Central Dogma of Molecular Biology

Understanding the 3-D structure of proteins is essential to understanding life

Algorithmic robotics and Motion Planing

Dynamic Maintenance of Kinematic Chains

Itay Lotan

Protein Folding

- Protein chain spontaneously collapses (folds) to compact 3-D structure

- Recurring structural elements:
  - α-helix
  - β-strand
  - loop

Chain of Amino-Acids

- 20 naturally occurring amino-acids

- The amino acids concatenate to form the protein

Molecule ≈ Robot

Conformational space

Ala, Pro, Trp, Arg
Monte Carlo Simulation (MCS)

- Popular method for sampling the conformation space of proteins
- Used for
  - estimating thermodynamic quantities
  - searching for low-energy conformations and the folded structure

MCS: How It Works

1. Perturb current conformation at random
2. Compute energy of new conformation
3. Accept with probability:
   \[ P(\text{accept}) = \min\{1, e^{-\Delta E/k_B T}\} \]
   \( \{ \Delta E \text{ - energy, } k_B \text{ - Boltzmann's constant, } T \text{ - temperature} \} \)

Requires \( \gg 10^9 \) steps to sample adequately

Energy Function

\[ E = \sum \text{bonded terms} + \sum \text{non-bonded terms} + \sum \text{solvation terms} \]

- Bonded terms
  - Bond-stretching, bending, rotating
  - \( O(n) \)
- Non-bonded terms
  - E.g., Van der Waals and electrostatic
  - Depend on distances between pairs of atoms
  - \( O(n^2) \) - Expensive to compute
- Solvation terms
  - May require computing molecular surface

Energy Function

\[ E = \sum \text{bonded terms} + \sum \text{non-bonded terms} + \sum \text{solvation terms} \]

- Bonded terms
  - Bond-stretching, bending, rotating
  - \( O(n) \)
- Non-bonded terms
  - E.g., Van der Waals and electrostatic
  - Depend on distances between pairs of atoms
  - \( O(n^2) \) - Expensive to compute
- Solvation terms
  - May require computing molecular surface

Non-Bonded Terms

- Energy terms go to 0 when distance increases
  - Cutoff distance (6 - 12Å)
- vdW forces prevent atoms from bunching up
  - Only \( O(n) \) interacting pairs

Finding Pairs

<table>
<thead>
<tr>
<th>Method</th>
<th>Update</th>
<th>Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brute Force</td>
<td>( \Theta(n) )</td>
<td>( \Theta(n^2) )</td>
</tr>
<tr>
<td>Grid</td>
<td>( \Theta(n) )</td>
<td>( \Theta(n) )</td>
</tr>
<tr>
<td>Spatially-adapted hierarchy</td>
<td>( O(n \log n) )</td>
<td>( \Theta(n) )</td>
</tr>
<tr>
<td>Chain-aligned hierarchy</td>
<td>( O(n) )</td>
<td>( \Theta(n^{4/3}) )</td>
</tr>
<tr>
<td>ChainTree</td>
<td>( O(k \log \frac{n}{k}) )</td>
<td>( \Theta(n^{4/3}) )</td>
</tr>
</tbody>
</table>

Challenge: find interacting pairs without enumerating all atom pairs?
Finding Pairs

<table>
<thead>
<tr>
<th></th>
<th>Update</th>
<th>Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brute Force</td>
<td>(\Theta(n))</td>
<td>(\Theta(n^2))</td>
</tr>
<tr>
<td>Grid</td>
<td>(\Theta(n))</td>
<td>(\Theta(n))</td>
</tr>
<tr>
<td>Spatially-adapted hierarchy</td>
<td>(O(n \log n))</td>
<td>(\Theta(n))</td>
</tr>
<tr>
<td>Chain-aligned hierarchy</td>
<td>(O(n))</td>
<td>(O(n^{4/3}))</td>
</tr>
<tr>
<td>ChainTree</td>
<td>(O(k \log \frac{n}{k}))</td>
<td>(O(n^{4/3}))</td>
</tr>
</tbody>
</table>

Grid Method

- Subdivide 3-space into cubic cells
- Compute cell that contains each atom center
- Represent grid as hash table

Asymptotically optimal in worst-case
Finding Pairs

<table>
<thead>
<tr>
<th>Update</th>
<th>Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brute Force</td>
<td>$\Theta(n)$</td>
</tr>
<tr>
<td>Grid</td>
<td>$\Theta(n)$</td>
</tr>
<tr>
<td>Spatially-adapted hierarchy</td>
<td>$O(n \log n)$</td>
</tr>
<tr>
<td>Chain-aligned hierarchy</td>
<td>$O(n)$</td>
</tr>
<tr>
<td>ChainTree</td>
<td>$O(k \log \frac{n}{k})$</td>
</tr>
</tbody>
</table>

Protein as Kinematic Chain

- links
- joints / DoFs (degrees of freedom)

Properties of kinematic chains

- Small changes $\Rightarrow$ large effects
Properties of kinematic chains

- Small changes ⇒ large effects
- Local changes ⇒ global effects
- Few DoF changes ⇒ long rigid sub-chains

ChainTree: A tale of two hierarchies

- Transform hierarchy: approximates kinematics of protein backbone at successive resolutions
- Bounding volume hierarchy: approximates geometry of protein at successive resolutions

Hierarchy of Transforms

Rigid Body Transform = Translation + Rotation

\[
T = \begin{bmatrix}
    t_0 \\
    t_1 \\
    t_2 \\
\end{bmatrix}
\quad R = \begin{bmatrix}
    r_{0,0} & r_{0,1} & r_{0,2} \\
    r_{1,0} & r_{1,1} & r_{1,2} \\
    r_{2,0} & r_{2,1} & r_{2,2} \\
\end{bmatrix}
\]
Hierarchy of Transforms

The ChainTree

Updating the ChainTree

Detecting Interactions

Update path to root:
- Recompute transforms that "shortcut" the DoF change
- Recompute BVs that contain the DoF change
- \( O(k \log_2(n/k)) \) work for \( k \) simultaneous changes

Pruning rules:
1. Prune search when distance between bounding volumes is more than cutoff distance
2. Do not search inside rigid sub-chains (no marked node between the tested nodes)
Computational complexity

- Updating:
  \[ O\left(k \log \frac{n}{k}\right) \]

- Searching:
  \[ \Theta\left(n^{4/3}\right) \] worst case bound

Much faster in practice

---

Proof

- Regularize the chain:
  - Replace all links by their bounding spheres
  - Make all sphere equal in size by growing smaller spheres
  - Further grow all links equally until each pair of consecutive links is in contact

No effect on complexity of the problem

---

Proof

**Assumption**: Distance between link centers > \( \delta \cdot r \)

**Proposition**: Each link overlaps at most \( M \) other links

\[ M_i = 27 \cdot (2^i)^2 / q \quad 0 < q < 1 \]

At level \( i \), replace \( r \) with \( 2^i \cdot r \)

**Lemma**: given two OBBs inside sphere of radius \( R \), their OBB fits inside a sphere of radius \( \sqrt{3} \cdot R \)

**Proof**: We will prove the existence of \( c \) by induction

We still need to show this is true for the ChainTree BVs

We will prove the existence of \( c \) by induction
Proof

**Base:** for levels 0 through 4 choose an appropriate $c = c_1$

**Induction step:**
- Each BV at level $i-5$ fits inside a sphere of radius $c \cdot 2^{i-5}r$
- A sphere of radius $2r$ is enough to bound all links inside a BV at level $i$
- No point in any of the 32 BVs at level $i-5$ is further than $2r + 2 \cdot c \cdot 2^{i-5}r$ from the center of this sphere
- No point in any of the 16 BVs at level $i-4$ is further than $\sqrt{3}(2r + 2 \cdot c \cdot 2^{i-5}r)$ from the center of this sphere
Proof

\( (2^i r + 2c \cdot 2^{-i-5} r) \cdot \sqrt{3}^i \leq c \cdot 2^i r \)

\[ \left( \frac{1}{\sqrt{3}} - 1 \right)^{-1} \leq c \]

\[ c = \max \left\{ \left( \frac{1}{\sqrt{3}} - 1 \right)^{-1}, c_i \right\} \]

Proof

Each link is a sphere of radius r
- Place d links along x-axis (starting at origin)
- Place d links parallel to y-axis
- Place d links parallel to z-axis
This constitutes a unit

Proof

- Place d/8 units one next to the other each translated by (2r, -2r, 0)
This constitutes a layer

Proof

- The point \( (2(d-1)r, (d-1)r, \sqrt{4/(d-1) r}) \) is contained inside convex hull of each unit
- There is a level in BVH where each unit has its own BV
- At this level all BV pairs intersect
- We have \( \frac{1}{2} \left[ \frac{d^2}{64} \right] \) overlaps: \( \Omega(n^{4/3}) \)
Experimental Setup

- Energy function:
  - Van der Waals
  - Electrostatic
  - Attraction between folded-state contacts
  - Cut off at 12Å
- 300,000 steps MCS with Grid and ChainTree
- **Steps are the same with both methods**
- Early rejection for large vdW terms

Test

Two-Pass ChainTree
(ChainTree+)

1st pass: small cutoff distance to detect steric clashes
2nd pass: normal cutoff distance

Computing the Energy

Recursively search ChainTree for interactions

Pruning rules:
1. Prune search when distance between bounding volumes is more than cutoff distance
2. Do not search inside rigid sub-chains (no marked node between the tested nodes)
Computing the energy

Computing the energy

Computing the energy

Computing the Energy

Computing the Energy

The EnergyTree
Extension: Interaction with Solvent

- Implicit solvent model: solvent as continuous medium, interface is solvent-accessible surface

E. Eyal, D. Halperin, Dynamic Maintenance of Molecular Surfaces under Conformational Changes. Proc. 21st ACM Symposium on Computational Geometry (SoCG05) 2005