1 An example of our final prompt setting for question generation

```
<s>[INST] You are a biologist. You will propose one question based
on [number of triples] facts and 1 given answer. Add an explanation.
    Two examples of generating questions based on facts and answers
are:
    1. Given a fact: [Long QT syndrome-1/ROMANO-WARD syndrome] [is
associated with] [KCNQ1]. The answer is: [KCNQ1]. The question is:
Which genes are affected in ROMANO-WARD syndrome?
    2. Given two facts: [CADASIL] [is caused mostly by] [missense mutations
in the NOTCH3 gen]; [Missense mutations in the NOTCH3 gene] [is associated
with] [a cysteine residue]. The answer is [Cysteine]. The question
is: Which amino acid residue appears mutated in most of the cases reported
with cadasil syndrome?
    [/INST]Now generate a question based on [number of triples] facts
and 1 given answer: [$x_{i}$] [$r_{i}$] [$x_{j}$]. The answer is [$x_{i}$].
Let's think step by step</s>.
```

The placeholders such as "number of triples", "x_i", "r_i", "x_j" refer to the number of triples in the subgraph, the head entity, the relation and the tail entity respectively.

2 An example of output from the model

```
 Fact 1: Squamous cell carcinoma of the corpus uteri is associated
with GMP catabolic process to IMP.
 Fact 2: GMP catabolic process to IMP is a metabolic pathway.\n3.
 Fact 3: Metabolic pathways are a series of chemical reactions that
occur in living organisms to maintain life.
 Given answer: squamous cell carcinoma of the corpus uteri.

 Question: What type of cancer is associated with the GMP catabolic
process to IMP in the metabolic pathways that maintain life in living
organisms?
 Explanation:
 1. We know that squamous cell carcinoma of the corpus uteri is associated
with GMP catabolic process to IMP.
 2. GMP catabolic process to IMP is a metabolic pathway."
```

3 An example of complete SPARQL to extract subgraphs based on the network motifs

```
SELECT DISTINCT ?subj ?prop1 ?obj1 ?prop2 ?obj2
    WHERE {{
    ?subj ?prop1 ?obj1.
    ?subj ?prop2 ?obj2.
    BIND(SHA512(CONCAT(STR(RAND()), STR(?item))) AS ?random) .
    FILTER (strstarts(str(?subj), 'https://zitniklab.hms.harvard.edu/projects/PrimeKG/'))
    FILTER (strstarts(str(?obj1), 'https://zitniklab.hms.harvard.edu/projects/PrimeKG/'))
    FILTER (strstarts(str(?obj2), 'https://zitniklab.hms.harvard.edu/projects/PrimeKG/'))
        }}
    ORDER BY ?random
    LIMIT 10000
```