

Symbolism

Table of contents

1 Macromolecules and Symbolism.....	2
2 A DNA example.....	2
3 Tree-like Structure.....	2
4 Table of Alphabets.....	3
5 Molecular Symbol Manager Accessor.....	3

1. Macromolecules and Symbolism

In the JPL, we handle mainly macromolecules. Here is a definition taken from <http://en.wikipedia.org/>. A macromolecule is a very large molecule most often created by some form of polymerization. In the context of biochemistry, the term may be applied to the four conventional biopolymers (nucleic acids, proteins, carbohydrates, and lipids). The constituent molecules from which macromolecules are assembled are called monomers (mono=single, meros=part).

A biopolymer is made of chemical elements. Each of these monomers belonging to a class of possible molecule. Human represents them as sequence of symbols as it is far more practical than handling chemical formulae.

In the JPL we've decided to decouple a macromolecule from its symbols representation for a simple reason: the language adopted to represent a macromolecules can be ambiguous. That means a sequence of symbols can be the representation of many macromolecules. Handling ambiguities and handling macromolecules are two different function that seemed logical to us to decouple.

2. A DNA example

DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds. The four bases found in DNA are adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T). These four bases are attached to the sugar/phosphate to form the complete nucleotide.

Moreover these bases are classified into two types; adenine and guanine are fused five- and six-membered heterocyclic compounds called purines (R), while cytosine and thymine are six-membered rings called pyrimidines (Y).

Thus, Adenin can be in the same time defined as a A or a R. The code is ambiguous as R code for A and G and Y for C and T.

3. Tree-like Structure

The formalism is represented as a tree-like structure with node symbols expressing ambiguities and terminal leaves for symbol directly mapping a monomer. We call this tree-like structure an alphabet over a specific set of monomers. Here is an example of alphabet for nucleotides:

Only A, C, G and T will map the specific monomer in the pool of molecules (see [coupling page](#)).

4. Table of Alphabets

Internally, a table of symbols stores and handles every created alphabets. By now, only amino-acid, deoxyribonucleotide and ribonucleotide alphabets are provided.

5. Molecular Symbol Manager Accessor

We define a main entry point toward all actual molecular manager that extends abstract class `JPLMolecularSymbolManager`. It contains a link to every managers, by now there are four: `JPLAAManager`, `JPLRiboNucManager`, `JPLDeoxyRiboNucManager` and `JPLPepTermManager`.

All these managers are accessed by their symbol type name. Here is an example:

```
JPLMolecularManagerAccessor accessor =
    JPLMolecularManagerAccessor.getInstance();

JPLMolecularSymbolManager aaManager =
    accessor.getManager("AminoAcid");

System.out.println(aaManager.lookupSymbol('A'));
```