Heme Protein Engineering and Evolution

Is a biophysical approach good enough or do I have to become an evolutionary biologist?

John S. Olson  
Department of Biochemistry and Cell Biology & W. M. Keck Center for Computational Biology, Rice University

Rational design of blood substitutes based on 3-dimensional structures and biochemistry (Eric Brucker)

Strategies for Protein Engineering

• 1. Random mutagenesis - Vary amino acids randomly to obtain new combinations and then select or screen for better gene products (like nature - “directed evolution”)

• 2. Comparative design - Examine animal and plant hemoglobins and myoglobins for more optimal properties and unusual active sites (“natural products” and bioinformatics).

• 3. Rational design - Use chemical mechanisms to design new active sites and more stable proteins (use of knowledge and “intelligent design”).
Sequence comparisons of globins from bacteria to man (weak homologies except in key structural (3') regions)

Charles Darwin in 1854

1. Travels on H. M. S. Beagle (1831-1836)
2. Published Zoology of the Voyage of the Beagle in 1840
3. Ideas of natural selection based on Darwin's finches discovered on the Galapagos Islands.
Darwin's *The Origin of Species* was published in 1859
(The original title was: “On the Origin of Species by Means of Natural Selection or The Preservation of Favoured Races in the Struggle for Life.”)

---

Gregor Mendel (1822-1884) - "Father" of Genetics

Chromosomes: They possess the characteristics of Mendelian segregation and inheritance during normal cell division and meiosis (sperm and egg formation).

Chromosomes: roughly equal amounts of protein and nucleic acid

1. Proteins were considered "smart" molecules, contain 20 different amino acids, and are capable of catalysis.

2. Nucleic acids were considered "dumb" structural molecules, contain only four nucleotides (A, T, G, C), and showed no enzymatic activity.
Oswald T. Avery worked on infectious bacteria and viruses at Rockefeller University, NY, including the Spanish flu.

Oswald T. Avery at age 67 proved that DNA is the genetic material:


Non-toxic pneumococcus, type II (small "rough" colonies) can be transformed into the toxic strain (large "smooth" colonies) by DNA isolated from type III bacteria.
The birth of modern biology (1944-1980):

Heredity can be understood at a molecular level without ambiguity, and evolution can be followed mechanistically.

Watson and Crick's Model of Double Helical DNA (1953)

1. X-ray diffraction patterns require a fibrous, helical structure (Franklin and Wilkins)
2. Structure had to explain Chargraff's rules for nucleic acid composition:
   \[ [A] = [T]; [G] = [C] \]
3. Diversity is explained by base sequences.
4. Specific and accurate replication of genetic material is explained by base pairing:
   \[ A\rightarrow T; G\rightarrow C \]

The Central Dogma of Molecular Biology (1950 - 1965)

The key question: what is the genetic code that specifies a protein gene product?

1-dimensional information

3-dimensional information

Nelson & Cox, Lehninger Principles of Biochemistry (4th Edition), Fig. 27-7
Gene Expression

DNA containing gene code or machine language program

Protein manufacturing unit or compiler

Final 3-dimensional structure which is the functional product

Synthetic Mb gene with restriction enzyme sites
(The “classic” 3/3 single domain globin)

Pst I

<table>
<thead>
<tr>
<th>DNA (genes)</th>
<th>mRNA (codons)</th>
<th>Protein (gene product)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA TCT AGA CAT CCA GGT AAC</td>
<td>mRNA U U G U</td>
<td>Phenylalanine Phe or F</td>
</tr>
<tr>
<td>Transcription</td>
<td>Translation</td>
<td></td>
</tr>
</tbody>
</table>

Gene Expression

DNA containing gene code or machine language program

Protein manufacturing unit or compiler

Final 3-dimensional structure which is the functional product

Val68

GTT ACC GTT TGA GCT GTC GGT CTG TGG CAG GGT GAA TGG CAG C

Pst I

1. GCGCTAATACTAAGAGGAGAAGAACACACACATGGTCTGCTCATGAGTTGCTGCTCATGATTGGCTGCT

Bgl II

<table>
<thead>
<tr>
<th>DNA TCT AGA CAT CCA GGT AAC</th>
<th>mRNA U U G U</th>
<th>Phenylalanine Phe or F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription</td>
<td>Translation</td>
<td></td>
</tr>
</tbody>
</table>

Val68

GTT ACC GTT TGA GCT GTC GGT CTG TGG CAG GGT GAA TGG CAG C

Pst I

1. GCGCTAATACTAAGAGGAGAAGAACACACACATGGTCTGCTCATGAGTTGCTGCTCATGATTGGCTGCT

Bgl II

<table>
<thead>
<tr>
<th>DNA TCT AGA CAT CCA GGT AAC</th>
<th>mRNA U U G U</th>
<th>Phenylalanine Phe or F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription</td>
<td>Translation</td>
<td></td>
</tr>
</tbody>
</table>

Val68

GTT ACC GTT TGA GCT GTC GGT CTG TGG CAG GGT GAA TGG CAG C

Pst I

1. GCGCTAATACTAAGAGGAGAAGAACACACACATGGTCTGCTCATGAGTTGCTGCTCATGATTGGCTGCT

Bgl II

<table>
<thead>
<tr>
<th>DNA TCT AGA CAT CCA GGT AAC</th>
<th>mRNA U U G U</th>
<th>Phenylalanine Phe or F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription</td>
<td>Translation</td>
<td></td>
</tr>
</tbody>
</table>
1. Change size
2. Change charge: acid (-) or base (+)
3. Change polarity: polar or apolar (hydrophobic)

<table>
<thead>
<tr>
<th>AA side chain</th>
<th>Molar Volume (cm³/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>34.8</td>
</tr>
<tr>
<td>Ser</td>
<td>51.9</td>
</tr>
<tr>
<td>Ala</td>
<td>52.0</td>
</tr>
<tr>
<td>Cys</td>
<td>65.0</td>
</tr>
<tr>
<td>Asp-</td>
<td>65.4</td>
</tr>
<tr>
<td>Asp0</td>
<td>65.4</td>
</tr>
<tr>
<td>Thr</td>
<td>68.4</td>
</tr>
<tr>
<td>Asn</td>
<td>69.6</td>
</tr>
<tr>
<td>Pro</td>
<td>72.6</td>
</tr>
<tr>
<td>Glu-</td>
<td>77.5</td>
</tr>
<tr>
<td>Glu0</td>
<td>77.5</td>
</tr>
<tr>
<td>Val</td>
<td>82.4</td>
</tr>
<tr>
<td>Gln</td>
<td>85.5</td>
</tr>
<tr>
<td>His+</td>
<td>90.4</td>
</tr>
<tr>
<td>His0</td>
<td>90.4</td>
</tr>
<tr>
<td>Met</td>
<td>96.9</td>
</tr>
<tr>
<td>Ile</td>
<td>99.1</td>
</tr>
<tr>
<td>Leu</td>
<td>99.1</td>
</tr>
<tr>
<td>Lys+</td>
<td>100.1</td>
</tr>
<tr>
<td>Phe</td>
<td>112.8</td>
</tr>
<tr>
<td>Tyr</td>
<td>114.7</td>
</tr>
<tr>
<td>Arg+</td>
<td>118.9</td>
</tr>
<tr>
<td>Trp</td>
<td>135.5</td>
</tr>
</tbody>
</table>

Apolar Amino Acids with different sizes and shapes (inside)

Amino Acids with charges and polarity (outside)
X-ray Crystallography: Visualization of the 3-D structures of proteins and nucleic acids

From crystals to X-ray diffractometer, computer, electron density maps, and molecular models
First three dimensional structures of proteins (1957-1960): hemoglobin (Hb) and myoglobin (Mb)

Nobel Prize Ceremonies, December 10, 1962

Physiology or medicine: Watson, Crick, and Wilkins for double helical DNA
Chemistry: Perutz and Kendrew for the 3-D structures of Proteins (Mb and Hb)
Bioinformatics: Using libraries created by sequencing genes from different species

<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>whelk</td>
<td>AKKLSRNHTA</td>
</tr>
<tr>
<td>aardvark</td>
<td>IQPLAQSHAT</td>
</tr>
<tr>
<td>pig</td>
<td>LTPLAQSHAT</td>
</tr>
<tr>
<td>mouse</td>
<td>IKPLAQSHAT</td>
</tr>
<tr>
<td>sheep</td>
<td>VKHLAESHAN</td>
</tr>
<tr>
<td>bovine</td>
<td>VKHLAESHAN</td>
</tr>
<tr>
<td>horse</td>
<td>LKPLAQSHAT</td>
</tr>
<tr>
<td>elephant</td>
<td>IQPLAQSHAT</td>
</tr>
<tr>
<td>whale</td>
<td>LKPLAQSHAT</td>
</tr>
<tr>
<td>dog</td>
<td>LKPLAQSHAT</td>
</tr>
<tr>
<td>chicken</td>
<td>LKPLAQTHAT</td>
</tr>
<tr>
<td>alligator</td>
<td>LKPLAKSHAL</td>
</tr>
<tr>
<td>tuna</td>
<td>LKPLANSHAT</td>
</tr>
<tr>
<td>shark</td>
<td>VKEAADTHIN</td>
</tr>
<tr>
<td>sea slug</td>
<td>LSQFAKEHVG</td>
</tr>
<tr>
<td>red clam</td>
<td>VKEFAVNHIT</td>
</tr>
<tr>
<td>fly larva</td>
<td>VNTFVASHKP</td>
</tr>
<tr>
<td>roundworm</td>
<td>AREIVDPHLR</td>
</tr>
</tbody>
</table>

- Required histidine (F8) for coordination to the iron atom
- Basic Structure of Single Domain Globins
- Heme group (red color)
- CD corner
His(H) at F8
Leu(L) at F4?
Wild-type Leu89 (F4) is conserved to "water-proof" the bottom or proximal portion of the heme pocket.

Water hydrolyzes the Fe-His(F8) bond causing the heme to dissociate and the protein to denature at 37°C.
Sequence comparisons of globins from bacteria to man

B10, C2, CD1, E7, E11, F4, and F8 are the key conserved amino acids to stabilize the heme group in either the 2/2 or 3/3 fold.
Myoglobin (SDgb) 3/3 AGH/BEF (C and D are small and variable)

Single domain globin (The classical myoglobin fold)

Myoglobin (SDgb) 3/3 AGH/BEF (C and D are small and variable)

Single domain globin (truncated fold)

Bacillus subtilis Hb 2/2 BE/GH (F is variable, A is small)
Chimeric (bi-domain) globin
3/3 fold

Globin domain

E. coli Flavohb (Hmp)

FAD

Heme

Reductase domain

VHb (bacterial) (Vitreoscilla stearcorium)

Myoglobin (chordate)

All are 3/3 Single domain globins

Animal globins (major functions are in O₂ transport and storage).

expression

hemoglobinn
myoglobin
cytoglobin
neuroglobin

<table>
<thead>
<tr>
<th></th>
<th>red blood cell</th>
<th>resident macrophage</th>
<th>neuron nucleus</th>
</tr>
</thead>
<tbody>
<tr>
<td>reoxygenation</td>
<td>26</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(P₅₀O₂ in mm Hg)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gene location</td>
<td>muscle (HbA)</td>
<td>muscle (HbA)</td>
<td>brain (HbA)</td>
</tr>
<tr>
<td></td>
<td>2285 (3)</td>
<td>2285 (3)</td>
<td>145g (5)</td>
</tr>
<tr>
<td>locus link</td>
<td>115677</td>
<td>115677</td>
<td>115677</td>
</tr>
<tr>
<td></td>
<td>168</td>
<td>168</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>0 (B)</td>
<td>0 (B)</td>
<td>0 (B)</td>
</tr>
</tbody>
</table>

Common 3/3 ancestor
Bacillus subtilis

Cyanobacterium

Aquifex aeolicus

Thermophile (3/3 ancestral)

Synechocystis

soil bacterium (2/2)

American Rice plant

Non-symbiotic plant Hb

soybean plant

Leghemoglobin (O₂ scavenging and transport) symbiotic Hb

Thermophile

American Rice plant

Soybean plant

Soybean root nodules

Nitrogen Fixation

10H⁺ + N₂ → 2NH₄⁺ + H₂ (16 ATP)

Leghemoglobin reduces [O₂] to ≤ 1 x 10⁻⁷ M, but still transports O₂ to the bacteria (a muscle myoglobin-like function).
Defense mechanisms against host macrophages and other sources of NO

- Flavohemoglobins (NO dioxygenation)
- Pathogenic fungi
- Soil bacterium (2/2)
- Thermophilic (3/3 ancestral)
- Cyanobacterium (2/2)
- Synechocystis
- Aquifex aeolicus
- Bacillus subtilis
- Aspergillus fumigatus
- Escherichia coli
- Bacillus subtilis
- Thermophile (3/3 ancestral)
- Cyanobacterium (2/2)
- Soil bacterium (2/2)
- Aquifex aeolicus
- Synechocystis
- Bacillus subtilis

**Fox, S. I. Human Physiology (7th Edition), Fig. 15.5**
NO is potentially very toxic

1. NO can inactivate aconitase at levels \( \leq 200 \text{ nM} \); shuts down TCA cycle (Gardner et al., 1997 J. Biol. Chem. 272, 25071-76)

2. NO can inhibit cytochrome oxidases at levels \( \leq 1 \text{ \( \mu \)M} \) (Sarti et al. (M. Brunori), 2000 Biochem. Biophys. Res. Com. 274, 183-7; Tania et al. (R. Poole), 2000, J. Biol. Chem. in press)

3. NO synthesis inhibits respiration in endothelial and smooth muscle cells (Clementi et al. (S. Moncada), 1999, PNAS 96, 1559-1562)

**NO dioxygenase (NOD) activity detoxifies NO (and O2)**

\[
2\text{Hb(FeII)} \text{O}_2 + 2\text{NO} \rightarrow 2\text{NO}_3^- + 2\text{Hb(FeIII)} \text{ (non-toxic)}
\]

FlavoHbs are expressed to detoxify NO, increasing the resistance of pathogenic and symbiotic microorganisms to host defense mechanisms. (Gardner et al., 1998, PNAS 95, 10378-10383; Stamler, Poole, and others)

**FlavoHbs** (Hmp, Yhb, AHb) are expressed in response to NO
NO dioxygenase activity for detoxification of 'NO

Mike Gustin (Rice), Paul Gardner (Children's Hospital, Cincinnati)
- flavoHbs from Candida albicans and Aspergillus fumigatus.

(Gardner et al. 2000 J. Biol. Chem. 275, 12581-9 and 31581-8)
The major animal globins participate in $O_2$ storage and transport and provide lessons for designing rHb-based blood substitutes.