Sector 1: Human-AI collaboration task description creation

Human-written task description
"Describe this molecule."

AI assistance

Diversity task descriptions
"Give me some details about this molecule."
"What can you tell me about this molecule?"
"Provide a brief overview of this molecule."
"Provide a description of this molecule."
"Could you provide a description of this molecule?
......

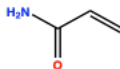**Sector 2: Information derivation from existing data**

Crawl

Data sources

Prompt input to AI

Key information
Molecule SMILES: C=CC(=O)N
CID: 6579   **Structure:** H_2N O
Description: The molecule is a colorless, odorless, crystalline solid that can react violently when...

Ask 5 questions and give answers that can be directed to the given paragraph.

Optimal format
Question 1: What is SGK1?
Answer 1: SGK1 is a serine/threonine kinase that works under acute...

Sector 3: Template-based conoversion of biological data into textual format

Biomolecular annotations
Gene Ontology (biological process)
Gene Ontology (cellular component)
DNA binding
....

Template
The designed protein must be able to regulate *{biological process}*.
The designed protein localizes to the *{cellular component}*.
A protein with *{DNA binding}* capability for targeted gene regulation.
......

Textual form
The designed protein must be able to regulate *signal transduction*.
The designed protein localizes to the *mitochondrion*.
A protein with *homeobox* capability for targeted gene regulation.
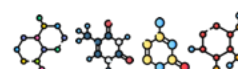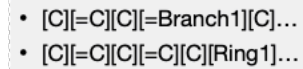......

Sector 4: Quality control

Molecule
Raw → Filter → Chemically valid → Robust descriptor
• [C][=C][C][=Branch1][C]...
• [C][=C][C][=C][C][Ring1]...

Protein
Manual Reviewed
Unreviewed
→ Annotation abundant scores → Sequence similarity clustering ≥90% <90%

Obtained biomolecules
⭐ *High quality*
⭐ *Chemically valid*
⭐ *Non-redundant*