



# Multi-objective De Novo Drug Design Via A Novel Evolutionary Algorithm

C.A. Nicolaou<sup>1,2</sup> & C.S. Pattichis<sup>1</sup>

<sup>1</sup> Computer Science Department, University of Cyprus

<sup>2</sup> Noesis Chemoinformatics, Nicosia, Cyprus

Hellenic Bioinformatics & Medical Informatics Conference

4-5 October, 2007



# Outline

1. Drug Discovery Overview
2. Computer-Aided Molecular Design
  - Overview
  - Design Constraints & Multi-objective Considerations
  - The De Novo Approach to Drug Discovery
3. A Novel De Novo Drug Design Approach
  - Current DND System Deficiencies
  - Multi-objective Memetic Graphs
  - Initial Results
4. Future Work

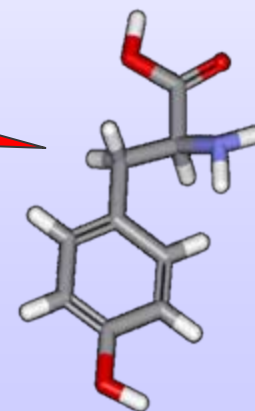
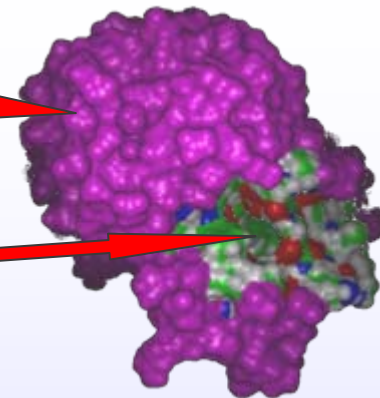


# 1. Drug Discovery Overview

- ◆ Drugs and Drug Action
- ◆ The “Lock and Key” Paradigm
- ◆ Drug Discovery Process

# Proteins, Receptors and Drugs

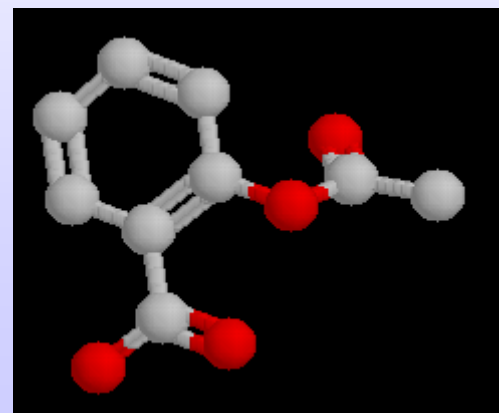
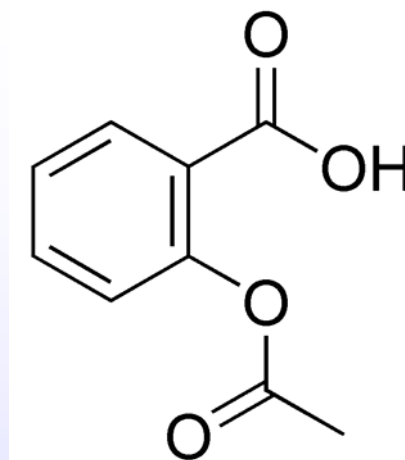
- ◆ Proteins: Macromolecules responsible for most of the functions in human cells
- ◆ Receptor/Target sites: Protein regions where other “natural” molecules bind to provoke or inhibit a function.
- ◆ Drugs/Leads/Hits: Externally administered molecules, capable of binding to receptor sites to modulate protein function



# Drugs

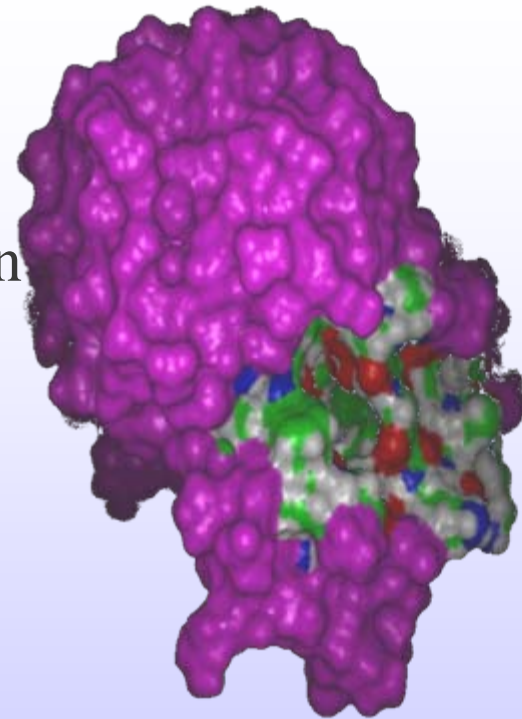
- ◆ Drugs are (typically) small molecules having a favorable biological effect on an organism.
- ◆ Drugs like all molecular compounds are flexible 3D structures
  - often represented by 2D or 3D-molecular graphs

Aspirin



# Drug Action

- ◆ Drugs act by binding to a pharmaceutical target, typically a protein.
- ◆ By binding they inhibit, stimulate, modulate the action of the protein and thus provoke a biological effect
- ◆ “Lock and Key” Paradigm





# The SAR Theory

- ◆ SAR – Structure Activity Relationships

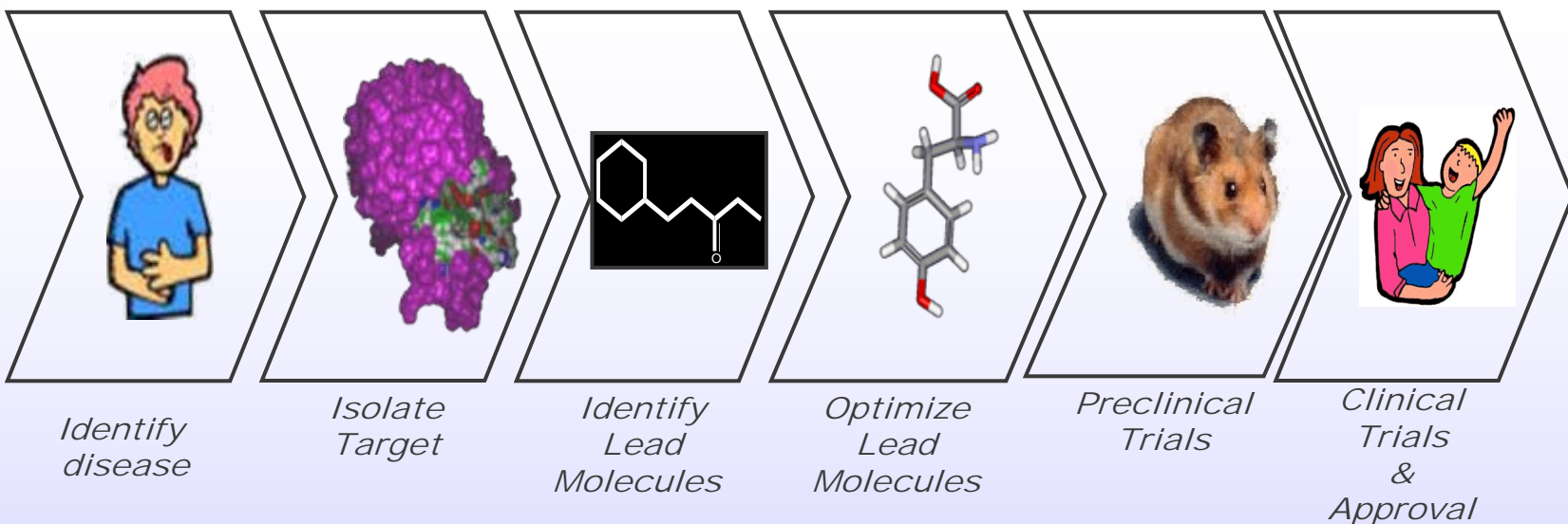
$$\textit{Activity} = f(\textit{structure})$$

- SPR – generalization (Property)
- STR – Toxicity

- ◆ Biological property can be explained by the molecular structure, e.g. the ability to complement and bind to the active site of the pharmaceutical target.

- Function Follows Form

# Drug Discovery Process



## Notes:

- Drugs must satisfy multiple criteria: multi-objective problem
- Hits Vs Leads Vs Drugs

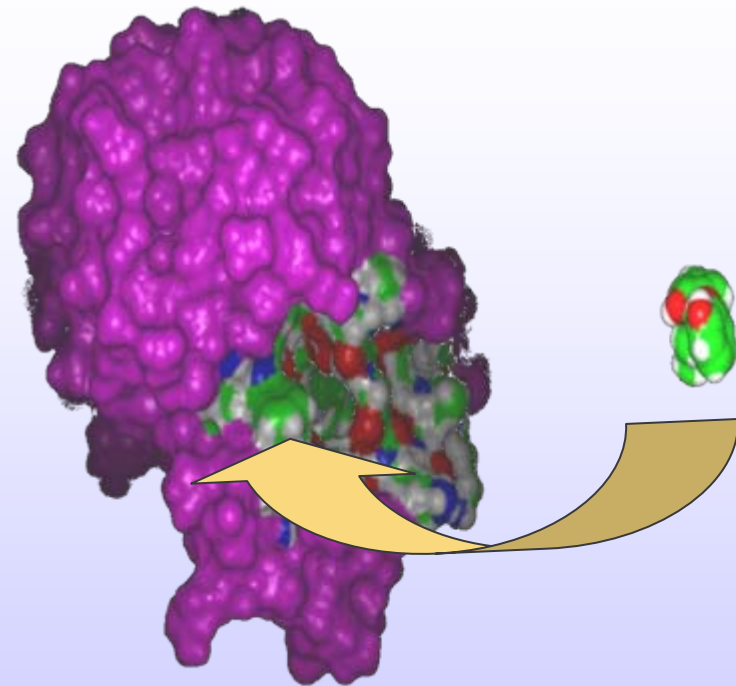


## 2. Computer-Aided Molecular Design

- ◆ Overview
- ◆ Design Constraints
- ◆ Multi-objective Considerations
- ◆ The De Novo Approach to Drug Discovery

# Computer-Aided Molecular Design [CAMD]

- ◆ To utilize computational intelligence to expedite drug discovery process
  - Increase efficiency of the process
  - Reduce failures, cost, time
- ◆ Problem is to find a molecule with:
  - Activity: “Lock & Key” concept
    - Lock = pharmaceutical target
    - Key = drug
    - Flexible lock AND flexible key
  - Additional drug requirements
    - Toxicity
    - Pharmacokinetics (ADME):
      - Absorption, Distribution, Metabolism, Excretion
    - Synthetic feasibility
    - Cost...
- ◆ True multi-objective problem



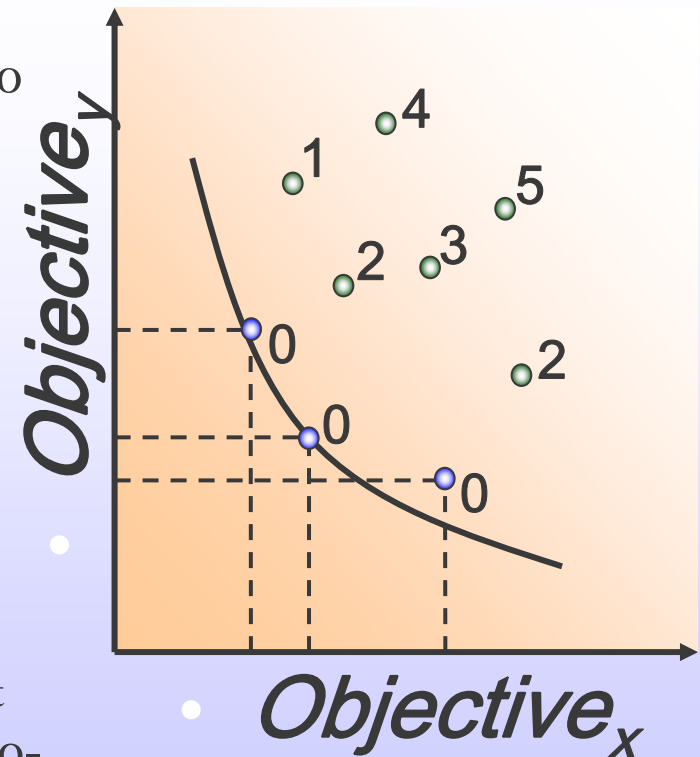


# CAMD Design Constraints

- ◆ Two main categories:
  - Receptor/Target based/constrained
    - Construct ligand that fits in a receptor
    - Majority, traditional
    - Requires detailed 3D structure of target (not always available)
  - Ligand/Analog based/constrained
    - Construct ligand similar to known ligands
    - Fewer representative methods
    - Requires known ligands (not always available)
- ◆ Challenges for any approach:
  - Vast solution space
  - Multi-objective nature of problem

# Optimization Considerations

- ◆ Some problems too complex/large to exhaustively search
- ◆ Optimization: Discovering an optimal solution to such problems
  - Balance solution quality and required effort
- ◆ Single-Objective Optimization
  - Taking into account one objective/criterion/goal
  - Searching for the global optima of objective
- ◆ Multi-Objective Optimization (MOOP)
  - Considering multiple objectives concurrently
  - Multiple competing objectives:
    - No single optimum possible
    - Multiple equivalent optima forming a trade-off/compromise surface known as the Pareto-front
- ◆ MOOP algorithms aim to approximate the Pareto-front effectively at the minimum computational cost.





# De Novo Drug Design [DND]

- ◆ Designing drug-like molecules from scratch
  - Little initial input in the form of constraints
- ◆ First applications appeared in the mid-90's
- ◆ Involves finding solutions to large multi-objective combinatorial problems for which an exhaustive search is difficult/impossible
  - Estimated  $10^{60}$  chemical structures using only C, N, O, S
- ◆ Known to be a “humbling experience” among chemoinformaticians



# Traditional DND Systems

- ◆ Single objective, i.e. focused on designing molecules taking into account only activity/“lock and key” paradigm
- ◆ Designed molecules either:
  - Fitting into a well known pharmaceutical target pocket, or,
  - Resembling a known drug/ligand
- ◆ Employed various optimization/approximation algorithms and/or exhaustive search when possible
- ◆ Often, produced molecules with severe limitations in practice due to lack of consideration of other pharmaceutically relevant properties/characteristics.

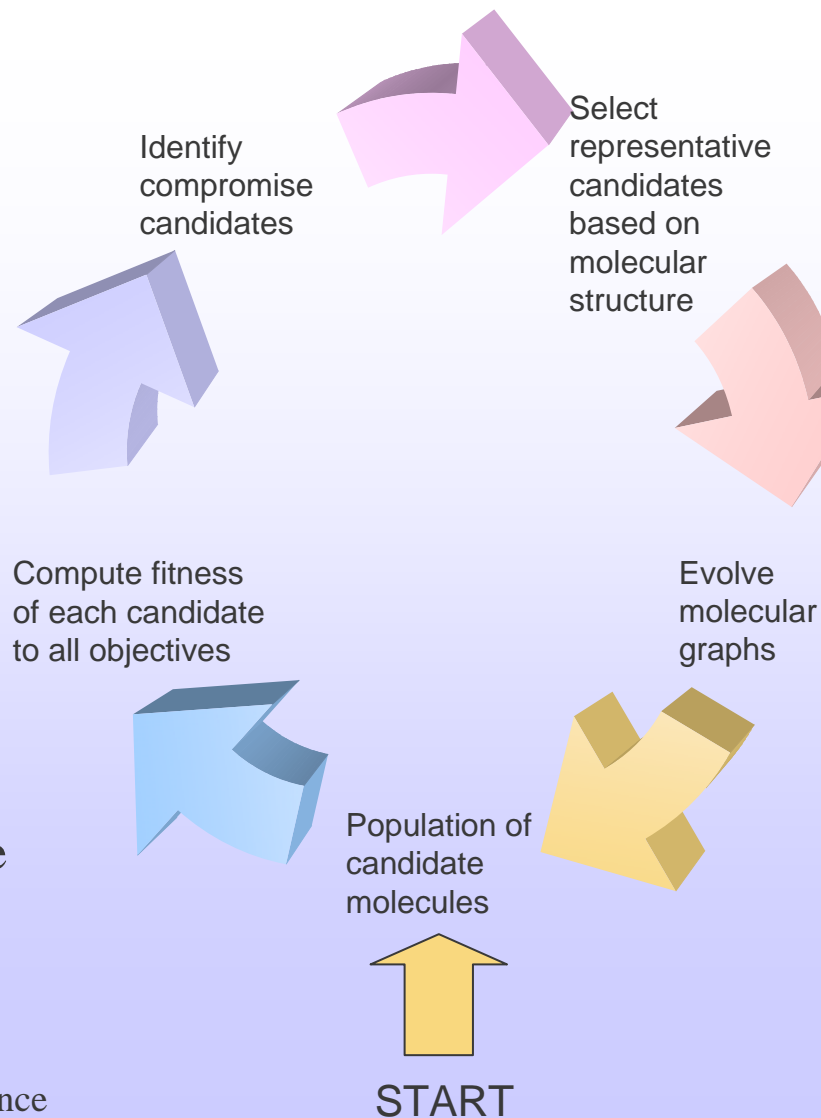


### 3. A Novel De Novo Drug Design Approach

- ◆ Multi-objective Memetic Graphs
- ◆ Initial Results

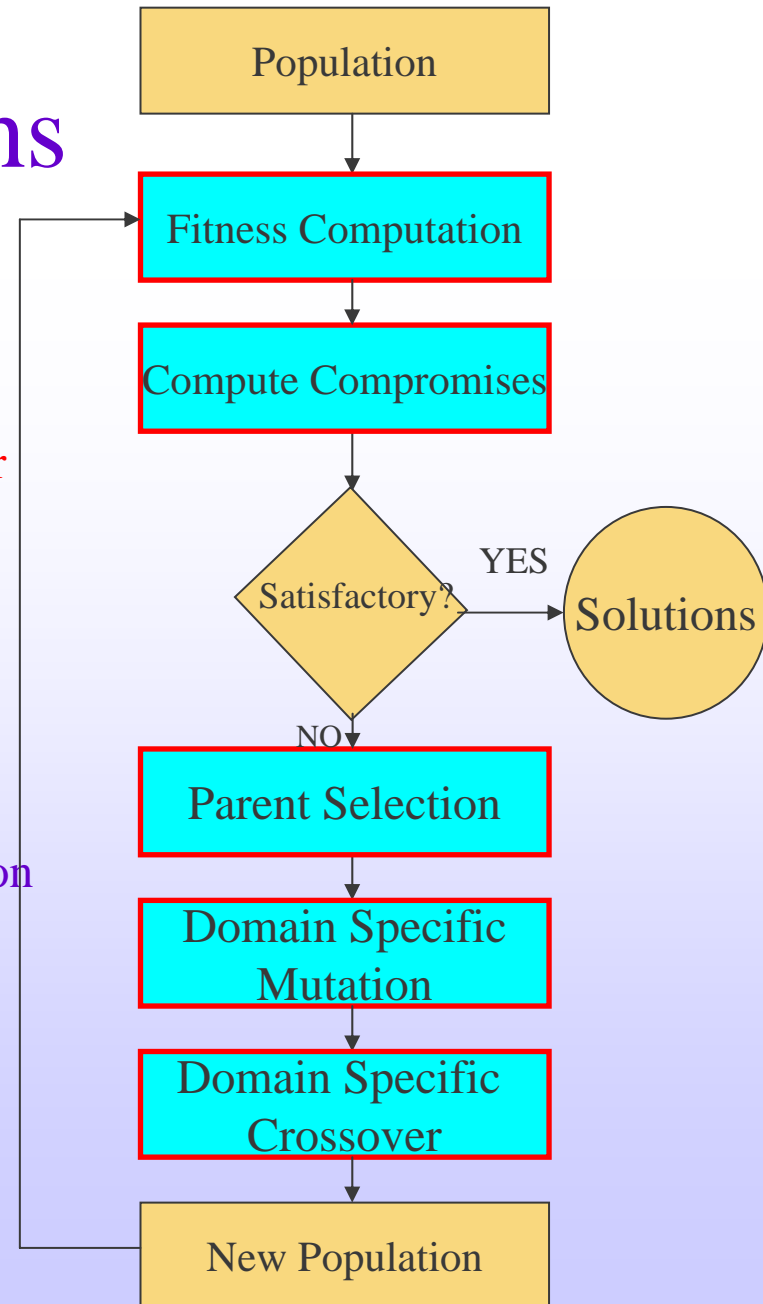
# Multi-Objective Memetic Graphs

- ◆ Graph chromosomes
  - Building blocks/genes
- ◆ Graph-specific evolution
  - Mutations on:
    - edges - bonds
    - vertices – atoms
    - ring systems
    - fragments/subgraphs
  - Crossover - exchange subgraphs
- ◆ Selection step involves:
  - Graph-based clustering
  - Picking representative molecular structures to ensure adequate coverage of objective compromise space/Pareto front



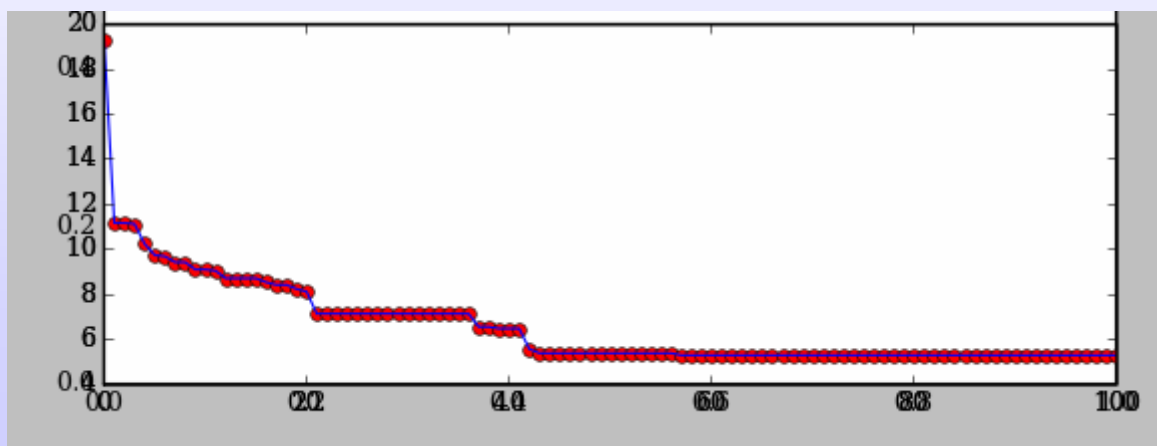
# MO Memetic Graphs

- ◆ Generate initial population
  - Graph/molecule format
- ◆ While Not Stop Condition:
  - Fitness computation for population for each objective
  - Identify set of compromise solutions/Pareto front
  - Select parents using:
    - Set of compromise solutions
    - Graph-based clustering/diversity analysis
  - Generate new solutions by reproduction of selected parents
    - Mutation
    - Crossover
  - Merge new solutions with current population to form new Population

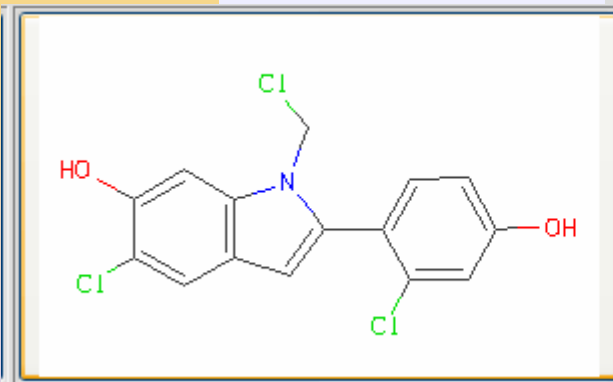
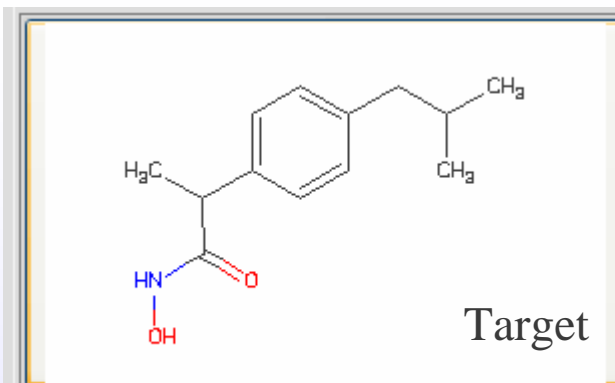


# Initial Results - 1

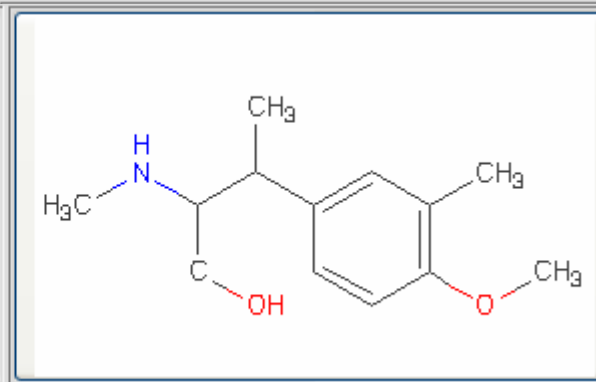
- ◆ Goal: Design molecules similar to ibuprofen
  - Single Objective
- ◆ Constraint based on molecular similarity (vector-based)
- ◆ Input:
  - Target molecule
  - A dissimilar set of molecules (~50 estrogens) forming the initial population
- ◆ Convergence: ~50 generations



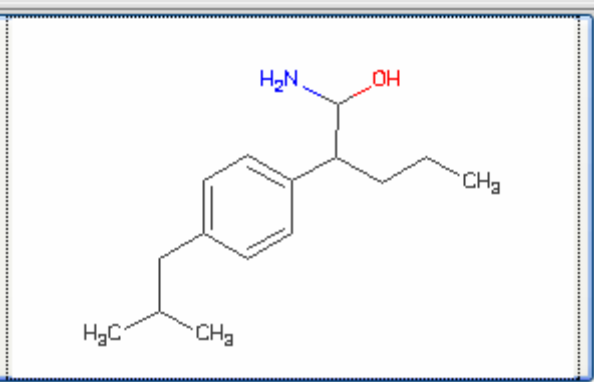
# Designing Molecules (1)



Generation 1



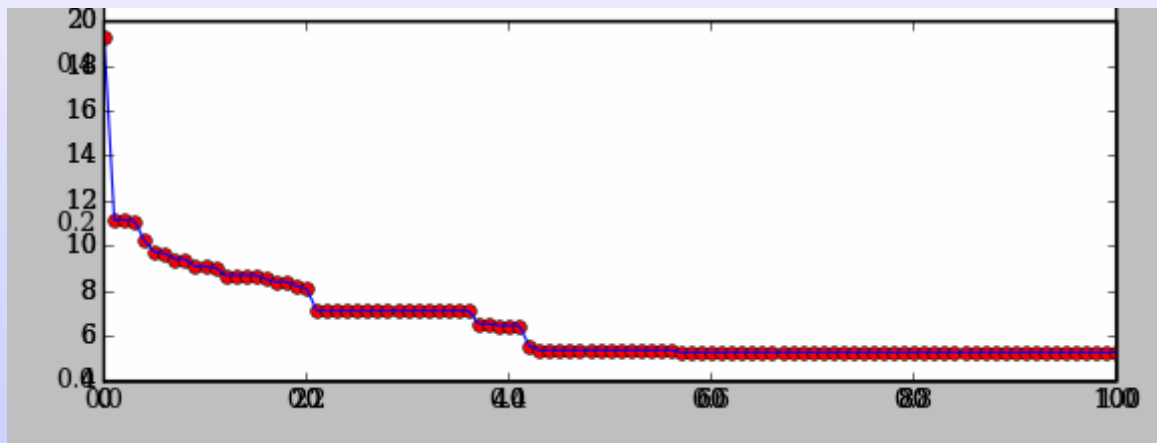
Generation 20



Generation 100

# Initial Results – 2

- ◆ Goal: Design molecules similar to ibuprofen
  - Single Objective
- ◆ Constraint based on molecular similarity (graph-based)
- ◆ Input:
  - Target molecule
  - A dissimilar set of molecules (~50 estrogens) forming the initial population
- ◆ Convergence: ~50 generations

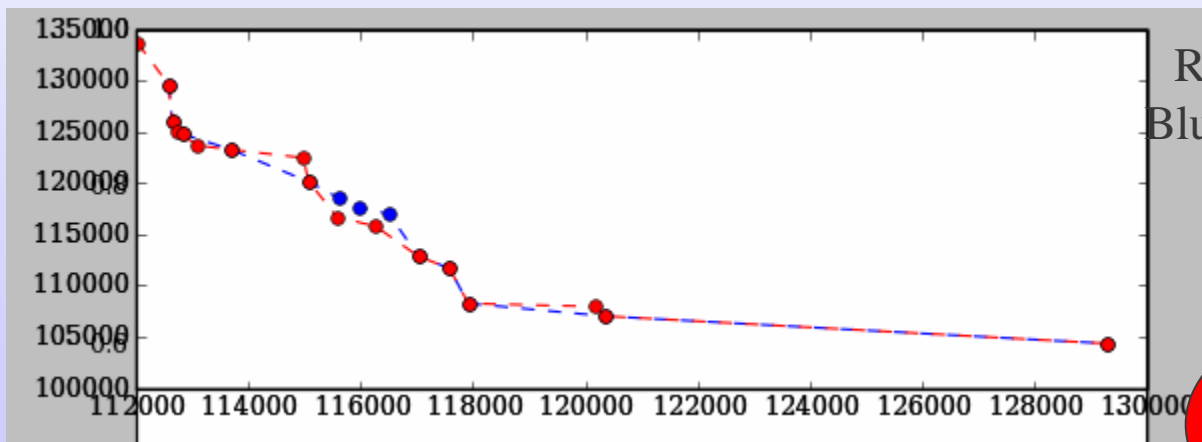




# Initial Results - 3

- ◆ Goal: Design molecular graphs satisfying:
  - Similarity to known ligand
  - Low complexity value
  - A bi-objective problem
- ◆ Constraint based on graph measures (structural similarity, complexity)
- ◆ Input: Known ligand, a set of 50 molecular graphs randomly generated
- ◆ Stabilization to a Pareto front: ~250 epochs
  - Variety of graphs: similar to input graph and more complex Vs less complex but not as similar

Current  
Research  
Effort



Red: Final pareto front  
Blue: Previous generation

Work In Progress!!!

## 4. Future Work





# Directions

- ◆ De Novo Design [Work in progress]
  - Include/test additional fitness functions
    - Toxicity, selectivity, ...
  - Memetic component enhancement
    - Incorporate additional elements of domain specific knowledge
  - Exploit self adaptation and negative information
    - Evolve strategy of new solution generation at the same time of evolving new solutions; record success in improvement of certain search space regions and exploit appropriately
    - Use information obtained by mutations not improving fitness
- ◆ Apply method on real data through collaborations



# Thanks for your attention!

## ◆ Contact

Christos Nicolaou

Medical Informatics Lab,

Computer Science Department

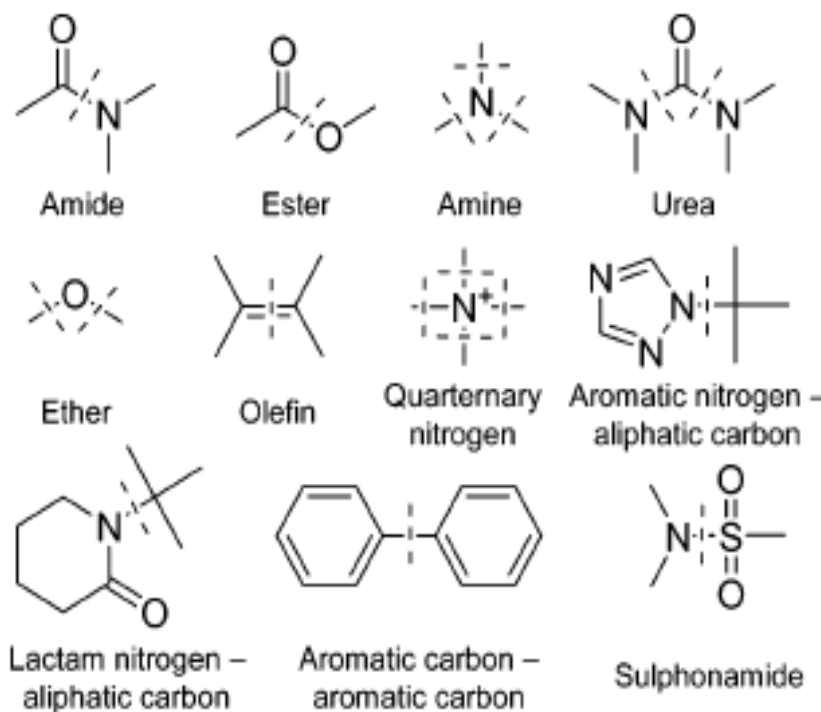
University of Cyprus

[cnicolaou@cs.ucy.ac.cy](mailto:cnicolaou@cs.ucy.ac.cy)

[cn@noesisinformatics.com](mailto:cn@noesisinformatics.com)

# Building Blocks

- ◆ Building blocks:
  - Privileged Molecular Fragments
  - Atoms
  - Bonds
- ◆ Molecular Fragments
  - RECAP Molecular fragmentation method
  - Substructure Mining
- ◆ Profiled to assess “privileged” nature

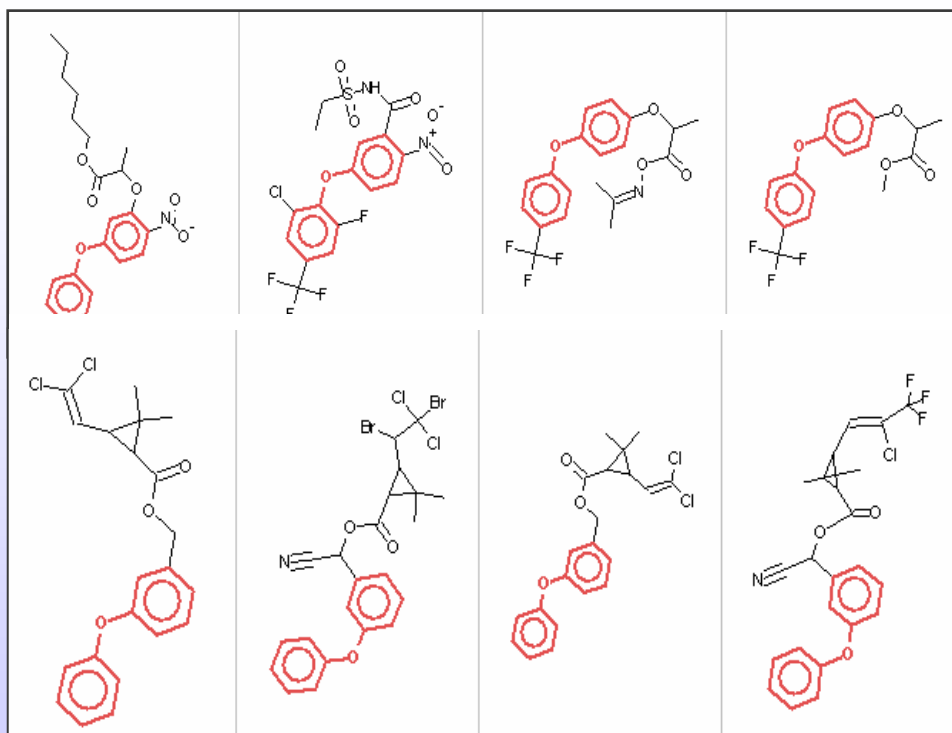


RECAP bond cleavage types

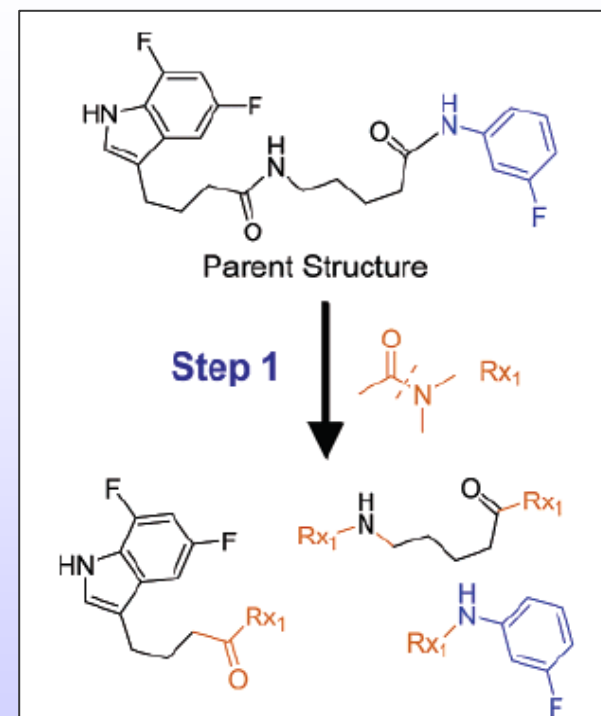


# Subgraph Identification

## SUBSTRUCTURE MINING



## RECAP

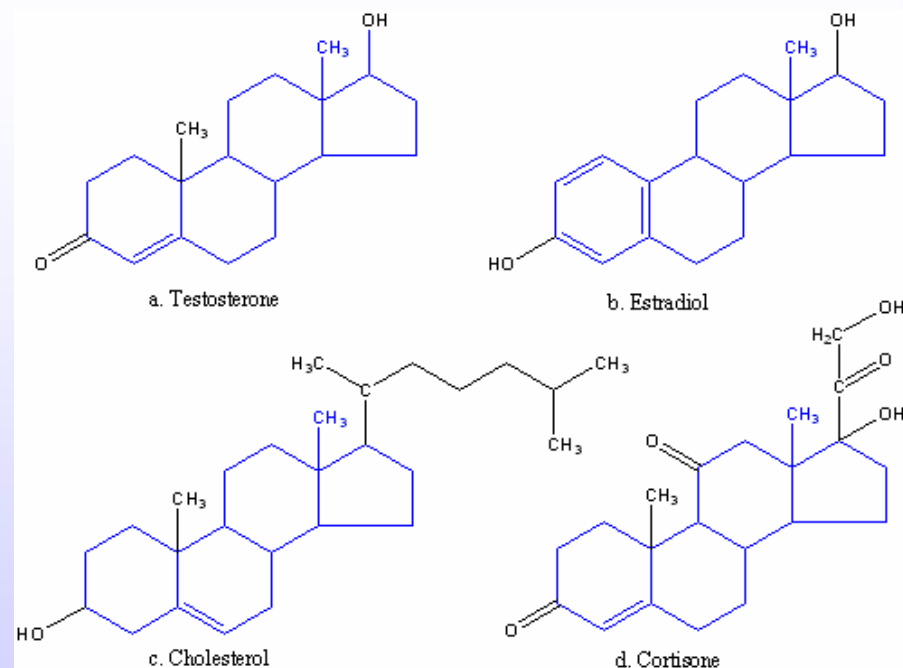


\*Fechner & Schneider, JCI, 42, 36, 2006.



# Graph-based Clustering

- ◆ Clustering based solely on molecular structure
- ◆ Guarantees that each and every cluster contains compounds sharing a substructure of substantial size
- ◆ May utilize:
  - Substructure mining
  - Graph-based similarity



**A steroid cluster**



# Identifying Compromise Solutions

