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 1: Molecular Computing
2. The First Experiment

Manchester Metropolitan University

Motivation

- “We...have to shift from electronics and physics to an approach in which chemistry is the fundamental technology. And the most sophisticated chemistry is biochemistry.” Tom Knight, MIT

Manchester Metropolitan University

Visionary

- As is often the case, Feynman was way ahead of his time in suggesting possibility of molecular-level computing
- Technology has lagged behind his vision
- Only realised in 1994, when Len Adleman demonstrated feasibility of computing with DNA molecules

Leonard M. Adleman, Molecular Computation of Solutions to Combinatorial Problems, *Science* **266**, pp. 1021-1024, 1994

DNA – deoxyribonucleic acid

- Ever since the Ancient Greeks, man has recognised that features of one generation are passed on to the next
- It wasn't until Mendel's work on garden peas that it was accepted that both parent contribute material that determines the characteristics of their offspring
- In the early 20th century, it was found that *chromosomes* make up this material
- Chemical analyses of chromosomes showed that they were made up of both *protein* and *deoxyribonucleic acid* (DNA)
- As proteins are "strings" over an "alphabet" of 20 characters, and DNA has only 4 subunits, it was believed that proteins were the genetic carriers
- It wasn't until Watson and Crick deciphered the genetic code that the mechanism of genetic transmission was found to lie with DNA

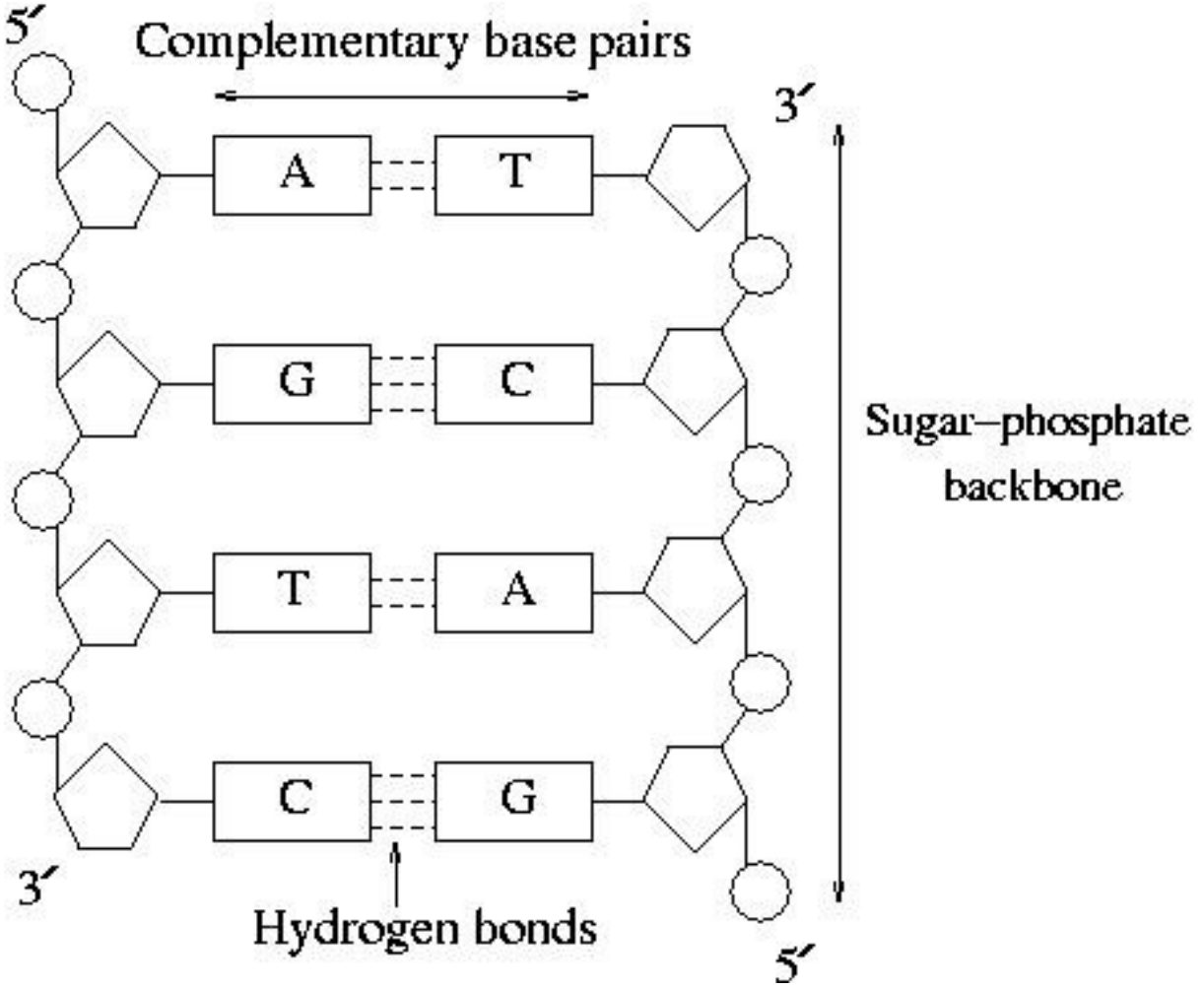
Manchester Metropolitan University

DNA

- DNA encodes the genetic information of cellular organisms
- It consists of chains, or *strands*, of bases
- There are four possible bases: (A)denine, (G)uanine, (C)ytosine and (T)hymine
- Each strand has an orientation, depicted by 3’ and 5’
- The classical *double helix* forms when two single strands bond via hydrogen bonds

Manchester Metropolitan University

The molecule of life



Manchester Metropolitan University

DNA structure

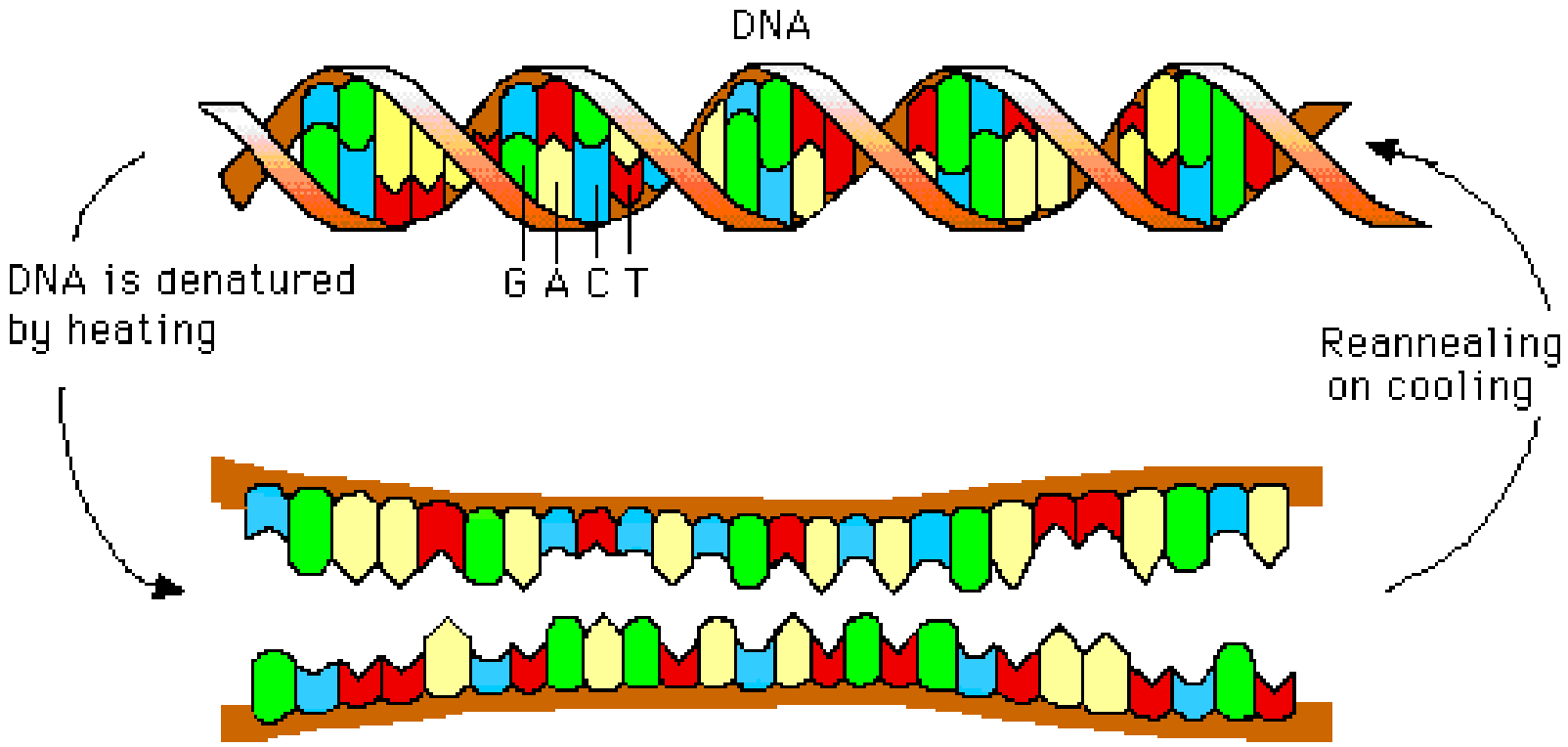
- The key thing to note about the structure of DNA is its inherent *complementarity*
- A binds with T and G binds with C
- One strand is therefore the “mirror image of another”
- Fundamental to its replication
- Complement of AGGCT is TCCGA
- Complement of TAGGA is ATCCT
- Complement of GATTACCA is CTAATGGT

  - *“It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.”*

A Structure for Deoxyribose Nucleic Acid, Watson J.D. and Crick F.H.C., *Nature* **171**, 737-738, 1953

Manchester Metropolitan University

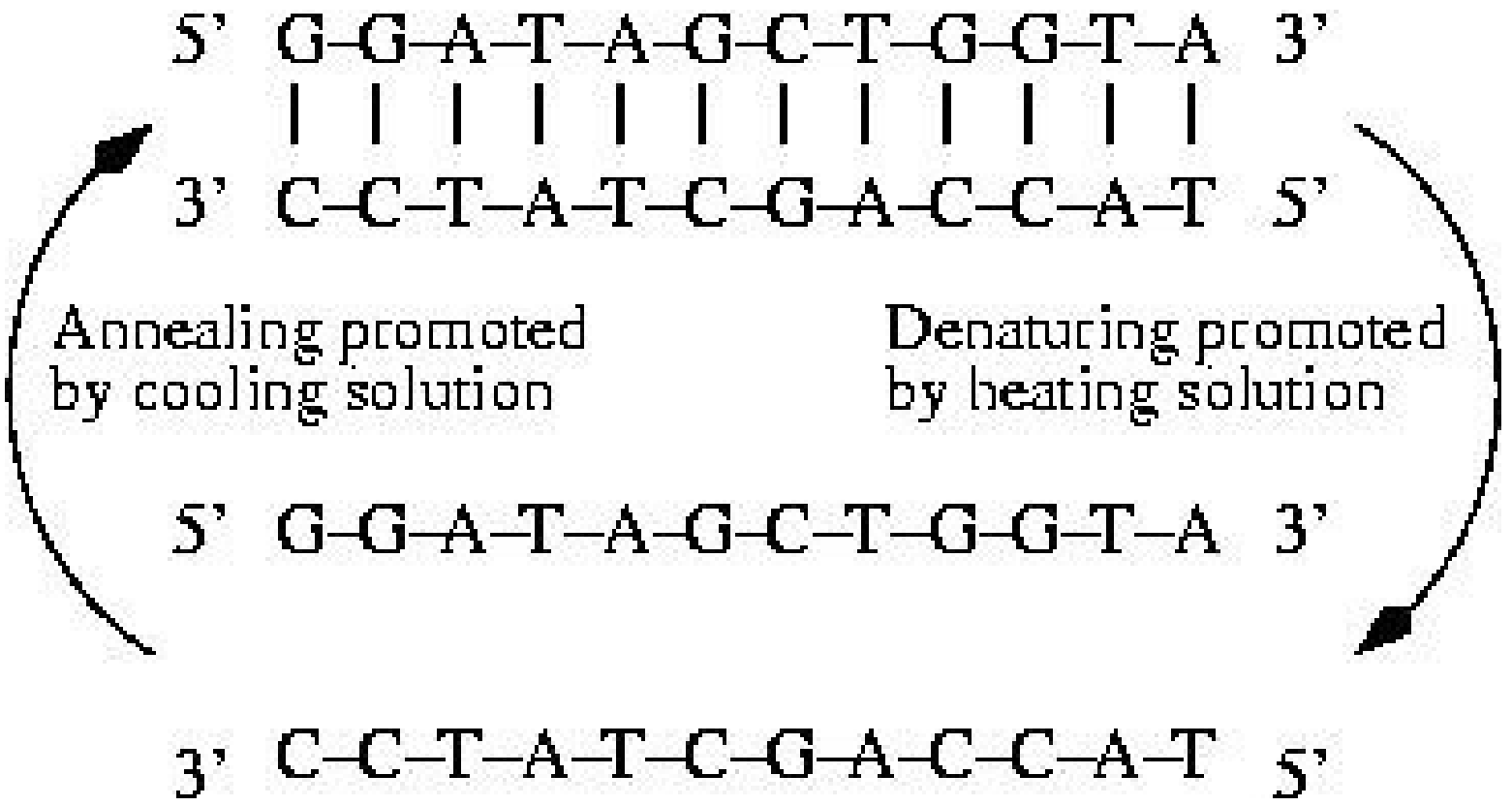
The molecule of life



Access Excellence

Manchester Metropolitan University

Another view



Manchester Metropolitan University

Operations on DNA

- We have at our disposal a wide variety of tools for the manipulation of DNA
- These include synthesising, sorting, chopping, extracting, etc.
- DNA can be made to order in the laboratory – important notion for DNA computation

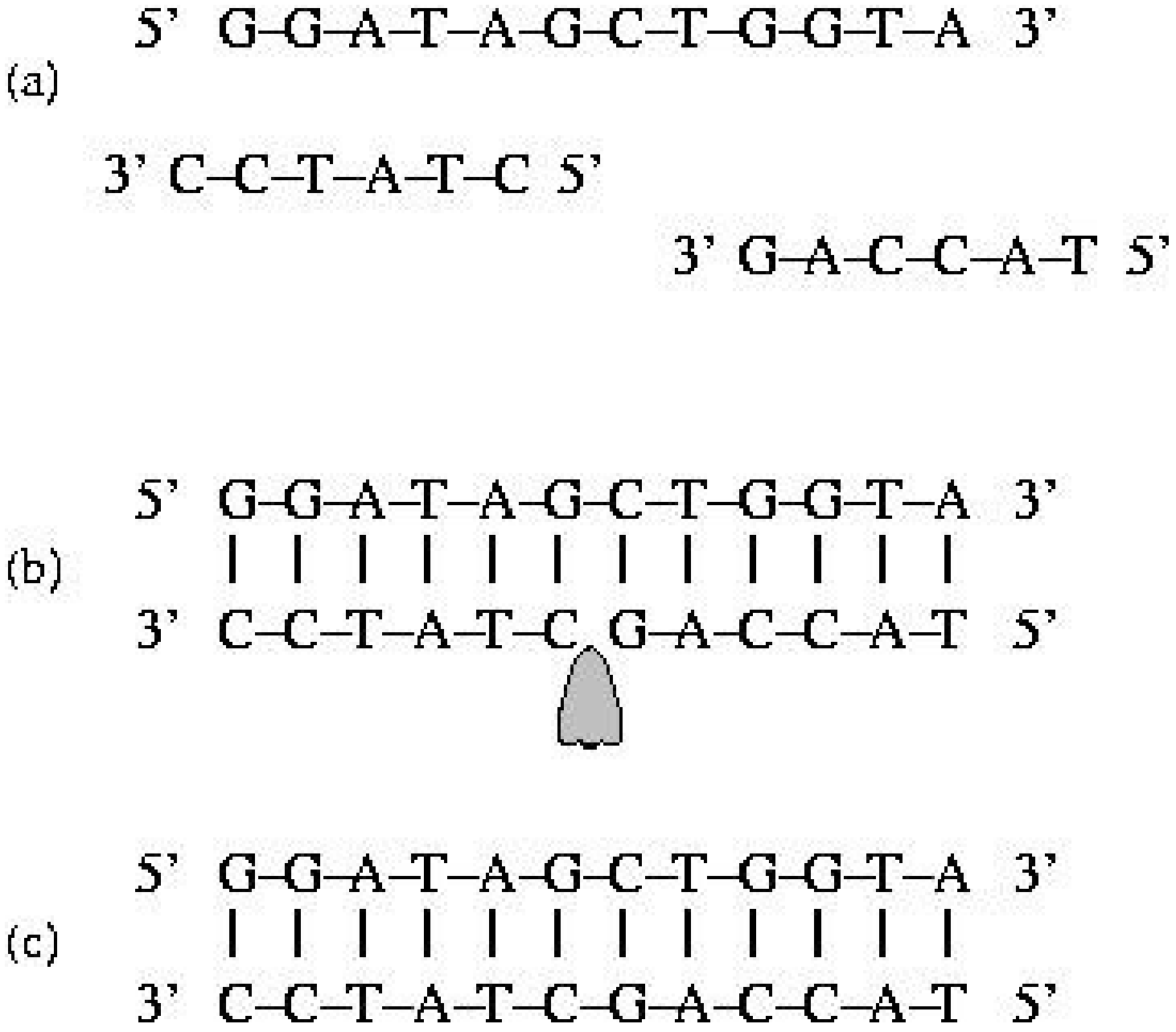
Manchester Metropolitan University

Ligation

- If a single strand contains a “nick” in it, this is known as a *discontinuity*
- Can be repaired by a class of enzymes known as *ligases*
- Allows us to create double-stranded complexes out of several different single strands – important for later

Manchester Metropolitan University

Ligase



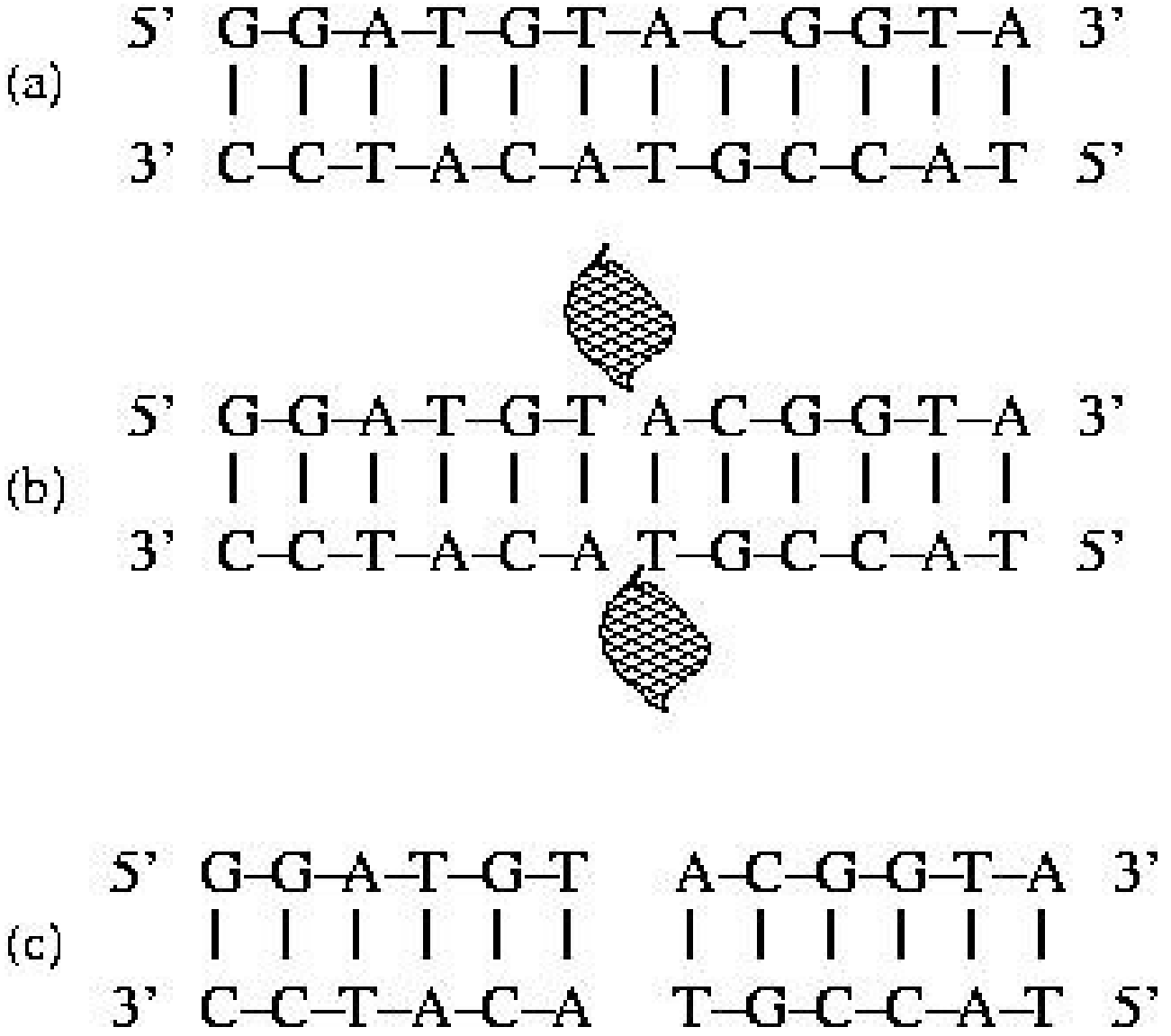
Manchester Metropolitan University

Restriction enzymes

- It is also possible to chop up double-stranded complexes into multiple pieces
- Useful for DNA sequencing and genetic engineering
- Use *restriction enzymes*, which recognise very specific subsequences of DNA, and chop the double-stranded molecule as a result
- Can leave “blunt” or “sticky” ends, depending on the enzyme

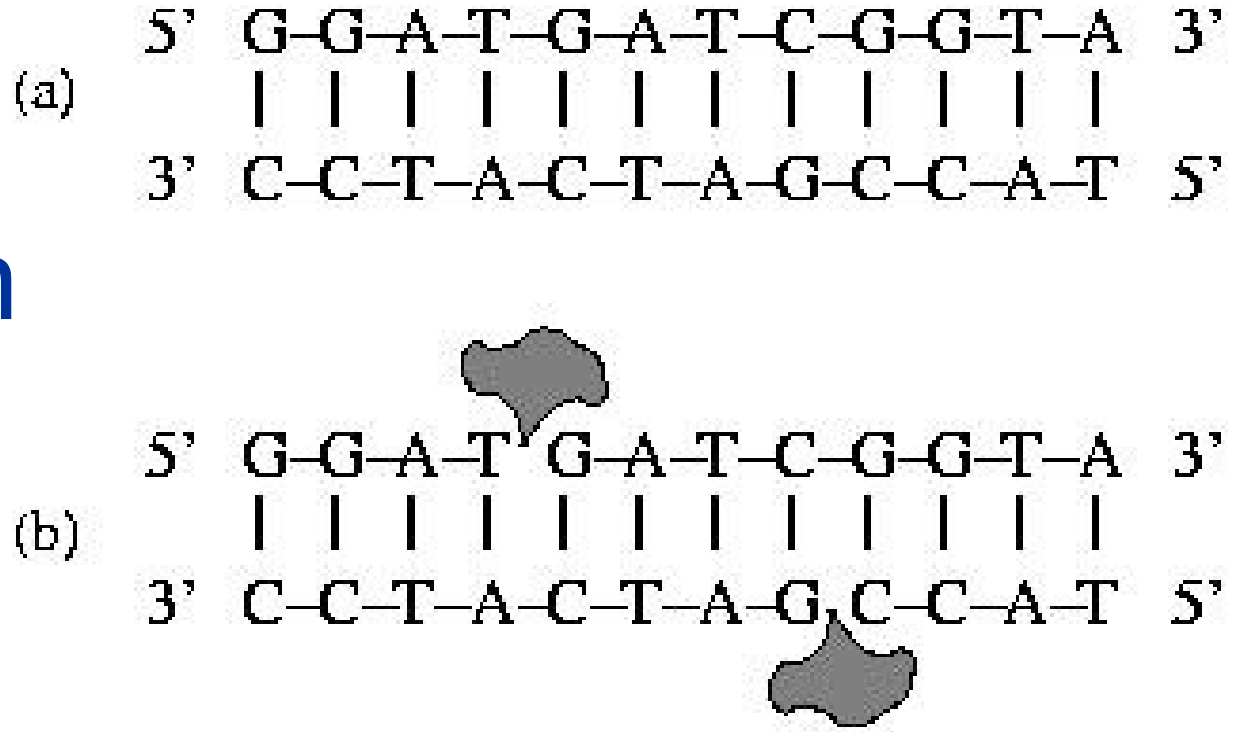
Manchester Metropolitan University

Restriction enzymes



Manchester Metropolitan University

Restriction enzymes

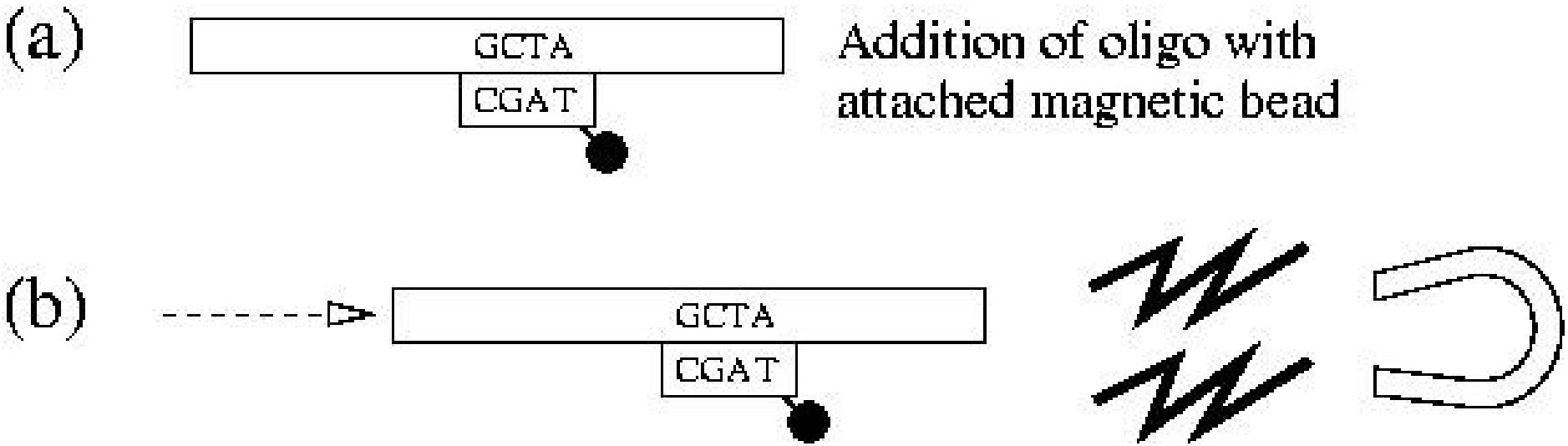


Manchester Metropolitan University

Separation

- It is sometimes useful to extract from a “pot” of DNA strands only those containing a certain sequence
- Rather like doing a Unix “grep”on a file, it only returns the lines of text containing the sequence you’re looking for
- Can be achieved using a technique known as magnetic bead separation, or affinity purification

Affinity purification



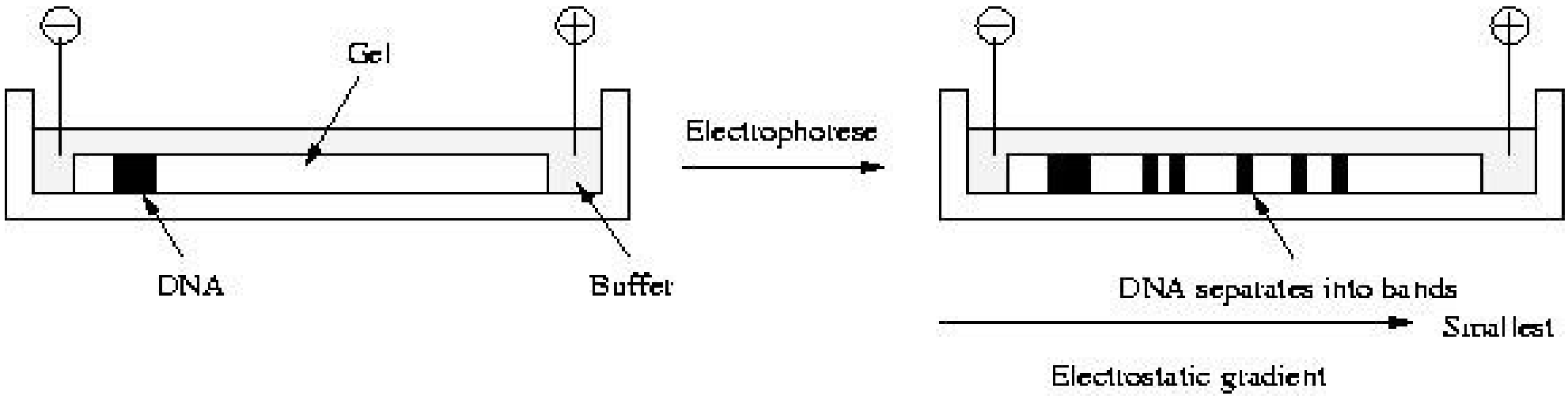
Manchester Metropolitan University

Sorting strands

- Another important technique is used to sort DNA strands on *length*
- Very important in DNA sequencing
- We use a technique known as *gel electrophoresis*
- *Electrophoresis* is the movement of molecules in a charged field
- DNA carries a negative charge, so it tends to be attracted to the anode (positive charge)
- In water, all strands move at the same rate
- In a gel, however, strands move at a rate that is proportional to their length (longer strands move more slowly than short strands)
- This is due to the porous nature of the gel

Manchester Metropolitan University

Gel electrophoresis



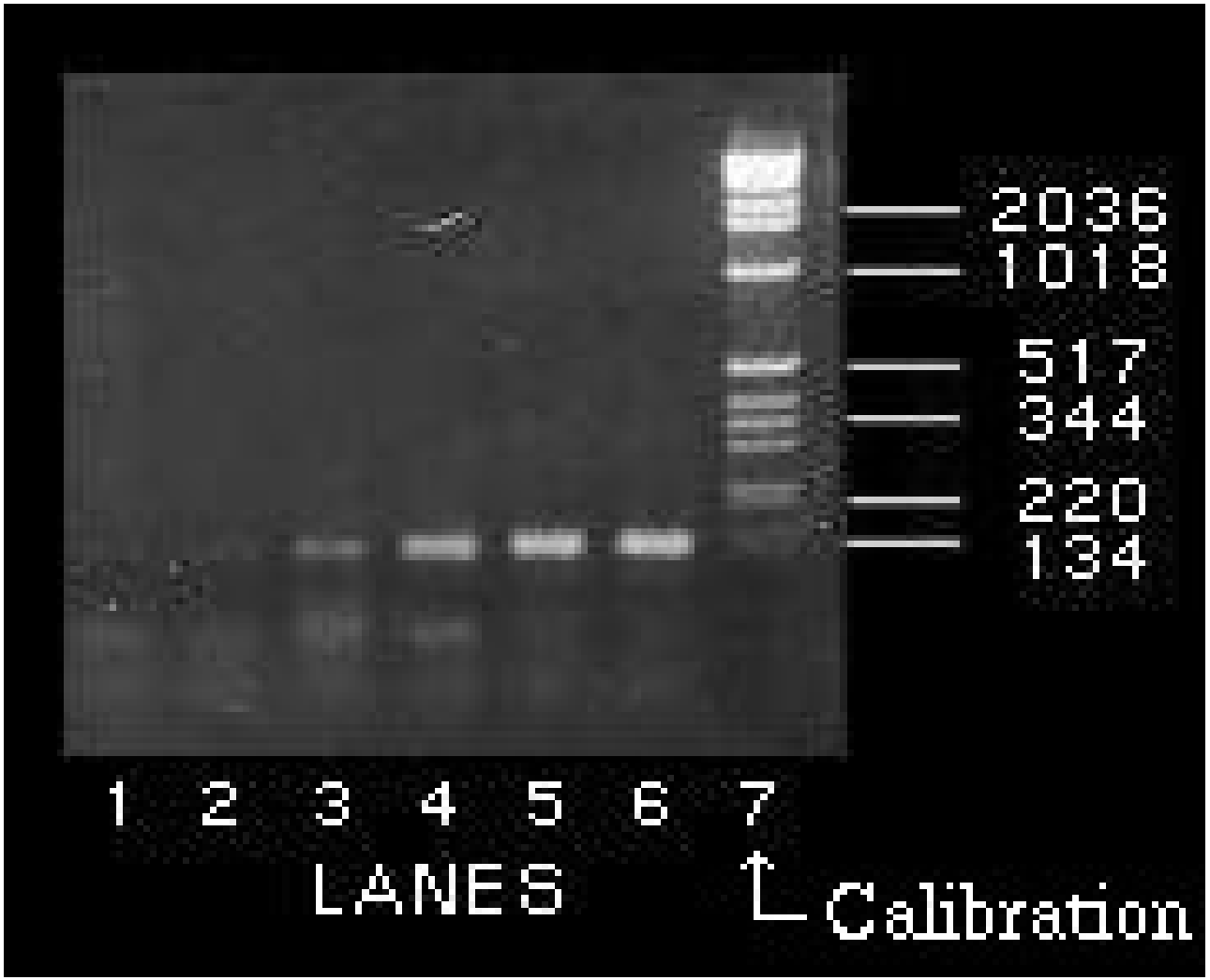
Manchester Metropolitan University

Visualisation

- Once the gel has run (usually overnight), we can stain the different bands of DNA with a fluorescent dye which is visible under UV light
- The gel is then photographed
- We can also cut out bands, thus retaining only DNA of a certain length; this can then be removed from the gel by soaking

Manchester Metropolitan University

Gel visualisation



Manchester Metropolitan University

DNA replication

- DNA can also be replicated, taking a single molecule and multiplying it a thousand-fold (litres, if necessary)
- Useful in forensics, as well as in general molecular biology
- We use a technique known as the *polymerase chain reaction* (PCR)
- Kary Mullis, its inventor, won the Nobel Prize for its discovery
- Uses enzymes known as *polymerases*, which, given an “anchor” point and free bases (“spare nucleotides”), extend the anchor point, creating DS DNA as it goes

Polymerases

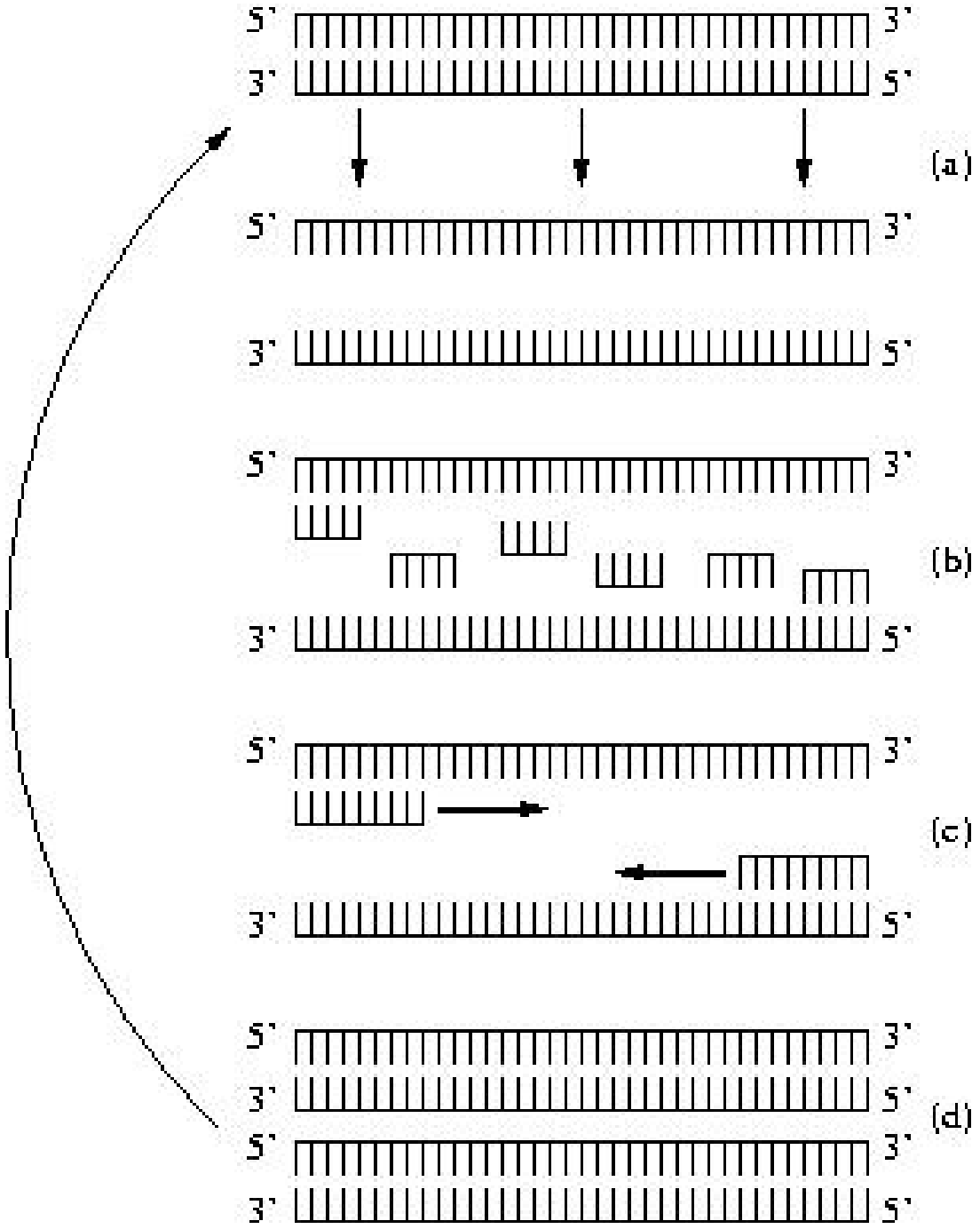
Manchester Metropolitan University

PCR continued

- Each cycle of PCR doubles the amount of target DNA (exponential growth)
- PCR uses polymerase to make copies of a *target sequence* that lies between two *known sequences* within single-stranded DNA
- We create *primers* (anchor points) at the beginning and end of the target sequence
- These are the *complement* of the regions delimiting the target sequence
- We add a large excess of these primers to the solution
- We then add the polymerase, which extends each strand, giving us 2X the target sequence in DS form
- Then heat the solution to break the DS form, giving us single strands, and repeat…

Manchester Metropolitan University

PCR illustrated



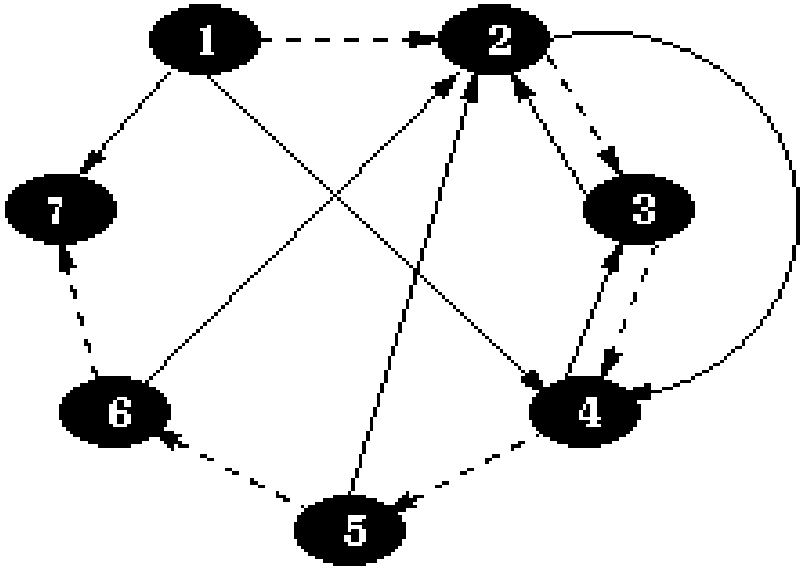
Manchester Metropolitan University

The first DNA computation

- Although Feynman's vision of molecular-scale computing was described in the 1950s, it took a lot longer for the technology to catch up
- Only realised in 1994, with the publication of Len Adleman's article in *Science*
- Paved the way for a whole new research field – annual conference is 14 years old now

The computation

- Adleman solved a small instance of a variant of the Travelling Salesman Problem, the *Hamiltonian Path Problem*
- Given a set of cities connected by roads, is there a tour starting at one city and ending at another that visits each city once and only once?



Manchester Metropolitan University

Complexity

- The HPP is an archetypal *NP-complete* problem
- Such problems are characterised by their having an exponential-sized search space (possible solutions)
- There may be trillions of possible solutions, the vast majority of which are incorrect, but a few of which might be valid

Manchester Metropolitan University

Complexity

- For example, the HPP has 2^n possible solutions, where *n* is the number of cities
- For 8 cities, there are 256 possible solutions
- For 20, there are 1,048576
- For 100, there are 1,267,650,600,228,229,401,496,703,205,376
- This is known as a *combinatorial explosion*

Manchester Metropolitan University

Complexity

- The problem is, we have no known fast algorithms for the NP-complete problems
- If we did, we could solve all NP-complete problems efficiently, as they can all be translated into instances of one another
- P=NP? (that is, can we solve NP-complete problems efficiently?) remains *the* open problem in Computer Science
- Most researchers believe that if a resolution is ever found, it will be negative (that is, we cannot solve these problems efficiently)

Manchester Metropolitan University

Adleman's solution

- The ultimate search for a "needle in a haystack" – generate all possible solutions to the problem, then throw away the ones that fail to meet certain criteria
- This is formally known as a *massively-parallel random search*
- Each possible solution is, in this case, represented as a strand of DNA

Manchester Metropolitan University

Adleman's algorithm

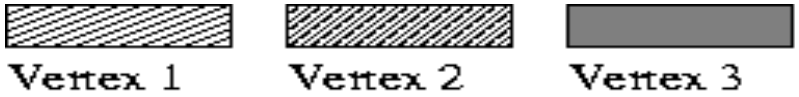
1) Generate strands encoding random paths, such that the HP is represented with high probability (use sufficient DNA to ensure this)
2) Remove all strands that *do not* encode a HP
3) Sequence what is left to discover the result

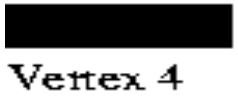
Manchester Metropolitan University

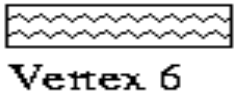
1. Generating paths

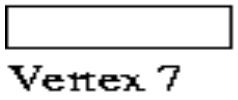
- Each town and city is assigned a unique, 20-base DNA sequence
- Sequences representing “roads” act as splints, binding together sequences representing their end-points (say, A and B) like Lego®
- The road strand is the complement of the *second* half of the strand representing A, followed by the complement of the *first* half of the strand representing B

Generating paths

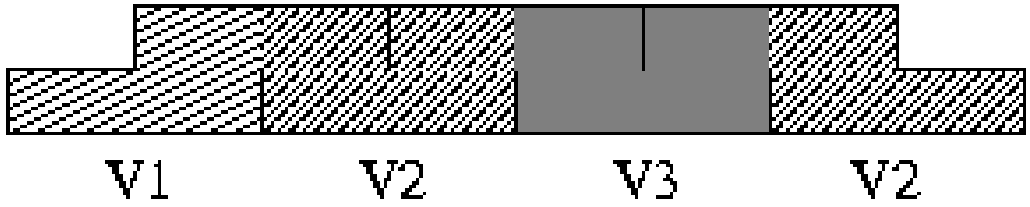


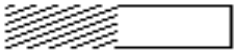


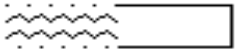




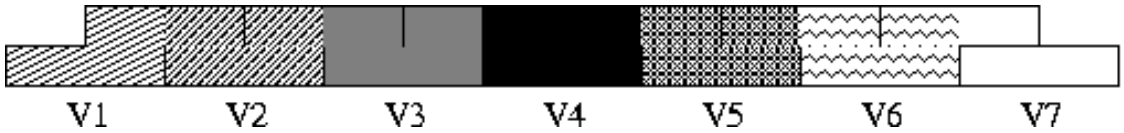
(a)







(b)



Manchester Metropolitan University

Generating solutions

- All strands are then mixed in solution, and they spontaneously bind together to form longer strands
- It is assumed that strand representing the HP is present with very high probability
- This approach solves the problem of generating an exponential number of different strands using a polynomial number of different initial strands

Manchester Metropolitan University

2. Remove illegal solutions

- Remove all strands that do not encode the HP
  - Wrong start/end point
  - Wrong length
  - Cities visited
- We know that the path must start at city 1 and end at city 7
- We therefore massively amplify only those strands that encode solutions that begin with the sequence encoding city 1 and end with the sequence encoding city 7
- How do we achieve this?

Manchester Metropolitan University

Use PCR

- Recall that PCR is used to amplify DNA sequences between two “tagged” regions
- We add strands corresponding to the complementary sequences of cities 1 and 7, then run the PCR
- Anything left is equivalent to slight noise in the system
- Now we have a population of strands encoding solutions that start at 1 and end at 7

Manchester Metropolitan University

Correct length?

- We know that the HP must visit 7 cities once and only once, therefore the path must be 7*20=140 base pairs long
- Any more, we must have visited a city twice, any less and we must have missed out a city
- How do we achieve this?

Manchester Metropolitan University

Use gel electrophoresis

- Run the solution of strands in a gel, along with a marker tube to distinguish strands of length 140
- These strands can then be removed from the gel
- We now have a population of strands that encode solutions starting and ending at the right place, and which visit exactly 7 cities

Manchester Metropolitan University

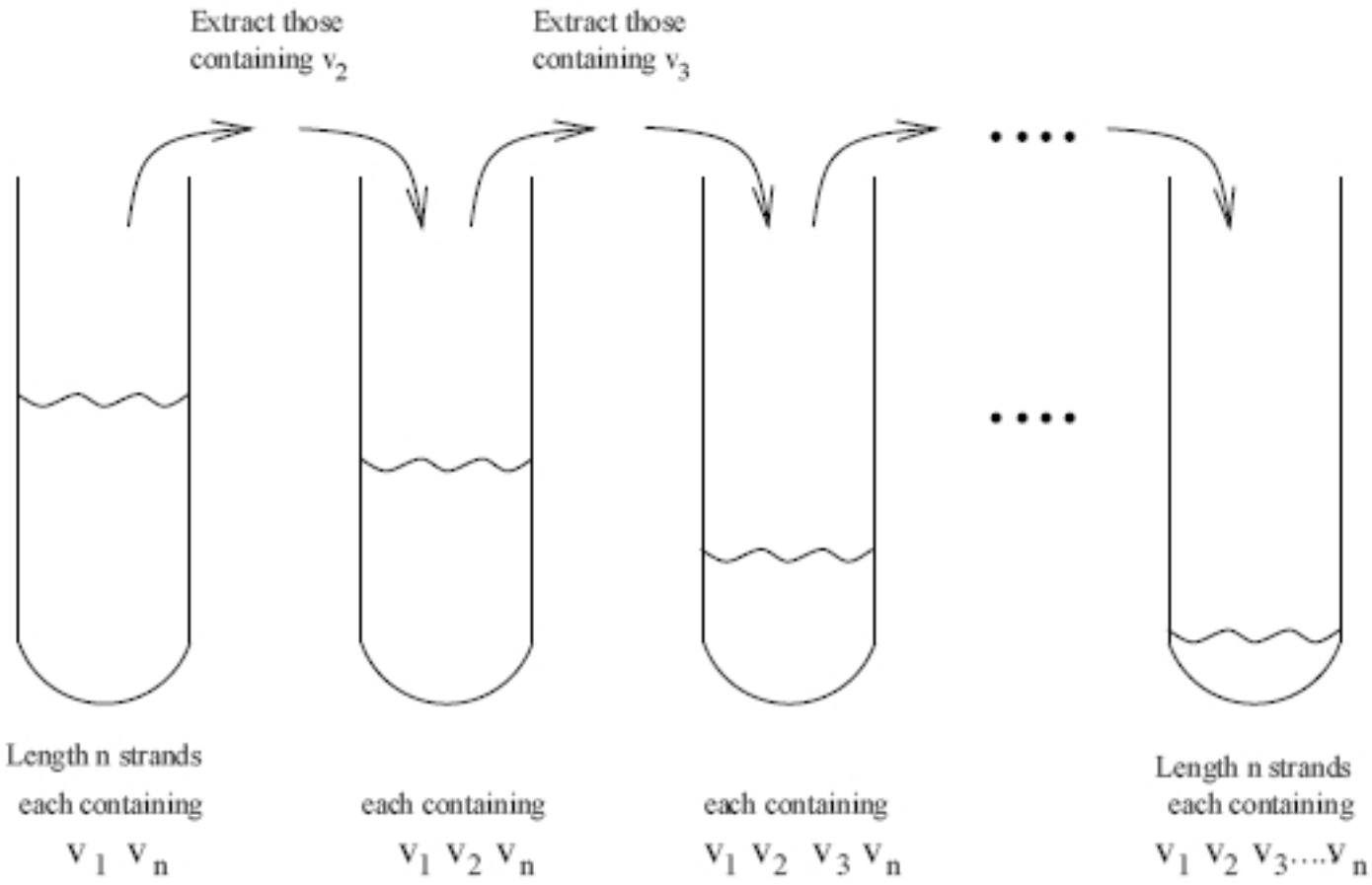
All cities visited?

- We now need to ensure that each city is visited at least once
- How do we achieve this?

Use magnetic bead separation

- We perform a series of extractions, one each for cities 2, 3, 4, 5 and 6
- We already know that cities 1 and 7 are represented
- Use the complementary sequence of city 2 as the “bait” on the magnetic fishing line, remove strands containing that sequence, and then use only those strands for the next separation
- Continue with an ever-shrinking tube until we are left only with strands that contain the sequences for cities 2, 3, 4, 5 and 6 (in addition to 1 and 7)

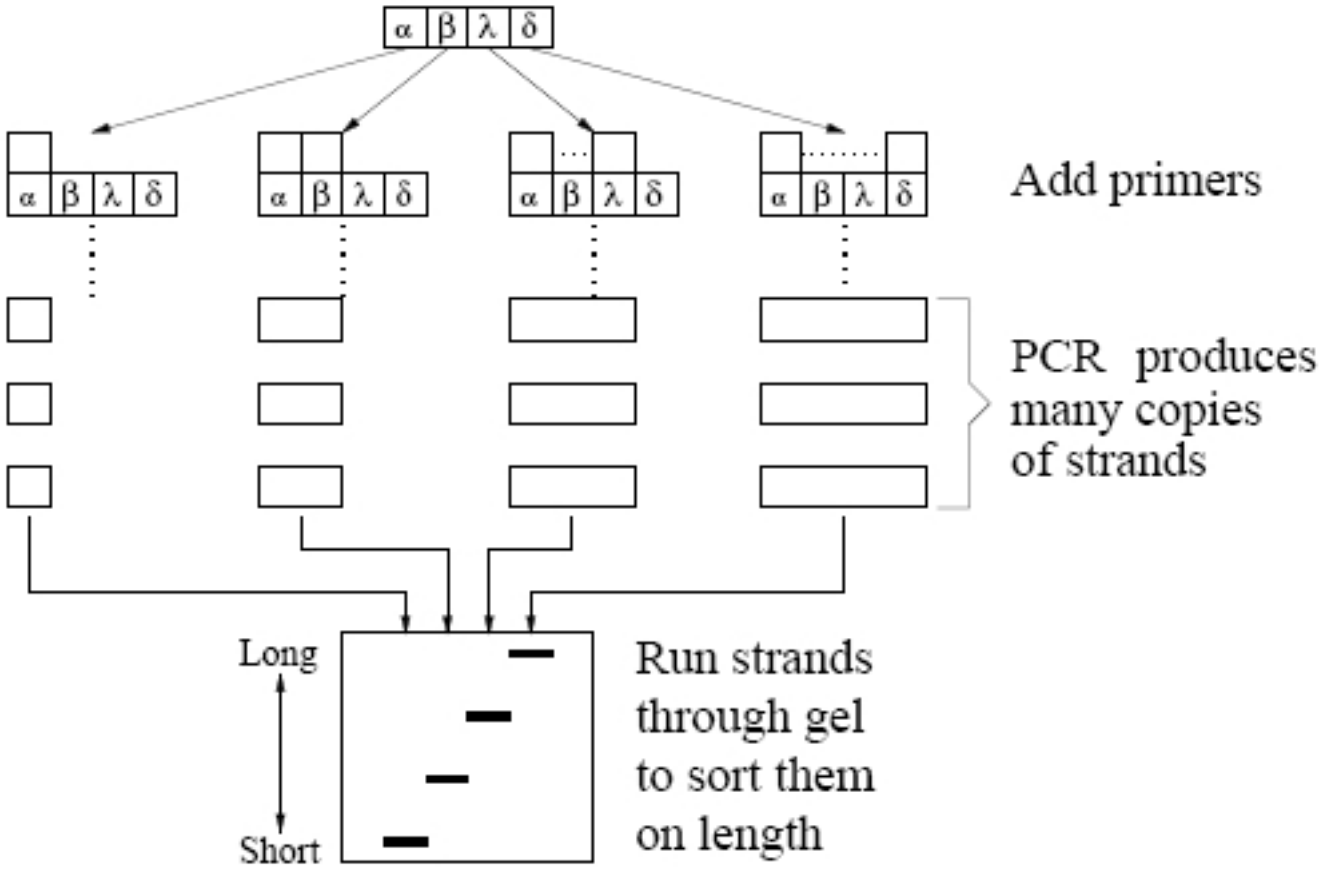
Manchester Metropolitan University



Manchester Metropolitan University

Confirm the result

Use graduated PCR to seek the *unique* HP.



Manchester Metropolitan University

Finally…

- Adleman's experiment worked, but he failed to carry it out on a graph that did *not* contain a HP
- It is also specific to the particular problem – he did not propose any way to solve other problems
- What was needed was a general model of molecular computation – a framework for the expression of *any number* of DNA algorithms
- Next lecture!