A machine learning based approach to sidechain optimization

Sabareesh Subramaniam
Sriram Natarajan
Alessandro Senes
Sidechain Optimization

Backbone

Add sidechains to achieve minimum energy configuration
Conformer Library
Computational Complexity - Combinatorial

# possible conformations = 10 * 3 * 10 * 3 * 10 * 9 = 81000
Varying sampling requirements

Exposed sidechain
Many isoenergetic
Conformations - Easy

Buried sidechain
Fewer isoenergetic
Conformations - Hard
Reduced Search Space and/or Better Energies

# possible conformations = 5 * 3 * 15 * 3 * 2 * 9 = 12150

By moving sampling from the easy positions to the hard ones, we could be more efficient (fast) and/or achieve better energies

But, can we predict if a position is easy or hard?
Yes, use machine learning

Associate each sidechain in the database with a feature vector $X$ and a label $Y$

$X = \{\varphi, \psi, SS, SE, PD, N, \ldots\}$
$Y = \{\text{Hard, Medium, Easy}\}$
Achieve better energies

The biased sampling scheme using machine learning achieves better energies on majority of the proteins
Thank You !!