
- Graphical model: fully connected graph
- Potts model



$P(\{\sigma_i\}_{i=1,\dots,L})$



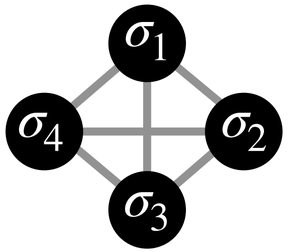
$$P(\{\sigma_i\}_{i=1,\dots,L}) = \frac{e^{\sum_{i=1}^{L} h_i(\sigma_i) + \sum_{i<j} J_{ij}(\sigma_i,\sigma_j)}}{Z}$$

Hopf *et al.,* Nature biotechnology, **2017**
Russ *et al.,* Science, **2020**
Morcos *et al.,* PNAS, **2011**
Weigt *et al.,* PNAS, **2009**
Ackley *et al.,* Cognitive science, **1985**
Cocco *et al.,* Rep. Prog. Phys., **2018**

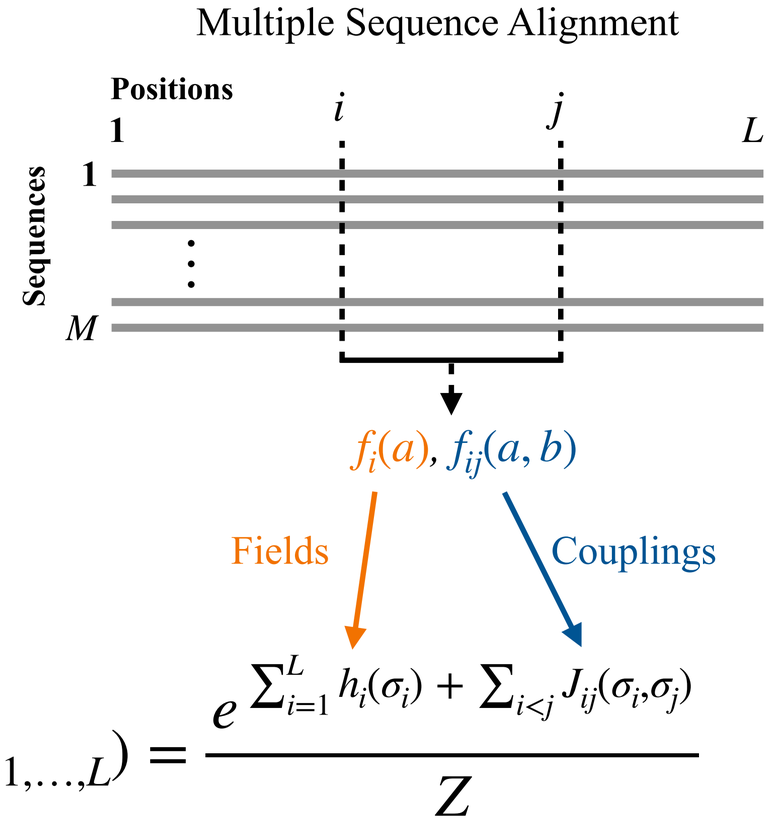
Modeling a protein family with a Boltzmann Machine (BM)

Principle

- Modeling of a protein family
- Graphical model: fully connected graph
- Potts model
- Trained to capture frequencies and pairwise frequencies (Maximum entropy approach)



$$P(\{\sigma_i\}_{i=1,\ldots,L})$$



$$P(\{\sigma_i\}_{i=1,\ldots,L}) = \frac{e^{\sum_{i=1}^{L} h_i(\sigma_i) + \sum_{i<j} J_{ij}(\sigma_i,\sigma_j)}}{Z}$$

Hopf *et al.,* Nature biotechnology, **2017**
Russ *et al.,* Science, **2020**
Morcos *et al.,* PNAS, **2011**
Weigt *et al.,* PNAS, **2009**
Ackley *et al.,* Cognitive science, **1985**
Cocco *et al.,* Rep. Prog. Phys., **2018**

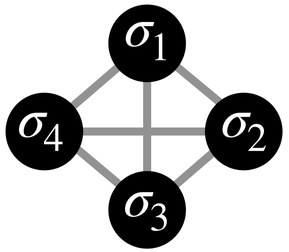
Modeling a protein family with a Boltzmann Machine (BM)

Principle

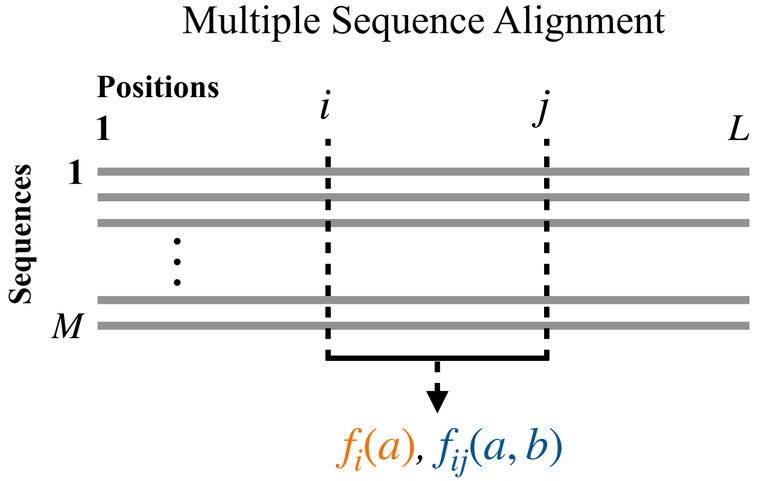
- Modeling of a protein family
- Graphical model: fully connected graph
- Potts model
- Trained to capture frequencies and pairwise frequencies (Maximum entropy approach)

Inference

- Parameters that maximize the probability of natural sequences (MLE)
- Other methods: Mean field, Pseudo-likelihood maximization, Autoregressive model…



$P(\{\sigma_i\}_{i=1,\dots,L})$



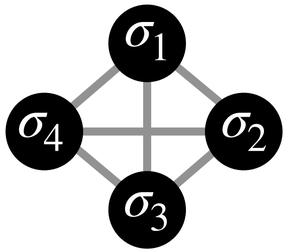
$$P(\{\sigma_i\}_{i=1,\dots,L}) = \frac{e^{\sum_{i=1}^{L} h_i(\sigma_i) + \sum_{i<j} J_{ij}(\sigma_i,\sigma_j)}}{Z}$$

Hopf *et al.,* Nature biotechnology, **2017**
Russ *et al.,* Science, **2020**
Morcos *et al.,* PNAS, **2011**
Weigt *et al.,* PNAS, **2009**
Ackley *et al.,* Cognitive science, **1985**
Cocco *et al.,* Rep. Prog. Phys., **2018**

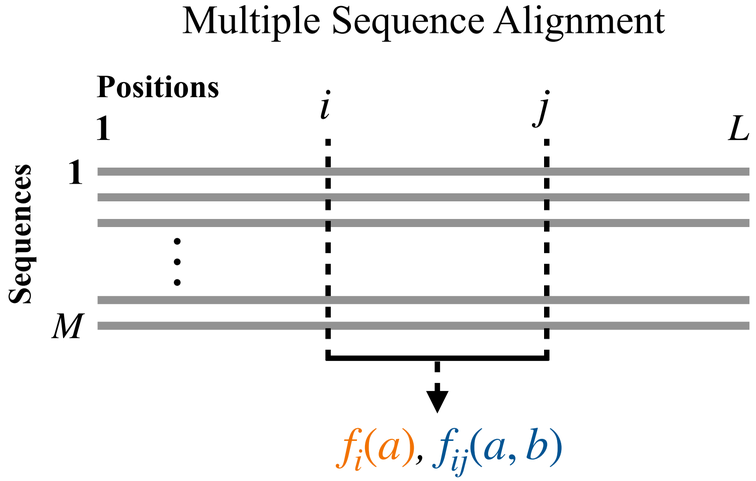
Modeling a protein family with a Boltzmann Machine (BM)

Principle

- Modeling of a protein family
- Graphical model: fully connected graph
- Potts model
- Trained to capture frequencies and pairwise frequencies (Maximum entropy approach)



$$P(\{\sigma_i\}_{i=1,\dots,L})$$



Inference

- Parameters that maximize the probability of natural sequences (MLE)
- Other methods: Mean field, Pseudo-likelihood maximization, Autoregressive model…

$$P(\{\sigma_i\}_{i=1,\dots,L}) = \frac{e^{\sum_{i=1}^{L} h_i(\sigma_i) + \sum_{i<j} J_{ij}(\sigma_i,\sigma_j)}}{Z}$$

Results

- Predict structural contacts
- Predict mutational effect
- **Generative with low-temperature sampling** (Russ et al., 2020)

$$P(\{\sigma_i\}_{i=1,\dots,L}) \sim e^{-\frac{E(\boldsymbol{h},\boldsymbol{J})}{T}} \text{ with } T < 1$$

Hopf *et al.,* Nature biotechnology, **2017**
Russ *et al.,* Science, **2020**
Morcos *et al.,* PNAS, **2011**
Weigt *et al.,* PNAS, **2009**
Ackley *et al.,* Cognitive science, **1985**
Cocco *et al.,* Rep. Prog. Phys., **2018**

The undersampling problem

Problem

- # parameters $\sim 10^5 - 10^7 \gg$ # sequences $\sim 10^2 - 10^5$
- Extreme statistics from undersampling $\rightarrow$ infinite parameters

Multiple Sequence Alignment

The undersampling problem

Problem

- # parameters $\sim 10^5 - 10^7 \gg$ # sequences $\sim 10^2 - 10^5$
- Extreme statistics from undersampling → infinite parameters

Regularization methods

- Remove parameters (Pruning, Alphabet reduction…)
- Modify statistics (pseudo-counts)
- Constrain parameters during optimization (L_p norm…)

Multiple Sequence Alignment

The undersampling problem

Problem

- # parameters $\sim 10^5 - 10^7 \gg$ # sequences $\sim 10^2 - 10^5$
- Extreme statistics from undersampling → infinite parameters

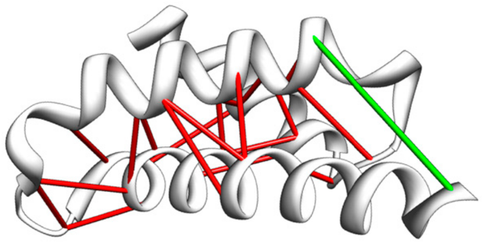
Regularization methods

- Remove parameters (Pruning, Alphabet reduction…)
- Modify statistics (pseudo-counts)
- Constrain parameters during optimization (L_p norm…)

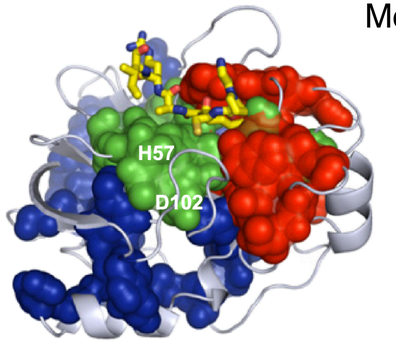
Importance of data's statistical structure

- Real data often have rich statistical structure
- Proteins: Correlated units of different sizes, magnitude…
- Uneven impact of undersampling on different statistical signatures (Kleeorin *et al.* 2023)

Multiple Sequence Alignment



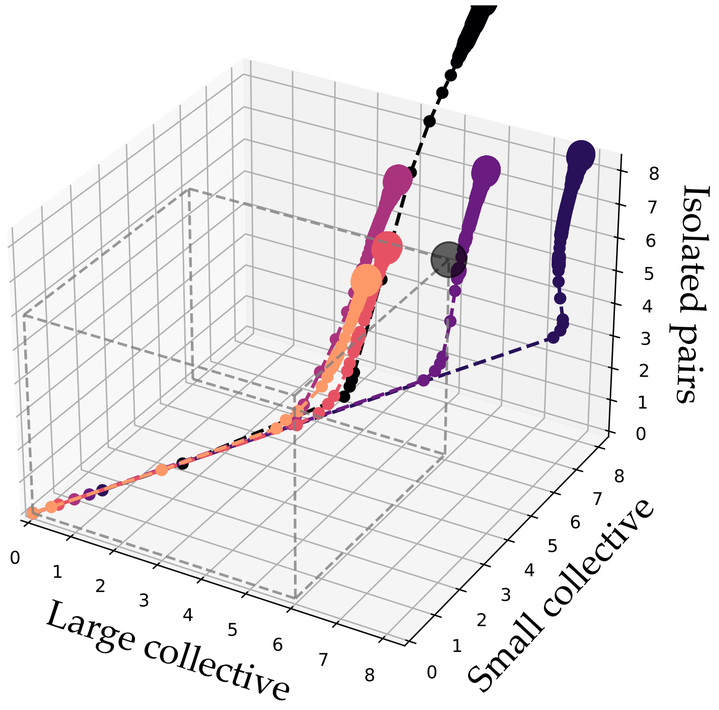
Morcos *et al.*, PNAS, **2011**



Halabi *et al.*, Cell, **2009**

The undersampling problem

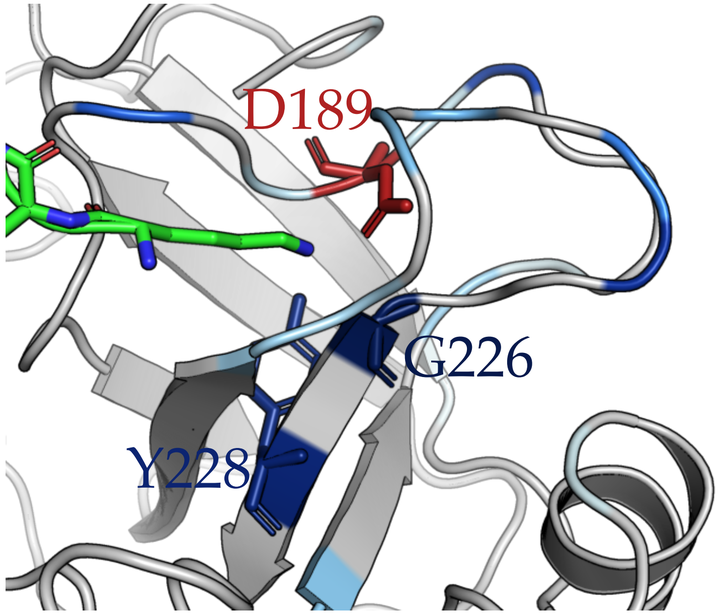
Overcoming undersampling induced biases



In collaboration with Emily Hinds, Yaakov Kleeorin, Rama Ranganathan (University of Chicago, USA)

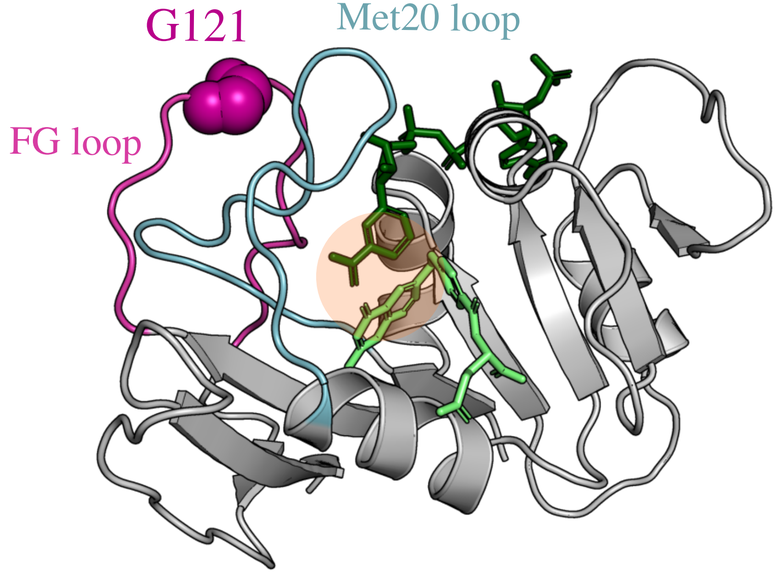
Investigate protein properties with statistical learning

Specificity mechanism in S1A family



In collaboration with Amaury Paveyranne, Timothé Lucas, Shoichi Yip, Clément Nizak (LJP, Sorbonne University, France)

Allosteric network of *E. Coli* DHFR



In collaboration with Paul Guenon, Damien Laage, Guillaume Stirnemann (ENS, France), Clément Nizak (LJP, France), Karolina Filipowska, Kim Reynolds (University of Texas, USA)

The Undersampling problem

In collaboration with Emily Hinds, Yaakov Kleeorin,
Rama Ranganathan (University of Chicago, USA)

The undersampling problem

I. Undersampling-induced biases (Kleeorin *et al.*, Cell system, 2023)

II. Generative Capacity of the Boltzmann Machine (Russ *et al.*, Science, 2020)

III. New inference method: Stochastic Boltzmann Machine

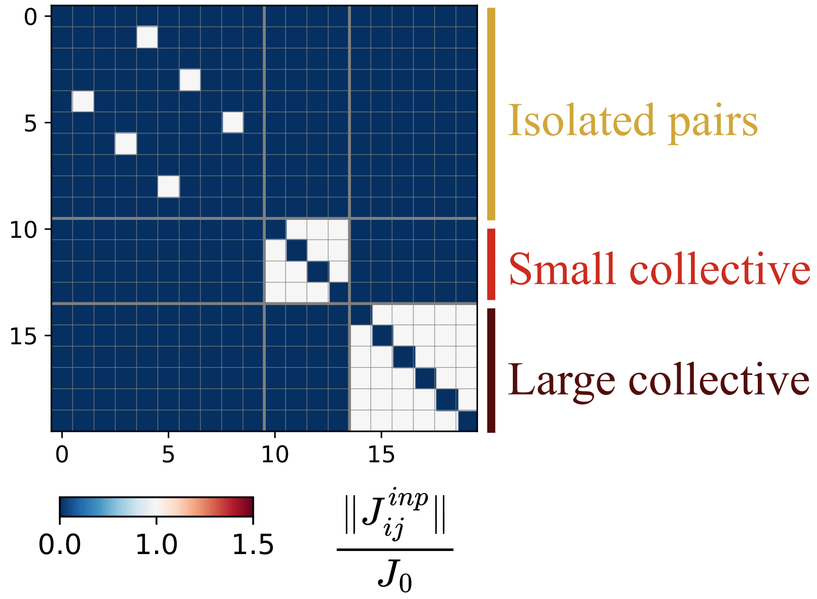
I. Undersampling-induced biases (Kleeorin *et al.* 2023)

Assess model performance with a toy model

Toy model features

- Boltzmann Machine model
- Correlated units of different sizes

$$P(\{\sigma_i\}_{i=1,\ldots,L}) = \frac{e^{\sum_{i=1}^{L} h_i(\sigma_i) + \sum_{i<j} J_{ij}(\sigma_i,\sigma_j)}}{Z}$$



positions $L = 20$, # states $q = 10$, $M = 300$

Kleeorin et al., Cell system **2023**

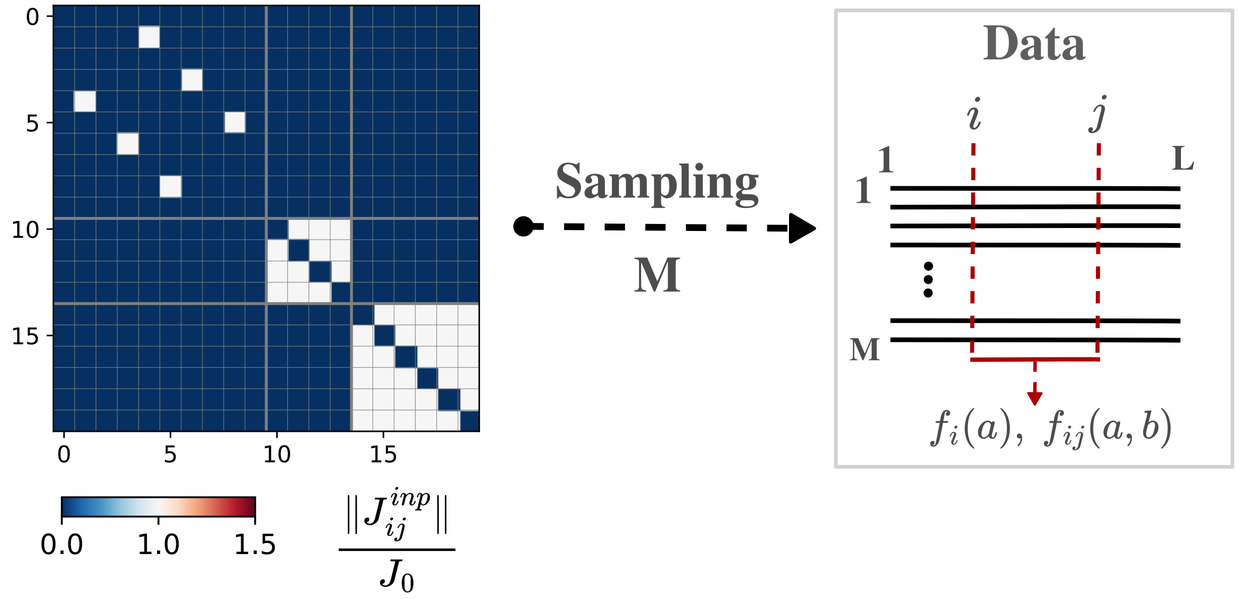
I. Undersampling-induced biases (Kleeorin *et al.* 2023)

Assess model performance with a toy model

Toy model features

- Boltzmann Machine model
- Correlated units of different sizes

$$P(\{\sigma_i\}_{i=1,\ldots,L}) = \frac{e^{\sum_{i=1}^{L} h_i(\sigma_i) + \sum_{i<j} J_{ij}(\sigma_i,\sigma_j)}}{Z}$$



positions $L = 20$, # states $q = 10$, $M = 300$

I. Undersampling-induced biases (Kleeorin *et al.* 2023)

Assess model performance with a toy model

Toy model features

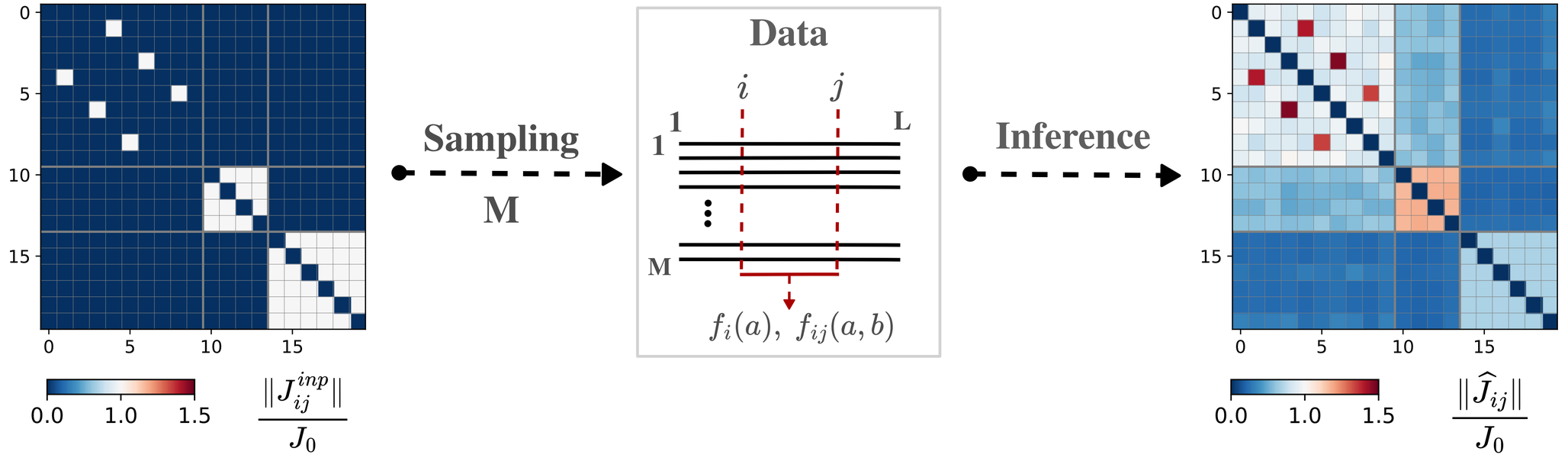
- Boltzmann Machine model
- Correlated units of different sizes

$$P(\{\sigma_i\}_{i=1,\ldots,L}) = \frac{e^{\sum_{i=1}^{L} h_i(\sigma_i) + \sum_{i<j} J_{ij}(\sigma_i,\sigma_j)}}{Z}$$

Inference:

- Undersampling regime
- Log-likelihood maximization with L_2 regularization

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \left[\frac{1}{M} \sum_{m=1}^{M} \log P(\boldsymbol{\sigma}^{(m)} \mid \boldsymbol{\theta}) \; - \lambda_J \|\boldsymbol{J}\|^2 - \lambda_h \|\boldsymbol{h}\|^2 \right]$$

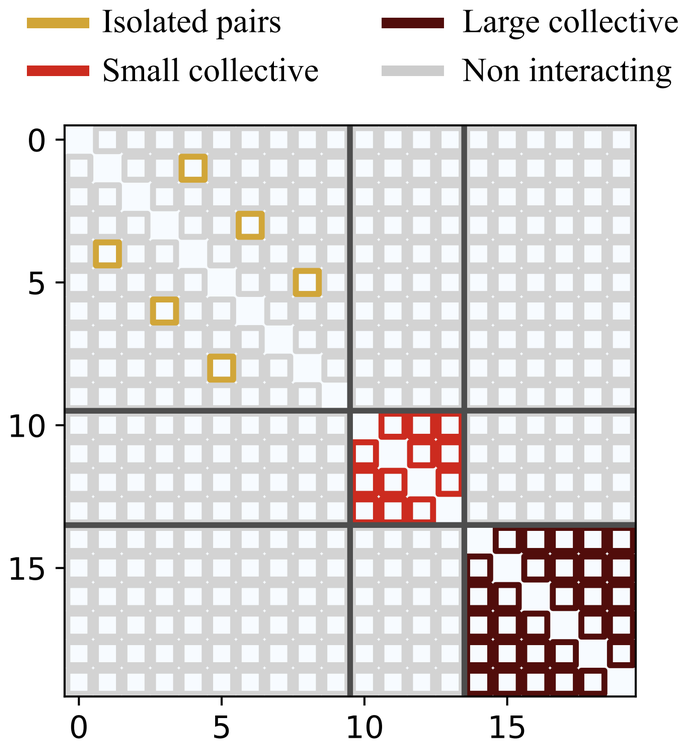


positions $L = 20$, # states $q = 10$, $M = 300$

Kleeorin et al., Cell system **2023**

I. Undersampling-induced biases (Kleeorin *et al.* 2023)

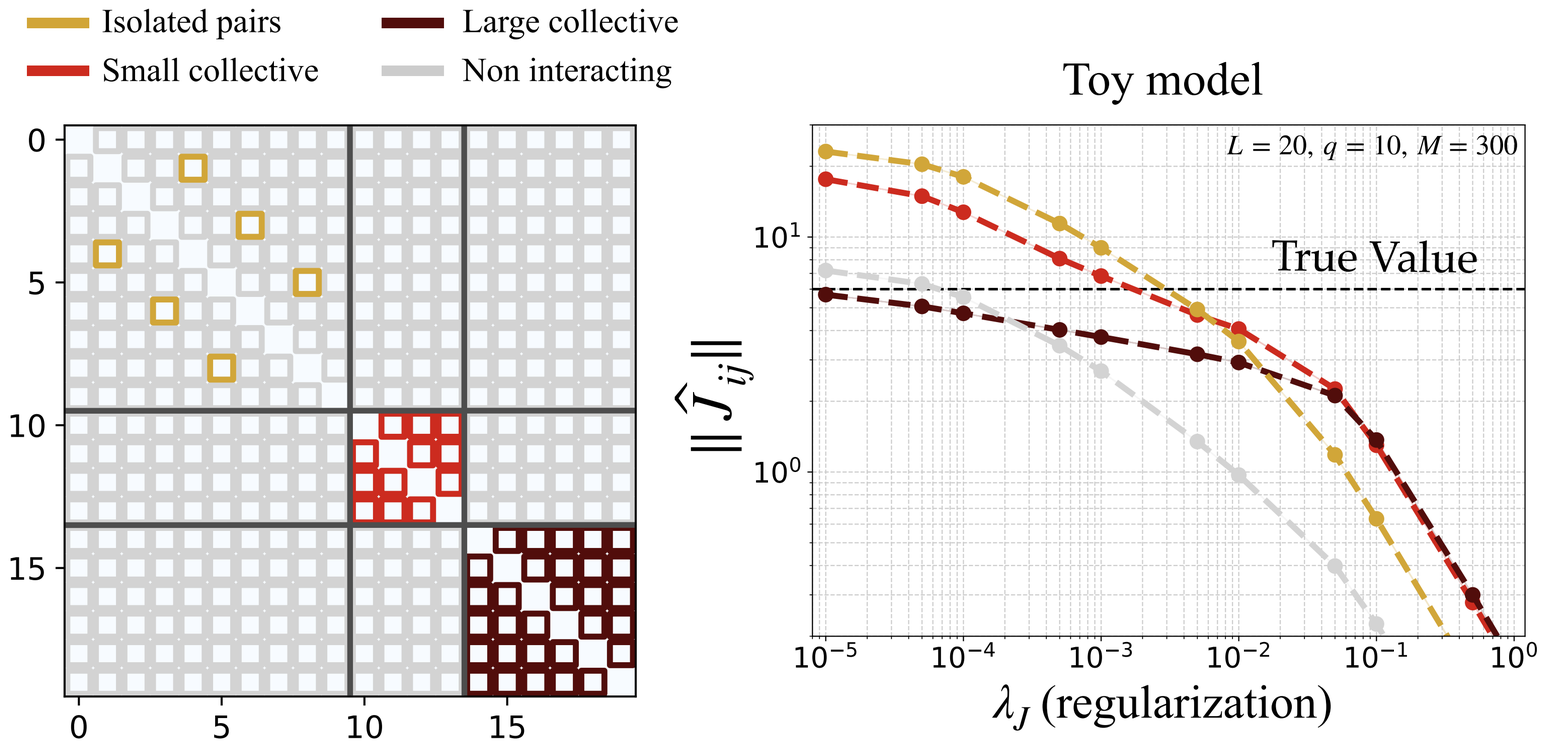
Inference as function of regularization strength



Figures made following the methodology of Kleeorin et al., Cell system **2023**

I. Undersampling-induced biases (Kleeorin *et al.* 2023)

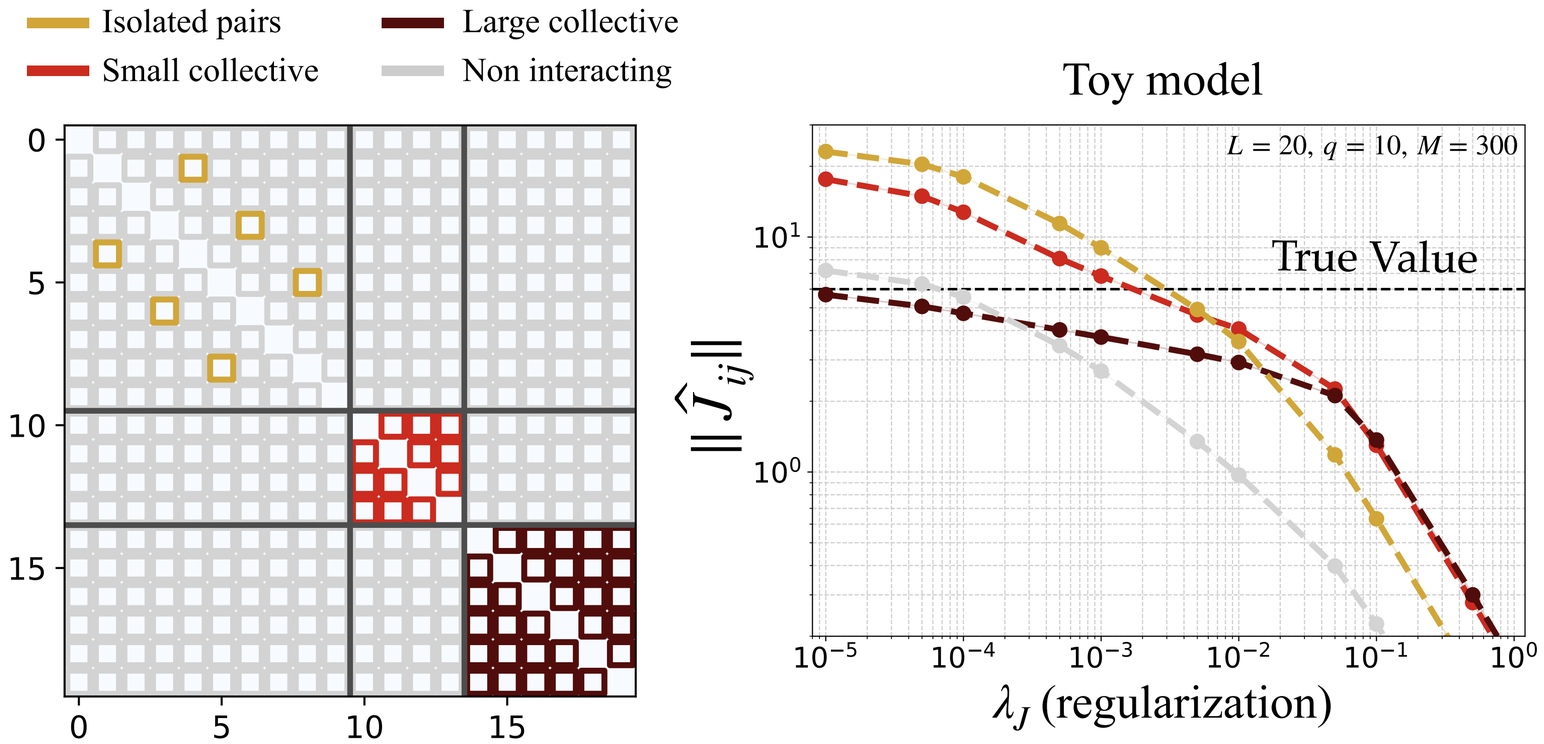
Inference as function of regularization strength



Figures made following the methodology of Kleeorin et al., Cell system **2023**

I. Undersampling-induced biases (Kleeorin *et al.* 2023)

Inference as function of regularization strength

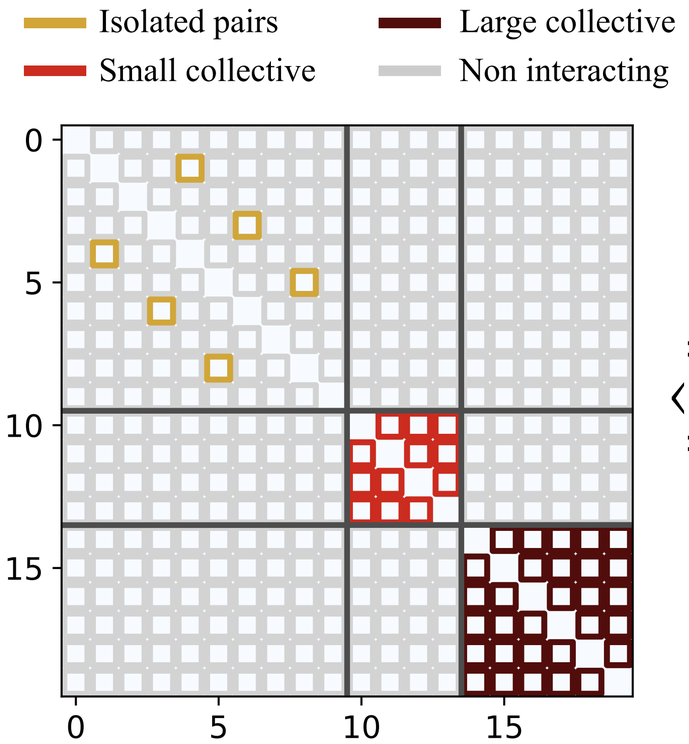


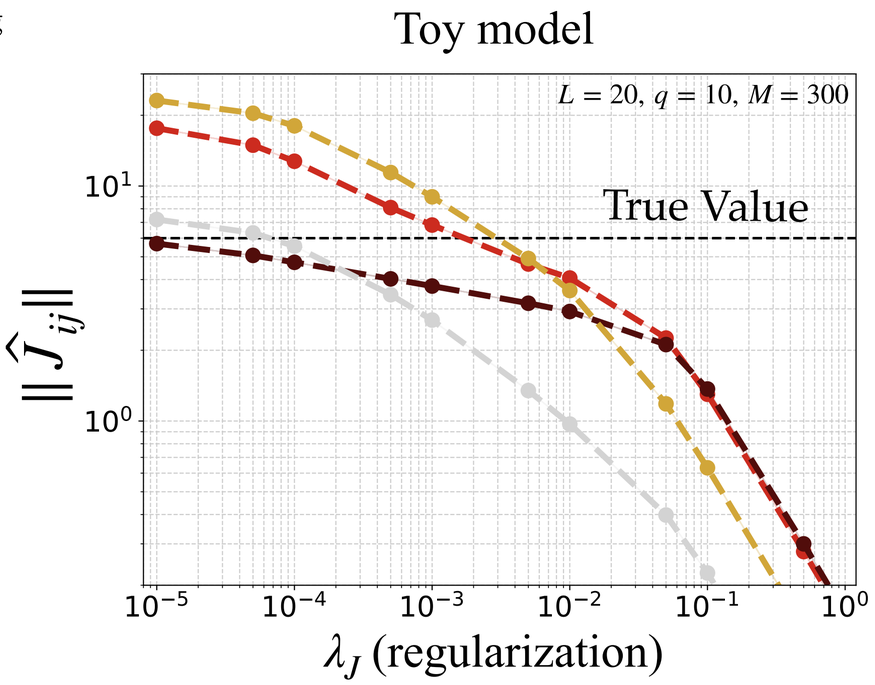
- **Systematic bias** between the estimation of collective modes and isolated interactions

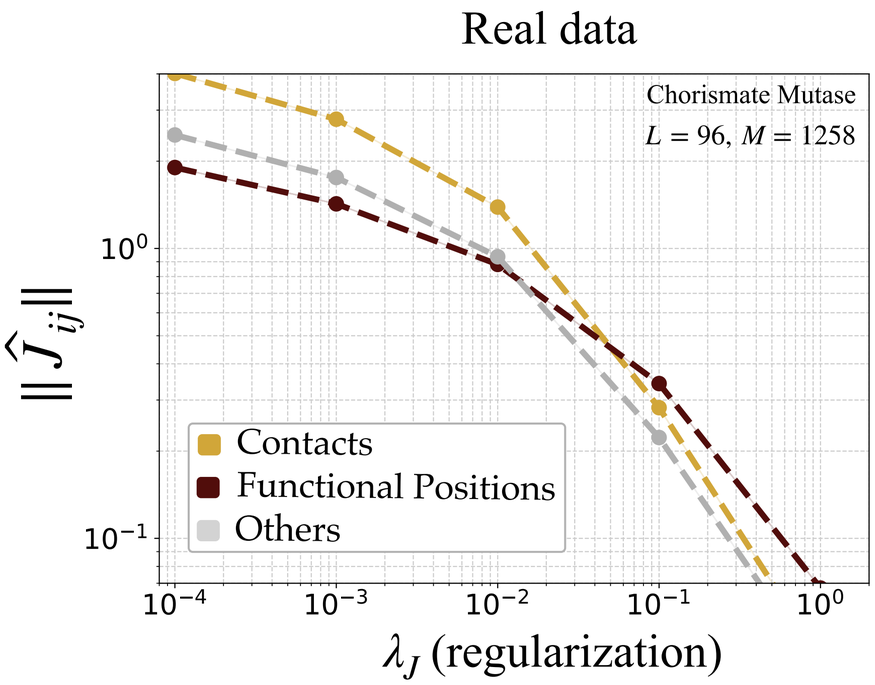
Figures made following the methodology of Kleeorin et al., Cell system **2023**

I. Undersampling-induced biases (Kleeorin *et al.* 2023)

Inference as function of regularization strength







- **Systematic bias** between the estimation of collective modes and isolated interactions
- Relevance for real protein data

Figures made following the methodology of Kleeorin et al., Cell system **2023**

Generative Capacity of the Boltzmann Machine (Russ *et al.* 2020)

Application to the Chorismate Mutase family

Fidelity

Do the artificial proteins share key properties with those observed in the training data?

Novelty

How much do the artificial sequences differ from the training data?

→ *Generalization capacity*

Diversity

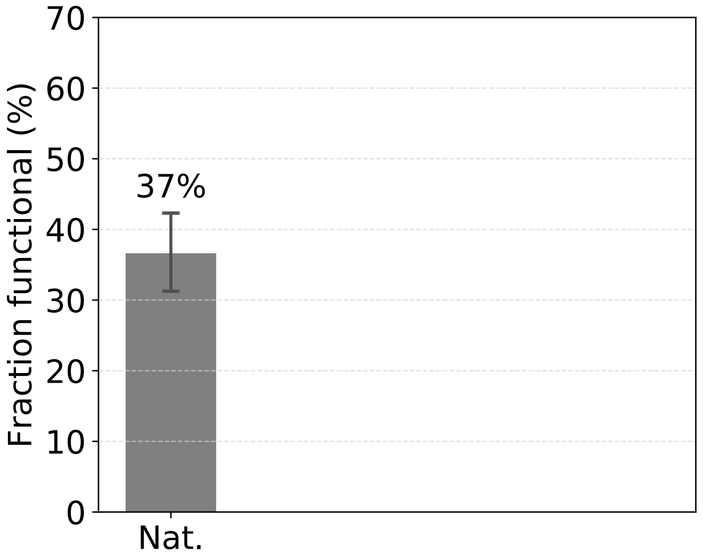
Do the artificial sequences span the same range of variability as the training data?

II. Generative Capacity of the Boltzmann Machine (Russ *et al.* 2020)

Application to the Chorismate Mutase family

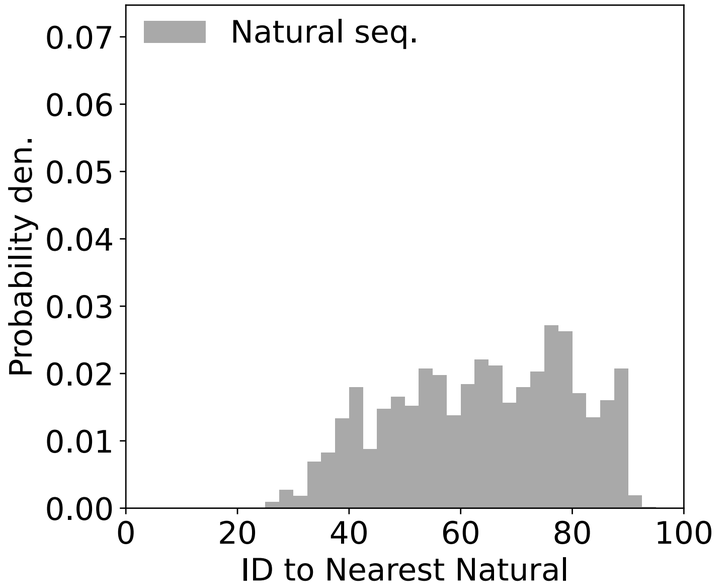
Fidelity

Fraction of functional sequences



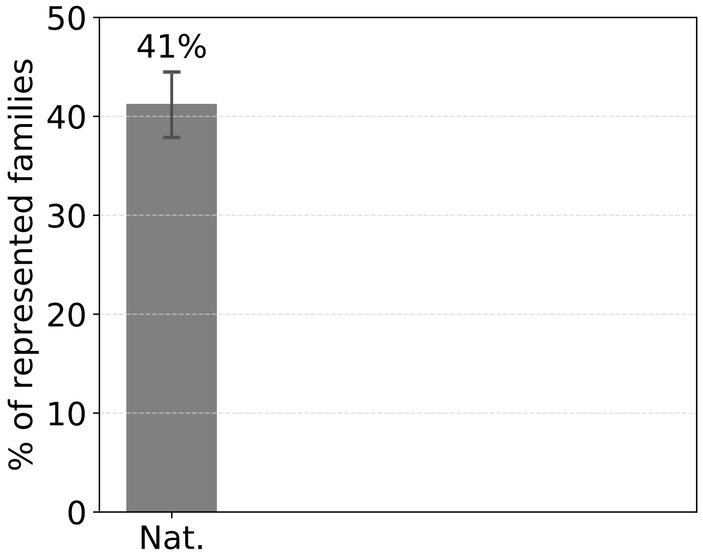
Novelty

Distribution of identity to the nearest natural sequence



Diversity

% of taxonomic families represented



L_2 regularization: $\lambda = 0.01$

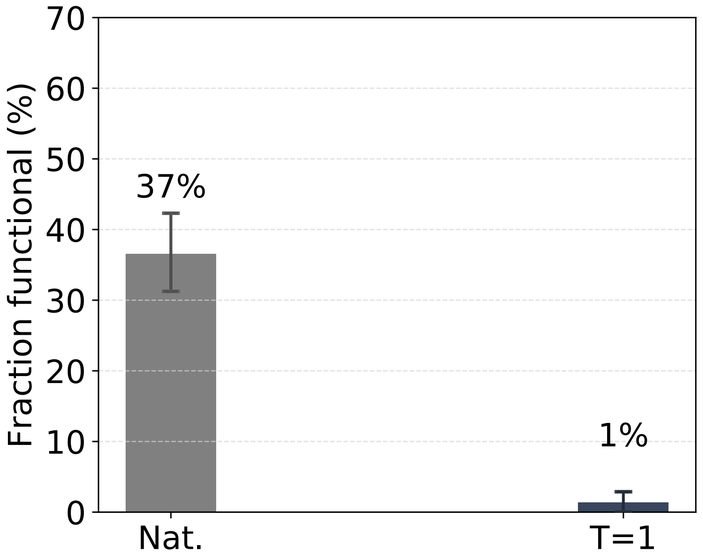
Figures made from Russ et al., *Science* **2020** data (sequences tested in *E. Coli)*

II. Generative Capacity of the Boltzmann Machine (Russ *et al.* 2020)

Application to the Chorismate Mutase family

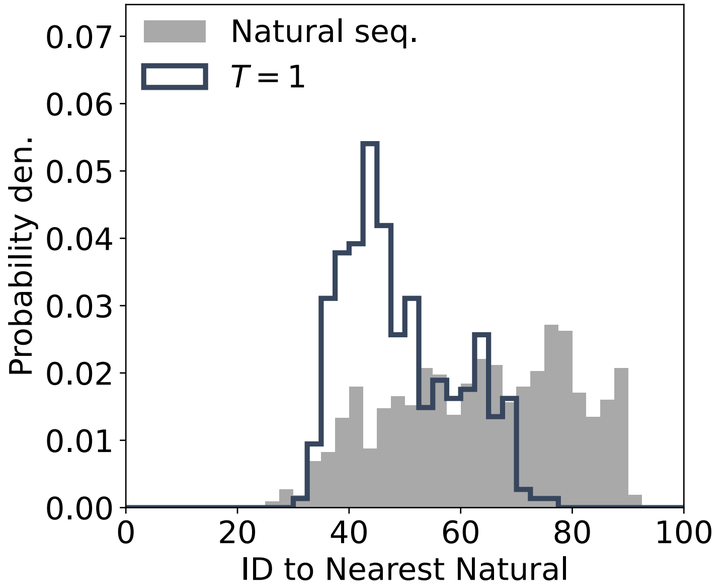
Fidelity

Fraction of functional sequences



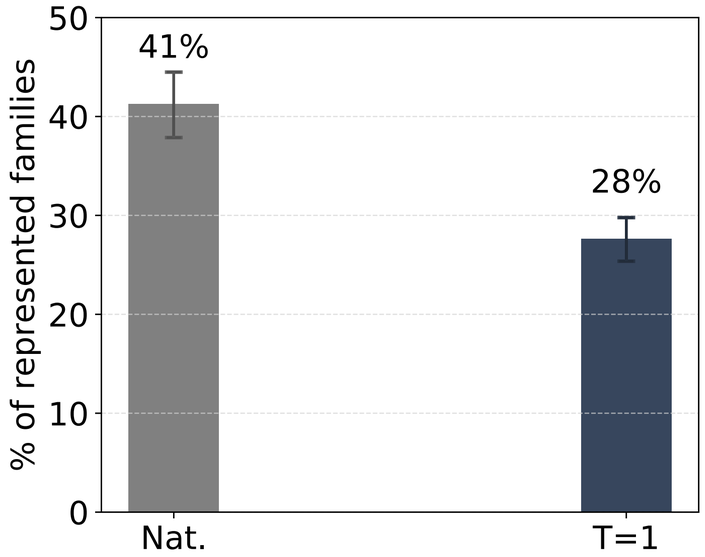
Novelty

Distribution of identity to the nearest natural sequence



Diversity

% of taxonomic families represented



L_2 regularization: $\lambda = 0.01$

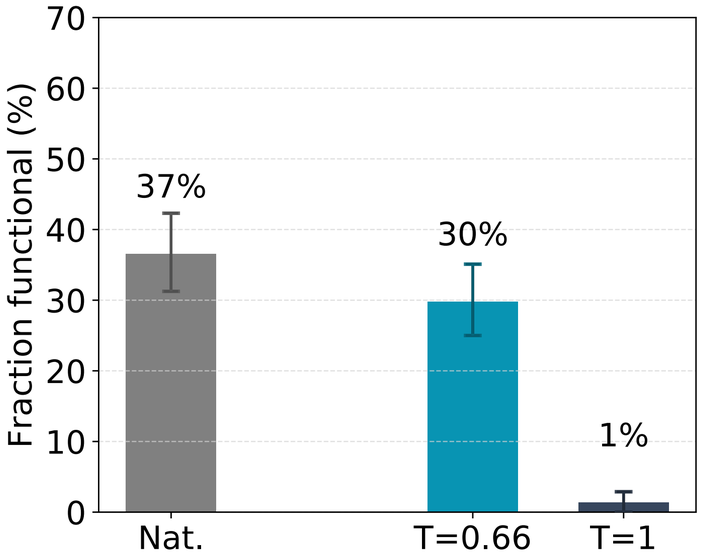
Figures made from Russ et al., *Science* **2020** data (sequences tested in *E. Coli)*

II. Generative Capacity of the Boltzmann Machine (Russ *et al.* 2020)

Application to the Chorismate Mutase family

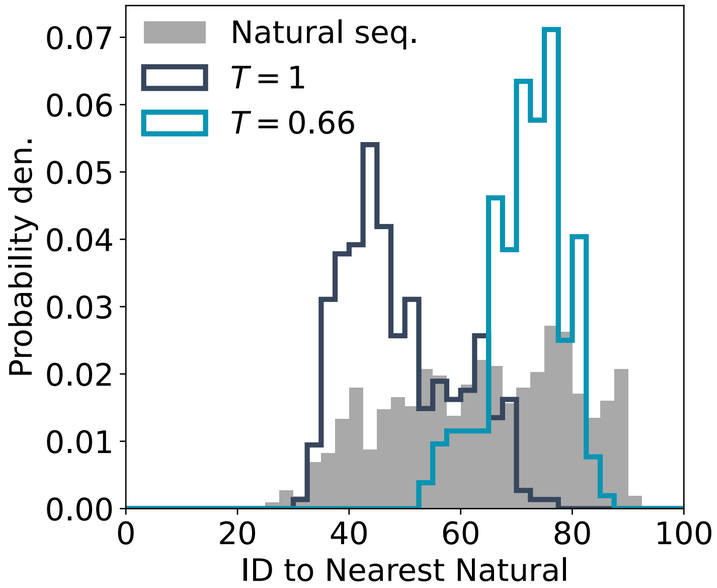
Fidelity

Fraction of functional sequences



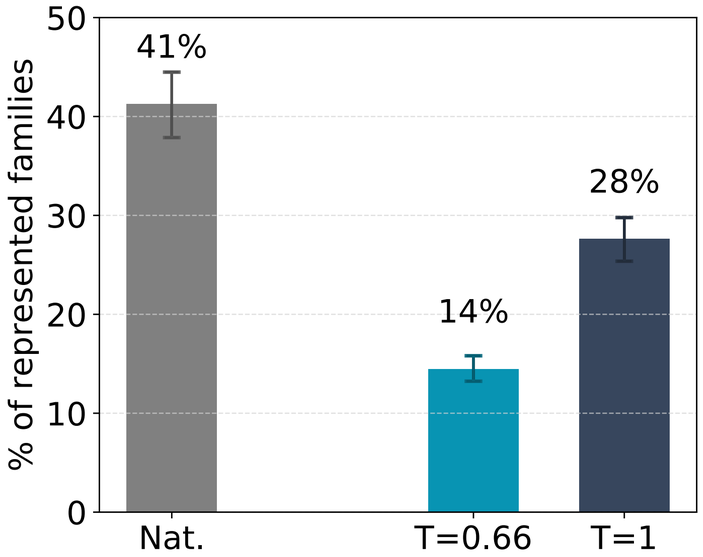
Novelty

Distribution of identity to the nearest natural sequence



Diversity

% of taxonomic families represented



L_2 regularization: $\lambda = 0.01$

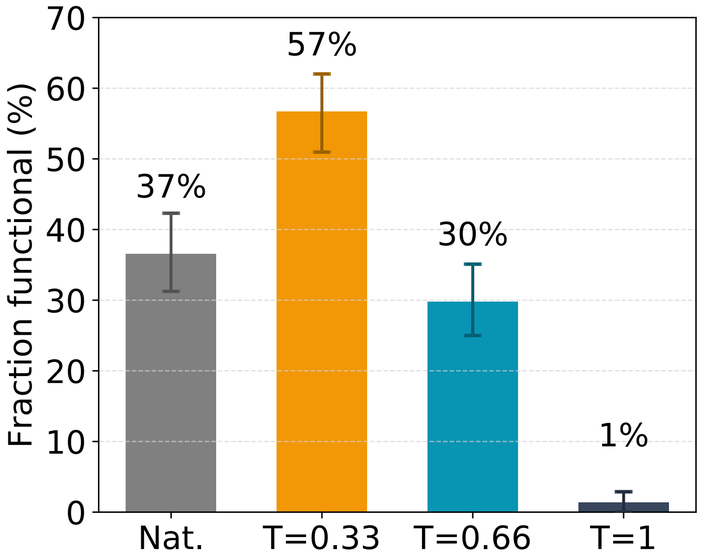
Figures made from Russ et al., *Science* **2020** data (sequences tested in *E. Coli)*

II. Generative Capacity of the Boltzmann Machine (Russ *et al.* 2020)

Application to the Chorismate Mutase family

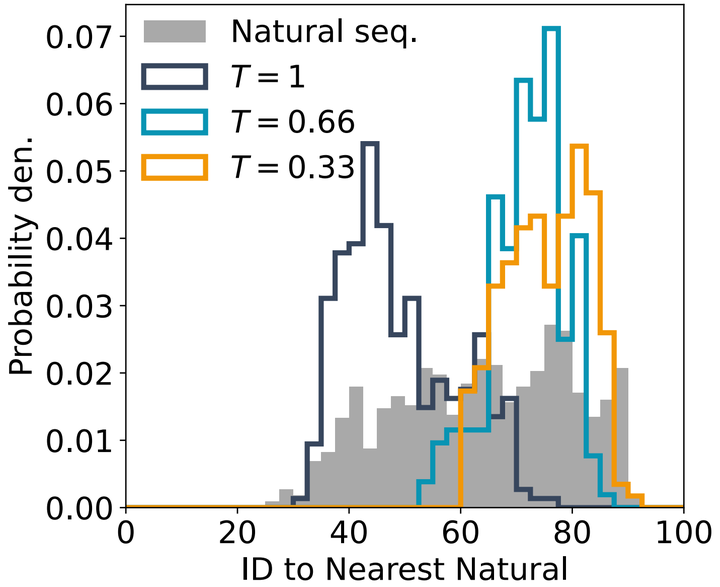
Fidelity

Fraction of functional sequences



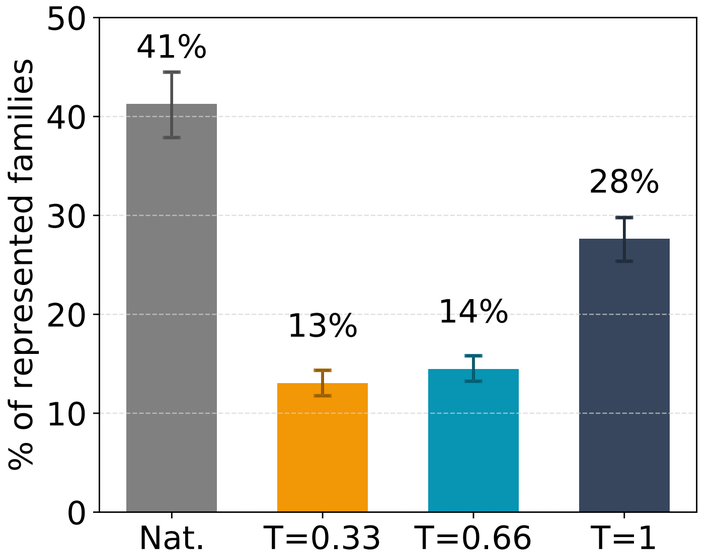
Novelty

Distribution of identity to the nearest natural sequence



Diversity

% of taxonomic families represented



- Functional sequences requires low-temperature sampling: $P(\{\sigma_i\}_{i=1,\dots,L}) \sim e^{-\frac{E(\boldsymbol{h},\boldsymbol{J})}{T}}$ with $T < 1$
- The gain in functionality comes at the cost of **reduced diversity** and **novelty**

L_2 regularization: $\lambda = 0.01$

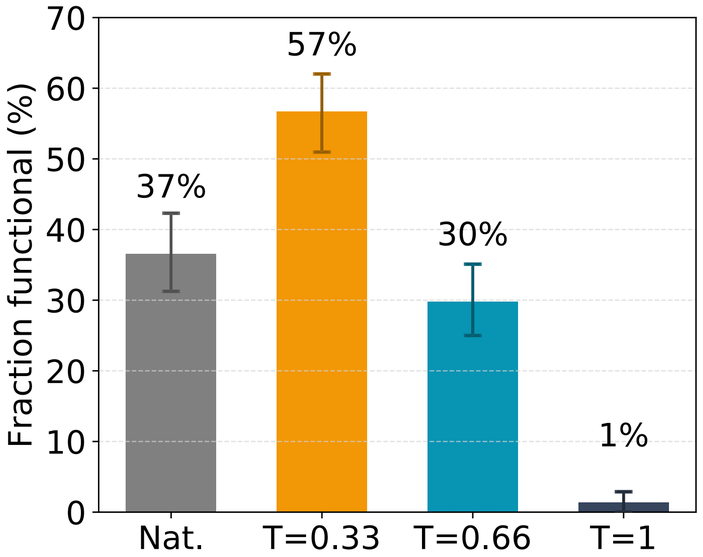
Figures made from Russ et al., *Science* **2020** data (sequences tested in *E. Coli)*

II. Generative Capacity of the Boltzmann Machine (Russ *et al.* 2020)

Application to the Chorismate Mutase family

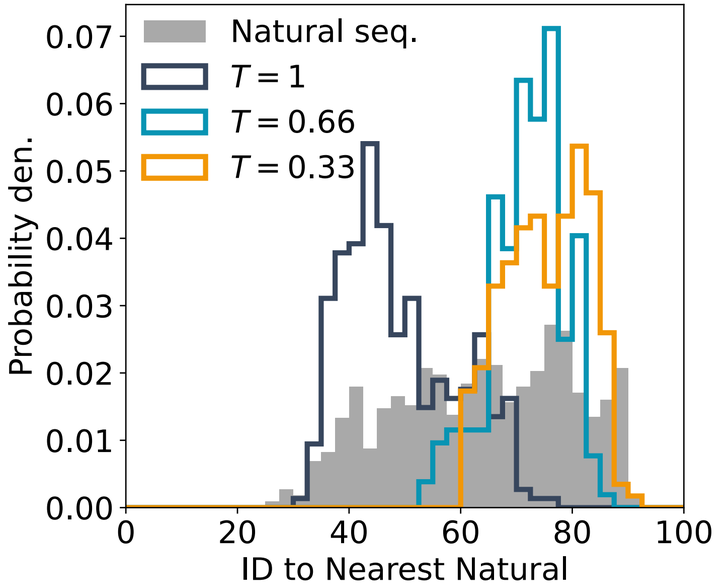
Fidelity

Fraction of functional sequences



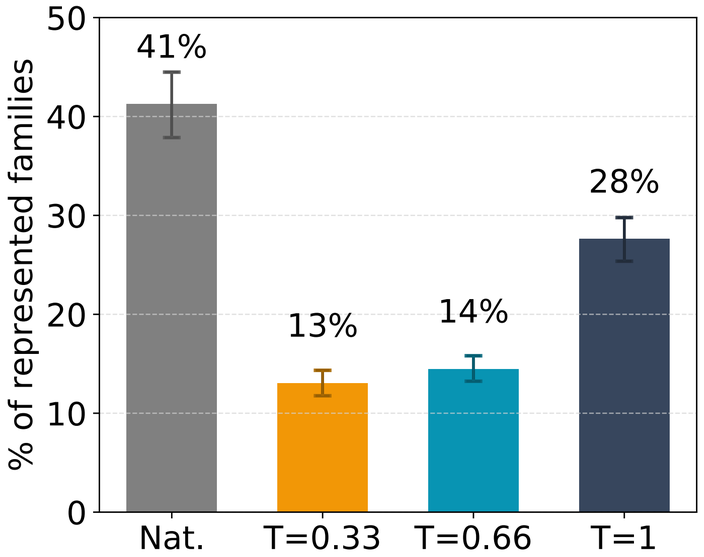
Novelty

Distribution of identity to the nearest natural sequence



Diversity

% of taxonomic families represented



- Functional sequences requires low-temperature sampling: $P(\{\sigma_i\}_{i=1,\dots,L}) \sim e^{-\frac{E(\boldsymbol{h},\boldsymbol{J})}{T}}$ with $T < 1$
- The gain in functionality comes at the cost of **reduced diversity** and **novelty**
- Success may depend on the experimental assay

L_2 regularization: $\lambda = 0.01$

Figures made from Russ et al., *Science* **2020** data (sequences tested in *E. Coli)*

III. Stochastic Boltzmann Machine

Assess model performance with a toy model

$$\boldsymbol{\theta} = \{\boldsymbol{J}, \boldsymbol{h}\} \qquad f(\boldsymbol{\theta}) = \frac{1}{M}\sum_{m=1}^{M} \log P(\boldsymbol{\sigma}^{(m)} \mid \boldsymbol{\theta}) \; - \lambda_J \|\boldsymbol{J}\|^2 - \lambda_h \|\boldsymbol{h}\|^2$$

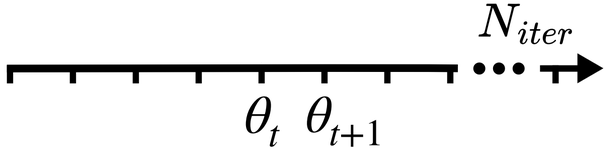
III. Stochastic Boltzmann Machine

Assess model performance with a toy model

$$\boldsymbol{\theta} = \{\boldsymbol{J}, \boldsymbol{h}\} \qquad f(\boldsymbol{\theta}) = \frac{1}{M}\sum_{m=1}^{M} \log P(\boldsymbol{\sigma}^{(m)} \mid \boldsymbol{\theta}) \; - \lambda_J \|\boldsymbol{J}\|^2 - \lambda_h \|\boldsymbol{h}\|^2$$

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta_t \boldsymbol{p}_t$$

1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$



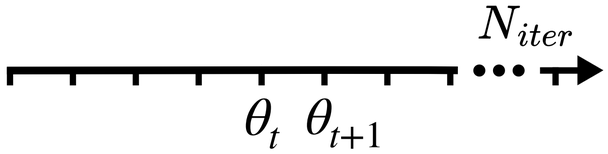
III. Stochastic Boltzmann Machine

Assess model performance with a toy model

$$\boldsymbol{\theta} = \{\boldsymbol{J}, \boldsymbol{h}\} \qquad f(\boldsymbol{\theta}) = \frac{1}{M}\sum_{m=1}^{M} \log P(\boldsymbol{\sigma}^{(m)} \mid \boldsymbol{\theta}) \; - \lambda_J \|\boldsymbol{J}\|^2 - \lambda_h \|\boldsymbol{h}\|^2$$

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta_t \boldsymbol{p}_t$$

1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$



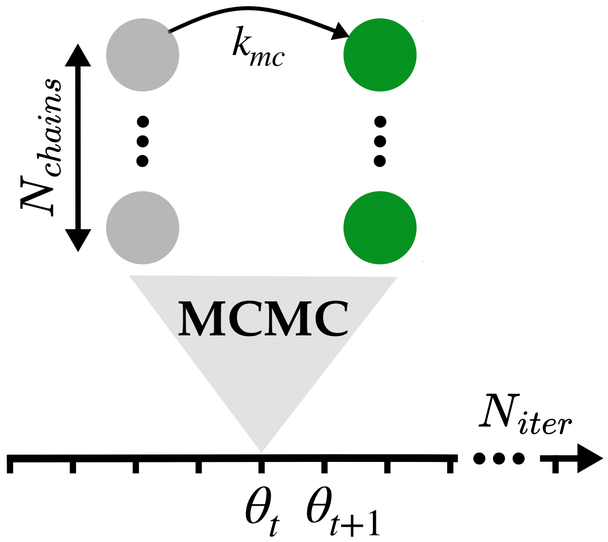
$$\frac{\partial f(\boldsymbol{\theta})}{\partial J_{ij}(a,b)} = f_{ij}(a,b) - \langle \delta(\sigma_i, a)\delta(\sigma_j, b) \rangle + \lambda_J J_{ij}(a,b)$$

Empirical Model

III. Stochastic Boltzmann Machine

Assess model performance with a toy model

$$\boldsymbol{\theta} = \{\boldsymbol{J}, \boldsymbol{h}\} \qquad f(\boldsymbol{\theta}) = \frac{1}{M}\sum_{m=1}^{M} \log P(\boldsymbol{\sigma}^{(m)} \mid \boldsymbol{\theta}) \; - \lambda_J \|\boldsymbol{J}\|^2 - \lambda_h \|\boldsymbol{h}\|^2$$



$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta_t \boldsymbol{p}_t$$

1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$

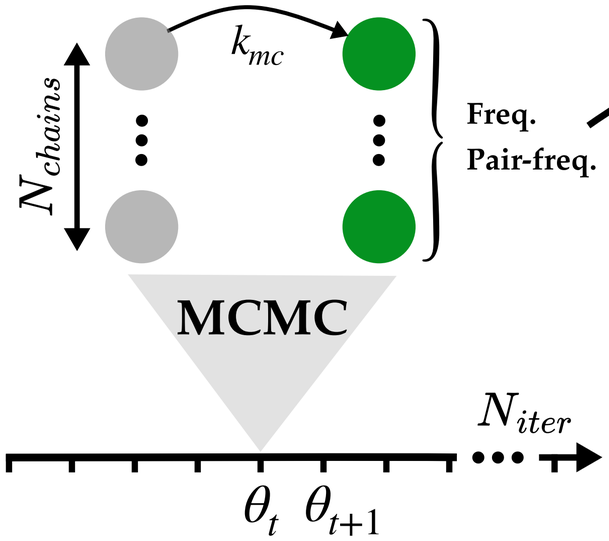
$$\frac{\partial f(\boldsymbol{\theta})}{\partial J_{ij}(a,b)} = f_{ij}(a,b) - \langle \delta(\sigma_i, a)\delta(\sigma_j, b) \rangle + \lambda_J J_{ij}(a,b)$$

Empirical Model

III. Stochastic Boltzmann Machine

Assess model performance with a toy model

$$\boldsymbol{\theta} = \{\boldsymbol{J}, \boldsymbol{h}\} \qquad f(\boldsymbol{\theta}) = \frac{1}{M}\sum_{m=1}^{M} \log P(\boldsymbol{\sigma}^{(m)} \mid \boldsymbol{\theta}) \; - \lambda_J \|\boldsymbol{J}\|^2 - \lambda_h \|\boldsymbol{h}\|^2$$



$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta_t \boldsymbol{p}_t$$

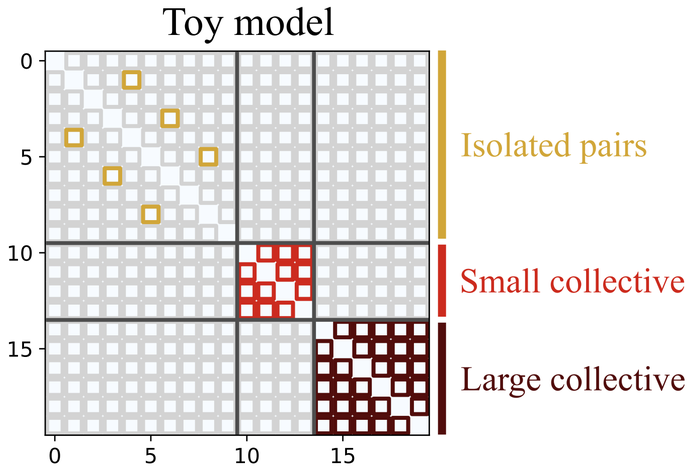
1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$

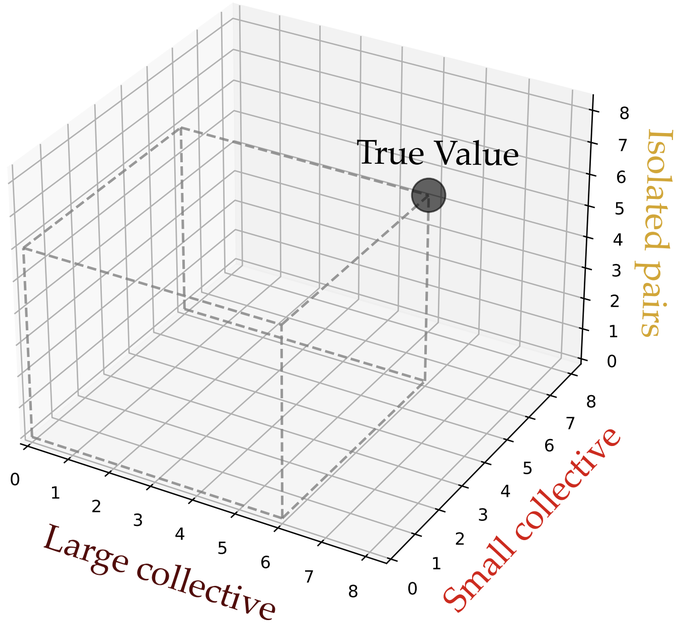
$$\frac{\partial f(\boldsymbol{\theta})}{\partial J_{ij}(a,b)} = f_{ij}(a,b) - \langle \delta(\sigma_i, a)\delta(\sigma_j, b) \rangle + \lambda_J J_{ij}(a,b)$$

Empirical Model

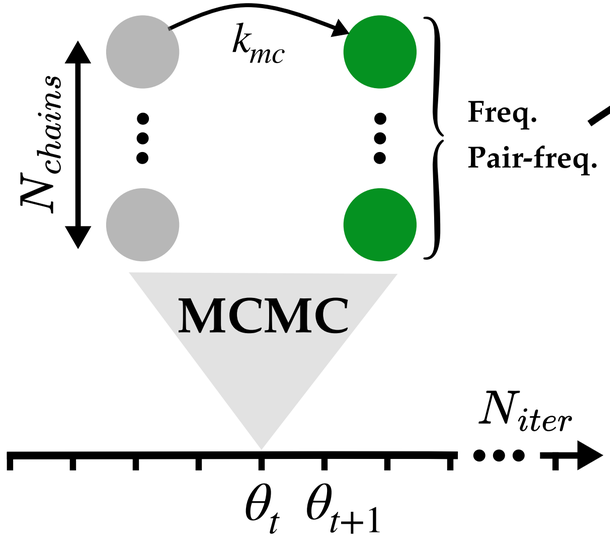
III. Stochastic Boltzmann Machine

Assess model performance with a toy model





$$\boldsymbol{\theta} = \{\boldsymbol{J}, \boldsymbol{h}\} \qquad f(\boldsymbol{\theta}) = \frac{1}{M}\sum_{m=1}^{M} \log P(\boldsymbol{\sigma}^{(m)} \mid \boldsymbol{\theta}) \; - \lambda_J \|\boldsymbol{J}\|^2 - \lambda_h \|\boldsymbol{h}\|^2$$



$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta_t \boldsymbol{p}_t$$

$$1^{st} \text{ order: } \boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$$

$$\frac{\partial f(\boldsymbol{\theta})}{\partial J_{ij}(a,b)} = f_{ij}(a,b) - \langle \delta(\sigma_i, a)\delta(\sigma_j, b) \rangle + \lambda_J J_{ij}(a,b)$$

Empirical Model

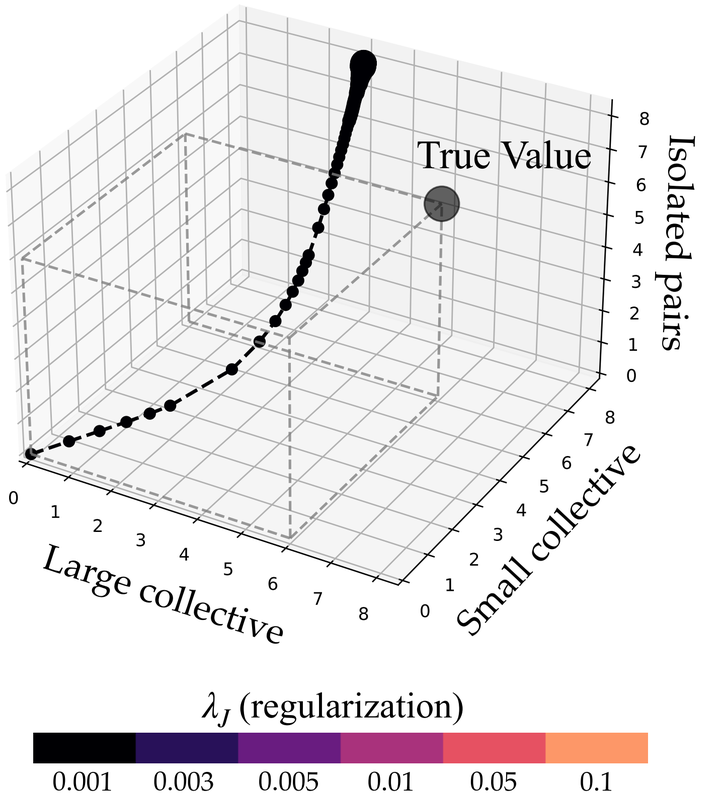
III. Stochastic Boltzmann Machine

Assess model performance with a toy model

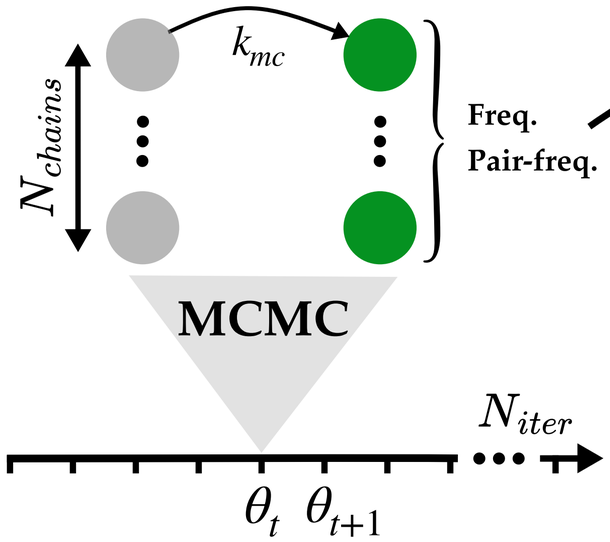
BM

L_2 regularization

1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$



$$\boldsymbol{\theta} = \{\boldsymbol{J}, \boldsymbol{h}\} \qquad f(\boldsymbol{\theta}) = \frac{1}{M}\sum_{m=1}^{M} \log P(\boldsymbol{\sigma}^{(m)} \mid \boldsymbol{\theta}) \; - \lambda_J \|\boldsymbol{J}\|^2 - \lambda_h \|\boldsymbol{h}\|^2$$



$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta_t \boldsymbol{p}_t$$

1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$

$$\frac{\partial f(\boldsymbol{\theta})}{\partial J_{ij}(a,b)} = f_{ij}(a,b) - \langle \delta(\sigma_i, a)\delta(\sigma_j, b)\rangle + \lambda_J J_{ij}(a,b)$$

Empirical Model

All models are inferred with $k_{\text{mc}} = 10^5$, $N_{\text{iter}} = 5000$

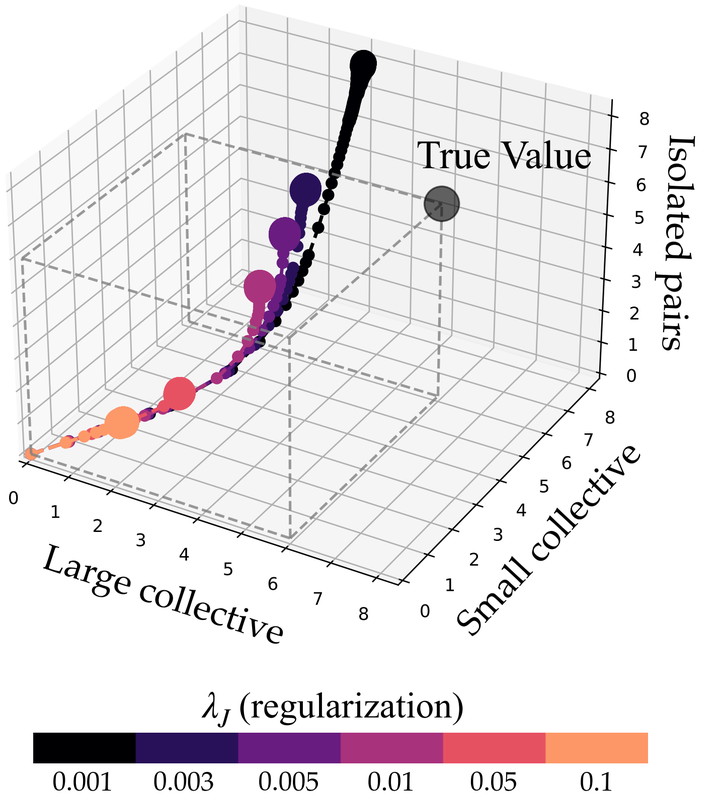
III. Stochastic Boltzmann Machine

Assess model performance with a toy model

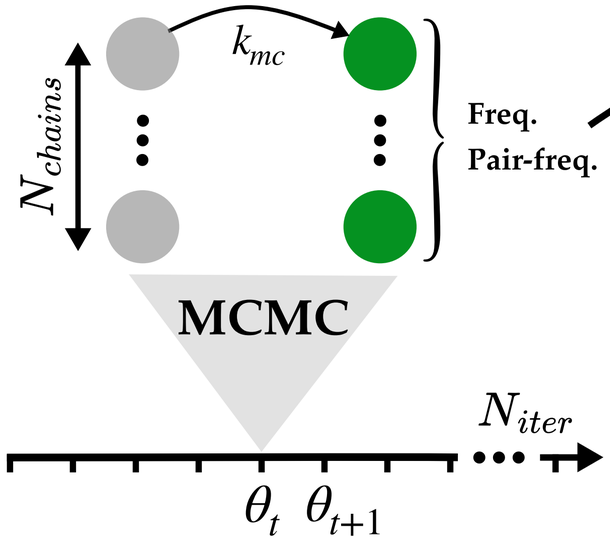
BM

L_2 regularization

1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$



$$\boldsymbol{\theta} = \{\boldsymbol{J}, \boldsymbol{h}\} \qquad f(\boldsymbol{\theta}) = \frac{1}{M}\sum_{m=1}^{M} \log P(\boldsymbol{\sigma}^{(m)} \mid \boldsymbol{\theta}) \; - \lambda_J \|\boldsymbol{J}\|^2 - \lambda_h \|\boldsymbol{h}\|^2$$



$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta_t \boldsymbol{p}_t$$

1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$

$$\frac{\partial f(\boldsymbol{\theta})}{\partial J_{ij}(a,b)} = f_{ij}(a,b) - \langle \delta(\sigma_i, a)\delta(\sigma_j, b)\rangle + \lambda_J J_{ij}(a,b)$$

Empirical Model

All models are inferred with $k_{\mathrm{mc}} = 10^5$, $N_{\mathrm{iter}} = 5000$

III. Stochastic Boltzmann Machine

BM

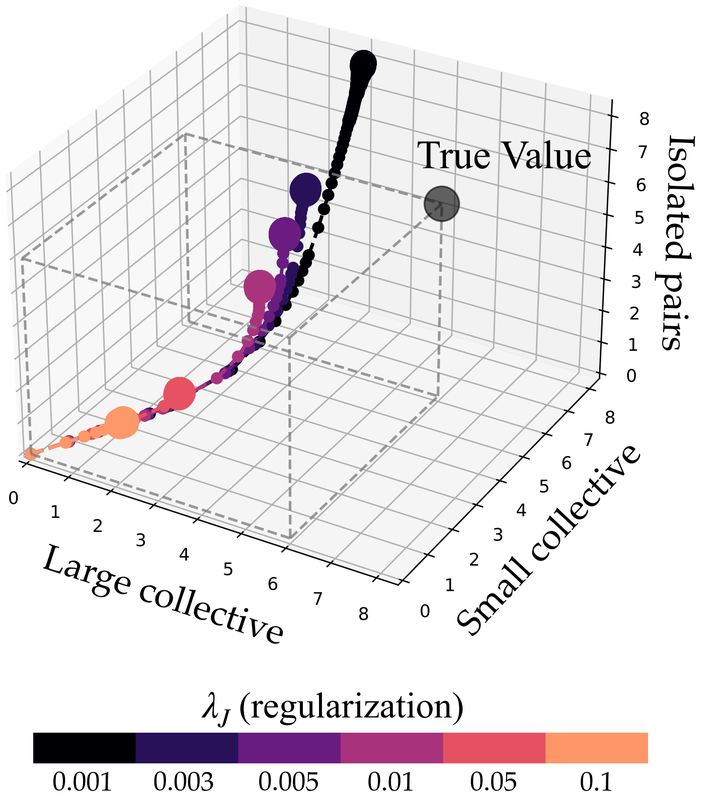
L_2 regularization

1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$

BM

L_2 regularization

2^{nd} order: $\boldsymbol{p}_t = -\boldsymbol{B}_k^{-1}\nabla f(\boldsymbol{\theta}_t)$



L-BFGS algorithm

Why?

- High dimensional space $\sim 10^4$
- 1^{st} order suboptimal because of anisotropic curvature

How?

- Curvature information from the *m* most recent iterations

$$\boldsymbol{p}_t = -\boldsymbol{B}_k^{-1}\nabla f(\boldsymbol{\theta}_t)$$

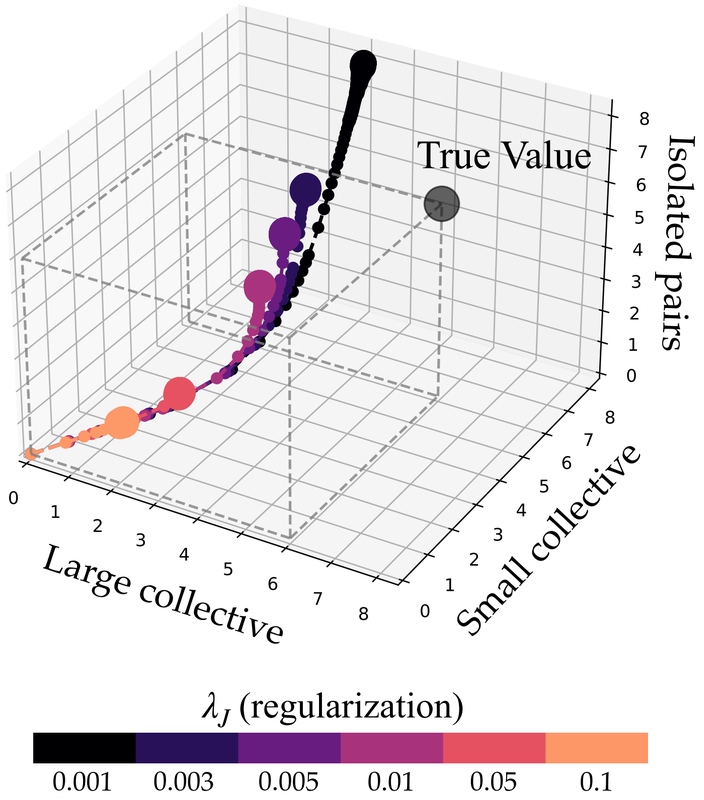
All models are inferred with $k_{\text{mc}} = 10^5$, $N_{\text{iter}} = 5000$

III. Stochastic Boltzmann Machine

BM

L_2 regularization

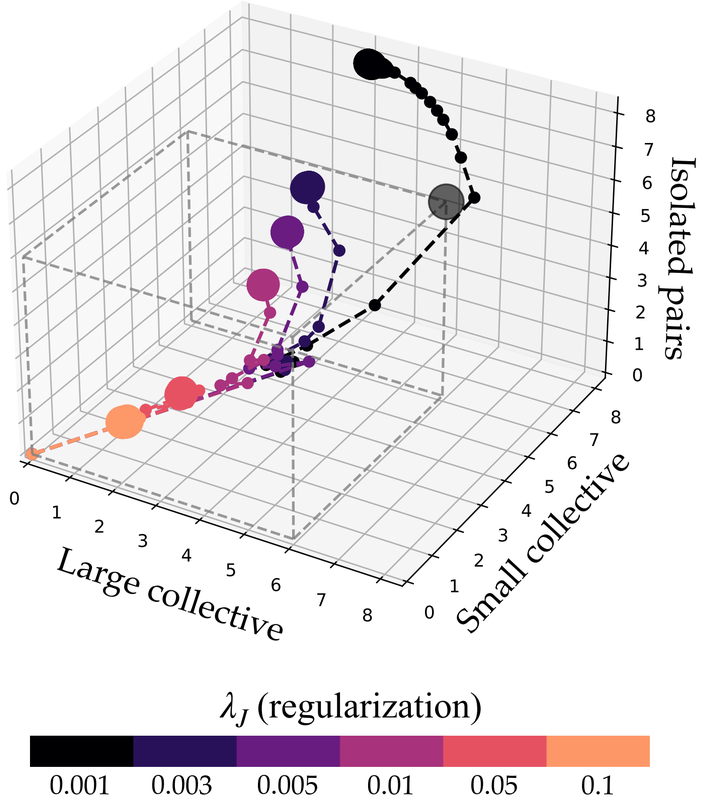
1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$



BM

L_2 regularization

2^{nd} order: $\boldsymbol{p}_t = -\boldsymbol{B}_k^{-1}\nabla f(\boldsymbol{\theta}_t)$



L-BFGS algorithm

Why?

- High dimensional space $\sim 10^4$
- 1^{st} order suboptimal because of anisotropic curvature

How?

- Curvature information from the *m* most recent iterations

$$\boldsymbol{p}_t = -\boldsymbol{B}_k^{-1}\nabla f(\boldsymbol{\theta}_t)$$

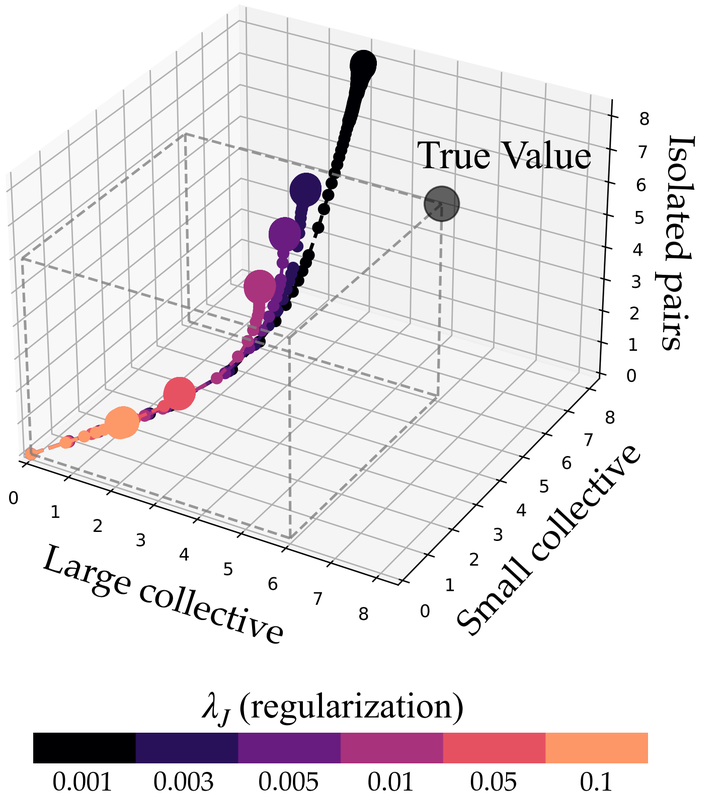
All models are inferred with $k_{\text{mc}} = 10^5$, $N_{\text{iter}} = 5000$

III. Stochastic Boltzmann Machine

BM

L_2 regularization

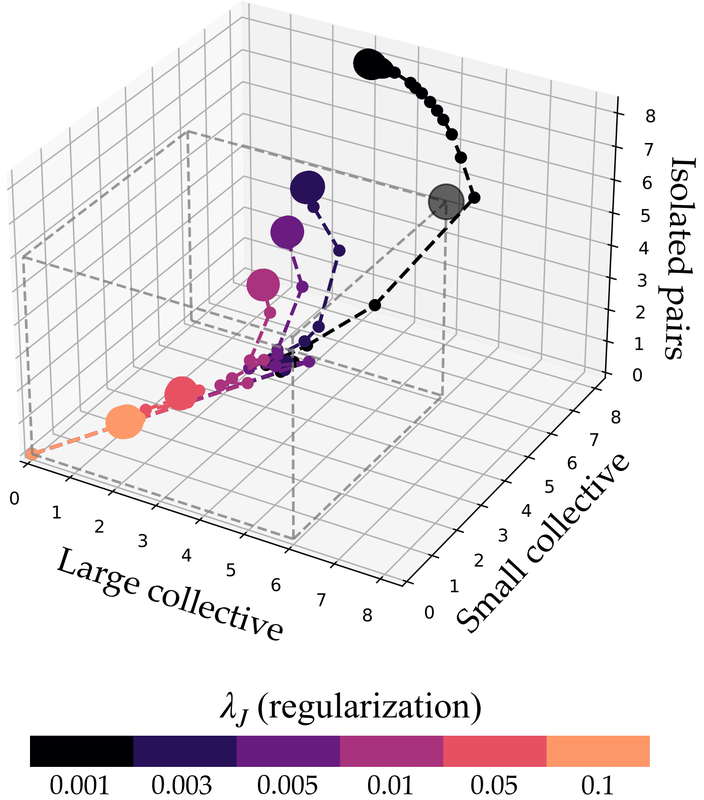
1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$



BM

L_2 regularization

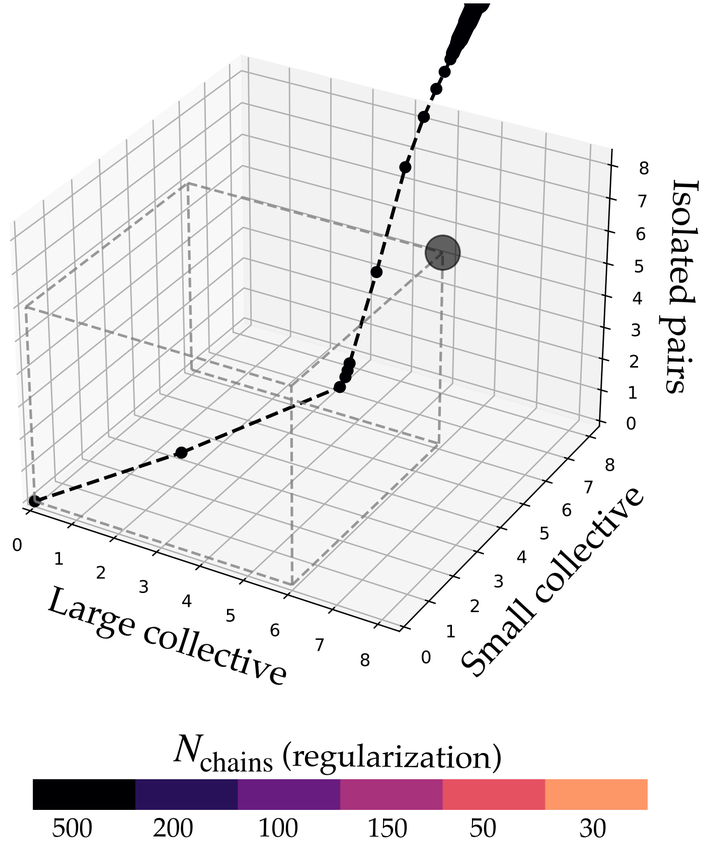
2^{nd} order: $\boldsymbol{p}_t = -\boldsymbol{B}_k^{-1}\nabla f(\boldsymbol{\theta}_t)$



Stochastic Boltzmann Machine

~~L_2 regularization~~

2^{nd} order: $\boldsymbol{p}_t = -\boldsymbol{B}_k^{-1}\nabla f(\boldsymbol{\theta}_t)$



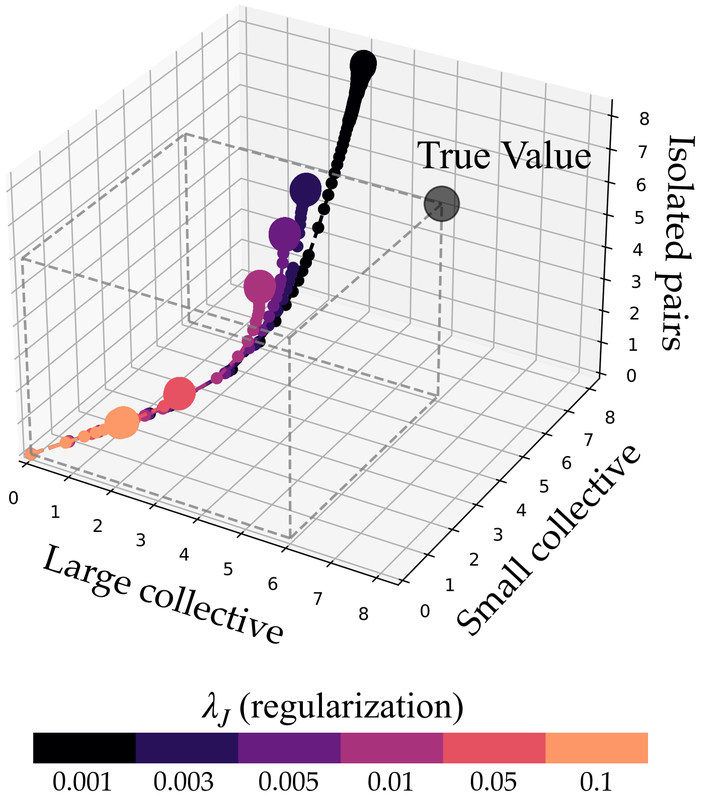
All models are inferred with $k_{\text{mc}} = 10^5$, $N_{\text{iter}} = 5000$

III. Stochastic Boltzmann Machine

BM

L_2 regularization

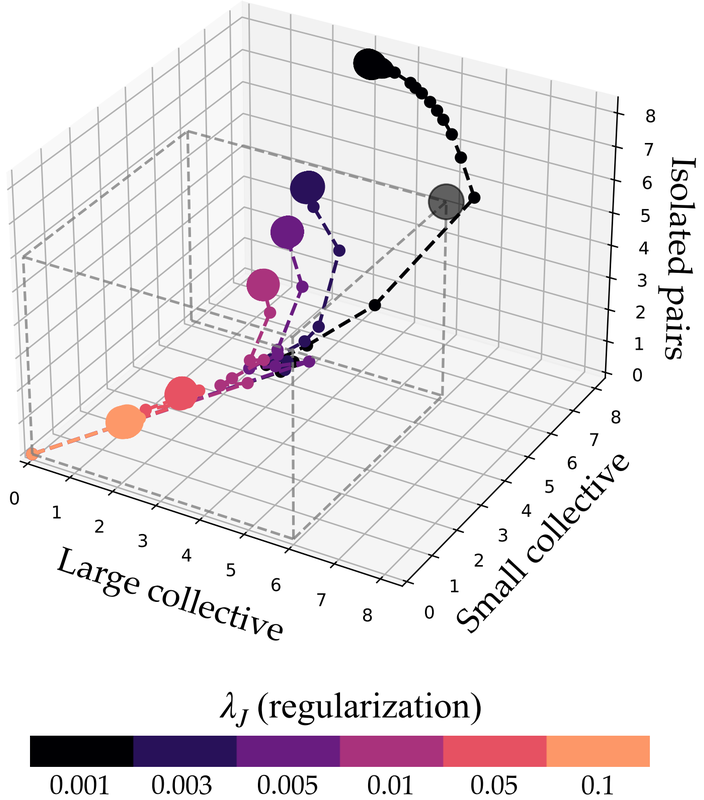
1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$



BM

L_2 regularization

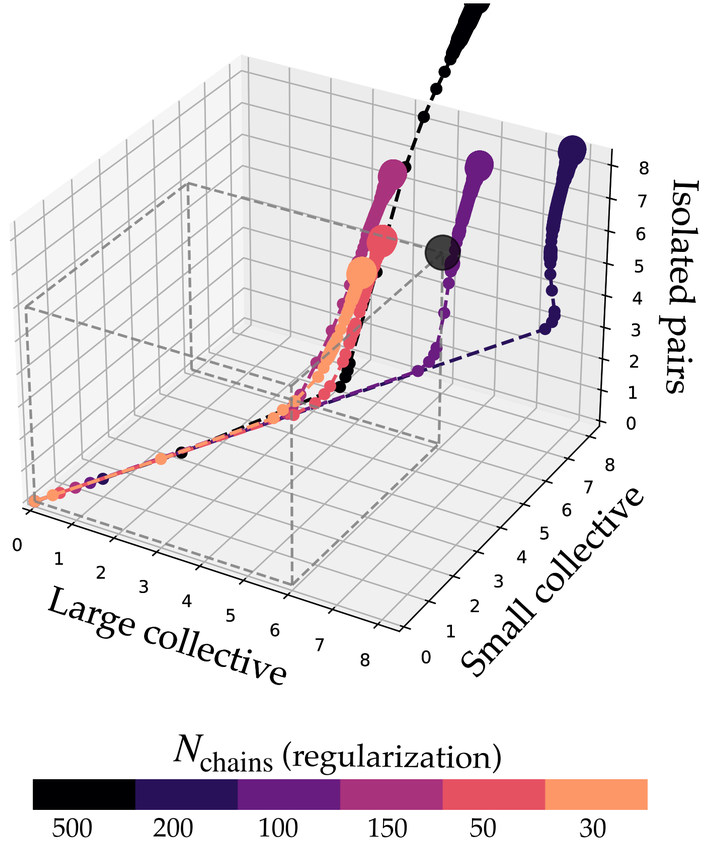
2^{nd} order: $\boldsymbol{p}_t = -\boldsymbol{B}_k^{-1}\nabla f(\boldsymbol{\theta}_t)$



Stochastic Boltzmann Machine

~~L_2 regularization~~

2^{nd} order: $\boldsymbol{p}_t = -\boldsymbol{B}_k^{-1}\nabla f(\boldsymbol{\theta}_t)$



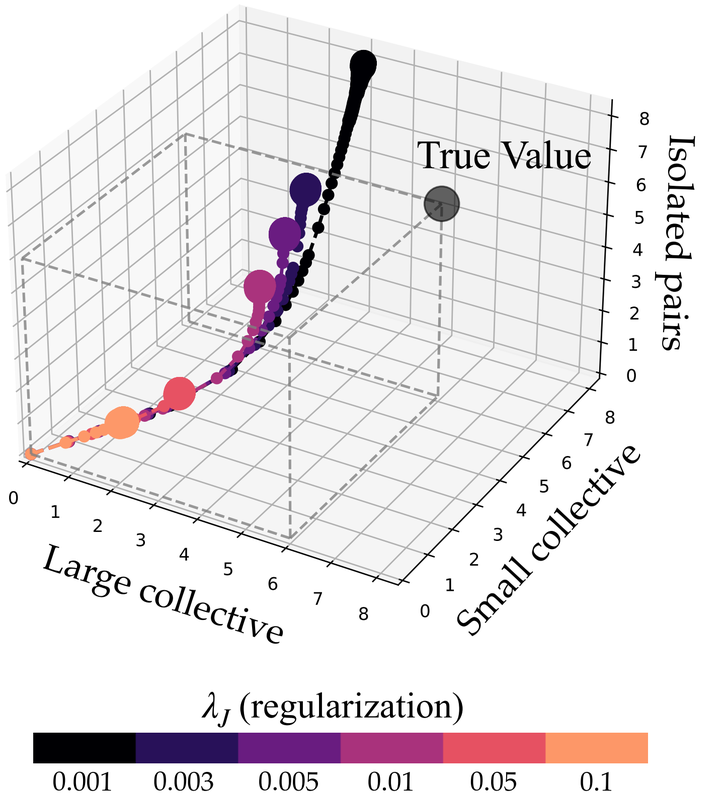
All models are inferred with $k_{\text{mc}} = 10^5$, $N_{\text{iter}} = 5000$

III. Stochastic Boltzmann Machine

BM

L_2 regularization

1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$



λ_J (regularization)

0.001 0.003 0.005 0.01 0.05 0.1

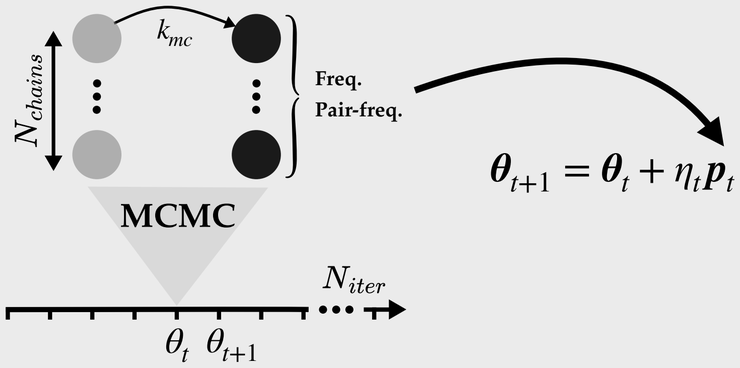
Lowering N_{chains}

Observation

- Training BM: capture statistics
- But statistics lack reliability because of undersampling

Strategy

- Intentionally undersample the model to mirror data undersampling

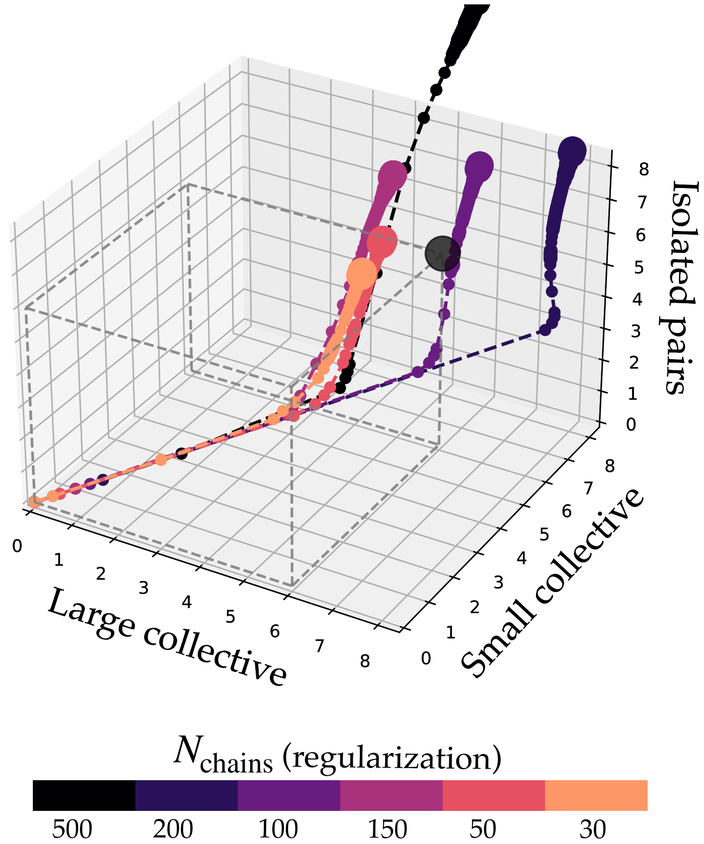


$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \eta_t \boldsymbol{p}_t$

Stochastic Boltzmann Machine

~~L_2 regularization~~

2^{nd} order: $\boldsymbol{p}_t = -\boldsymbol{B}_k^{-1} \nabla f(\boldsymbol{\theta}_t)$



N_{chains} (regularization)

500 200 100 150 50 30

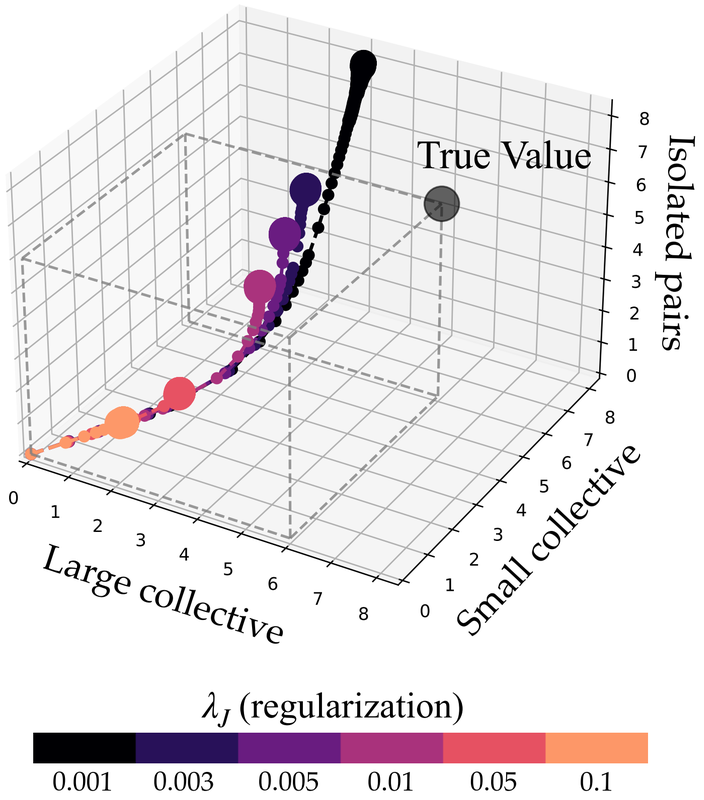
All models are inferred with $k_{\text{mc}} = 10^5$, $N_{\text{iter}} = 5000$

III. Stochastic Boltzmann Machine

BM

L_2 regularization

1^{st} order: $\boldsymbol{p}_t = -\nabla f(\boldsymbol{\theta}_t)$



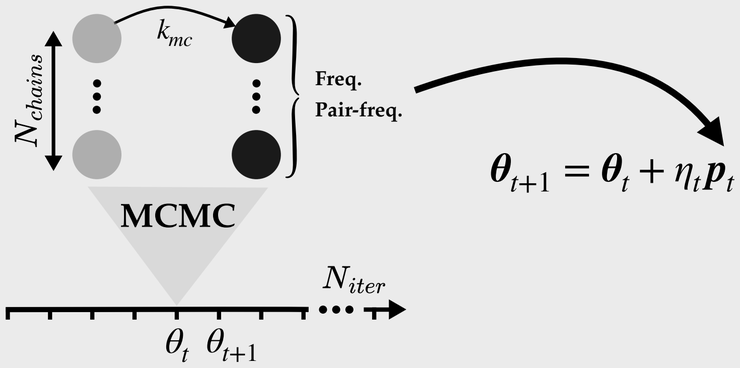
Lowering N_{chains}

Observation

- Training BM: capture statistics
- But statistics lack reliability because of undersampling

Strategy

- Intentionally undersample the model to mirror data undersampling

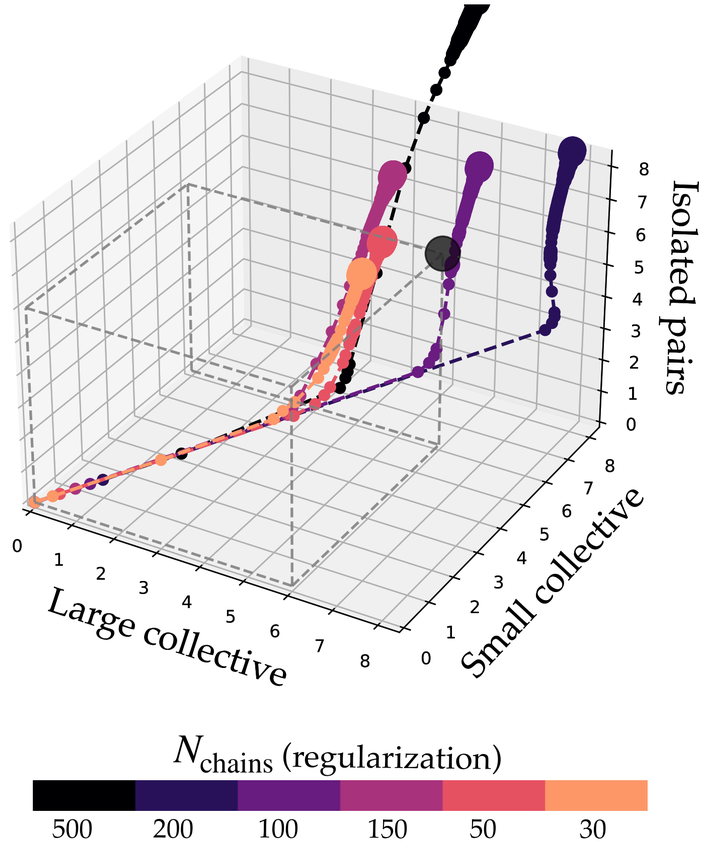


$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \eta_t \boldsymbol{p}_t$$

Stochastic Boltzmann Machine

~~L_2 regularization~~

2^{nd} order: $\boldsymbol{p}_t = -\boldsymbol{B}_k^{-1}\nabla f(\boldsymbol{\theta}_t)$

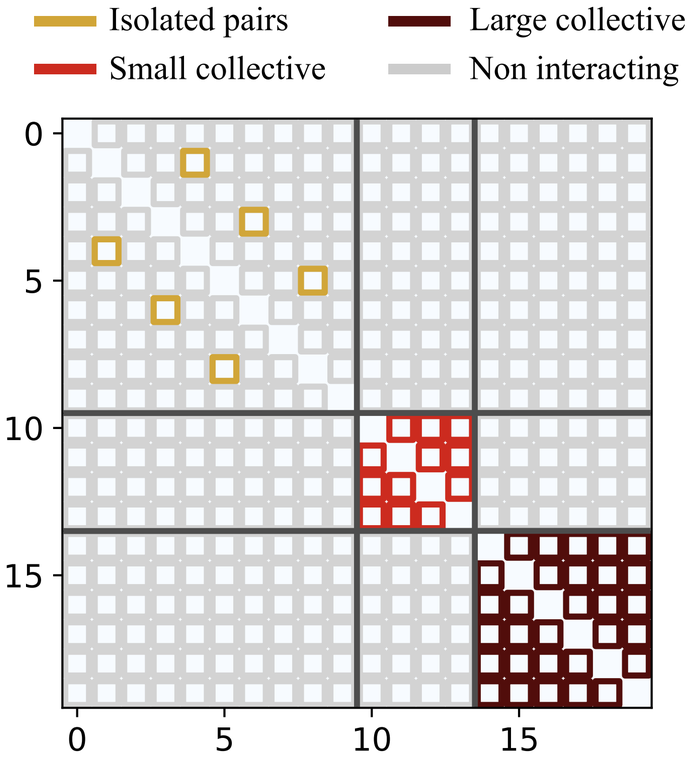


- Model undersampling: N_{chains}
- Slow down convergence: $m \sim 1$
- Early stopping: N_{iter}

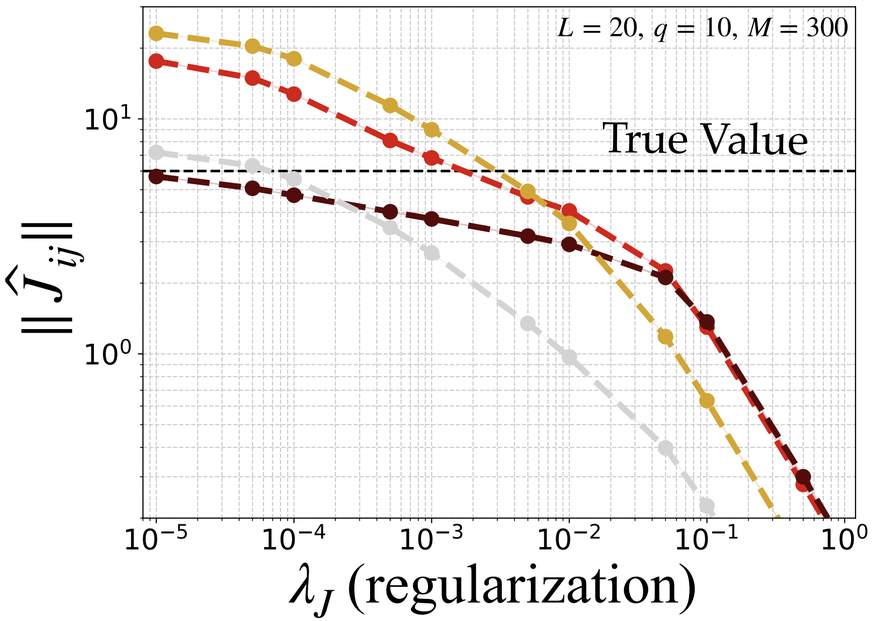
All models are inferred with $k_{\text{mc}} = 10^5$, $N_{\text{iter}} = 5000$

III. Stochastic Boltzmann Machine

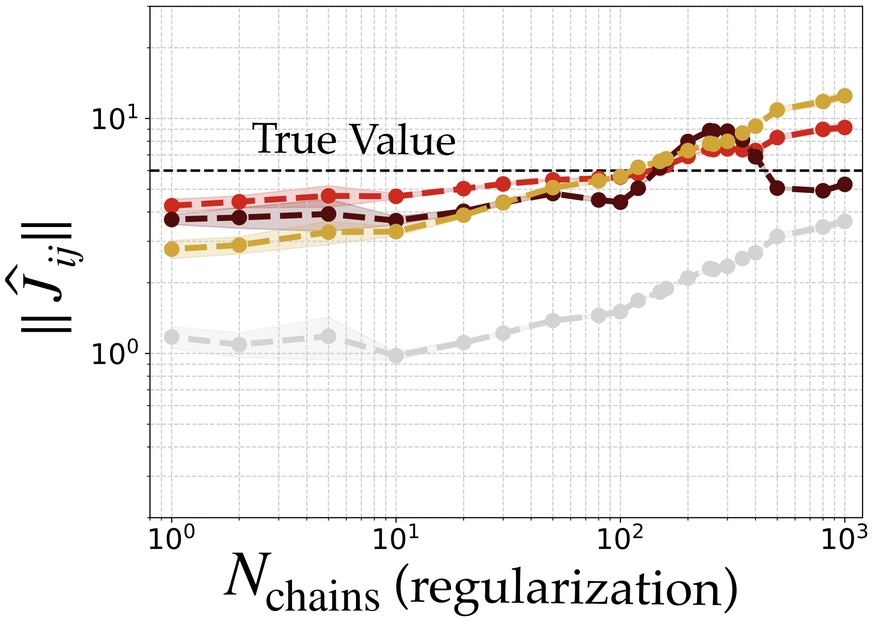
Assess model performance with a toy model



BM



SBM



Stochastic Boltzmann Machine

- Toy model: correct undersampling-induced biases
- Real data?

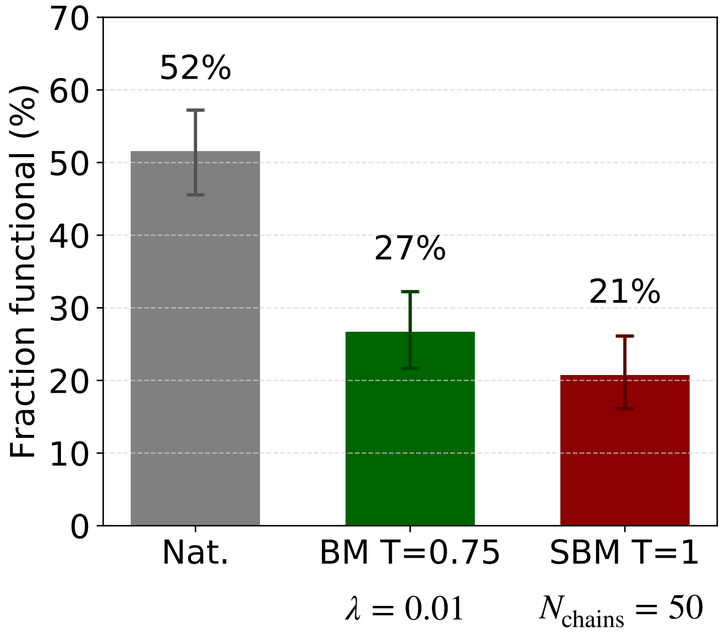
Figures made following the methodology of Kleeorin *et al.*

III. Stochastic Boltzmann Machine

Relevance to real proteins: application to the chorismate mutase family

Fidelity

Fraction of functional sequences



Novelty

Distribution of identity to the nearest natural sequence

Diversity

% of taxonomic families represented

The experiments were performed by Emily Hinds

The models are inferred with $k_{\text{mc}} = 10^5$, $N_{\text{iter}} = 400$, $\theta = 0.3$

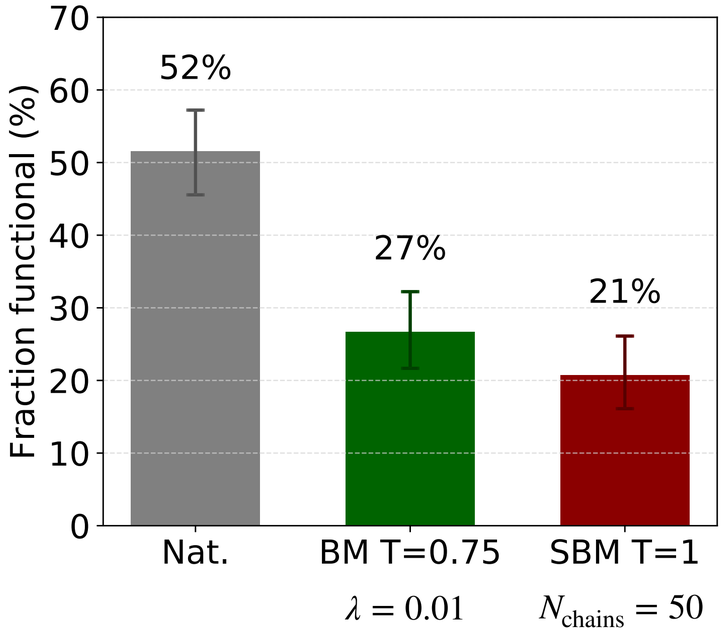
890, 193 and 180 sequences were tested for the natural dataset, the SBM and the BM models respectively

III. Stochastic Boltzmann Machine

Relevance to real proteins: application to the chorismate mutase family

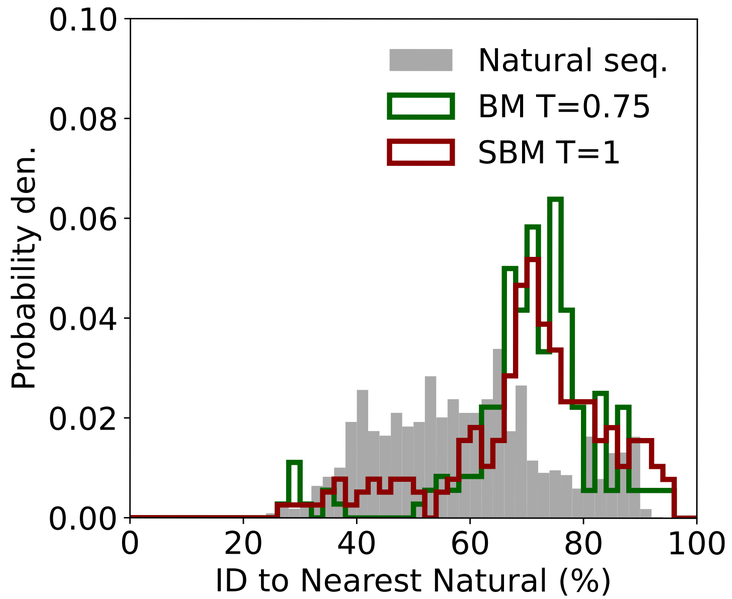
Fidelity

Fraction of functional sequences



Novelty

Distribution of identity to the nearest natural sequence



Diversity

% of taxonomic families represented

The experiments were performed by Emily Hinds

The models are inferred with $k_{\text{mc}} = 10^5$, $N_{\text{iter}} = 400$, $\theta = 0.3$

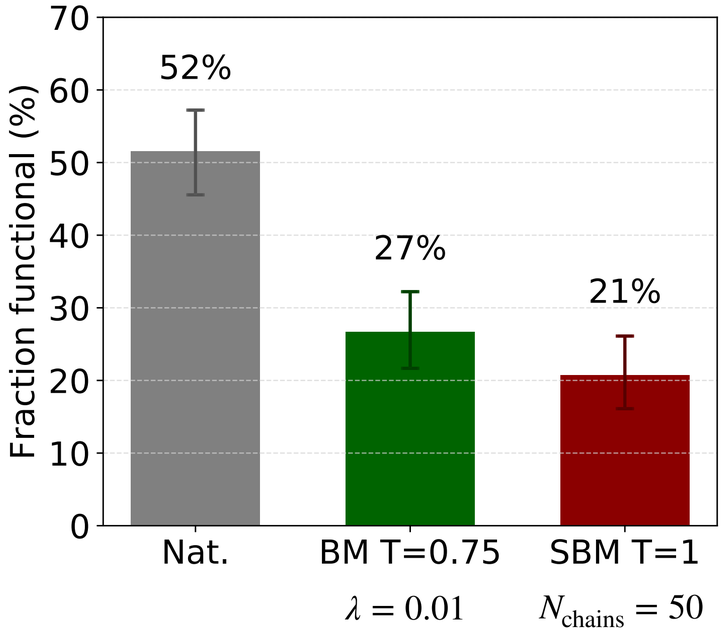
890, 193 and 180 sequences were tested for the natural dataset, the SBM and the BM models respectively

III. Stochastic Boltzmann Machine

Relevance to real proteins: application to the chorismate mutase family

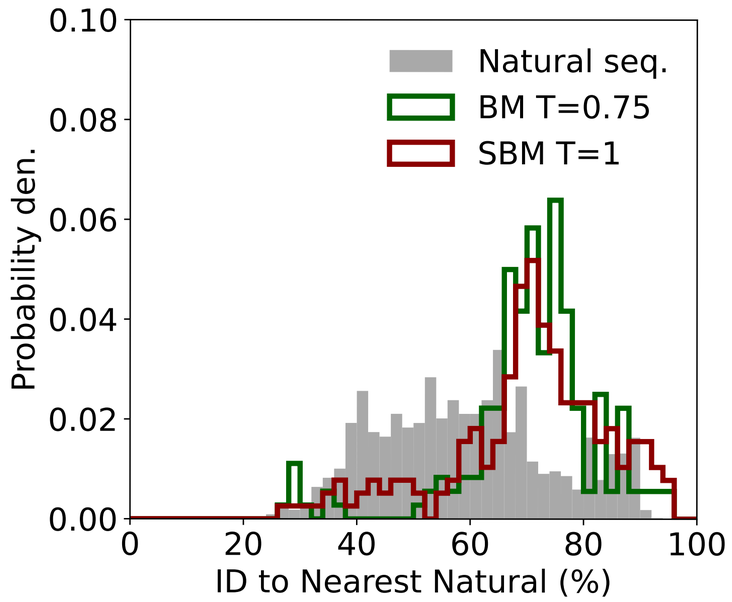
Fidelity

Fraction of functional sequences



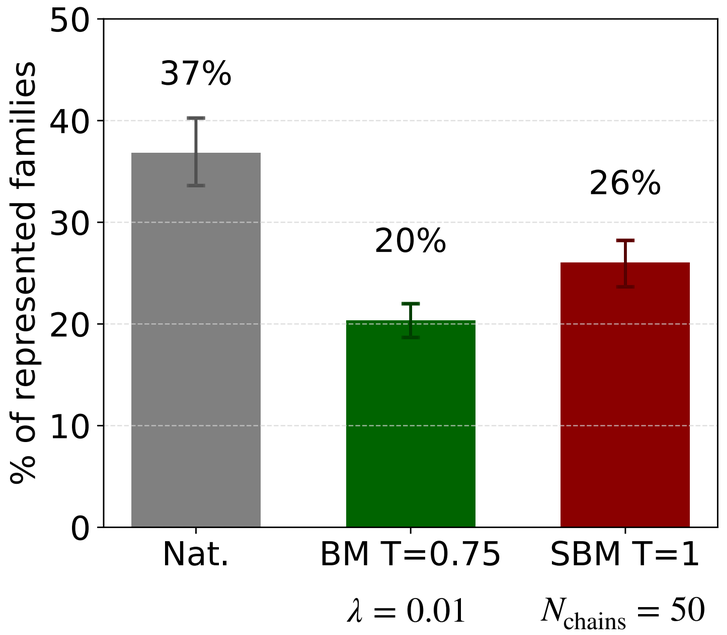
Novelty

Distribution of identity to the nearest natural sequence



Diversity

% of taxonomic families represented



Stochastic Boltzmann Machine

- Toy model: correct undersampling-induced biases
- Real data: generative without low-temperature sampling

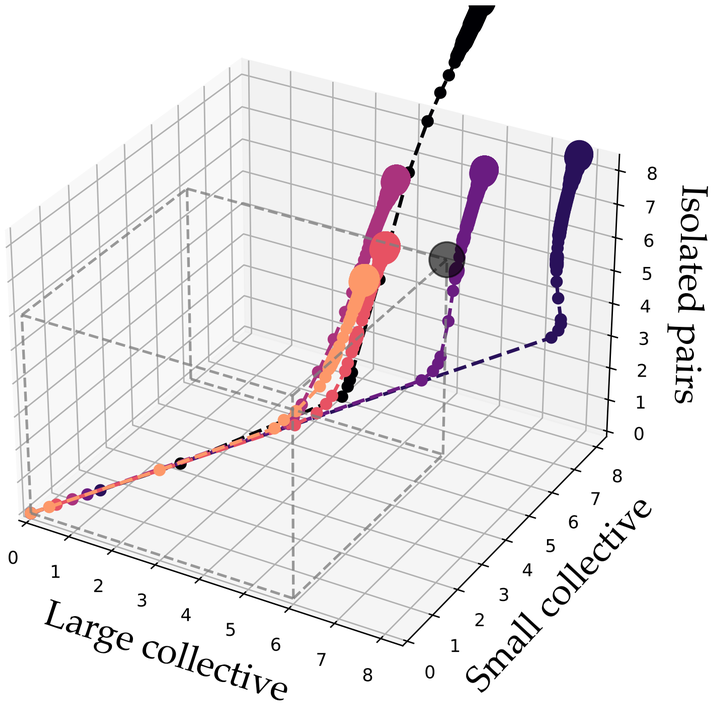
The experiments were performed by Emily Hinds

The models are inferred with $k_{\text{mc}} = 10^5$, $N_{\text{iter}} = 400$, $\theta = 0.3$

890, 193 and 180 sequences were tested for the natural dataset, the SBM and the BM models respectively

The undersampling problem

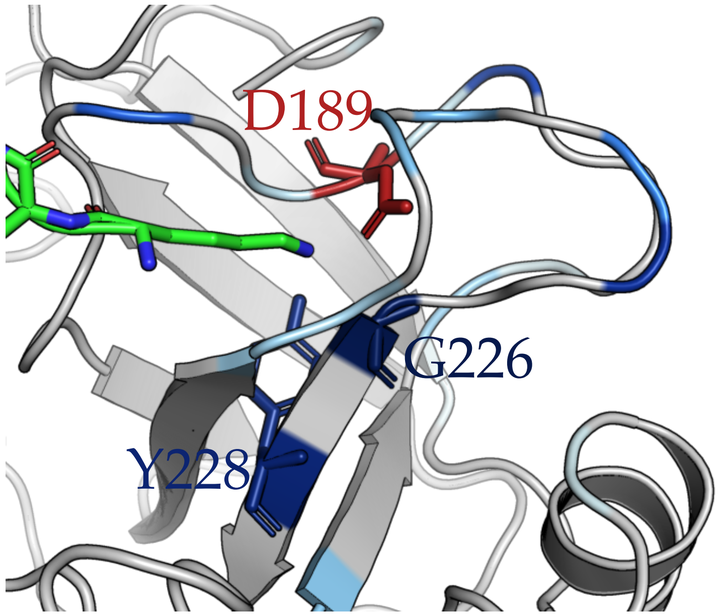
Overcoming undersampling induced biases



In collaboration with Emily Hinds, Yaakov Kleeorin, Rama Ranganathan (University of Chicago, USA)

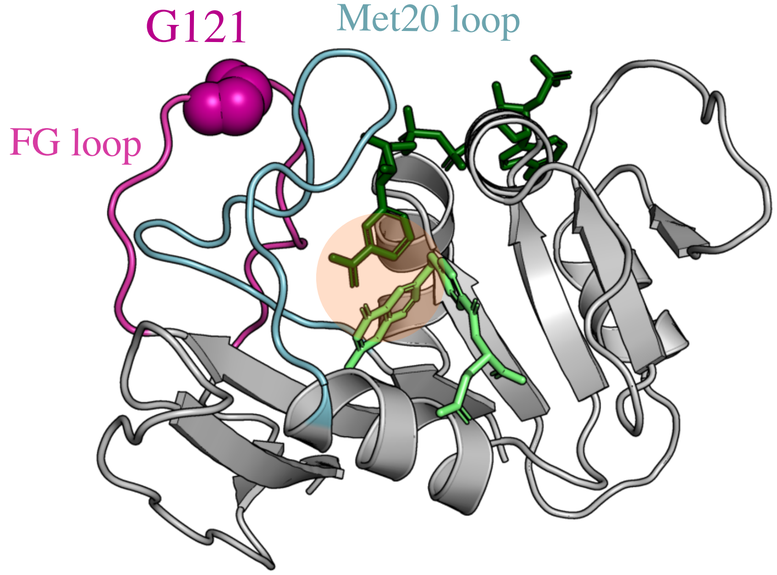
Investigate protein properties with statistical learning

Specificity mechanism in S1A family



In collaboration with Amaury Paveyranne, Timothé Lucas, Shoichi Yip, Clément Nizak (LJP, Sorbonne University, France)

Allosteric network of *E. Coli* DHFR



In collaboration with Paul Guenon, Damien Laage, Guillaume Stirnemann (ENS, France), Clément Nizak (LJP, France), Karolina Filipowska, Kim Reynolds (University of Texas, USA)

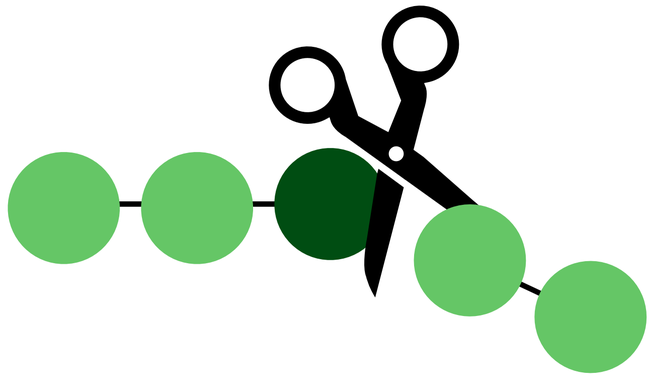
Investigate the determinant of specificity within S1A family

In collaboration with Amaury Paveyranne, Timothé Lucas, Shoichi Yip, Clément Nizak (LJP, Sorbonne University, France)

S1A serine proteases protein family

The basics

- Catalyzes peptide bound hydrolysis
- Many substrate specificities



Trypsin

Positively charged

Chymotrypsin

Aromatic

Di Cera, IUBMB **2009**
Hedstrom, Chemical reviews **2002**
Barrett *et al.*, **2012**

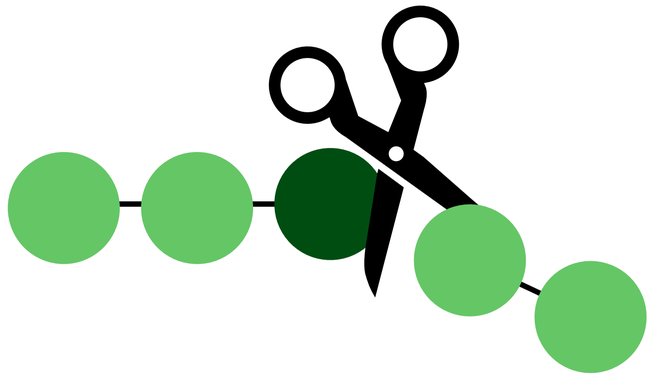
S1A serine proteases protein family

The basics

- Catalyzes peptide bound hydrolysis
- Many substrate specificities

Specificity

- Cleavage after a specific amino-acid
- Efficiency can vary up to 10^5-fold depending on the amino acid
- Mechanism not fully understood

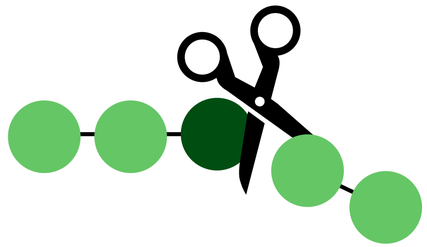


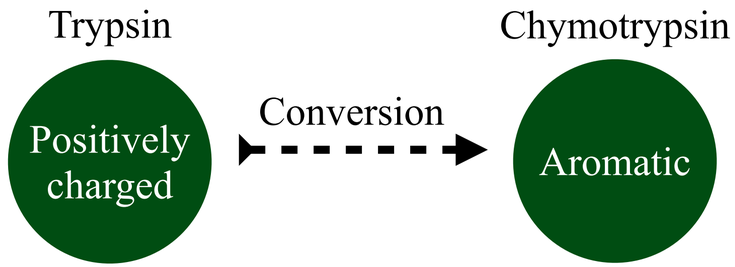
Trypsin

Chymotrypsin

Di Cera, IUBMB **2009**
Hedstrom, Chemical reviews **2002**
Barrett *et al.*, **2012**

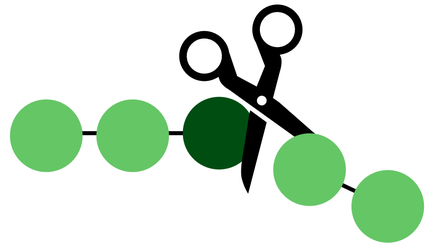
A rare success in specificity conversion: Trypsin → Chymotrypsin





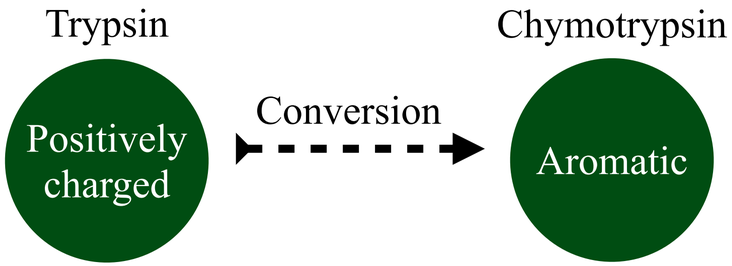
Gráf et al., PNAS **1988**
Hedstrom et al., *Science* **1992**
Hedstrom et al., *Biochemistry* **1994**

A rare success in specificity conversion: Trypsin → Chymotrypsin

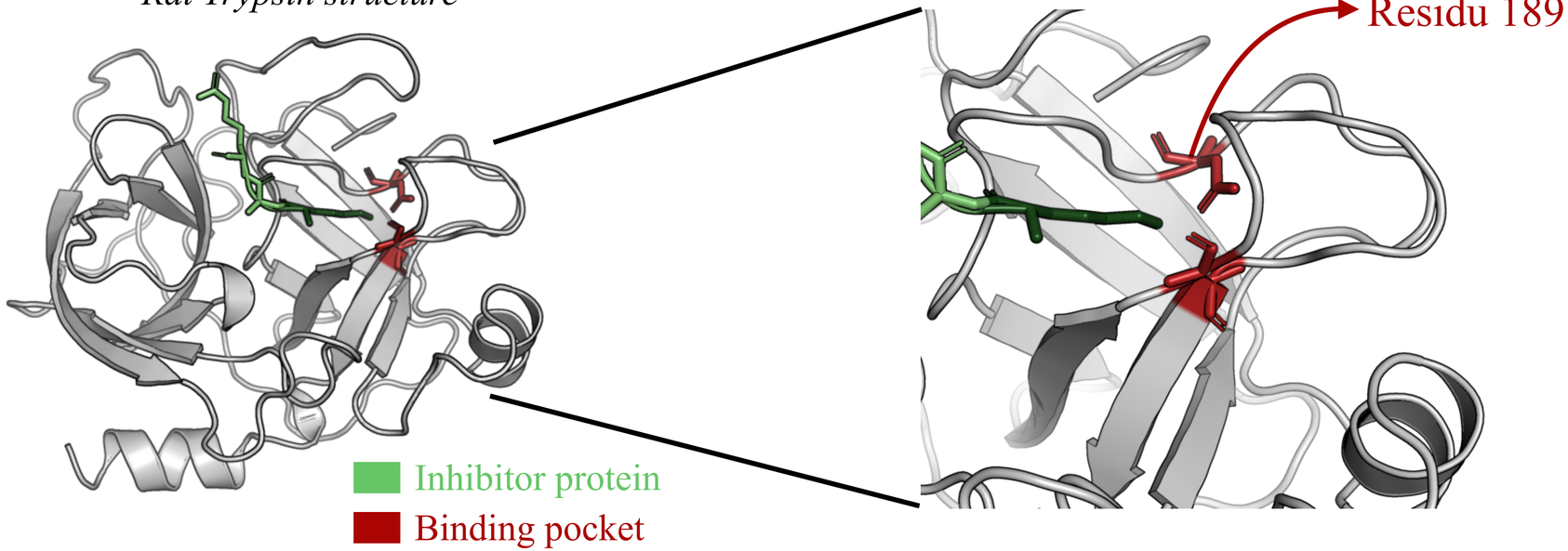


Structural approach

- Focus on binding pocket
- Residu 189 sufficient to convert specificity? No (Gráf *et al.* 1988)

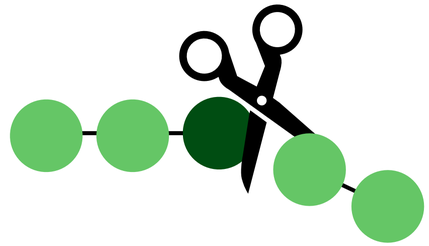


Rat Trypsin structure



Gráf et al., PNAS **1988**
Hedstrom et al., *Science* **1992**
Hedstrom et al., *Biochemistry* **1994**

A rare success in specificity conversion: Trypsin → Chymotrypsin

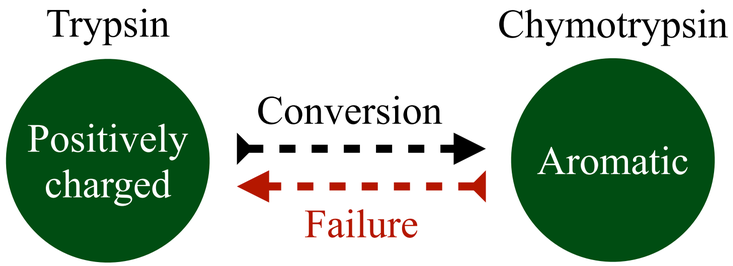


Structural approach

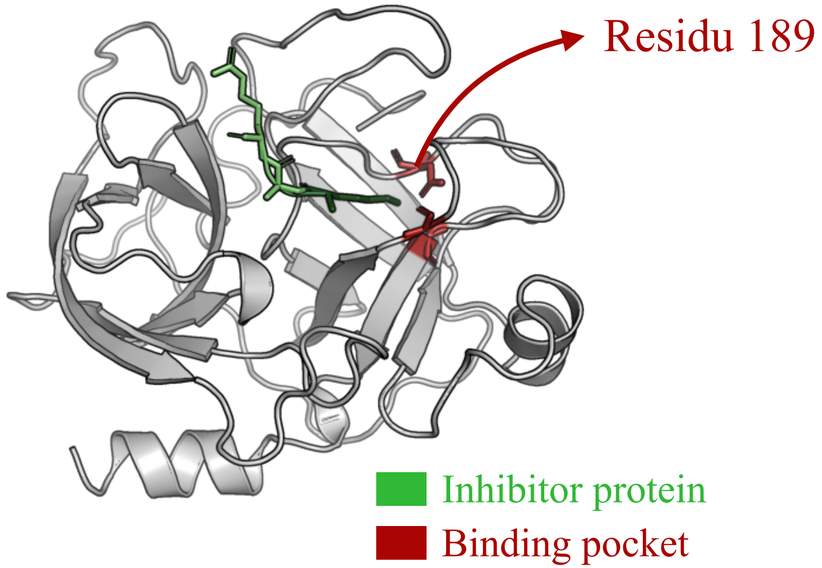
- Focus on binding pocket
- Residu 189 sufficient to convert specificity? No (Gráf *et al.* 1988)

Evolutionary approach

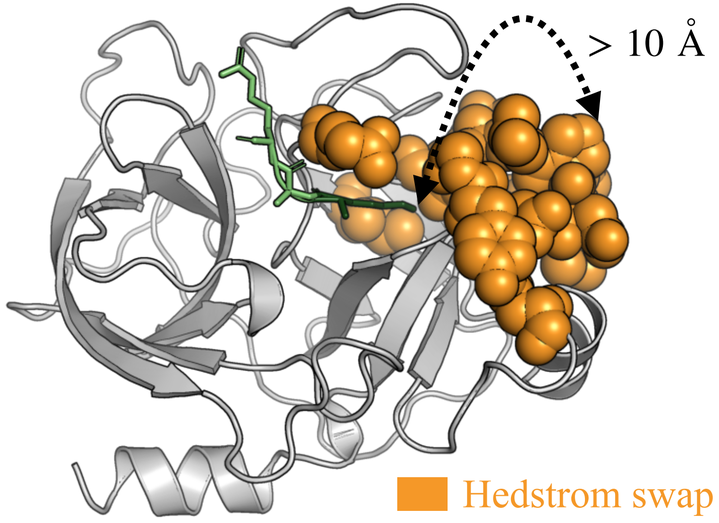
- Analysis of trypsin and chymotrypsin sequences
- Conversion with 16 mutations (Hedstrom *et al.* 1994)



Rat Trypsin structure

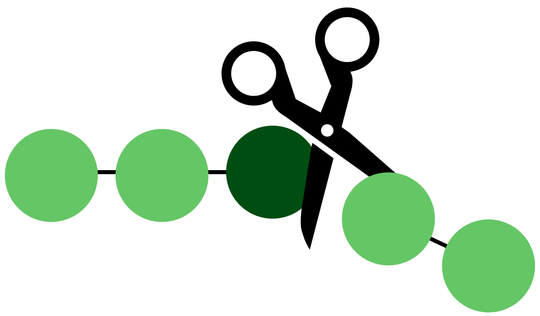


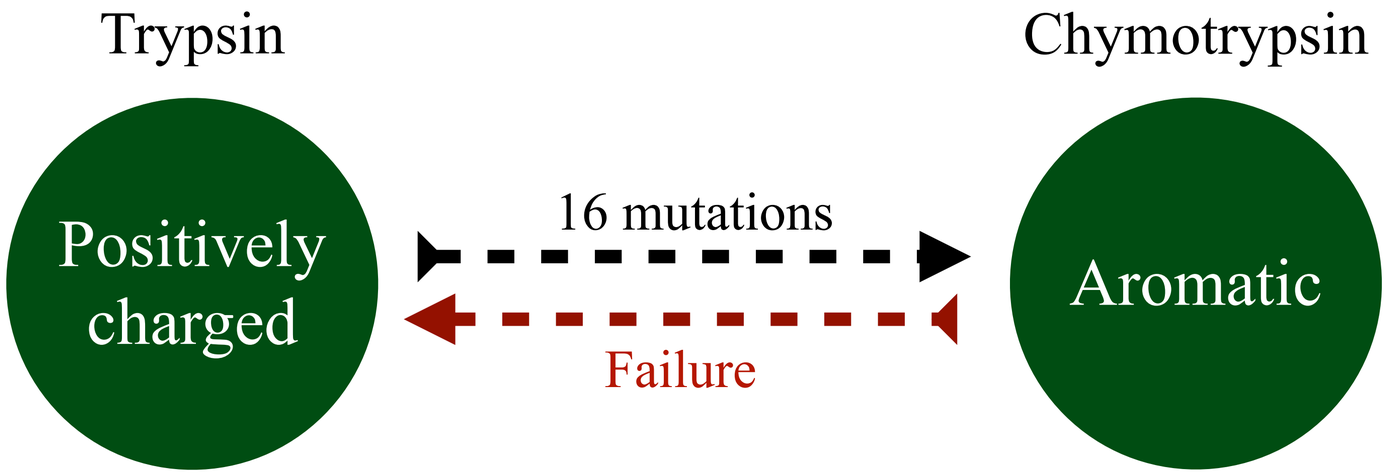
Rat Trypsin structure



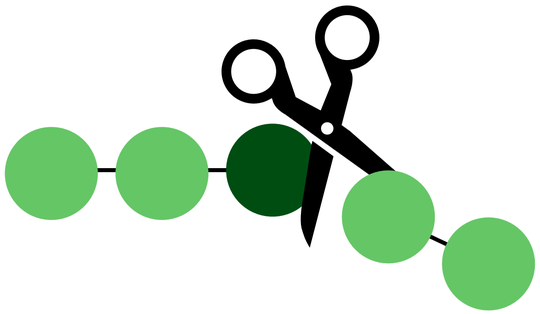
Gráf et al., PNAS **1988**
Hedstrom et al., *Science* **1992**
Hedstrom et al., *Biochemistry* **1994**

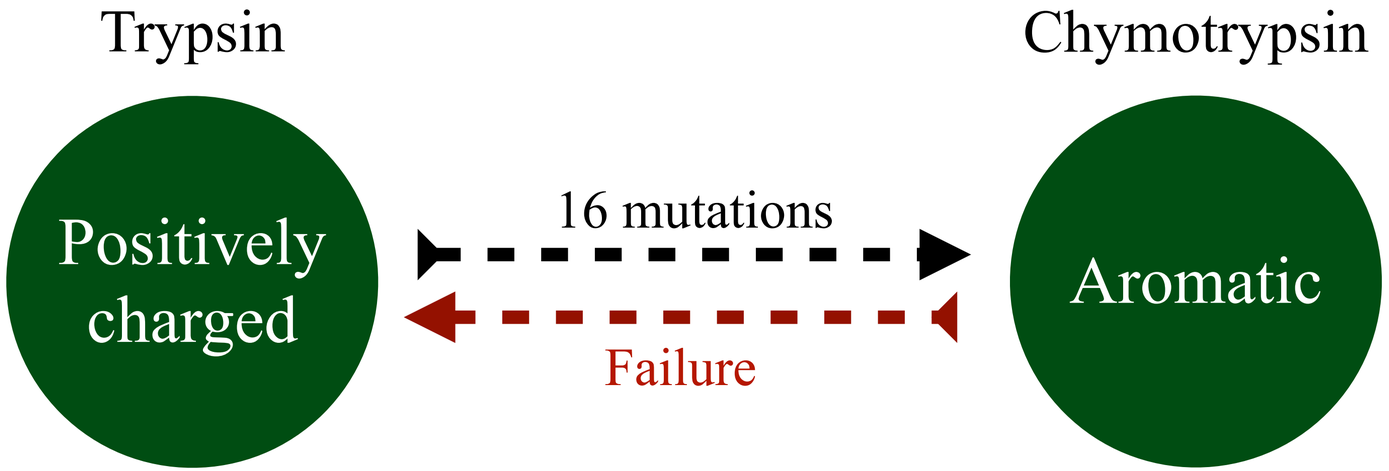
Boltzmann Machine-guided specificity conversion





Boltzmann Machine-guided specificity conversion

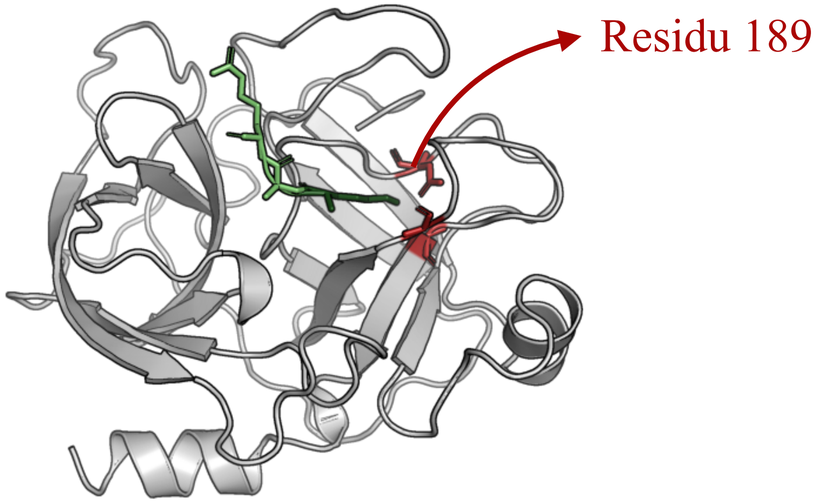




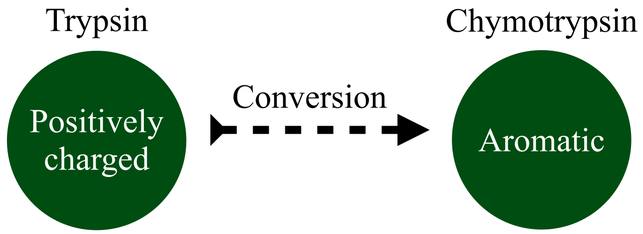
Our approach

- Impose a mutation at site 189
- Predict compensatory mutations using Boltzmann Machine model
- Identify positive epistatic interactions using model couplings

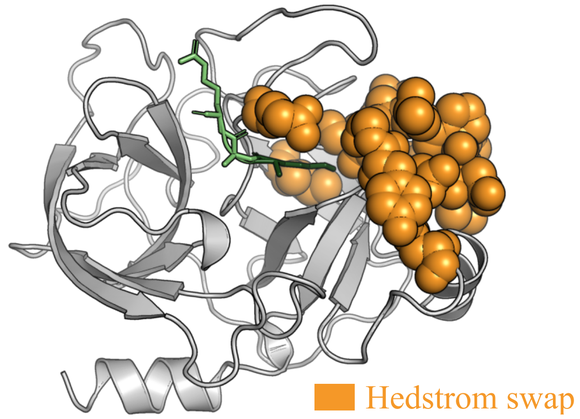
Rat Trypsin structure

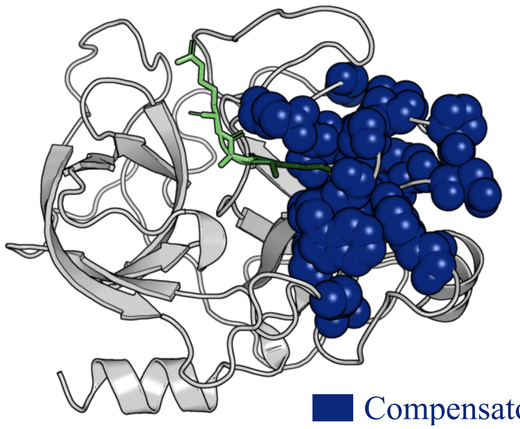


Boltzmann Machine-guided specificity conversion

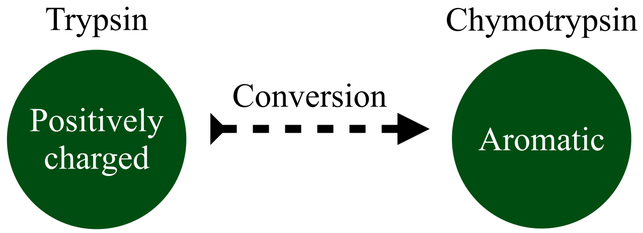


In silico comparison with Hedstrom

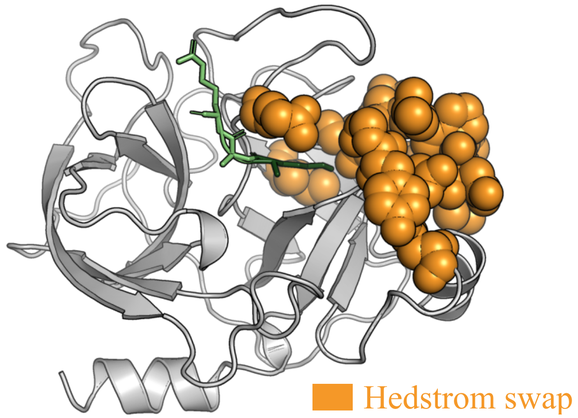


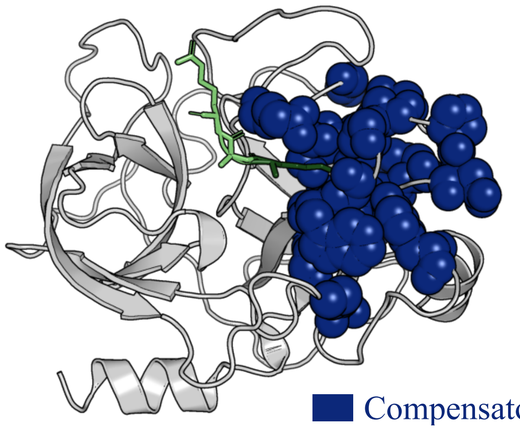


Boltzmann Machine-guided specificity conversion



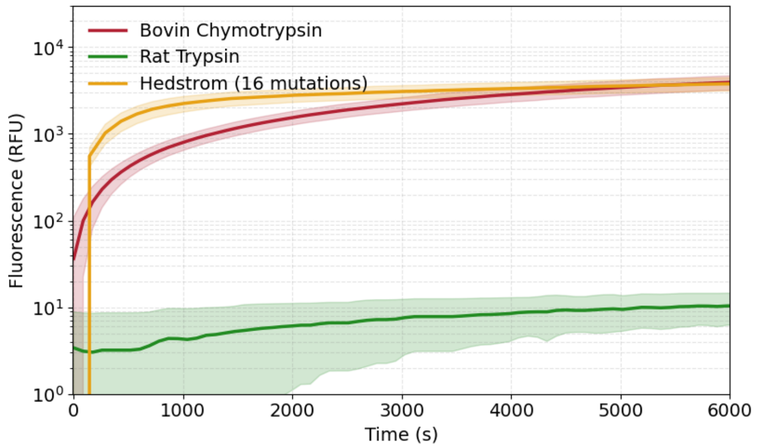
In silico comparison with Hedstrom

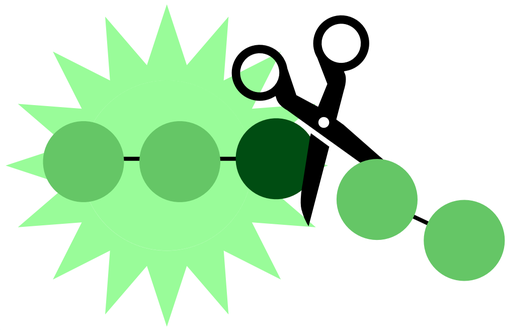




Preliminary experimental results

Chymotrypsin activity

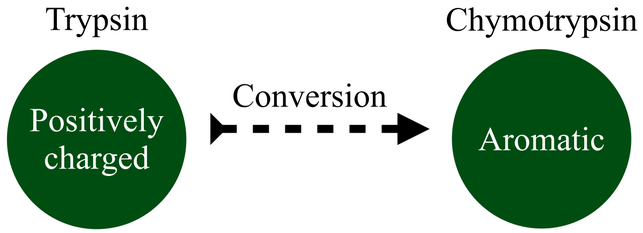




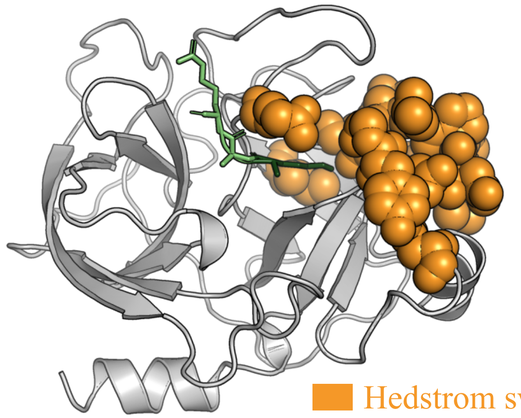
Not possible to make quantitative comparisons here

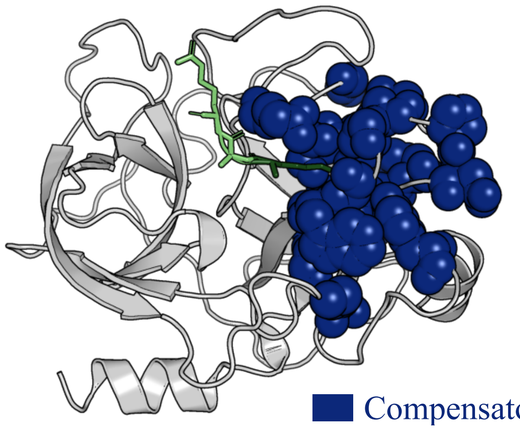
The experiments were performed by Amaury Paveyranne, Timothé Lucas and Shoichi Yip

Boltzmann Machine-guided specificity conversion



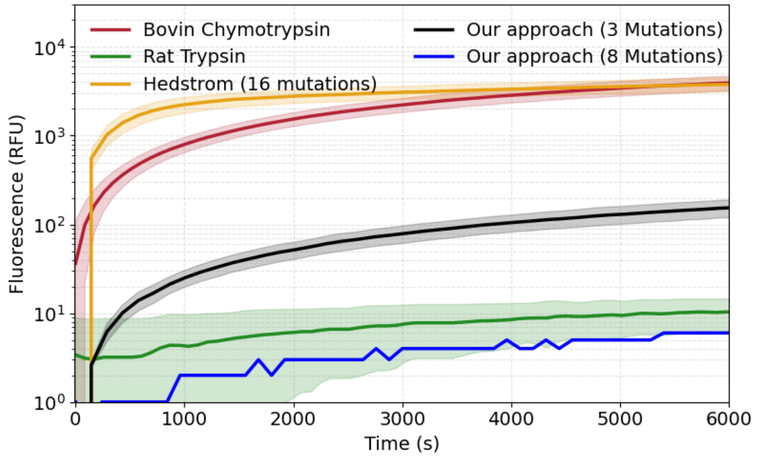
In silico comparison with Hedstrom

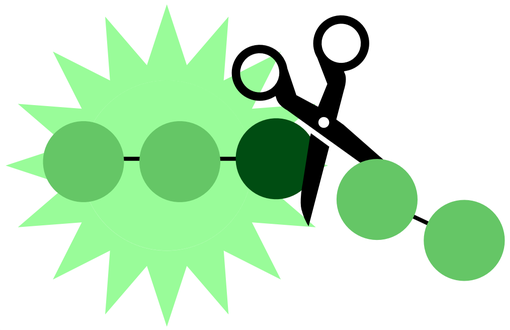




Preliminary experimental results

Chymotrypsin activity



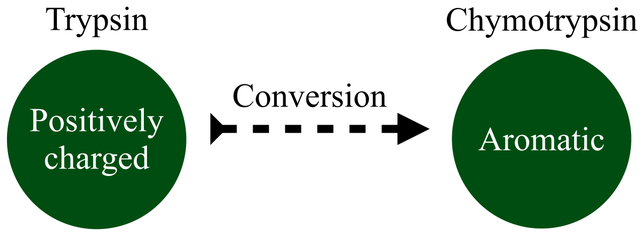


Not possible to make quantitative comparisons here

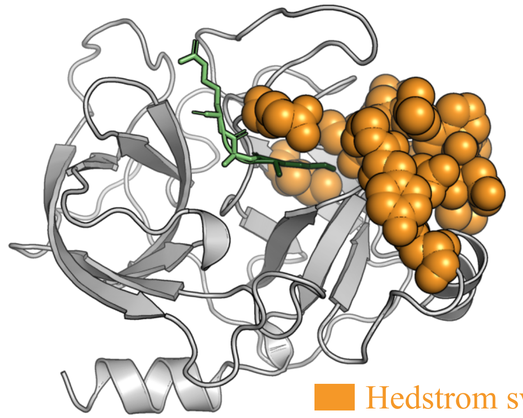
- Evidence of chymotrypsin activity 3 mutations away from rat trypsin

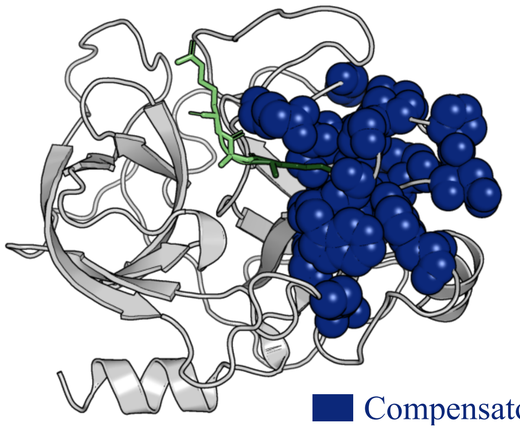
The experiments were performed by Amaury Paveyranne, Timothé Lucas and Shoichi Yip

Boltzmann Machine-guided specificity conversion



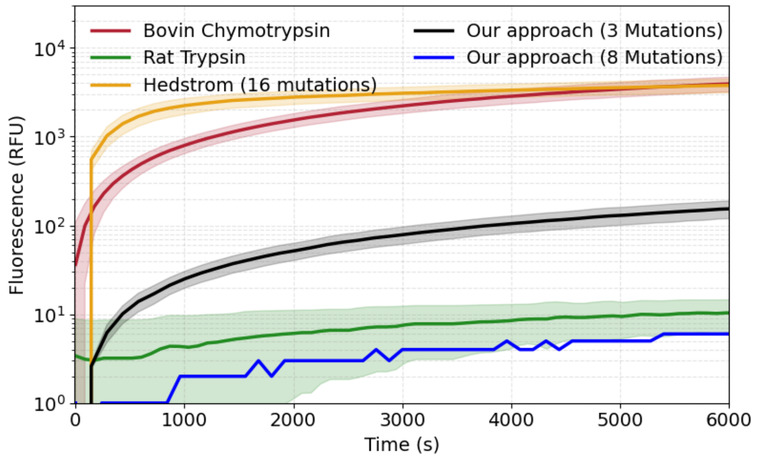
In silico comparison with Hedstrom



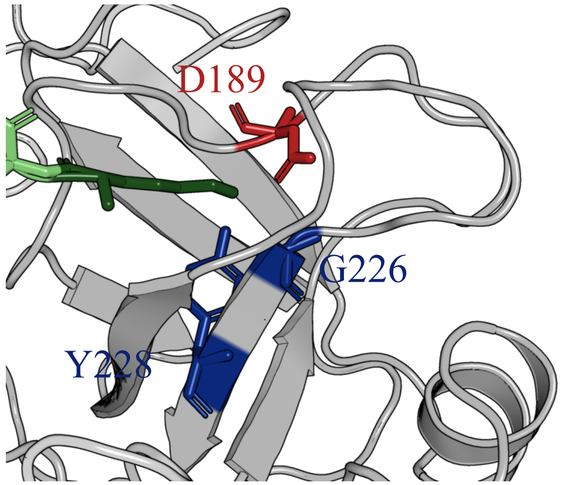


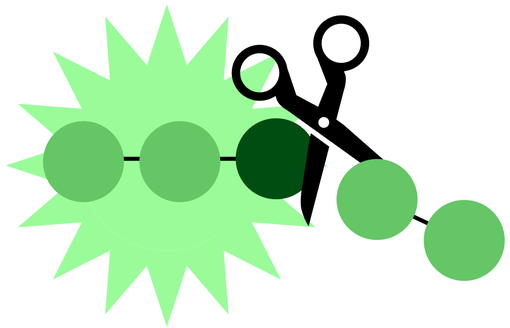
Preliminary experimental results

Chymotrypsin activity



Zoom on the binding pocket



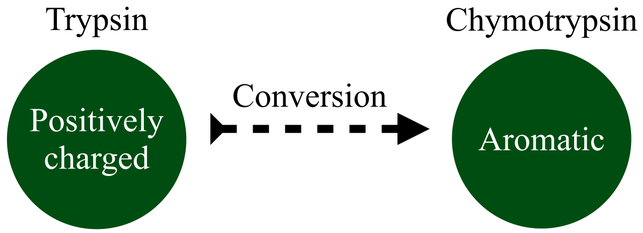


Not possible to make quantitative comparisons here

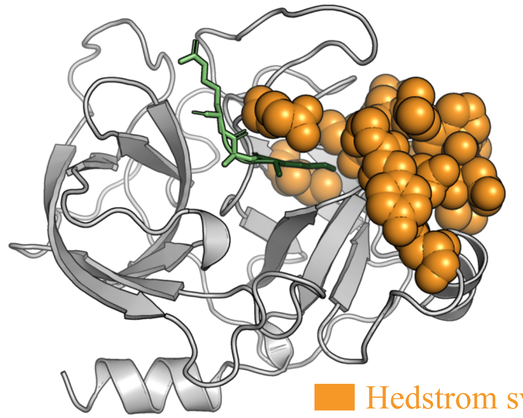
- Evidence of chymotrypsin activity 3 mutations away from rat trypsin
- Mutations located in the binding pocket

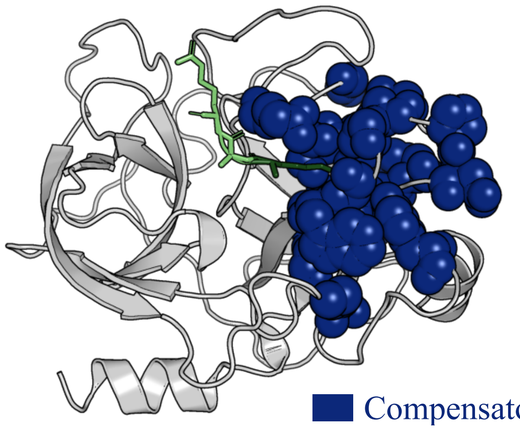
The experiments were performed by Amaury Paveyranne, Timothé Lucas and Shoichi Yip

Boltzmann Machine-guided specificity conversion



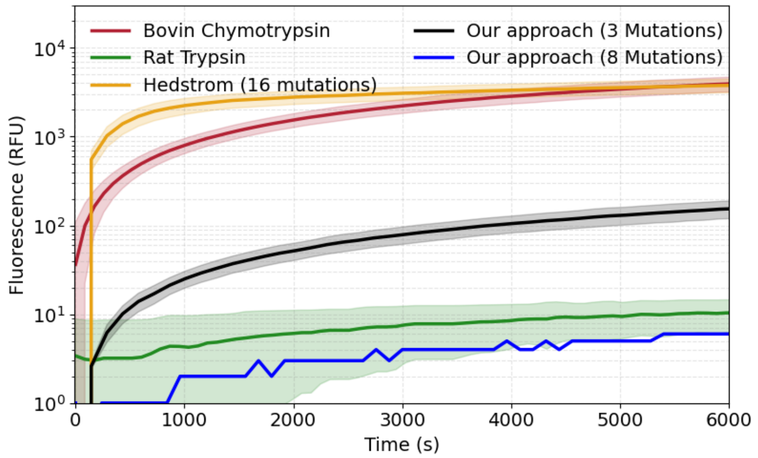
In silico comparison with Hedstrom



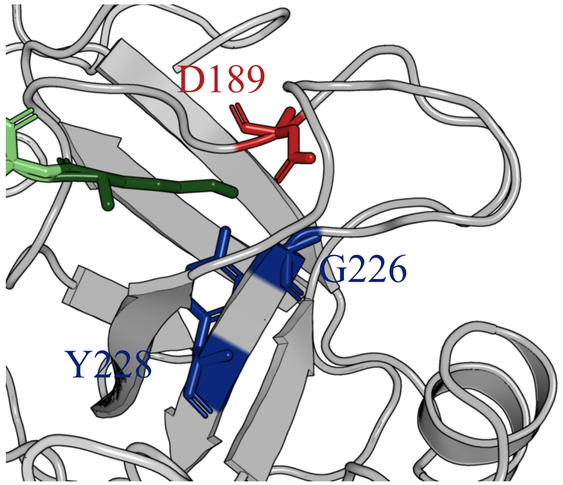


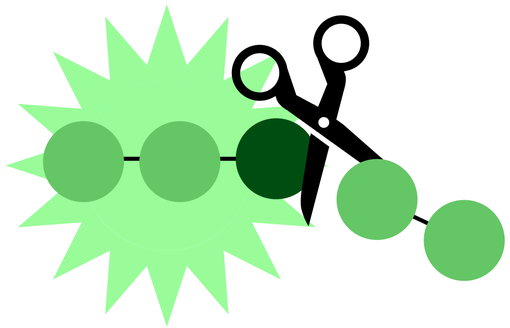
Preliminary experimental results

Chymotrypsin activity



Zoom on the binding pocket





Not possible to make quantitative comparisons here

- Evidence of chymotrypsin activity 3 mutations away from rat trypsin
- Mutations located in the binding pocket

Other results & next steps:

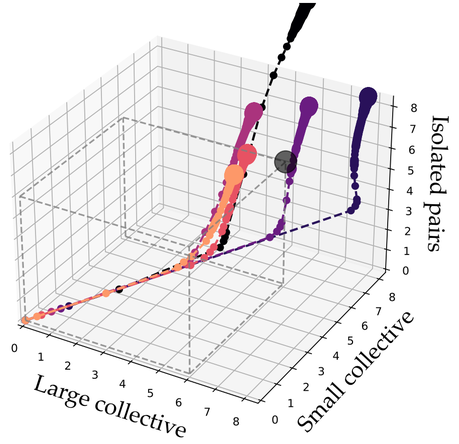
- More precise characterization of functional mutants
- Other conversions

The experiments were performed by Amaury Paveyranne, Timothé Lucas and Shoichi Yip

The undersampling problem

How to infer rich statistical structure from limited data?

Stochastic Boltzmann Machine



Toy model: correct undersampling-induced biases

Real data: model generative without modification

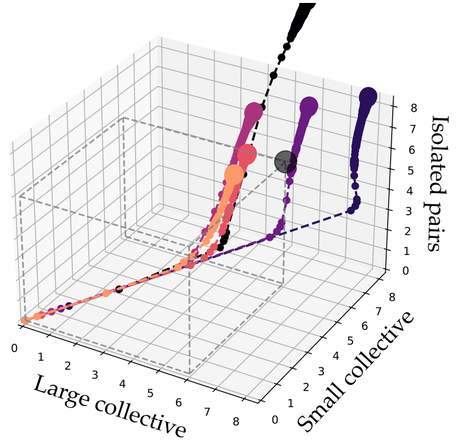
Theoretical understanding?

In collaboration with Emily Hinds, Yaakov Kleeorin, Rama Ranganathan (University of Chicago, USA)

The undersampling problem

How to infer rich statistical structure from limited data?

Stochastic Boltzmann Machine



Toy model: correct undersampling-induced biases

Real data: model generative without modification

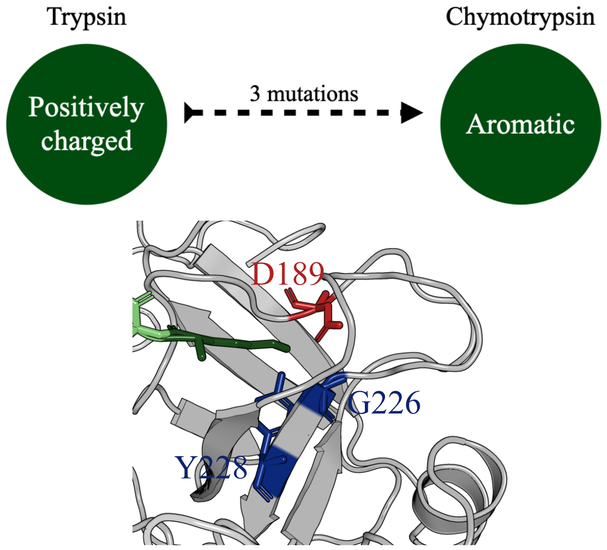
Theoretical understanding?

In collaboration with Emily Hinds, Yaakov Kleeorin, Rama Ranganathan (University of Chicago, USA)

Investigate protein properties with statistical learning

Specificity conversion in S1A family?

Predict compensatory mutations with Boltzmann Machine



Evidence of chymotrypsin activity
3 mutations away from rat trypsin

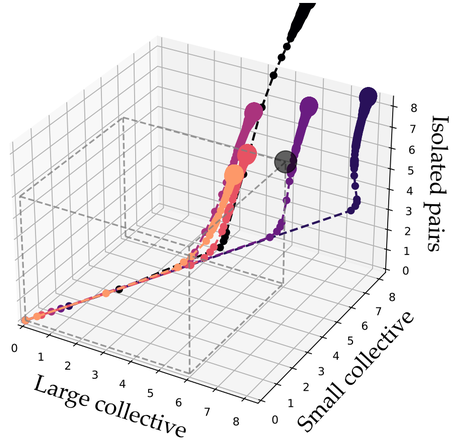
Better experimental characterization

In collaboration with Amaury Paveyranne, Timothé Lucas, Shoichi Yip, Clément Nizak (LJP, Sorbonne University, France)

The undersampling problem

How to infer rich statistical structure from limited data?

Stochastic Boltzmann Machine



Toy model: correct undersampling-induced biases

Real data: model generative without modification

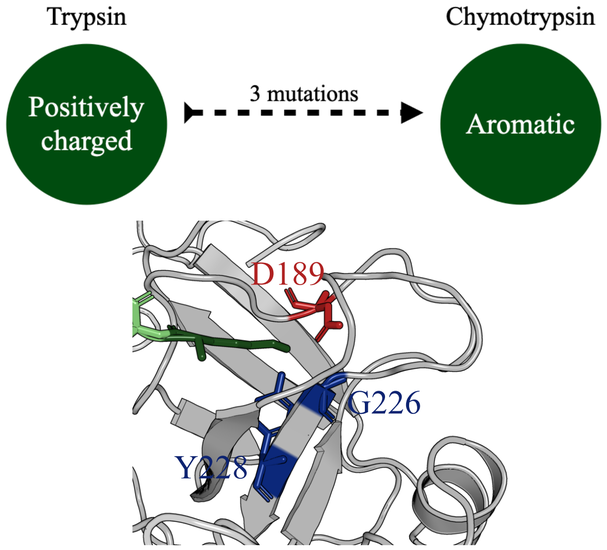
Theoretical understanding?

In collaboration with Emily Hinds, Yaakov Kleeorin, Rama Ranganathan (University of Chicago, USA)

Investigate protein properties with statistical learning

Specificity conversion in S1A family?

Predict compensatory mutations with Boltzmann Machine

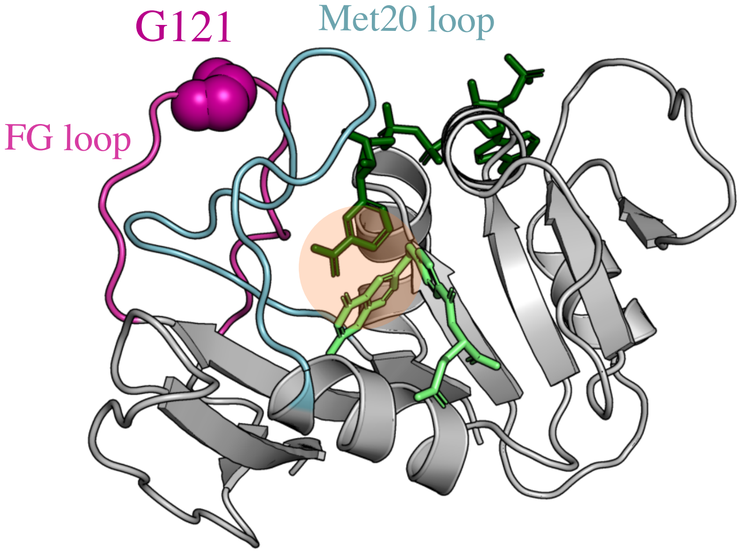


Evidence of chymotrypsin activity 3 mutations away from rat trypsin

Better experimental characterization

In collaboration with Amaury Paveyranne, Timothé Lucas, Shoichi Yip, Clément Nizak (LJP, Sorbonne University, France)

Allosteric network of *E. Coli* DHFR



Support for an allosteric mechanism obtained through molecular dynamics simulations

In collaboration with Paul Guenon, Damien Laage, Guillaume Stirnemann (ENS, France), Clément Nizak (LJP, France), Karolina Filipowska, Kim Reynolds (University of Texas, USA)

Merci

Collaborators

Rama Ranganathan
Emily Hinds
Nadav Benhamou Goldfajn
Miranda Moore
Yaakov Kleeorin
Amaury Paveyranne
Timothé Lucas
Shoichi Yip
Clément Nizak
Paul Guenon
Damien Laage
Guillaume Stirnemann
Karolina Filipowska
Kim Reynolds

Nantes & co

Ma famille, mes amis
Les p’tits lapins
Les colocs
Snoopy, Sido, Pinpin

Directeurs de thèse

Ivan Junier
Olivier Rivoire

Jury

Anne-Florence Bitbol
Cyril Furtlehner
Thierry Mora
Marco Ribezzi

Paris

Le laboratoire Gulliver
Le bureau 414 du LJP

Grenoble

Le laboratoire TIMC
L'équipe TrEE

Chicago

Ranganathan Lab’