Uses of DNA Sequence Data

Marcel Jaspars

Marine Biodiscovery Centre, Department of Chemistry

University of Aberdeen, Scotland, UK

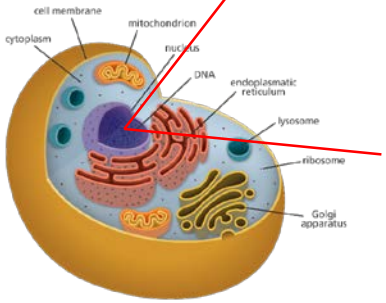
m.jaspars@abdn.ac.uk

Outline

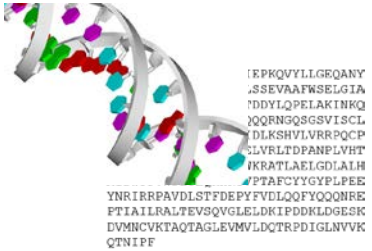
- Production of DSI
  - Genetic resources
  - The central dogma of molecular biology
  - The Origin of DSI
- The rise of ‘omics’
- What is being sequenced and why
  - Example – fish identification
- DSI usage
  - Non-linearity of DSI usage
  - Example - enzyme modification to make pharmacuticals
  - Synthetic biology to make pharmaceuticals
- Conclusions

Production of DSI - Genetic Resources

Biological resource

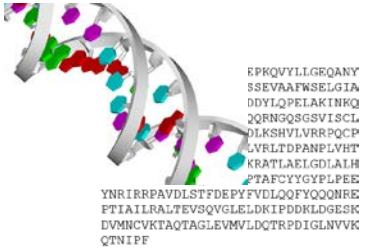


Cells (one or more)



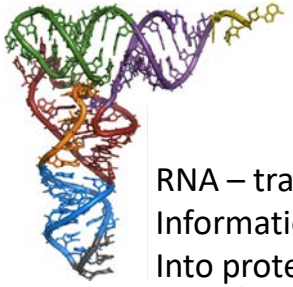
DNA (GR/Functional Units of Heredity)

Production of DSI - The Central Dogma of Molecular Biology



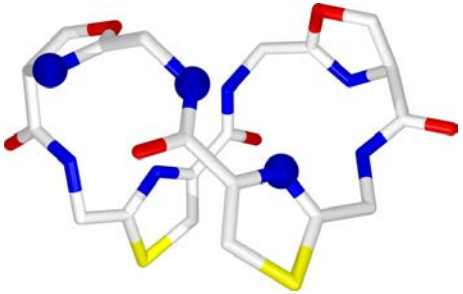
DNA (GR/Functional Units of Heredity) (Nucleotides)

Transcription

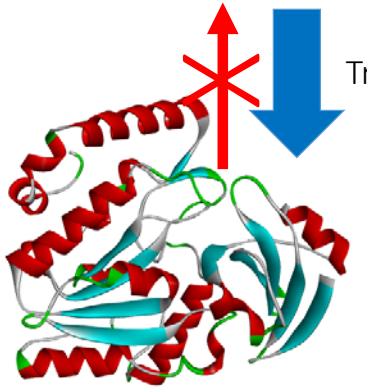


RNA – translates Information from DNA Into proteins (Nucleotides)

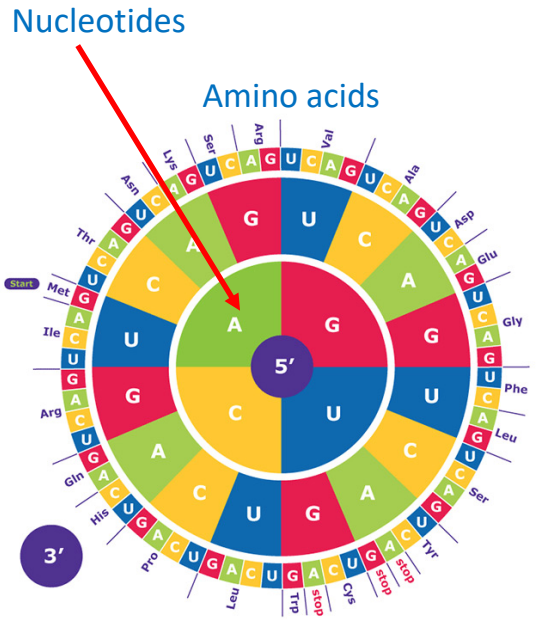
Translation



Metabolites

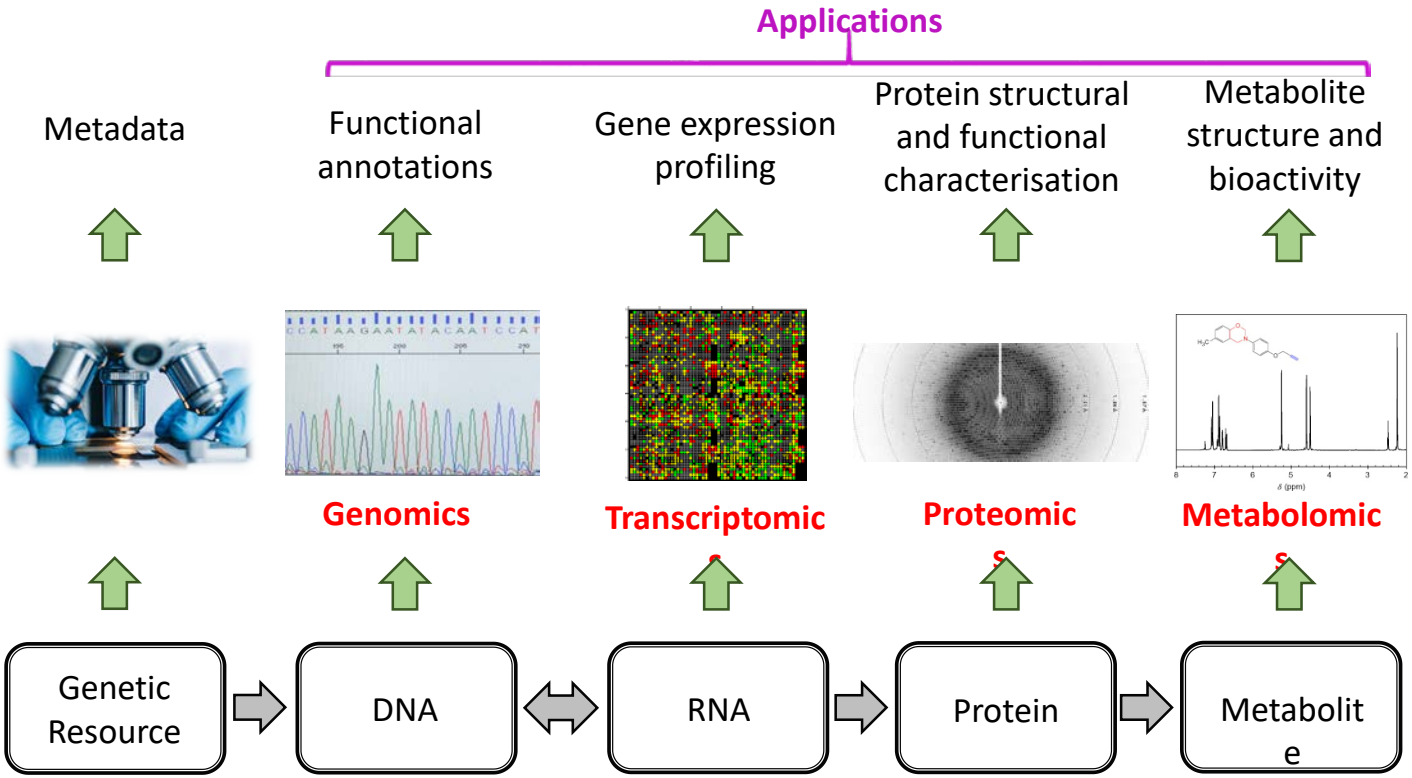


Proteins/Enzymes (Amino acids)



DNA Code is Degenerate

The Origin of Digital Sequence Information

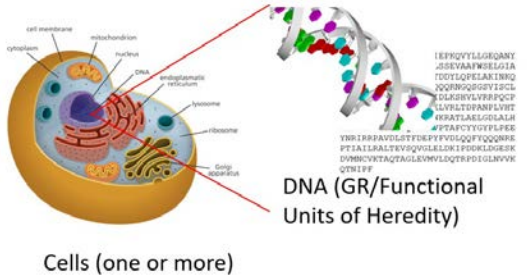


Dr Wael Houssen

The Rise of ‘Omics’

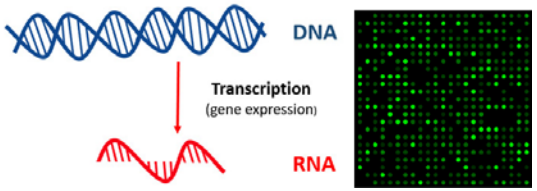
Genomics

The genome is the complete set of genes or genetic material present in a cell or organism.



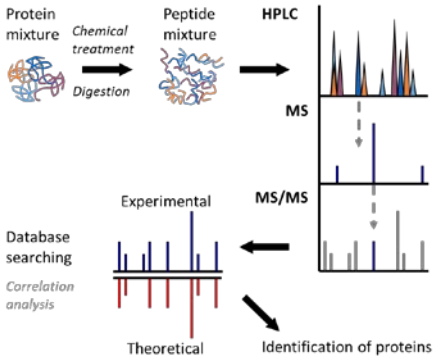
Transcriptomics

The transcriptome is a measure of which genes are expressed under any given set of conditions.



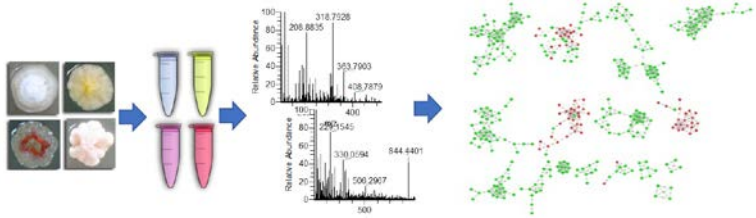
Proteomics

The proteome is the totality of all proteins produced or modified under any given set of conditions.



Metabolomics

The metabolome is the totality of small molecule metabolites produced under any given set of conditions.

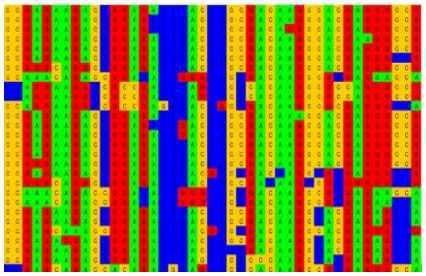


What is Being Sequenced and Why?

Ecosystems

Understand how life forms interact with each other and their environment.

Populations

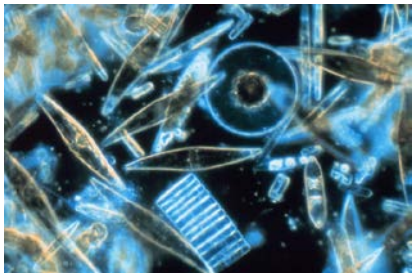


Understand variation within or between related species to learn more about their biology through comparative genomics.

Biodiversity

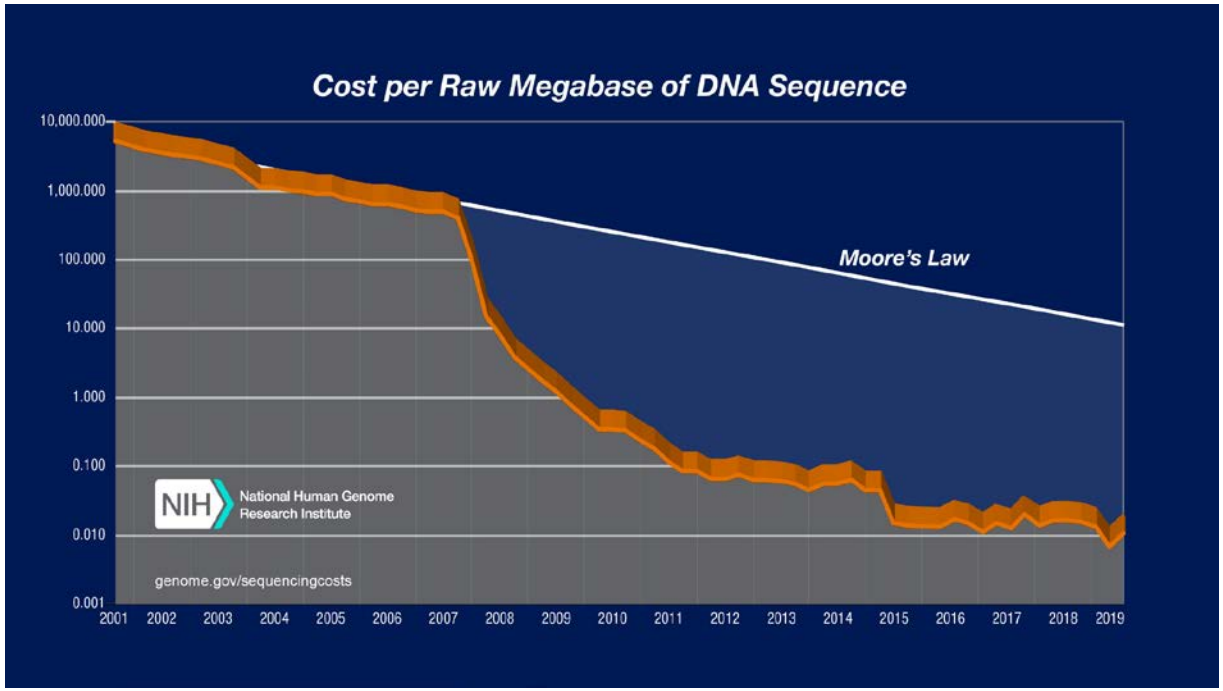
Learn about life forms and preserve the genetic data before they become extinct.

Metagenomes



Understanding the microbial communities and their environments

Sequencing Costs



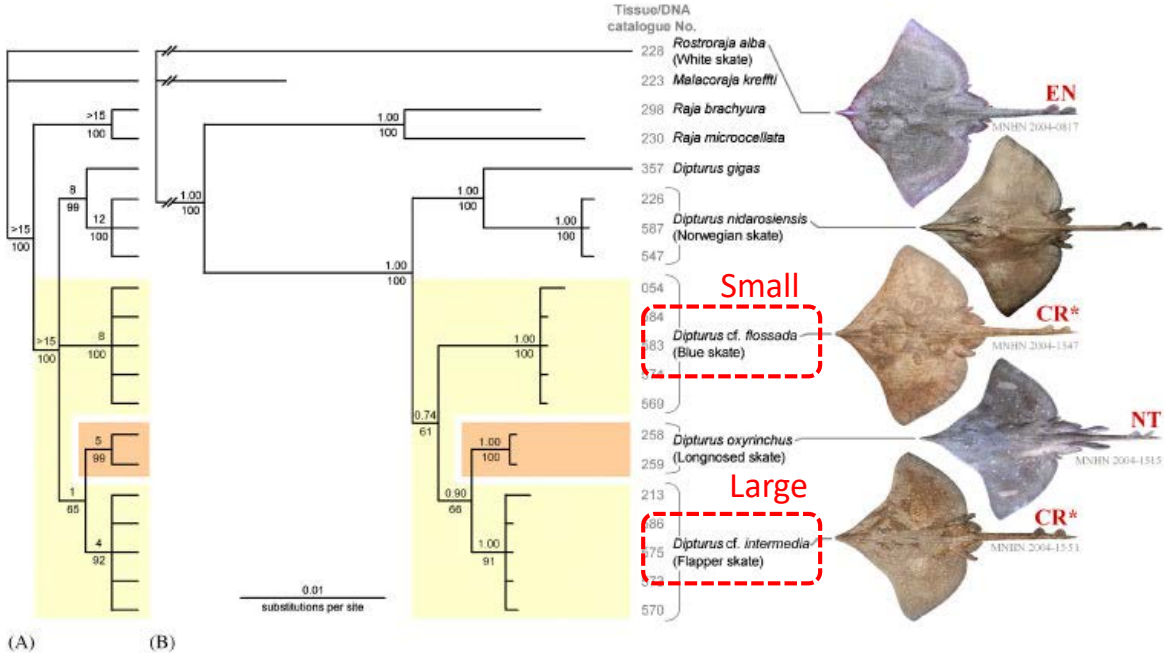
Cost of sequencing continues to decline

Number of sequences escalates – **data avalanche**

Accurate interpretation and annotation of sequences needs to improve

Example of Application of DNA Sequences
Mistaken Identity of Fish

- Declining Catches of Common Skate *'Dipturis batis'*
- Taxonomic discrepancies
- mtDNA gave phylogenetic tree
- Two species
  - Small one common
  - Large one rare
- Information used for setting catch limits

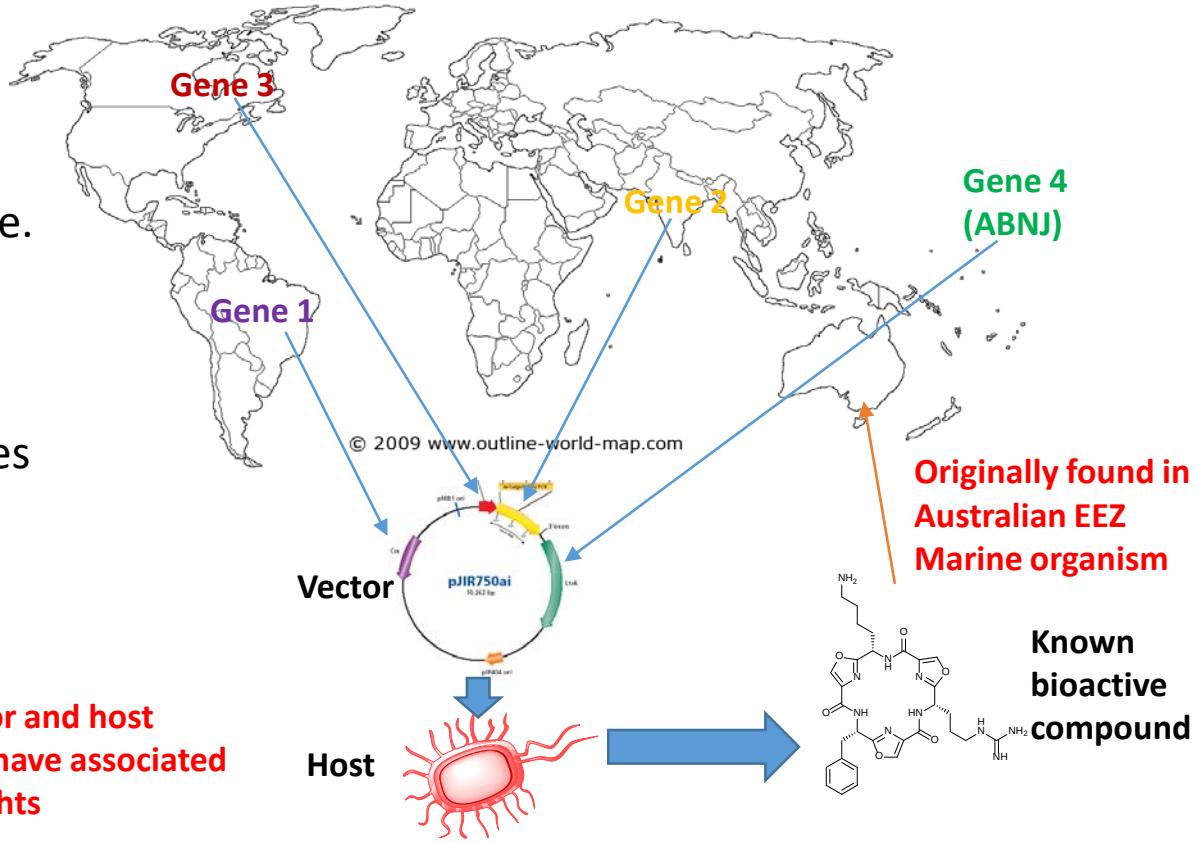


Aquatic Conserv: Mar. Freshw. Ecosyst. 20: 319–333.

DSI Usage is Not Straightforward or ‘Linear’

Sequences used can be the result of analysing multiple sequences and finding the most useful one.

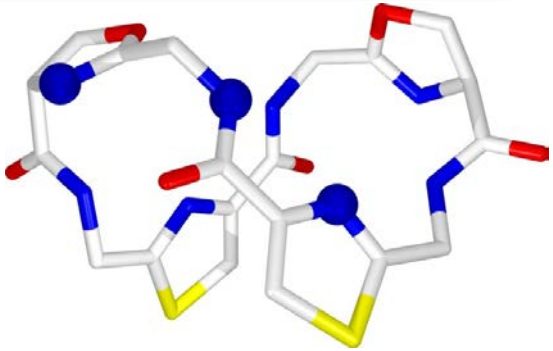
Some sequences can be designed or engineered based on natural sequences and the process can take several iterations.

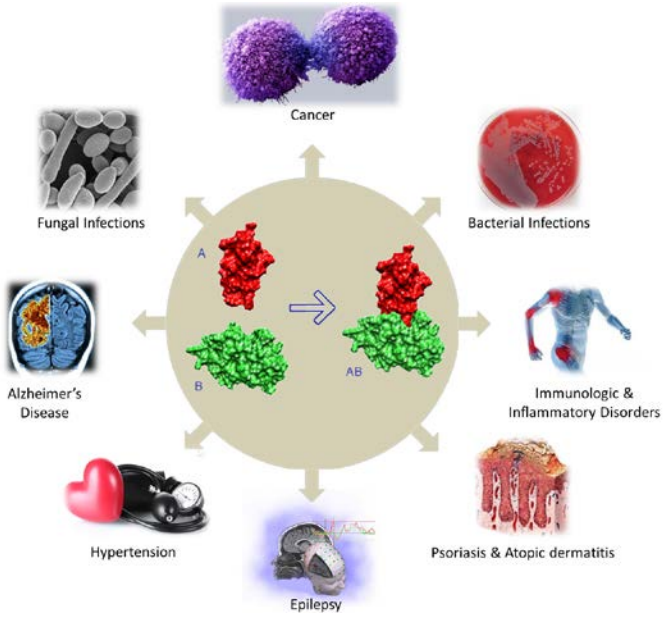


GyreOx THERAPEUTICS

Gyrocycle™ highly modified macrocyclic peptides - effective alternatives to therapeutic peptides

Taking the Lead from Nature



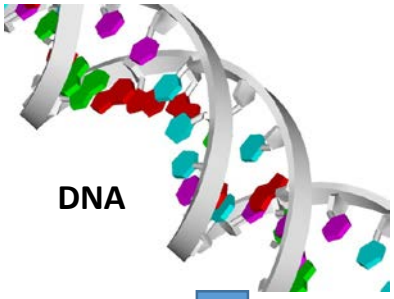


Macrocycles may be effective against a range of complex diseases

GyreOx THERAPEUTICS

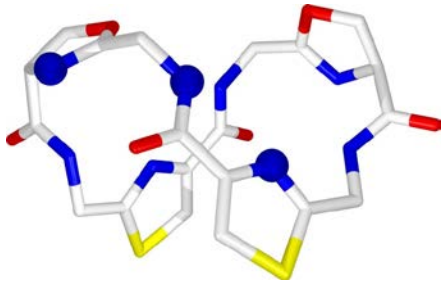
Gyrocycle™ highly modified macrocyclic peptides - effective alternatives to therapeutic peptides

Using DNA Sequence Data



DNA

Protein



Macrocycle

GyreOx THERAPEUTICS

Gyrocycle™ highly modified macrocyclic peptides - effective alternatives to therapeutic peptides

Using Protein Sequence Data

MQPTALQIKPHFHVEIIEPKQVYLLGEQGNHALTGQLY
LSRLVEKGYLTEVAPELSLEVAAFWSELGIAPSVVAEG
VSDPKAPKAPKAGDSTAQLQVVLTDDYLQPELAAINKE
HCLAQRLRGNREVEASVLQQKRALQERNGQNKNGAVSC
NAIAPGTARFPTLAGKIFTFNQTTLELKAHPLSRRPQC
RATTPEQTVQKYQHLIGPITGVVTELVRISDPANPLVH
DSQSRASGLCEAIERYSGIFLGDEPRKRATLAELGDLA
PHRFAASQAIDWTPLWSLTEQKHKYVPTAICYYNYLLF
RDSVALWWYNRLRRPEVELSSFEEPYFLQLQQFYRSQN
IGFGAHLDPKIAILRALTEVSQVGLELDKVPDEKLDGE
RWSDDIYTDVMACVEMAKVAGLETLVLDQTRPDIGLNV
PLAEAEMNPTNIPF

Original Protein Sequence

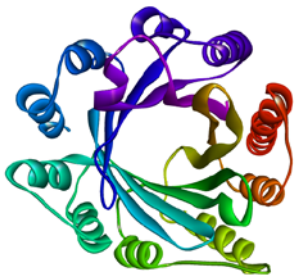
Native enzyme

MQSTTLLQIKPHFHIEVIEPKQVYLLGEQGNHALTG
VLNRLAEKGYLTEATPDLSPEVAAFWTELGIAPTVA
PVQNAS MQSTPLLQIQPHFHVEVIEPKQVYLLGEQANY
RLRGNR VLDRLAEKGYLTEAAPELSSEVAAFWSELGIA
TALFPT VQTSTEAGSPTALNVVLTDDYLQPELAKINKQ
QTLQKY LRGNRI MQSTPLLQIQPHFHVEVIEPKQVYLLGEQANY
GFCEAV VFFPTI VLDRLAEKGYLTEAAPELSSEVAAFWSELGIA
QAIDWT TVQKY(VQTSTEAGSPTALNVVLTDDYLQPELAKINKQ
WYNRLS LCEAI LRGNREVEASVLQQKQAQQQRNGQSGSVISCL
DPTIGI AIDWT VFFPTLDGKIITFNHTVIDLKSHVLVRRPQCP
TDVMTC YNRIR TVQKYQHLVSPITGVVTELVRLTDPANPLVHT
NPMNIP PTIAI LCEAIERYSGIFQGDEPWKRATLAELGDLALH
DVMNC AIDWTPVWSLTEQKHKYVPTAFCYYGYPLPEE
QTNIP YNRIRRPAVDLSTFDEPYFVDLQQFYQQQNRE
PTIAILRALTEVSQVGLELDKIPDDKLDGESK
DVMNCVKTAQTAGLEVMVLDQTRPDIGLNVVK
QTNIPF

Use DNA databases to find analogues and make comparisons

MQPTALQIKPHFHVEIIEPKQVYLLGEQGNHALTGQLY
LSRLVEKGYLTEVAPELSLEVAAFWSELGIAPSVVAEG
VSDPKAPKAPKAGDSTAQLQVVLTDDYLQPELAAINKE
HCLAQRLRGNREVEASVLQQKRALQERNGQNKNGAVSC
NAIAPGTARFPTLAGKIFTFNQTTLELKAHPLSRRPQC
RATTPEQTVQKYQHLIGPITGVVTELVRISDPANPLVH
DSQSRASGLCEAIERYSGIFLGDEPRKRATLAELGDLA
PHRFAASQAIDWTPLWSLTEQKHKYVPTAICYYNYLLF
RDSVALWWYNRLRRPEVELSSFEEPYFLQLQQFYRSQN
IGFGAHLDPKIAILRALTEVSQVGLELDKVPDEKLDGE
RWSDDIYTDVMACVEMAKVAGLETLVLDQTRPDIGLNV
PLAEAEMNPTNIPF

Choose sequence(s)
Targeted modifications
Gene Synthesis



Engineered Enzyme

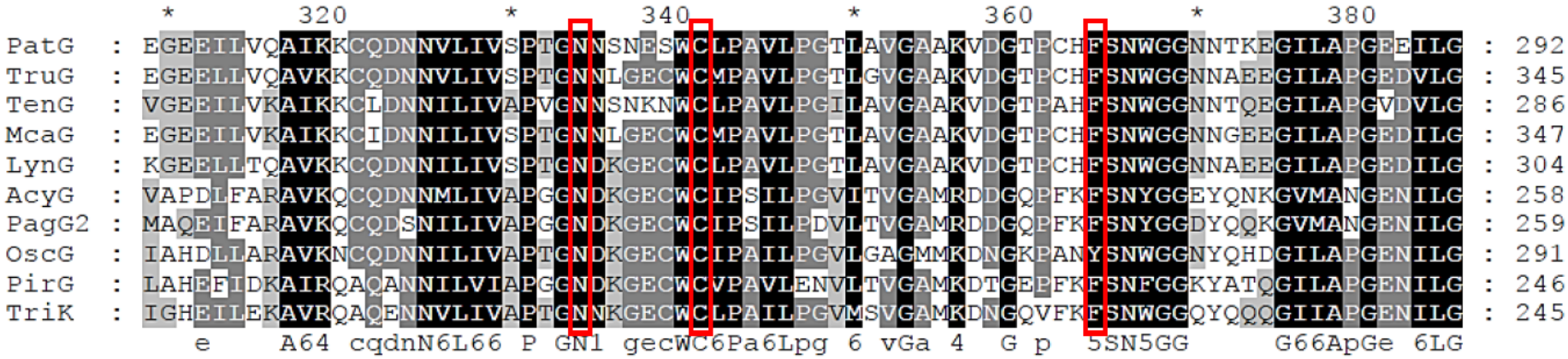
Desired T range
Desired pH range
Desired specificity

Can be produced at High titres

GyreOx THERAPEUTICS

Gyrocycle™ highly modified macrocyclic peptides - effective alternatives to therapeutic peptides

Using Protein Sequence Data



- Many sequences with a variety of origins have the same properties
- If sequences have >30% similarity they are likely to have the same function
- Sequences can be codon-optimised/engineered making them hard to trace back to the source

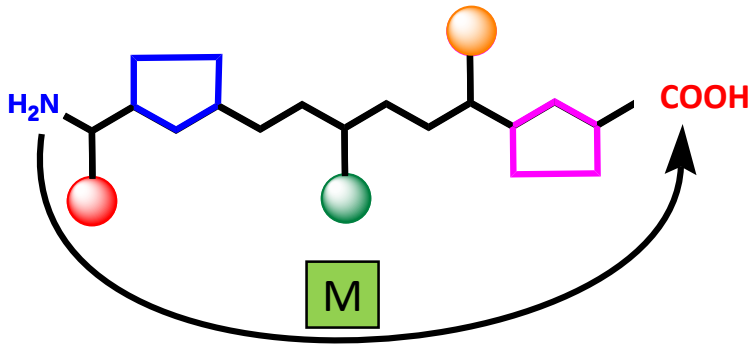
Protein Engineering vol.12 no.2 pp.85–94, 1999

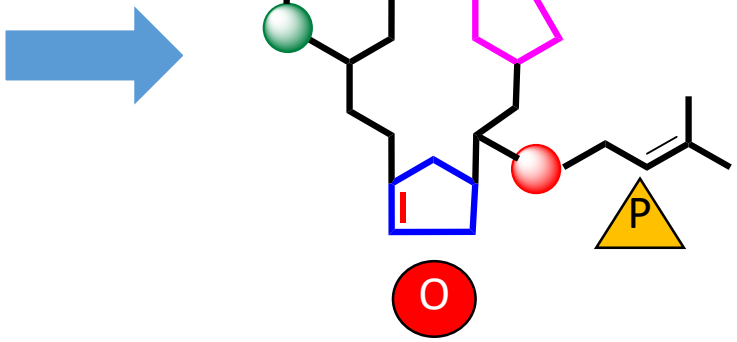
Cristina Crivac

GyreOx THERAPEUTICS

Gyrocycle™ highly modified macrocyclic peptides - effective alternatives to therapeutic peptides

Using Engineered Enzymes





H Heterocyclase

M Macrocyclase

O Oxidase

P Prenylase

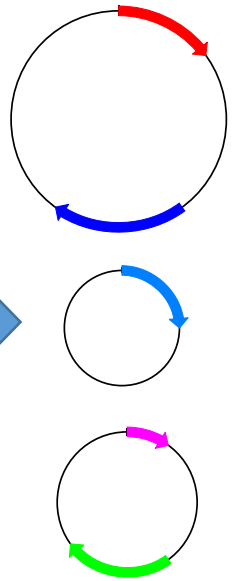
Synthetic Biology/Engineering Biology

(Design-Build-Test-Analyse-Rebuild)

Together with:

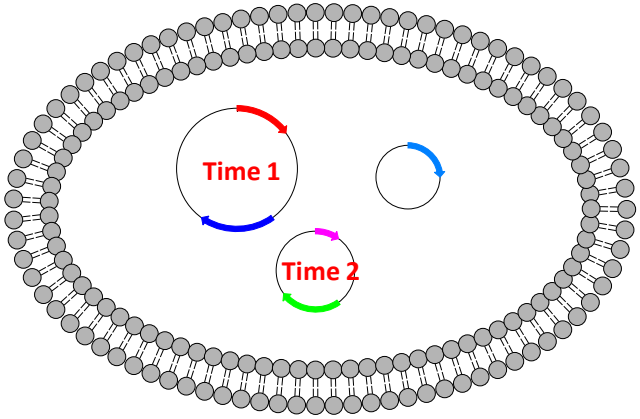
Ingenza

DNA Sequences

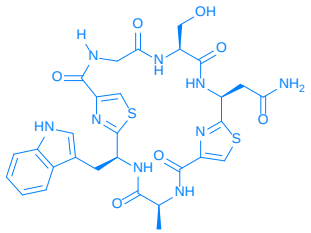


Native/modified/synthetic DNA
From different origins/based on many sequences

Substrate plasmid



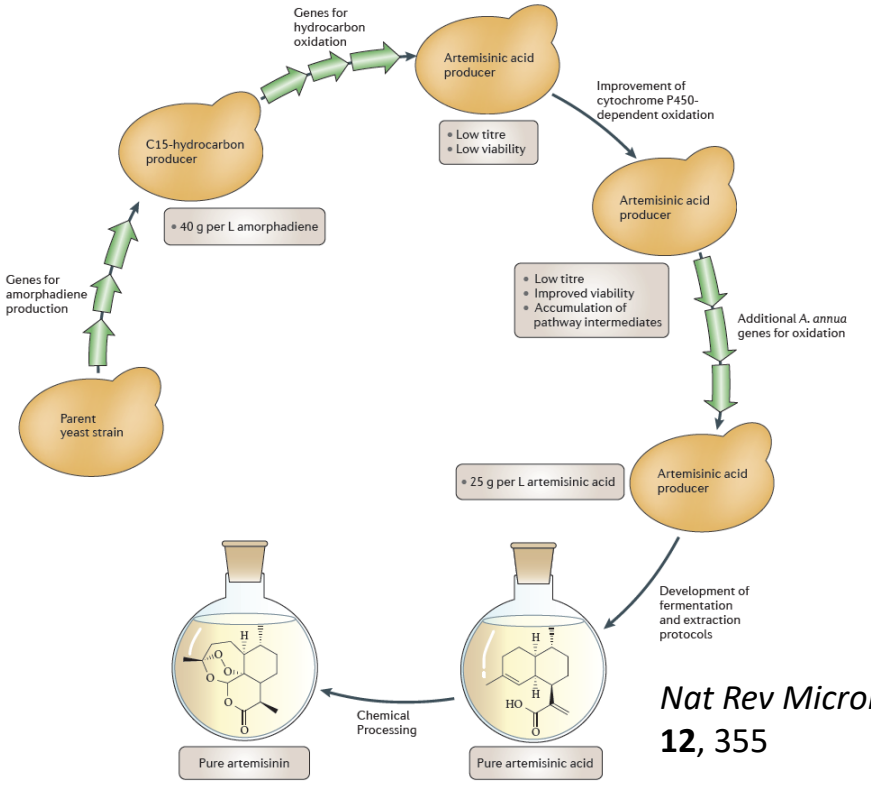
Engineered cell
Optimised production of product
Exports product



Desired
'Unnatural product'

Examples of Application of 'DSI' WHO Essential Medicines Made Using SynBio

- Malaria killed 445,000 people globally in 2016
- Quality-assured artemisinin-based combination therapies are a major part of the strategy against malaria
- SynBio process developed over 7 years with multi million US$ input
- Produces 25 g/L of precursor that needs synthetic finishing – could meet 25% of global demand



Nat Rev Microbio **12**, 355

Conclusions

- Many uses of DSI in basic and applied research e.g.
  - Conservation
  - Food safety
  - Vaccines/Pharmaceuticals
  - Industrial enzymes etc.
- Usage of DSI is not straightforward or linear
- Multiple genes sequences used from databases to create final product
  - Some to understand the function of genes
  - Some are incorporated in native or modified form
  - Many genes deposited many years ago
  - **What is the relative contribution of each sequence analysed/used?**