

Bioinformatics CSM17 Week 6: DNA, RNA and Proteins

- Transcription (reading the DNA template)
- Translation (RNA -> protein)
- Protein Structure

Transcription - reading the data

- enzyme - transcriptase
- gene opens up and...
- messenger RNA (for each gene) is formed
- *introns* (non-coding regions) are removed
- mRNA moves off through nuclear pores...
- ...to the ribosomes
- Note:
 - DNA sense strand (mRNA has T->U)
 - and antisense strand (complement of mRNA)

Transcription - reading the data

A T G C A T G C Sense Strand
(Partner)

| | | | | | | |

T A C G T A C G Antisense
(Template)

Transcription - reading the data

A T G C A T G C Sense Strand
(Partner)

| | | | | | | |

| | | | | | | |
T A C G T A C G Antisense
(Template)

Transcription - reading the data

A T G C A T G C Sense Strand
(Partner)

| | | | | | | |

A **U** G C A **U** G C Messenger RNA

| | | | | | | |

T A C G T A C G Antisense
(Template)

RNA (Ribose Nucleic Acid)

- single chain
- sugar-phosphate backbone
- nucleotide bases (nt)
- Adenine (A)
- Guanine (G)
- Cytosine (C)
- **Uracil (U)** [DNA has Thymine (T)]
- plant viruses, retroviruses (e.g. HIV)

Translation - reading the mRNA

- transfer RNA (tRNA) for each amino acid
 - clover-leaf structure
- codons and anticodons (triplets of bases)
- long chains of amino acids are built up

		Second Position							
		U		C		A		G	
First Position	U	UUU	Phe (F)	UCU	Ser (S)	UAU	Tyr (Y)	UGU	Cys (C)
		UUC		UCC		UAC		UGC	
		UUA	Leu (L)	UCA		UAA	Stop	UGA	Stop
		UUG		UCG		UAG	Stop	UGG	Trp (W)
	C	CUU	Leu (L)	CCU	Pro (P)	CAU	His (H)	CGU	Arg (R)
		CUC		CCC		CAC		CGC	
		CUA		CCA		CAA	Gln (Q)	CGA	
		CUG		CCG		CAG		CGG	
	A	AUU	Ile (I)	ACU	Thr (T)	AAU	Asn (N)	AGU	Ser (S)
		AUC		ACC		AAC		AGC	
		AUA		ACA		AAA	Lys (K)	AGA	Arg (R)
		AUG	Start	ACG		AAG		AGG	
	G	GUU	Val (V)	GCU	Ala (A)	GAU	Asp (D)	GGU	Gly (G)
		GUC		GCC		GAC		GGC	
		GUA		GCA		GAA	Glu (E)	GGA	
		GUG		GCG		GAG		GGG	

Third Position

Amino Acids

- 20 in total
- building blocks of proteins

Proteins & Polypeptides

- polypeptides are short
- proteins are long
- enzymes are proteins that act as catalysts
- haemoglobin (in blood) is a protein

Protein Structure

- strings of amino acids
- primary (first) structure - AA sequence
- secondary (second) structure - e.g. Spiral
- tertiary (third) structure - complex shape
 - has a large influence on function

Mutation



Mutation

- changes of nucleotide bases
- caused by
 - ionizing radiation, mutagenic chemicals
- usually harmful (damaging)
- may be
 - single base (changing one amino acid)
 - frameshift (more serious)

Sequence Comparison

Simple Alignment

match score: 1

mismatch score 0

A A T C T A T A

A A G A T A

4 + 0 = 4 (best)

A A T C T A T A

A A G A T A

1 + 0 = 1 (worst)

A A T C T A T A

A A G A T A

3 + 0 = 3

Sequence Comparison

Simple Alignment with gap penalties

match score: 1 *mismatch score 0* *gap penalty -1*

A A T C T A T A

A A G - A T - A

$$3 + 0 - 2 = 1 \quad (\text{worst})$$

A A T C T A T A

A A - G - A T A

$$5 + 0 - 2 = 3 \quad (\text{equal$$

best)

A A T C T A T A

A A - - G A T A

$$5 + 0 - 2 = 3 \quad (\text{equal$$

best)

A A T C T A T A

- A A G A T A -

$$1 + 0 - 2 = -1 \quad (\text{worst})$$

Sequence Comparison

Simple Alignment with origination and length penalties

match score: 1 mismatch score 0

origination penalty: -2 length penalty -1

A A T C T A T A

A A - G - A T A 5 + 0 - 4 - 2 = -1 (worst)

A A T C T A T A

A A - - G A T A 5 + 0 - 2 - 2 = 1 (best)

Origination penalty is applied for starting a series of gaps

Length penalty is also applied for each gap

Databases

- NCBI Genbank
www.ncbi.nlm.nih.gov/Genbank/index.html
- EBI European Bioinformatics Institute
www.ebi.ac.uk
- EMBL European Molecular Biology Lab.
www.embl.de/
- ...where you can conduct BLAST – Basic Local Alignment Search Tool – searches

Useful Websites

- GeneStudio Sequence Converter
 - www.genestudio.com/seqverter.htm
- Mitochondrial maps
 - www.mitomap.org
- European Molecular Biology Laboratory
 - www.embl.org
- BioEdit
 - www.mbio.ncsu.edu/BioEdit/bioedit.html

References & Bibliography

- Watson, J. (2003). *DNA- the Secret of Life*. Heinemann. ISBN 0434011169
- Gibas, C. & Jambeck, P. (2001). *Developing bioinformatics computer skills*. O'Reilly, USA. pp. 4-8, 22-29, 159-167, 215-22, 296-303. ISBN 1-56592-664-1 (paperback)
- Hofstadter, D. (1979) *Gödel, Escher, Bach: An Eternal Golden Braid*, Penguin Books, London, UK. pp. 514-524
- Rose, S. (1979, reprinted 1987) *The Chemistry of Life*. Penguin Books (Pelican), London. (paperback).
- Vander, A., Sherman, J. & Luciano, D. (1994). *Human Physiology*. Chapter 2 & Chapter 4. McGraw-Hill, New York. ISBN 0-07-066992-9 (paperback)
- Lewin, B. (1997). *Genes*. (6th edition) Oxford University Press, UK. (Chapters 4 & 11)