Overview

The objective of this work is to [R] create a framework for future advancements for protein design via K* optimization (ie. K*MAP) leveraging a depth-first branch-and-bound algorithm over AND/OR search spaces, and [R] form a foundation for efficient algorithms solving such protein design problems.

Contributions:
1. Formulation of K*MAP as a graphical model
2. wMBE-K*, a weighted mini-bucket scheme for K*MAP enhanced with a domain partitioning scheme
3. AOBB-K*, a powerful, general MMAP algorithms, [R] explore new related branched heuristics, and [R] create a foundation for efficient algorithms solving such protein design problems.

Problem

Redesign of proteins to form higher affinity complexes

K* Objective
An approximation of binding affinity between molecules (based on the biological association constant known as KJ)

\[ K'(r) = \frac{Z_r(f)}{Z_r} \]

\[ Z_r = \sum_{c} \exp(-E_c(c)/RT) \]

\[ r = \text{amino acid assignments to residues} \]
\[ c = \text{possible conformations given} r \]
\[ E = \text{energy given configuration} \]
\[ R = \text{universal gas constant} \]
\[ T = \text{absolute temperature (Kelvin)} \]

Graphical Model Formulation

\[ M = < X, D, F > \]

R’s capture amino acid assignment for residue

C’s capture all (amino acid, rotation) combinations possible at its corresponding R

Constraints between corresponding C’s and R’s ensure consistent assignments

Objective

Find amino acid assignments to the residues that maximize K*

\[ Z_r(R_1...R_n) = \sum_{C_1...C_n} \prod_{i=1}^{n} \mathcal{C}_i(R_i, C_i) \]

\[ = \sum_{C_1...C_n} \prod_{i=1}^{n} \exp(-E_i(c_i)/RT) \]

\[ K'(R_1,...,R_n) = Z_r(R_1,...,R_n) \]

\[ \text{task: K*MAP} = \max_{R_1...R_n} K'(R_1,...,R_n) \]

wMBE-K*

Precompiled K* heuristic to guide search, based on dynamic programming message-passing Mini Bucket Elimination

AND/OR Search

Graphical Model Network Possible Pseudo Tree

Directed tree (based on a variable ordering) that branches when conditional independences exist given assignments to ancestors. The pseudo tree is used to construct the AND/OR search space

Domain-Partitioned wMBE-K*

Strategy to improve lower bounds generated by wMBE-K*

- Let X, Y, and Z be three variables
- Let \( \text{obj} = \sum_i f(x,y) \cdot g(x,z) \)
- Let \( X' = \{ x \in X | g(x,z) \neq 0 \} \) s.t. \( c_{xy} = \min_{x \in X} g(x,z) \)
- Since \( c_{xy} > 0 \), we can derive...

\[ \text{obj} = \sum_i f(x,y) \cdot g(x,z) \]

\[ \text{obj} = \sum_i f(x,y) \cdot g(x,z) + \sum_i f(x,y) \cdot g(x,z) \]

\[ \text{obj} = \sum_i f(x,y) \cdot g(x,z) \geq c_{xy} \cdot \sum_i f(x,y) > 0 \]

AOBB-K*

- Branch-and-bound algorithm over AND/OR search spaces
- AOBK* is exact
- Can use wMBE-K* to guide search
- Exploits determinism by using constraint propagation
- Incorporates a constraint enforcing biologically relevant solutions

Subunit Stability Constraints

Condition to enforce the stability of each subunit to be no less than a given threshold \( \tau \) from that of the wild-type stability

\[ Z_{subunit}(r) > \tau \]

\[ Z_{subunit}(r)^{\text{wt}} - \exp(-5\tau) \]

Stability of naturally occurring version

Infusing Determinism: \( \gamma \)-Underflows

Replace extremely unfavorable assignments with hard constraints taking exploiting the strength of constraint propagation

- Let \( x \) be a non-negative function
- Consider \( \gamma \cdot r \)
- Then the \( x \) underflow of \( f(x) \)

\[ f(x) = \begin{cases} f(x) & f(x) \geq \gamma \tau \\ 0 & \text{otherwise} \end{cases} \]

Summary

- AOBK* shows promise vs. state-of-the-art BBK*
- Competitive runtime
- Can find better solutions
- \( \gamma \)-AOBK* can greatly improve runtimes
- AOBK* has scalability issues as the number of MAP variables increase

Acknowledgements

Special Thanks:
- Thomas Schier [INRAE, France]
- Donald Brown and members of his lab [Duke University] including: Graham Holt, Jonathan Jou and Nathan Gurrin
- Supported in part by: NSF Grant IIS-2008516