Reply to a review of

“Modelling of DNA Mismatch Repair with reversible process calculus”

by Stefan Kuhn and Irek Ulidowski submitted for publication in the Journal of Theoretical Computer Science, 2021

The text of the review is included for completeness and is displayed using normal font. Our responses and comments are displayed in italic font.

In this paper the authors show that an abstraction of an important biological mechanism, i.e., the DNA Mismatch Repair, may be modelled in a reversible process calculus. The abstraction here ignores the complicated geometrical structure of the DNA and also the mechanisms required for making local part of interest of the DNA available the reactions (e.g., gene expressions).

Nevertheless, it is an good step forward in linking process calculi to biological processes.

To this aim the authors reshape a former process calculus CCB (the Calculus of Covalent Bonding), previously introduced and use by the authors for a related correction mechanism - Base Excision Repair. The extension requires allowing prefixing with collections of bonding sits and a more careful operational semantics.

The paper is interesting, correct, and of interest for this journal. I is also well written - an extra

value of the paper is the good survey of the process algebra based approach to model biological processes.

I strongly recommend its publication in this journal. A few more comments and suggestions for improving the presentation are inserted below.

**General comments**:

1. The process terms describing the systems are long and somehow difficult to follows. For instance, the given example, involving only 6 pairs of bases, results in a half a page formula (page 22). While there is some effort to simplify the term, the notation is still long and much of the text in pages 23-29 contains redundant information. A suggestion here is the use something similar to the global-local handling notation used in the K-framework (see, e.g., https://kframework.org/).

>I understand this is asking for some formal short notation. I will check that kframework.

IU: is it realistic to use it in this paper?

I am not sure how we could use it. The problem is that what we use is already minimal in that sense that only parts which will change at some point are shown. We could of course shorten it more, but that would mean things change between steps, which I think is not good. We could even do that informally, just saying “parts not involved are left out in each step”. Probably K framework could formalize this, but I don’t think it would add much (apart from additional overhead, we would need to specify rules in K etc.) and not really help with reading it.

2. Quite often, graphs are useful in grasping intuition on process terms. Maybe developing a

graphical representation of the CCB processes may simplify understanding and pave a way towards including geometrical information on DNA shape.

>This asks not for some graphics (which we have got), but a graphical version of the calculus. I will look at this. Generally, both comments may not be easy to do without needing more space.

IU: I would not like to propose any new graphical representation of the calculus. Let us think more how to improve further representation of the system on pages 23-29.

What we could try is to make a more formal notation for the bonds. We have three types:

1. Unchanged (normal line)
2. New (bold)
3. Broken (dotted)

If we use that everywhere (which would mean to make e. g. some bonds bold in Figure 10) we could call this a proper graphical representation.

**Detailed comments** (page, global line numbers):

p2,25: DNA errors often occur when DNA is used for reading the information about the proteins to be synthesized in the cells.

>Not sure about this. We are talking about wrong replicated DNA. It can happen that a protein is produced wrongly from DNA (or RMA), but that is not a matter of "gene repair". Might be interesting, but it is different.

IU: maybe we need to add a comment or clarification to satisfy the reviewer. Write this please and let me check.

*There are indeed many different sources of DNA damage, including damage during transcription, i. e. production of mRNA from DNA. We have decided here to focus on DNA replication errors. There are certainly more potential issues to discuss, including DNA repair as part of transcription (*[*https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1470299/*](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1470299/)*). We believe that for this paper focusing on one case is justified.*

p5,110: bee

>Corrected.

p6,130: Maybe a short clarification of the distinction strong-weak actions is useful at this point.

>Not sure what exactly the reviewer is thinking about. The distinction is really in the operational semantics, and those follow later.

IU: I think it might be good to explain this in words as well in addition to be just used in SOS rules. I can look into this.

p6,140: : "use α, s and s, s′ to denote a concatenation" - the notation is unclear; include some parentheses ( ), or " " to separate the terms from the text

>Not sure what's best. If we say "use (Î±, s) and (s, sâ€²)" it looks like the () are part of the syntax. And more spaces is odd. We have the italics. Doesn't that show it?

IU: Let me think about it.

P7, 150: "A collection µ is either σ or σ | σ" - if the is a BN-like syntax, should be σ | µ

>I think that's a correct remark, it needs changing. IU: Agree. I will do it.

p7,155: The syntax includes also recursive definition, not included in that one-line description

>Don't understand this. The "one-line description" is P ::= S | mu:P | P|Q | P\L, I assume - this has recursion?

IU: I think it is still about line 155. I will think more.

p8,185: "(s; b).C | P | C | C | P | C \ L" - confusion of two roles of "|"

>I think we need to use spaces similar to p7. IU: Yes. Let me sort it out.

p10,Fig3: b does not appear in this rule

>b is in the s rule.

p11,Fig4: similar

>we use beta here. IU: let me sort it out.

p12,Fig5: similar somehow, it is unclear why do you use condition \* and \*\*; why not simply

include those particular cases explicitly?

>for me, this was just because of space. IU: I think we just add \* and \*\* to rules as requested. But, personally, I don’t think it will improve matters. SK don’t think either, but if they insist on it

p12,Ex3: it seems to be a confusion here between b,d,f - check this example again

>Corrected. uses b and d (not b and b) since b is weak and d is not.

p13,Fig6: what is "t" in concert2 act?

>That should be eta. Yes, I will change it.

p16,275: Is this strong-weak action explanation useful at p6,130?

>I am not sure if just moving this is useful. And if we want to give the reason etc. on 130 as well, it means reorganising a lot. I will take care of this.

p22,top: "The MMR system in full detail..." Actually, this is the result after all reactions in the

previously described MMR system are used. It shows an order of firing transitions in the original system to arrive at this form, directly linked to the picture in Fig.9

>I think the reviewer believes that the "previously described system" (p 21) is different from this. It is not, for my understanding. There is just less detail. Not sure how to word this best. It does link to fig. 9, but we say so in line 395. IU: please try to explain this more fully, and let me check if I understand it better afterwards.

*We believe there are three descriptions the reviewer is considering here, two of them as processes, one graphical. Firstly, there is the process following the sentence “The system shown in Figure 9 is modelled in CCB as follows:”, secondly the process following the sentence “The MMR system in full detail (but without restriction) is given below”, and thirdly Figure 9. All of those actually describe the same system. The first process description just has the process names from the previous pages written, whereas the second process description has the full definitions, including bonds, which are in the figures as well. It does not imply there is a development from the first to the second process. We do not consider how we arrived at this state, and give no order of transitions. The first process description is intended to give an overview, the second the details, which are shown in the Figure as well.*

p30,Fig14: "the offending base C\_3" - is G\_3

>Corrected.