Important Points in Drug Design based on Bioinformatics Tools

http://www.geocities.com/bioinformaticsweb/drugdiscovery.html

History of Drug/Vaccine development

- **Plants or Natural Product**
  - Plant and Natural products were source for medical substance
  - Example: foxglove used to treat congestive heart failure
  - Foxglove contain digitalis and cardiotonic glycoside
  - Identification of active component
- **Accidental Observations**
  - Penicillin is one good example
  - Alexander Fleming observed the effect of mold
  - Mold(Penicillium) produce substance penicillin
  - Discovery of penicillin lead to large scale screening
  - Soil micoorganism were grown and tested
  - Streptomycin, neomycin, gentamicin, tetracyclines etc.

Important Points in Drug Design based on Bioinformatics Tools

- **Chemical Modification of Known Drugs**
  - Drug improvement by chemical modification
  - Pencillin G -> Methicillin; morphine->nalorphine
- **Receptor Based drug design**
  - Receptor is the target (usually a protein)
  - Drug molecule binds to cause biological effects
  - It is also called lock and key system
  - Structure determination of receptor is important
- **Ligand-based drug design**
  - Search a lead ocompound or active ligand
  - Structure of ligand guide the drug design process

Important Points in Drug Design based on Bioinformatics Tools

- **Identify Target Disease**
  - Identify and study the lead compounds
  - Marginally useful and may have severe side effects
- **Refinement of the chemical structures**
  - Detect the Molecular Bases for Disease
  - Detection of drug binding site
  - Tailor drug to bind at that site
  - Protein modeling techniques
  - Traditional Method (brute force testing)

Genetics Review

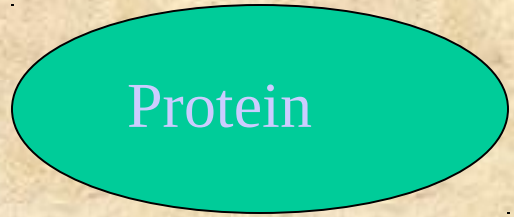
DNA: TACGCTTCCGGATTCAA

transcription

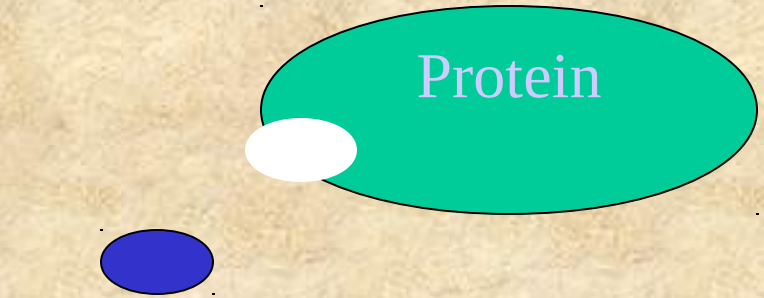
RNA: AUGCGAAGGCCUAAGUU

translation

Amino Acids: PIRLMQTS

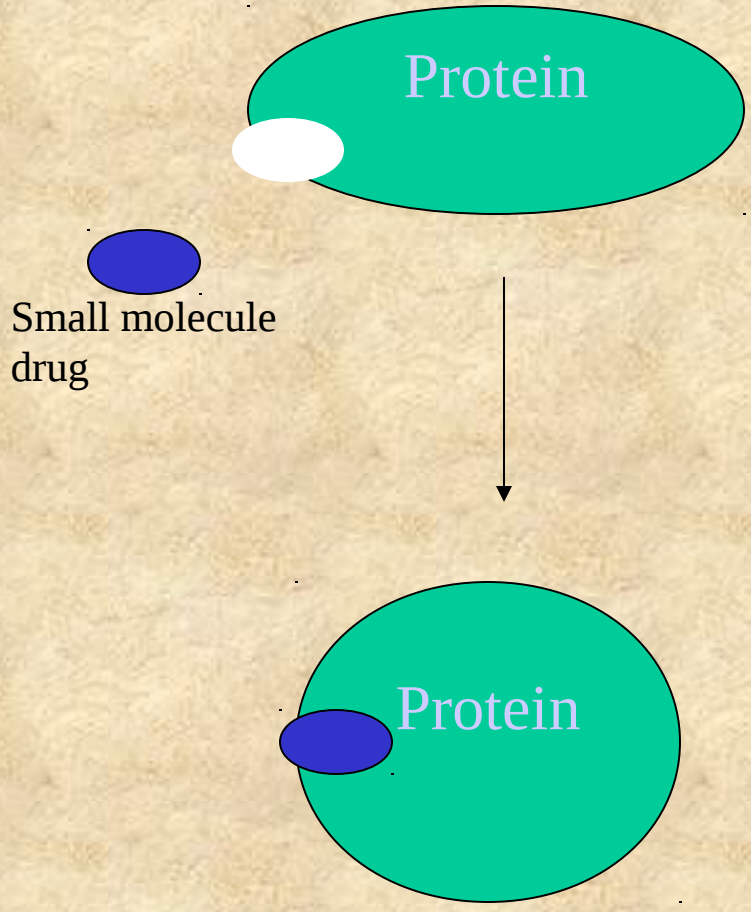


Overview Continued – A simple example



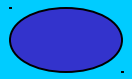
Small molecule drug

Overview Continued – A simple example



Protein disabled … disease cured

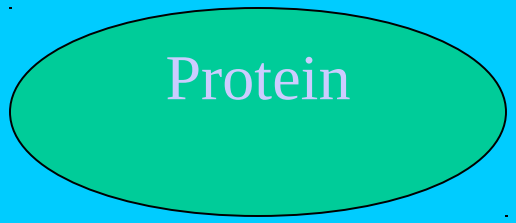
Chemoinformatics



Small molecule drug

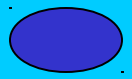
- Large databases

Bioinformatics



- Large databases

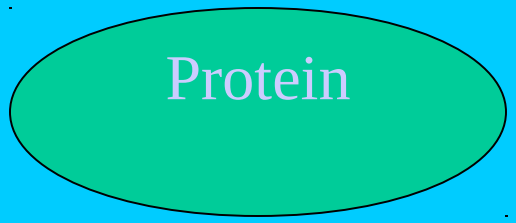
Chemoinformatics



Small molecule drug

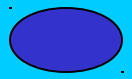
- Large databases
- Not all can be drugs

Bioinformatics



- Large databases
- Not all can be drug targets

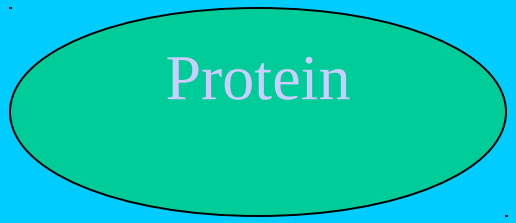
Chemoinformatics



Small molecule drug

- Large databases
- Not all can be drugs
- Opportunity for data mining techniques

Bioinformatics



- Large databases
- Not all can be drug targets
- Opportunity for data mining techniques

Important Points in Drug Design based on Bioinformatics Tools

- Application of Genome
  - 3 billion bases pair
  - 30,000 unique genes
  - Any gene may be a potential drug target
  - ~500 unique target
  - Their may be 10 to 100 variants at each target gene
  - 1.4 million SNP
  - 10^{200} potential small molecules

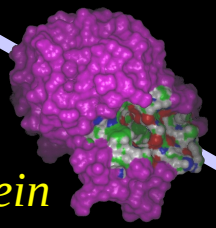
Important Points in Drug Design based on Bioinformatics Tools

- **Detect the Molecular Bases for Disease**
  - Detection of drug binding site
  - Tailor drug to bind at that site
  - Protein modeling techniques
  - Traditional Method (brute force testing)
- **Rational drug design techniques**
  - Screen likely compounds built
  - Modeling large number of compounds (automated)
  - Application of Artificial intelligence
  - Limitation of known structures

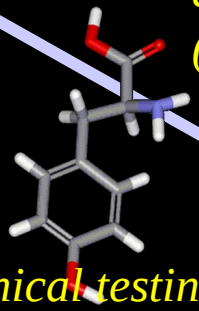
Important Points in Drug Design based on Bioinformatics Tools

- **Refinement of compounds**
  - Refine lead compounds using laboratory techniques
  - Greater drug activity and fewer side effects
  - Compute change required to design better drug
- **Quantitative Structure Activity Relationships (QSAR)**
  - Compute functional group in compound
  - QSAR compute every possible number
  - Enormous curve fitting to identify drug activity
  - chemical modifications for synthesis and testing.
- **Solubility of Molecule**
- **Drug Testing**

Identify disease



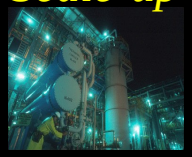
Isolate protein involved in disease (2-5 years)



Find a drug effective against disease protein (2-5 years)

Preclinical testing (1-3 years)

Scale-up



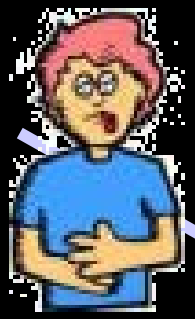
File IND

Human clinical trials (2-10 years)

Formulation

File NDA

FDA approval (2-3 years)



Identify disease

GENOMICS, PROTEOMICS & BIOPHARM.

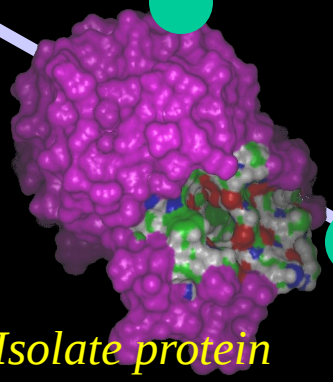
Potentially producing many more targets and "personalized" targets

HIGH THROUGHPUT SCREENING

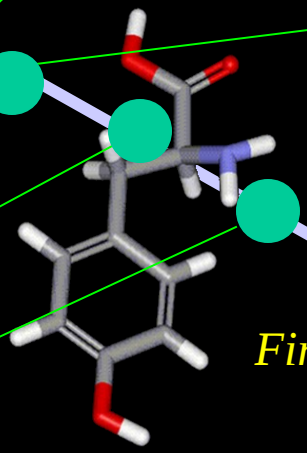
Screening up to 100,000 compounds a day for activity against a target protein

VIRTUAL SCREENING

Using a computer to predict activity



Isolate protein



Find drug

COMBINATORIAL CHEMISTRY

Rapidly producing vast numbers of compounds

MOLECULAR MODELING

Computer graphics & models help improve activity

Preclinical testing

IN VITRO & IN SILICO ADME MODELS

Tissue and computer models begin to replace animal testing

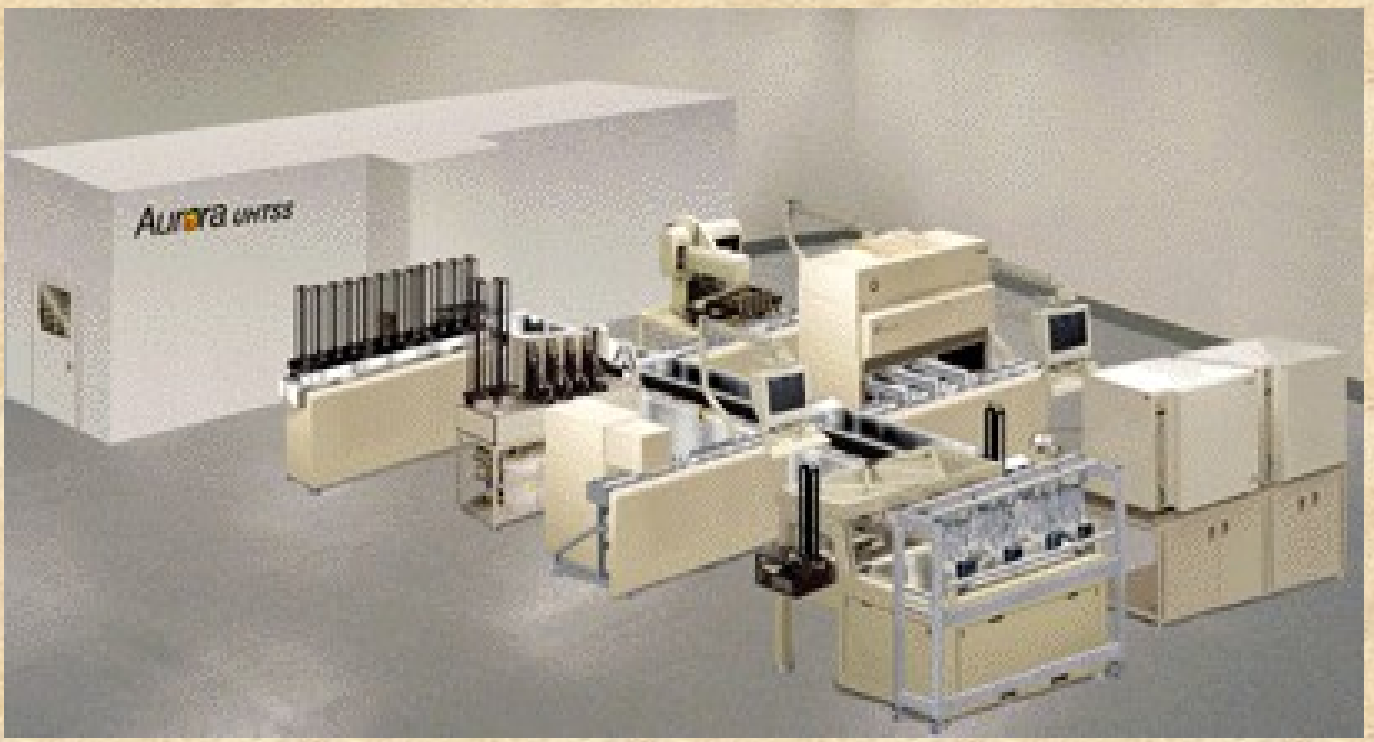
1. Gene Chips

- “Gene chips” allow us to look for changes in protein expression for different people with a variety of conditions, and to see if the presence of drugs changes that expression
- Makes possible the design of drugs to target different phenotypes

Biopharmaceuticals

- Drugs based on proteins, peptides or natural products instead of small molecules (chemistry)
- Pioneered by biotechnology companies
- Biopharmaceuticals can be quicker to discover than traditional small-molecule therapies
- Biotechs now paring up with major pharmaceutical companies

2. High-Throughput Screening



Screening perhaps millions of compounds in a corporate collection to see if any show activity against a certain disease protein

High-Throughput Screening

- Drug companies now have millions of samples of chemical compounds
- High-throughput screening can test 100,000 compounds a day for activity against a protein target
- Maybe tens of thousands of these compounds will show some activity for the protei
- The chemist needs to intelligently select the 2 - 3 classes of compounds that show the most promise for being drugs to follow-up

Informatics Implications

- Need to be able to store chemical structure and biological data for millions of datapoints
  - *Computational representation of 2D structure*
- Need to be able to organize thousands of active compounds into meaningful groups
  - *Group similar structures together and relate to activity*
- Need to learn as much information as possible from the data (data mining)
  - *Apply statistical methods to the structures and related information*

3. Computational Models of Activity

- Machine Learning Methods
  - E.g. Neural nets, Bayesian nets, SVMs, Kahonen nets
  - Train with compounds of known activity
  - Predict activity of “unknown” compounds
- Scoring methods
  - Profile compounds based on properties related to target
- Fast Docking
  - Rapidly “dock” 3D representations of molecules into 3D representations of proteins, and score according to how well they bind

4. Combinatorial Chemistry

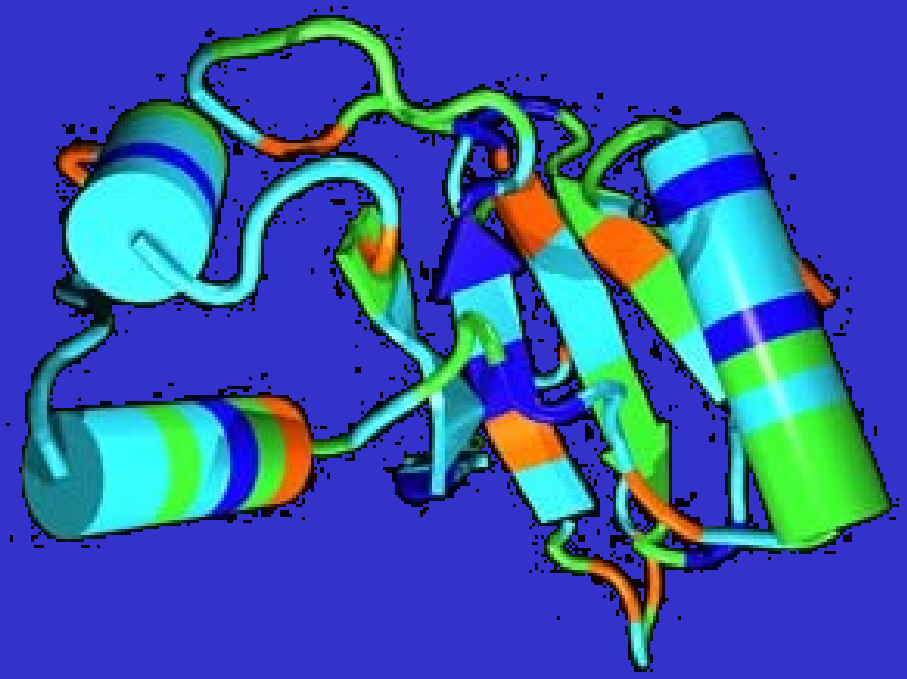
- By combining molecular “building blocks”, we can create very large numbers of different molecules very quickly.

- Usually involves a “scaffold” molecule, and sets of compounds which can be reacted with the scaffold to place different structures on “attachment points”.

Combinatorial Chemistry Issues

- Which R-groups to choose
- Which libraries to make
  - "Fill out" existing compound collection?
  - Targeted to a particular protein?
  - As many compounds as possible?
- Computational profiling of libraries can help
  - "Virtual libraries" can be assessed on computer

5. Molecular Modeling

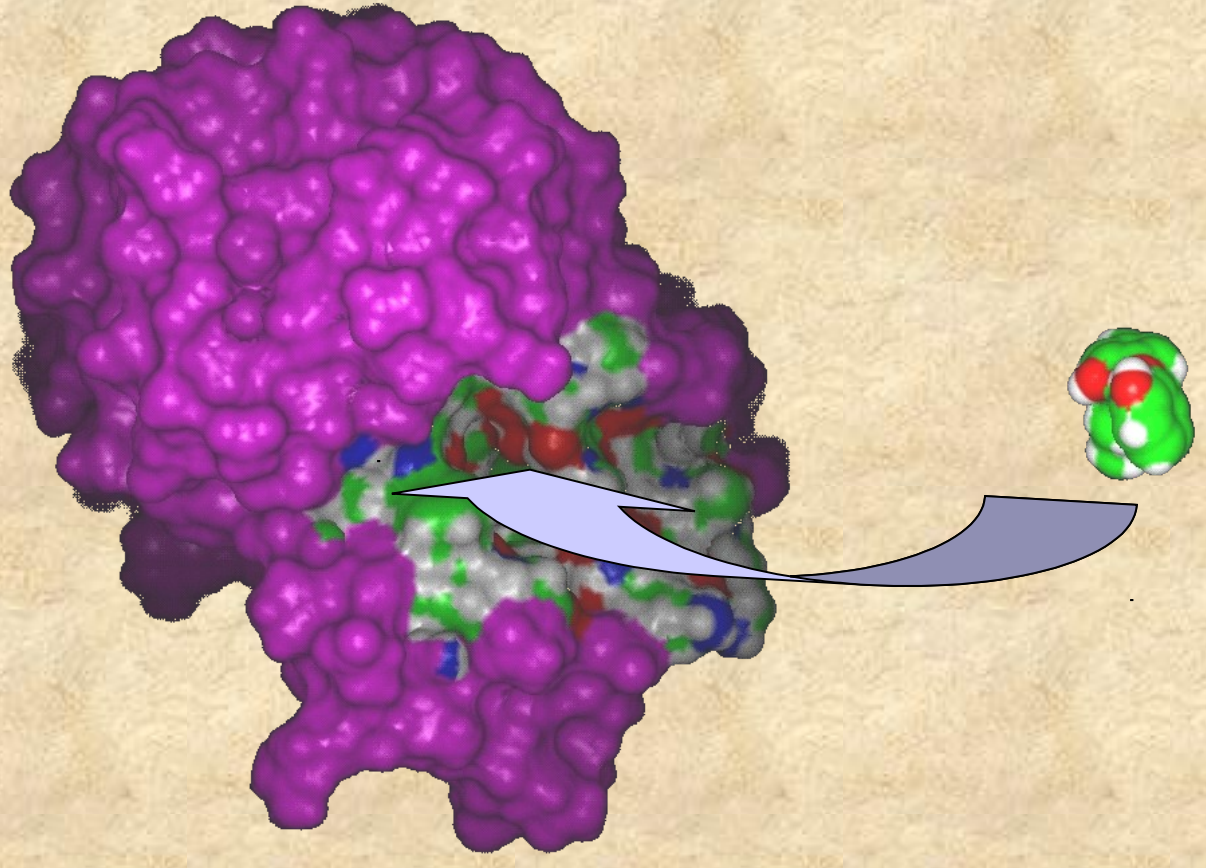


- *3D Visualization of interactions between compounds and proteins*
- *“Docking” compounds into proteins computationally*

3D Visualization

- X-ray crystallography and NMR Spectroscopy can reveal 3D structure of protein and bound compounds
- Visualization of these “complexes” of proteins and potential drugs can help scientists understand the mechanism of action of the drug and to improve the design of a drug
- Visualization uses computational “ball and stick” model of atoms and bonds, as well as surfaces
- Stereoscopic visualization available

“Docking” compounds into proteins computationally



6. In Vitro & In Silico ADME models

- Traditionally, animals were used for pre-human testing. However, animal tests are expensive, time consuming and ethically undesirable

- ADME (Absorbtion, Distribution, Metabolism, Excretion) techniques help model how the drug will likely act in the body

- These methods can be experemental (*in vitro*) using cellular tissue, or *in silico*, using computational models

In Silico ADME Models

- Computational methods can predict compound properties important to ADME, e.g.
  - LogP, a liphophilicity measure
  - Solubility
  - Permeability
  - Cytochrome p450 metabolism
- Means estimates can be made for millions of compouds, helping reduce “atrittion” – the failure rate of compounds in late stage

Size of databases

- Millions of entries in databases
  - CAS : 23 million
  - GeneBank : 5 million
- Total number of drugs worldwide: 60,000
- Fewer than 500 characterized molecular targets
- Potential targets : 5,000-10,000