Molecular and Cellular Computing

Lecture series at Universidad Politécnica de Madrid

Martyn Amos
Department of Computing and Mathematics
Manchester Metropolitan University
United Kingdom

http://www.martynamos.com

Day 2: From *in vitro* to *in vivo*
2. Laboratory implementations

Manchester Metropolitan University

Introduction

- We have seen several theoretical models of DNA computation
- Now we will introduce several laboratory implementations
- We begin with a theoretical model that has been successfully realised using molecules

Manchester Metropolitan University

Tile Assembly Model

- A *constructive* model of DNA computation, based on the principle of *self-assembly*
- Molecular self-assembly gives rise to a vast number of complexes in nature, including crystals and the DNA double helix itself
- The growth of such structures is controlled, at a fundamental level, by *natural* computational processes

Manchester Metropolitan University

TAM

- Due to Rothemund and Winfree, the TAM is a formal model for the assembly of complexes such as proteins or DNA on a square lattice
- Extends the theory of tiling by Wang tiles to encompass the physics of self-assembly
- Within the model, computations occur by the self-assembly of square tiles, each side of which may be labelled

P.W.K. Rothemund and E. Winfree. The program-size complexity of self-assembled squares. Proc. STOC 1999, pp. 459-468.

Manchester Metropolitan University

Tile Labels

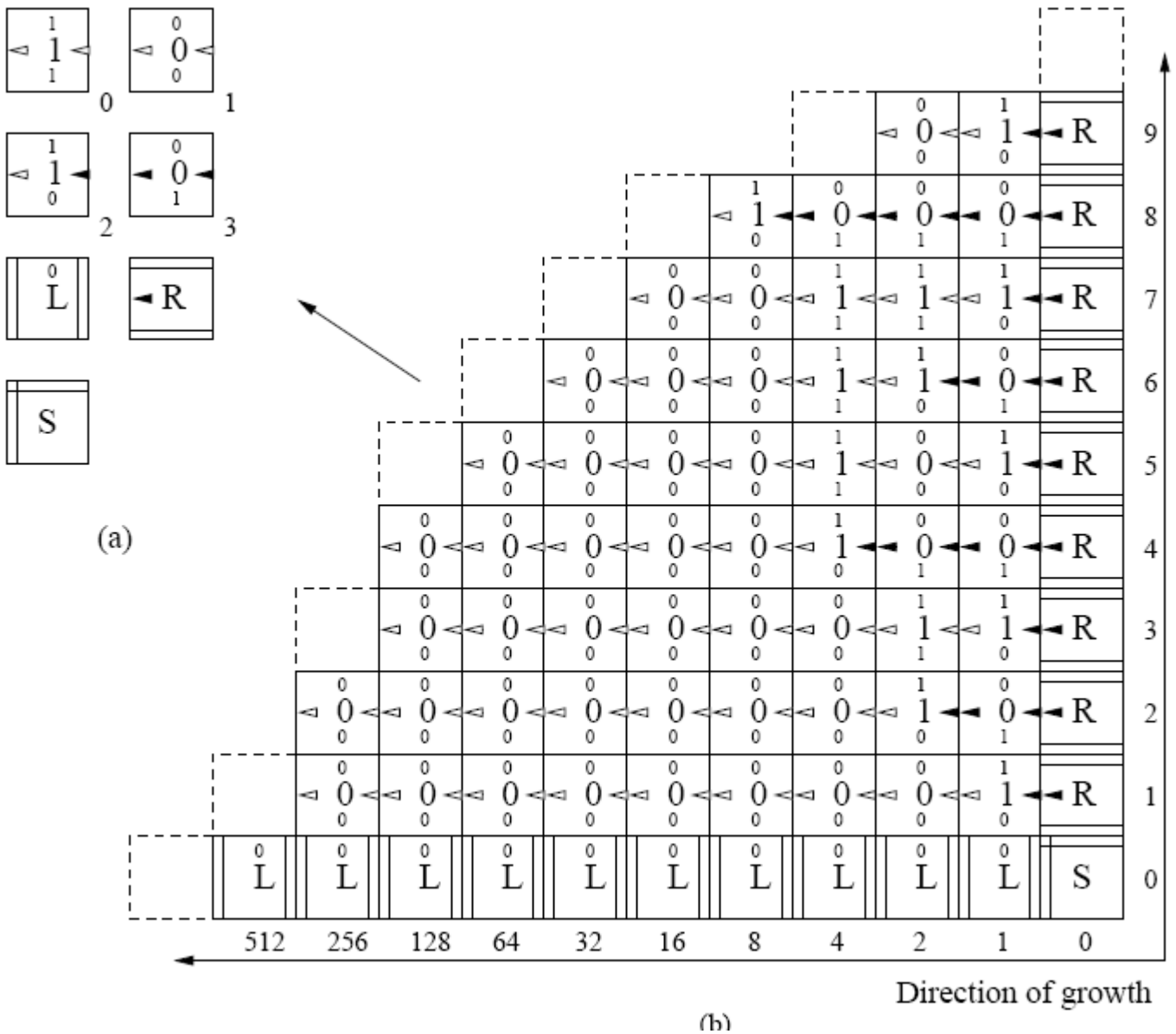
- Different labels represent ways in which tiles may bind together, the strength (or “stickiness”) of the binding depending on the *binding strength* associated with each *side*
- Rules within the system are therefore encoded by selecting tiles with specific combinations of labels and binding strengths
- We assume the availability of an unlimited number of each tile – a computation begins with a specific seed tile, and proceeds by the addition of single tiles
- Tiles bind together to form a growing complex representing the state of the computation only if their binding interactions are of sufficient strength (that is, the complex is stable)

Manchester Metropolitan University

Example – binary counting

- We have seven different tiles, four rule tiles (labelled either “1” or “0”, two border tiles (“L”, “R”) and a seed tile (“S”)
- Sides depicted by a *single* line have binding strength=1, those with a *double* line, of 2
- We impose the following important restrictions:
  - A tile may only be added to the assembly if it is held in place by a combined binding strength of 2
  - A labelled tile (rule) may only be added if the labels on its side match those of its proposed neighbour

Manchester Metropolitan University



Manchester Metropolitan University

TAM in action

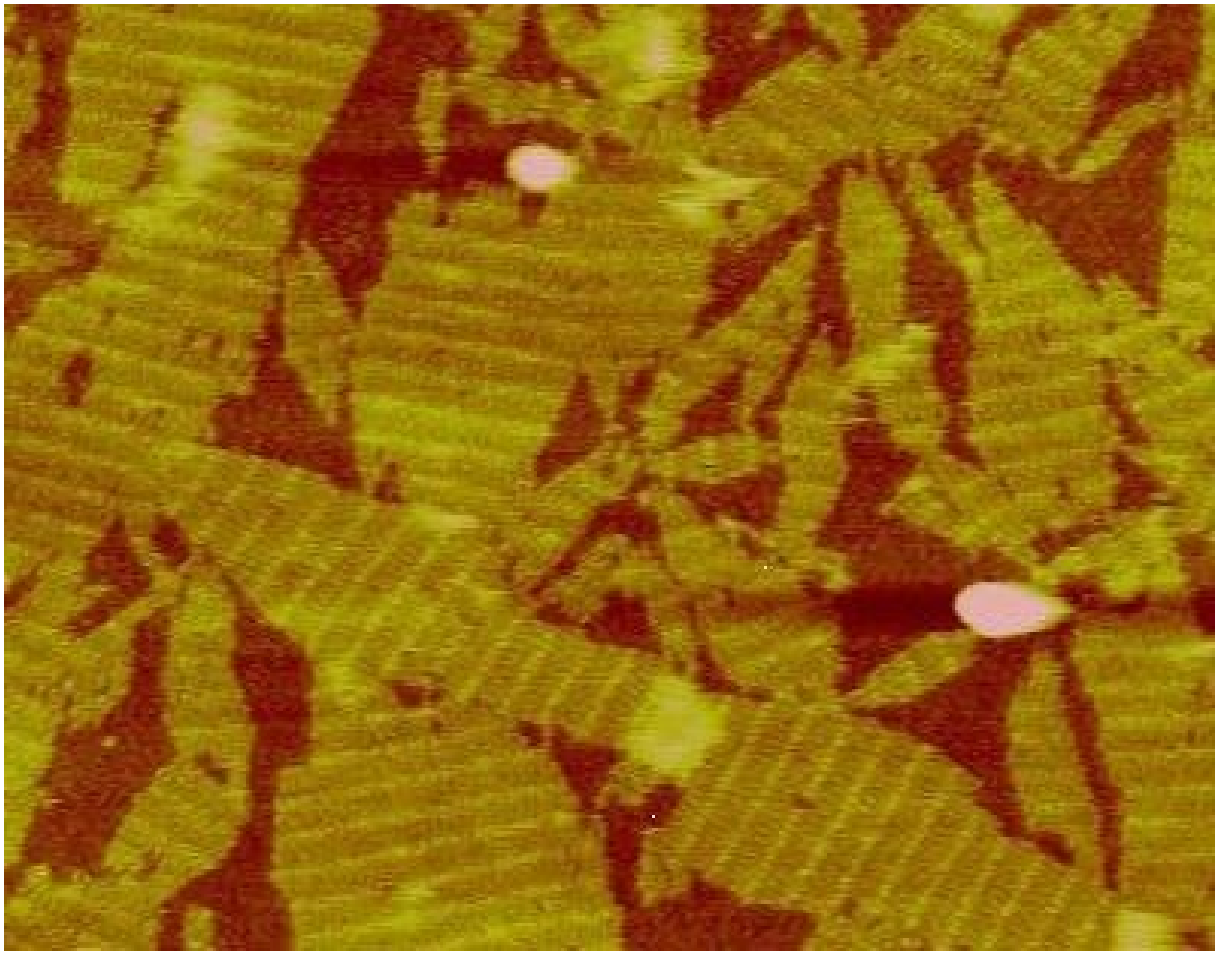
- Two rule tiles cannot bind, since their combined bond strength is only 1
- Crystallized by the seed tile, a “scaffold” of L and R tiles emerges to support the assembly, as a result of the binding strengths
- At the beginning, with one S, one L and one R, the only tile that matches is rule 2, so it is added to the complex, and so on...

Manchester Metropolitan University

TAM in the lab

- Excitingly, the model has actually been demonstrated in the lab
- Details beyond the scope of this lecture, but 2-dimensional “pads”, made up of a number of DNA strands, have been shown to self-assemble into a programmed lattice
- By altering the binding interactions between different molecules, arbitrary sets of tiles may be constructed

Manchester Metropolitan University



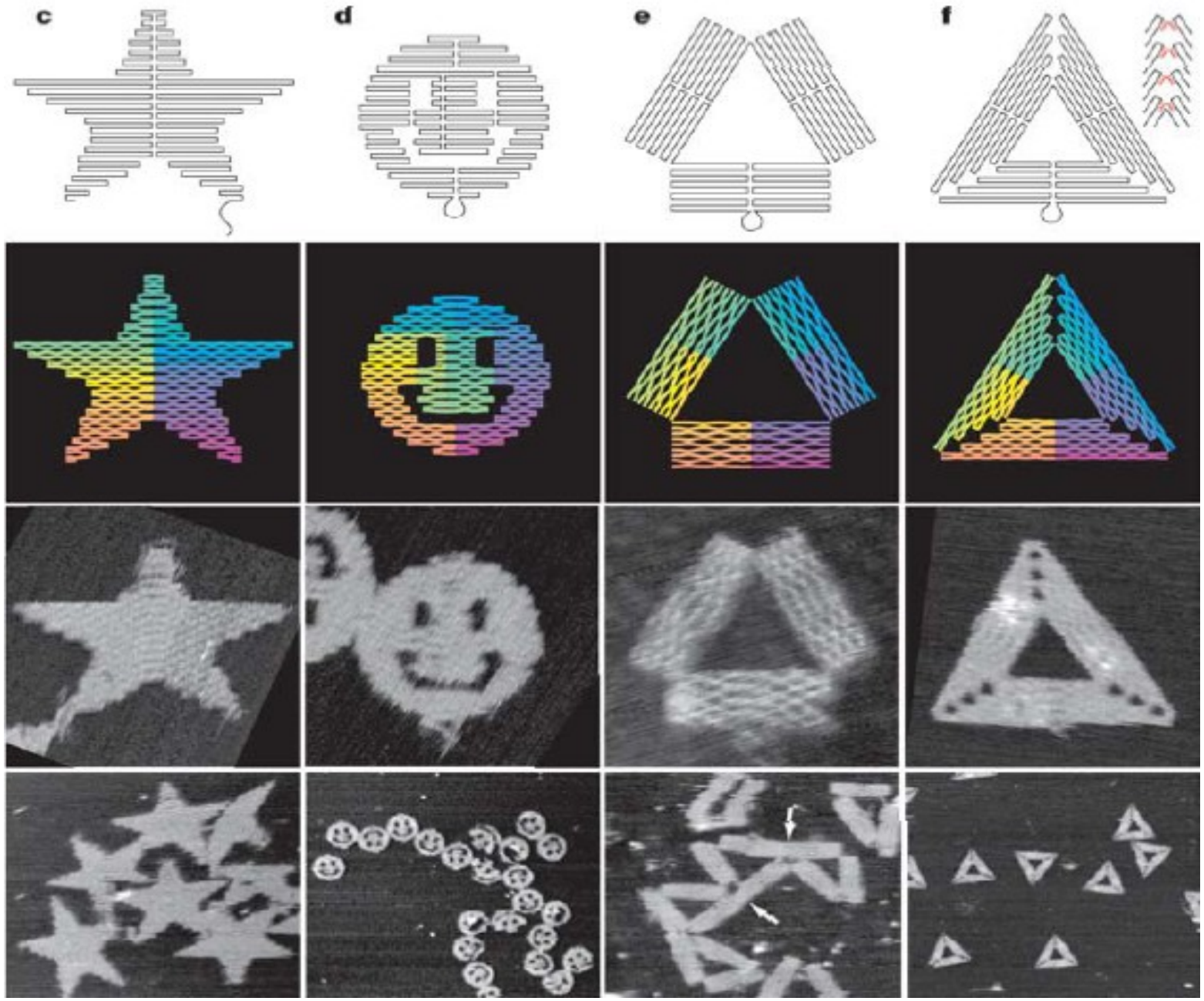
Design and self-assembly of two-dimensional DNA crystals.
Erik Winfree, Furong Liu, Lisa Wenzler, Nadrian C. Seeman.
Nature **394**, 539-544, 1998

Manchester Metropolitan University

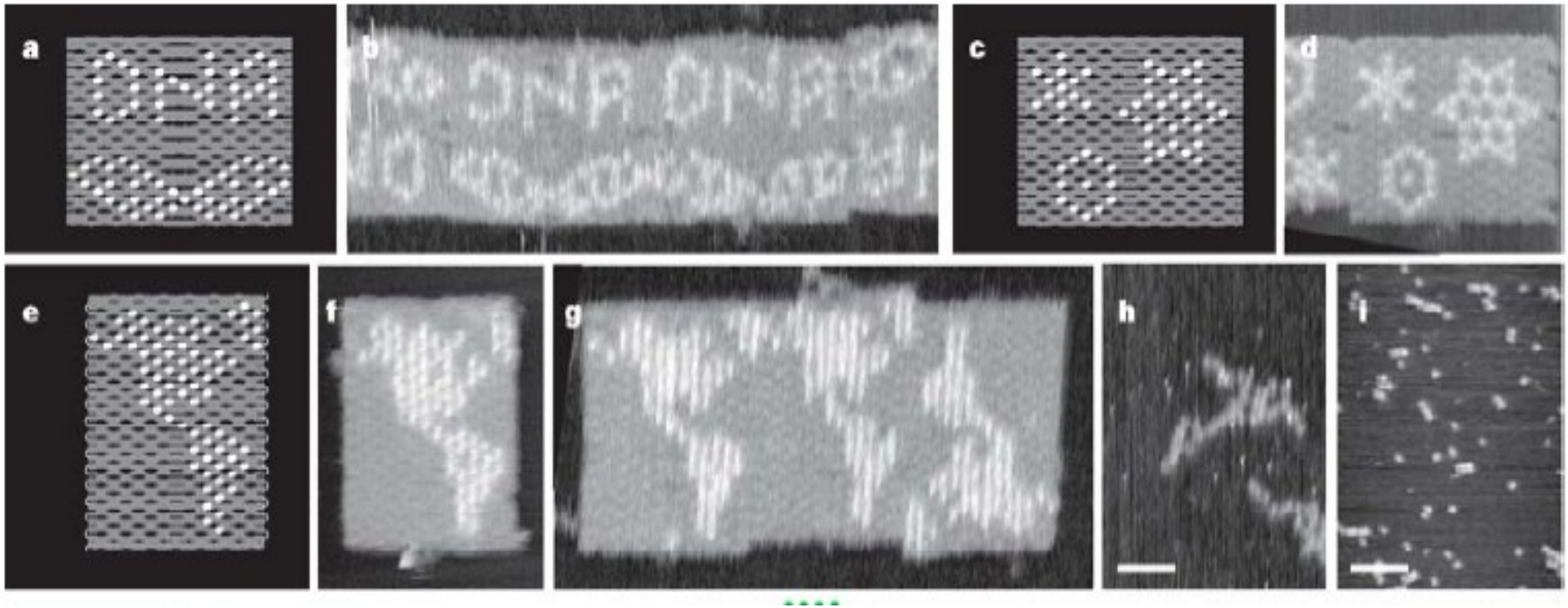
DNA origami

- Takes the entire 7kb single-stranded genome of M13mp18 phage DNA and folds it into a two-dimensional structure
- Uses computer program that takes in a desired shape and calculates position (and therefore sequence) of short “staple” strands that pin the ssDNA in place
- Possible applications as nano-scale “breadboard” for attaching components, or for making nano-wells for reaction vessels

Manchester Metropolitan University



Manchester Metropolitan University

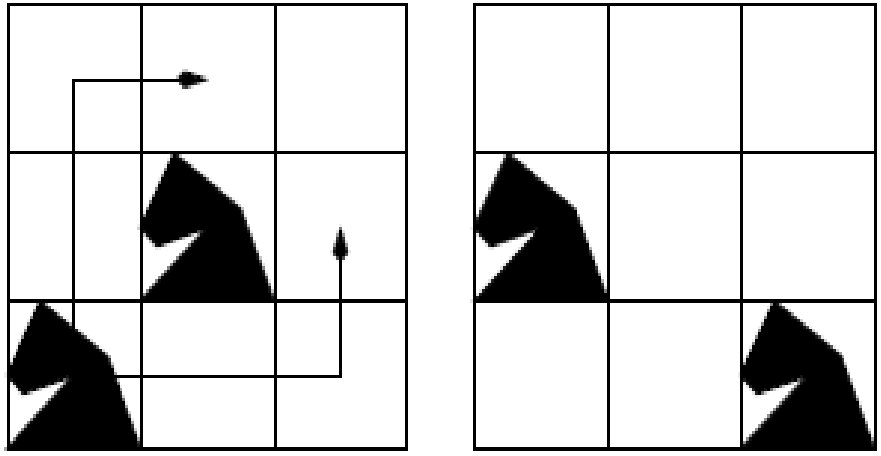


Folding DNA to create nanoscale shapes and patterns. Paul W. K. Rothemund, *Nature* **440**, 297-302, 16 March 2006

Manchester Metropolitan University

Other implementations

- Faulhammer *et al*. described a solution to a variant of SAT that uses RNA rather than DNA
- The so-called “Knight Problem” seeks configurations of knights on an *n*x*n* chess board such that no knight is attacking another



Chess Games

- Uses the “mark and destroy” paradigm, but utilises RNA (see later lecture) plus Rnase H digestion – acts as a “universal restriction enzyme”
- Used a 3x3 board, with variables *a-i* representing the squares (1=knight present, 0=knight absent)
- Problem can be represented as an instance of SAT

simplified (using the rules of Boolean algebra) to ((NOT h AND NOT f) OR NOT a) AND ((NOT g AND NOT i) OR NOT b) AND ((NOT d AND NOT h) OR NOT c) AND ((NOT c AND NOT i) OR NOT d) AND ((NOT a AND NOT g) OR NOT f)

a	b	c
d	e	f
g	h	i

Manchester Metropolitan University

Chess Games

- Construct initial library of all possible 9-bit binary strings
- For each square, sequentially, split the library into two tubes, labelled 1 and 2. After destruction, 1 contains strands encoding a knight at that square, and 2 those strands that do not
- In tube 1, destroy any strands that have *no knight* at that square, as well as any strands that have a knight in an attacking position
- In tube 2, destroy any strands that *have* a knight at that square
- Pool tubes 1 and 2, and repeat for next square
- Any remaining strands encode a legal “board”

Manchester Metropolitan University

Results

- Out of 127 knights placed on 43 boards sampled, only one was placed illegally, giving an overall success rate of 97.7%
- Experiment played a valuable role in getting research over to a general audience

Dirk Faulhammer, Anthony R. Cukras, Richard J. Lipton, and Laura F. Landweber. Molecular computation: RNA solutions to chess problems. *Proceedings of the National Academies of Science*, 97(4):1385–1389, 2000.

Manchester Metropolitan University

Computing on Surfaces

- Another experiment using the mark and destroy paradigm, but this time with strands tethered to a support, rather than being allowed to float freely in solution
- Idea is to simplify the automation of the (potentially very many) repetitive chemical processes required during the performance of an experiment
- Reported a DNA-based solution to a small instance of SAT: (w OR x OR y) AND (w OR NOT y OR z) AND (NOT x OR y) AND (NOT w or NOT y)

Manchester Metropolitan University

Computing on Surfaces

- Strands representing each of the 16 unique assignments were synthesized, and each set was then fixed to a specific region of a gold-coated surface using chemical attachment
- The algorithm proceeds in a round of “mark”, “destroy”, “unmark” operations, the idea being to destroy strands that do not satisfy each clause in turn
- Marked strands are *protected* from destruction by using exonuclease to “chew up” unmarked strands

Manchester Metropolitan University

Results

- Results clearly showed that only legal assignments were retained at the end of the experiment
- There are, however, issues of scalability imposed by the nature of the 2D surface

Liu. Q, et al. DNA computing on surfaces. *Nature* **403**:175-179, 2000

Manchester Metropolitan University

Gel-Based Computing

- A much larger (20 variable) instance of 3-SAT was successfully solved by Adleman's group in 2002
- The largest problem instance successfully solved by a DNA computer; indeed, “this computational problem may yet be the largest solved by nonelectronic means.”

Braich, R.S et al, Solution of a 20-variable 3-SAT problem on DNA computer, *Science* **296**:499-502, 2002

Manchester Metropolitan University

Gel-Based Computing

- Architecture related to the Sticker Model, but only using separation (no application of stickers)
- Probes immobilised in gel-filled modules, and strands are pulled through them by electrophoresis
- Strands are removed (retained in the module) by virtue of their “sticking” to the probes, with other strands free to pass through the module for further processing

Manchester Metropolitan University

Potential Benefits

- Use of electrophoresis *minimizes* the number of (messy, error-prone) laboratory operations (DNA shear, etc.)
- Since strands are not (deliberately) damaged in any way, they, together with the gel modules, are potentially *reusable* for multiple computations
- Whole process is potentially *automatable* – fully integrated DNA computer?

Manchester Metropolitan University

Instance Solved

- 20 variable, 24 clause SAT instance:

$$\Phi = (\neg x_{13} \vee x_{16} \vee x_{18}) \wedge (x_5 \vee x_{12} \vee \neg x_9) \wedge (\neg x_{13} \vee \neg x_2 \vee x_{20}) \wedge (x_{12} \vee x_9 \vee \neg x_5) \wedge (x_{19} \vee \neg x_4 \vee x_6) \wedge (x_9 \vee x_{12} \vee \neg x_5) \wedge (\neg x_1 \vee x_4 \vee \neg x_{11}) \wedge (x_{13} \vee \neg x_2 \vee \neg x_{19}) \wedge (x_5 \vee x_{17} \vee x_9) \wedge (x_{15} \vee x_9 \vee \neg x_{17}) \wedge (\neg x_5 \vee \neg x_9 \vee \neg x_{12}) \wedge (x_6 \vee x_{11} \vee x_4) \wedge (\neg x_{15} \vee \neg x_{17} \vee x_7) \wedge (\neg x_6 \vee x_{19} \vee x_{13}) \wedge (\neg x_{12} \vee \neg x_9 \vee x_5) \wedge (x_{12} \vee x_1 \vee x_{14}) \wedge (x_{20} \vee x_3 \vee x_2) \wedge (x_{10} \vee \neg x_7 \vee \neg x_8) \wedge (\neg x_5 \vee x_9 \vee \neg x_{12}) \wedge (x_{18} \vee \neg x_{20} \vee x_3) \wedge (\neg x_{10} \vee \neg x_{18} \vee \neg x_{16}) \wedge (x_1 \vee \neg x_{11} \vee \neg x_{14}) \wedge (x_8 \vee \neg x_7 \vee \neg x_{15}) \wedge (\neg x_8 \vee x_{16} \vee \neg x_{10})$$

with a unique satisfying assignment of

$$x_1 = F, x_2 = T, x_3 = F, x_4 = F, x_5 = F, x_6 = F, x_7 = T, x_8 = T, x_9 = F, x_{10} = T, x_{11} = T, x_{12} = T, x_{13} = F, x_{14} = F, x_{15} = T, x_{16} = T, x_{17} = T, x_{18} = F, x_{19} = F, x_{20} = F.$$

Manchester Metropolitan University

The Solution

- “Mix and split” technique similar to that used for chess game solution used to generate 300-base library sequence for each of the 1,048,576 possible assignments
- For each clause, a gel module is constructed containing probes designed to capture only those strands that satisfy the clause
- So the first module contains probes to capture X_{13}=F, X_{16}=T, X_{18}=T
- Retained strands act as input to the next module
- The final module contains only strands that satisfy the entire formula

Manchester Metropolitan University

Other Notable Results

- Guarnieri *et al* (1996) described a DNA-based algorithm for binary addition. Used ssDNA to add two non-negative binary numbers. Notable in that it involved DNA for direct numeric representation and addition, as opposed to for massively-parallel random search

All references are in the bibliography supplied with the lectures.

Manchester Metropolitan University

Other Notable Results

- Tendency of DNA molecules to self-anneal was exploited by Sakamoto et al (2000)
- Solved a small instance of SAT by encoding the formula in “literal strings”
- A formula is satisfiable if there exists a literal string that does not contain any variable together with its negation
- If each variable is encoded as a sequence of DNA, with its negation as the WCC, then any strands containing a variable and its negation will self-anneal, forming a “hairpin” structure
- Such structures can be easily distinguished from other strands
- Only requires temperature cycling, no physical manipulation
- But, requires an exponential number of literal strings

Manchester Metropolitan University

Other Notable Results

- Algorithmic self-assembly (as previously described) has been used to build tiled structures that execute four steps of a logical XOR operation (Mao et al, 2000)
- Uses triple-crossover molecules – four strands that interact to form three double helices in a roughly planar structure, with sticky regions
- Successful, but scalability questionable, due to requirement for proper hairpin formation in very long single-stranded molecules

Manchester Metropolitan University

Other Notable Results

- Benenson *et al* (2003) built on their earlier work on the construction of biomolecular machines
- Describe the construction of a molecular automaton that uses the process of DNA backbone hydrolysis and strand hybridization, fuelled by the potential free energy stored in the DNA itself
- “DNA acts as both data and fuel”

Manchester Metropolitan University

Other Notable Results

- Stojanovic and Stefanovic (2003) describe a molecular automaton that plays tic-tac-toe (noughts and crosses) against a human
- Automaton is a collection of deoxyribozymes implementing a Boolean network, arranged in a 3x3 well formation
- Player signals a move by adding a short strand, and the “computer's” move is signalled by fluorescence in a particular well
- Plays a perfect strategy, can never be beaten
- Also see work of Ehud Shapiro and his group