

---

---

# Optimization in the SaAB team

We also do statistics, machine learning, genomics, text processing & system biology

INRA, MIA Toulouse

---

---

# Background: discrete optimization

- using Computer Science, Complexity theory
    - data structures (heaps, suffix arrays, binary decision diagrams...)
  - Graphs, modeling and computing with graphs
    - EuGène is a gene finder based on constrained shortest path + parameter estimation by stochastic optimization (GAs)
    - CarthaGène (genetic mapping) is using the famous Lin-Kernighan algorithm to order markers for building RH and genetic maps
  - Dynamic Programming
    - CarthaGène (genetic mapping) is using an accelerated EM algorithm using a fast dedicated DP algorithm for the E phase
  - (Integer) Linear Programming
    - Optimal Crop allocation with ILP
-

---

# Our specialties (optimization)

- Constraint programming
    - discrete variables
    - constraints (not necessarily linear)
    - A home made solver: MiniCSP
  - Satisfiability (boolean/propositional logic)
    - George participated in “Glucose” design (SAT solver)
  - Graphical model optimization
    - Markov Random Fields (discrete)
    - Bayesian nets
    - Cost Function networks
    - Our award winning solver: toulbar2 (on mulcyber)
-

---

# We love to go from theory to final applications

- Genetics
    - genetic and RH mapping
    - large cyclic pedigree correction
    - haplotype prediction
  - Genomics
    - Gene prediction and genome annotation (pro and eukaryotes)
    - structured RNA gene finding
  - Agronomy
    - Spatio temporal crop allocation (PhD)
  - Structural Biology
    - Global Minimum Energy Conformation (PhD, PostDoc)
-

---

# Just a glimpse on recent results

- Given a (rigid) protein backbone
- Find the minimum possible energy conformation, allowing side chains to move and amino-acids to mutate
- Targets synthesis of new enzymes & nano technologies
- Huge non linear problems (search space >  $10^{200}$ )
- Famous US molecular modelling suite: Rosetta
- Uses “Biased Monte Carlo” (aka Simulated annealing)

Toulbar2 is able to prove the incapacity of Rosetta SA to get close to the optimal solution...even after 2 years of CPU.

---

# Guaranteed Discrete Energy Optimization on Large Protein Design Problems

David Simoncini,<sup>†</sup> David Allouche,<sup>†</sup> Simon de Givry,<sup>†</sup> Céline Delmas,<sup>†</sup> Sophie Barbe,<sup>‡,§,⊥</sup>  
and Thomas Schiex<sup>\*,†</sup>





