**PREP:**

* ERO - really need to think about chemical reagent hierarchy - and if reagents are molecules or compounds
* right now, organic macromolecule is a subclass of organic compound, which is wrong to say that a molecule can be a type of compound (the definition says that this is an organic compound with a large molecular mass . . . and this is ok - compounds/substances are said to have molecular masses)

1. **General question regarding how complex I should be making things?**
   * As my attempt at an antibody model demonstrates, I am having trouble focusing on a level of modeling detail/complexity. I was struck by your comment last week that my original antibody model was a good 'thought exercise', but that the level of complexity in it would never be useful.
   * Without any application or use case and in need of some guiding principles for ReO, my approach has been to think about the most useful classes/concepts that a reagent ontology should reference, and structure them in the most intuitive and understandable way possible, and link them to any related entities according to a realist view (ie reflecting reality). As you noted, this has resulted in a complex, OBI-like ontology that may be more detailed and complicated than will EVER be practically useful. So how do you suggest I proceed . . . what guiding principles should I set my sights on if reality is too complex to be practically used, and I have no application or use case? I don’t want to waste time building complex models such as the antibody model if they will never be used .

* + **Related issue :** Too many axes of classification can lead to reticulate ontology structure (if duplicate each axis of classification at each branch point). But at same time we want to provide multiple axes for classification to classify reagents in different ways to accommodate different perspectives and needs. How deal with this? - ie how balance need for multiple axes with also not wanting too complex/reticular an ontology structure?
    - clear evidence for this in cell and cell culture hierarchies . . .and now dealing with it in molecular entity/reagent modeling

1. **Review Molecular Entity/Reagent Modeling Approach and cmap**
   * **General Approach:** the classification scheme I have built for reagent types of molecular entities is based on the following general approach :

- Modeling ABOVE the core reagent molecular entity classes (ie protein, nucleic acid, amino acid, nucleotide, etc) my approach has been, as always, to provide high level context using established classes/hierarchies from external ontologies (ChEBI, PRO, SOM,OBI, etc) (ie import these classes/hierarchies an assert the corresponding core molecular reagent general/universal classes within this context to provide maximal interoperability for reasoning with these resources and linking to data annotated with these resources.

- Modeling DOWN from these core reagent molecular entity classes, I aim to provide subclassing (often along multiple axes) that reflects what I perceive to be the most useful and intuitive perspective ***to researchers*** who are using reagents of these core types, according to what they are, how they are made, and how they are used - to generate the most practical and understandable set of classes to be used for annotation of and reasoning with data.

- For ex, the following axes are applied to model alpha-amino acids and mononucleotides

1. **natural vs unnatural** (ie represents a molecular composition and arrangement found in nature, or not)

2. **canonical vs non-canonical** (represents a type that is normally incorporated into information biopolymers, or not . . . 20 canonical aas or ACTGU for nts)

3. **bearing of some structural modification** (ie structure diverges from some reference 'unmodified' structure)

4. **nature of modification** (based on nature of modified moieties for residues, or identity of modified residue for aas/nts).

Axes are applied in this order, as I think this best reflects the relative importance of these distinctions, and best parallels the precedence of (fundamental but typically implicit/tacit/unrecognized) mental processes involved in / distinctions made in designing and execution an experiment, or looking for research resources.

* + **CMAP tour - review principles above in action**
    - **lots to discuss here, including the compound vs molecule issue, and reviewing hierarchies for polypeptides/polynucleotides and their analogs, as well as amino acids/nucleotides and their analogs.**
    - **looking for broad/high level critique at this point, regarding classification axes, scope, theoretical approach, etc - many of the details of the cmap are test cases or examples that may not make the final cut**

1. **Time permitting - talk more about antibody model**

**-----------------**

* + **Molecular entity/reagent modeling**

* + **Compound vs molecule entity issue** ([Chemical / molecular reagents](onenote:..\PROJECTReO\_Modeling%20Issues.one#Chemical%20\%20molecular%20reagents&section-id={210C13BD-9159-4514-AA1E-7F6F12B47094}&page-id={86D6CF88-C889-4B56-9ADA-1B0AEF7E9FBD}&end&base-path=C:\Documents%20and%20Settings\brushm\My%20Documents\OneNote%20Notebooks))
    - also review my proposed structures for the molecular entity reagent branch

* + **Overview of Definitions and Classification Axes :**
    - **For top level molecular entities**
      * related : macromolecular reagent class that subsumes nucleic acid, protein , lipid, and carbohydrate reagents. this would be contrasted with organic small molecule reagents which are not polymeric.

* + **For nts/aa monomers:**
    - aas/nts defined on some structural basis (ie based on a general molecular template)
    - FOCUS is on mononucleotides and alpha amino acids, as these represent substituent components of information biopolymers
      * **alpha amino acid (a-aa) =** organic small molecule containing an amine group and carboxylic acid group centered around an alpha carbon atom, that also links to a side chain specific to each type of amino acid (Note that this is defined loosely enough such that unnatural small molecules that meet this requirement can qualify as aas (as opposed to enumerating all natural aas that would qualify) )
      * **mononucleotide =** This class represents any organic small molecule having part exactly 1 nucleobase, exactly 1 sugar, and 1-3 phosphate groups attached to a single sugar carbon (and such that the molecule as a whole has a defined structure regarding the arrangement of these components). Unnatural molecules can still qualify as mononucleotides as long as they meet the criteria above.
        + Note that I only model mononucleotides, as these represent precursors of nucleic acids and are thus of most biological interest.
        + Note that I could even specify the 5' carbon (vs the 2' or 3') if I want to focus ONLY on mononucleotides that comprise nucleic acids.

* + **Classification Axes**

1. natural vs unnatural/analog

2. canonical vs non-canonical

3. modified vs unmodified

\*Note that the natural vs unnatural comes first, because I need to group all unnatural monomers in one class and infer these to be analogs . . . if i applied this axis second, it could still work but i would have to infer > 1 class or unnatural monomers to be analogs (eg canonical and non-canonical unnatural monomers)

* + **For Monomer Analogs:**
    - **Monomer analogs -** 'molecular analog' is a 'molecular reagent role', so aa analogs are any org sm molecule intended to mimic an aa. analogs are by definition not natural. aa analogs can themselves be aas if they fit the structural qualifications above, or may not be aas if they do not

* + **Axes of Classification**
    - canonical vs non-canonical
    - backbone vs side chain/base
      * note that there is no concept yet of a modified aa- analog in the same sense as we use 'modified' in describing 'modified aas'. This is because analogs are not natural, but rather engineered from some specification. Thus, there is no natural reference entity to say that a modified analog varies from
      * however, we could say that there is such a concept of a modified-aa analog - ie an analog intended to mimic/ be a surrogate for a modified aa
      * alternatively, we could define modified-aa analog as an analog that is derived from (ie starts with) a natural aa which is structurally/chemically modified in some way

* + **For Polymers:**
    - **Axes of Classification:** different than for monomers, and different b/t polynucleotides and polynucleotides
      1. **polypeptides**
         1. origin (protein = ribosome generated, synthetic peptide = chemically synthesized)
         2. for proteins, endogenous vs bioengineered

\*Main eq class here =EXPERIMENTALLY modified polypeptide

* + represent INTENTIONALLY modified polypeptides (where reference starting point is a polypeptide comprised only of unmodified aa residues).

Experimentally modified polypeptides are ENGINEERED to contain at least one naturally modified aa residue or aa analog residue (or non-canonical aa residue?)

* + structurally modified
  + sequence modified

* 1. **polynucleotides**

* 1. **Polymer analogs: protein analog and nucleic acid analog**
     + analogs exist only for natural polymers (so no need for polynucleotide or polypeptide analogs)
     + **Two types:**
       1. complete analog= all monomer residues are themselves analogs (e.g. a morpholino)
       2. analog-bearing/pseudo analog - at least one but not all monomer residues are analogs (e.g. a protein where one type of aa us replaced by an analog aa during translation, or where some label is post-translationally attached)

***OK that nucleic acids/proteins with only 1 analog residue (ie 1 unnatural modification or label) = nucleic acid analogs/protein analogs?***

* 1. note the difference in interpretation here between 'monomers' and 'polymers'
     + monomers defined at molecular/atomic level - based on specific structure and atomic arrangement **(ie defined by a molecular formula**). . . so any alteration of this is 'unnatural'. polymers defined more loosely (ie not based on a molecular formula)- but based on having monomers of a certain type (and in a certain order to be a specific protein such as shh)- but must ALL monomers be of a type for it to be polypeptide/nucleic acid? or just some?

* 1. **Idea of making mononucleotides and mononucleotide analogs types of nucleic acid reagents, and similarly, alpha amino acids and their analogs types of protein reagents?**

* 1. **Still not clear on what to do to prep ReO for ERO - or by when - need CT here**

* 1. **Modeling of 'modified' and experimentally modified entities**
     + modified used for monomers and residues - here modified is in reference to a unmodified canonical monomer
     + exp modified used with polymers - exp modified implies synthetic (b/c output of a planned process) but not necessary unnatural (b/c experimental modification could mimic a natural modification, as indicated by subclasses (ie often, researchers generate some protein and then use a kinase to phosphorylate some residue of interest . . . of course, other residues may be modified, but we can only classify here based on what is specified in the specification . . . ie these are specification based classifications, not necessary reflective of entire reality

* 1. **How refer to/model parts of larger molecules**
     + **look at examples between nucleic acid, nucleotide reside, and nucleotide**
     + e.g. ok to say a nucleotide has\_part some nucleoside? or has moiety? . . . would be nice if the former way ok, because wouldn’t have to duplicate hierarchies for distinct/independent molecular entities and their moieties when part of some larger molecule (e.g. ribose and ribose moiety)
     + also, wrt residues, wikipedia says that in biochemistry this is reserved for cases of monomers that comprise polymers . . .
     + but in end, would be weird if we allowed ourselves to say has\_part some ribose but not has\_part some nucleotide (should either pursue a residue/moiety approach for all, or not for all)



