University of Munich
Computer Science

R. Backofen  S. Will  P. Clore

an algorithmic approach

folding

Quantifying the hydrophobic force in protein
Outline

- Medium-sized proteins (E. Coli RecA, Zeb, Pythrococin)
- W. Gilbert's database of ancient conserved regions
- GA-predicted conformation and PDB data
- Comparison of RMSE between distance matrix of autotomorphism groups
- Energy computation for arbitrary lattices, using
  Energy computation (O(n) energy computation for arbitrary lattices)
- Hybrid Genetic Algorithm (GA)
  3-dimensional face-centered cubic lattice (FCC)
- Energy - pairwise contact potentials using Woose polar
functions not fully correct.

Moreover, studies by M. Teeter indicate energy
Molecular dynamics can simulate $10^{-7}$ seconds of folding.

...nonpolar residues to minimize contact with solvent.

Coal: quantitatively hydrogen-bonding, etc.

...hydrogen-bonding, etc.

(dipole-dipole and sphere packing), Tennant-Jones,
with terms for Coulomb (electrostatic force), van der Waals

\[ E = E_{\text{Coulomb}} + E_{\text{Van Der Waals}} + E_{\text{Tennant-Jones}} + E_{\text{hydrogen bonding}} + \ldots \]

Energy function

Energy

Hypotheses: native state is the conformation which minimizes...
\[ \begin{array}{c}
else \quad 0 \\
I = H = ?_I \ \ \ \ \ I 
\end{array} \] = \mathcal{C} \mathcal{B} \\
\quad \text{and} \\
\begin{array}{c}
else \quad 0 \\
I = |\mathcal{C}_I - ?_I| \ \ \ \ I 
\end{array} = (\mathcal{C}_I ?_I) \nabla \\
\quad \text{where} \\
\left(\mathcal{C}_I ?_I\right) \nabla ^{u \geq \mathcal{C} \geq ?} \int \mathcal{B} - = \mathcal{E} \quad (1) \\
\text{Contact energy for DIII's HPE model:} \\
\bullet \]
\[
\frac{\text{ASP (13.0)}}{x - 13} = d
\]

and ASP (13.0).

Requirement values: 4.8 - 13.0, i.e., Eq. (4.9).

where \( p \) is normalized polar requirement, whose polar

\[
\sum_{u \leq f > t \geq 1} \nabla \cdot \rho \cdot d \cdot d = R
\]

Normalized Polar Requirement
### Face-Centered Cubic (FCC) Lattice, using relative directions.

<table>
<thead>
<tr>
<th>Direction</th>
<th>0</th>
<th>3</th>
<th>2/3</th>
<th>6/3</th>
<th>5/6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>6/3</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>3/2</td>
<td></td>
<td>6/3</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>3/3</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>3/2</td>
<td></td>
<td>6/3</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>6/3</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>3/2</td>
<td></td>
<td>6/3</td>
</tr>
</tbody>
</table>

#### 2. Lattice Models
Euclidean distance in conformation $c$.

$|q_u - r_u| = q_{ip}$

2. Compute $DP(p)_{ip}$, where

self-avoiding walk in the FCC lattice.

1. Use $G_A$ to determine predicted conformation $c$ as a normalized polar requirement values.

the average $q$-carbon positions in $T/30$-size regions, and average the average $q$-carbon positions in $T$ to represent $0$ to $30$ size regions. By taking

requirement values. For more than $30$-$40$ residues, contact PDB

Input: $q$-carbon coordinates from PDB and normalized polar

Pseudocode
3. Compute $D_{RD}$.

4. Compute $RSDM_{HP, D_{PDB}}$.

successive $\alpha$-carbons in linear chain.

in PDB conformation, where $\bar{\sigma}$ is average distance between

\[
\bar{\sigma} = \frac{\sigma_{ii}}{\sigma_{ii} - \sigma_{jj}}
\]

where $D_{PDB} = (\bar{\sigma})$, where
\[
\frac{E \left( \frac{RMSD_{rc}}{RMSD_{rc} - RMSE_c} \right)}{|\{ RC \, : \, RMSE_c < RMSE_{rc} \}|}
\]

Hyd. meas. 2 is

Hydrophobic force. Hyd. meas. 1 is

Output: \( RMSE_{rc} \) between conformation \( C \) found by CA and

Conformation from PDB data, and percent contribution of the

Conformation between conformation \( C \) found by CA and
<table>
<thead>
<tr>
<th>Name</th>
<th>Energy</th>
<th>rmsd</th>
<th>hyd.meas. 1</th>
<th>hyd.meas. 2</th>
<th>s.d. (hyd.meas. 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aat</td>
<td>-12.01337</td>
<td>0.92424</td>
<td>99.94</td>
<td>61.37</td>
<td>0.47</td>
</tr>
<tr>
<td>acid</td>
<td>-11.39643</td>
<td>0.88077</td>
<td>99.94</td>
<td>61.37</td>
<td>0.47</td>
</tr>
<tr>
<td>acyl</td>
<td>-11.18783</td>
<td>1.05592</td>
<td>99.87</td>
<td>60.7</td>
<td>0.47</td>
</tr>
<tr>
<td>aldeh</td>
<td>-15.96610</td>
<td>1.08664</td>
<td>99.23</td>
<td>55.21</td>
<td>0.47</td>
</tr>
<tr>
<td>alcoh</td>
<td>-12.57532</td>
<td>1.17564</td>
<td>99.96</td>
<td>61.71</td>
<td>0.47</td>
</tr>
<tr>
<td>alcoh</td>
<td>-12.50156</td>
<td>0.92853</td>
<td>99.6</td>
<td>52.81</td>
<td>0.47</td>
</tr>
<tr>
<td>alcoh</td>
<td>-11.31886</td>
<td>1.09688</td>
<td>99.87</td>
<td>60.7</td>
<td>0.47</td>
</tr>
<tr>
<td>alk</td>
<td>-11.98783</td>
<td>1.09688</td>
<td>99.23</td>
<td>55.21</td>
<td>0.47</td>
</tr>
<tr>
<td>alcoh</td>
<td>-13.67139</td>
<td>0.88077</td>
<td>99.94</td>
<td>61.37</td>
<td>0.47</td>
</tr>
<tr>
<td>alcoh</td>
<td>-10.42641</td>
<td>0.99311</td>
<td>99.94</td>
<td>61.37</td>
<td>0.47</td>
</tr>
<tr>
<td>alk</td>
<td>-10.31886</td>
<td>1.09688</td>
<td>99.23</td>
<td>55.21</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>8.3</td>
<td>9.3</td>
<td>6.4</td>
<td>7.6</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>7.6</td>
<td>4.4</td>
<td>3.7</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>6.3</td>
<td>3.4</td>
<td>4.5</td>
<td>6.7</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>8.5</td>
<td>9.9</td>
<td>9.9</td>
<td>9.9</td>
<td>9.9</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
<td>6.7</td>
<td>4.5</td>
<td>3.4</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>8.5</td>
<td>9.9</td>
<td>9.9</td>
<td>9.9</td>
<td>9.9</td>
</tr>
<tr>
<td></td>
<td>0.4704</td>
<td>1.2</td>
<td>2.T</td>
<td>99.67</td>
<td>1.1136</td>
</tr>
<tr>
<td>-----</td>
<td>--------</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>1.027</td>
<td>99.67</td>
<td>1.074412</td>
<td>1.13577</td>
<td>1.161589</td>
</tr>
<tr>
<td></td>
<td>0.477</td>
<td>99.67</td>
<td>1.074412</td>
<td>1.13577</td>
<td>1.161589</td>
</tr>
<tr>
<td></td>
<td>4.693</td>
<td>69.67</td>
<td>1.074412</td>
<td>1.13577</td>
<td>1.161589</td>
</tr>
<tr>
<td></td>
<td>0.473</td>
<td>69.67</td>
<td>1.074412</td>
<td>1.13577</td>
<td>1.161589</td>
</tr>
<tr>
<td></td>
<td>0.471</td>
<td>69.67</td>
<td>1.074412</td>
<td>1.13577</td>
<td>1.161589</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Nr.</td>
<td>Energy</td>
<td>RMSD</td>
<td>hyd.meas.1</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>--------</td>
<td>------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>2reb</td>
<td>4</td>
<td>-12.384</td>
<td>1.061</td>
<td>48.15</td>
<td></td>
</tr>
<tr>
<td>2reb</td>
<td>3</td>
<td>-12.524</td>
<td>1.146</td>
<td>97.79</td>
<td></td>
</tr>
<tr>
<td>2reb</td>
<td>2</td>
<td>-12.266</td>
<td>1.167</td>
<td>97.42</td>
<td></td>
</tr>
<tr>
<td>2reb</td>
<td>1</td>
<td>-12.055</td>
<td>1.062</td>
<td>99.25</td>
<td></td>
</tr>
<tr>
<td>2act</td>
<td>3</td>
<td>-13.545</td>
<td>1.126</td>
<td>99.25</td>
<td></td>
</tr>
<tr>
<td>2act</td>
<td>2</td>
<td>-12.974</td>
<td>1.039</td>
<td>99.85</td>
<td></td>
</tr>
<tr>
<td>2act</td>
<td>1</td>
<td>-13.846</td>
<td>1.123</td>
<td>99.85</td>
<td></td>
</tr>
<tr>
<td>Leca</td>
<td>4</td>
<td>-15.409</td>
<td>2.046</td>
<td>53.85</td>
<td></td>
</tr>
<tr>
<td>Leca</td>
<td>3</td>
<td>-16.552</td>
<td>2.119</td>
<td>49.49</td>
<td></td>
</tr>
<tr>
<td>Leca</td>
<td>2</td>
<td>-15.612</td>
<td>2.061</td>
<td>52.25</td>
<td></td>
</tr>
<tr>
<td>Leca</td>
<td>1</td>
<td>-16.554</td>
<td>2.122</td>
<td>48.15</td>
<td></td>
</tr>
</tbody>
</table>
of naive \( O(u^2) \) algorithm for both

computation and \( O(\log u) \) test for self-avoiding walk, instead

for arbitrary 2- and 3-dimensional lattices, has \( \Theta(u^2 \log u) \) energy

Our contribution: uses lattice automorphisms, oct-trees, works

(left), \( \mathcal{R} \) (right).

2-dimensional square lattice. Relative directions \( \mathcal{S} \) (straight), \( \mathcal{T} \)

Unger-Moulton used hybrid genetic algorithm for folding in

Methods
Hybrid GA

place c in next generation

\[
\frac{1}{\exp(-e_{ave} (t) + \text{random}(0,1))} > p(c) \Rightarrow \text{ave} = \text{average}(P(w), P(f))
\]

\[
\text{ave} = \text{average}(P(w), P(f))
\]

produce child c by crossover of m,f

select 2 chromosomes, m,f

while n > p

0=n

potentiate mutation

++

repeat

\[
(x \in P(t)) \text{ best} = \max_{x \in P(t)} P(x)
\]

initialize population \( P(t) \) of random calls

0 = t
\[
\begin{pmatrix}
0 \\
1 \\
0
\end{pmatrix} = \mathcal{T} \omega
\]

\[
\begin{pmatrix}
0 \\
0
\end{pmatrix} = \mathcal{F} \omega
\]

Vector \( \omega \):

Relative move in element of \( \mathcal{D}, \mathcal{I}, \mathcal{R}, \mathcal{T} \) with corresponding relative moves in cubic lattice
A sequence \( w \in \{ \mathcal{D}, \cup, \mathcal{R}, \mathcal{T}, \mathcal{I} \} \) is called a relative move sequence.

\[
\begin{pmatrix}
1 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0
\end{pmatrix} = \alpha
\]

\[
\begin{pmatrix}
0 & 1 \\
1 & 0 \\
0 & 0 \\
0 & 0
\end{pmatrix} = \mathcal{H}
\]

\[
\begin{pmatrix}
1 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0
\end{pmatrix} = \Omega
\]
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix} = B
\]

Thus
\[
\begin{pmatrix}
0 \\
0 \\
1
\end{pmatrix}
\]
rotation matrices transforming into \( \psi \) are \( \{D', R', T', H\} \) for \( m \) and \( B \)
where \( I_3 \) is identity and \( B \) for \( m \)

\[
_{\text{def}} \begin{cases}
_{\text{def}} m' m = m \quad m' B \circ (m) \quad m' B \circ (m) = (m)^\psi \\
_{\text{def}} e = m \quad \text{id}
\end{cases}
\]
$$B_r = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

$$B_u = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$$B_l = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$
Given sequence \( m \), conformation \( c \) is found when

\[
\begin{pmatrix}
0 & 0 & 1 \\
0 & 1 & 0 \\
1 & 0 & 0
\end{pmatrix}
\]

and for all \( \lambda \geq \lambda' \geq \lambda \),

\[
|m| \geq \lambda' \geq \lambda \Rightarrow
\begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix} = [\lambda']_c
\]

and

\[
?m_n \circ (1-?m \cdots 1m)^w \mathcal{E} + [1 - ?]c = [\lambda]c
\]
Let \( u = \text{FULLDL} \). The resulting conformation is

Example
Let $\mathcal{W}$ be the rotation of $\mathbb{Z}$ such that for all $m \geq 0$ and all $\eta$, we have $|\mathcal{W} m| \geq \eta$. Let $m = \mathcal{W} m'$, and let $c = \text{con}(m')$ and $c' = \text{con}(m')$.

Then there exists a sequence $\{0, 0, 0, \ldots\}$ that satisfies

$$
\begin{pmatrix}
0 \\
0 \\
1
\end{pmatrix} \circ (m)^{\text{qassem}} = [I - \eta]c - [\eta]c'
$$

Then $\mathcal{W}$ is given by

**Theorem** 1. Let $m = m_1 \cdot m_2 \cdot \ldots \cdot m_n$, and let $c = \text{con}(m_1)$. Let $c' = \text{con}(m')$. Then

$$
[\eta]c = \mathcal{W} m = \mathcal{W} m_1 \cdot \mathcal{W} m_2 \cdot \ldots \cdot \mathcal{W} m_n.
$$
\[(\mathcal{W} \circ [\varphi]) : |m| \geq \varphi \geq \mathfrak{I} + |\mathfrak{I}m| \forall m \in \mathbb{Z} \]\n
and \[(\mathcal{W} = [\varphi]) : |\mathfrak{I}m| \geq \varphi \geq 1 \forall m \in \mathbb{Z} \]

Corollary 1 (Mutation) Let \(w = \lim_{n \to \infty} w_n \) and \(w' = \lim_{n \to \infty} w'_n \) be
{\{ T \in \mathcal{B} \mid \varnothing \circ T \} = T \text{ the } \mathcal{B} \text{ of } \mathbb{R}^{n} \text{ is an isometry of } \mathbb{R}^{n} \}.

Lattice automorphism \( B \) is an isometry of \( \mathbb{R}^{n} \) with the property

Lattice automorphism \( B \) is an isometry of \( \mathbb{R}^{n} \) with the property

Lattice automorphism \( B \) is an isometry of \( \mathbb{R}^{n} \) with the property

\begin{align*}
\text{Definition 2 (Lattice Automorphism)} \quad & \text{Let } I \text{ be a lattice. A } \\
& \text{The vectors } v_1, \ldots, v_m \text{ are called the basis of } I.
\end{align*}

\begin{align*}
& \text{If } n \in I \text{ and } n \in L, \text{ then } n + v \text{ and } n - v \text{ are also in } I.
\end{align*}

\begin{align*}
& \text{That lattice generated by these vectors is the smallest set } L \subset \mathbb{R}^{n} \text{ such}
\end{align*}

\begin{align*}
& \text{Definition 1 (Lattice)} \quad \text{Let } v_1, \ldots, v_m \text{ be vectors in } \mathbb{R}^{n} \text{. The}
\end{align*}

\begin{align*}
& \text{Lattices and Lattice Automorphisms}
\end{align*}
\[ \{ w \mathbf{Z} \in \mathcal{T} \mid \mathcal{T} \circ \mathcal{W} \} = \mathcal{T} \] Then \( \mathcal{T} \) is the matrix of \( \mathcal{T} \) as the basis of the lattice \( \mathcal{T} \). Let \( \mathcal{T} \) be the corresponding generator. Let \( \mathbf{v}_1, \ldots, \mathbf{v}_m \) be the column vectors forming a basis of lattice \( \mathcal{T} \). The Proposition 2 (Integral Representation)
Example 1 Consider the 2-dimensional, hexagonal lattice \( \mathbb{A}_2 \). The center (0,0) and the six nearest points of \( \mathbb{A}_2 \) are as follows:
The point has integral representation \((\frac{\mathbf{z}}{\mathbf{S}}, \frac{\mathbf{z}}{\mathbf{I}-}) = \mathbb{P}\) if

\[
\begin{pmatrix}
\frac{\mathbf{z}}{\mathbf{S}} & 0 \\
\frac{\mathbf{z}}{\mathbf{I}} & 1 
\end{pmatrix} = \mathcal{V} \mathcal{W}
\]

is matrix is a basis of the generator. The matrix is \((\frac{\mathbf{z}}{\mathbf{S}} \frac{\mathbf{z}}{\mathbf{I}-}) = \mathcal{V} \mathcal{W}\) and \((0, 1) = \mathbb{P} \mathbb{W} \mathbb{N}\)
\[
\begin{cases}
1 + \text{len} \mod\; l & \text{else} \\
1 + \text{len} \mod\; n & \bar{h} = \bar{h}
\end{cases}
\]

\[
= (z, \bar{h}', x)_{\text{COORD}}
\]

Efficient implementation of pivot moves
```
endpoint
end

RETURN, not self-avolating

else

if COORDS[X][Y][Z][VOLD] == 0
    COORDS[X][Y][Z][VOLD] = [T][X]
    (Z', Y', X') = [T][X]

    if COORDS[X][Y][Z][VOLD] == 0 or COORDS[X][Y][Z][VOLD] <= [T][X]
        (Z', Y', X') = A[CONF[STATE] - CONF[STATE]] + CONF[STATE]

for i=STATE to STATE+1 do
    (random(STATE, STATE+1] + CONF[STATE])

    COORDS[STATE+1][STATE] = PIVOT(STATE, STATE+1)
```

```