A Probabilistic Approach to Description of Molecular Biological Processes on DNA and Their Object Oriented Simulation

U. Hatnik, T. Hinze, M. Sturm

Dresden DNA Computation Group

email: dnacomp@tcs.inf.tu-dresden.de

www: http://wwwtcs.inf.tu-dresden.de/dnacomp

Contents

State of the Art in DNA Computing

vision

- establish a universal biocomputer in theory and laboratory
- biocomputer based on a formal model should feature by
  - computational completeness (universality), reliability
  - high operational speed using massive data parallelism
  - high storage capacity and density, persistence of stored data
  - DNA reusability, energy efficient processing without mech. wear

challenges

- making DNA operations error resistent reducing side effects
- bridging the gap between formal models of DNA computing and lab-reality $\longrightarrow$ our approach

Gap between Models and Lab-Reality

side effects of DNA operations

- not controllable, unreproducible, stochastically occurring effects of molecular biological processes used as DNA operations
- can sum up in sequences of DNA operations
- lead to unexpected, unprecise, unreproducible or even unusable final results of experimental DNA computations

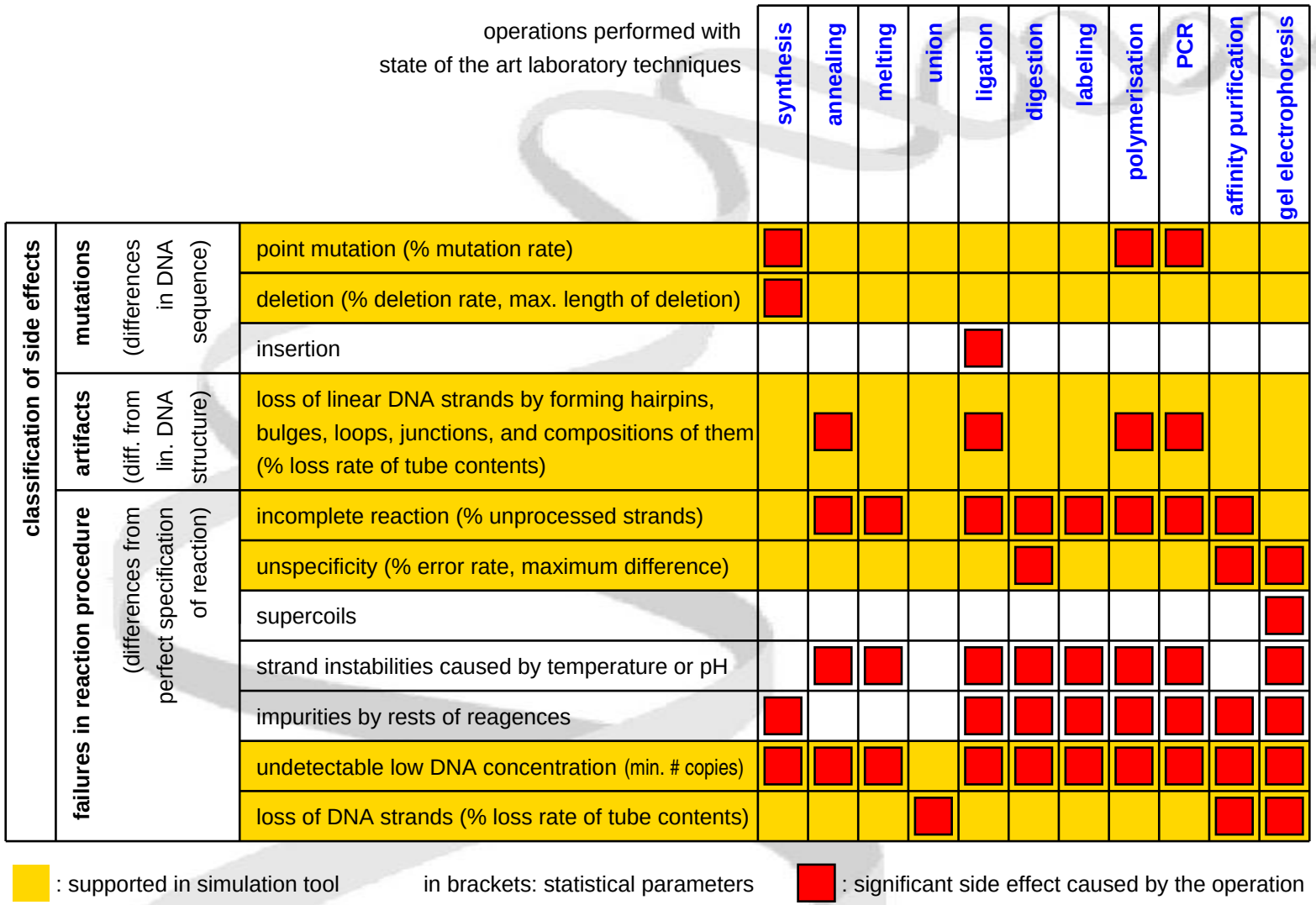
frequently used abstractions of formal models of DNA computing

- only linear DNA used as data carrier (words of formal languages)
- unrestricted approach; arbitrary (∞) number of strand copies
- unique result strands detectable absolutely reliable
- all DNA operations performed completely and reproducible

idea to bridge the gap

- specification of DNA operations on molecular level
- include side effects specified by statistical parameters into the description of DNA operations $\rightarrow$ probabilistic approach

Side Effects of DNA Operations



A Probabilistic Approach to DNA Computing

properties

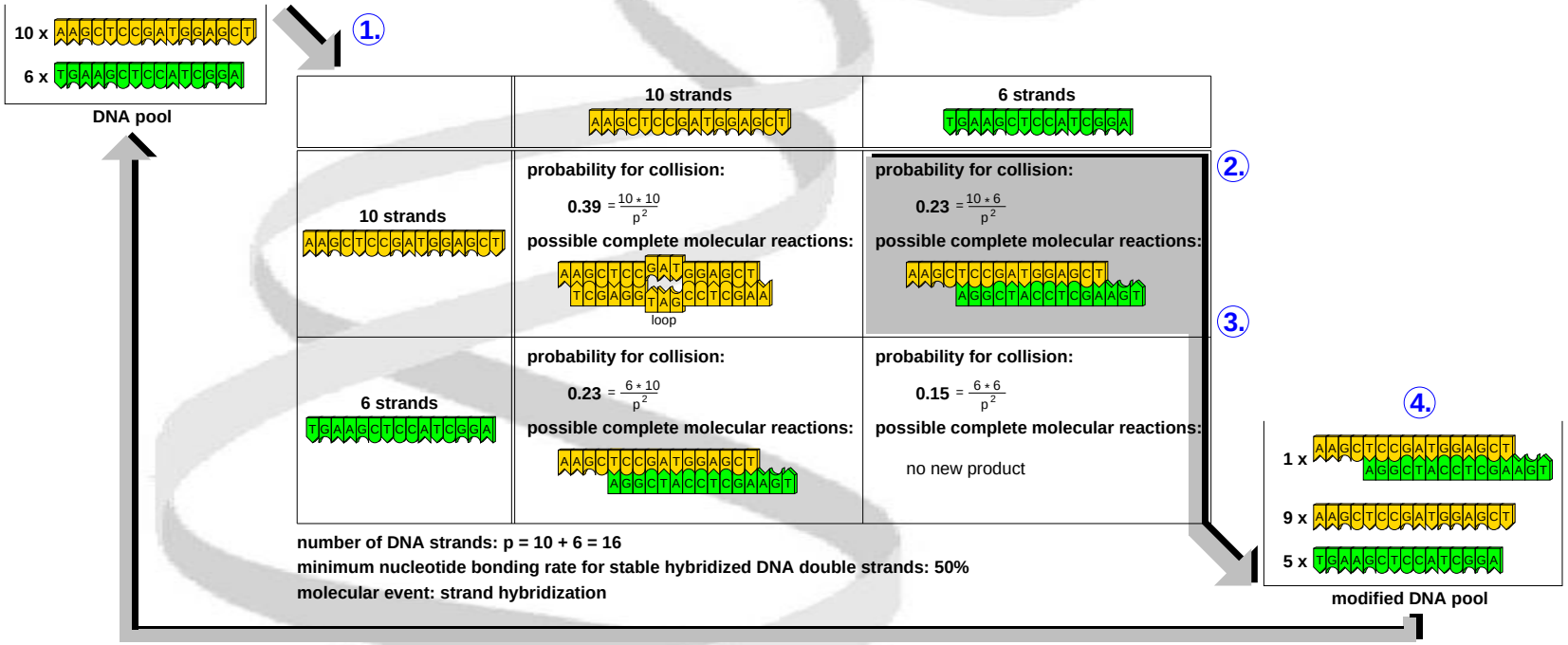
- multiset based, nondeterministic, restricted model
- description of DNA operations on level of single nucleotides and strand end labels $\rightarrow$ operation param., side effect param.
- recently supported: synthesis, annealing, melting, union, ligation, digestion, labeling, polymerisation, PCR, affinity purification, gel electrophoresis; formal description by prog. language

operation control

- iteration of molecular events, probability-controlled
- probabilities of molecular events depend on: DNA pool, number of strand copies, operation parameters, side effect parameters
- iteration terminates iff empty list (matrix) of possible mol. events

A Probabilistic Approach to DNA Computing

iteration exemplified by annealing (simplified)

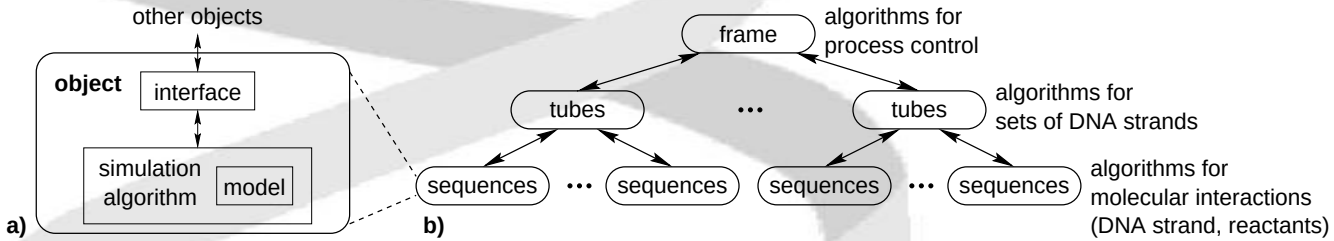


1. Create list (matrix) of molecular events and their probabilities including side effects
2. Select one molecular event randomly with respect to the probability distribution
3. Determine all possible reaction products from this molecular event and select one of them
4. Modify DNA pool

An Object Oriented Simulation Tool

main features

- specification of DNA operations on the level of single nucleotides and strand end labels using probabilistic approach
- number of strand copies considered → concentrations of different DNA strands and their influence to the behaviour in op. process
- each DNA operation processed inside a virtual test tube collecting a multiset of DNA strands, several test tubes supported
- each DNA operation characterized by a set of specific parameters and side effect parameters
- arbitrary sequences of DNA operations including propagation of side effects can be visualized and logged
- Java, simulation tool requires at least Java Development Kit 2.0



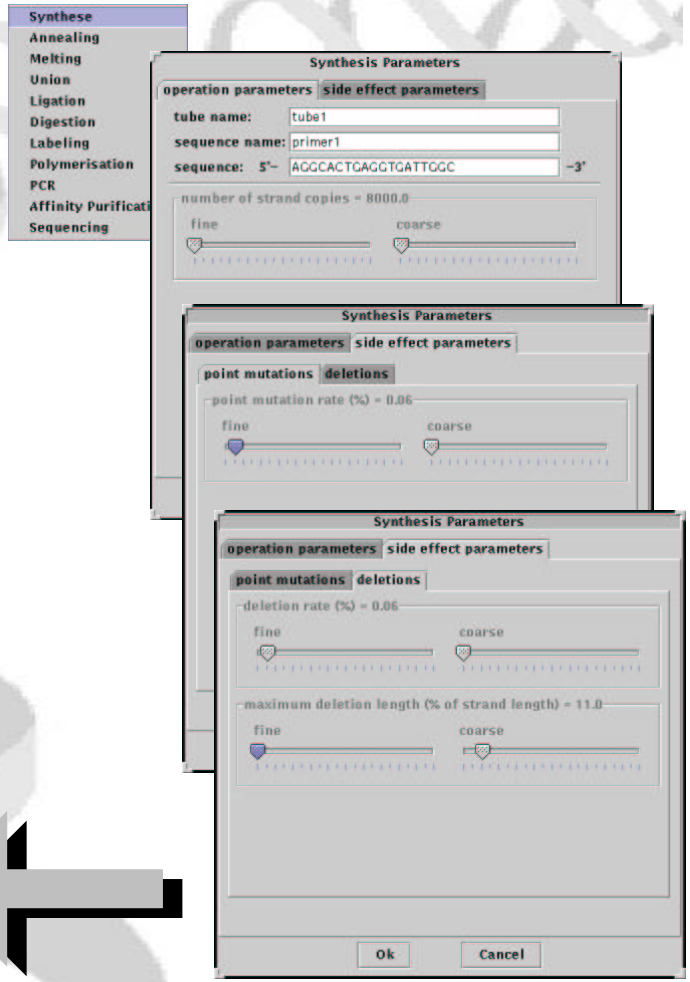
Selected DNA Operations – Synthesis

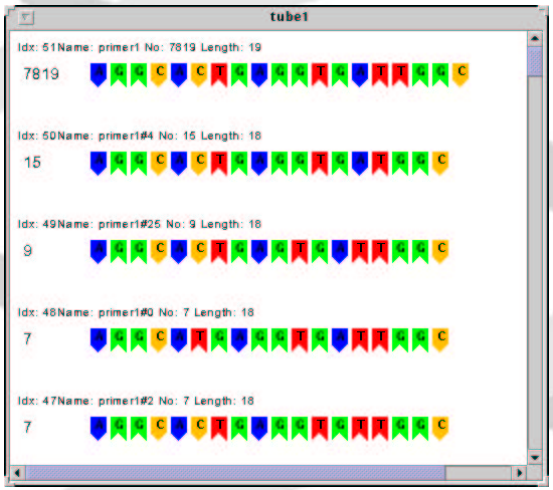
operation parameters:

tube name: tube1
nucleotide sequence (5'-3'): AGGCACTGAGGTGATTGGC
number of strand copies: 8 000

side effect parameters:

point mutation rate: 0.06%
deletion rate 0.06%
maximum deletion length: 11% of strand length





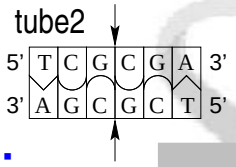
output of test tube contents

Selected DNA Operations – Digestion

operation parameters:

tube name: tube2

recognition sequence and restriction site:

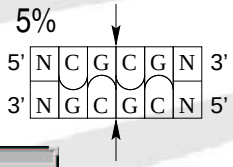


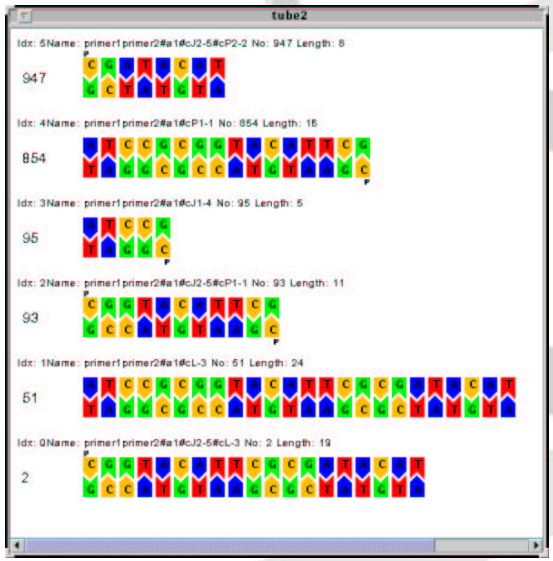
side effect parameters:

rate of not executed molecular cuts: 5%

rate of star activity (unspecificity): 5%

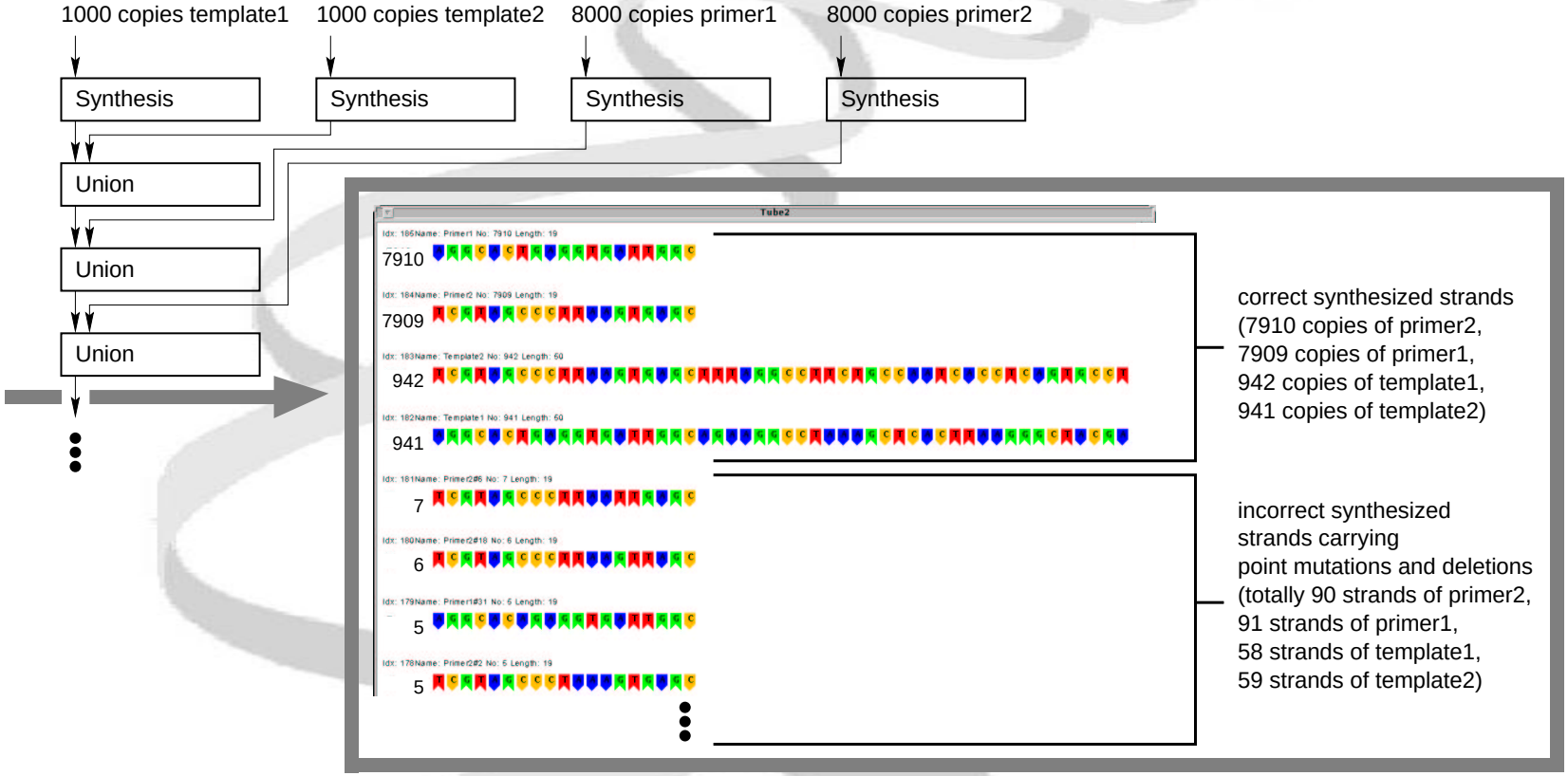
wildcarded recognition sequence:



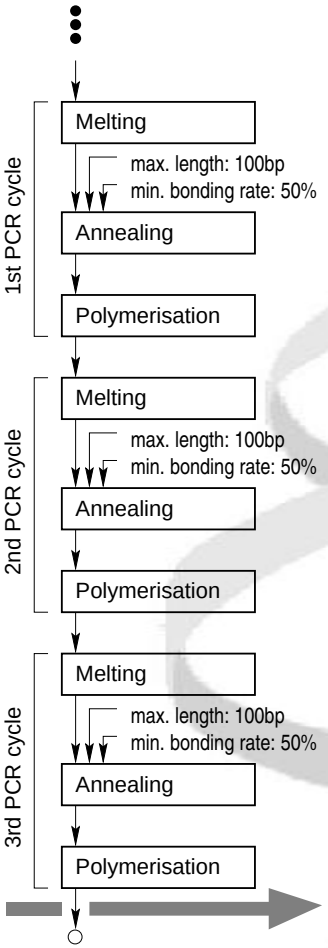


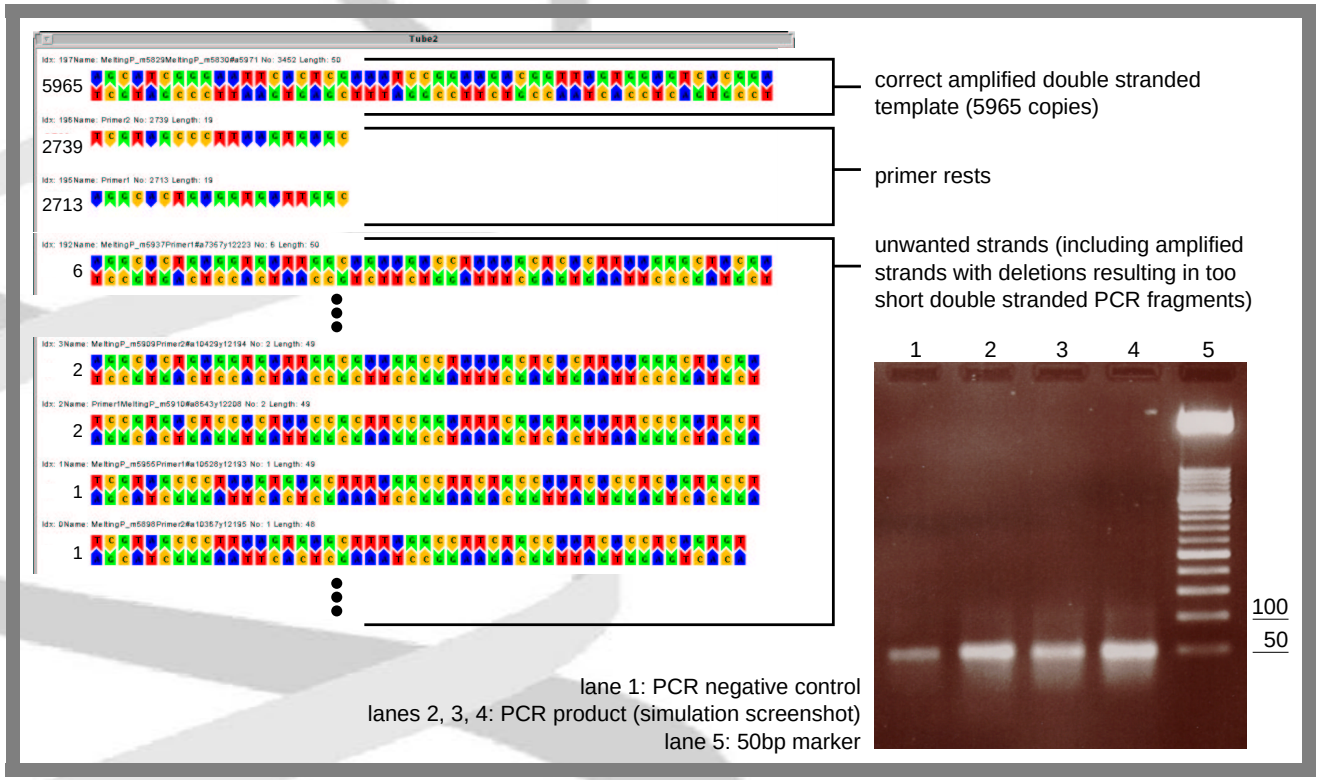
output of test tube contents

A PCR Example (I)



A PCR Example (II)





Conclusions

results

- first formal model of DNA computing considering side effects of DNA operations using a probabilistic, restricted, and multiset based approach
- contribution to bridge the gap between experimental and theoretical DNA computing
- supports experimental setup of DNA algorithms as well as implementations of models for DNA computation
- prediction of experimental results and cost effective optimization of error reducing and error compensating operation sequences
- object oriented simulation tool based on this approach enables a flexible, interoperable, and ergonomic model handling

further work

- extension to additional side effects concerning nonlinear DNA
- application to the implementation of distributed splicing systems