Proteins

- Consist of linear polymers built from series of up to 20 different L-α-amino acids
- Perform a vast array of functions within living organisms
  - catalyzing metabolic reactions, replicating DNA, responding to stimuli, and transporting molecules from one location to another
- Enzyme: Catalyzes chemical reactions, and are highly specific and accelerate only one or a few chemical reactions
  - High turnover number: Product molecules /enzyme molecule/sec
    - ex) Acetylcholinesterase: hydrolyzes acetylcholin to cholin and acetate at ~ $5 \times 10^5$ /sec

- Primary structure: the amino acid sequence
- Secondary structure: regularly repeating local structures stabilized by hydrogen bonds. Alpha helix, beta sheet and turns
- Tertiary structure: the overall shape of a single protein molecule
- Quaternary structure: the structure formed by several subunits
Protein structures

Figure 2

LEVELS OF PROTEIN STRUCTURE
Primary Structure

Secondary Structure

β-Sheet
α-Helix

Tertiary Structure

Quaternary Structure

Particle Sciences

bovine rhodopsin
human telomere protein
leucine rich repeat protein
nucleoside transporter
mouse cadherin

β/α-barrel form (TIM)
α/β superhelix (ribonuclease inhibitor)
5-propeller form (lectin)
solenoid form (transferase)
3D structure of the protein myoglobin showing turquoise alpha helices. This protein was the first to have its structure solved by X-ray crystallography. Towards the right-center among the coils, a prosthetic group called a heme group (shown in gray) with a bound oxygen molecule (red).
Method to develop more useful or valuable proteins

- Alteration of a single amino acid residues at specific site
- Insertion or deletion of a single amino acid residue
- Alteration or deletion of a segment or an entire domain
- Generation of a novel fusion protein
- Incorporation of unnatural amino acids at specific site
Why Protein Engineering?

- Proteins/Enzymes: Evolved for host itself, not for human
  - Most proficient catalysts with high specificity

- Need further improvement for practical use:
  - Substrate specificity
  - Binding affinity
  - Stability
  - Catalytic activity
  - Folding/Expression level etc..

- Goal of protein engineering:
  - Design of protein/enzyme with desired function and property for practical applications

Designer proteins/Enzymes

Therapeutic proteins, Industrial enzymes
Technology Development

Random approach
- Screening from nature
- Random mutations

Structure-based rational approach
- Structure-function relationship
- Site-directed/saturation mutagenesis

Evolutionary approach
- Directed evolution
  - Accumulation of beneficial mutations
  - No structural data
  - HTS system
  - Construction of diverse library

Computational (in silico) method
- Virtual screening of large sequence space
- Large structural data: >~30,000
- High computing power
- Mechanistic knowledge

Combinatorial approach
- Structure-based design
- Evolutionary method
- Computational method
New version of therapeutic proteins

• EPO with longer plasma half-life by incorporation of N-glycosylation

• Faster-acting insulin by modification of amino acid sequence

• Slow-acting insulin

• Faster-acting tissue plasminogen activator (t-PA) by removal of three of the five native domains → higher clot-degrading activity

• Ontak: A fusion protein consisting of the diphtheria toxin linked to IL-2
  → Selectively kill cells expressing an IL-2 receptor
  → Approved for the treatment of cutaneous T cell lymphoma in 1999 in US
One of the most important post-translational modifications (PTMs): N-glycosylation / O-glycosylation in Mammalian / Yeast

Essential roles in *in vivo*: Biological activity, folding, solubility, protease resistance, immunogenicity, signal transduction, and pharmacokinetics

Carbohydrates on cell surface: Cell signaling, cell attachment, cell adhesion, recognition, and inflammation

ex) Tamiflu: Neuramidase inhibitor, an antiviral drug to prevent the spread of influenza A and B (flu) virus between cells in the body by stopping the virus from cutting ties with its host cell

About 60% of therapeutic proteins are glycoprotein
- Therapeutic proteins: 140 approved
- EPO
Glycan profile: very complex and varies broadly, depending on cell type and production conditions: Glycan moiety, occupation number, length of glycosylation chain.

Therapeutic proteins require proper glycosylation for biological efficacy.

Analysis of glycan profile, its role/function in vivo, glycosylation pathway, and property of glycoproteins are a key to Glycobiology.
Monosaccharide structures

Fucose, Fuc  
\( \text{C}_6\text{O}_5\text{H}_{12} \)

Galactose, Gal  
\( \text{C}_6\text{O}_6\text{H}_{12} \)

Mannose, Man  
\( \text{C}_6\text{O}_6\text{H}_{12} \)

Sialic Acid, NAcNeuA  
\( \text{C}_{11}\text{O}_9\text{NH}_{19} \)

N-Acetylgalactosamine, GalNAc  
\( \text{C}_8\text{O}_6\text{NH}_{15} \)

N-Acetylglucosamine, GlcNAc  
\( \text{C}_8\text{O}_6\text{NH}_{15} \)
## N-Linked glycan structures

<table>
<thead>
<tr>
<th>N-Linked Structures</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>complex sialylated fucosylated diantennary</td>
<td>$C_{90}O_{65}N_6H_{146}$</td>
</tr>
<tr>
<td>complex sialylated fucosylated triantennary</td>
<td>$C_{115}O_{83}N_8H_{186}$</td>
</tr>
<tr>
<td>complex sialylated fucosylated tetraantennary</td>
<td>$C_{140}O_{101}N_{10}H_{226}$</td>
</tr>
<tr>
<td>high mannose (man 7)</td>
<td>$C_{64}O_{49}N_2H_{106}$</td>
</tr>
<tr>
<td>hybrid</td>
<td>$C_{83}O_{62}N_2H_{136}$</td>
</tr>
<tr>
<td>common core str. subunit</td>
<td>$C_{34}O_{25}N_2H_{56}$</td>
</tr>
<tr>
<td>sialylated branch subunit</td>
<td>$C_{25}O_{18}N_2H_{40}$</td>
</tr>
<tr>
<td>branch subunit</td>
<td>$C_{14}O_{10}N_1H_{23}$</td>
</tr>
</tbody>
</table>

### Residue Composition

<table>
<thead>
<tr>
<th>Residue</th>
<th>Symbol</th>
<th>Residue Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sialic Acid</td>
<td>□</td>
<td>$C_{11}O_8N_7H_{17}$</td>
</tr>
<tr>
<td>Galactose</td>
<td>○</td>
<td>$C_6O_3H_{10}$</td>
</tr>
<tr>
<td>N-Acetylglucosamine</td>
<td>□</td>
<td>$C_8O_5N_3H_{13}$</td>
</tr>
<tr>
<td>Mannose</td>
<td>○</td>
<td>$C_6O_3H_{10}$</td>
</tr>
<tr>
<td>Fucose</td>
<td>△</td>
<td>$C_6O_4H_{10}$</td>
</tr>
</tbody>
</table>
Erythropoietin (EPO)

- **Glycoprotein : Growth factor** (166 amino acids, MW 36 kDa) produced in kidney → Promote the formation of red blood cells (erythrocytes) in the bone marrow

- **Carbohydrate moiety :** *in vivo* activity, stability, solubility, cellular processing and secretion, immunogenicity

- Three N-glycosylation sites and one O-glycosylation site

- About 50 % of EPO’ secondary structure : α-Helix

- Carbohydrate content : ~ 40 %
Glycosylation pattern of EPO
History of the EPO development

• **1971**: First purified from the plasma of anaemic sheep

• **1985**: Produced by recombinant DNA technology

• **1989**: Approved by FDA for treatment of anemia resulting from chronic kidney disease and cancer treatment (chemotherapy and radiation)

• **Total sales**: $11 billion (2005)

• **Major EPO brands**: generic form
  - Epogen by Amgen ($2.5 billion)
  - Procrit by Ortho Biotech ($3.5 billion)
  - Neorecormon by Boehringer-Mannheim ($1.5 billion)
As the patent becomes expired, Amgen wanted to prolong the market share by developing a new version of EPO by protein engineering.

- **Aranesp**: Introduction of two additional N-glycosylation sites
  - Which site of EPO?
    - A prolonged serum half-life from 4-6 up to 21 hrs
    - What benefit to patients?

- Launched in 2001 → Current sale: $3.5 billion
Atorvastatin (Lipitor)

• Lipitor : Brand name by Pfizer

• A competitive inhibitor of HMG-CoA reductase (3-hydroxy-3-methylglutaryl CoA reductase)
  - HMG-CoA reductase catalyzes the reduction of 3-hydroxy-3-methylglutaryl-coenzymeA (HMG-CoA) to mevalonate, which is the rate-limiting step in hepatic cholesterol biosynthesis.

• Inhibition of the enzyme decreases de novo cholesterol synthesis, increasing expression of low-density lipoprotein receptors (L receptors) on hepatocytes.
  → Increase in LDL uptake by the hepatocytes, decreasing the amount of LDL-cholesterol in the blood.

• Like other statins, atorvastatin also reduces blood levels of triglycerides and slightly increases the levels of HDL-cholesterol.
• The largest selling drug in the world: $12.9 billion in 2006
• Generic drug: Simvastatin by Merck

Lipitor
Doctors and patients began switching to a cheaper generic alternative drug called simvastatin from Merck.

Pfizer launched a campaign including advertisements, lobbying efforts, and a paid speaking tour by Dr. Louis W. Sullivan, a former secretary of the federal Depart. of Health and Human Services, to discourage the trend.

Studies show that at commonly prescribed doses Lipitor and simvastatin are equally effective at reducing LDL cholesterol.

Pfizer has begun promoting a study, conducted by Pfizer’s own researchers, concluding switching increased the rate of heart attacks among British patients.
Economic process using an enzyme with higher efficiency

- Halohydrin dehalogenase (HHDH)
  - Catalyze the nucleophilic displacement of a halogen by a vicinal hydroxyl group in halohydrins, yielding an epoxide
  - Interconverts halohydrins and epoxides

- Manufacture of ethyl (R)-4-cyano-3-hydroxybutyrate (HN) from ethyl ethyl (S)-4-chloro-3-hydroxybutyrate (ECHB)
  - NH: Starting material for the production of the cholesterol-lowering drug: Atorvastatin (Lipitor)

- The specifications for the chemical and enantio-purity of HN are tightly controlled. The hydroxyl in HN is defined the second stereo-center in atorvastatin (Lipitor), and high chemical purity is essential for downstream chemistry

**Essential for more economic process**
Halohydrin and Epoxide

- A type of organic compound or functional group in which one carbon atom has a halogen substituent, and an adjacent carbon atom has a hydroxyl substituent.

- An epoxide is a cyclic ether with three ring atoms

General structure of a halohydrin, where $X = I, \text{Br}, F, \text{or Cl}$
Figure 1  HHDH interconverts halohydrins and epoxides$^{18}$ and can accept alternative nucleophiles$^{17}$. Chemical cyanation is typically performed at elevated temperatures (80 °C) and pH 9. HHDH catalyzes the single-vessel enzymatic conversion of ethyl (S)-4-chloro-3-hydroxybutyrate (ECHB) to ethyl (R)-4-cyano-3-hydroxybutyrate (HN) in the presence of cyanide at pH 7 where no chemical reaction is observed. The HCl produced in the reaction causes the pH to drop, although both the pH and the cyanide concentration can be maintained by running the reaction in a pHstat with NaCN as the base.
Design criteria

- Enzyme source
  - Expression of HHDH from *Agrobacterium radiobacter* in *E. coli*
  - Volumetric productivity: $6 \times 10^{-3}$ g product/L/hour/gram of biocatalyst

- Requirement for implementing the enzyme process at commercial scale
  - Yield: Complete conversion (100%) of at least 100 g per liter substrate
  - Volumetric productivity: $> 20$ g product/liter/hour/gram of biocatalyst

Figure 2  Multivariate optimization of enzymes. At any point, ~50 mutations (variables) are evaluated in the combinatorial libraries (in the hopper). The best variants for the desired activity, and a fraction of less improved variants, are sequenced and analyzed as described. After ProSAR analysis, individual mutations are parsed into four classes: 'Beneficial', which are fixed into the population by retention in the next round parental enzyme, 'Potentially beneficial', which are sent back into the hopper for retesting, 'Deleterious', which are discarded and 'Neutral', which have little or no effect on protein function and are discarded. The amount of the diversity under investigation is maintained by addition of additional diversity discovered, for example, through rational design, homologous sequences, saturation or PCR mutagenic libraries or other evolution programs.
Agrobacterium radiobacter halohydrin dehalogenase with its substrate (white). Integration of computational analysis with experimental screening to identify 37 mutations (yellow) that increase enzymatic activity ~4,000-fold.
Increase in the volumetric productivity: ~4,000 fold

Development of a more economic process
Structural and functional features of Immunoglobulin Ab

Diagram showing the structure of an antibody with labeled regions such as Fv, Fab, and Fc, and highlighting the variable and constant regions.
• Reduced immunogenicity: Humanization, Human Ab
• Improved affinity: Engineering of variable domains (< nM)
• Antibody fragment: Fab, single-chain Fv (scFv), minibody, diabody
• Novel effector function: Conjugation with radioisotope, cytotoxic drug
• Improved effector function: Fc engineering
• Longer half-life: Fc engineering (FcRn binding site)
• Bi-specific antibody
• In the placenta, FcRn facilitates transport of mother's IgG to the growing fetus.
• Binds IgG at acidic pH of (<6.5), but not at neutral or higher pH.
• FcRn can bind IgG from the intestinal lumen (the inside of the gut) at a slightly acidic pH: Efficient unidirectional transport to the basolateral side (inside the body) where the pH is neutral to basic (pH 7.0–7.5).
• FcRn also plays a role in adult salvage of IgG through its occurrence in the pathway of endocytosis in endothelial cells.
• FcRn in the acidic endosomes bind to IgG internalized through pinocytosis, recycling it to the cell surface, releasing it at the basic pH of blood, thereby preventing it from undergoing lysosomal degradation.
• This mechanism may provide an explanation for the greater half-life of IgG in the blood compared to other isotypes.
Endocytosis

• Process in which the cell takes in macromolecules and particulate matter by forming new vesicle from the plasma membrane.
• Proteins involved in the processes are different, and the events of endocytosis look like the reverse of exocytosis.
• A small area of the plasma membrane sinks inward to form a pocket. As the pocket deepens, it pinches in, forming a vesicle containing material that had been outside the cell.
• **Receptor-mediated endocytosis:** Extracellular substances (ligands) bind to these receptors. When binding occurs, the coated pit forms a vesicle containing the ligand molecules. After this ingested material is liberated from the vesicle, the receptors are recycle to the plasma membrane by the same vesicle.
• Phagocytosis: A cell engulfs a particle by wrapping pseudopodia around it and packaging it within a membrane-enclosed sac large enough to be classified as a vacuole. The particle is digested after the vacuole fuses with a lysosome containing hydrolytic enzymes.
• **Pinocytosis**: A mode of endocytosis in which small particles are brought into the cell, forming an invagination, and then suspended within small vesicles.
- These pinocytotic vesicles subsequently fuse with lysosomes to hydrolyze (break down) the particles.
SMART-IgG Technology

Novel antibody recycling technology by engineering endosomal antigen dissociation

- Conventional antibody against membrane-bound receptor: Receptor-bound antibody $\rightarrow$ Receptor-mediated endocytosis $\rightarrow$ Degradation in lysosome $\rightarrow$ Receptor-mediated antibody clearance $\rightarrow$ Low efficacy and high dose of antibody

- Recycling antibody: Engineering antibody to dissociate the antigen in acidic endosome (pH-dependent antigen binding).
  $\rightarrow$ Free antibody binds FcRn and will be recycled back to the cell surface
  $\rightarrow$ enables a single antibody to bind the antigen multiple times by pH-dependent antigen dissociation within acidic endosome.

• **Tocilizumab:**
  - Monoclonal antibody for human IL-6 receptor
  - Clinically used for treating rheumatoid arthritis
pH-dependent IL-6 R binding

- Rational design of tocilizumab showing pH-dependent IL-6R binding by protein engineering
  - Tyrosine in CDR of ticilizumab interacting with IL-6 R was rationally mutated to histidine to induce repulsion with IL-6 R at acidic pH
  - pKa of His: ~ pH 6.0
    - At pH 7.4, His interacts with Arg
    - At pH 5-6, electrostatic repulsion with cationic Arg
- About 20-fold lower affinity at pH 6.0, while no change at pH 7.4
- Engineering to increase the binding affinity for FcRn

\[
\text{His} \quad \text{Histidine} \quad \text{pK}_a \approx 6.5
\]
Engineering of Ab for humanization
Antibody Fragments

- Better tissue penetration, fast clearance, no effector function
- Local delivery
- Production in bacterial system
Lucentis

• A monoclonal antibody fragment (Fab) derived from the same parent mouse antibody as Avastin

• Much smaller than the parent molecule and has been affinity matured to provide stronger binding to VEGF-A

• An anti-angiogenic protein to treat the "wet" type of age-related macular degeneration (AMD, also ARMD), Common form of an age-related vision loss.

• Cost $1,593 per dose, compared to Avastin that cost $42.

• Developed by Genentech and is marketed in the United States by Genentech and elsewhere by Novartis
<table>
<thead>
<tr>
<th>Approved Year</th>
<th>Product</th>
<th>Target</th>
<th>Indication</th>
<th>Company</th>
<th>Market size(07)*</th>
<th>Antibody Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>Rituxan</td>
<td>CD20</td>
<td>Non-Hodgkin’s lymphoma</td>
<td>Genentech</td>
<td>4,603</td>
<td>Chimeric</td>
</tr>
<tr>
<td>1998</td>
<td>Herceptin</td>
<td>Her2/neu</td>
<td>Breast cancer</td>
<td>Genentech</td>
<td>4,047</td>
<td>Humanized</td>
</tr>
<tr>
<td>1998</td>
<td>Synagis</td>
<td>RSV</td>
<td>RSV prophylaxis</td>
<td>MedImmune</td>
<td>1,100</td>
<td>Humanized</td>
</tr>
<tr>
<td>1998</td>
<td>Remicade</td>
<td>TNF-α</td>
<td>RA, Chron’s disease</td>
<td>J&amp;J</td>
<td>5,234</td>
<td>Chimeric</td>
</tr>
<tr>
<td>2002</td>
<td>Humira</td>
<td>TNF-α</td>
<td>RA</td>
<td>Abbott</td>
<td>3,064</td>
<td>Human</td>
</tr>
<tr>
<td>2003</td>
<td>Raptiva</td>
<td>CD11a</td>
<td>Psoriasis</td>
<td>Genentech</td>
<td>211</td>
<td>Humanized</td>
</tr>
<tr>
<td>2004</td>
<td>Erbitux</td>
<td>EGFR</td>
<td>Colorectal cancer</td>
<td>Imclone</td>
<td>1,336</td>
<td>Chimeric</td>
</tr>
<tr>
<td>2004</td>
<td>Avastin</td>
<td>VEGF</td>
<td>Colorectal cancer</td>
<td>Genentech</td>
<td>3,335</td>
<td>Humanized</td>
</tr>
<tr>
<td>2006</td>
<td>Vectibix</td>
<td>EGFR</td>
<td>Colorectal cancer</td>
<td>Amgen</td>
<td>170</td>
<td>Human</td>
</tr>
</tbody>
</table>

*Data Monitor ‘Monoclonal 2008*