Unifying notions of biochemistry

* All proteins are made of 20 main amino acids.

* All nucleic acids are made of 4 nucleotides. Virtually unlimited combinations are possible.

* Adenosine triphosphate (ATP) is a universal source of free energy.

* The Genetic Code is carried by DNA in all organisms (except some viruses), and mediates the accumulation and transmission of biological information.

* All biomolecules and processes are the result of evolution.

* Biochemical processes and structures are defined by very precise compositions, conformations and interactions. Minor changes can have drastic biological effects. e.g. D vs. L chirality. Cellulose vs. carbohydrate.

* Organisms are compartmentalized by membranes that mediate matter and energy exchange with the environment.

* The biological milieu is aqueous.

* Biological molecules are built of light atoms: C, H, O, N, P, S
  ions Na⁺, K⁺, Ca²⁺, Mg²⁺
  some Fe, Co and Mn
Proteins

There are 20 main amino acids:

\[
\begin{align*}
H_3^+N-\overset{\text{H}}{\overset{\text{R}}{\text{C}}}-\overset{\text{R}}{\text{C}}O^- \\
\end{align*}
\]

*zwitterionic at physiological pH (for non-ionizable R)
* only L-configuration exists in proteins

The peptide bond

\[
\begin{align*}
H_3^+N-\overset{\text{R}}{\text{C}}-\overset{\text{R}}{\text{C}}O^- + H_3^+N-\overset{\text{R}}{\text{C}}-\overset{\text{R}}{\text{C}}O^- & \rightarrow H_3^+N-\overset{\text{R}}{\text{C}}-\overset{\text{R}}{\text{C}}O^- + H_2O \\
\end{align*}
\]

equilibrium strongly in favor of this side

Only L-symmetry exists

Order convention

\[
\begin{align*}
\text{start @ amino} & \rightarrow \text{carboxyl} \\
H_3^+N-\overset{\text{R}}{\text{C}}-\overset{\text{R}}{\text{C}}O^- & \\
\end{align*}
\]
The 20 common amino acids

aliphatic side chains

![Glycine (Gly)](glycine.png)
![Alanine (Ala)](alanine.png)
![Valine (Val)](valine.png)
![Leucine (Leu)](leucine.png)
![Isoleucine (Ile)](isoleucine.png)

aliphatic hydroxyl side chains

![Serine (Ser)](serine.png)
![Threonine (Thr)](threonine.png)

aromatic side chains

![Phenylalanine (Phe)](phenylalanine.png)
![Tyrosine (Tyr)](tyrosine.png)
![Tryptophan (Trp)](tryptophan.png)

basic side chains

![Lysine (Lys)](lysine.png)
![Arginine (Arg)](arginine.png)
![Histidine (His)](histidine.png)

acidic side chains

![Aspartate (Asp)](aspartate.png)
![Glutamate (Glu)](glutamate.png)

sulfur-containing side chains

![Cysteine (Cys)](cysteine.png)
![Methionine (Met)](methionine.png)

amide side chains

![Asparagine (Asn)](asparagine.png)
![Glutamine (Gln)](glutamine.png)

proline

![Proline (Pro)](proline.png)
Protein Functions

* All known enzymes are proteins. Virtually all biochemical reactions are catalyzed by enzymes. Proteins themselves exert feedback control in metabolic integration.
* Transport and storage. Proteins transport many small molecules and ions. e.g. Hemoglobin-oxygen
* Motility, muscle action
  cilia, flagella, muscle; actin/myosin
* Structural material
  bone, tendon, skin; collagen and elastin
* Immune system
  anti-bodies, self-identification
* Generation and transmission of nerve impulses
  reception and transmission across synapses
* Control elements of DNA
  Gene regulation and expression
* Chaperones
  protein folding
Levels of Enzyme Organization

**Primary structure:** The sequence of amino acids specifying the enzyme.

**Secondary Structure:** The regularities of the three dimensional sub-structures.

α-Helix

* α-helix backbone is the peptide backbone: functional groups stick out.

* All the main chain CO and NH groups are H-bonded, staggered by 4 units.

* There is a 1.5Ang. vertical rise and 100° of turn between successive a.a. 3.6 a.a./turn, or 5.4 Ang/turn

* All α-helices are right-handed.

* Two and more helices can wind together to form fibrous bundles; e.g. keratin in hair, myosin in muscle.
Protein secondary structure, cont'd.

β-sheets

* Protein backbone is fully extended, all-trans. 3.5 Ang. between adjacent a.a.

* sheets can be parallel or anti-parallel and are between different polypeptide strands

* Protein strands can make sharp reverses in direction via β-turns
Levels of Protein Structure

**Tertiary structure:** The overall 3-D structure of the protein

(From Biochemistry, by L. Stryer, Freeman Pub.)

**Protein folding problem:** Predict complete secondary, tertiary and quaternary structure from the primary structure.

**Protein chaperones** - Molecules that discriminate between slowly-folding and misfolded proteins, aid in correct folding and localize with aggregates to inhibit their formation.

**Protein aggregate formation** occurs frequently in vitro, is suspected to occur in vivo, and are related to amyloid diseases, such as Alzheimers.
DRUG DISCOVERY

PROMISCUITY DOESN'T PAY

Chemists find mechanism behind false positive drug screening 'hits'

SOME molecules culled from compound libraries look like promising drug leads. They're good at inhibiting a particular disease-causing enzyme. But soon it becomes apparent that these molecules aren't all selective; they inhibit everything under the sun. Chemists call them "promiscuous."

And it's a bitter pill to swallow when you discover that you've invested months, even years, in a promising inhibitor, only to discover that it's biologically useless.

A new study, however, has identified what may be happening to a significant percentage of these compounds—knowledge that could save pharmaceutical companies much fruitless pursuit.

Some of the molecules, it seems, are collecting in large clumps known as aggregates that inhibit enzymes by gumming them up, rather than "docking" monomodois into an enzyme pocket, report Brian K. Shoichet, chemistry professor at Northwestern University, his graduate student Susan L. McGovern, and their colleagues (J. Med. Chem., 45, 1712 (2002)). Fortunately, it's simple enough to test for aggregates right off the bat.

"It's important to understand the molecular basis for false positive activity, and this paper makes an important contribution," says Mark Murcko, chief technology officer at Vertex Pharmaceuticals.

The pharmaceutical research community has already pricked electron microscopy, they indeed saw aggregates 10 to 100 times bigger than the enzyme.

Next, they persuaded the pharmaceutical company Pharmacia to give them 35 compounds from its libraries. Tests soon found that 20 of those compounds formed aggregates.

"People can begin to think more about the problem, now that there's some data," says Gerald M. Maggiora, a senior computational scientist at Pharmacia's lab. "Both experimental and computational screening teams should be aware of this extensive set of false positives," says Irwin D. Tack.

The fact that some compounds are just bad eggs can be a common knowledge among pharmaceutical chemists. But putting a mechanism to the problem has proved difficult. Scientists have had a number of suspicions. For example, it's been thought that some compounds might be acting as covalent inhibitors, chemically reacting with enzymes.

"This work represents an intriguing observation as to why certain hits in high-throughput screening don't make it beyond the screening stage," says Yvonne Martin, a computational chemist at Abbott Laboratories.

Shoichet's group made the discovery while searching for a penicillinase inhibitor. Most of the hits from their screening program were promiscuous.

Several unusual characteristics pointed them toward the problem's source. Most significantly, when they increased the enzyme's concentration, the molecules' ability to inhibit disappeared. The logical explanation, Shoichet says, was aggregates. "Everything clicked when we did that," he says. And with transmission elec-
Protein Denaturation

Loss of 3-D structure due to effects such as pH, Temperature, or denaturing agents such as urea and Guanidine-HCl.

3-D globular protein $\rightarrow$ random coil
biologically active $\rightarrow$ biologically inactive

In many cases re-naturation can proceed spontaneously

Total # of possible protein conformations $N=3^n$, $n=$ # of a.a.
e.g. for $n=100$, $N=3^{100}=5\times10^{47}$ different conformations!

yet the protein primary sequence is normally sufficient to
determine the 3-D folded state:

$I_{to\ fold} = \log_2(3^{100}) = 158$ bits needed to specify the native conformation

In the sequence itself there are $I_{sequence} = \log_2(20^{100}) = 432$ bits
i.e. enough information in the sequence to specify the structure.
Simple enzyme model: Michaelis-Menten Kinetics

\[ E + S \xrightarrow{k_1} E + S \xrightarrow{k_3} E + P \]

Production rate \[ V = \frac{d[P]}{dt} = k_3[ES] \]

@ steady state rate \([ES] = \text{constant}\)

\[ k_1[E][S] = (k_2 + k_3)[ES] \]

Total enzyme conc. \([E]_T = [E] + [ES] \)

Solving for \([ES]\) leads to

\[ V = \frac{k_3[S][E]_T}{k_M + [S]}, \quad k_M = \frac{k_2 + k_3}{k_1}, \quad V_{\max} = k_3[E]_T \]
Nucleic Acids - DNA

Purine Bases

Pyrimidine Bases

Model of a guanine-cytosine base pair.

Skeletal model of double-helical DNA.
The structure repeats at intervals of 34 Å, which corresponds to ten residues on each chain.

Structure of part of a DNA chain.

Diagram of one of the strands of a DNA double helix, viewed down the helix axis. The bases (all pyrimidines here) are inside, whereas the sugar-phosphate backbone is outside. The tenfold symmetry is evident. The bases are shown in blue and the sugars in red.
# The Genetic Code

<table>
<thead>
<tr>
<th>First position (5' end)</th>
<th>Second position</th>
<th>Third position (3' end)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cys</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td></td>
<td>His</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arg</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td></td>
<td>His</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arg</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td></td>
</tr>
</tbody>
</table>
Adenylates

1) Energy Source: Adenosine Triphosphate (ATP)

\[
ATP + H_2O \rightarrow ADP + P_i + H^+ \quad \text{(ATP-asc)}
\]

\[\Delta G \approx -12 \text{ kcal/mole}\]

2) Reducing Power: Nicotinamide Adenine Dinucleotide (NAD\(^+\))

Flavin Adenine Dinucleotide (FADH\(_2\))

3) Activation: Coenzyme A (CoA)