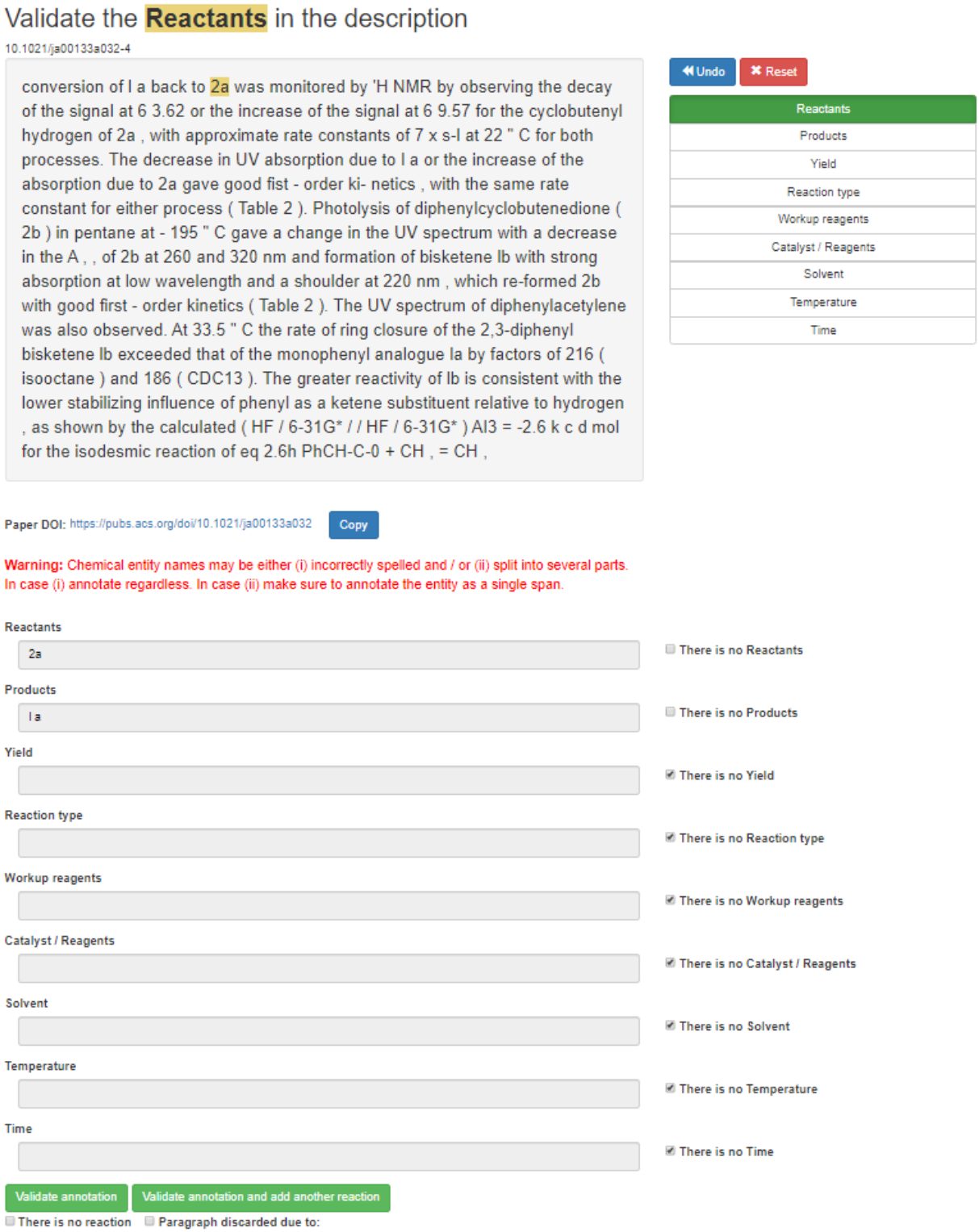
Annotation validation guide

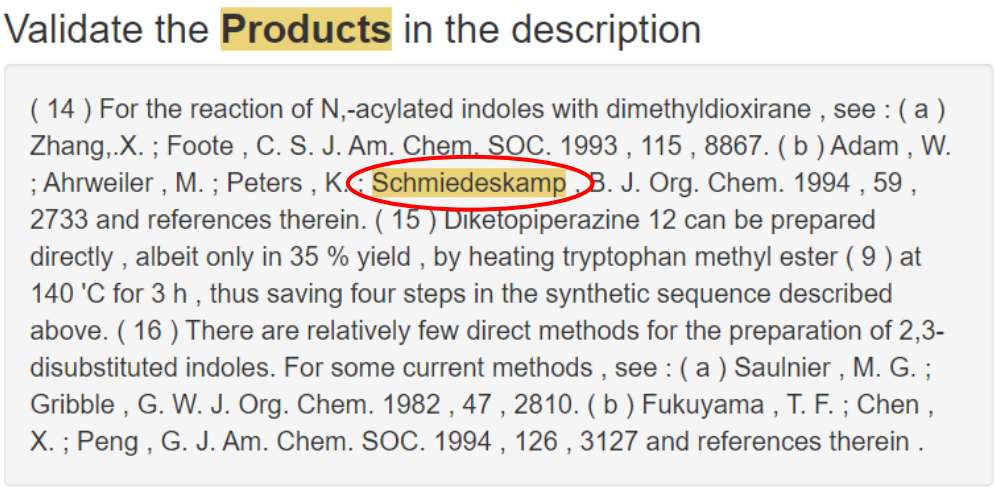
A. INTRODUCTION

The goal of this tool is the fine-tuning and validation of the annotations produced by the chemical literature annotation tool.



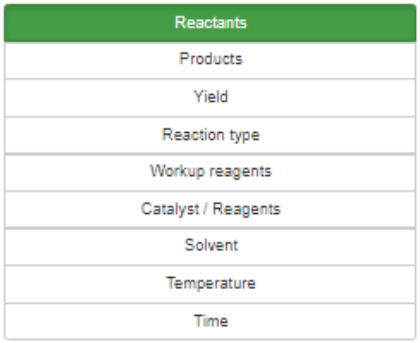
B. VALIDATION PROCESS

1. The text paragraph to be validated will be shown in the text window on the left side of the screen.
   - The paragraph shown may include from none to multiple chemical reactions.
   - The validation tool will automatically go through each and every reaction annotated in the paragraph, one at a time.
   - The annotated entities are automatically highlighted in order to facilitate the validation.

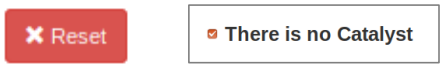


FOR EACH PARAGRAPH REPEAT SETPS 2 – 6 BELOW:

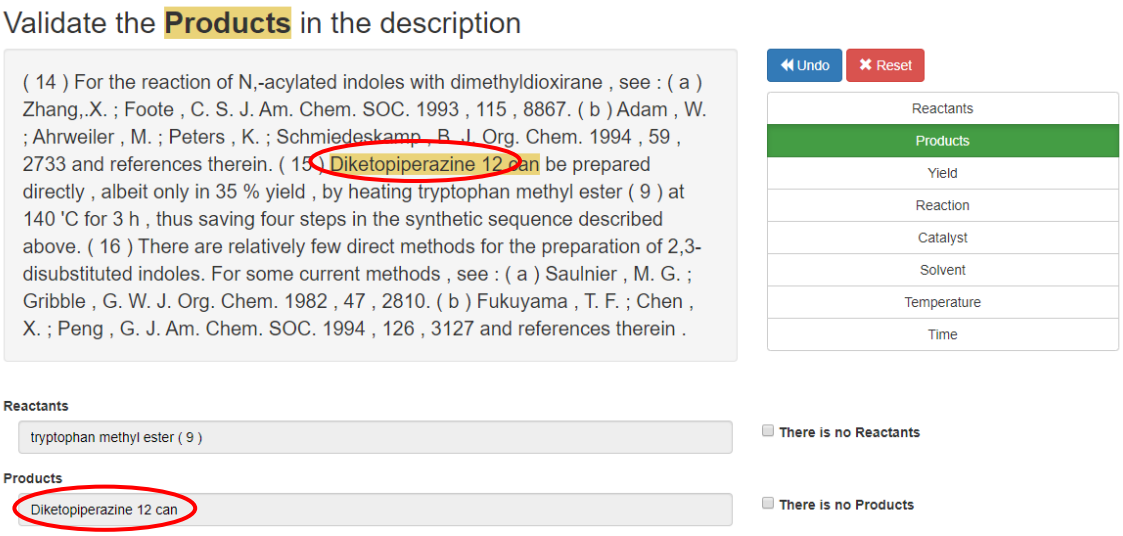
2. If a change is required, select (highlight in green) in the window on the right the entity class to be modified.



3. Proceed as per one of the following three options depending on the required type of modification:
   1. If there should be no annotation at all, reset it by clicking the reset button and check the "There is no XXX" box next to the entity class:

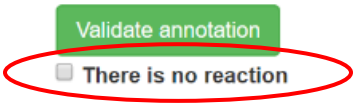


2. If the entity class contains no annotation but one or more entities should be included, add them by selecting the relevant word(s) or span(s) in the paragraph.

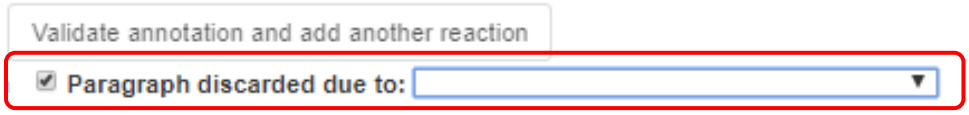


3. If the entity class contains an incorrect annotation that should be substituted, reset the field and annotate the entity class as described above.

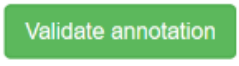
4. In case that, after the review, no entity should be annotated, the **'There is no reaction'** box must be checked in order to proceed to the next annotation.



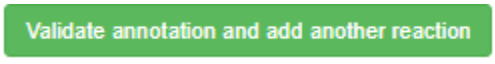
5. In case that the paragraph should be discarded, the 'Paragraph discarded due to' box must be checked and a reason chosen from the dropdown list in order to proceed to the next annotation.



6. Once the current reaction has been validated, and no additional reactions must be added, click on **'Validate annotation'** in order to review the next annotation.



7. In case additional reactions must be added, click on **'Validate annotation and add another reaction'**



D. CORRECTIONS

The **'Undo'** and **'Reset'** buttons allow to make corrections during the validation process:

- Undo: Undoes the last action
- Reset: Clears the active entity field

E. GENERAL ANNOTATION RULES

This section describes the general annotation rules for the different types of entities.

Legend:

- 'Annotate'
- '~~Do not annotate~~'

1. General

- In case a paragraph is describing more than one reaction:
    - Discard multi-step reactions.
    - If a paragraph describes multiple reactions, label them as separate reactions (e.g. 4a and 4b were reacted to form 5a and 5b, it should be two reactions: 4a → 5a and 4b → 5b). If it describes a reaction with multiple products, it should be annotated as a single reaction.
    - Annotate all entities explicitly linked to the reaction being annotated
    - Annotate as well those entities implicitly linked to the reaction being annotated, which are mentioned in other reactions within the paragraph.

      Example:

      "On the basis of conventional Knoevenagel reactions, we initially investigated reaction conditions using benzaldehyde (1a) and dimethyl malonate (2a) as a model substrate (Table 1). When a reaction was performed with 10 mol % of InBr3 in toluene at 60 °C for 8 h, only 3% of Knoevenagel product 3aa was detected (entry 1) [Reaction 1]. Thus, to promote the initial abstraction of the activated proton, the addition of 1 equiv of several bases to the reaction mixture was examined. Consequently, when the primary amine 2-aminoethanol was added, the yield was remarkably increased to 61% (entry 2) [Reaction 2]."

      Annotate [Reaction 2] as follows:

        - **Reactants**: benzaldehyde, (1a), dimethyl malonate, (2a) [Reaction 1]

- **Products**: Knoevenagel, 3aa [Reaction 1]
- **Yield:** 61% [Reaction 2]
- **Reaction:** Knoevenagel [Reaction 1]
- **Catalyst:** InBr3 [Reaction 1], primary amine 2-aminoethanol [Reaction 2]
- **Solvent:** toluene [Reaction 1]
- **Temperature:** 60 [Reaction 1]
- **Time:** 8h [Reaction 1]

- For reversible reactions, annotate as two reaction if the reverse reaction is explicitly described (e.g. different conditions are used to reverse the reaction)
- Do not annotate expected results but actual results. Example:
  - 'Since CuCl(PPh 3) is a widely studied and common copper salt, the active catalytic species was expected to be ~~CuCN (PPh 3)~~ in our optimized reaction condition'
- Annotate not only the reactions obtained in the actual experiments but also past reactions described. Example:
  - 'Both the Ullmann and Buchwald-Hartwig aminations are well-known copper-catalyzed crosscoupling reactions between an aryl halide and an amine'

2. **Conditions**

- Do annotate compound expressions when they form a single word. Examples:
  - 'Copper-catalyzed crosscoupling reactions'
  - 'Phosphine-catalyzed [3 + 2] annulation'
- Do not annotate references to reaction conditions. Examples:
  - '~~condition A~~ shows the highest reactivity'
- Do annotate generic reaction conditions. Example:
  - 'Elevated temperature'
- Do annotate physical units. Example:
  - '32 C', '32 °C'', '25 F'
  - '25 min', '8 h', '10 hours'

3. Chemical entities (reactants, products, catalysts, reagentes, and solvents)

- ALL ENTITIES
  - Annotate both the entity and its corresponding reference when they are shown together. Example:

        - 'We started our investigation by examining the Bu3P catalyzed reaction of 2-hydroxyisoindoline-1,3-dione (1a) and ethyl propiolate (2a)'
    - Annotate references (to Schemes, Tables, Figures, etc.) to entities. Example:
        - 'The reaction of 1a with 2a in the presence of Bu3P in DMF…'
    - Annotate all references to an entity when shown together. Example:
        - pivalonitrile (2,2,-dimethylpropanenitrile, bp 105 °C)
    - Only label the chemicals, without additional details about concentration, state, etc. E.g. "1% aqueous FeCl3", only label "FeCl3"
    - If the text mentions a range of chemicals, only annotate those explicit text (e.g. 4g – 4i were used to form 5a – 5c: only annotate two reactions 4g->5a and 4i->5c). If the range is not separable, e.g., 4a-c or 4a-4c (without space around the "-"), discard the reaction
    - Default to reactant if unsure between reactant and catalyst/reagent. Default to catalyst/reagent if unsure between catalyst/reagent and solvent

- PRODUCTS AND REACTANTS
    - Do not annotate references to entity types but to specific entities. Examples:
        - 'the initial ~~hexameric-PPh3-bounded-copper-cyanide~~ species'
        - 'the expected ~~alcohols~~'
        - 'the corresponding ~~alkene~~ ~~3aa~~'
        - 'however, only ~~colorless crystals~~ were obtained'
    - Do not annotate generic references to entities. Examples:
        - '~~the product~~', '~~final product~~', '~~the reactant~~'
        - '~~the corresponding product~~'
        - '~~same starting materials~~'
    - Do not annotate entity classes but only specific entities. Examples:
        - 'While the ~~C-2 arylated product~~ (~~4aa~~) was the major product'
        - '~~Benzoxazoles bearing substituents with diverse electronic properties such as methyl (1b-1d), phenyl (1e), and chloro (1f) groups'~~
        - 'effectively promotes the Knoevenagel condensation of ~~aromatic / aliphatic / heteroaromatic aldehydes~~ with a ~~variety of activated methylene compounds~~'

- REAGENTS, CATALYSTS AND SOLVENTS
    - Annotate even if the text only describes the type of chemical (e.g. "Pd catalyst": annotate "Pd"; "in-situ generated catalyst": annotate "in-situ generated").

4. **Yields**

- Annotate generic yield expressions. Examples:
    - ‘similar yield’, ‘lower yield’, ‘higher yield’, ‘diminished yield’, ‘improved reaction yield’
- Annotate all yield measures in case more than one is provided. Examples:
    - ‘Higher yield (62%)’
- Alternative expressions of yield: **TBD – Chemistry**
    - ‘increased catalytic efficiency’
    - ‘improved reactivity’
    - ‘perfect selectivity’
    - ‘significant effect’
    - ‘highest conversion’
    - ‘no reactivity’
    - ‘no further reaction observed’

5. Reaction type

- Annotate if the description is a verb (e.g. A is oxidized to B: annotate "oxidized" as the reaction type).