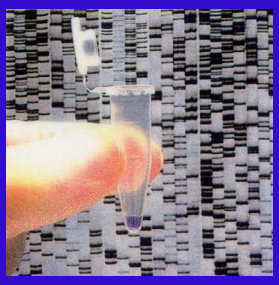
SCIENCE

Bioinformatics



Prof Peter O'Donoghue

1

Chemical basis of life

DNA —transcription→ RNA —translation→ Protein

DNA —replication→ DNA

2

What is DNA?

deoxyribonucleic acid

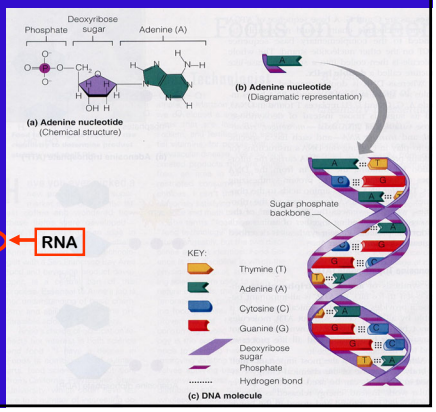
linear polymer of 4 nucleotides

- purines
  - adenine A
  - guanine G
- pyrimidines
  - cytosine C
  - thymine T (uracil U)

RNA

complementary bases
G - C
A – T

double helix



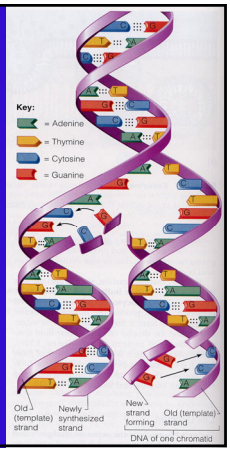
3

DNA replication

double strand unfolds (unzip weak H bonds)

new strands assembled using single strands as templates for replication proteins & DNA polymerases

⇒ like mirror-image processing



4

Bioinformatics

- design PCR primers (e.g. 20 bp probes)
  (forward and reverse, sandwiching desired gene fragment)
- search databases for sequences specific to organism
  (assumes sequences are in database)??
  (if not, design nested primers)

5

Bioinformatics

- determine PCR run conditions for designed primers
- calculate %GC content to determine melting temperature
  (count total number, count G, C, A, T)
  (calculate optimal Tm)

6

Oligonucleotide probes

Melting Temperature (Tm) of probe
(defined as temperature at which half DNA strands are in double-helical state and half in random-coil state)

Need to determine Tm for:
- PCR
- Southern/Northern blots
- *In situ* hybridzation

Tm affected by:
- Probe sequence (GC content)
- Probe length
- Salt concentration
- Strand concentration
- Denaturants (formamide, DMSO)
- Hybridization conditions

7

Oligos Tm

Wallace Rule: (for oligos 14-20 bp in 0.9M NaCl)

$Tm\ (°C) = 2(A+T) + 4(G+C)$

(where A = number of A bases in probe
T = number of T bases in probe,
G = number of G bases in probe
C = number of C bases in probe)

Wallace-Ikatura Rule: (for short oligos < 14 bp)

$Tm\ (°C) = 2(L+G+C)$

(where L = length of probe
G = number of G bases in probe
C = number of C bases in probe)

Show they are the same!

8

Oligos Tm

Wallace-Ikatura Rule: (for oligos longer than 14 bp)

$Tm\ (°C) = 64.9 + [41(G+C-16.4)] / L$

(where L = length of probe,
G = number of G bases in probe
C = number of C bases in probe)

More complex equations (for probes longer than 50 bp)

$Tm\ (°C) = 81.5 + 16.6\ logM + 41(XG+XC) – 500/L – 0.62F$

(where M = molar concentration of monovalent cations,
XG and XC = mole fractions of G and C in probe,
L = length of shortest strand,
F = molar concentration of formamide)

9

Bioinformatics

- amplify DNA by PCR
- characterize amplicon
  (count total number of nucleotides)
  (count number of G, C, A, T)

10

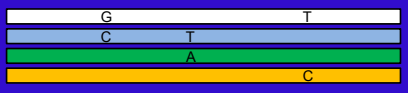
DNA sequences

Amylovorax dehorityi (novel endosymbiotic ciliate from stomach of kangaroo)

```
0      AACCTGGTTA ATCCTGCCAG TAGTCATATG CTTGTCTCAA AGACTAAGCC
50     ATGCATGTCT AAGTATAAAT AACTACACAG TAAAACTGCG AATGGCTCAT
100    TAAAACAGTT ATAGTTTATT TGATACATTA AATGGATAAC TGTAGAAAAA
150    CTAGAGCTAA TACATGCTGA GGCCGCAAGG TCGTATTTAT TAGATATTCC
200    AATTAAGGTG AATCATAATA ACTTCGCAAA TCACGATTTT GTCGTGATAA
250    ATCATCCAAG TTTCTGCCCT ATCATGCTTT CGATGGTAGT GTATTGGACT
300    ACCATGGCTT TTACGGGTAA CGGGGAATTA GGGTTCGATT CCGGAGAAGG
350    AGCCTGAGAA ACGGCTACTA CATCTACGGA AGGCAGCAGG CGCGTAAATT
400    ACCCAATCCT GACTCAGGGA GGTGGTGACA AGATATAACG ACGTGATTAA
450    AATCGCGATT GGTAGTGAGG GTTTCCTACA CCGAACCACT AGTACGATTA
500    GAGGGCAAGT CTGGTGCCAG CAGCCGCGGT AATTCCAGCT CTAATAGCGT
550    ATATTAAAGT TGCTGCAGTT AAAAAGCTCG TAGTTGGATT TCAAGGATTA
600    TAATCACCTT CTGGTGACTA TACCCTACTA CCCTTTTAGG TGTTACTGTG
650    AGAAAATTAG AGTGTTTAAA GCAGGCTATT GCAAGAATAC ATTAGCATGG
700    AATAACGAAT GTGTTTAGAA TCTTGGTTAA TTCTAGACGC GGTTAATAGG
750    GACAGTTGGG GGCATTAGTA TTTAATAGTC AGAGGTGAAA TTCTTGGAAT
800    TTGTTAAAGA CCTAACTTAT GCGAAAGCAT TTGCCAAGGA TGTTTTCATT
850    AATCAAGAAC GAAAGATAGG GGATCAAAGA CAATCAGATA CTGTCGTAGT
```

11

Bioinformatics



- compare sequence to others in database
  (locate differences)
  (enumerate differences)

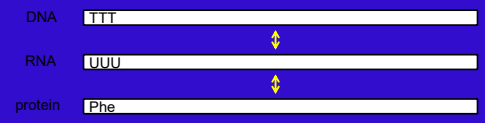
12

Comparing DNA sequences

```
                    10        20        30        40
Sa. muris           AACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCT--TAAAGATTAAG
Is. intestinalis    ..................................--C.....C....
Is. prostoma        ..................................--C.....C....
D. ruminantium      .........A........................--C.....C....
B. coli             ..................................--C.....C....
En. caudatum        ....................A.............--C.....C....
P. multivesiculatum ..................................--C.....C....
E. maggii           ..................................--C.....C....
D. dentatum         ..................................--C.....C....
Ep. caudatum        ....................A-............--C.....C....
O. maggii           ....................-.............--C.....C....
C. edentatum        ..................................--C.....C....
P. turniae          .........A........................--C.....C....
Po. roundi          ..................................--C.....C....
Ma. ennuensis       ..................................--C.....G....
Ma. yalanbense      ..................................--C.....C....
Am. dehorityi       .........A........................--C.....C....
Am. dogieli         ..................................--C.....C....
Bi. tasmaniensis    .........A........................--C.....C....
Ba. smalesae        ....................AG............--C.....C....
```

13

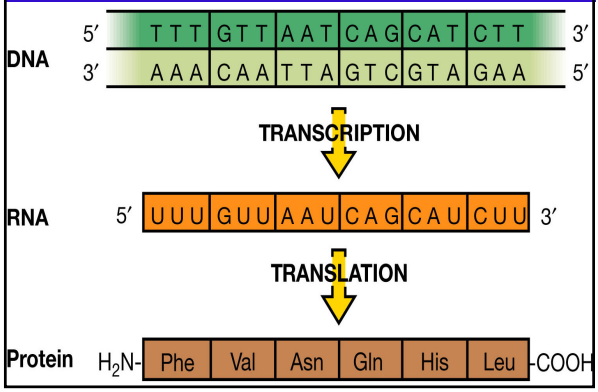
Bioinformatics



- read genome (triplet codons)
- transform DNA to RNA (T→U), or reverse
- transform RNA to protein, or reverse

14

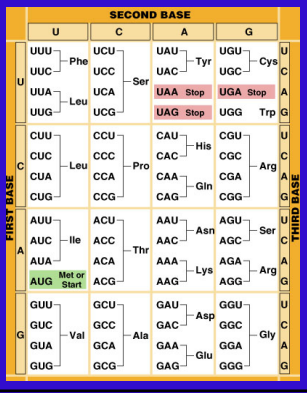
Transcription & Translation



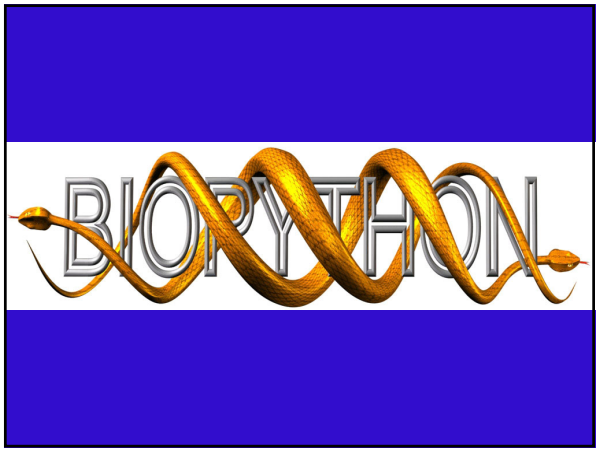
15

20 amino acids

Ala	alanine
Arg	arginine
Asn	asparagine
Asp	aspartic acid
Cys	cysteine
Glu	glutamic acid
Gln	glutamine
Gly	glycine
His	histidine
Ile	isoleucine
Leu	leucine
Lys	lysine
Met	methionine
Phe	phenylalanine
Pro	proline
Ser	serine
Thr	threonine
Trp	tryptophan
Tyr	tyrosine
Val	valine



16



17