What is the Protein Side-Chain Positioning Problem?

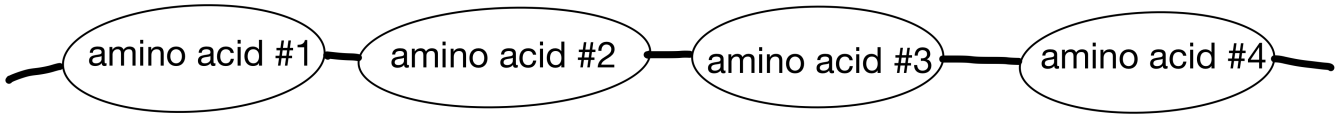
Protein Side-Chain Positioning, (SCP), is one of the most important subproblems of the protein structure prediction problem. SCP has many applications, e.g. to e.g., [Bur15]: ligand binding; and protein-protein docking with backbone flexibility. We now see how to formulate it as a mixed integer program, (MIP).

What is a Protein?

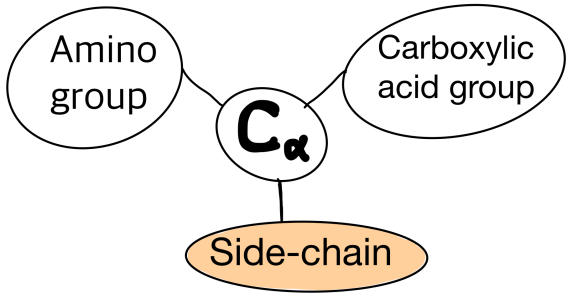
A protein

A collection of chains of amino acids

- **Chain**:
  a sequence of amino acids



- **Amino acid**:
  building blocks of a protein
  consists of $\{C, H, O, N\}$



Protein Side-Chain Positioning Problem (**SCP**)

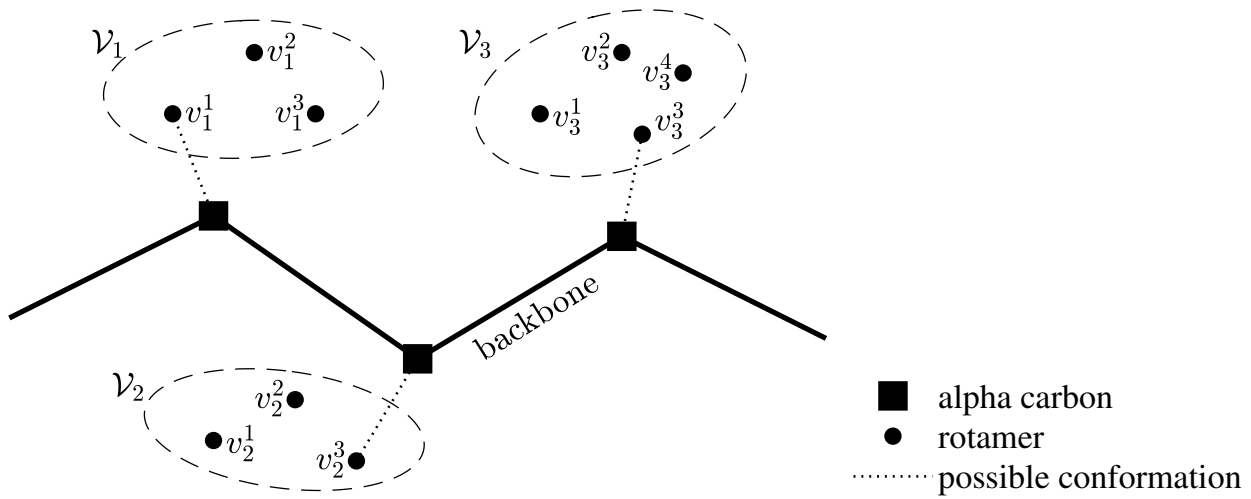
Task

Given

- a collection of sets of candidates for each side-chain of an amino acid
- pair-wise energy values between all candidates

Do

- Pick exactly ONE candidate from each set
- while yielding the minimum energy



Applications & Roadmap

Applications

- One of the most important subproblems of the protein structure prediction problem.
- Examples: ligand binding, protein-protein docking, etc

Roadmap

- Formulate as a quadratic integer program
- Relax the hard problem
- Solve using continuous optimization methods
- Round the approximate solution
- Use branch and bound and cutting planes to improve solution

Problem Formulation as **IQP**

Given a collection of disjoint sets $\mathcal{V}_i, i \in [p]$. $|\mathcal{V}_i| = m_i$,
$n_0 = \sum_{i=1}^{p} m_i$ and $\mathcal{V} = \cup_{i=1}^{p} \mathcal{V}_i$.
$\mathcal{V}_i$: rotamer set, elements in $\mathcal{V}_i$ rotamers.

Goal

1. Select *exactly one* rotamer v_i from each set $\mathcal{V}_i$,
2. Minimize the sum of
   1. the weights (energy) on the edges between chosen rotamers, and
   2. the energy between each chosen rotamer and the backbone

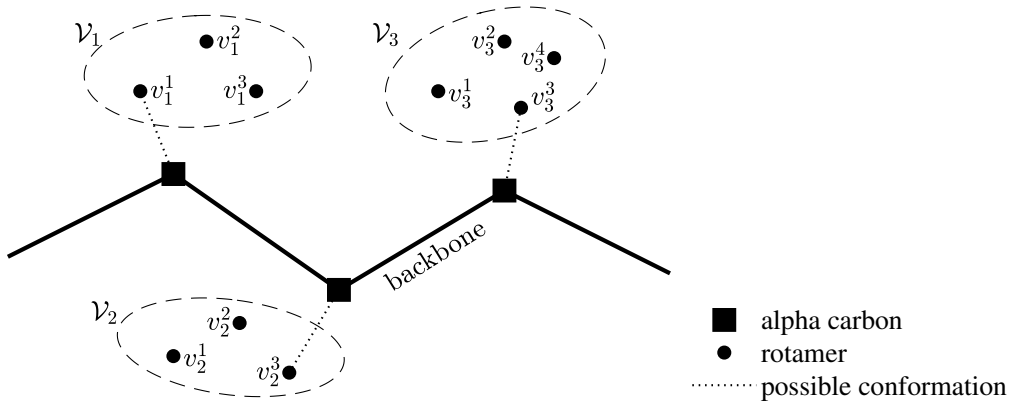


Figure: A Diagram of the Protein Side-Chain Positioning Problem

Problem Formulation as **IQP**

$$
\begin{aligned}
p^*_{\mathbf{IQP}} := \quad & \min \quad && \sum_{u,v} E_{uv} x_u x_v \\
& \text{s.t.} && \sum_{u \in \mathcal{V}_k} x_u = 1, \quad k = 1, \ldots, p \\
& && x = (x_u) \in \{0,1\}^{n_0}.
\end{aligned}
$$

Equivalently,

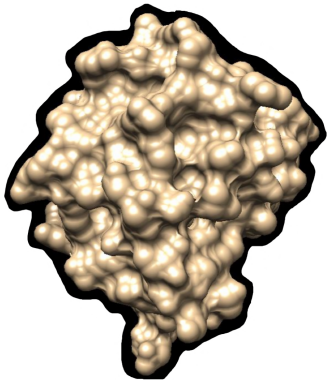
$$
\begin{aligned}
p^*_{\mathbf{IQP}} := \quad & \min \quad && x^T E x \\
& \text{s.t.} && Ax = \bar{e}_p \\
& && x = \begin{bmatrix} v_1^T & v_2^T & \ldots & v_p^T \end{bmatrix}^T \in \{0,1\}^{n_0} \\
& && v_i \in \{0,1\}^{m_i}, \ i = 1, \ldots, p,
\end{aligned}
\tag{\textbf{IQP}}
$$

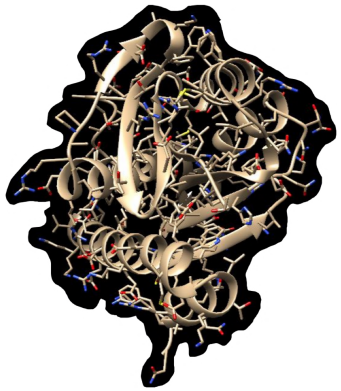
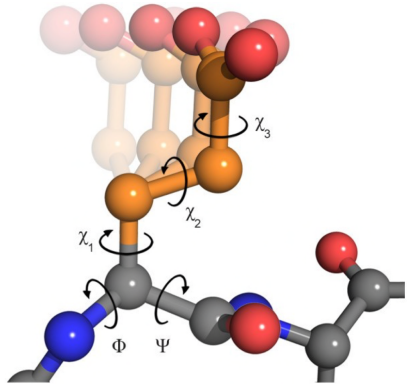
where

$$
A := \text{blkdiag}(\bar{e}_{m_1}^T, \bar{e}_{m_2}^T, \cdots, \bar{e}_{m_p}^T) \in \mathbb{R}^{p \times n_0},
$$

Strategy

The **SCP** problem is NP-hard $\rightarrow$ Work with e.g., **SDP** relaxation



F.J. Burkowski.
Computational and Visualization Techniques for Structural Bioinformatics Using Chimera.
Chapman & Hall/CRC mathematical and computational biology series. Chapman and Hall/CRC, London, 2015.