Formation reactions and relative stabilities of proteins

Jeffrey M. Dick

School of Earth and Space Exploration
Arizona State University
(presentation made in October 2010)

August 22, 2011
Motivation

Background
Protein geochemistry is concerned with the occurrence and variation of proteins in all environments on Earth.

Example
Amino acid differences between mitochondrial and nuclear proteins
Motivation

Background
Protein geochemistry is concerned with the occurrence and variation of proteins in all environments on Earth.

Hypothesis
Molecular evolution is a type of chemical reaction.

Outline
1. Protein Formation Reactions
2. Calculating Relative Stabilities
3. Natural Experiment
Folding Reactions

Cytochrome C

Ribonuclease A
Folding Reactions

- Folding as a conformational process
- Stability referenced to unfolded protein
- Cellular/laboratory timescales

ΔG° ≈ -8kcal/mol
ΔG° ≈ -12kcal/mol

Cytochrome C

Ribonuclease A
Formation Reactions

- Formation as a chemical process
- $\Delta G_f^\circ$: Standard Gibbs energy of formation from the elements

<table>
<thead>
<tr>
<th>Protein</th>
<th>Formula</th>
<th>$\Delta G_f^\circ$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytochrome C</td>
<td>C$<em>{517}$H$</em>{825}$N$<em>{143}$O$</em>{150}$S$_{4}$</td>
<td>$\approx$ -3650</td>
</tr>
<tr>
<td>Ribonuclease A</td>
<td>C$<em>{575}$H$</em>{909}$N$<em>{171}$O$</em>{193}$S$_{12}$</td>
<td>$\approx$ -4960</td>
</tr>
</tbody>
</table>
Formation Reactions

- Formation as a chemical process
- Stability referenced to inorganic species
- $\Delta G_r^\circ$: Standard Gibbs energy of reaction
- Overall energy change is independent of mechanism

<table>
<thead>
<tr>
<th>Cytochrome C</th>
<th>Ribonuclease A</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{517}H_{825}N_{143}O_{150}S_4$</td>
<td>$C_{575}H_{909}N_{171}O_{193}S_{12}$</td>
</tr>
<tr>
<td>539$O_2$</td>
<td>571.5$O_2$</td>
</tr>
<tr>
<td>517$CO_2$</td>
<td>575$CO_2$</td>
</tr>
<tr>
<td>194$H_2O$</td>
<td>186$H_2O$</td>
</tr>
<tr>
<td>143$NH_3$</td>
<td>171$NH_3$</td>
</tr>
<tr>
<td>4$H_2S$</td>
<td>12$H_2S$</td>
</tr>
</tbody>
</table>

$\Delta G_r^\circ \approx 55980 \text{ kcal/mol}$

$\Delta G_r^\circ \approx 59800 \text{ kcal/mol}$
### Protein Formation Reactions

**Calculating Relative Stabilities**

**Natural Experiment**

---

## Residue Equivalents

<table>
<thead>
<tr>
<th>Cytochrome C</th>
<th>Ribonuclease A</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{1}{104} \text{C}<em>{517}\text{H}</em>{825}\text{N}<em>{143}\text{O}</em>{150}\text{S}_{4} )</td>
<td>( \frac{1}{124} \text{C}<em>{575}\text{H}</em>{909}\text{N}<em>{171}\text{O}</em>{193}\text{S}_{12} )</td>
</tr>
<tr>
<td>5.18O₂</td>
<td>4.61O₂</td>
</tr>
<tr>
<td>4.97CO₂</td>
<td>4.64CO₂</td>
</tr>
<tr>
<td>1.86H₂O</td>
<td>1.50H₂O</td>
</tr>
<tr>
<td>1.38NH₃</td>
<td>1.38NH₃</td>
</tr>
<tr>
<td>0.04H₂S</td>
<td>0.10H₂S</td>
</tr>
</tbody>
</table>

\( \Delta G_r^\circ \approx 538 \text{ kcal/mol} \)

\( \Delta G_r^\circ \approx 482 \text{ kcal/mol} \)

---

- Reactions normalized by protein length
- Energetic meaning of reaction coefficients

---

**Environment & Energy**

Shift to lower O₂ potential more strongly decreases Gibbs energy of formation of CYC than RNAS1.
Residue Equivalents

- Reactions normalized by protein length
- Energetic meaning of reaction coefficients
- Transformation reaction; cellular to evolutionary timescales

Environment & Energy
Shift to lower $O_2$ potential more strongly decreases Gibbs energy of formation of CYC than RNAS1.
Group additivity of aqueous species properties for amino acid residues [Dick et al., 2006]
Standard Gibbs Energies

- Group additivity of aqueous species properties for amino acid residues [Dick et al., 2006]
- Protein size dependence of standard Gibbs energies

<table>
<thead>
<tr>
<th>Protein</th>
<th>Size (Residues)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LYSC Lysozyme</td>
<td>129</td>
</tr>
<tr>
<td>CYC Cytochrome C</td>
<td>104</td>
</tr>
<tr>
<td>RNAS1 Ribonuclease A</td>
<td>124</td>
</tr>
<tr>
<td>MYG Myoglobin</td>
<td>153</td>
</tr>
</tbody>
</table>

- Gibbs energies of folding [Privalov and Khechinashvili, 1974] are 1% or less of energies of formation.
Computational Plan

- Calculate standard Gibbs energies of proteins from amino acid group additivity and equations of state for aqueous species.
- Proceed provisionally without energy of protein folding.
  - Per mole of protein, energy of folding is small compared to Gibbs energy of chemical formation reaction.
  - If all proteins are folded, energy of folding tends to cancel in relative stability calculations.
Metastable Equilibrium

- Chemical affinity is opposite of Gibbs energy change due to reaction progress,

\[ dG = -SdT + VdP - Ad\xi \]

- \( G \): Gibbs energy
- \( S \): Entropy
- \( T \): Temperature
- \( V \): Volume
- \( P \): Pressure
- \( A \): Chemical Affinity
- \( \xi \): Reaction Progress

Equal-activity reference state
- More stable: higher affinity (\( A \))

Equal-affinity reference state
- More stable: higher activity (\( a \))
Metastable Equilibrium

- Chemical affinity is opposite of Gibbs energy change due to reaction progress,
- computed as a function of standard Gibbs energy, reaction stoichiometry and chemical activities.

\[ dG = -SdT + VdP - Ad\xi \]

\[ A = 2.303RT \log\left(\frac{K}{Q}\right) \]

- $dG$ Chemical Affinity
- $K$ Equilibrium Constant
- $Q$ Activity Product
- $R$ Gas Constant
Metastable Equilibrium

- Chemical affinity is opposite of Gibbs energy change due to reaction progress,
- computed as a function of standard Gibbs energy, reaction stoichiometry and chemical activities.

\[ dG = -SdT + VdP - Ad\xi \]

\[ A = 2.303RT \log(K/Q) \]

\[ A = -\Delta G_r \]
\[ K = 10^{(-\Delta G^\circ_r / 2.303RT)} \]
\[ Q = 10^{\sum \nu \log a} = \prod a^\nu \]

\( \nu \) Reaction Coefficient
\( a \) Chemical Activity
Metastable Equilibrium

- Chemical affinity is opposite of Gibbs energy change due to reaction progress,
- computed as a function of standard Gibbs energy, reaction stoichiometry and chemical activities.
- Maxwell-Boltzmann distribution allows for a transformation between reference states.

\[ dG = -SdT + VdP - Ad\xi \]

\[ A = 2.303RT \log\left(\frac{K}{Q}\right) \]

\[ \frac{a}{\sum a} = \frac{e^{A/RT}}{\sum e^{A/RT}} \]

Equal-activity reference state
More stable: higher affinity \( A \)
Metastable Equilibrium

- Chemical affinity is opposite of Gibbs energy change due to reaction progress,
- computed as a function of standard Gibbs energy, reaction stoichiometry and chemical activities.
- Maxwell-Boltzmann distribution allows for a transformation between reference states.
- When the chemical affinities of the formation reactions are all equal, the proteins are in metastable equilibrium.

\[ dG = -SdT + VdP - Ad\xi \]

\[ A = 2.303RT \log\left(\frac{K}{Q}\right) \]

\[ \frac{a}{\sum a} = \frac{e^{A/RT}}{\sum e^{A/RT}} \]

**Equal-activity reference state**
More stable: higher activity \((a)\)

**Equal-affinity reference state**
More stable: higher affinity \((A)\)
Computational Plan

- Calculate standard Gibbs energies of proteins from amino acid group additivity and equations of state for aqueous species.
- Proceed provisionally without energy of protein folding.
- Calculate equilibrium activities of proteins.
  - Start with chemical affinities in equal-activity reference state.
  - Use reference state transformation to calculate equilibrium activities.

Investigate temperature, oxidation potential, pH, other chemical potentials.
Explore the protein universe using model systems.

CHNOSZ is the software package used for the preceding calculations.
Formation of Ionized Unfolded CYC\textsubscript{BOVIN}

\[
4.97\text{CO}_2 + 1.38\text{NH}_3 + 0.04\text{H}_2\text{S} + 1.86\text{H}_2\text{O} + 0.08\text{H}^+ \rightarrow \frac{1}{104}\text{C}_{517}\text{H}_{825}\text{N}_{143}\text{O}_{15}\text{S}^{+0.08} + 5.18\text{O}_2
\]

- Ionization of proteins using additivity

[more info?]
Formation of Ionized Unfolded CYC_BOVIN

\[ 4.97\text{CO}_2 + 1.38\text{NH}_3 + 0.04\text{H}_2\text{S} + 1.86\text{H}_2\text{O} + 0.08\text{H}^+ \rightarrow C_{4.97}H_{7.93}N_{1.38}O_{0.14}S_{0.04} + 5.18\text{O}_2 \]

- Ionization of proteins using additivity
- Write per-residue formulas.
Formation of Ionized Unfolded CYC_BOVIN

4.97CO₂ + 1.38NH₃ + 0.04H₂S + 1.86H₂O + 0.08H⁺ → C₄.97H₇.93N₁.38O₀.₁₄S⁺₀.₀₈ + 5.1₈O₂

$$\log K = -\Delta G^o / 2.303RT$$

$$= -393.0$$

- Ionization of proteins using additivity
- Write per-residue formulas.

$$T \quad 25 ^\circ C$$

$$P \quad 1 \text{ bar}$$
Formation of Ionized Unfolded CYC_BOVIN

4.97CO₂ + 1.38NH₃ + 0.04H₂S + 1.86H₂O + 0.08H⁺ → C₄.97H₇.93N₁.38O₀.₁₄S⁺₀.₀₈ + 5.18O₂

\[ \log Q = \log a_{C₄.97H₇.93N₁.38O₀.₁₄S⁺₀.₀₈} + 5.18 \log f_{O₂} - 4.97 \log a_{CO₂} - 1.38 \log a_{NH₃} - 0.04 \log a_{H₂S} - 1.86 \log a_{H₂O} - \log a_{H⁺} \]

\[ = -393.4 \]
Protein Formation Reactions

Calculating Relative Stabilities

Formulation Properties

<table>
<thead>
<tr>
<th>Protein</th>
<th>log $K$</th>
<th>log $Q$</th>
<th>log $a$</th>
<th>$A/2.303RT$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LYSC</td>
<td>$-361.6$</td>
<td>$-357.4$</td>
<td>0</td>
<td>$-4.17$</td>
</tr>
<tr>
<td>CYC</td>
<td>$-393.0$</td>
<td>$-393.4$</td>
<td>0</td>
<td>$0.34$</td>
</tr>
<tr>
<td>RNAS1</td>
<td>$-352.6$</td>
<td>$-348.4$</td>
<td>0</td>
<td>$-4.27$</td>
</tr>
<tr>
<td>MYG</td>
<td>$-407.6$</td>
<td>$-408.6$</td>
<td>0</td>
<td>$0.96$</td>
</tr>
</tbody>
</table>

Equal activity reference state: MYG is more stable.
# Formation Properties

<table>
<thead>
<tr>
<th>Protein</th>
<th>log $K$</th>
<th>log $Q$</th>
<th>log $a$</th>
<th>$A/2.303RT$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LYSC</td>
<td>−361.6</td>
<td>−362.0</td>
<td>−4.62</td>
<td>0.45</td>
</tr>
<tr>
<td>CYC</td>
<td>−393.0</td>
<td>−393.5</td>
<td>−0.10</td>
<td>0.45</td>
</tr>
<tr>
<td>RNAS1</td>
<td>−352.6</td>
<td>−353.1</td>
<td>−4.72</td>
<td>0.45</td>
</tr>
<tr>
<td>MYG</td>
<td>−407.6</td>
<td>−408.0</td>
<td>0.51</td>
<td>0.45</td>
</tr>
</tbody>
</table>

- Equal activity reference state: MYG is more stable.
- Equal affinity reference state: MYG is still more stable!
### Formation Properties

<table>
<thead>
<tr>
<th>Protein</th>
<th>log $K$</th>
<th>log $Q$</th>
<th>log $a$</th>
<th>$A/2.303RT$</th>
<th>log $a_{\text{protein}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LYSC</td>
<td>−361.6</td>
<td>−362.0</td>
<td>−4.62</td>
<td>0.45</td>
<td>−6.73</td>
</tr>
<tr>
<td>CYC</td>
<td>−393.0</td>
<td>−393.5</td>
<td>−0.10</td>
<td>0.45</td>
<td>−2.12</td>
</tr>
<tr>
<td>RNAS1</td>
<td>−352.6</td>
<td>−353.1</td>
<td>−4.72</td>
<td>0.45</td>
<td>−6.81</td>
</tr>
<tr>
<td>MYG</td>
<td>−407.6</td>
<td>−408.0</td>
<td>0.51</td>
<td>0.45</td>
<td>−1.68</td>
</tr>
</tbody>
</table>

- Equal activity reference state: MYG is more stable.
- Equal affinity reference state: MYG is still more stable!
- Molality of residue $=$ molality of protein * protein length
- Activity of residue $=$ activity of protein * protein length (assuming ideality)
Equilibrium Activity Diagrams

- Metastable equilibrium activities with total activity of residues $= 4$
- We are using $T = 25 \, ^\circ C$ and $pH = 7$

It Shows ...

MYG is relatively most stable at $\log f_{O_2} = -80$. 
Equilibrium Activity Diagrams

- Metastable equilibrium activities as a function of $\log f_{O_2}$ with total activity of residues $= 4$
- $\log f_{O_2}$ can be converted to other measurements of oxidation-reduction potential.

It Shows ...

Relative stability is sensitive to oxidation potential.
Computational Plan

- Calculate standard Gibbs energies of proteins from amino acid group additivity and equations of state for aqueous species.
- Proceed provisionally without energy of protein folding.
- Calculate equilibrium activities of proteins.
  - Relative stabilities depend on the species (chemical compositions, standard Gibbs energies).
  - Relative stabilities depend on the environment (temperature, pressure, activities/fugacities of basis species).
Elongation Factor Tu

- EF-Tu from *Escherichia coli*
Elongation Factor Tu

- EF-Tu’s from *Escherichia coli*, *Thermotoga maritima*, *Thermus thermophilus*
- and reconstructed by maximum likelihood (ML) stem of bacterial tree, stem of mesophilic bacteria, and Alternative tree topology [Gaucher et al., 2003]
- This tree built using parsimony (PHYLIP software), 394 aligned amino acids.

Elongation Factor Tu

- EF-Tu’s from *Escherichia coli*, *Thermotoga maritima*, *Thermus thermophilus*
- and reconstructed by maximum likelihood (ML) stem of bacterial tree, stem of mesophilic bacteria, and Alternative tree topology [Gaucher et al., 2003]
- This tree built using parsimony (PHYLIP software), 394 aligned amino acids.
Equilibrium activity diagram

In most reactions, proteins from higher temperature favored by lower $\log f_{O_2(g)}$. 

- Equilibrium activity diagram
- In most reactions, proteins from higher temperature favored by lower $\log f_{O_2(g)}$. 

Equilibrium Activity Diagrams
Equilibrium Activity Diagrams

- Equilibrium predominance diagram
- At constant $f_{O_2(g)}$, increasing $T$ tends to favor “ML-stem”.

![Equilibrium Activity Diagram](image)
Equilibrium Activity Diagrams

- Equilibrium predominance diagram
- \( \log f_{O_2} = -60 \)
**Equilibrium Activity Diagrams**

- Equilibrium predominance diagram
- $T = 25 \, ^\circ \text{C}$

**Multidimensionality**

What about $\log a_{\text{H}_2\text{O}}$, $\log a_{\text{CO}_2}$, etc.?
Computational Plan

- Calculate standard Gibbs energies of proteins from amino acid group additivity and equations of state for aqueous species.
- Proceed provisionally without energy of protein folding.
- Calculate equilibrium activities of proteins.
- Investigate temperature, oxidation potential, pH, other chemical potentials.
  - Are reducing conditions associated with hotter environments?
  - The system is multidimensional; could also vary the chemical potentials of carbon, nitrogen, sulfur.
Computational Plan

- Calculate standard Gibbs energies of proteins from amino acid group additivity and equations of state for aqueous species.
- Proceed provisionally without energy of protein folding.
- Calculate equilibrium activities of proteins.
- Investigate temperature, oxidation potential, pH, other chemical potentials.
- Explore the protein universe using model systems.
- CHNOSZ is the software package used for the preceding calculations.


Protein Ionization (CYC_BOVIN)

- Net charges computed additively [Dick et al., 2006] using temperature-dependent sidechain pKa values.
Protein Ionization (CYC_BOVIN)

- Net charges computed additively [Dick et al., 2006] using temperature-dependent sidechain pK_a values.
- Also affects standard Gibbs energies of the ionized proteins.
Eh-pH diagram for proteins

Dashed lines indicate stability limits of H₂O:

\[ \log f_{O_2} = 0 \text{ (upper), } \log f_{O_2} = -83.1 \text{ (lower)} \]
Oxygen Fugacity

- Eh-pH diagram for proteins
- Convert between Eh and log $f_{O_2}$ using law of mass action for
  \[ H_2O \rightleftharpoons \frac{1}{2}O_2 + 2H^+ + 2e^- \]
- \[ \text{Eh} = \frac{RT}{F} \text{pe} = -\frac{RT}{F} \log a_{e^-} \]