ENERGY-BASED MODELS FOR ATOMIC-RESOLUTION PROTEIN CONFORMATIONS

Bowen Jing , Stephan Eismann , Patricia Suriana, Raphael J.L. Townshend, Ron O. Dror

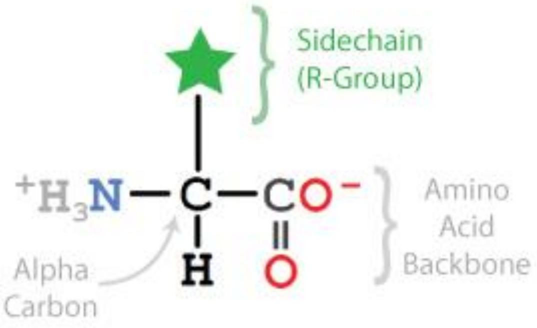
Stanford University

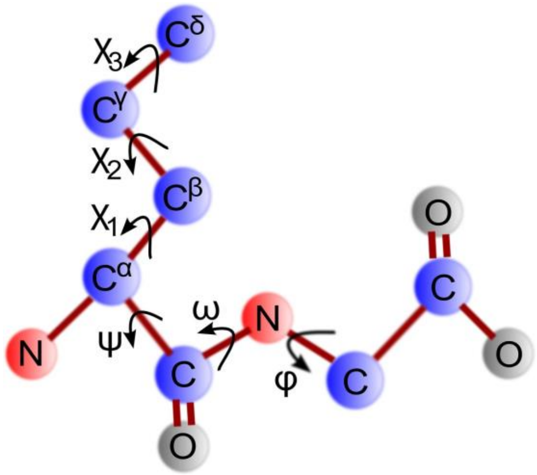
Presented by: Md Hossain Shuvo

Virginia Tech

Background

Protein conformation





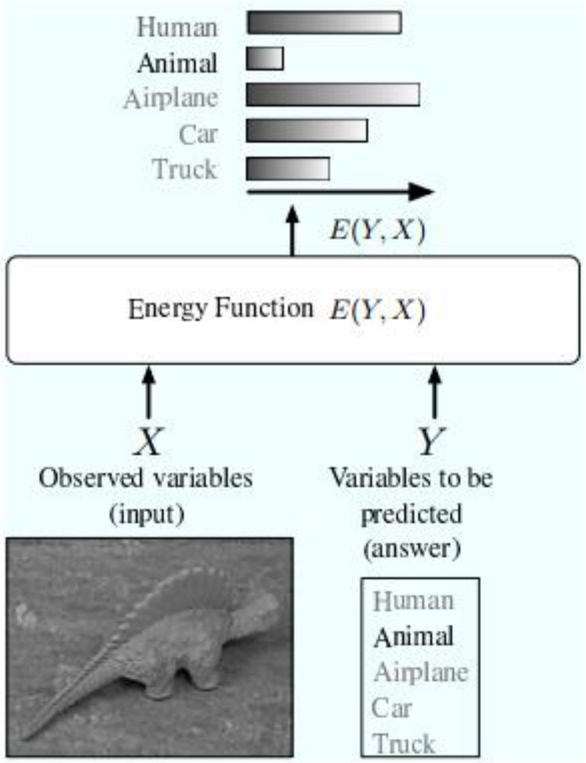
Energy-based models (EBMs)

$$p_\theta(x) = \exp(-E_\theta(x))/Z(\theta),$$

where $Z = \int \exp(-E_\theta(x))\,dx$

$$\begin{aligned} L_{\mathrm{ML}}(\theta) &= \mathbb{E}_{x\sim p_D}[\log p_\theta(x)] \\ &= \mathbb{E}_{x\sim p_D}[E_\theta(x) - \log Z(\theta)] \end{aligned}$$

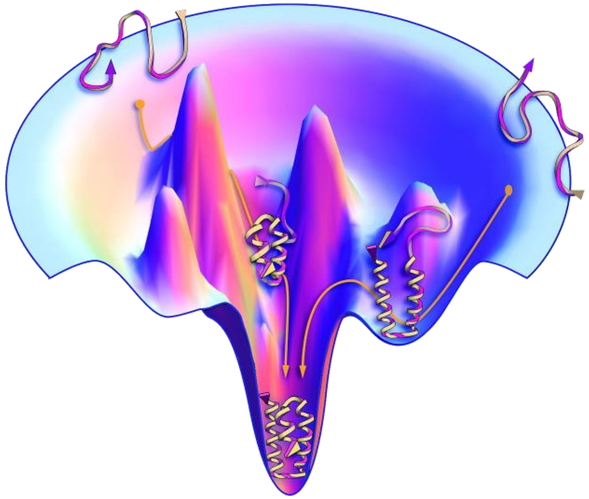
$$\begin{aligned} \nabla_\theta L_{\mathrm{ML}} \approx\ & \mathbb{E}_{x^+\sim p_D}[\nabla_\theta E_\theta(x^+)] \\ & - \mathbb{E}_{x^-\sim p_\theta}[\nabla_\theta E_\theta(x^-)] \end{aligned}$$



https://cbm.msoe.edu/
https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-S14-S10/figures/1
http://yann.lecun.com/exdb/publis/pdf/lecun-06.pdf

Motivation

Protein folding



Learning energy function

- Force field
- Statistical potentials
- Rosetta

Learn the energy function directly from data using generative modeling, EBMs

https://fold.it/portal/node/2005623

Problem definition

Rotamer recovery

Given:

set of surrounding atoms ,k (Context atoms) for a residue

k = 64

Train:

$Y^* = \text{argmin}_{Y \in \mathcal{Y}} E(Y, X).$

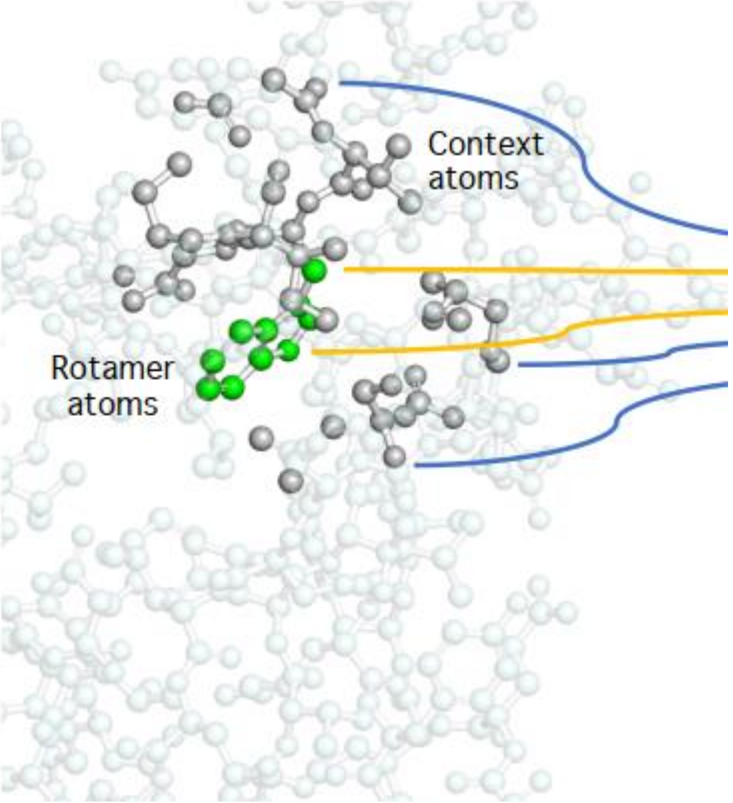
Sample from rotamer library

Energy function: $E_\theta(x, c) = f_\theta(A(x, c))$

Loss function : $\mathcal{L}(\theta) = -E_\theta(x, c) - \log Z_\theta(c)$

Predict:

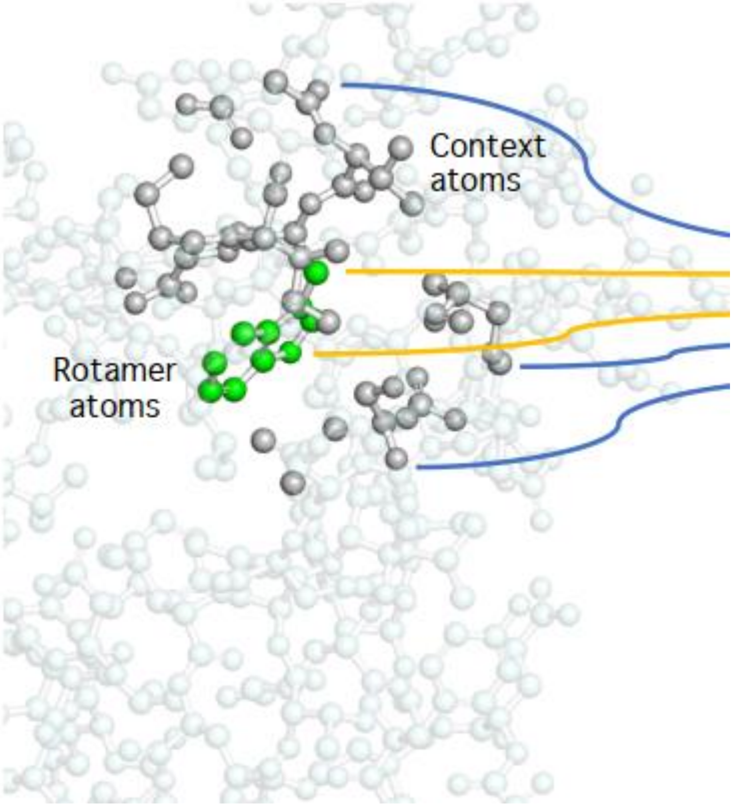
Rotamer atoms



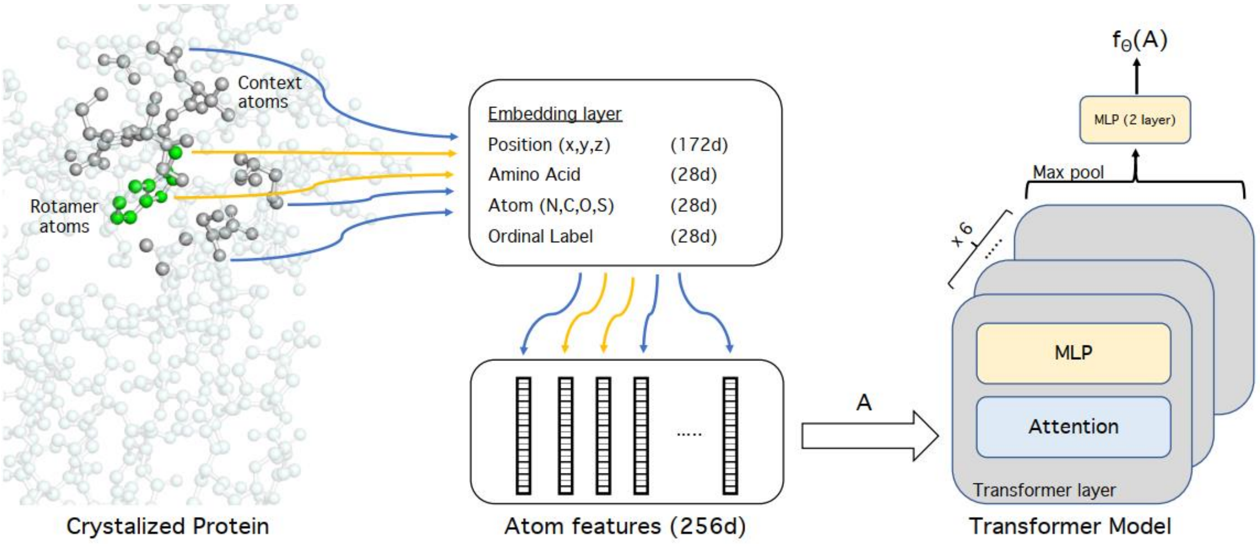
Problem setup

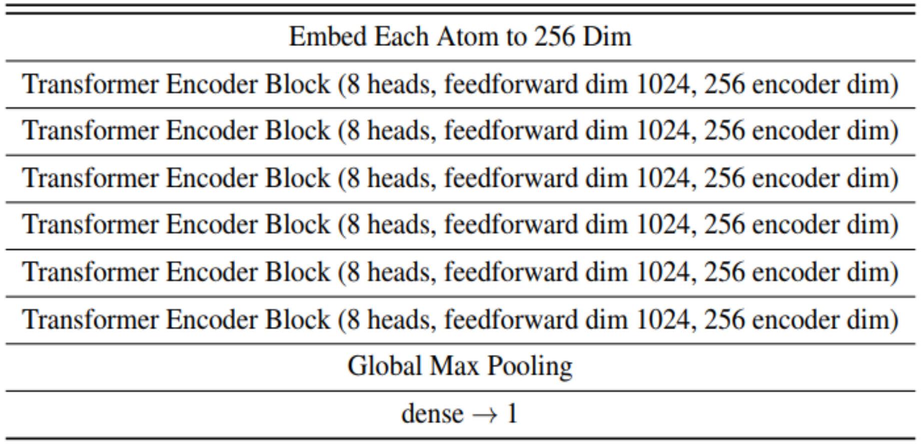
Atom input (context atoms) representations

- Cartesian coordinates (x,y,z)
- Categorical features: N/C/O/S
- Ordinal label: type of N/C/O/S
- Type of the amino acid



Architecture





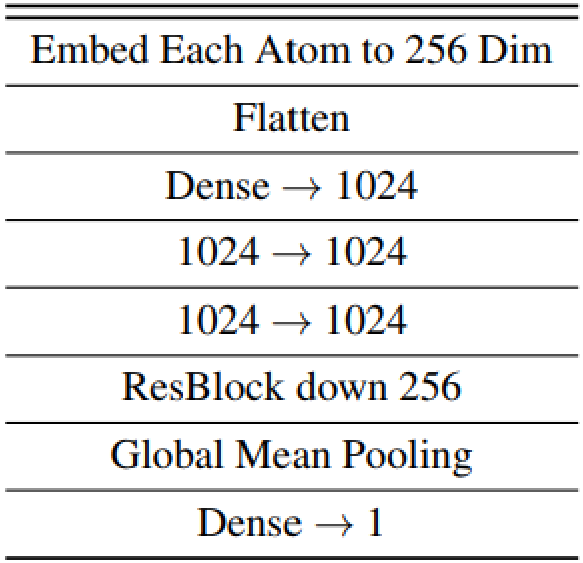
Embed Each Atom to 256 Dim
Transformer Encoder Block (8 heads, feedforward dim 1024, 256 encoder dim)
Transformer Encoder Block (8 heads, feedforward dim 1024, 256 encoder dim)
Transformer Encoder Block (8 heads, feedforward dim 1024, 256 encoder dim)
Transformer Encoder Block (8 heads, feedforward dim 1024, 256 encoder dim)
Transformer Encoder Block (8 heads, feedforward dim 1024, 256 encoder dim)
Transformer Encoder Block (8 heads, feedforward dim 1024, 256 encoder dim)
Global Max Pooling
dense → 1

Figure A3: Atom Transformer Model (6 Transformer Encoder Blocks)

Additional parameters:

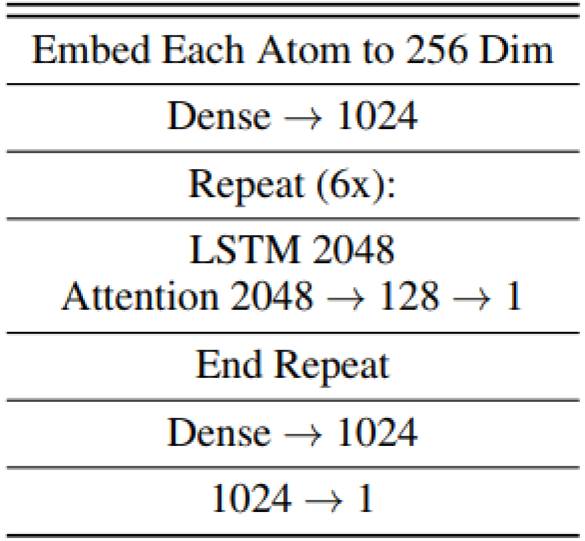
- No dropout used during the training
- Uses Layer normalization

Baseline models



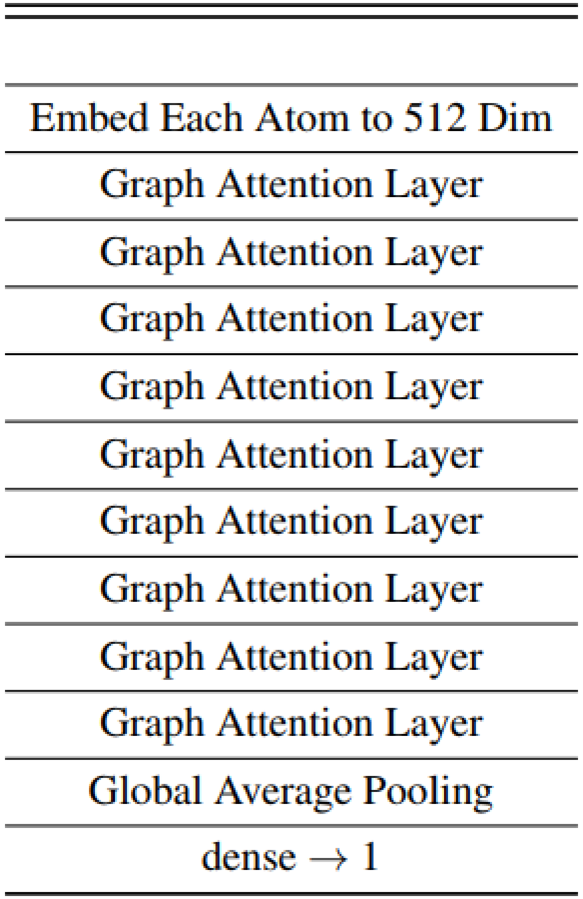
Embed Each Atom to 256 Dim
Flatten
Dense → 1024
1024 → 1024
1024 → 1024
ResBlock down 256
Global Mean Pooling
Dense → 1

(a) Fully Connected Model



Embed Each Atom to 256 Dim
Dense → 1024
Repeat (6x):
LSTM 2048 Attention 2048 → 128 → 1
End Repeat
Dense → 1024
1024 → 1

(b) Set2Set Model (6 Permutation Invariant Blocks)



Embed Each Atom to 512 Dim
Graph Attention Layer
Graph Attention Layer
Graph Attention Layer
Graph Attention Layer
Graph Attention Layer
Graph Attention Layer
Graph Attention Layer
Graph Attention Layer
Graph Attention Layer
Global Average Pooling
dense → 1

Graph network

Datasets

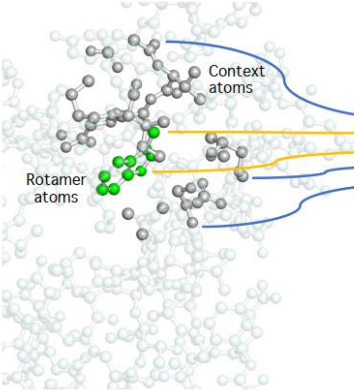
High-resolution PDB structures from CullPDB database

- Resolution finer than 1.8Å
- Sequence identity < 90%
- R-value < 0.25
- Total train proteins: 12,473
- Total train proteins: 129
- Sequence identity <= 25%

Training steps

Algorithm 1 Training Procedure for the EBM

```
Input: Rotamer library $q(x|c)$, Training set of proteins $D$
for Protein $d_i$ of $D$ do
    ▷ Sample random amino acid from $d_i$
    $R \sim d_i$
    ▷ Set positive sample to 64 nearest neighbor atoms of carbon beta of R
    $c^+ \leftarrow \mathrm{NN}_{64}(R)$
    ▷ Generate N negative samples from the rotamer library
    $c^- \leftarrow q(x|c^+)$
    ▷ Compute loss of model (logsumexp across all negative samples)
    $L_{ml} = E(c^+;\theta) + \mathrm{logsumexp}(-E(c^+;\theta), -E(c_0^-;\theta), -E(c_1^-;\theta), \ldots, -E(c_N^-;\theta))$
    ▷ Minimization step of $L_{ml}$ using Adam optimizer
    $\theta \leftarrow \theta - \nabla_\theta L_{ml}$
end for
```



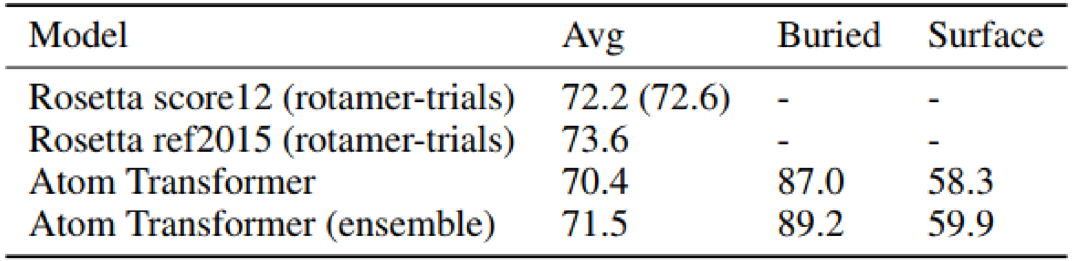
Evaluation metric

- Percentage of rotamer recovery
- Successful rotamer recovery:

  sampled_chi – true_chi < 20Å
- Sampling strategies:
  - Discrete sampling
  - Continuous sampling

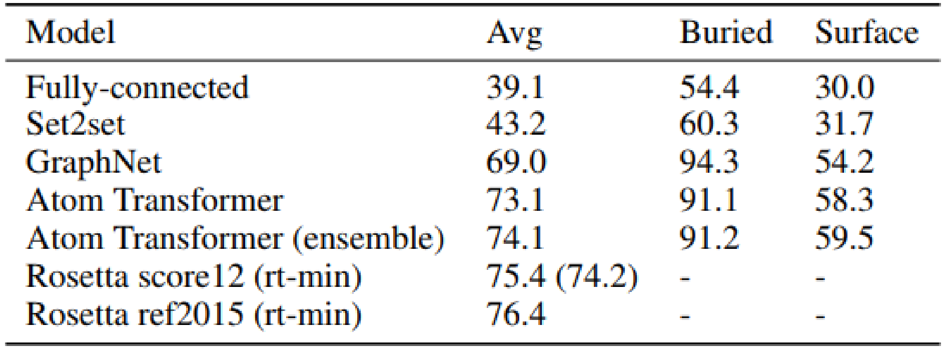
Benchmark

Discrete sampling



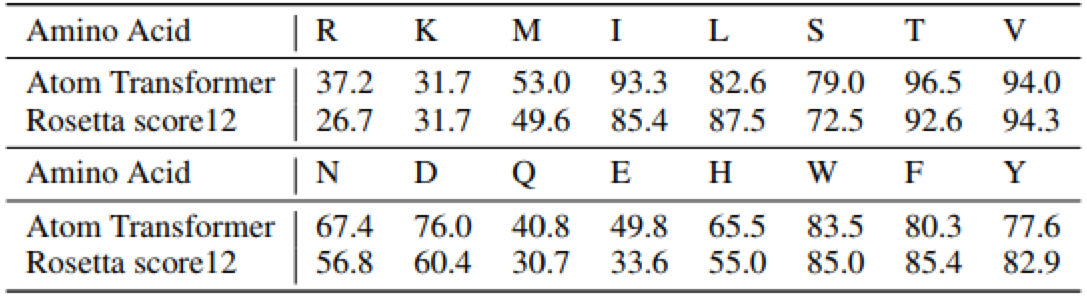
Model	Avg	Buried	Surface
Rosetta score12 (rotamer-trials)	72.2 (72.6)	-	-
Rosetta ref2015 (rotamer-trials)	73.6	-	-
Atom Transformer	70.4	87.0	58.3
Atom Transformer (ensemble)	71.5	89.2	59.9

Continuous sampling



Model	Avg	Buried	Surface
Fully-connected	39.1	54.4	30.0
Set2set	43.2	60.3	31.7
GraphNet	69.0	94.3	54.2
Atom Transformer	73.1	91.1	58.3
Atom Transformer (ensemble)	74.1	91.2	59.5
Rosetta score12 (rt-min)	75.4 (74.2)	-	-
Rosetta ref2015 (rt-min)	76.4	-	-

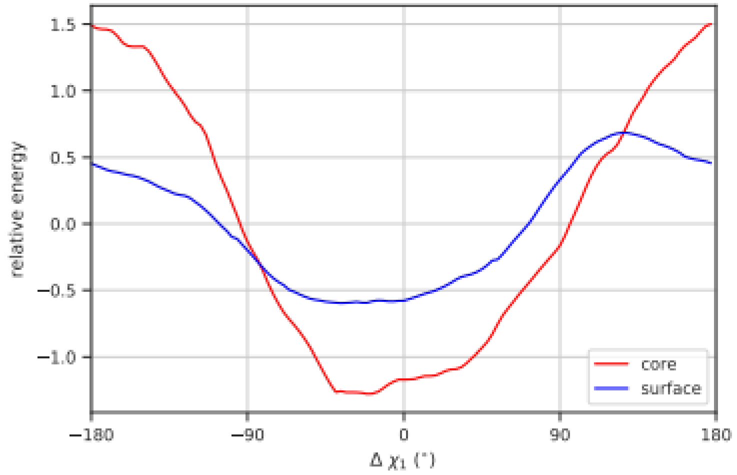
Rotamer recovery by amino acid



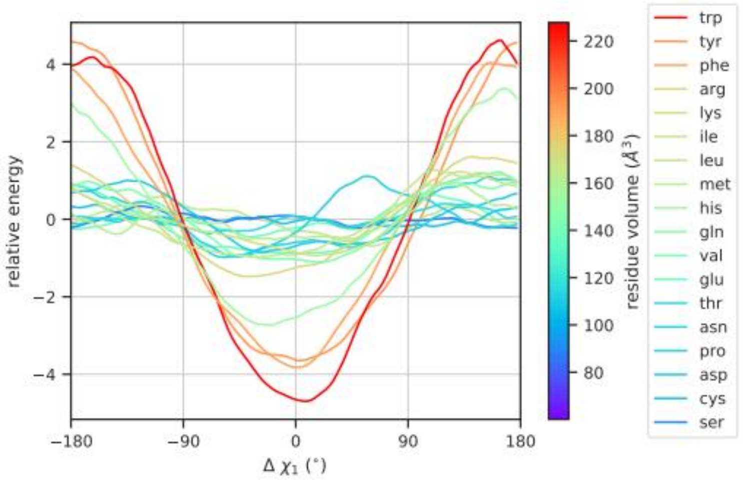
Amino Acid	R	K	M	I	L	S	T	V
Atom Transformer	37.2	31.7	53.0	93.3	82.6	79.0	96.5	94.0
Rosetta score12	26.7	31.7	49.6	85.4	87.5	72.5	92.6	94.3
Amino Acid	N	D	Q	E	H	W	F	Y
Atom Transformer	67.4	76.0	40.8	49.8	65.5	83.5	80.3	77.6
Rosetta score12	56.8	60.4	30.7	33.6	55.0	85.0	85.4	82.9

Energy visualization

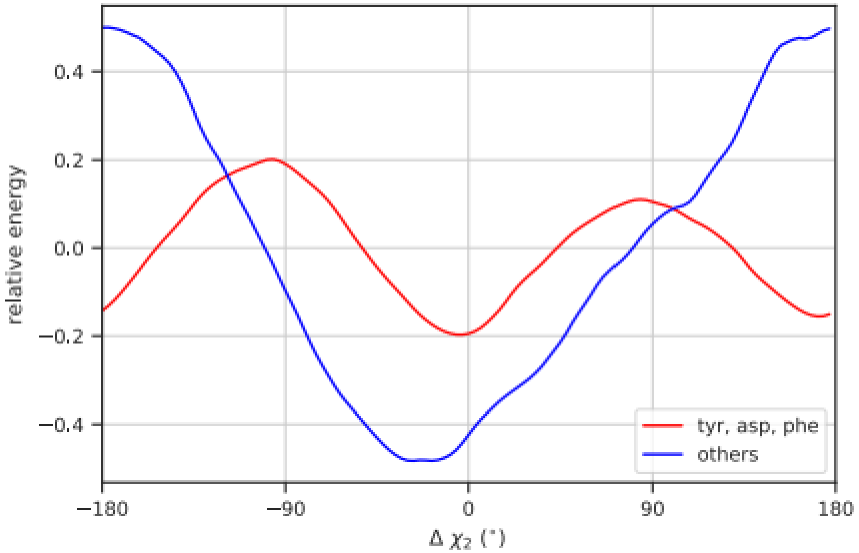
Core vs Surface residue energies

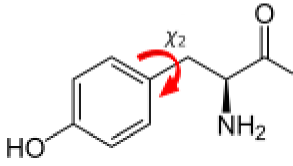


Residue size vs energy well

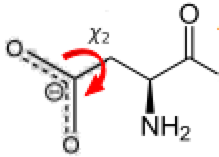


Symmetries of amino acids

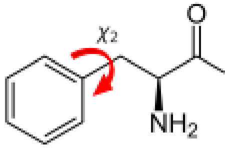




Tyr



Asp



Phe

Summary and observations

- Learns energy function directly from the data using EBMs
- Discovers relevant features automatically
- Performance
- Evaluating methods