Refining Knowledge Based Potentials of Proteins

Stephan Bernhart

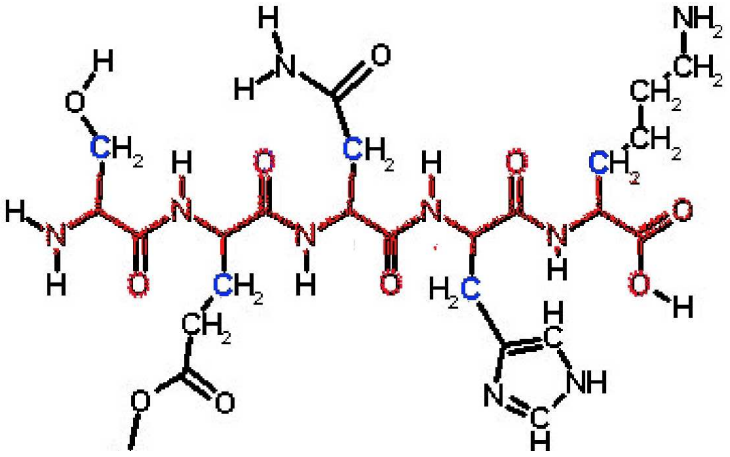
Institute for Theoretical Chemistry and Structural Biology

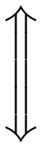
University of Vienna

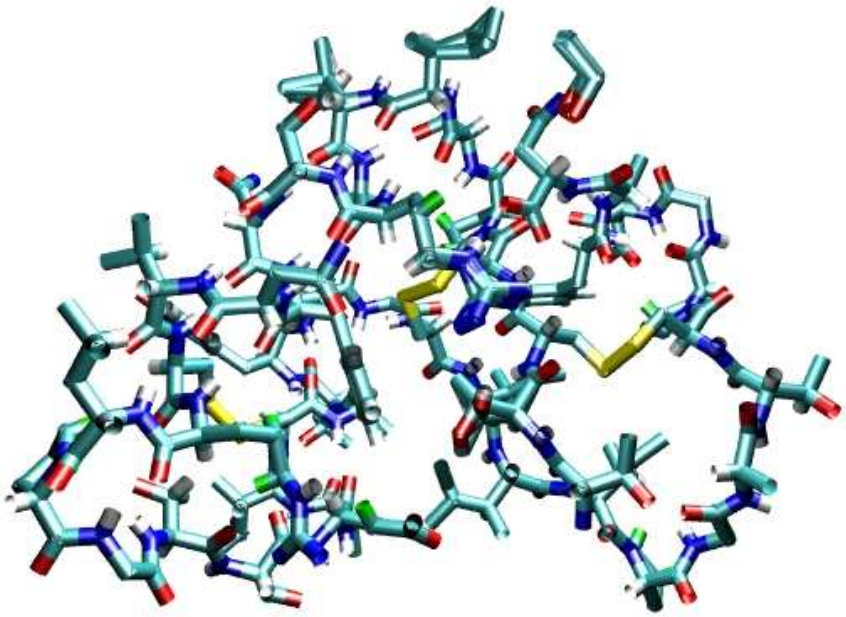
http://www.tbi.univie.ac.at/ berni

Bled, 2002

Proteins







WHY?

- Finding errors in empirically obtained 3-d structures

- Predicting effects of aa mutation on structure

- Inverse folding: structure $\Rightarrow$ sequence

- Ultimately: sequence $\Rightarrow$ structure

"Physical" potential field

Amber

$$
\begin{aligned}
E_{\text{total}} &= \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \\
&+ \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \\
&+ \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \\
&+ \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \right] + \\
&+ \sum_{\text{H-bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right]
\end{aligned}
$$

Knowledge based Potentials

Using Boltzmann's Principle:

Frequently observed states correspond to low energy states

Sippl

$$E_{total} = \sum_{\text{a}<b,s} -kTln \left[\frac{f_{(r)}^{abs}}{f_{(r)}^{s}} \right]$$

Tropsha

$$E_{\text{total}} = \sum_{\text{quadruples}} q_{\text{i}jkl}$$

Possible Problems

'Physical' Potential Field

- Vast energy landscape

- Rugged energy landscape

Knowledge based potential

- Size of Database

- Reliability of Database

Sippl's Approach

Pair interactions depending on:

- type of amino acids a, b
- type of atom c, d
- distance in sequence between a, b: k
- spatial seperation of a, b: r

r is a continouus variable $\longrightarrow$ intervalls
$\Rightarrow f_r^{abcdk}$

redundant information $\longrightarrow$ reference state:
f_r^{cdk}

$$\begin{aligned} \Delta E_r^{abcdk} &= E_r^{abcdk} - E_r^{cdk} = \\ &= -kTln\left(\frac{f_r^{abcdk}}{f_r^{cdk}}\right) \end{aligned}$$

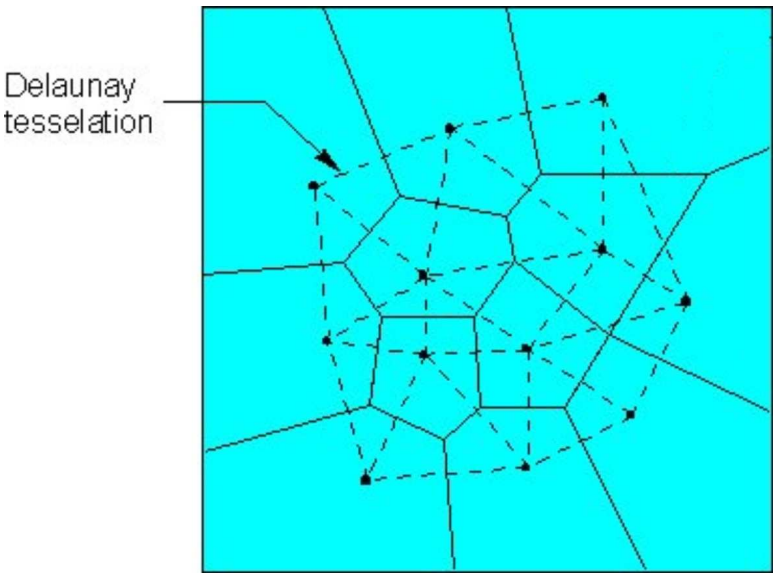
Tropsha's Approach

Filling protein space with tetrahedra

$$E_{\text{total}} = \sum_{\text{tetrahedra}} log\left(\frac{f_{\text{i}jkl}}{p_{\text{i}jkl}}\right)$$

$$p_{\text{i}jkl} = Ca_i a_j a_k a_l$$

Delaunay Triangulation



2 Dimensions: Triangles vertices i, j, k

3 Dimensions: Tetrahedra vertices i, j, k, l

C_α as vertices

Expanding Tropsha's Potential

$$
\begin{aligned}
\sum_{\text{tetrahedras}} q_{\text{ijkl}} &= \sum_{\text{aa}} q_{aa} + \\
&+ \quad \Sigma_{\text{pairs}} \quad q_{pairs}|(q_{aa1}, q_{aa2}) + \\
&+ \quad \Sigma_{\text{triangles}} \quad q_{trian}|(q_{pair1}, q_{pair2}, q_{pair3}) + \\
&+ \quad \Sigma_{\text{4-tupel}} \quad q_{tetra}|(q_{triangle1}, q_{tr2}, q_{tr3}, q_{tr4})
\end{aligned}
$$

$$
\begin{aligned}
q_{pair_{ab}}|(q_a, q_b) &= \frac{f_{pair_{ab}}}{p_{pair_{ab}}|(p_a,p_b)} \\
p_{pair_{ab}}|p_a,p_b &= a\ p_a\ p_b
\end{aligned}
$$

$$
\begin{aligned}
q_{abc}|(q_{ab}, q_{ac}, q_{bc}) &= \frac{f_{triangle_{abc}}}{p_{triangle_{abc}}|(p_{ab}, p_{ac}, p_{bc})} \\
p_{triangle_{abc}}|(p_{ab}, p_{ac}, p_{bc}) &= a\ p_{ab}\ p_{ac}\ p_{bc}\ p_{triangle} \\
p_{triangle} &= N_{triangles}/N_{triples}
\end{aligned}
$$

Tesselated Protein with Water Shell