

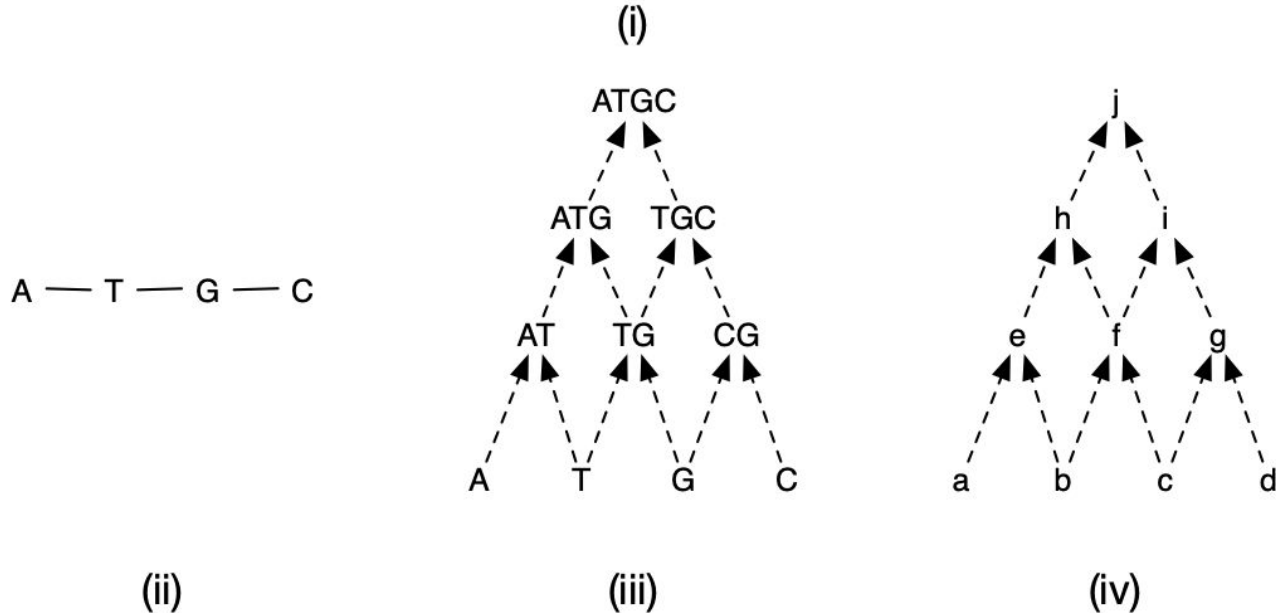
A BioSequence Ontology From Molecular Structure

Jona Thai, Michael Gruninger

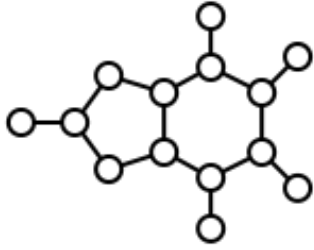
University of Toronto

What makes a gene?

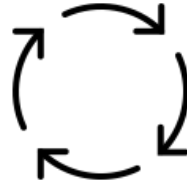
ATG CAA TGG GGA AAT ACC AGG TCC GAA CTT ATT GAG GTA AGA CAG ATT TAA
ATGC AAT GGG GAA ATA CCA GGT CCG AAC TTA TTG AGG TAA GAC AGA TTT AA
AT GCA ATG GGG AAA TAC CAG GTC CGA ACT TAT TGA GGT AAG ACA GAT TTA A



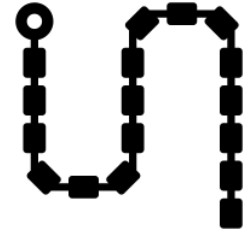
Central Research Questions



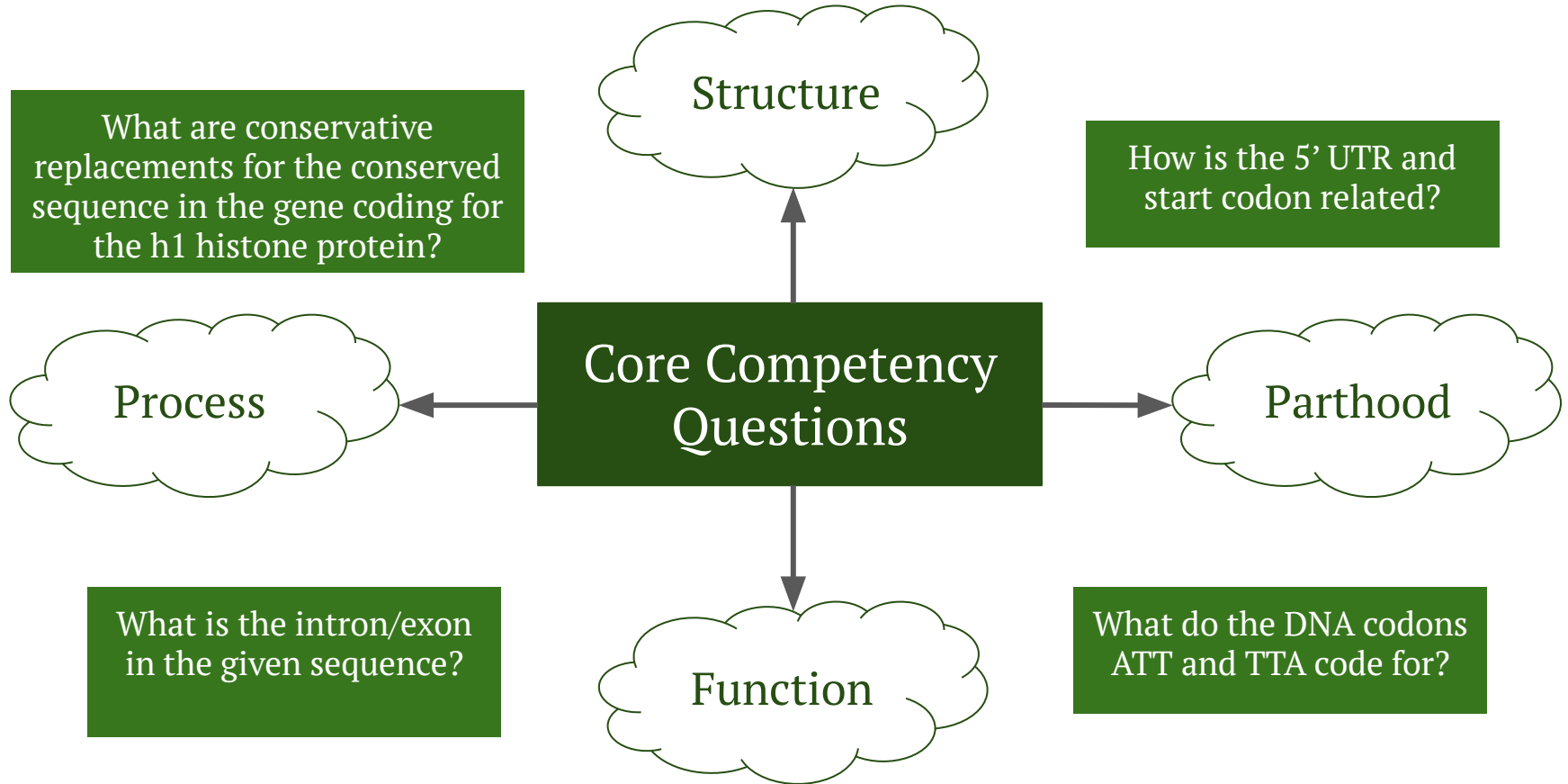
How can we represent gene sequences as molecular structures?



What is the representation for linear and circular gene sequences?



What is the mereotopology of gene sequences?



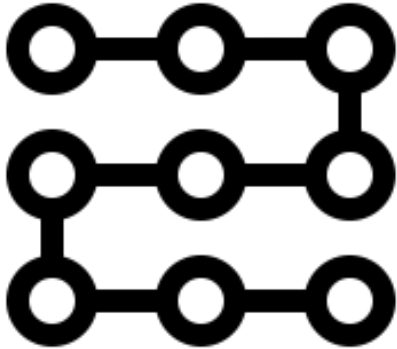
The Mereology

$$(\forall x \forall y (C(x, y) \equiv (\exists z (sum(x, y, z)))))$$

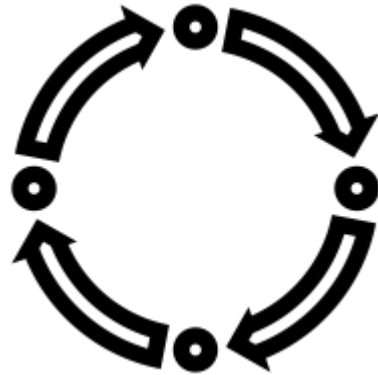
$$(\forall x \forall y \forall z (sum(x, y, z) \supset (\forall u (C(u, z) \equiv (C(u, x) \vee C(u, y)))))$$

$$(\forall x \exists y \exists z (atom(y) \wedge sum(y, z, x)))$$

Between-ness

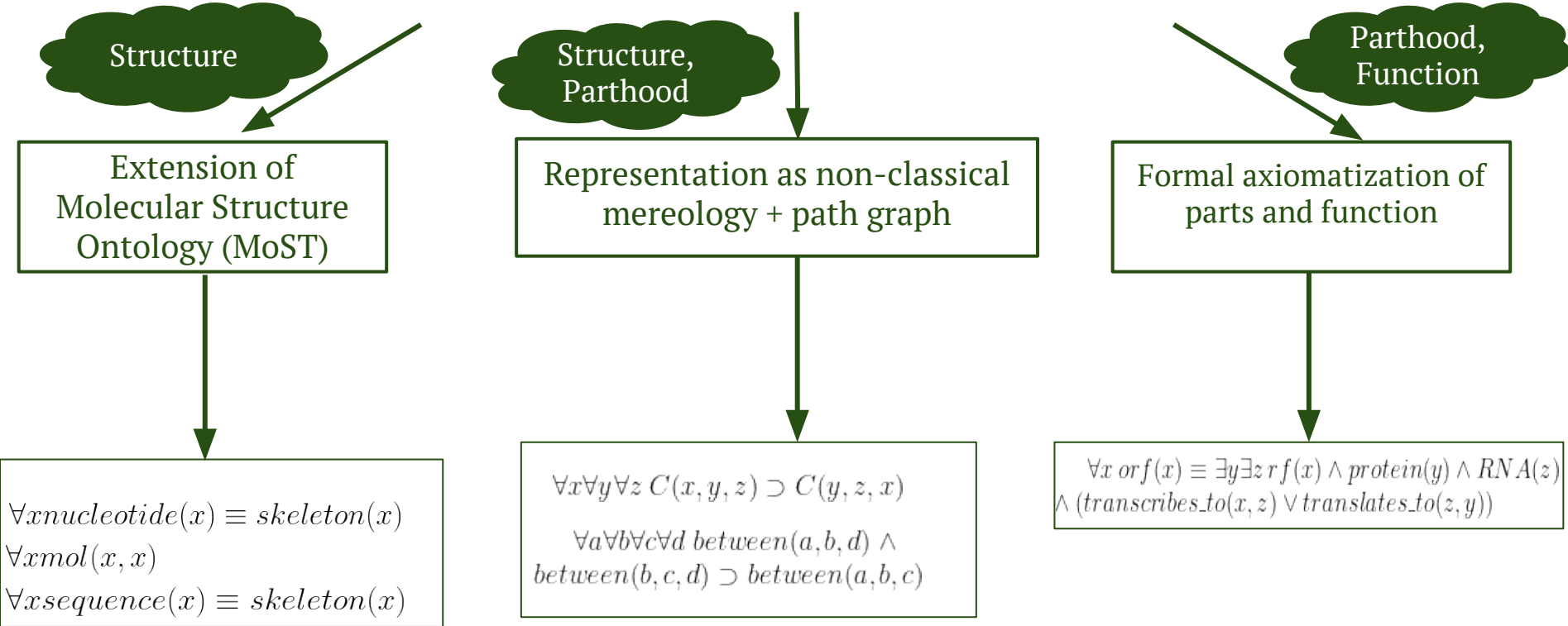


Linear



Cyclic

The BioSequence Ontology (BSO)



Evaluation

1. How is the 5' UTR and start codon related?

$$\forall x \text{start_codon}(x) \supset \exists y 5'UTR(y) \wedge \text{mol}(y, x)$$

2. What do the DNA codons ATT and TTA code for?

$$\forall x \forall y \text{ATT}(x) \wedge \text{TTA}(y) \supset \exists z \text{protein}(z) \text{code_for}(x, z) \wedge \text{code_for}(y, z)$$

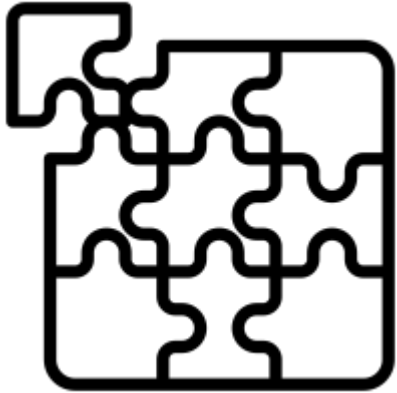
3. What are the introns and exons in the given sequence? *

$$\forall x \text{sequence}(x) \supset \exists i \exists e \text{intron}(i) \wedge \text{exon}(e) \wedge \text{part}(i, x) \vee \text{part}(e, x)$$

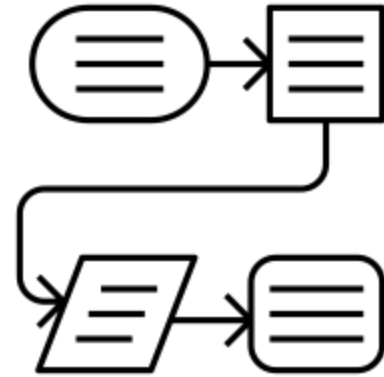
4. What are conservative replacements for the conserved sequence in the gene coding for the h1 histone protein? *

$$\forall s \text{conserved_sequence}(s) \forall h \text{histoneh1}(h) \wedge \text{part}(s, x) \supset \exists y \text{conserved_sequence}(y) \wedge s \neq y \wedge (\text{part}(y, x) \wedge \neq \text{part}(s, x))$$

Next Steps

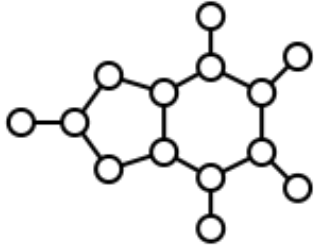


Mapping to other gene sequence
ontologies

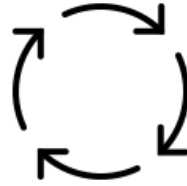


Building the BioSequence
Process Ontology

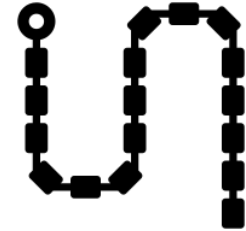
Conclusion



Built a reusable gene ontology extended from the Molecular Structure Ontology (MoST)



Defined separate and explicit axioms for cyclic and linear between-ness



Axiomatized a non-classical mereology that was the basis for other gene part definitions

Thank you, questions?