References:
KORP: knowledge-based 6D potential for fast protein and loop modeling
José Ramón López-Blanco, Pablo Chacón

Improving the orientation-dependent statistical potential using a reference state.
Liu Y1, Zeng J, Gong H.

To read
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094334
Aplications/datsets
https://www.ncbi.nlm.nih.gov/pubmed/30414728
https://www.ncbi.nlm.nih.gov/pubmed/30036062
http://chaconlab.org/

Maria
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2978081/
https://www.ncbi.nlm.nih.gov/pubmed/28921375

Phone
Proteins. 2014 Oct;82(10):2383-93. doi: 10.1002/prot.24600. Epub 2014 Jun 3.
Improving the orientation-dependent statistical potential using a reference state.
Liu Y1, Zeng J, Gong H.

Potentials of Mean Force for Protein Structure Prediction Vindicated, Formalized and Generalized by Thomas Hamelryck ,
PLOS Published: November 10, 2010 https://doi.org/10.1371/journal.pone.0013714

Probabilistic metric space to model orientation dependent states
(for fast protein and loop modelling)

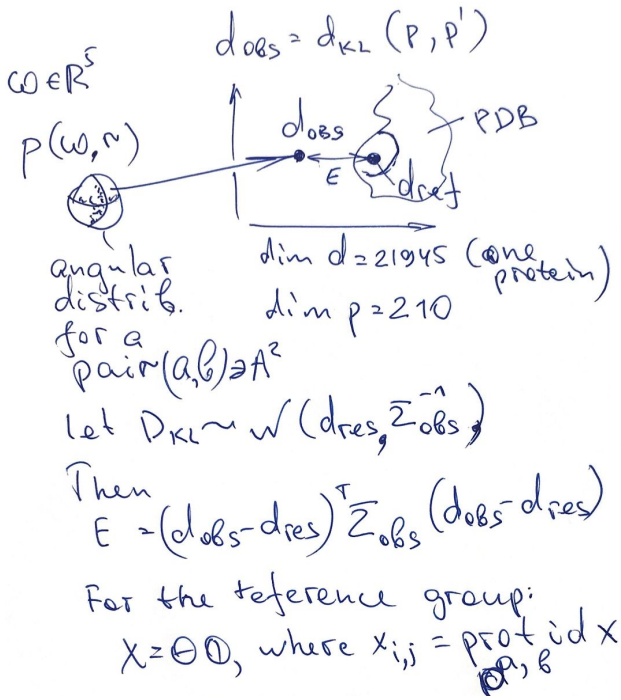
The project goal is to rank the set of reference proteins, generated from the same set of amino-acids, according to the statistical similarity of the observed protein structure to the structures of reference proteins from the PDB.

Briefly, one has to assess does an observed protein looks like a typical reference protein or not, according to some criterion, which are not constructed yet. We have to construct these criterions and to propose the assessment method.

The basic element of the methodology is the probability distribution of the mutual orientations between

1) two amino-acids
2) an amino-acid a small molecule.

Here the first case follows.



Denote by $\mathcal A^2\ni(a,b)$ the alphabet all unsorted pairs of amino-acids, 20 amino-acids make 210 pairs. Denote by $p^{a,b}(\omega,r)$ the (joint, conditional, marginal: see Problems) probability distribution of the angular orientation ω given the distance r. Denote by $d = D_KL(p,p^\prime) the distance between two distributions, 210 pairs of amino-acids make 21945 distances. Let $d\sim\mathcal{N}(d_\text{ref}, \Sigma_\text{obs}). So this distance vector produce the metric space to rank proteins according to the quadratic form

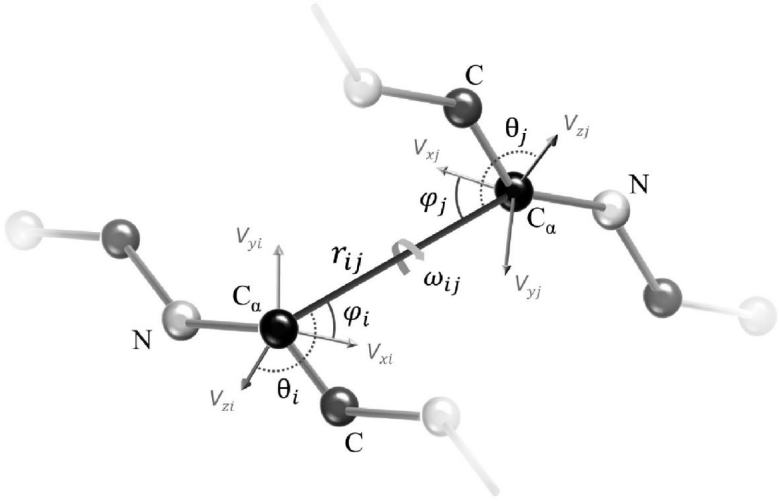
$$
E = (\mathbf{d}_\text{obs} - \mathbf{d}_\text{ref})^\mathsf{T}) \Sigma_\text{ref}^{-1} ().
$$

how far an observed protein from the nearest stable reference subset from the PDB. The reference subset is extracted using the collaborative filtering (topic modeling). The cartesian product is formed by the sets {A^2} and proteins, described by the vector p.

Urgent to check:
The distribution $p(\omega,r)$ must tend to uniform when r tends to $\inf$. In fact after $r>r^*$.

Problems to be resolved:

1) the distance r is not a random variable
2) analysis of the energy distribution function



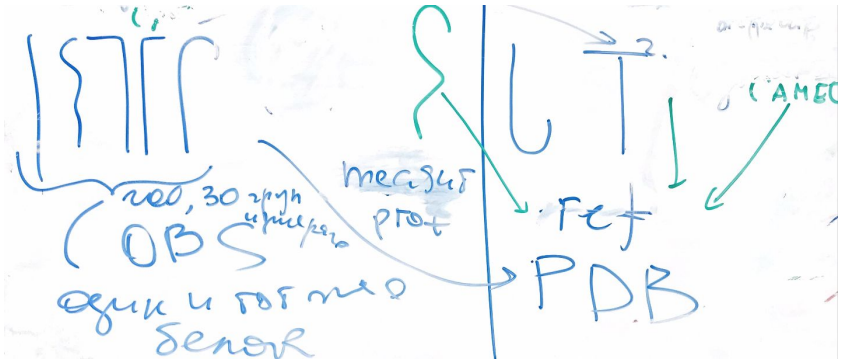
Analysis:
This problem statement answers to the following questions

1. How to avoid unnecessary aggregation making the protein description?
2. How to include pairs of couples in the model?
3. How to ensure instability of the reference protein description?

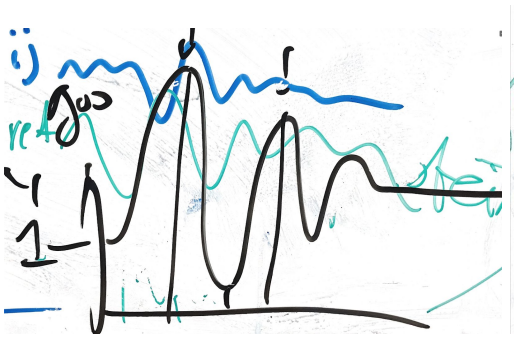
Both approaches consider the great variability of proteins.

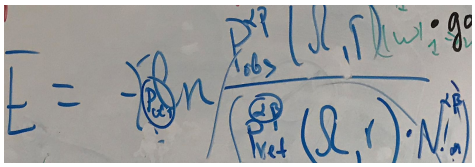
Appendix figures:

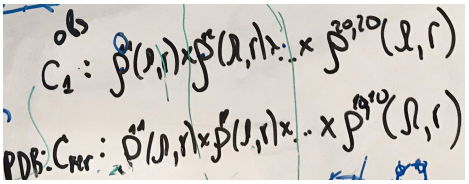
Motivation for the PDB and CASP challenge



Motivation for the Energy approach



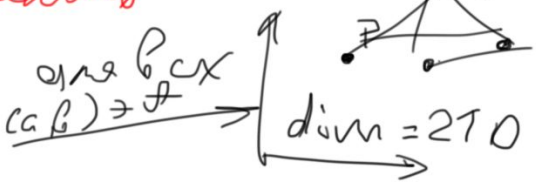




$A^2 = \{A, \dots Z\} \Rightarrow (a,b)$ #210

probabilist metric space

$p(\omega, r)$



$\bar{E} = p(d_{obs}, d_{ref})$

t-divergence

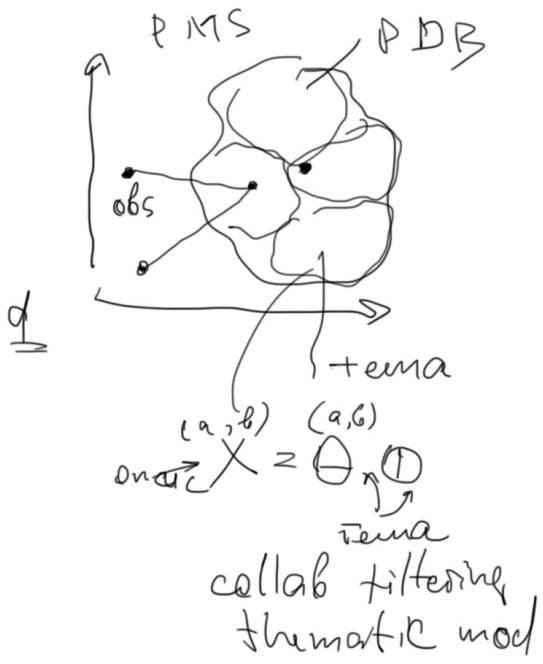
DKL 21945

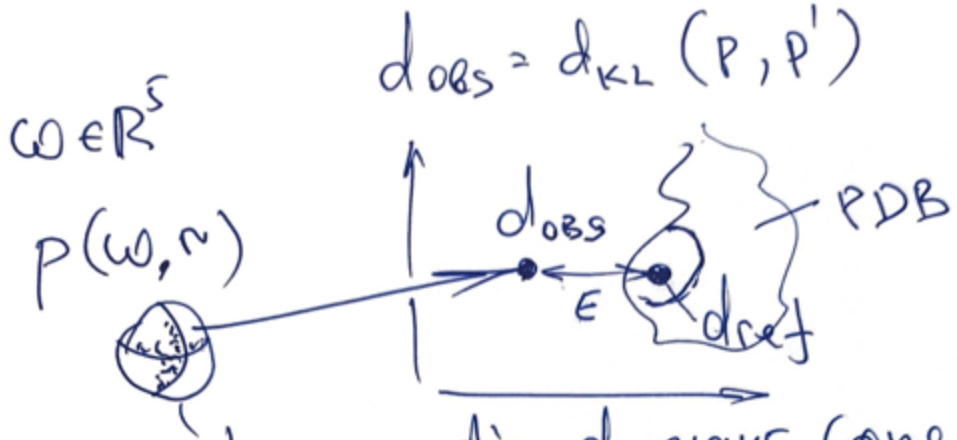
$d(P, P')$ dim

$\bar{E} = (d_{obs} - d_{ref})^T \sum_{obs} (-$

Reference (a,b)

сов $d_{obs} = \sum_{obs}$


PMS
PDB
obs
+ema
(a,b)
(a,6)
ema
collab filtering
thematic mod



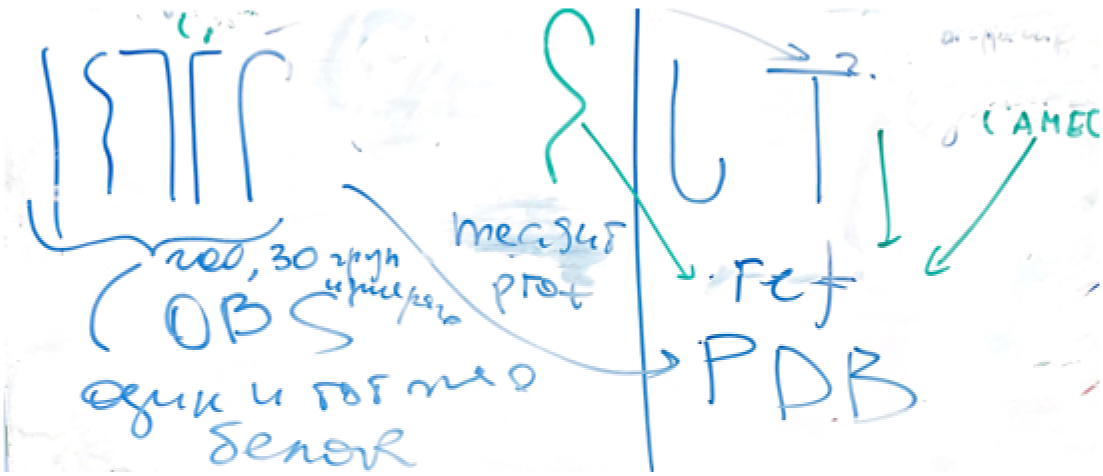
let $D_{KL} \sim \mathcal{N}(d_{res}, \bar{\Sigma}^{-1}_{obs})$

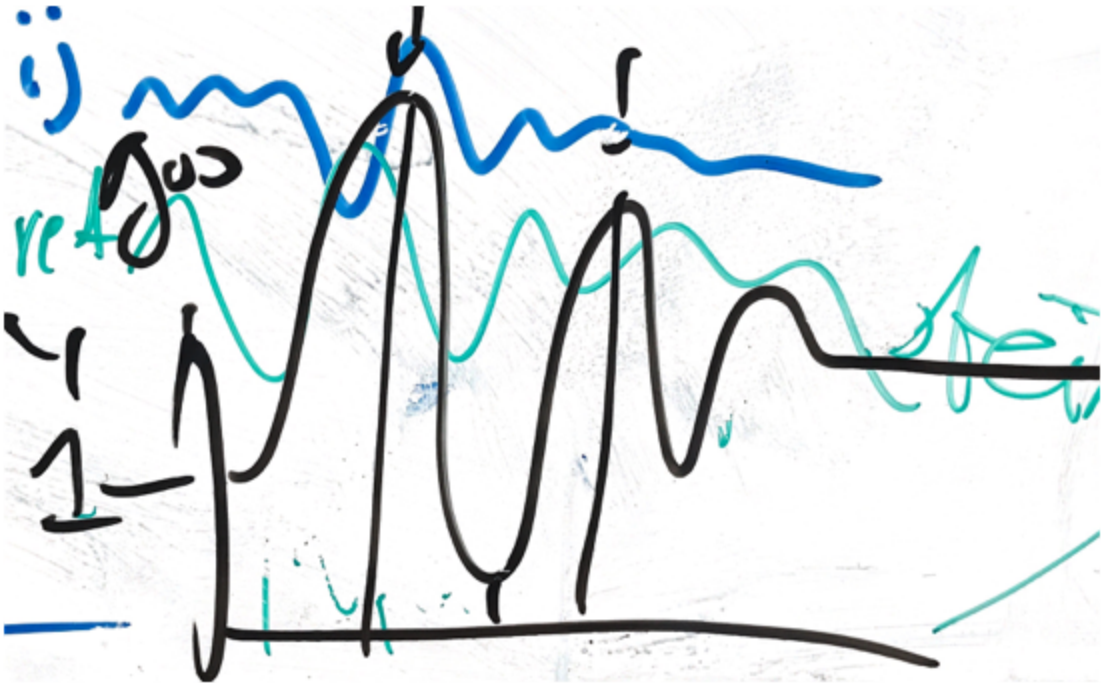
Then

$$E = (d_{obs} - d_{res})^T \bar{\Sigma}_{obs} (d_{obs} - d_{res})$$

For the reference group:

$X = \Theta D$, where $x_{i,j} = \text{prot id } x$


OBS
ref
PDB
CAMEO




geom

Structure Factors

$F(\rho(r)) = g(\vec{q})$

Tests

Convex PL

Graph-

based

Docking