CAMD and GAs in Rational Drug Design

Christopher Michael Lorenz

BioChemistry 118Q

June 4, 2001

What is CAMD?

CAMD stands for Computer Aided Molecular Design

The design of new molecules based on desired properties

Focused on modeling drugs and biological receptors the drugs bind to so that better binding, and more potent drugs can be developed

Why use CAMD?

Avoids tedious lab work by using computers to model molecules and their properties

Computers can develop new structures and determine whether those structures could serve a specific purpose much faster than humans

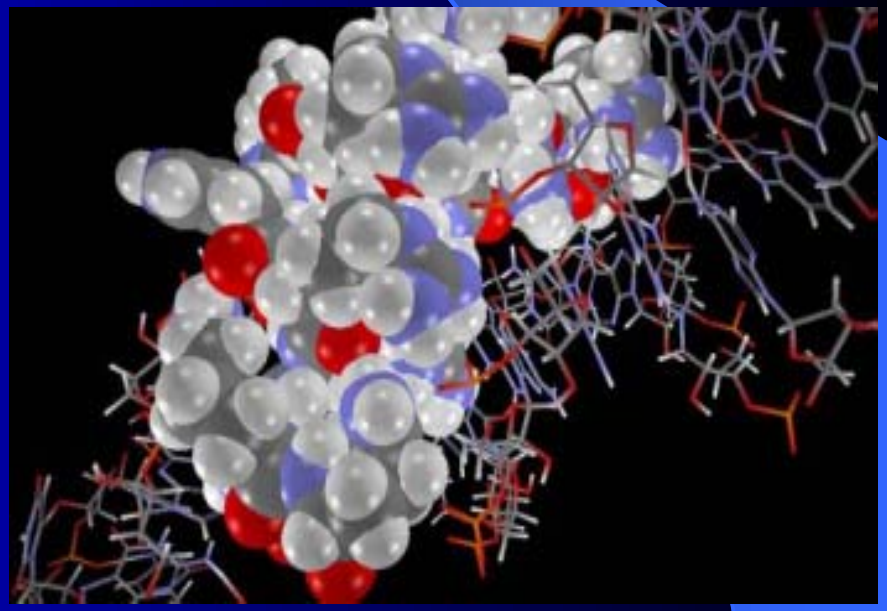
Computers can imitate the millions of years of random variation and natural selection that specialized the molecular structures that gave rise to compounds such as morphine, penicillin, digitalis, and tamoxifen

Allows discoveries of a random, almost ‘accidental’ nature

CAMD Success:

Development of an HIV protease inhibitor by Dupont Merck

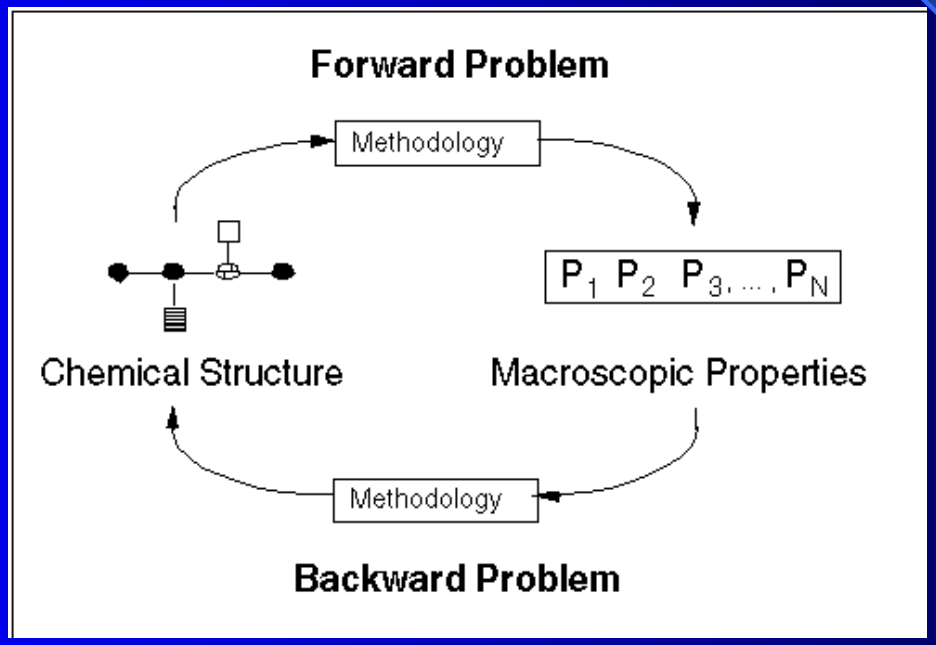
Developed completely on a computer by studying the molecular properties favorable to such an inhibitor, and then designing a molecule to meet the necessary requirements



HIV REV bound to RNA.

CAMD addresses two problems:

Forward: the computation of macroscopic properties given the molecular structure



Backward: identification of the appropriate molecular structure given the desired properties

How does CAMD address these problems?

Genetic Algorithms (GAs)

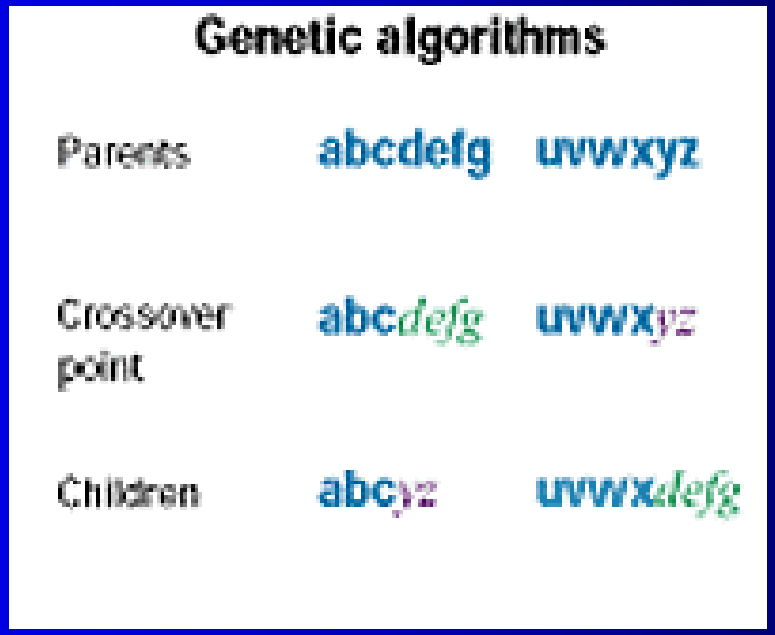
What are genetic algorithms?

GAs are computer programs that apply optimization methods of evolution (mutation, crossover, replication, etc.) to generations of populations of computer code "chromosomes”

Genetic Algorithms

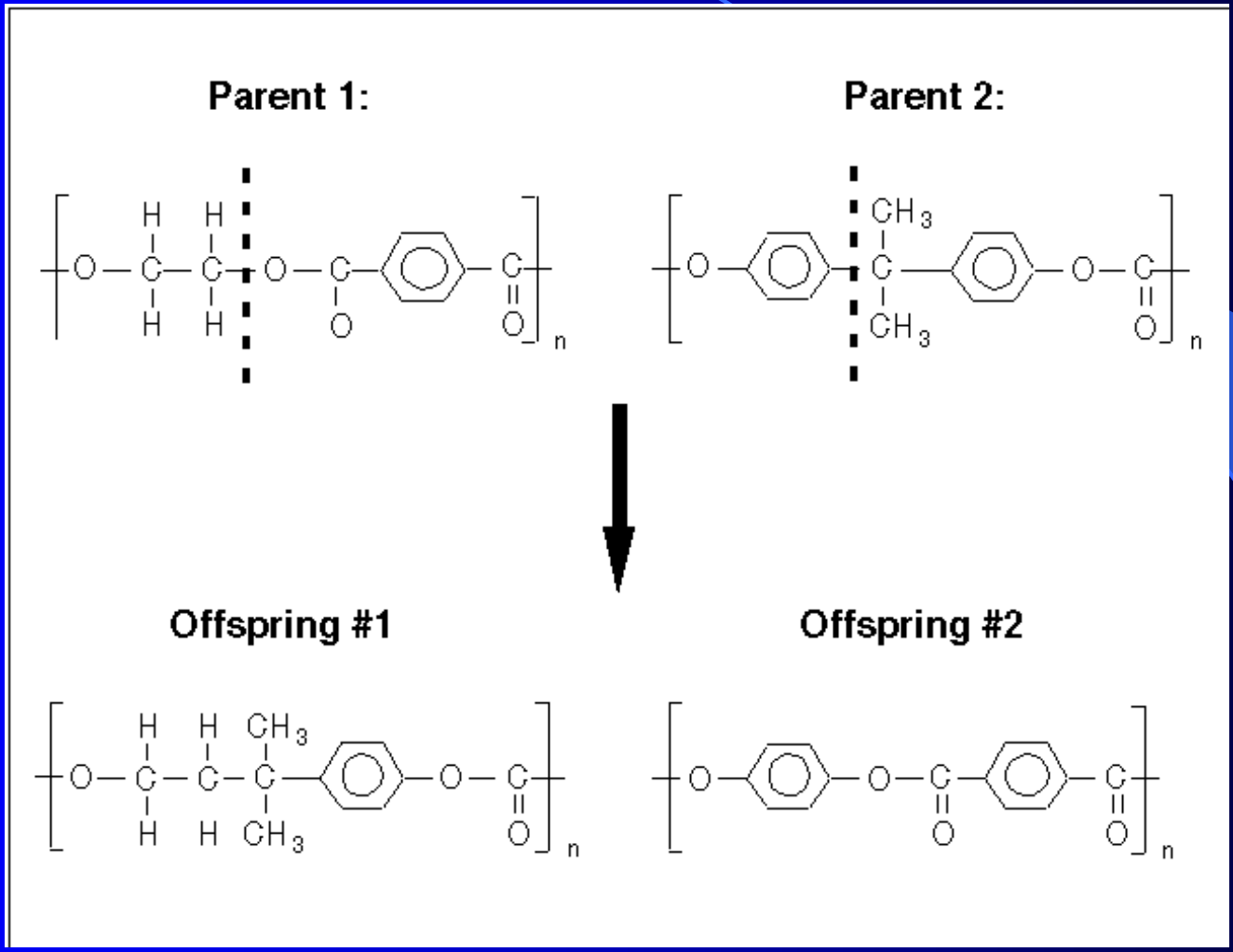
Genetic algorithms manipulate genetic material, but instead of DNA, this genetic material is some other linear string of symbols which can represent base pairs, codons, amino acids, or molecular structures

What happens?



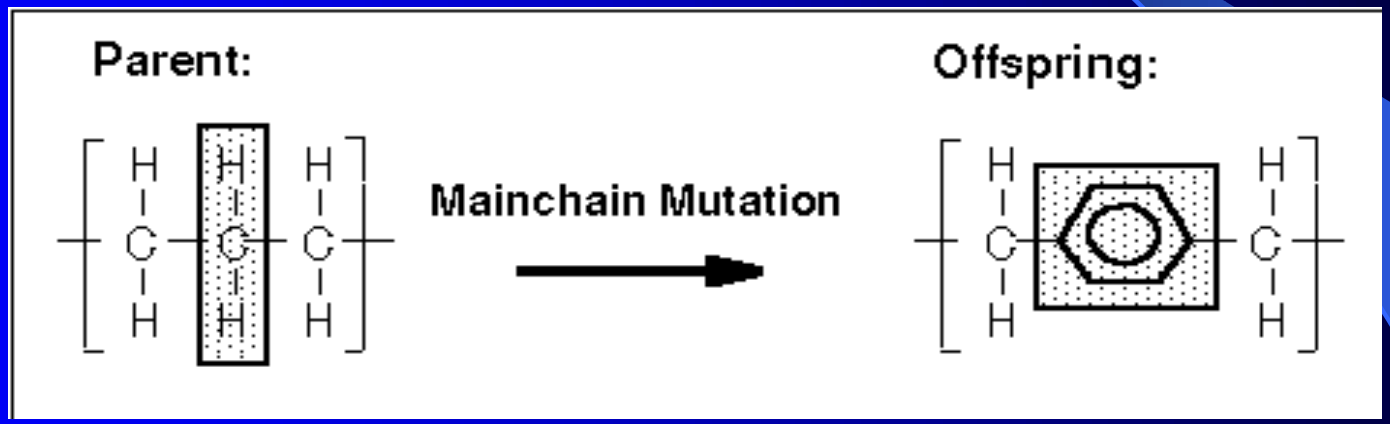
Genetic operators (crossovers, mutations, etc.) occur, and fittest offspring pass on to next generation

Crossover: An Example



In this example, crossover occurs after position three of parent 1 and position two of parent 2

Mutation: An Example



The —CH2— is replaced by a benzene ring

Other Operators

- Blending
- Insertion
- Deletion
- Hopping

How are the fittest offspring determined?

Population members are ranked by a *fitness function*, which could include parameters such as bond angles and energy values that reflect the structure's stability

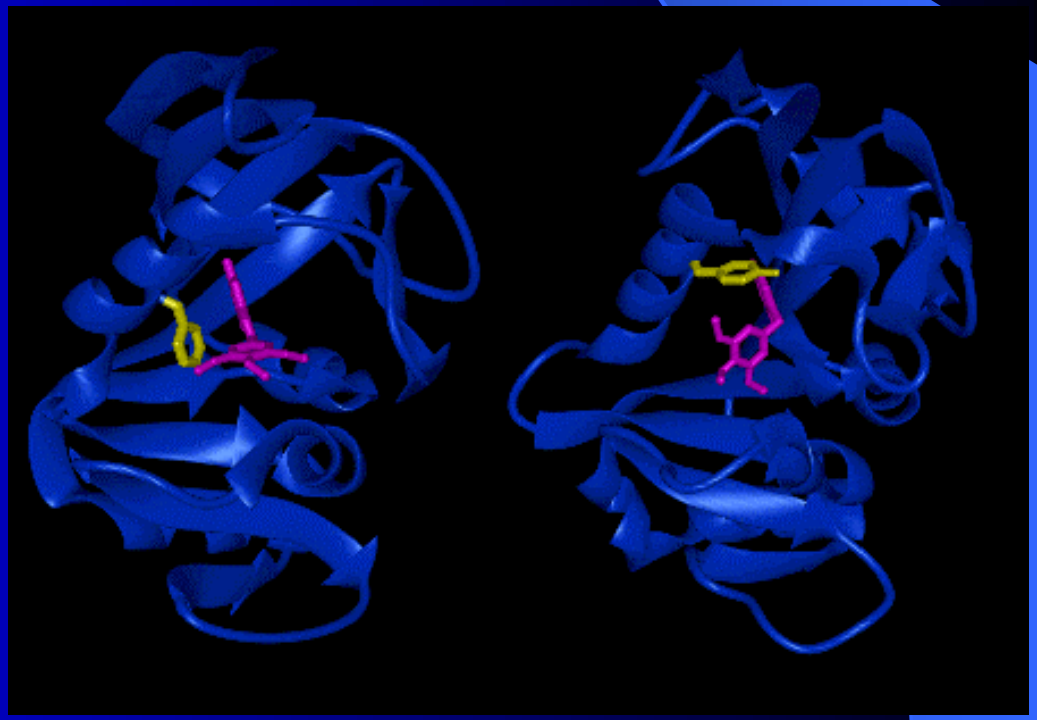
The fitness function can estimate and rank the docking abilities of ligands and receptors—the poorest docking compounds are removed, and the remainder are modified genetically and continue through the loop

The members in the generation with the highest level of fitness become the optimal designs, and will have a higher expected number of offspring

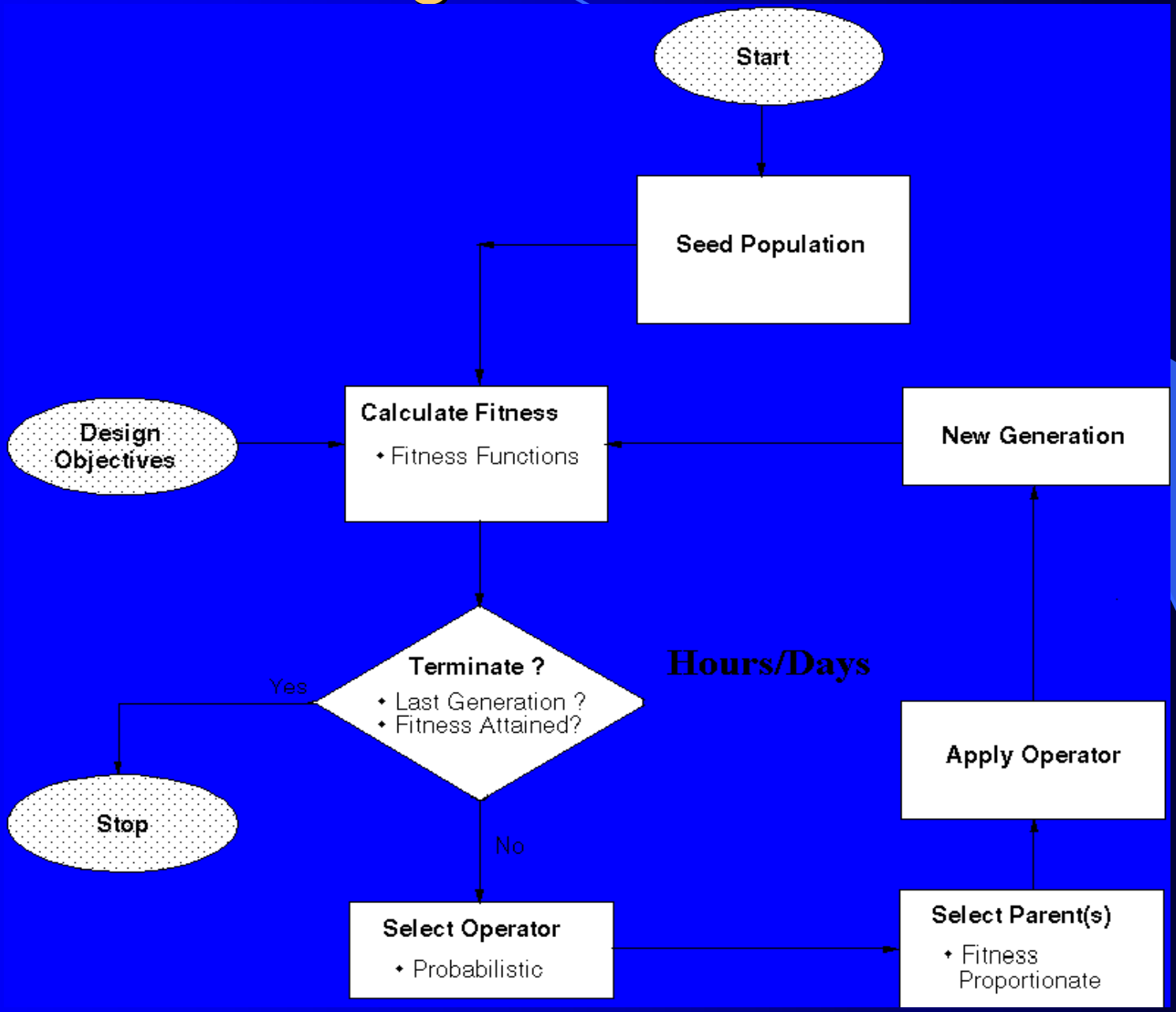
Fitness Functions: Difficulties

Extremely complex:

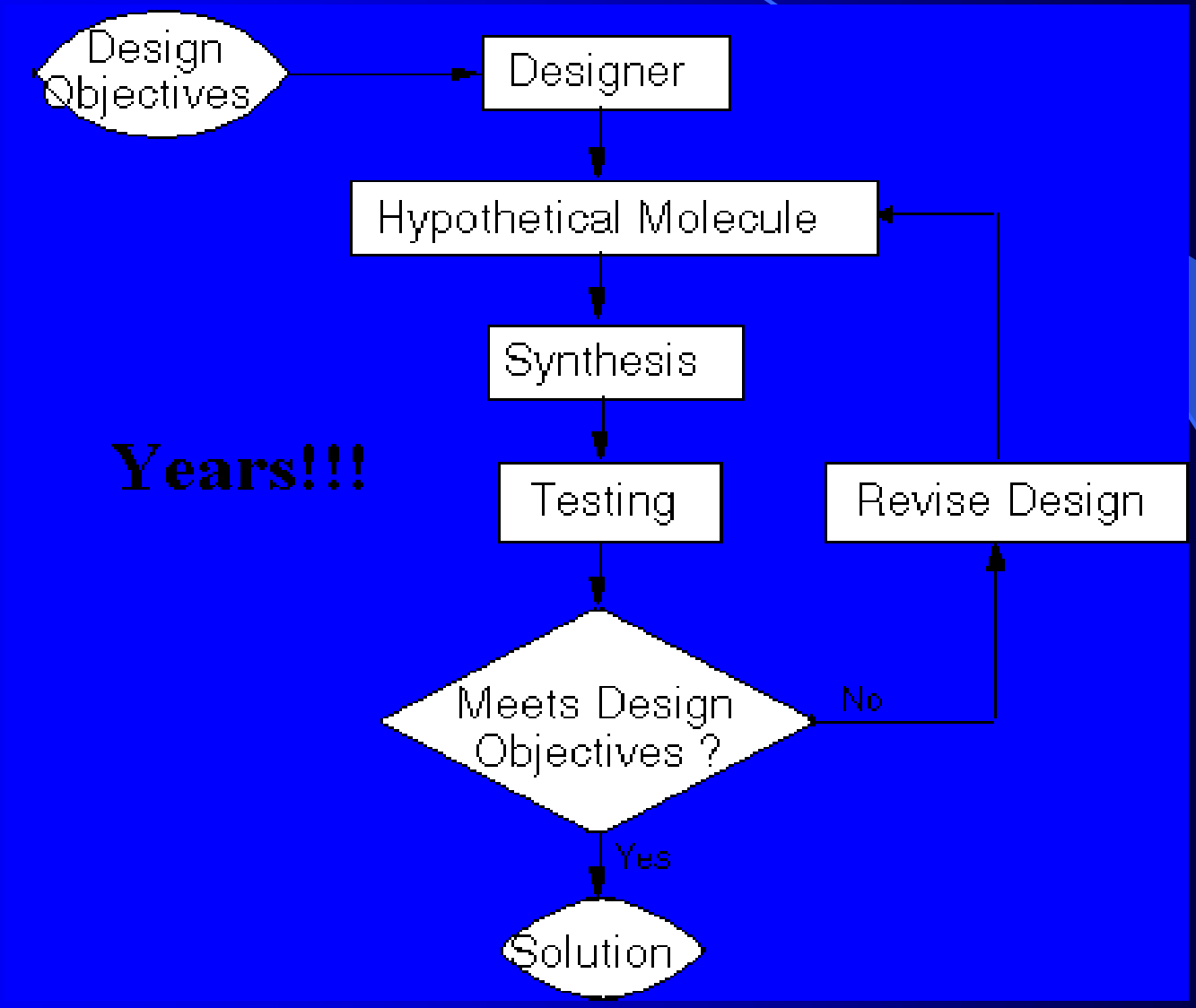
- Must build molecules *and* calculate properties
- Determine effects of placing molecule on the receptor
- Account for 3-D aspect of the molecule and its pathway to the receptor



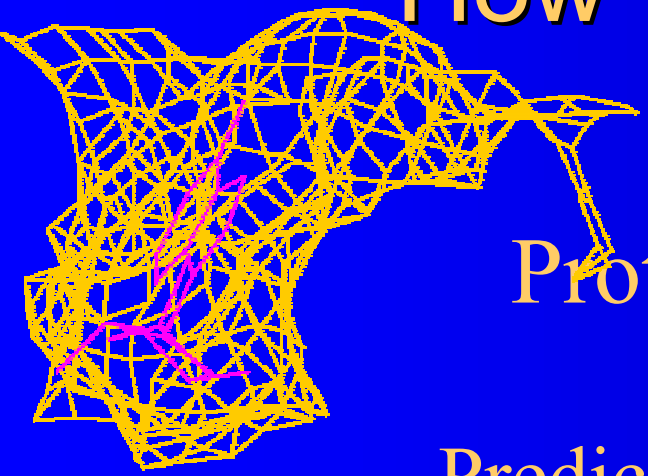
Genetic Algorithm Framework

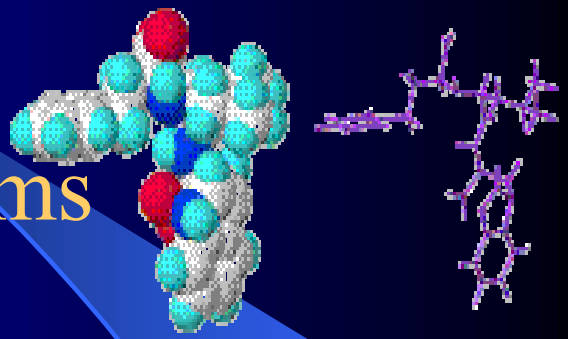


Much better than:



How Will *You* Make Money?





Protein Simulation Programs

Predict structures based on sequences

Predict how ligands will dock into protein structures