SAND2000-0007

Evolutionary Hybrids for Flexible Ligand Docking (Electronic Notes)

William E. Hart
Optimization/Uncertainty Estimation Dept
Sandia National Laboratories
wehart@sandia.gov
www.cs.sandia.gov/~wehart

In Collaboration with
Chris Rosin (TSRI), Richard K. Belew (UCSD)
Garrett Morris (TSRI)

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

Sandia National Laboratories

Overview

1. Evolutionary algorithms (EAs)

2. EA Hybrids
   - Design
   - Motivation
   - Dynamics
   - Applications

3. Drug Docking with AutoDock
   - Problem formulation
   - Empirical evaluation of EA hybrids

Sandia National Laboratories

Evolutionary Algorithms — Overview

Idea: a multi-point search strategy inspired from natural evolutionary processes

Motivation:

- Selection of better individuals (Darwin)
- Generation of variants from selected individuals (Mendel)

Analogous Search Mechanisms:

- Stochastic competitive selection
- Selective recombination (a.k.a. crossover)

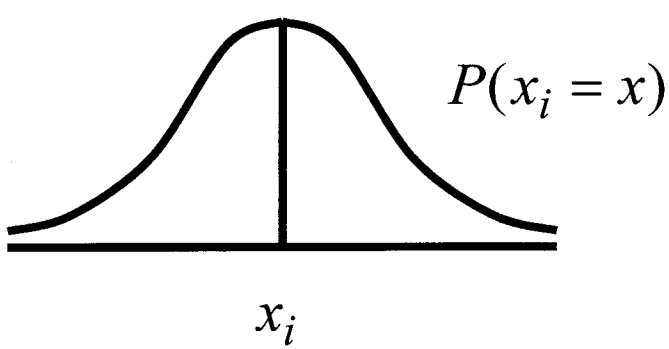
Evolutionary Algorithms: genetic algorithms, genetic programming, evolutionary programming, evolutionary strategies

Methods for generating variants

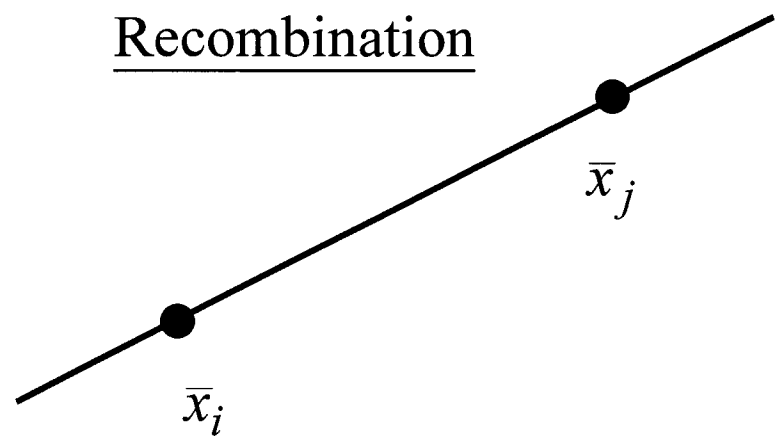
- Mutation - generate a simple change
- Recombination - combine two or more search points (a.k.a. parents)

Example: searching R^n

Mutation



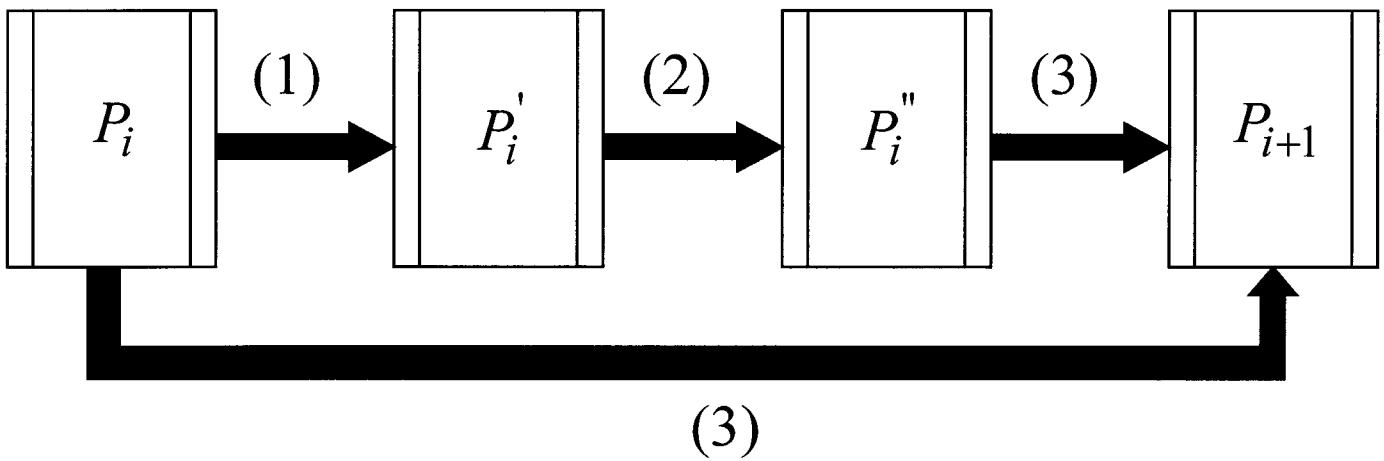
Recombination



Sandia National Laboratories

Evolutionary Algorithms

Main Loop



(1) Stochastic competition

(2) Generate new variants

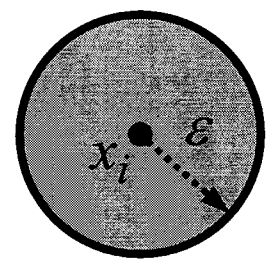
(3) Replacement competition

Local Search

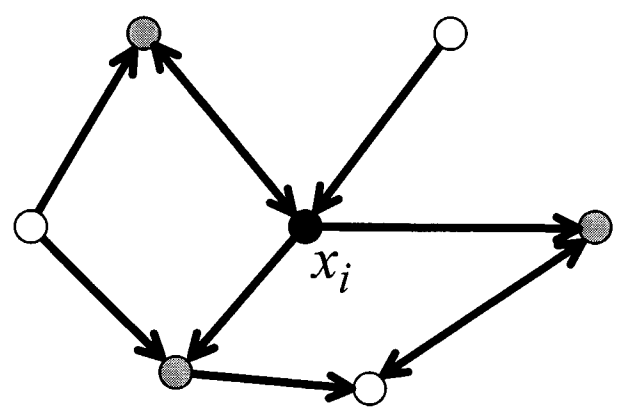
Idea: iterative examine solutions in a local neighborhood and select one that has a lower value.

Examples of local neighborhoods:

- An ε-close ball (in R^n)



- Adjacent nodes in a graph



Sandia National Laboratories

EA Hybrids with Local Search

Motivation

Idea: utilize local search to refine solutions in each iteration of the EA.

Biological Motivation

- Organisms utilize phenotypic adaptations (e.g. learning) to adapt to their environment
- Evolution modifies an individual's genotype across generations
- Phenotypic adaptations smooth the evolutionary landscape

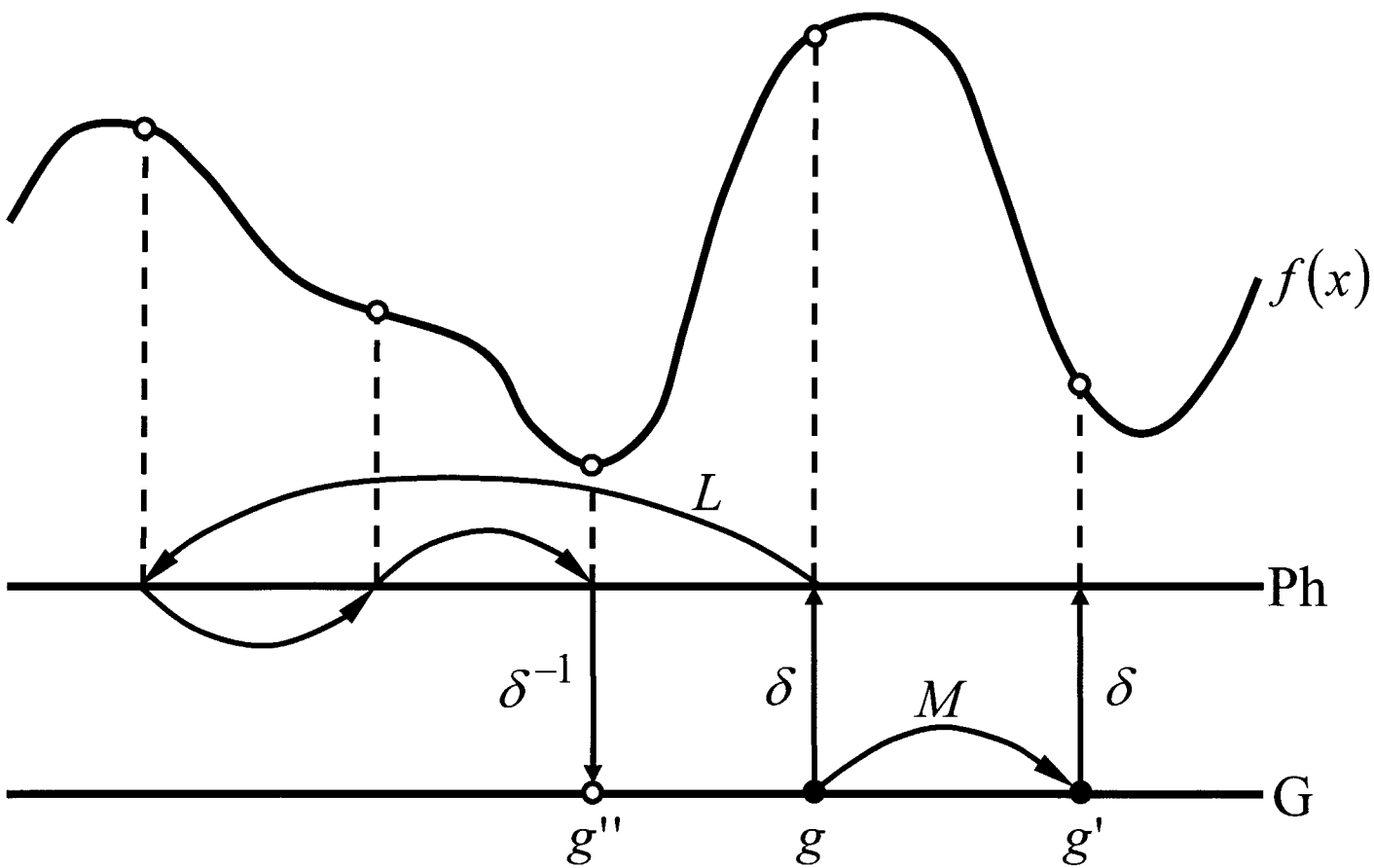
Mathematical Motivation

- Local search quickly refines solutions
- Global search ensures that all areas of the search domain are adequately explored
- Most effective global search methods combine global and local search

Sandia National Laboratories

Model

Local search can distinguish between a genotypic and phenotypic search space through a developmental map $\delta : G \rightarrow Ph$.



Model (cont'd)

Lamarckian Local Search

- modifies the fitness value and solution parameters
- is typical used for optimization

NonLamarckian Local Search

- modifies only the fitness value
- can be useful when the EA's selection pressure is "too strong"
- can be viewed as a "repair mechanism for constrained problems

EA Hybrids

Design

Pseudo Code

Initialize a population of points $\{x_0^0, \ldots, x_0^N\}$
Perform local search?
Repeat
    Perform competitive selection to select $\{\hat{x}_t^0, \ldots, \hat{x}_t^N\}$
    Apply evolutionary operators (mutation, recombination)
    Apply local search (A)
    Replace previous population of points
    Apply local search (B)
Until stopping rule satisfied

Sandia National Laboratories

EA Hybrids

Design (cont'd)

Design Issues

- Duration of local search
- Type of local search (robustness!)
- How local search is initialized
- What points are refined with local search
- Lamarckian or nonLamarckian local search

Hybrid designs:

memetic algorithms

parallel genetic algorithms

Lamarckian genetic algorithms

genetic local search

genetic hillclimbing

EA-LS hybrids

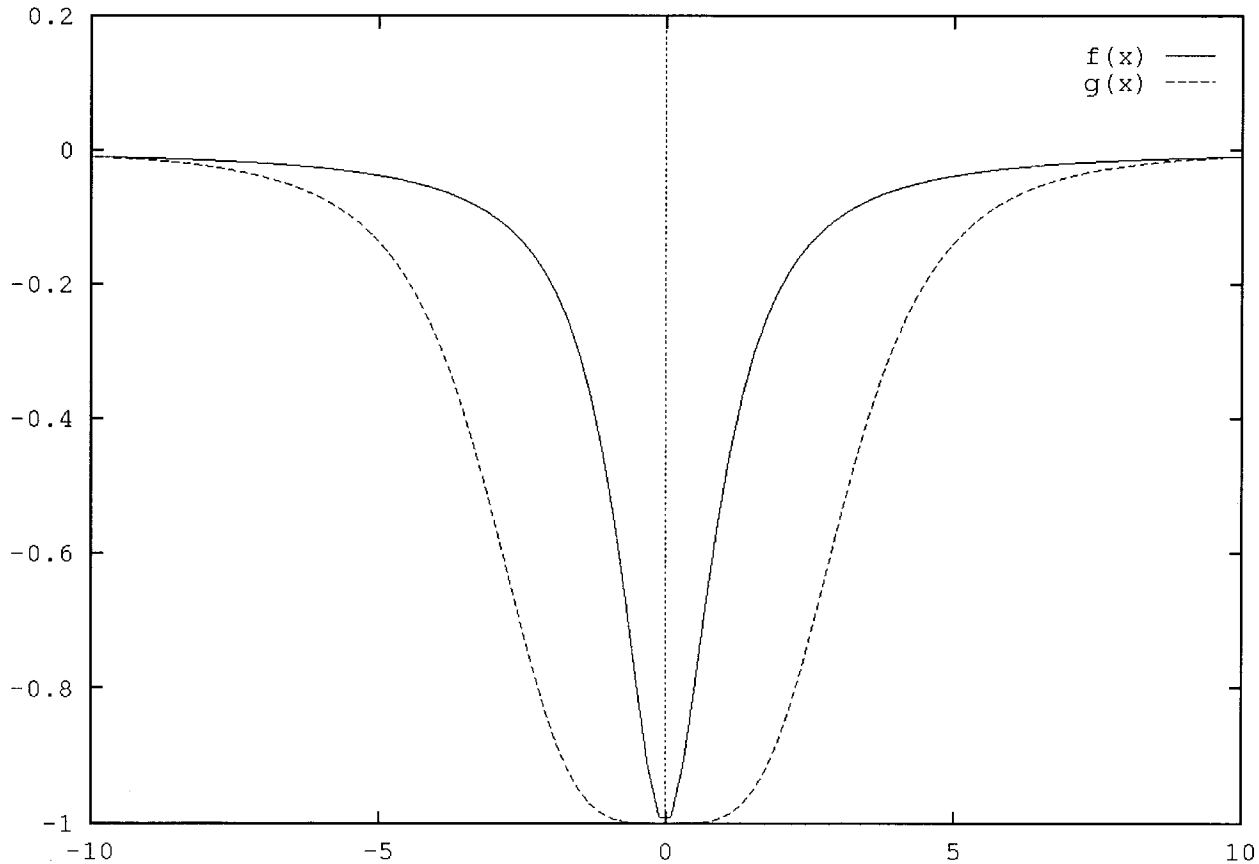
Fitness Transformation

Local search can play a variety of roles in EAs (Hart et al, 94)

I. Fitness transformation: nonLamarckian EA hybrids

f(x) Func

g(x) Transformed func



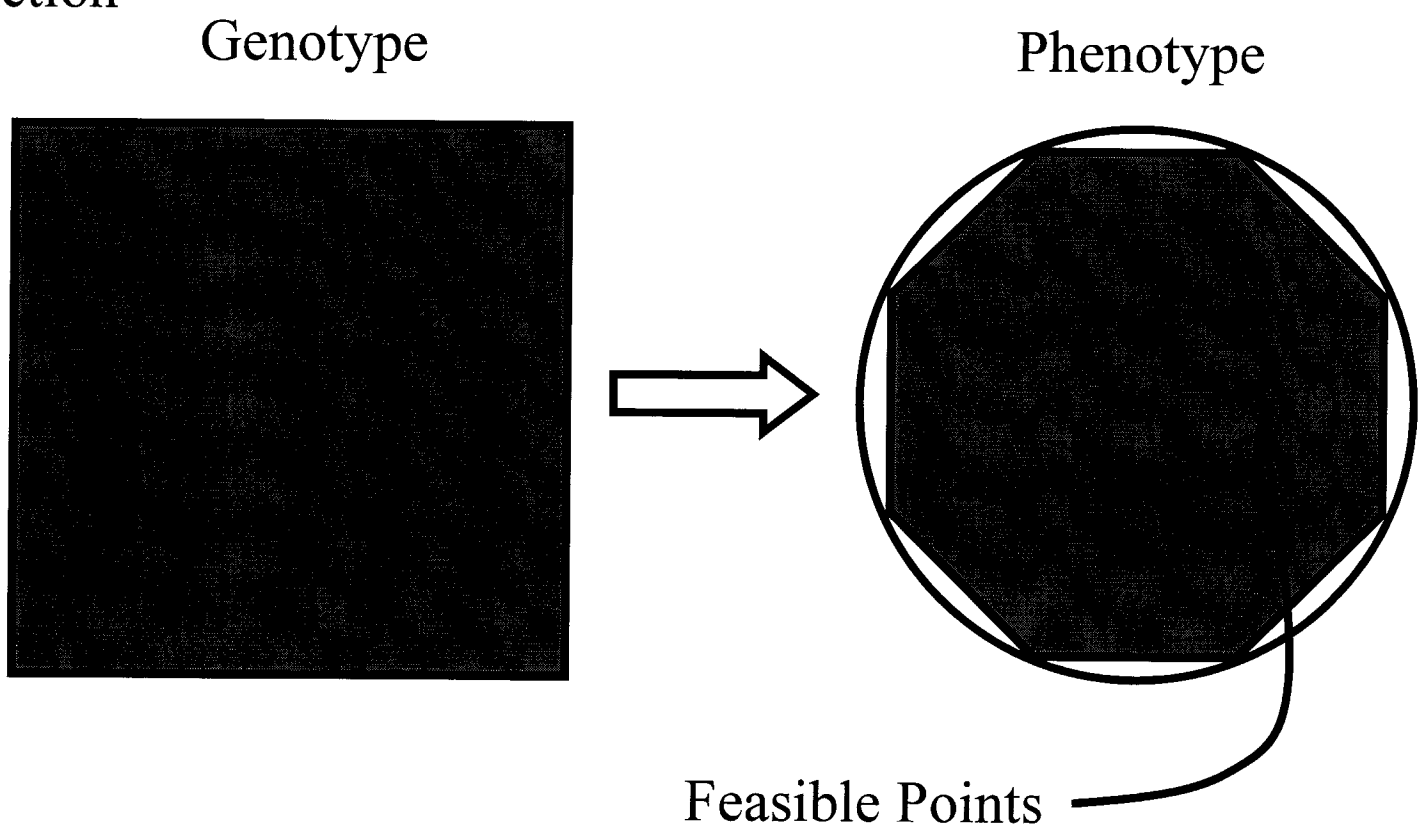
II. Refinement operator: Lamarckian EA hybrids

Extreme Case

- all solutions are mapped to their respective local minima
- recombination and mutation define transition between local minima
- Example: genetic local search

Repair Operator

III. Repair operator: nonLamarckian EAs with a genotype-phenotype distinction



Example: evolving grammars (Kammeyer and Belew, 97)

Applications of EA Hybrids

Thin Film Metrology (Land et al, 97)

Flowshop Scheduling (Ishibuchi et al, 97)

Drug Docking (Rosin et al, 97; Morris et al, 98; Hart et al, 99)

Reliable Communication Networks (Dengiz et al, 97)

QAP (Mühlenbein, 91)

Neural Network (Belew et al, 91; Keesing and Stork, 91; Gruau and Whitley, 93)

[See also www.ing.unlp.edu.ar/cetad/mos/memetic_home.html]

Sandia National Laboratories

Motivation

Goal: develop docking tools for structure-based drug discovery

AutoDock Method

- Flexible ligand
- Rigid receptor
- Conformational search used to find good docking configurations

The AutoDock Potential Energy Model

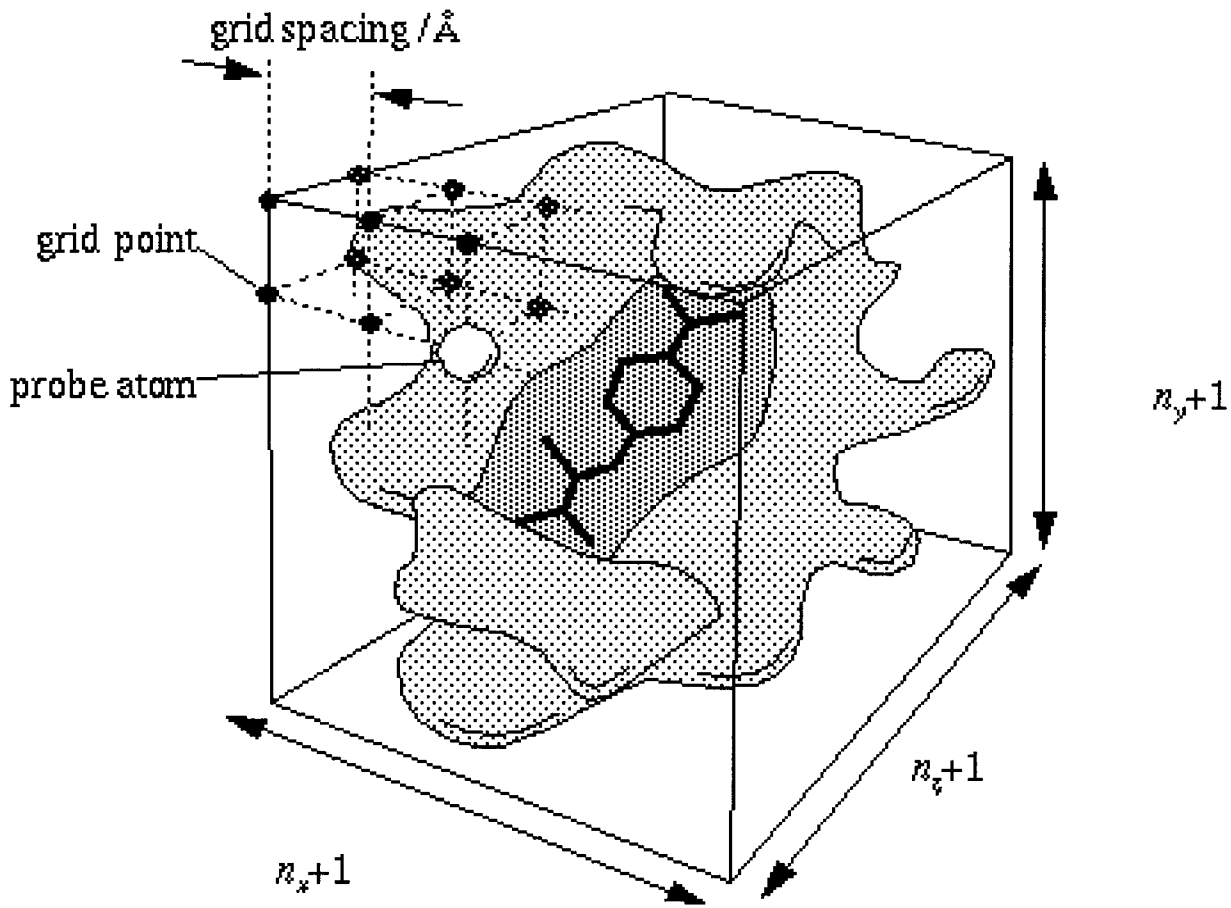
Energy Potential:

- Long-range electrostatic forces
- Short-range weak van der Waals attractive forces
- Hydrogen bonds (12-10 LJ potential)
- Unfavorable conformational entropy
- Desolvation measure (Stouten et al '93)

Sandia National Laboratories

Computing Potential Energies

Ligand-protein pairwise interaction energies are precalculated



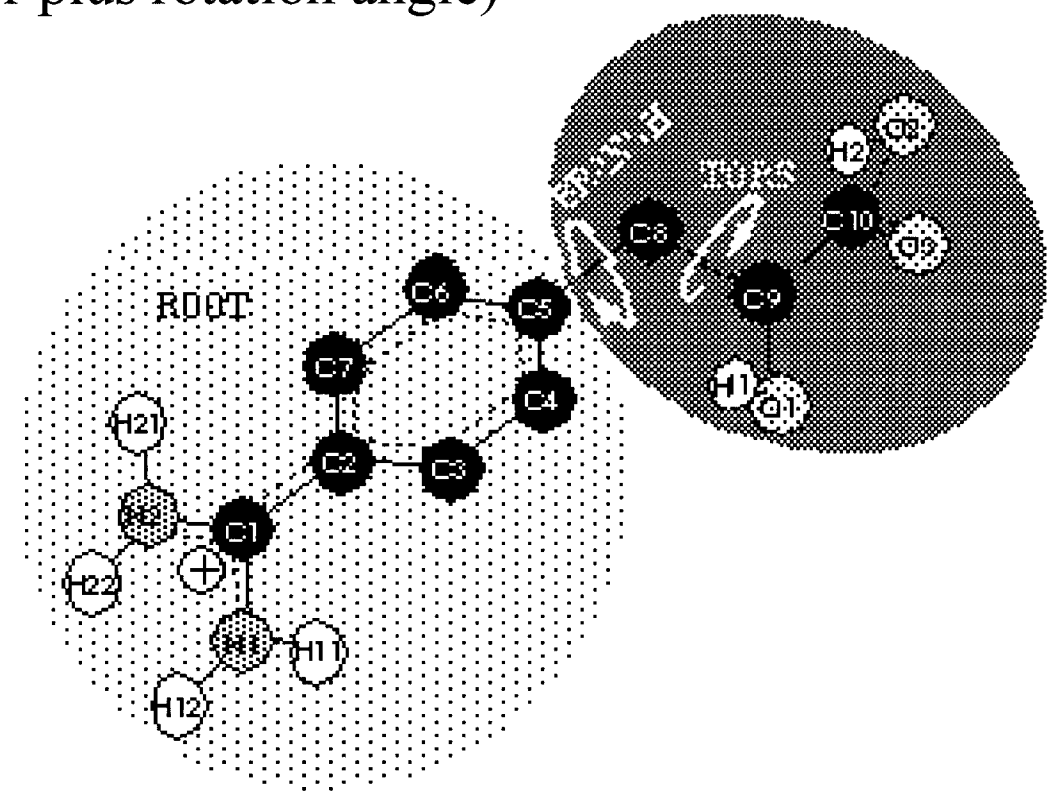
Grid-based method rapidly computes interpolated energies

Sandia National Laboratories

Docking Formulation

Search Dimensions

- Cartesian Coordinates (3 dimensions)
- Quarternion (unit vector plus rotation angle)
- Torsion angles



Docking Formulation (cont'd)

Optimization Formulation

- Unconstrained variables
- Quadratic penalty for coordinate bound constraints
- Quarternion vector renormalized

Note: this objective is nondifferentiable

Sandia National Laboratories

Optimization Methods

Simulated annealing

- Steepest descent that accepts inferior points with probability

$$e^{-\Delta E / T},$$

  where ΔE is the difference in energy and T is a "temperature" constant
- Good for "walking through" shallow local minima

Evolutionary algorithms

- Evolutionary programs
- Evolutionary pattern search
- EA hybrids

Sandia National Laboratories

Evolutionary Programming

Evolutionary operator: mutation

- Idea: add a scaled random variable to each dimension
  - Normal: $\mathrm{N}(0,\sigma_i)$
  - Cauchy: $\mathrm{C}(0,\sigma_i)$

Self-adaptation

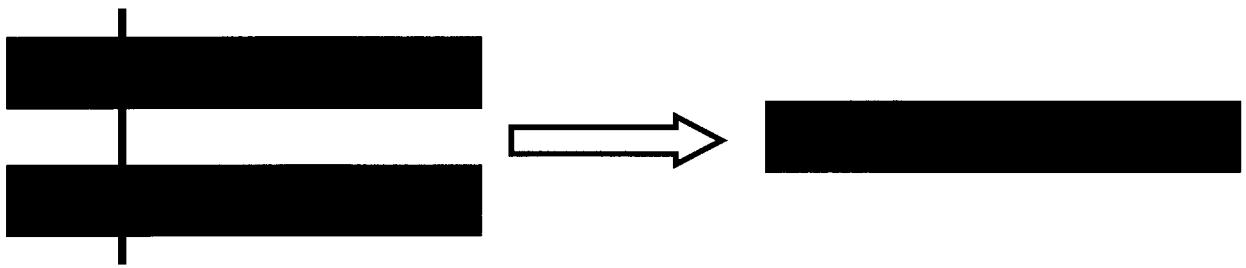
- Mutation scale adapted (co-evolved) with the search parameters
- Idea: multiply scale by a random variable
  - Gaussian: $+\alpha\,\mathrm{N}(0,1)$
  - Lognormal: $\exp(\tau'\mathrm{N}(0,1)+\tau\,\mathrm{N}_i(0,1))$

Sandia National Laboratories

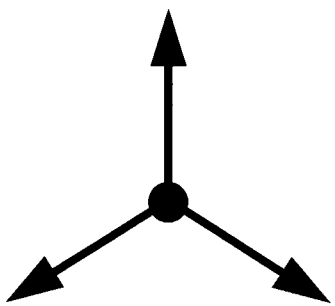
Evolutionary Pattern Search

An adaptive EA

- Evolutionary operators: coordinate-wise crossover and mutation



- Uses a global mutation step scale parameter Δ_t
  - For all individuals
  - For all dimensions
- Uses a fixed pattern of mutation offsets



Sandia National Laboratories

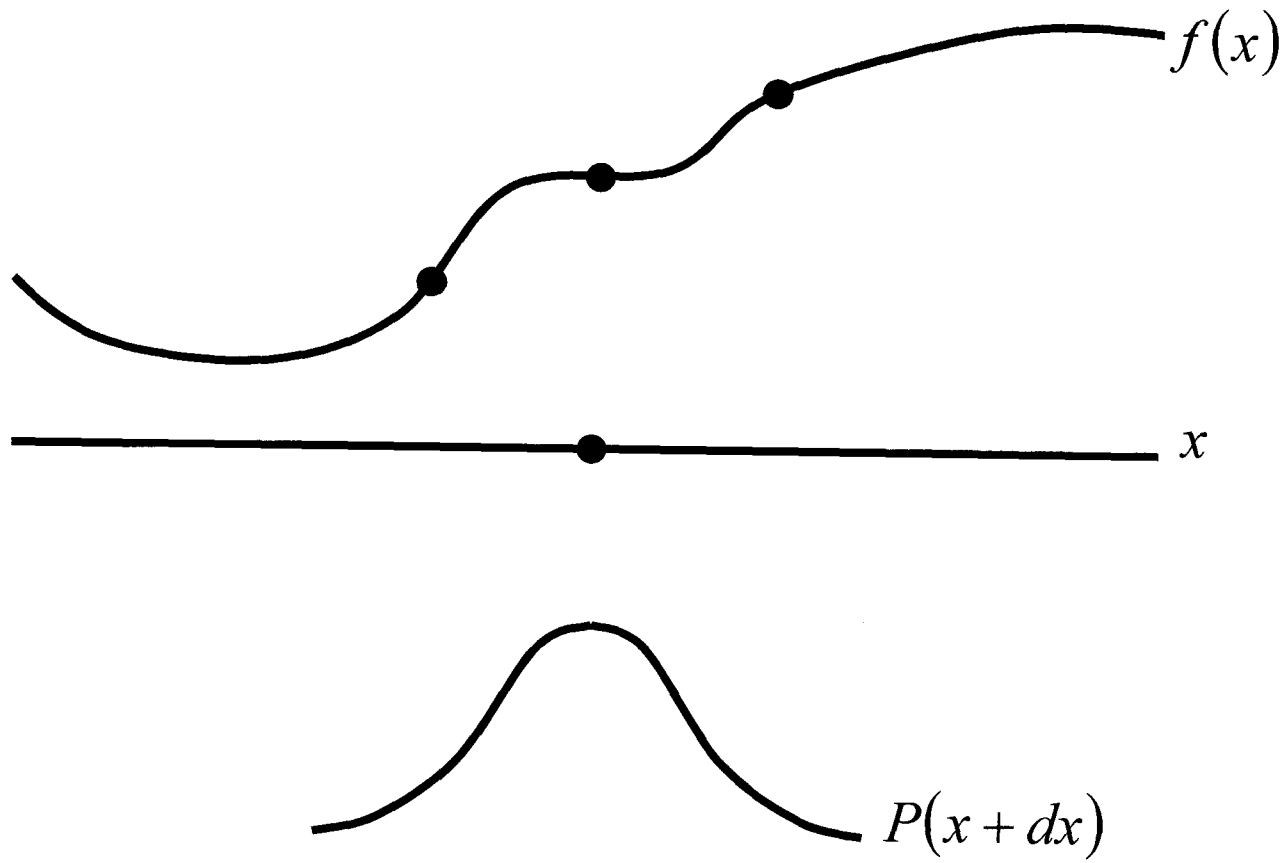
EA Hybrids for Docking

Experimental options

- Cauchy mutation and coordinate-wise crossover
- Infrequent local search
- Local search
  - Solis-Wets
  - Pattern Search
- Fixed local search duration
  - 300 function evaluations
  - 3000 function evaluations
- Step-length initialization
  - Fixed
  - Adaptive

Sandia National Laboratories

Solis-Wets Local Search



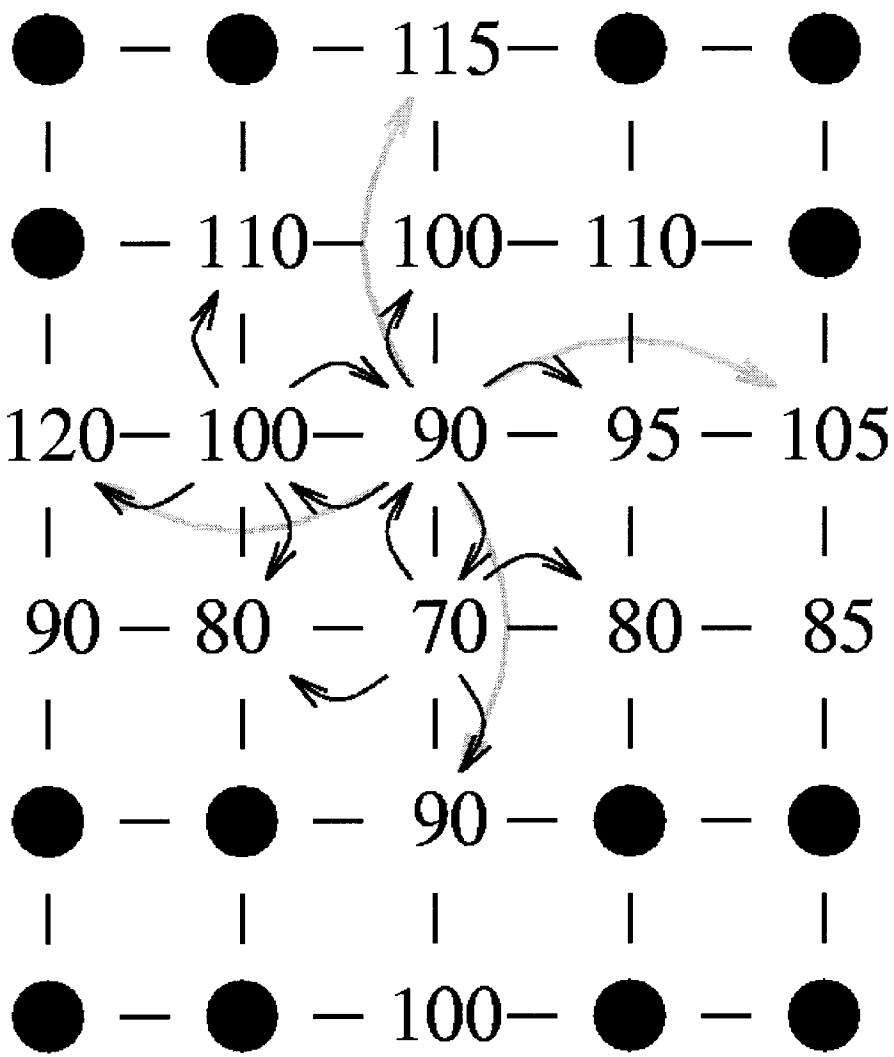
Solis-Wets Local Search (cont'd)

Properties

- Cheap local search (especially in high dimensions)
- Can rapidly adjust the step length parameter
- Convergence properties unclear
- Does not have robust stopping rules

Sandia National Laboratories

Pattern Search Local Search



Deterministic NESW Search

Step 1

Step 2

Step 3

Step 4

Pattern Search Local Search (cont'd)

Properties:

- More expensive search (especially in high dimensions)
- Slowly adjust the step length parameter
- Can guarantee that $P\left(\liminf_{k\to\infty}\left|\nabla f\left(x_k\right)\right|=0\right)=1$ on c.d.f.'s
- Reasonable stopping rules exist

Sandia National Laboratories

Experiments

Six Docking Test Problems

Ligand/Protein Complex	PDB Code	Number of Torsions	Number of Dimensions
β-Trypsin/Benzamidine	3ptb	0	7
Cytochrome P-450cam/Camphor	2cpp	0	7
McPC-603/Phosphocholine	2mcp	4	11
Streptavidin/Biotin	1stp	5	12
HIV-1 protease/XK263	1hvr	10	17
Influenza Hemagglutinin/sialic acid	4hmg	11	18

Sandia National Laboratories

Relative Performance

	Hybrid EA		EPSA 0.0		EPSA 0.5		EP	
	Min	Mean	Min	Mean	Min	Mean	Min	Mean
1hvr	-21.42	-21.40	-21.39	-6.33	-21.23	-10.55	-21.09	3.77
1stp	-10.17	-10.15	-8.89	-5.65	-10.06	-6.09	-9.74	-5.17
2cpp	-7.36	-7.36	-7.35	-7.19	-7.36	-7.22	-7.36	-7.26
2mcp	-5.55	-5.52	-5.34	-3.41	-5.32	-3.43	-5.00	-2.64
3ptb	-8.19	-8.16	-8.19	-7.93	-8.19	-8.08	-8.17	-7.09
4hmg	-7.91	-7.85	-7.00	-3.54	-6.97	-4.55	-7.00	-2.66

Note:

- All EAs were terminated after 1.5M function evaluations
- The EPSAs were also terminated with a step-size criterion

Hart, 99; Hart, Rosin, Belew and Morris, 99

Sandia National Laboratories

Relative Performance (cont'd)

EA hybrids

- Better mean energy than previous methods
- New best energies for 5 of the 6 test problems
- Lower mean rmsd than previous methods (less than one!)

Local search evaluation

- Pattern search more consistent than Solis-Wets
- Long searches particularly important for pattern search
- Initial step length of local search not a significant factor

Sandia National Laboratories

Length of Optimization

Question: Are 1.5M function evaluations really needed?

Stall Statistic: The last iteration where an improvement of more than 0.0001 was found

Results

- 2cpp,3ptb: Terminate within 400,000 function evaluations
- 1hvr,1stp,2mcp,4hmg: Require at least 1M function evaluations

Note: preliminary experiments suggest that a multistart method could be more efficient

Sandia National Laboratories

Future Work

Reconsider problem formulation

- Used constrained optimizers
- Periodic boundary constraints
- Work around singularities in problem representation (Quarternions)
- Rescale search w.r.t. torsion angles (scalability of method?)

Exploit local smoothness

- More rapid convergence
- Guarantee convergence to a local optimum

Sandia National Laboratories

Acknowledgements

Evolutionary Programming

- Kumar Chellapilla (UCSD)

Evolutionary Pattern Search

- Keith Hunter (Sandia)

AutoDock

- Art Olson (TSRI)
- Garret Morris (TSRI)

EA Hybrids

- Rik Belew (UCSD)
- Chris Rosin (TSRI)
- Natalio Krasnogor (UW England)

Sandia National Laboratories