

BOKEI: Bayesian Optimization using Knowledge of correlated torsions and Expected Improvement for conformer generation

Lucian Chan¹, Geoffrey R. Hutchison² and Garrett M. Morris^{1,*}

¹Oxford Protein Informatics Group, Department of Statistics, University of Oxford, 24-29 St Giles', Oxford, OX1 3LB, UK;

²Department of Chemistry and Chemical Engineering, University of Pittsburgh, 219 Parkman Avenue, Pittsburgh, PA, 15260, USA

1. Introduction

We have previously shown that our Bayesian Optimization Algorithm (BOA) is more efficient in finding the lowest-energy conformer than exhaustive systematic search and uniform random search.¹

Many conformer sampling tools assume independence of rotatable bonds and nearest neighbour effects are generally ignored, including BOA.

However, adjacent torsion angles are often naturally constrained, usually by steric clashes, but sometimes by favorable intramolecular interactions.

2. Method

Bayesian Optimization Algorithm (BOA)

Input: Initial observations, D .

Step 1: Select next point, x , by optimizing acquisition function over the Gaussian Process (GP).

Step 2: Evaluate the objective function at x .

Step 3: Augment the data, D , and update the GP.

Step 4: Repeat from *Step 1*.

Correlated Torsions

Bivariate von Mises distributions are used to jointly model the adjacent torsion angles.

These torsion distributions are used to constrain the search.

Acquisition Function

Knowledge-based Expected Improvement²:

where M is the number of correlated torsions found in a molecule, and \mathcal{M} is the mixture model of the torsion angle pairs in pattern, m .

Benchmarking

533 diverse small molecules selected from the Platinum dataset³ and the dataset assembled by Ebejer *et al*⁴. was used to benchmark the performance of the three search algorithms: (i) BOA with KEI (BOKEI), (ii) BOA with standard EI (BOA-EI) and (iii) Genetic Algorithm in Open Babel (GA). A fixed number of MMFF94 energy evaluations was used for all search methods.

3. Results

BOKEI consistently finds lower energy conformations sooner than BOA. The energy gap between BOKEI and BOA-EI diminishes as the number of iterations increases, as they converge to same global minimum.

In our tests, BOKEI consistently finds lower energy conformations than BOA-EI for all sets of rotatable bonds, and it outperforms GA for molecules up to ten rotatable bonds.

Our analysis of structures in COD revealed higher-order correlated torsions (see Fig 3). The torsion angles of the substructure (see SMARTS pattern) are restricted, as the C=O...H-N forms intramolecular hydrogen bonds, giving a pseudo six-membered ring.

4. Conclusions

BOKEI finds lower energy conformations in fewer iterations than other global optimization methods, including BOA-EI and GA, especially given a small number of energy evaluations.

Information about correlated torsions in small molecules helps constrain the search.

Further understanding the influence of intramolecular interactions will help improve the efficiency in sampling low-energy conformers.