Non-enzymatic recombination of RNA: Ligation in loops

DOI: 10.1016/j.bbagen.2017.10.019

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Biochimica et Biophysica Acta (BBA)-General Subjects

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Title: Non-enzymatic Recombination of RNA: Ligation in Loops

Article Type: Regular Paper

Keywords: non-enzymatic RNA ligation; RNA world; origin of life; molecular dynamics.

Corresponding Author: Professor Marina A Zenkova, Ph.D, Doctor of Sciences
Corresponding Author's Institution: Institute of Chemical Biology and Fundamental Medicine SB RAS
First Author: Yaroslav Staroseletz

Order of Authors: Yaroslav Staroseletz; Sergey Nechaev; Elena Bichenkova, Ph.D; Richard Bryce, Ph.D; Catherine Watson; Valentin Vlassov, Ph.D, Academician of the RAS; Marina A Zenkova, Ph.D, Doctor of Sciences

Abstract: Background
While the RNA world hypothesis is widely accepted, it is still far from complete: the existence of self-replicating ribozyme, consisting of potentially hundreds of nucleotides, is a core assumption for the majority of RNA world models. The appearance of such long RNA molecules under prebiotic conditions is not self-evident. Recombination seems to be a plausible way of creating RNA diversity, resulting in the appearance of functional RNAs, capable of self-replicating.

Methods
We report here on the study of recombination process modeled with two 96 nts RNA fragments. Detection of recombination products was performed with RT-PCR followed by TA-cloning and Sanger sequencing.

Results
A wide range of recombinant products was detected. We found that (i) the most efficient ligation was observed for RNA species forming bulges or internal loops, with ligation partners located within the loop; (ii) a strong preference was observed for formation of a few types of major products with a large variety of minor products; (iii) ligation could occur with participation of either 2',3'-cyclophosphate or 5'ppp; (iv) the presence of key reaction components, i.e. 5'ppp-RNAs, enabled the formation of additional types of product; (v) molecular dynamics simulations of one of the most abundant products suggests that the ligation results in a preferable formation of 2'-5' rather than 3'-5' linkages.

Conclusions
The study demonstrates regularities of new RNA molecules formation with non-enzymatic recombination process.

General significance
Our findings provide new data supporting the RNA World hypothesis and showing the way of new RNA sequences emergence under prebiotic conditions.
Response to Reviewers: We are pleased to revise and update the article entitled ‘Non-enzymatic Recombination of RNA: Ligation in Loops’ in light of the appended reviewer comments for consideration by the BBA General Subjects journal. We are very grateful to the reviewers for critical reading and constructive suggestions.

Please find our below point by point replies on the reviewer’s comments.

Reviewers’ #1 comments.

The reaction drawn in the graphical abstract could be clearer.

Respond to Reviewers comment No.1
Corrected. We created new variant of graphical abstract trying to make it as clear as possible.

On the bottom left of the graphical abstract figure, the 5' O that attacks the phosphate looks like an O-, but if I understand right, it is an O in another phosphodiester bond, so it should look like a bond, not a minus (like the O- in the phosphate group).

Respond to Reviewers comment No.2
Thank you for turning our attention to this point. 5’O that attacks the phosphate is 5’OH, which is formed either at cleavage stage or belongs to initial RNA molecules.

I think there is a new strand (green below) created by the cleavage of the bottom strand. This is ignored in the graphical abstract.

Respond to Reviewers comment No.3
Yes, a new strand is created by the consecutive cleavage and ligation events. We try to clarify this in the new edition of the graphical abstract.

It would be helpful if this figure could show all parts of the strands. Also the 'zooming in' from the top of the figure to the bottom is not very clear - how does the chemistry part fit into the schematic RNA structure part? My cartoon below illustrates what I think is happening, but I do not see it clearly in any of the figures in this paper, so I may have misunderstood. Is there is a new cyclic 2'3' bond at the end of the green strand? The circle is my attempt to illustrate this.

Respond to Reviewers comment No.4
The cartoon suggested by the reviewer is correct with two amendments. Firstly, red strand with 2',3' cyclic phosphate is not presented in the reaction mixture from the very start, but is formed at the cleavage stage of recombination process. Secondly, the recombination consists of at least of two main stages, namely cleavage and ligation. Thus green strand with a new cyclic 2',3' bond is the result of the green-blue strand cleavage and red-blue strand is the result of red and blue strands ligation.

I would like to know whether this reaction is reversible. If you interchange a cyclic bond and a phosphodiester bond it should be $\Delta G = 0$ -
because it is the same both ways at the level of the chemical bonds. However, if the end of the green strand is different from the red, then the reaction will have a preferred direction. (I now see there is a discussion about 'energy neutral' reactions on p17 - but this is coming only at the end of the paper).

Respond to Reviewers comment No.5
The end of the green strand is the same as the end of the red one therefore the reaction is reversible. Discussion of recombination as a series of neutral reactions passes through the manuscript: p.4 line 6, p.18 line 24.

The 'selection of modules' paragraph in the middle of p15 seems relevant to the previous point - the M2 fragment 'wins' over the other because it has a stronger secondary structure. So it seems that you are saying that the forward and backward reactions do not proceed at equal rates because of properties of the structures, even if the formation and breaking of chemical bonds is the same in the two directions?

Respond to Reviewers comment No.6
Yes, exactly, properties of the RNA structure (both initial and formed after cleavage) determine the relative rates of forward and back reactions. M2 fragment “wins” in competition because it forms more stable complex with a fragment of HIV RNA, than other fragments.

Is there a way to study a single pair of reactions - forward and back - and measure their relative rates? This would show that they are not equal, even if they are equivalent at the level of the chemical bonds.

Respond to Reviewers comment No.7
Yes, there are ways to study a single pair of reaction and measure their relative rates. For instance, similar measurements were done for paired cleavage/ligation reaction catalyzed by hammerhead ribozyme (Nelson JA 2005). Indeed, this work has to be done for each specific structure identified in this work, and we plan to perform this investigation in a future.

I don't understand how to see from 3C that the M2 fragment wins (line38 of p15). What are the two structures being compared here before and after the reaction?

Respond to Reviewers comment No.8
Our suggestion that M2 fragment “wins” is based on the fact that multiple recombinant products include this fragment. M2 fragment 5’-UGAGGA-3’ shown at 3C “competes” with 5’-AGGGA-3’ fragment from HIV. At the moment, quantitative comparison of these two structures is not possible. All what we know are initial system state of the system, consisting of two 96-nts RNA molecules, HIV and M2, and final state of the system, consisting of a wide range of HIV and M2 fragments and recombinant molecules, including those shown at Fig. 3C. We have no opportunity to reconstruct the path of product 3C formation from HIV and M2 in all details, step by step, with all intermediate stages. This path includes at least two cleavage events and one ligation event, though the number of cleavage events can be greater. Thus, the fact that M2 fragment “wins” in this competition shows
the state of the system at the stage of analysis but not the particular recombination reaction.

In general, in all the pictures of secondary structures in Figs 3, 4 etc, I am not sure where the recombination is occurring.

Respond to Reviewers comment No.9
Corrected, new linkages formed upon ligation are marked by dot lines in Fig. 3.

In the discussion of ligation (bottom of p15), it says ligation occurs mostly in the internal loops etc - here we are still talking about recombination, right? The title of this paper calls it recombination!

Respond to Reviewers comment No.10
Yes, we are talking about recombination because it consists of at least two main stages cleavage of initial RNA molecules and ligation of fragments of these RNA molecules.

But in the next paragraph it mentions 'non-enzymatic ligation in RNA loops' in the work of Bartel. This case is really a ligation with a triphosphate at the end, not a recombination with a cyclic phosphate at the end. In the ligation case, there is no equivalent of the green strand, and ΔG is not zero for this reaction.

Respond to Reviewers comment No.11
In the above mentioned work Bartel with co-workers investigated the ligation reaction only with 5′-triphosphate as activated form of the phosphate. Thus, the data on this work partly correspond to the conditions when 5′-ppp-RNAs was used in the reaction. The differences between the recombination products formed under conditions when 5′-OH or 5′-ppp-RNAs were used are discussed throughout the manuscript (see p. 13 2.11 G4 products). We divide recombination reaction into two main stages namely cleavage and ligation. In reaction series with dephosphorylated RNAs the only activated form of phosphate is a cyclic phosphate, derived at the cleavage stage, therefore ligation occurs with participation of 2′,3′ cyclic phosphate and 5′-OH. In reaction series with 5′-ppp-RNAs ligation occurs with 5′-ppp and 3′-OH as well that results in formation of larger number of new RNA molecules.

In the work of Lehman there is always a single G at the end of one strand that it lost when the recombination happens.

Respond to Reviewers comment No.12
Relevant discussion at the end of the manuscript (see p. 21 line 22)

Is it actually possible for a cyclic phosphate to react with the end of the other strand too? In that case there would be no green strand for the recombination either. Please make all these things clear.

Respond to Reviewers comment No.13
Yes, it is actually possible for a cyclic phosphate to react with the end of the other strand. On the reviewers’ cartoon the red strand
theoretically can react both with the end of green and blue strands. In the latter case the reaction may occur after green-blue strand cleavage. In practice the possibility of this or that reaction occurrence depends on the RNA structures.

An important point about recombination reactions is that the mean length of the strands is not changed in any one reaction. For example, in the cartoon above, the mean length of the strands is just the sum of the lengths of the red, green and blue, divided by 2. This is the same before and after. Recombination cannot therefore systematically create long sequences, however it can change the distribution of lengths. It is likely to broaden the distribution of lengths: if you start with all medium length sequences, then you might form a distribution with both very short and very long. In this sense, recombination might favour the formation of long ones that are functional. One case like this was calculated by Wu and Higgs (2011) - Comparison of the roles of nucleotide synthesis, polymerization and recombination in the origin of autocatalytic sets of RNAs. Astrobiology 11: 895-906. Here it was shown that if recombination occurs at equal rates at all points, the expected distribution of lengths is exponential. However, the distribution expected from many simple polymerization reactions is also exponential. In that case, adding random recombination reactions does not change the distribution. Wu and Higgs gave an example related to the Azoarcus ribozyme where recombination is not equal at all points, where the distribution can be shifted by recombination. I think it is clear from this paper that recombination generates diversity if you start from a small number of starting sequences. It is not fully clear that recombination creates more diversity than would already be created by random polymerization. Hence it is not really clear that recombination 'solves' the problem of the origin of the RNA world. As recombination does not change the mean length, there has to be a way of making fairly long strands by polymerization before recombination can act on these. In other words - where do the M2 and HIV sequences come from in the first place?

Respond to Reviewers comment No.14
Thank you for valuable remarks. We added this discussion and relevant references to manuscript (p. 3, lines 17-21; p. 5, lines 4-12, marked by green).

Some typos: In Fig 1 - Route not Root. p22 - routes not routs. In figure captions - dangling not dandling.

Respond to Reviewers comment No.15
Corrected

Reviewer #2 comments.

In this manuscript the authors characterize various mechanisms, products, and kinetics of spontaneous RNA-RNA recombination among 96-nt RNAs. This is a very thorough work, well done, well written, and well referenced. It has direct relevance to plausible prebiotic pathways in
the RNA world leading from short, abiotic RNA oligomers to larger catalytic RNAs. I can recommend publication provided the authors address a few salient points:

1. My main comment is that the authors should compare their loop recombination reactions (route B in Figure 1) to the recombination reactions described by the Di Mauro group (e.g., Pino et al., 2013 Entropy). This is not my work, but it strikes me that they may be discussing the same type of reaction. The Di Mauro group has presented two types of reactions, one a recombination (which I think is very relevant here) and a second one that is a ligation (which I personally remain unconvinced of, but in any event may not be relevant here). In any event, a contrast between the reactions observed in the current manuscript and those discussed by the Di Mauro group is a requisite. In particular, the Product Pr-1 seems like a direct analog to the Italian group's model.

Respond to Reviewers comment No.1
Thank you for valuable comment. We compared our system and recombination products with those described by the Di Mauro group. This discussion and relevant references were added to manuscript. See P. 17 and 18, starting line 24, marked by green.

2. The authors should discuss briefly how prebiotically relevant are 96-mers. It seems like they based this strictly off of some previous work, including their own work. The previous papers from this group (Lutay & Vlassov) did have short oligomers that were prebiotically relevant in that they were all under 20 nucleotides in length to start, the length that is close to the maximum for abiotic polymerization that has been achieved in vitro. However, by using 96mers they have created the question of whether the same kinds of reactions are achievable with smaller strands, and the question of whether the length impacts the reaction, which seems to be a possibility. On the other hand, the 96mers have the advantage of being known viral RNAs that can be worked into cells; they have also worked with them previously, according to a citation. So there is that practical aspect, even if the theoretical questions linger. The size issue played a role when they discuss their molecular cloning - they say that the variety of products is very large. So in the future it might be simpler to use smaller RNAs. One might imagine that there were more than just prebiotic reasons to do this, but the point of the paper seems to talk about prebiotic chemistry.

Respond to Reviewers comment No.2
We added discussion on prebiotic relevance of 96-mers RNAs to the manuscript. P. 5, lines 4-12, marked by green.

3. It was interesting that they suggested the 3' OH of the ligating fragment was 60 times more active in GA tandem mismatch bearing loops in terms of forming the phosphodiester bond than compared to the 2' hydroxyl which is 6 times more reactive toward activated phosphate esters in an unconstrained aqueous environment. I have wondered on occasion if the 2' OH of the Italian group's model could form the stem loop as readily as the 3' OH. Please discuss.

Respond to Reviewers comment No.3
We added discussion on comparative reactivity of 2'-OH and 3'-OH groups to the manuscript. See P. 18, lines 7-11, marked by green.
We thank you for all comments.

Yours faithfully,
Zenkova Marina
Institute of Chemical Biology and Fundamental Medicine
Prof, **Lennart Nilsson**
Executive Editor, BBA General Subjects
Karolinska Institutet,
Department of Biosciences and Nutrition (BioNut), H2
Novum 141 83 Huddinge, Sweden.

Dear Professor Nilsson,

We are pleased to submit the attached the revised manuscript entitled *Non-enzymatic Recombination of RNA: Ligation in Loops* for consideration by the BBA General Subjects journal.

This submission presents the study of the spontaneous, non-enzymatic RNA recombination, modelled with two 96 nucleotides fragments of viral RNAs, *i.e.* HIV-1 RNA and Influenza virus M2 protein RNA. A wide range of recombinant products was detected and analysed in terms of frequency of occurrence and structural preferences of the ligation site. Using molecular dynamics simulations, we examined the most abundant recombination product in order to evaluate the probability of formation either 2'-5' or 3'-5'-linkages and understand the molecular mechanisms of non-enzymatic recombination.

**THE SIGNIFICANCE AND POTENTIAL IMPACT**

This pilot study provides the position to explore the fundamental principles of molecular interactions and regulatory mechanisms of RNA recombination and investigate the key factors promoting these reactions at the molecular level. Recombination is a contributing factor to the rapid evolution of viruses and the emergence of new, pathogenic strains. The insight into this area may shed the light on the origin of rapid, uncontrollable mutations of RNA viruses and
disclose the processes underlying rapid appearance of new viral strains. As such, this knowledge may also influence future strategies for the prevention and treatment of pathogenic viral diseases.

THE INTEREST TO THE WIDE GENERAL READERSHIP

A wider scientific community may benefit from this new knowledge, as non-enzymatic recombination remains more enigmatic than its enzyme-mediated counterpart, and the structural rules by which it occurs are still largely unknown. The understanding of the molecular mechanisms of non-enzymatic recombination of RNAs may not only explain some aspects of the origin of life, but also provide a vital clarification of enormous structural and functional diversity of RNAs, which at some stage of prebiotic life development seemed to carry out both the task of storage of genetic information and the catalytic roles required for survival of primitive self-replicating systems.

We would like to confirm that this work is original and has not been published elsewhere and is not currently under consideration for publication elsewhere.

The manuscript has been read and approved by all the authors (i.e. Yaroslav Staroseletz, Sergey Nechaev, Elena Bichenkova, Richard Bryce, Catherine Watson, Valentin Vlassov and Marina Zenkova).

Thank you for considering this submission for us.

Yours faithfully,

Prof. Marina Zenkova
Institute of Chemical Biology and Fundamental Medicine
Response to Reviewers

We are pleased to revise and update the article entitled ‘Non-enzymatic Recombination of RNA: Ligation in Loops’ in light of the appended reviewer comments for consideration by the BBA General Subjects journal. We are very grateful to the reviewers for critical reading and constructive suggestions.

Please find our below point by point replies on the reviewer’s comments.

**Reviewers’ #1 comments.**

The reaction drawn in the graphical abstract could be clearer.

*Corrected. We created new variant of graphical abstract trying to make it as clear as possible.*

On the bottom left of the graphical abstract figure, the 5’ O that attacks the phosphate looks like an O- , but if I understand right, it is an O in another phosphodiester bond, so it should look like a bond, not a minus (like the O- in the phosphate group).

*Thank you for turning our attention to this point. 5’O that attacks the phosphate is 5’OH, which is formed either at cleavage stage or belongs to initial RNA molecules.*

I think there is a new strand (green below) created by the cleavage of the bottom strand. This is ignored in the graphical abstract.

*Yes, a new strand is created by the consecutive cleavage and ligation events. We try to clarify this in the new edition of the graphical abstract.*

It would be helpful if this figure could show all parts of the strands. Also the ‘zooming in' from the top of the figure to the bottom is not very clear - how does the chemistry part fit into the schematic RNA structure part? My cartoon below illustrates what I think is happening, but I do not see it clearly in any of the figures in this paper, so I may have misunderstood. Is there a new cyclic 2’3’ bond at the end of the green strand? The circle is my attempt to illustrate this.

*The cartoon suggested by the reviewer is correct with two amendments. Firstly, red strand with 2’,3’ cyclic phosphate is not presented in the reaction mixture from the very start, but is formed at the cleavage stage of recombination process. Secondly, the recombination consists of at least of two main stages, namely cleavage and ligation. Thus green strand with a new cyclic 2’,3’ bond is the result of the green-blue strand cleavage and red-blue strand is the result of red and blue strands ligation.*
I would like to know whether this reaction is reversible. If you interchange a cyclic bond and a phosphodiester bond it should be $\Delta G = 0$ - because it is the same both ways at the level of the chemical bonds. However, if the end of the green strand is different from the red, then the reaction will have a preferred direction. (I now see there is a discussion about 'energy neutral' reactions on p17 - but this is coming only at the end of the paper).

The end of the green strand is the same as the end of the red one therefore the reaction is reversible. Discussion of recombination as a series of neutral reactions passes through the manuscript: p.4 line 6, p.18 line 24.

The 'selection of modules' paragraph in the middle of p15 seems relevant to the previous point - the M2 fragment 'wins' over the other because it has a stronger secondary structure. So it seems that you are saying that the forward and backward reactions do not proceed at equal rates because of properties of the structures, even if the formation and breaking of chemical bonds is the same in the two directions?

Yes, exactly, properties of the RNA structure (both initial and formed after cleavage) determine the relative rates of forward and back reactions. M2 fragment “wins” in competition because it forms more stable complex with a fragment of HIV RNA, than other fragments.

Is there a way to study a single pair of reactions - forward and back - and measure their relative rates? This would show that they are not equal, even if they are equivalent at the level of the chemical bonds.

Yes, there are ways to study a single pair of reaction and measure their relative rates. For instance, similar measurements were done for paired cleavage/ligation reaction catalyzed by hammerhead ribozyme (Nelson JA 2005). Indeed, this work has to be done for each specific structure identified in this work, and we plan to perform this investigation in a future.

I don't understand how to see from 3C that the M2 fragment wins (line38 of p15). What are the two structures being compared here before and after the reaction?

Our suggestion that M2 fragment “wins” is based on the fact that multiple recombinant products include this fragment. M2 fragment 5’-UGAGGGA-3’ shown at 3C “competes” with 5’-AGGGA-3’ fragment from HIV. At the moment, quantitative comparison of these two structures is not possible. All what we know are initial system state of the system, consisting of two 96-nts RNA molecules, HIV and M2, and final state of the system, consisting of a wide range of HIV and M2 fragments and recombinant molecules, including those shown at Fig. 3C. We have no opportunity to reconstruct the path of product 3C formation from HIV and M2 in all details, step by step, with all intermediate stages. This path includes at least two cleavage events and one ligation event, though the number of cleavage events can be greater. Thus, the fact that M2 fragment “wins” in this competition shows the state of the system at the stage of analysis but not the particular recombination reaction. Image below shows competing structures.
In general, in all the pictures of secondary structures in Figs 3, 4 etc, I am not sure where the recombination is occurring.

Corrected, new linkages formed upon ligation are marked by dot lines in Fig. 3.

In the discussion of ligation (bottom of p15), it says ligation occurs mostly in the internal loops etc - here we are still talking about recombination, right? The title of this paper calls it recombination!

Yes, we are talking about recombination because it consists of at least two main stages cleavage of initial RNA molecules and ligation of fragments of these RNA molecules.

But in the next paragraph it mentions 'non-enzymatic ligation in RNA loops' in the work of Bartel. This case is really a ligation with a triphosphate at the end, not a recombination with a cyclic phosphate at the end. In the ligation case, there is no equivalent of the green strand, and \( \Delta G \) is not zero for this reaction.

In the above mentioned work Bartel with co-workers investigated the ligation reaction only with 5'-triphosphate as activated form of the phosphate. Thus, the data on this work partly correspond to the conditions when 5'-ppp-RNAs was used in the reaction. The differences between the recombination products formed under conditions when 5'-OH or 5'-ppp-RNAs were used are discussed throughout the manuscript (see p. 13 2.11 G4 products). We divide recombination reaction into two main stages namely cleavage and ligation. In reaction series with dephosphorylated RNAs the only activated form of phosphate is a cyclic phosphate, derived at the cleavage stage, therefore ligation occurs with participation of 2’,3’ cyclic phosphate and 5’-
OH. In reaction series with 5'-ppp-RNAs ligation occurs with 5'-ppp and 3’-OH as well that results in formation of larger number of new RNA molecules.

In the work of Lehman there is always a single G at the end of one strand that it lost when the recombination happens.

Relevant discussion at the end of the manuscript (see p. 21 line 22)

Is it actually possible for a cyclic phosphate to react with the end of the other strand too? In that case there would be no green strand for the recombination either. Please make all these things clear.

Yes, it is actually possible for a cyclic phosphate to react with the end of the other strand. On the reviewers’ cartoon the red strand theoretically can react both with the end of green and blue strands. In the latter case the reaction may occur after green-blue strand cleavage. In practice the possibility of this or that reaction occurrence depends on the RNA structures.

An important point about recombination reactions is that the mean length of the strands is not changed in any one reaction. For example, in the cartoon above, the mean length of the strands is just the sum of the lengths of the red, green and blue, divided by 2. This is the same before and after. Recombination cannot therefore systematically create long sequences, however it can change the distribution of lengths. It is likely to broaden the distribution of lengths: if you start with all medium length sequences, then you might form a distribution with both very short and very long. In this sense, recombination might favour the formation of long ones that are functional. One case like this was calculated by Wu and Higgs (2011) - Comparison of the roles of nucleotide synthesis, polymerization and recombination in the origin of autocatalytic sets of RNAs. Astrobiology 11: 895-906. Here it was shown that if recombination occurs at equal rates at all points, the expected distribution of lengths is exponential. However, the distribution expected from many simple polymerization reactions is also exponential. In that case, adding random recombination reactions does not change the distribution. Wu and Higgs gave an example related to the Azoarcus ribozyme where recombination is not equal at all points, where the distribution can be shifted by recombination. I think it is clear from this paper that recombination generates diversity if you start from a small number of starting sequences. It is not fully clear that recombination creates more diversity than would already be created by random polymerization. Hence it is not really clear that recombination 'solves' the problem of the origin of the RNA world. As recombination does not change the mean length, there has to be a way of making fairly long strands by polymerization before recombination can act on these. In other words - where do the M2 and HIV sequences come from in the first place?

Thank you for valuable remarks. We added this discussion and relevant references to manuscript (p. 3, lines 17-21; p. 5, lines 4-12, marked by green).

Some typos: In Fig 1 - Route not Root. p22 - routes not routs. In figure captions - dangling not dandling.

Corrected
Reviewer #2 comments.

In this manuscript the authors characterize various mechanisms, products, and kinetics of spontaneous RNA-RNA recombination among 96-nt RNAs. This is a very thorough work, well done, well written, and well referenced. It has direct relevance to plausible prebiotic pathways in the RNA world leading from short, abiotic RNA oligomers to larger catalytic RNAs. I can recommend publication provided the authors address a few salient points:

1. My main comment is that the authors should compare their loop recombination reactions (root B in Figure 1) to the recombination reactions described by the Di Mauro group (e.g., Pino et al., 2013 Entropy). This is not my work, but it strikes me that they may be discussing the same type of reaction. The Di Mauro group has presented two types of reactions, one a recombination (which I think is very relevant here) and a second one that is a ligation (which I personally remain unconvinced of, but in any event may not be relevant here). In any event, a contrast between the reactions observed in the current manuscript and those discussed by the Di Mauro group is a requisite. In particular, the Product Pr-1 seems like a direct analog to the Italian group's model.

Thank you for valuable comment. We compared our system and recombination products with those described by the Di Mauro group. This discussion and relevant references were added to manuscript. See P. 17 and 18, starting line 24, marked by green.

2. The authors should discuss briefly how prebiotically relevant are 96-mers. It seems like they based this strictly off of some previous work, including their own work. The previous papers from this group (Lutay & Vlassov) did have short oligomers that were prebiotically relevant in that they were all under 20 nucleotides in length to start, the length that is close to the maximum for abiotic polymerization that has been achieved in vitro. However, by using 96mers they have created the question of whether the same kinds of reactions are achievable with smaller strands, and the question of whether the length impacts the reaction, which seems to be a possibility. On the other hand, the 96mers have the advantage of being known viral RNAs that can be worked into cells; they have also worked with them previously, according to a citation. So there is that practical aspect, even if the theoretical questions linger. The size issue played a role when they discuss their molecular cloning - they say that the variety of products is very large. So in the future it might be simpler to use smaller RNAs. One might imagine that there were more than just prebiotic reasons to do this, but the point of the paper seems to talk about prebiotic chemistry.

We added discussion on prebiotic relevance of 96-mers RNAs to the manuscript. P. 5, lines 4-12, marked by green.

3. It was interesting that they suggested the 3' OH of the ligating fragment was 60 times more active in GA tandem mismatch bearing loops in terms of forming the phosphodiester bond than compared to the 2' hydroxyl
which is 6 times more reactive toward activated phosphate esters in an unconstrained aqueous environment. I have wondered on occasion if the 2' OH of the Italian group's model could form the stem loop as readily as the 3' OH. Please discuss.

We added discussion on comparative reactivity of 2’-OH and 3’-OH groups to the manuscript. See P. 18, lines 7-11, marked by green.

We thank you for all comments.

Yours faithfully,

Zenkova Marina
Institute of Chemical Biology and Fundamental Medicine
Graphical Abstract (for review)

Cleavage

Complex formation

Ligation

Recombination

\[ 5'\text{-OH} - \text{-}2',3'\text{p} = 5'\text{-OH} \]

Base
Highlights

- Cell-free RNA recombination results in formation of wide range of products;
- Ligation could occur with participation of either 2’,3’-cyclophosphate or 5’-ppp as activated form of substrate;
- Molecular dynamics simulations suggests a preferable formation of 2’-5’- rather than 3’-5’-linkages.
Staroseletz Yaroslav

**Non-enzymatic Recombination of RNA: Ligation in Loops**

Staroseletz Yaroslav¹*, Sergey Nechaev¹*, Elena Bichenkova², Richard A Bryce², Catherine Watson², Valentin Vlassov¹ and Marina Zenkova¹

¹Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of Russian Academy of Sciences, 8 Lavrentiev Avenue, Novosibirsk, Russia. 630090.

²Division of Pharmacy and Optometry, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Oxford Road, Manchester, M13 9PT, UK

*These persons contributed equally to this work

Reprint requests to: Marina Zenkova, Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of Russian Academy of Sciences, 8 Lavrentiev Avenue, Novosibirsk 630090, Russia, e-mail: marzen@nibochnsc.ru.

**Keywords:** non-enzymatic RNA ligation, RNA world, origin of life, molecular dynamics.
Staroseletz Yaroslav

Abstract

Background

While the RNA world hypothesis is widely accepted, it is still far from complete: the existence of self-replicating ribozyme, consisting of potentially hundreds of nucleotides, is a core assumption for the majority of RNA world models. The appearance of such long RNA molecules under prebiotic conditions is not self-evident. Recombination seems to be a plausible way of creating RNA diversity, resulting in the appearance of functional RNAs, capable of self-replicating.

Methods

We report here on the study of recombination process modeled with two 96 nts RNA fragments. Detection of recombination products was performed with RT-PCR followed by TA-cloning and Sanger sequencing.

Results

A wide range of recombinant products was detected. We found that (i) the most efficient ligation was observed for RNA species forming bulges or internal loops, with ligation partners located within the loop; (ii) a strong preference was observed for formation of a few types of major products with a large variety of minor products; (iii) ligation could occur with participation of either 2′,3′-cyclophosphate or 5′-ppp; (iv) the presence of key reaction components, i.e. 5′ppp-RNAs, enabled the formation of additional types of product; (v) molecular dynamics simulations of one of the most abundant products suggests that the ligation results in a preferable formation of 2′-5′- rather than 3′-5′-linkages.

Conclusions

The study demonstrates regularities of new RNA molecules formation with non-enzymatic recombination process.

General significance

Our findings provide new data supporting the RNA World hypothesis and show the way of new RNA sequences emergence under prebiotic conditions.

1. Introduction

The RNA world hypothesis states that systems based on an RNA genome and RNA catalysts preceded current forms of life [1,2]. Several model systems of the RNA world have been described in detail [3–6]. The central notion of these models is an autoreplicase enzyme, an RNA molecule that can catalyze its own construction through template-directed polymerization of nucleotides. The plausibility of this theory was confirmed by several experimental findings. The strongest argument for the theory’s accuracy is the creation of artificial ribozymes capable
of synthesizing a spectrum of RNA sequences, like aptamers, ribozymes [7] and tRNA [8]. Moreover, amplification of short RNA templates in an RNA-catalyzed form of the PCR was shown for one type of ribozyme [8]. Other classes of in vitro selected ribozymes display such prebiotically important activities as nucleoside synthesis, triphosphorylation and polymerization of nucleoside triphosphates [9,10]. Dozens of recently created ribozymes with a wide variety of catalytic activities provide promise that a self-replicating ribozyme will be discovered.

Nevertheless, these RNA World models display various shortcomings [5,11]; in particular, the means of origin of such systems is unclear. One could envisage the following route to development of the RNA world: nucleotide synthesis from inorganic substances, formation of an RNA pool and then appearance of functional RNAs in this pool. Two possible ways could lead to the formation of the RNA pool: polymerization of nucleotides or recombination of RNA fragments. Polymerization of activated nucleotides produces RNA fragments not longer than 50 nucleotides [12,13], or the same order of magnitude [14] whereas the size of a presumable RNA replicase ribozyme is some hundreds of nucleotides in length [4,7–10]. The gap between short abiotic RNA fragments and long catalytically proficient molecules could readily be met by spontaneous RNA recombination occurring via SN2-mechanism through formation of a 2',3'-cyclophosphate [15,16]. Distribution of lengths of RNA fragments formed by polymerization remains exponential with or without recombination but only if recombination occurs at equal rates at all points [17]. If the recombination occurs not randomly but at certain number of “hot points”, it can shift distribution of lengths, making possible formation of fragments long enough to show a capacity for catalysis.

Since the first demonstrations of the RNA recombination phenomenon using poliovirus model in the 1960s [18,19], a great number of studies have been devoted to RNA recombination, reviewed elsewhere [16,20,21]. The most widely accepted model of RNA recombination is a ‘copy choice’ recombination. In this process, the RNA polymerase that mediates viral replication switches from one template to another during synthesis, while remaining bound to the nascent nucleic acid chain [21]. Non-replicative RNA recombination has also been demonstrated experimentally [22–24]. It can be explained by transesterification reactions, not necessarily involving an enzymatic activity. Some of the crossover loci are clustered, thus indirectly indicating the involvement of structures similar to known ribozymes in the recombination reaction [22]. The co-transfection of replicating and nonreplicating rubella virus RNA transcripts containing nonoverlapping deletions did restore the infectious virus [23].

A series of studies published by Lehman’s group revealed fascinating phenomena of RNA recombination. In particular, it was demonstrated that four special RNA fragments of 40-
60 nucleotides can self-assemble with the formation of the Azoarcus intron I [25]. The latter, in turn, catalyzes RNA oligonucleotide recombination. There are some constraints concerning the sequence of the substrates; nevertheless, hammerhead ribozyme and ligase ribozyme were obtained from nonfunctional RNA fragments with Azoarcus group I intron as a catalyst [26,27]. These results show that entire catalytic RNA motifs could be built up from smaller RNAs through a series of energy neutral reactions [26]. These data are in accord with the model suggested by Briones [28]: in this model, a fraction of simple RNA fragments, notably hairpin-like ones, are obtained by random polymerization on mineral surfaces; these fragments could display RNA ligase activity and catalyze the assembly of larger, eventually functional RNA molecules [28].

All these studies indicate that RNA recombination was likely to play an important role in the early stages of evolution. But the open question is, could RNA recombination occur without the involvement of an external complex catalyst (either enzyme or ribozyme) and lead to the increase of RNA length and complexity. The major step in this direction was made in the works of Chetverin. Nonhomologous recombination between fragments of a viral RNA in a cell-free system was demonstrated using the molecular colonies technique [29,30].

In our earlier works with chimeric DNA/RNA oligonucleotides, the chemical mechanism for a new phosphodiester bond formation was investigated. It was shown that the reaction intermediates possessed a 2',3'-cyclophosphate and a 5'-OH group, and that the ligation proceeded through the mechanism of intermolecular transesterification [31–33]. In a previous paper, we reported studies of the consecutive cleavage/ligation reaction of two 96nt long dephosphorylated RNA fragments, catalyzed by Mg$^{2+}$. Similarly to [31–33] the reaction was performed in the presence of oligodeoxyribonucleotide template (ON), which guided the recombination within selected RNA sites (5’-OH RNAs, ON+). The products of recombination were amplified, isolated and their nucleotide sequences were determined. Based on the Sanger sequencing data, some preliminary conclusions on the characteristic properties and mechanisms of recombination events were made [34].

In this study, we examined reaction routes and range of recombinant products in other settings and compared range of novel recombinant products with those obtained under previously investigated conditions (5’-OH-RNAs, ON+). Firstly, we studied the system behavior in the absence of the template (5’-OH-RNAs, ON-) and found that free spontaneous (template-independent) recombination in our system also existed and was efficient. Then, we used 5’-ppp-RNAs instead of 5’-OH-RNAs both in the absence (5’-ppp-RNAs, ON-) and in the presence of the template (5’-ppp-RNAs, ON+). This resulted not only in appearance of new products of
recombination, but surprisingly influenced the ratio between the products. Finally, we examined
the most abundant recombination product in order to evaluate the probability of formation either
2'-5' or 3'-5'-linkages using molecular dynamics simulations.

Rationale behind the choice of RNA fragment is as follows. Appearance of 96-nts RNA
fragments under prebiotic conditions is not proven yet however it is plausible similar to
recombination-mediated formation of 198-nts Azoarcus intron I ribozyme from four 40-60-nts
fragments [25]. As mentioned above random recombination is not very useful in formation of
long RNA structures so the concern is to find short RNA structures (up to 50 nts according to
[12]) prone to recombination preferably occurring at certain sites resulting in formation of long
RNA sequences. Prevalence of such structures is low however it gets higher the longer RNA
fragment formed [35]. Thus it seems reasonable to “shift” the limit of the search space from 50
nts to, for instance, 100 nts.

2. Results

2.1. Model system and reaction conditions.

Non-enzymatic reaction of RNA recombination was investigated with the use of two 96nt
RNA fragments of viral RNAs as substrates (96 nts fragments of influenza virus M2 RNA and
HIV-1 RNA, here and after M2-RNA and HIV-RNA, respectively) [34]. We hypothesize here
that non-enzymatic recombination of two RNA fragments occurs through preliminary cleavage
of substrates into shorter molecules, which are expected to be more eligible for ligation than
intact substrates. In our reaction system, cleavage and ligation pathways are not spatially
separated and occur, without any specific control, in one reaction system under constant thermal
and chemical conditions. However, we will separate all the reactions into elementary steps in our
further discussion to clarify the mechanisms of non-enzymatic processes.

In the first step, RNA molecules undergo cleavage in the presence of magnesium ions,
known to catalyze these reactions through several mechanisms. Any RNA molecules contain
"fragile" sites, whose location is predetermined by the RNA structure. The RNA substrates used
in this study were investigated earlier with a view to locate "fragile" sites, which are particularly
unstable under cleavage conditions. Location of these sites was recently discussed [34]. As a
result of the cleavage step, a set of molecules comprising 2',3'-cyclophosphate at the 3'-end is
formed. All of these molecules can be regarded as potential substrates of the ligation reaction,
occurring through the mechanism of intermolecular transesterification and yielding a new
phosphodiester bond between the reacting RNA molecules. The described mechanism of the
transesterification reaction is generally acknowledged as a primary one, as far as 2',3'-
cyclophosphate is essentially a self-activated form of a substrate, and should be considered as the
form of activated substrate in the absence of other chemical agents, such as BrCN, EDAC, etc [31,33,36].

Bearing in mind that these processes are random to a certain extent, ligation proceeding through the pathway described above should lead to the production of a large number of new molecules. Previously we limited the freedom of RNA substrates to react with random RNA "partner" with the use of oligodeoxyribonucleotide (ON-template) as a template for bringing together two certain RNA fragments, formed at the cleavage step, and the use of dephosphorylated RNAs (5'-OH-RNAs, ON-+) [34]. In this study we used these conditions as a control of reproducibility of cleavage/ligation events and identified recombinant products emerged in other conditions (5'-OH-RNAs, ON-, 5'-ppp-RNAs, ON+, 5'-ppp-RNAs, ON-).

At the preliminary step formation of RNA/DNA duplexes was examined via incubation of M2- and HIV-RNAs in the presence of 5'-32P-ON-template at 37˚C under two conditions: for 2 h in the absence of magnesium and for 72 h in its presence (Supplemental Fig. S1).

ON-template/M2-RNA duplex is readily formed upon incubation for 2 h (product A), while HIV-RNA is not prone to form this heteroduplex. Nevertheless, prolonged incubation (72 h) leads to formation of several duplexes of template with HIV-RNA fragments of different lengths (the major product is marked as "C"). This indicates the more facile formation of duplexes between template and partially cleaved RNA, experiencing less spatial complications for approaching the template. In the lanes corresponding to M2-RNA incubated 72 h with the template, one extra major product (B) is formed in addition to that detected upon 2 h incubation. In addition, a number of minor products, corresponding to a wide range of RNA lengths, are observed between bands “A” and “B” (Supplemental Fig. S1).

2.2 Reaction scheme

The suggested scheme of processes leading to the formation of recombinant products is depicted in Fig. 1. As it was shown previously, upon RNA incubation in the presence of Mg2+, a pool of activated substrates bearing 2',3'-cyclophosphosphate is formed (Fig. 1 1(a), 2(a), 2(b)). These substrates can be either free or complexed with the ON-template. Their self-assembling (root B) or association with the ON-template (root A) can lead to the formation of different complexes, i.e. those which are ready for ligation (Fig. 1 3(a)), those which are undergoing further cleavage in single-stranded regions formed within the complex with a template (Fig. 1 3(b)), and those which are self-assembling independently from the ON-template (Fig. 1 3(c), 3(d)). When the appropriate spatial complex for ligation is attained, the 5'-OH group of a 5'-substrate “attacks” the 2',3'-cyclophosphosphate of a 3'-substrate, with formation of a new phosphodiester bond (Fig. 1 4(b), 5(a,b)).
We note here that RNA complexes and recombinant products shown in Fig. 1 are limited to those where HIV-RNA is a 3'-substrate, and M2-RNA is a 5'-substrate. Actually, different RNA molecules can emerge in the reaction system, but an approach applied to detect recombinant products allows us to amplify only those formed by the pairs of substrates discussed above. The method of detection was described in detail in our previous work [34], and here only its basic principles are discussed. In the first stage, a pool of RNA molecules obtained after recombination is used as a mixture of templates for a reverse transcription reaction. Reverse transcription is performed with a use of sequence-specific M\textsubscript{rev} primer. In the second stage, we ran PCR from the obtained pool of cDNAs using primers H\textsubscript{for} and M\textsubscript{rev} (H – for HIV-RNA, and M – for M2-RNA specific primers). From these two stages, we obtain a mixture of PCR products corresponding to recombinant RNAs that saved the regions corresponding to H\textsubscript{for} and M\textsubscript{rev} primers in their structures and were formed by HIV-RNA fragment as a 3'-substrate and M2-RNA as a 5'-substrate.

2.3 Detection of recombinant products: RT-PCR

In PCR, we used 5'-\textsuperscript{32}P-labelled M\textsubscript{rev} primer to ensure detection of all the amplicons via radioautographing of gels. Fig. 2 shows an autoradiogram of gel after separation of PCR products obtained from the recombinant RNA products incubated under different recombination conditions. Those conditions involved (i) use of 5'-triphosphorylated and 5'-OH-RNAs; (ii) 30 and 100 mkl total reaction volume; (iii) in the presence (A1, A2, C, E) and in the absence (B, D, F) of ON-template; and (iv) PCR matching to non-incubated recombination reaction mixtures (E, F). We used a partially degraded at purine sites amplicon of M2-RNA, synthesized with M\textsubscript{for} and 5'-\textsuperscript{32}P-M\textsubscript{rev} primers, as a ladder.

cDNA was used in dilutions 1:1, 1:10\textsuperscript{3}, 1:10\textsuperscript{6}, and PCR included 28 cycles of amplification. For a positive control, we run 25 cycles of PCR with primers M\textsubscript{for}/M\textsubscript{rev} and cDNA mixture in a dilution 1:10\textsuperscript{3}. Negative controls involved reactions under three conditions: (1) PCR with primers H\textsubscript{for}/M\textsubscript{rev}, with reverse transcription performed in the absence of any RNA; (2) PCR in the absence of cDNA, with primers M\textsubscript{for}/M\textsubscript{rev}; and (3) PCR in the absence of cDNA, with primers M\textsubscript{for}/M\textsubscript{rev}. We separated radioactive amplicons in 10% denaturing PAAG. Titles of the experimental series used in the capture of Fig. 2 will be used further for marking relevant experimental models.

As shown on Fig. 2, ligation products can be divided into two major types: (1) those formed under all reactions conditions (X-9 group of products) and (2) those formed only in the presence of one of the key components of the reaction system. These components include ON-template (X-1, X-2, and X-5) or triphosphate group at the 5'-terminus of M2-RNA (X-1 through
X-4). It is noteworthy that products of the first type are formed in different proportions depending on specific reaction conditions, though this dependence is not strict. Lengths of products X-5, X-6, X-7 and X-8 are very close to the lengths of supposed products of template-dependent ligation. However, at least some of them seem to be formed in the absence of template as well, indicating the obvious limit of using electrophoretic methods for the explanation of sophisticated recombination events. Comparison of series A1 and A2 shows that the reaction is insensitive to the total volume increase, and 30 µl total volume used in most reactions is sufficient for the formation of recombinant products.

Comparative analysis of products in series A1 and B shows that X-6 disappears when the ON-template is excluded from the reaction system. A similar observation can be made for products X-1 and X-2; however correlation with the presence of template is less evident in the latter case.

Series C generates the only major product, X-6, and thus indicates the high selectivity of reaction toward formation of this specific product. Remarkably, the only difference between series A1 and C was the use of 5’-OH-RNA in series C. Therefore, in this group we exclude the participation of one of the reaction “components”, 5’-triphosphate group, and that leads to a sharp reduction in the diversity of the products, particularly so of the longer fraction. In summary, these data indicate that X-6 ligation occurs with participation of 2’,3’-cyclophosphate as an active group, and ligation is active only in the presence of ON-template (compare C, B and D series in Fig. 2).

The results of all negative controls are satisfactory; therefore formation of significant amounts of non-specific products in reverse transcription reaction and PCR is not observed.

2.4 Molecular cloning

The data presented in Fig. 2 indicate that the variety of products obtained under different reaction conditions is very large, and the electrophoretic method does not allow us to elucidate the mechanism and patterns of formation of recombinant products. In addition, it is not sensitive enough to detect the majority of products in the reaction mixture. Therefore we decided to sequence PCR products and to align the obtained sequences with those of initial RNA substrates. However, that required preliminary isolation of each PCR product from the mixture of amplicons. To achieve this goal, we used the TA cloning procedure, described previously [34] in detail and involving cloning of plasmids containing T-vector part (bearing sticky ddT-ends) and amplicon-part (bearing sticky mononucleotide 3’-A-ends). After transformation of competent E.coli cells with a pool of plasmids containing different types of TA-inserts and plating at Amp-containing LB agar, we obtained a set of colonies, each consisting of bacterial cells with a
specific type of plasmid. Colonies were screened for TA insert size via bacterial colony PCR with primers M13/pUC_for and M13/pUC_rev, directed to a permanent plasmid region outside of the insert area. A representative electrophoresis after separation of PCR products is shown in Supplemental Fig. S2. We divided all the products into groups on the basis of their lengths; from one to twelve colonies within each group were selected for plasmid sequencing. PCR products from 4 reaction groups were subjected to this procedure: A1, B, C and F (see Fig. 2 for keys). Group F was chosen as a negative control, group C, described in our previous work, as a positive one. In total, plasmids in 379 colonies were analyzed via bacterial colony PCR (81 colony in group A, 134 in B, 106 in C, and 58 in F), and the number of sequenced colonies was 35, 57, 27 and 19, respectively. After alignment with parent RNAs, 50 unique products were detected. The Sanger sequencing data are summarized in Table 1; the nucleotide sequence of each product in the area between primers H_for and M_rev and statistics of occurrence of each product in all the experimental series were determined.

2.5 Classification of recombinant products

The data presented in Table 1 demonstrate that the diversity of products arising from RNA recombination is very wide. Indeed, some products can be regarded as major ones (e.g. Pr-1, Pr-4, Pr-13, Pr-23); others were detected only in a few bacterial colonies (Pr-6, Pr-14, Pr-15, etc.); the rest are unique ones and were found only in a single colony (Pr-2, Pr-10, Pr-11). Each product can be also found in one or several experimental series (A, B, C and F). The first straightforward criterion for a product’s classification is the requirement of template for its formation. Group G1 (template-dependent, Pr-2 through Pr-9) unites products of ligation within bulges and internal loops, formed in a partially complementary complex of RNAs with ON-template. In this case, the template facilitated an approach of ligation of RNA fragments, but the ligation occurred in single-stranded regions of RNAs, bulged out of the heteroduplex in the elements of imperfect secondary structure of RNA. A large number of detected products were formed outside of the complementary complexes of RNAs with ON-template as a result of template-independent ligation (all other groups, Pr-10 through Pr-50). Herein and thereafter "template-independent" ligation means ligation occurring without the participation of ON-template, while this template may be present in the reaction mixture and/or some regions of RNA substrates may serve as templates for self-ligation. Pr-1 stands apart from all other products, since it can be formed as a result of both template-dependent and template-independent ligation.

In order to analyze the range of template-independent products, we classify them according to their distribution in experimental series (A, B, C and F). However, some products of
template-independent reaction were extremely similar to each other (with a difference in a few nucleotides in length and sequence) and were presented by too large number of colonies to be referred to as a random convergence of RNA fragments. Therefore, those products (Pr-10 through Pr-19) were attributed to a particular group G2. It contains products of template-independent ligation in RNA bulges and internal loops, formed within self-assembled structures of RNA containing double-stranded regions. Group G3 is a mixture of products with moderate frequency and encountered mainly in 3 or 4 experimental series, except Pr-23, that was found only in the F series. However, as it is one of the major products, we separate it from rare products of the G5 group. The G4 group contains products formed with participation of 5’-ppp and G5, combining residually all other rare products found only in one experimental series. All groups of recombinant products are described in detail below.

2.6 Product Pr-1

Pr-1 is a unique product, readily formed in series A and C in the presence of the ON-template, and its formation was described in detail in [34]. However, as summarized in Table 1, twice as much of relevant colonies are formed in the series "B", where no template is used. Moreover, Pr-1 is a major product in that experimental series. Therefore we assume that this product attains a very stable tertiary structure, which may facilitate ligation of the reacting RNA substrates. Two alternative secondary structures (Fig. 3 A,B) were found for this product with RNAfold [37]. In the first one, ligating 3' and 5' dangling ends of HIV-RNA and M2-RNA (1 nt and 3 nts, respectively) are located in the end of 10-pairs long RNA stem, turned into stem loop when a new phosphodiester bond is formed (Fig. 3 A). Another structure is an RNA stem-loop, formed by two RNA fragments. Within this complex, one 5'-nucleotide of a 5'-substrate is dangling, and ligation is accompanied by the formation of 1 nt bulge in the ligation site. This situation is similar to that in Pr-2, but in the current case a fragment of HIV-RNA serves both as a substrate and as a template for assembling of ligation intermediary complex.

2.7 G1 products

Group G1 includes 7 types of product: Pr-2 through Pr-9. All of them are formed exclusively in the presence of the ON-template, and ligation occurs within bulge loops of a size one (Pr-2, Pr-3), three (Pr-4, Pr-5) and six (Pr-6) nucleotides, as well as within 2 and 3 nts internal loops, formed by the RNAs and ON-template (Pr-7 and Pr-8, respectively). Most of these products were described and discussed by us earlier [34]. The main unexpected result from the analysis of this group is the strongest preference for formation of one type of product. Pr-4 was found in 101 colonies, 87 out of them belonging to the C series, 14 belonging to the A series, whereas all other products were found at best in 5 colonies. Thus ligation may occur in
different structures like 1- or 6-nts bulges either 2- or 3-nts internal loops (possibly, in some other structures, which have not been detected in our research) but the most favorable one is 3-nts bulge with 1- and 2-nts dangling ends, for 3’ and 5’ ligating fragments, respectively.

2.8 G2 products

Products of group G2 (Pr-10 through Pr-19, Fig. 3 C) have the largest degree of reciprocal similarity in comparison to products from other groups. Lengths of the relevant amplicons vary in the range of 104-110 nucleotides, and all the modelled secondary structures contain the same and very stable prolonged domain, shown in Fig. 3 C. That domain is formed mostly by a fragment of HIV-RNA, and a 7 bp duplex with M2-RNA fragment is formed at its 5’-end. Remarkably, the secondary structure of the initial HIV-RNA, used as a substrate, remains unchanged in this region, and hence this fragment serves as a feedstock for a template and as a 3’-side substrate for ligation at the same time. Therefore, it is reasonable that products formed in this sort of complex are present in all the experimental series, except a negative control one. The absence of any products from this group in the control series F, which refers to non-incubated recombination reaction mixtures, denotes a time requirement for the formation of those products. As highlighted above, cleavage of initial RNAs is a process occurring at moderate rates, and the lack of products in the control experiment argues that cleavage and, consequently, simplification and adjustment of secondary structure, is essential for the formation of stable intermediates.

Depending on the product type, dangling 3’ and 5’ RNA ends at the ligation site of G2 products have different lengths, and their sequences are provided in Table 1, along with a total number of colonies referring to each product in the experimental series. As shown in the table, the 3’-end can be 0, 1, 3, 4, 5, or 6 nucleotides long, while 5’-end is normally shorter and consists at most of two nucleotides. In the case of products Pr-11, Pr-13 and Pr-19, the 5’-end is fixed by the complementary interactions with an opposite strand, and in Pr-11, it is one nucleotide shorter than in Pr-13 and Pr-19.

The preference for the product formation here is striking, although not so strongly marked as in the G1 group. Between the major product (Pr-13, 39 colonies) and the unique products (Pr-10 and Pr-11, single colony each) several products were found with moderate frequency. The distribution of products between experimental series is also interesting. G2 products are most present in series B (46%, 61 colonies out of 134), twice more than in series A (23%, 19 colonies out of 81) and poorly represented in C series (4 out of 106). Therefore, the presence of template shifts the product formation process to the template-dependent products, decreasing the percentage of G2 products in series A. The low number of G2 products in the C series (dephosphorylated RNAs) is harder to explain. All explanations of this phenomenon serve
not more than speculation; however, it’s possible that formation of G2 products involves several stages, including formation of intermediate products like Pr-25 or Pr-26, which demand 5’-ppp-M2-RNA. Thus, by excluding 5’-ppp from reaction system, we can prevent formation of G2 products.

2.9 Molecular modelling of G2 products

Our primary goal here was to investigate the structural rules and molecular mechanisms of non-enzymatic RNA recombination using computational methods. To achieve this we used unrestrained Molecular Dynamics simulations focusing on behaviour around the ligation site for Pr-13 and three shortened models in explicit aqueous solvent (Fig. 4 for two dimentional representations and Fig. 5 for 3D structures). The choice of the model was due to the fact that Pr-13 is the most abundant product among template-independent ones, and the further development of our study intends to be concentrated mainly on template-independent recombination. All models preserve the same ligation site and differ in distal stem and loop regions, thus showing the influence of those regions on the geometry of the ligating groups. The root-mean-squared deviation (RMSD) in coordinates of backbone atoms, with respect to the initial modelled structure, was calculated over the time-course of the simulation. We also evaluate the backbone RMSD of the loop region of each model, relative to the NMR structure of the HIV-1 A-rich hairpin loop (Fig. 6).

The four model RNAs demonstrated relatively large fluctuations in backbone RMSD from its corresponding starting structure (Fig. 6): for example for the 53nt Pr-13 product, the average RMSD value is $6.0 \pm 1.7$ Å over the course of the 100 ns simulation (Table 2). The fluctuations observed here are suggestive of a high degree of flexibility within the models. The average backbone RMSD for the loop region, however, is considerably lower, ranging from 1.5 to 2.2 Å across the four models (Table 2). This indicates that the loop region is rather stable, agreeing well with the NMR model of [38], and that the mobility arises from the stem part of the structures.

To evaluate the likelihood of opportunities for recombination events within each model, the distance between the potential ligation partners was measured at each time point (Fig. 7). This was taken as the distance $d$ between the 2’,3’ P atom of 3’cytidine (HIV-1 RNA) and the O5’ atom of 5’uridine (M2-RNA). For Pr-13, the average interatomic distance over 100 ns simulation is $7.41 \pm 2.61$ Å (Table 2). While the mean distance between the terminal atoms of the proposed ligation partners is relatively high, there is a large amount of fluctuation in this distance over the course of 100 ns. As such, there are multiple occurrences during the trajectory where this 2’,3’ P⋯O5’ distance comes within the required 3 - 5 Å distance for ligation to be
possible. Sampling of distances below 4 Å is evident from the time series of $d$ for all four RNA models (Fig. 7).

Another consideration for the likelihood of non-enzymatic ligation is the angle at which the ligation partners approach one another. The requirement for non-enzymatic cleavage is that the attacking and leaving groups should occupy axial positions around the central phosphorus atom, equating to a O2’-P-O5’ angle of approximately 150 - 180°. As non-enzymatic ligation is thought to occur as a reverse of the cleavage reaction, proceeding through the same 5-membered transition state, it should follow that the 150 - 180° angle of incidence is also an important factor here. The attacking group, in this case, would be O5’ of the 5’-terminal U residue of M2-RNA and the leaving group may be either O2’ (resulting in 3’,5’-linked product) or O3’ (resulting in 2’,5’-linked product) of the 3’-terminal C residue of HIV-1 RNA. The angle of approach between the attacking and leaving groups was measured over the course of 100ns molecular modeling run. Assuming a distance of 3 - 5 Å and an angle of 150 - 180° to be necessary for spontaneous ligation to occur, pairwise comparison of the possible opportunities for ligation, forming either the natural 3’,5’-linked product (Fig. 8) or 2’,5’-linked product (Fig. 9), was performed for each model. Once again, it is apparent that all four models sample in this reactive region, both for formation of the natural 3’,5’-linked product (Fig. 8) and 2’,5’-linked product (Fig. 9). However, it appears from the 100 ns trajectories that in general the reactive region of the 2’,5’-linked product is more frequently visited relative to the 3’,5’-linked product (see Discussion below).

2.10 G3 products

G3 group is a mix of five different products. The most stable secondary structure of the Pr-21 is shown in Fig. 3 D. It is similar to that found for Pr-1 in Fig. 3 A, but in the case of Pr-21, the 3’ dangling end is 5 nucleotides long, and the 5’ end is paired with the HIV-RNA strand. Pr-24 probably has a similar structure, but its 3’ dangling end is 3 nucleotides long. Pr-22 may be just an PCR artifact, considering that it was found 8 times in the control series versus 7 times in three other series. Pr-23 was formed just by forward and reverse primers without any odd nucleotides in between, is presented by 26 colonies in series F and is not found in any other series. Probably it is formed under PCR template deficient conditions.

2.11 G4 products

These products appear as a result of ligation with the 5’- end of the initial M2 RNA, and require the presence of 5’-phosphate for recombination to occur. Although the statistics of occurrence of these products is very poor (5 colonies in case of Pr-26 and 1 colony for each of the remaining ones), it can be suggested that in this case another activated group, 5’-triphosphate,
Staroseletz Yaroslav

is involved in the reaction, instead of 2',3'-cyclophosphate; this provides an alternative pathway of recombination reaction in the current experiments.

2.12 G5 products

Group G5 combines all other products characterized by rarity of occurrence on the one hand and occurrence in only one experimental series on the other. Therefore it is difficult to make firm conclusions on the analysis of this group of products. For example Pr-39, Pr-43 and Pr-50 were found only in A series (in the presence of ON), but it is difficult to rationalise how the template directs formation of these products. Pr-44 - Pr-49 were detected only in the F series, but it is reasonable to suppose that they can be formed in the other series.

RNAFold analysis of G5 products showed great diversity of their secondary structures. The structures of some products look quite reasonable; for example, ligation occurs in 1- (Pr-40) and 3-nts (Pr-50) bulges or internal loops (Pr-35, Pr-37). The common secondary structures and the structures of the ligation sites of these molecules are similar to ones of abundant products from other groups and thus appearance of such products is not surprising. The appearance of many other G5 products especially long ones like Pr-38 and Pr-42 is more difficult to explain in the light of their secondary structures. Ligation occurs with 1- (Pr-38) or 3-nts (Pr-42) 3'-dangling end between two helical stems and one can imagine a large degree of freedom of ligation partners (Fig. 3 E). On the one hand, this structure does not preclude ligation; however, the ligation partners are out of effective distance for interaction almost all the time. However distant tertiary interactions could influence this situation considerably.

2.13 RNA cleavage patterns

To confirm the hypothesis of preferable synthesis of certain products due to the primary cleavage of RNA substrates in a limited number of sites, we studied the kinetics of both HIV and M2 RNA cleavage in the presence of magnesium ions under conditions reproducing the non-enzymatic recombination reaction, with the only exclusion that the two RNAs were treated separately. Both RNAs were labeled with $5'$-$^{32}$P-pCp and incubated in the presence and in the absence of the ON-template for 0, 2, 5, 24, 48 and 72 h. Mixtures of cleavage products were separated in 15% denaturing PAAG. Autoradiogram of the gel is shown in Supplemental Fig. S3. The figure shows that both RNAs have their own prominent "fragile" sites, and the cleavage depth at those sites rises as the incubation progresses. Such sites and their location in RNAs (counting both from 3' and 5'-ends) are shown on the sides of Supplemental Fig. S3. It is also remarkable that cleavage at several sites occurs mostly or only in the presence of ON-template: after nucleotides 33-43 and 45-63 in HIV-RNA and after 41-47 in M2-RNA. All fragile sites
Staroseletz Yaroslav

were juxtaposed with secondary structures of RNAs, and the locations of those areas are shown in Supplemental Fig. S4.

As shown in Supplemental Fig. S4, the upper stem of HIV-RNA, formed by nucleotides 24-60, is subjected to cleavage only when is partially unfold by the ON-template. Given that the template is absent from the reaction mixture (series B in Fig. 2), the stem remains almost intact and can lead to formation of products referring to group G2. Minor stems (formed by nucleotides 62-72 and 73-88) are cleaved under any conditions, and it supports the data provided in [39]. M2-RNA contains fewer numbers of template-sensitive "fragile" sites than HIV-RNA, they are located in the area 64-51 nts and most probably undergo cleavage due to unfolding of 54-71 nts stem by ON-template. Other sites of cleavage are located mostly in single-stranded regions of RNA, susceptible to rupture and generally coinciding with data demonstrated earlier [40] for the same RNA fragment, cleaved by artificial nucleases.

3 Discussion

3.1 RNA ligation in loops.

Recombination is a prebiotically plausible route to the emergence of a repertoire of functional RNA species. Whether it was the principal route is debatable but it is likely that it contributed to the development of the RNA world. In this work, we aimed to explore the free recombination process by modelling it with a simple system which included two RNA fragments and Mg$^{2+}$ as a catalyst. Here we discuss the observed features of this process.

There are several factors influencing the range of recombinant products formed. The most important factor is the selection of “building blocks” or RNA modules, capable of self-assembly. Predisposition to self-assembly is an inherent feature of RNA, enabling it to form multi-motif functional complexes (ribozymes). Self-assembly of a Group I intron from four inactive oligonucleotide fragments through RNA-directed recombination events was demonstrated in [27,41].

Selection of modules (“building blocks”) for the creation of recombinant products listed in Table 1 occurs both at cleavage and ligation steps of recombination reaction. As we have shown previously [34] and demonstrate in the current work (Supplemental Fig. S3 and S4), location of “fragile sites” is the major criterion for selection of building blocks at the cleavage step. However, an ease of dissociation of RNA fragments from each other and competition between fragments for the binding site determines availability of this blocks for the ligation reaction. For example, M2 fragment in Fig. 3 C competes with “leaving” HIV fragment and “wins” due to formation of longer duplex. When available, blocks are selected again, and those which form the most stable intermediate complex due to formation of a duplex with a template...
Staroseletz Yaroslav

or some kind of intramolecular duplex will benefit. The last selection stage is the ability to ligate
fragments by placing the reacting groups (2',3'-cyclophosphate and 5'-OH) to achieve the
optimal orientation, so-called in-line conformation [42]. Eventually, the success of RNA
fragments in proceeding through all of these stages determines the relative yield of the various
recombination products.

Disposition of reacting groups in the optimal orientation depends on an RNA structure.
As we have shown in our previous work [34] and demonstrate in the current study, RNA ligation
occurs mostly in the internal loops, bulge-loops and the end of stem structures, turned into
hairpin upon ligation. Different kinds of loops are inherent structural motifs of almost all known
long RNAs, and their presence gives a definite advantage to RNA over DNA in terms of ability
to possess diverse catalytic functions. Loops, stem-loops and pseudo knots are critical elements
for every known complex ribozyme [43–45], and certainly would be featured in a self-replicating
ribozyme. Therefore, it is possible that under prebiotic conditions, RNA ligation resulted in the
appearance of stem-loop structures capable of further ligation and rearrangement with formation
of catalytically active RNA molecules.

Non-enzymatic ligation in RNA loops was discussed in the works of David Bartel [46].
Seven families of RNA ligases were isolated from a pool of random RNA sequences. The
authors have reported the generation of 3 classes of ribozymes, in accordance with the ligation
mechanism involved in their formation. Formation of class II ligases was the closest to the one
we have demonstrated in the current study. A surprising feature of the class II catalytic motif is
that the continuous base pairing across the ligation junction is replaced by an internal loop
involving the ligator and these loops were formed by tandem GA mismatches, which were
shown earlier to be unexpectedly strong stabilizers of loop regions within Watson-Crick helices
[47]. Another interesting feature is that the 3' OH of the ligating fragment was 60 times more
active in GA tandem mismatch bearing loops in terms of formation of a new phosphodiester
bond than 2'-OH, although the 2'-hydroxyl is well known to be at least 6 times more reactive
than 3'-OH toward activated phosphate esters in an unconstrained aqueous environment [48,49].
However, exclusion of a hairpin structure located in the ligator 2-8 nucleotides away from the
internal loop in question resulted in a rapid increase of rates of 2',5' linkage formation over the
initial levels [46]. Hence the prevalence of 2',5' or 3',5' bond formation upon non-enzymatic
ligation in loops does not depend entirely on the type, size and nucleotide composition of a loop.
Reciprocal orientation of ligating groups is highly influenced by the interactions with other
motifs of the same or other RNAs, and therefore general universal predictions on type of the
bond seem to be senseless without a consideration of specific spatial setting of every single
structure. In the current study, we did not aim to investigate the type of newly formed phosphodiester bond experimentally, because our product detection scheme was based on the analysis of sequences of RT-PCR products, while the information about the linkage type is lost in transition from RNA to DNA. However, this information was evaluated via molecular dynamics simulation, albeit for only one product. Based on the structure of product Pr-13 (Fig. 3 C), which is 53nt in length, we also created 3 shortened models, 38-, 33-, 27-nts long (Fig. 4 and 6). Using a simultaneous co-incidence of a O2’-P-O5’ bond angle of 150 - 180˚ and a distance between reacting atoms of 3 – 5 Å as requirements for spontaneous ligation to occur, we found that 2’,5’- linked products are generally predicted as favored, based on the sampling of these 100 ns simulations (Fig. 8, 9). The 33nt model showed the strongest preference for 2’,5’-linked products over 3’,5’ ones whereas 53nt and 27nt models demonstrated only a slight preference for 2’,5’-linked products. The 38nt model in contrast is favorable for 3’,5’- linked products to be formed (Fig. 8, 9). Taking into account the identity in sequences at the ligation site of the four models, it appears that distal regions of the RNA molecule influence the structure of the ligation site.

Another feature of spatial architecture in ligation sites, discussed in [46], was less relevant to the current work, but nevertheless is notable. Class I ligases required fixing one ligation substrate in a fully complementary complex with a distal part of a ribozyme, while the second substrate partner had four nucleotide long dangling end, flexible enough to access the first substrate and to accomplish ligation reaction [46]. These data correspond well with the mechanism of formation of products of G2 group (see Fig. 3 C), where the major product, Pr-13, is formed in a similar complex, with the only difference that the 5’ rather than 3’ end is fixed in a Watson-Crick helix, and the dangling end is 3 nts long.

Ligation in terminal loops was demonstrated in direct experiments in [50] using a model system consisting of two synthetic 24mers ribooligonucleotides G24 and C24 and variants thereof. The authors showed that incubation of 5’P-G24-3’OH and 5’OH-C24-3’OH in water results in formation of two products: C24G24 and C24G. 3’OH group of C24 performs nucleophilic attack at either terminal 5’-phosphate of G24 (ligation) or internucleotide phosphate 5’GpG23-3’ (recombination). The second reaction is not similar to recombination process described here because it involves only one transesterification act without intermediate cyclic phosphate formation whereas in our model recombination involves two independent transesterification acts namely cleavage leading to the formation of cyclic phosphate and ligation when cyclic phosphate acts as activated phosphate form. The major difference between the models is the use of 5’-phosphate (in contrast to 2’,3’-cyclophosphate
or 5’-ppp). The point is that 5’-phosphate is not activated form of phosphate so the ligation reaction in model system described in [50] is not energy neutral.

Authors assume that ligation occurs in 5-nts loop that is consistent with our results. However it is rather a suggestion based on computations then an established fact and we are entitled to suppose that ligation in this model system may also occur for instance in 4-nts or shorter loops.

The question whether 3’-OH group or 2’-OH acts as nucleophile is also beyond the scope of Italian’ group paper [49] though the authors assume 3’-OH group as nucleophile by default. As it was shown earlier [46,51] and here this question couldn’t be solved in general terms. Each structure provides benefits for the choice of either 3’-OH or 2’-OH or both in the certain ratio.

Non-enzymatic recombination attracted the attention of progenitors of the RNA world hypothesis after the finding that this kind of RNA rearrangement is a critical step in life cycles of a number of viroids [52,53]. Even bacterial tRNAs that contain self-splicing group I introns have their sequences interrupted at the same anticodonic location, giving strong evidence to the notion that tRNAs originated in several pieces and needed to be assembled through recombination events [15].

Another stimulus that provoked the study in this direction was the gradual understanding that ribozyme with replication activity will unlikely be found in nature and that, in turn, led to the conclusion that long catalytically active RNA molecules (ribozymes) had to evolve from short RNA fragments through the reactions of cleavage and intermolecular ligation (named collectively “recombination”) [28].

There is one especially attractive feature of recombination: it can be performed in an energy-neutral fashion, when the number and type of chemical bonds are unchanged in the reaction. Hence only a slight perturbation of the environment is enough to disturb the equilibrium and to drive the reaction substantially in either direction [15]. Certainly there must be local and global environmental conditions, i.e. divalent metal ions, temperature, pH or surfaces, which could force this reaction back or forward.

In spite of the high current interest in ribozymes, not many studies were done for the investigation of features on non-enzymatic ligation. Due to low yields of these reactions, model systems described in literature are limited to the use of polyN substrates or templates [54], short and chimeric oligonucleotides [55], use of templates and preformed substrate/template duplexes. Intuitively, these limitations are used to guide the reaction to the synthesis of one major product. However, as we have shown in this work, the information is lost when these kinds of restrictions
are applied. Given the opportunity to evolve in any energetically allowed direction, reaction of RNA recombination yields a number of surprising products whose formation was not predicted and therefore was inhibited in the previous restricted experimental schemes. Specifically, we speculate that ligation of RNAs in loops is more preferable than within fully complementary complexes with a template.

3.2 Similarities between spontaneous recombination and ribozyme’s cleavage and ligation reactions.

We have discussed RNA ligase activity only in cis conformational arrangements. However we cannot exclude that some of the recombinant products are formed as a result of in trans action of some catalytically active RNA elements spontaneously formed in a pool of RNA upon incubation.

The possibility of evolving complex and highly active RNA ligases has been documented in vitro, using as the starting material pools of random RNA sequences [46], natural ribozymes such as group I introns [56], or previously evolved RNA ligases [57]. Those experiments have shown that the ligation and exchange of RNA structural domains through recombination can be used to engineer new functional RNAs [28,58,59].

The probability of hairpin ribozyme-like structures presence in a pool of random RNA fragments (40 nts length) was shown to be relatively high and its probable participation in recombination event should be taken into account. It was hypothesized that stable hairpin structures could have been endowed with a certain degree of primordial RNA ligase activity [28,35].

In the early 1990s it was shown that group I ribozymes in Tetrahymena and sunY broken into 3 parts can assemble in in trans complexes without a covalent contiguous RNA chain [60–63]. Such trans complexes could perform ligation and splicing reactions on exogenous RNA substrates and could thereby promote the assembly of fragments complementary to the ribozymes themselves [41,60,61].

Computer simulations performed by Manrubia et al. show that upon hairpin-catalyzed ligation of two preformed RNA modules, up to 1% of the assembled molecules retain the previous modular structure of their building blocks [64]. That is what we observe in case of G2 group of products (see Fig. 3 C) that possess unchanged structural fragment of HIV and therefore are abundant in the reaction mixture under any reaction conditions.

3.3 Cleavage and ligation rates.

We have discussed above that recombination reactions that we performed in our study consist of two major steps: cleavage and ligation. Although they are considered as reciprocal
reactions in chemical terms, and theoretically equilibrium between them should be achieved, this does not happen in the reaction system we use. The reason is that RNAs have multiple cleavage and ligation sites, and cleavage at one site may be followed by ligation at the other. Generally speaking, products of ligation reaction are different from the substrates of cleavage reaction. Hence equilibrium is never attained, and ongoing cleavage/ligation events yield in continuous enriching the pool of RNA with new molecules. RNA recombination is a dynamic process, and any complementary complex can undergo dissociation, leading to fragments’ association with new RNA molecules and so forth. Release of RNA from complex with template or with another RNA molecule has two effects. Firstly, secondary structure of RNA is being changed and thus location of “fragile” sites changes as well. Secondly, single-stranded RNA becomes very sensitive to cleaving conditions (i.e. presence of divalent metal ions and high pH). Therefore yields of recombination products depend not only on the ease of formation of ligation substrates and the probability of their convergence, but also on the stability of double-stranded complexes that preserve them from rapid degradation. Stability of these complexes, however, may also inhibit ligation reaction as a result of a need for diffusion of the ligated product off the template [41]. However, some RNAs ligating in cis possess an ability to adapt to the “structural” conditions of reaction and perform self-ligation in two alternative complexes. It was shown that the use of simple self-replicating systems of the form A + B → T, where A and B are the substrates that bind to a complementary template T, and become joined to form a molecule that is identical to the template [65]. Selected substrates A and B were designed in such a way that they could ligate with or without a template. Adjustment was made by varying the length of complementary stem formed between each of RNA substrates and RNA template. Evidently, conditions can be found when one or another mechanism will be preferable [65].

Similar observations were made for Pr-1 in our study. It was demonstrated to be formed at relatively high yields (in comparison to other products) both in the presence and in the absence of template. Surprisingly, yields were almost the same under different conditions (26% of all clones in series A and 28% in series B). Hence, fragments forming these products adjust to the reaction conditions. Although the exact structures of intermediates remain unknown, a suggestion can be made that even in the presence of template self-ligation (i.e. template independent ligation) is preferable (see Fig.4 A and B for suggested intermediate structures).

As we have highlighted earlier, direction of reaction is determined by some general reaction conditions, i.e. temperature, pH, metal ion (catalyst), etc. The raise of pH (from 7.0 to 9.5) leads to the increase of yield of products of cleavage/ligation reaction, but the degradation of
RNA at alkaline pH is also increased [31]. As for temperature formation of long RNA molecules was registered within 20-37°C [31].

Due to complicacy of the elaborated system, we did not expect obtaining of high yields of products in the current study. That determined the choice of indirect detection method which, in turn, made the yields impossible to be measured 37°C, pH 8.0 for 72 h. Reaction conditions chosen (37°C, pH 8.0 for 72 h) were suggested to ensure that both cleavage and ligation reaction can take place in the system, that dissociation of relatively short complexes is possible and that RNA substrates are not totally degraded during the reaction.

3.4 Similarities between spontaneous recombination and splicing.

As we have highlighted above, numerous “checkpoints” necessary for completion of ligation yield in the presence of major and minor products in the reaction mixture after 72 h of incubation. Obviously, major products are of everyone’s primary interest as long as they reflect the most important pathways that take place in the reaction. Nevertheless some minor products seem very curious and inspire us for continuing research.

In the RNA pool, there are several groups of products with (possibly) similar secondary structures that differ in several nucleotides at the ligation site (see Table 1). In addition to the hypothesis proposed earlier (formation of RNA fragments of slightly different lengths at the cleavage stage), a new suggestion about the possible inclusion of splicing events in the formation of shorter products from longer ones can be made now.

33 products of 50 detected in the pool of RNAs after recombination reaction were referred to the group G5 that unites all minor ligation products with undefined intermediate complexes. However, two products, Pr-47 and Pr-28, deserve a special attention. When compared to the sequences of products Pr-3 and Pr-25 respectively, one can see that the first two products differ from the latest two by the presence of one (Pr-47) or two (Pr-28) guanines at the ligation site, right between the HIV and M2 sections. Noteworthy, 5’ terminus of M2 is not cleaved prior to ligation in this case, i.e. M2 fragment ligates with its intact 5’-end and odd G nucleotides do not originate from 5’ end of M2. However, they are not present in the HIV sequence as well, thus proving that two ligation events occurred in these products, and exterior nucleotides were inserted as linkers from the outside. Guanines are known to play a key role in Group I Intron splicing, where they facilitate degradation of exon-intron phosphodiester bond and thus relieve 3’-OH group of exon to attack the bond between intron-Exon II [66]. Although that mechanism does not imply presence of guanine in a processed RNA, the “mediator” role that guanine plays in ligation can be relevant to our study. It should be noted that all four discussed products are minor and were detected only in one bacterial colony each. Hence, no
conclusions about preferences of formation and relative yields can be made with this poor
statistics. Nevertheless, there is one more interesting point to be discussed in relevance to these
products. Pr-3 and Pr-25 (without G) are formed in A series only (template-dependent
recombination). Pr-47 and Pr-28 (with G inserts) are detected only in F series which is
essentially a control group with no incubation under recombination conditions. However all
these products are formed only in reactions where 5'-ppp-M2 RNA was used. The first
suggestion is that they are formed due to a completely different mechanism where not 2',3’-
cyclophosphate, but 5’-triphosphate acts as an activated form of phosphate that ensures the
energy supply for transesterification reaction. This group was shown to be able to perform this
role in a number of works [67–69]. The second suggestion is that products with G inserts
actually appear during reverse transcription reaction performed at 42 C. If this is the case, 5’-
triphosphate might attack 3’-OH group of free dGTP (that is present at high concentrations in
reverse transcription reaction) and thus extend the length of M2 RNA in one
deoxyribonucleotide, with 5’-triphosphate at the end. This pathway can be undergone several
times, as it probably happens in case of Pr-28 (GG insert). Finally, 5’-triphosphate of terminal
guanine of extended M2 fragment attacks relevant site of HIV RNA (which is ironically a
“fragile” site as one can see from Supplemental Fig. S3) and completes recombination mission.
Again, this happens only in F series (non-incubated mixture), not in A series (incubated for 72
h). Possible explanation for that can be found in the fact that in recombination reaction presence
of free GTP is very unlikely, and the only option for M2 RNA is to attack HIV RNA via with its
nascent 5’-triphosphate. Those M2 RNA fragments that did not choose this route will most
probably lost their terminal 5’-triphosphate as a result of hydrolysis or cleavage of one or several
nucleotides at 5’ termini, resulting in the loss of ability to accept dGTP in reverse transcription
reaction.

Formation of non-specific RNA recombination products in reverse transcription reaction
is obviously limited to those described above, as far as 2’, 3’-cyclophosphates can accept no
more than one dNTP due to the lack of 2’-OH in dNTPs, and 5’-triphosphates, as shown above,
are present in the mixture in negligible quantities. However, the described pathways, if true, can
be suggested as a demonstration of first steps to the development of elementary splicing
machineries, which were supported by enzymes later and finally became ubiquitous in the
modern biological world.

4 Conclusion

We investigated the recombination process between two 96 nucleotides (nts) RNA
molecules under certain conditions (pH 8.0, Mg\(^{2+}\), 37°C, 3 days incubation) and found several
remarkable regularities, namely, preference for formation of a few types of major product with a large variety of minor products, switching reaction routes when adding such components as 5’ppp-RNAs and ON-template, and preference for formation of 2’-5’- rather than 3’-5’-linkages. We are sure that it is a very important but a tiny step on a really uncharted territory of non-enzymatic RNA recombination.

The understanding of the molecular mechanisms of non-enzymatic recombination of RNAs may not only explain some aspects of the origin of life, but also provide a vital clarification of enormous structural and functional diversity of RNAs. More importantly, recombination involving RNA sequence exchange and rearrangements between and within RNA molecules are believed to contribute to the evolution and diversity of RNA viruses and might be responsible for the production of defective interfering RNAs. The insight into the fundamental principles of molecular interactions and regulatory mechanisms of RNA recombination and the understanding at the molecular level of the key factors promoting these reactions may shed the light on the origin of rapid, uncontrollable mutations of RNA viruses, which may facilitate production of new viral strains.

5 Materials and methods
5.1 Enzymes and chemicals

[γ-32P]-ATP and 32pCp (specific activity >3000 Ci/mmol) were from Biosan Co. (Russia). Enzymes: calf intestinal alkaline phosphatase and endonuclease Fok I were purchased from “SibEnzyme”, Russia. T7 RNA polymerase, Taq DNA polymerase, M-MuLV DNA polymerase were produced in this Institute. T4 polynucleotide kinase and T4 RNA Ligase were from Fermentas. TA cloning was proceeded using “InstAclone™ PCR Cloning Kit” (Fermentas). Plasmid DNA was isolated with “GenElute™ Plasmid Miniprep Kit” (Sigma). BigDye v.1.1 sequencing mixture was from Applied Biosystems. All the chemicals were of analytical or ACS grade, and autoclaved Milli-Q™ water was used for all procedures. M2-RNA of influenza virus was synthesized by T7 transcription from plasmid pSVK3M2, this Institute collection. HIV-1 fragment was obtained via T7 transcription of plasmid pHIV-2, kindly provided by Prof. H.J. Gross (University of Wurzburg, Germany).

5.2 Oligonucleotides

Oligonucleotides were synthesized using standard automated solid-phase methods and purified by reverse-phase liquid chromatography in this Institute. Primers were as follows: M2-96for – ACAAGCTTTAATACGACTCATA GGGCGCTTCTACGGAAGGAGTACC (T7 promoter region is underlined), M2-96rev – CGAGACAAAATGACTGTCGTCGTCAGC, Mfor –
Staroseletz Yaroslav

CTACGGAAGGATCCTGA, $M_{\text{rev}}$ – ATGACTGTCGTCAGCATC, $H_{\text{for}}$ – GAGATCCCTCAGACCACT, M13/pUC_for – GTAAAACGACGGCCAGT, M13/pUC_rev – CAGGAAACAGCTATGAC. ON-template – CGATC CACAGCCTACCGTCTAGAGTAAAGC (overhangs are underlined).

5.3 5’- and 3’-labeling

Primer $M_{\text{rev}}$ and ON-template were 5’-end labeled with $^{32}$P using T4 polynucleotide kinase and [$\gamma$-$^{32}$P]-ATP according to the manufacturer’s manual as described previously [34]. Radioactive RNAs were prepared by labeling with [5’-$^{32}$P]Cytidine 3’;5’-bisphosphate ($^{32}$pCp) using T4 RNA ligase according to the manufacturer’s manual as described previously [70]. Labeled oligonucleotides and RNAs were purified in 8% denaturing polyacrylamide gel (dPAAG).

5.4 Plasmid linearization

Plasmid pHIV-2 was linearized with Fok I endonuclease according to the manufacturer’s manual. After extraction with TE-phenol-chloroform (1:1) and chloroform-isoamyl alcohol (24:1) DNA from the aqueous phase was precipitated with ethanol and dissolved in water.

5.5 Amplification of M2-DNA fragment

M2-DNA fragment was amplified from the plasmid pSVK3M2 using primers M2-96$_{\text{for}}$ and M2-96$_{\text{rev}}$ as described in [33]. Amplification product (121 bp) was purified by phenol extraction (TE-phenol : chloroform (1:1)) and precipitated from the aqueous phase with ethanol. Dissolved in water precipitates were used as templates for T7 transcription.

5.6 T7 transcription

T7 transcription for the synthesis of M2-RNA and HIV-RNA from M2-DNA (obtained by amplification of pSVK3M2 plasmid fragment) and linearized plasmid pHIV-2, respectively, was performed as described in [33]. Both RNA transcripts were purification in 8% native PAAG and were eluted from the gel with 3 mL of elution solution (0.6 M NH$_4$Ac (pH 5.0), 0.1 mM EDTA, 0.1% SDS) on a shaker at 4°C for 4 h. Eluted RNA was precipitated with ethanol and dissolved in water.

5.7 Dephosphorylation of RNA

A reaction mixture (total volume 50 µL) containing 8 µg of M2-RNA or HIV-RNA, 50 mM bis(tris(oxymethyl)methylamino)propane HCl buffer (pH 8.0), 2% formamide, 0.2% SDS, 1.25 mM DTT and 1.5 u of Calf intestinal alkaline phosphatase (CIP) was incubated at 37°C for 45 minutes. Then 1.5 u of CIP was added again to the reaction mixture and it was additionally incubated under the same conditions. After incubation, mixture was subjected to twofold extraction with phenol- chloroform (1:1), then chloroform-isoamyl alcohol (24:1). RNA was
precipitated from the aqueous phase by 95% ethanol, washed twice with 75% ethanol, dissolved in water and stored at -20°C.

5.8 Non-enzymatic cleavage/ligation reaction

<table>
<thead>
<tr>
<th>Reaction mixtures</th>
<th>(total volumes 30 µL) containing</th>
</tr>
</thead>
<tbody>
<tr>
<td>bis(tris(oxy)methylamino)propane HCl buffer (pH 8.0), 0.5 µM dephosphorylated or 5'-ppp-bearing M2-RNA and HIV-RNA, 1.0 µM of the ON-template (in template-directed reactions) and 5 mM MgCl₂, were incubated at 37°C for 3 days. After incubation, EDTA was added to the ligation mixtures in a concentration equal to that of MgCl₂. Volumes of all probes were adjusted to 100 µL with water, and RNA was precipitated with ethanol in the presence of 0.3 M NaAc and 10 µg of glycogen. Precipitates were washed with ethanol, dissolved in 15 µL of water and used in reverse transcription reaction.</td>
<td></td>
</tr>
</tbody>
</table>

5.9 Reverse transcription and PCR of ligation products

The DNA template for PCR amplification was produced from the ligation mixture of recombinant RNA in reverse transcription reaction. A reaction mixture with an initial volume of 12 µL, containing 1.25 µM of specific reverse primer and 1 µL of mixture of products of the cleavage/ligation reaction, was subjected to denaturation at 70°C for 5 min and then chilled on ice for 3 min. After that NTPs of a final concentration 1 mM and M-MuLV buffer were added, volume was adjusted to 18 µL with water, and mixtures were incubated at 37°C for 5 min. Finally, 20 u of M-MuLV polymerase were added and the mixture of a final volume 20 µL was incubated at 42°C for 60 min. The reaction was terminated by heating at 70°C for 10 min. Obtained cDNA was directly used in PCR. PCR was performed from reverse transcription (RT) products using primers H_for and M_rev. In a positive control (to confirm M2-RNA presence in the reaction mixture) primers M_for and M_rev were used. PCR mixtures contained 67 mM Tris-HCl (pH 8.0), 16 mM (NH₄)₂SO₄, 1.5 mM MgCl₂, 0.01% Tween-20, 0.25 mM dNTPs, 0.8 µM each PCR primer, 1 µL of RT-product in dilutions 1:1, 1:10³, 1:10⁶ and 2 u of Taq DNA polymerase in a total volume of 20 µL. PCR was started with 3 min of preincubation at 95°C and followed by 25 cycles of amplification (1 min, 94°C; 1 min, 56°C; 1 min, 72°C). Final cycle contained prolonged elongation step (72°C, 10 min) for complete formation of sticky A-ends. PCR products were used directly for TA cloning. In parallel, we run PCR with 5'-³²P-labeled M_rev primers for visualization purposes. In this case aliquots of PCR products were mixed with loading buffer and applied on 10% dPAAG electrophoresis. Visualization was accomplished by exposing dried gels on FX Imaging Screen, followed by screen scanning on Molecular Imager FX-PRO Plus (Bio-Rad) and images processing in Quantity One v.4.2.3 Software (Bio-Rad). Amplification products were identified by electrophoresis on 10% dPAAG using DNA ladder,
Staroseletz Yaroslav

obtained via chemical cleavage of M2-80 radioactive PCR product (synthesized using primers M_for and 5'-32P-M_rev, product length 80 bp) at adenine and guanine sites. Reaction mixture of a volume 30 µL, containing 10 µL of 5'-32P-labeled M2-80, and 20 µL of 3% diphenylamine solution in formic acid, was incubated at 37°C for 5 and 10 minutes. At the end of incubation, 100 µL of water was added and reaction mixtures were subjected to twofold extraction with 300 µL of diethyl ether. DNA from the aqueous phase was precipitated with 10 volumes of 2% LiClO₄ solution in acetone, pellets were washed with acetone, dissolved in water and then added to the loading solution.

5.10 TA cloning

The PCR products were cloned using the "InsTAclone™ PCR Cloning Kit" (Fermentas), according to the manual provided by the supplier similar to [33].

5.11 Bacterial colony PCR

Fragments of individual bacterial colonies were dissolved in 20 µL of PCR master mix, containing 67 mM Tris-HCl (pH 8.0), 16 mM (NH₄)₂SO₄, 1.5 mM MgCl₂, 0.01% Tween-20, 0.2 mM dNTPs, 0.3 µM of primers M13/pUC_for and M13/pUC_rev, and 2 u of Taq DNA polymerase. PCR was started with 3 min of preincubation at 95°C and followed by 27 cycles of amplification (30 s, 94°C; 30 s, 51°C; 1 min, 72°C). PCR products were analyzed in 8% dPAAG and visualized by staining of gels with ethidium bromide.

5.12 DNA sequencing

Sequencing of constructs possessing the DNA inserts was performed via Sanger sequencing reaction with mixture BigDye v.1.1 (Applied Biosystems) followed by further analysis of products at automatic gel analyzer ABI 3130xl (Applied Biosystems). Detailed protocol is published in [33]. Sequences were determined in Sequence Scanner Software v. 1.0 (Applied Biosystems) and aligned with initial RNA sequences using Vector NTI v.10.0.1 Software (Invitrogen).

5.13 Molecular Modelling

The loop region of each of the selected models was based on the 1.92 Å resolution NMR structure of the HIV-1 A-rich hairpin loop [38] (PDB code 1BVJ). Nucleotides G156 – C172 of the original experimental structure correspond to nucleotides G18 – C34 of our full-length 53-mer model. The extended stem region of the 53nt model was created by manipulation of a standard double helix, generated via the make-na web server [71]. Nucleotides were then excised as appropriate from the stem opposite the predicted single-stranded regions and the single nucleotide bulge (C38); this bulge was modelled as a mirror image of the bulged nucleotide (C158) found in the experimental structure [38]. The 38nt, 33nt and 27nt models were
constructed via removal of the appropriate number of nucleotides from the stem region of the original 53nt model. As much of the experimentally derived geometry as possible was retained at each stage. Models were constructed and manipulated using Yasara [72].

The 3’ terminal cytidine 2’,3’-cyclic phosphate residue of the HIV-1 recombination partner was modelled from based on the geometry of a standard cytidine residue, using Yasara [72]. Geometry optimisation was performed and electrostatic potentials calculated at the HF/6-31G* level, using Gaussian 09 [73]. Restrained electrostatic potential (RESP) charges were fitted using the antechamber module of AmberTools 13 [74] and the general AMBER force field (GAFF) [75] used for non-electrostatic parameters.

Simulations were carried out using the ff99 force field with bsc0 [76] and XOl3 [77] modifications for RNA. A box of TIP3P water molecules [78], extending 12 Å in all directions beyond the solute, was used. Potassium ions (radius of 1.705 Å and well depth of 0.19368 kcal/mol) and chloride ions (radius of 2.513 Å and well depth of 0.03559 kcal/mol) [79] were added, to a final K+ concentration of 200 mM.

To equilibrate each system, energy minimisation was followed by 20 ps MD in the NVT ensemble, to heat the system to 100 K, with subsequent 200 ps MD in NPT with 25 kcal/(mol.Å²) positional restraints on solute atoms, while the system was heated from 100 to 300 K. Short (100 ps) simulations with gradually decreasing restraints of 20, 10, 5, 2 and 1 kcal/(mol.Å²) applied to the solute were followed by 150 ps with positional restraints of 0.5 kcal/(mol.Å²) and 150 ps unrestrained MD. Production simulations of 100 ns for each system were performed using the pmemd module of AMBER 12 [74]. The particle mesh Ewald (PME) method [80] (was used to treat long-range electrostatic interactions. The SHAKE algorithm [81] was used to constrain covalent bonds involving hydrogen atoms, along with a 2 fs time step. Simulations were carried out in the NPT ensemble, with periodic boundary conditions and a temperature of 300 K maintained using a Langevin thermostat [82].

Acknowledgements

This research has been supported by the Russian Science Foundation (Grant No. 14-44-00068) and Presidium of RAS research program ‘Molecular and Cellular Biology’ 6.1.
References


Staroseletz Yaroslav


Staroseletz Yaroslav


Staroseletz Yaroslav


### TABLES

**Table 1.** Nucleotide sequences of products of HIV-RNA and M2-RNA recombination in the regions between \( H_{\text{for}} \) and \( M_{\text{rev}} \) PCR primer binding sites.

<table>
<thead>
<tr>
<th>Group</th>
<th>Name</th>
<th>Amp. length</th>
<th>Sequence (between primers)</th>
<th>Number of colonies in experimental series</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>G0</td>
<td>Pr-1</td>
<td>55</td>
<td>([H_{\text{FOR}}] \text{CUAGACGGUGAGUGCUUG}[M_{\text{REV}}])</td>
<td>21</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Pr-2</td>
<td>56</td>
<td>([H_{\text{FOR}}] \text{CUAGACGGUAGUGCUUG}[M_{\text{REV}}])</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr-3</td>
<td>56</td>
<td>([H_{\text{FOR}}] \text{CUAGACGGUAGUGCUUG}[M_{\text{REV}}])</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr-4</td>
<td>58</td>
<td>([H_{\text{FOR}}] \text{CUAGACGGUAGAGUGCUUG}[M_{\text{REV}}])</td>
<td>14</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Pr-5</td>
<td>76</td>
<td>([H_{\text{FOR}}][H_{\text{FOR}}] \text{CUAGACGGUAGAGUGCUUG}[M_{\text{REV}}])</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>Pr-6</td>
<td>61</td>
<td>([H_{\text{FOR}}] \text{CUAGACGGUAGCAGAGUGCUUG}[M_{\text{REV}}])</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr-7</td>
<td>55</td>
<td>([H_{\text{FOR}}] \text{CUAGACGGAGAGUGCUUG}[M_{\text{REV}}])</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr-8</td>
<td>54</td>
<td>([H_{\text{FOR}}] \text{CUAGACGGUGUGCUGU}[M_{\text{REV}}])</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr-9</td>
<td>58</td>
<td>([H_{\text{FOR}}] \text{CUAGACGGUAGAGUGCUUG}[M_{\text{REV}}])</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr-10</td>
<td>104</td>
<td>([H_{FOR}])CUAGACGGUGUAAAAAUCUCUAG CAGUGGAAUGAGGGAAGAAUACGAAA GAAACAGCAGAGUCUGUG [(M_{REV})]</td>
<td>1</td>
<td>-/A</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-------------------------------------</td>
<td>----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>Pr-11</td>
<td>105</td>
<td>([H_{FOR}])CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCGAGGGAAAGAAUACGAAA GGAACAGCAGAGUCUGUG [(M_{REV})]</td>
<td>1</td>
<td>CGC/-1</td>
<td></td>
</tr>
<tr>
<td>Pr-12</td>
<td>105</td>
<td>([H_{FOR}])CUAGACGGUGUAAAAAUCUCUAG CAGUGGUAUGAGGGAAGAAUACGAAA GGAACAGCAGAGUCUGUG [(M_{REV})]</td>
<td>3</td>
<td>-/UA</td>
<td></td>
</tr>
<tr>
<td>Pr-13</td>
<td>106</td>
<td>([H_{FOR}])CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCGCUGAGGGAAGAAUACGAAA GGAACAGCAGAGUCUGUG [(M_{REV})]</td>
<td>12 27</td>
<td>CGC/-</td>
<td></td>
</tr>
<tr>
<td>Pr-14</td>
<td>107</td>
<td>([H_{FOR}])CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCAUGAGGGAAGAAUACGAAA GGAACAGCAGAGUCUGUG [(M_{REV})]</td>
<td>4 2</td>
<td>CGC/A</td>
<td></td>
</tr>
<tr>
<td>Pr-15</td>
<td>108</td>
<td>([H_{FOR}])CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCGCAUGAGGGAAGAAUACGAAA GGAACACAGCAGAGUCUGUG [(M_{REV})]</td>
<td>5 1</td>
<td>CGCC/A</td>
<td></td>
</tr>
<tr>
<td>Pr-16</td>
<td>109</td>
<td>([H_{FOR}])CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCGCAUGAGGGAAGAAUACGAAA GGAACACAGCAGAGUCUGUG [(M_{REV})]</td>
<td>10</td>
<td>CGCCCA/A</td>
<td></td>
</tr>
<tr>
<td>Pr-17</td>
<td>110</td>
<td>([H_{FOR}])CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCGCGAGGGAAAGAAUACGAAA CGAAAGGAACAGCAGAGUCUGUG [(M_{REV})]</td>
<td>1</td>
<td>CGCCCGA/A</td>
<td></td>
</tr>
<tr>
<td>Pr-18</td>
<td>105</td>
<td>([H_{FOR}])CUAGACGGUGUAAAAAUCUCUAG CAGUGGCAUGAGGGAAGAAUACGAAA GGAACAGCAGAGUCUGUG [(M_{REV})]</td>
<td>4</td>
<td>C/A</td>
<td></td>
</tr>
<tr>
<td>Pr-19</td>
<td>104</td>
<td>([H_{FOR}])CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGAGGGAAGAAUACGAAA GGAACAGCAGAGUCUGUG [(M_{REV})]</td>
<td>7 6</td>
<td>C/-</td>
<td></td>
</tr>
<tr>
<td>Pr-20</td>
<td>37</td>
<td>([H_{FOR}]) G [(M_{REV})]</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pr-21</td>
<td>42</td>
<td>([H_{FOR}]) CUAGAC [(M_{REV})]</td>
<td>4 6</td>
<td>2</td>
<td>5 nts terminal loop</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Pr-22</td>
<td>38</td>
<td>$[H_{FOR}]$ AU $[M_{REV}]$</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Pr-23</td>
<td>36</td>
<td>$[H_{FOR}]$ $[M_{REV}]$</td>
<td></td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Pr-24</td>
<td>39</td>
<td>$[H_{FOR}]$ CUA $[M_{REV}]$</td>
<td>2</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Pr-25</td>
<td>170</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAGCAGUGGAGCACCAGGGGACUUUAAAAGUGAAAGAUACAGGGGGGCUUCUACGGAAAGAGACUGAGUCAUGAGGAAAGAAUACUAAGGAACAGCGAGAGUCUGUG $[M_{REV}]$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr-26</td>
<td>169</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAGCAGUGGAGCACCAGGGGACUUUAAAAGUGAAAGAUACAGGGGGGCUUCUACGGAAAGAGACUGAGUCAUGAGGAAAGAAUACUAAGGAACAGCGAGAGUCUGUG $[M_{REV}]$</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Pr-27</td>
<td>169</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAGCAGUGGAGCACCAGGGGACUUUAAAAGUGAAAGAUACAGGGGGGCUUCUACGGAAAGAGACUGAGUCAUGAGGAAAGAAUACUAAGGAACAGCGAGAGUCUGUG $[M_{REV}]$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr-28</td>
<td>172</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAGCAGUGGAGCACCAGGGGACUUUAAAAGUGAAAGAUACAGGGGGGCUUCUACGGAAAGAGACUGAGUCAUGAGGAAAGAAUACUAAGGAACAGCGAGAGUCUGUG $[M_{REV}]$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr-29</td>
<td>39</td>
<td>$[H_{FOR}]$ GAU $[M_{REV}]$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr-30</td>
<td>39</td>
<td>$[H_{FOR}]$ GUG $[M_{REV}]$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr-31</td>
<td>44</td>
<td>$[H_{FOR}]$ CUGCUUGUG $[M_{REV}]$</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr-32</td>
<td>45</td>
<td>$[H_{FOR}]$ CUAGACGGU $[M_{REV}]$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**G4**

**G5**
<table>
<thead>
<tr>
<th>Pr-33</th>
<th>55</th>
<th>$[H_{FOR}] [H_{FOR}] A [M_{REV}]$</th>
<th>1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr-34</td>
<td>63</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUGCUGU [M_{REV}]</td>
<td>3</td>
<td>1 nt from 6 nts internal loop</td>
</tr>
<tr>
<td>Pr-35</td>
<td>75</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCCCGCUGUG [M_{REV}]</td>
<td>3</td>
<td>6 nts assymetric internal loop</td>
</tr>
<tr>
<td>Pr-36</td>
<td>87</td>
<td>$[H_{FOR}]$CUACCUGAGUCUAUGAGGGAAGA GUAUAGCAAAGGAAACACAGCAGAGCUUGU [M_{REV}]</td>
<td>2</td>
<td>5 nts terminal loop</td>
</tr>
<tr>
<td>Pr-37</td>
<td>91</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG CAGUGUAUCGAAAGGAACACAGCAGAGCUGUG [M_{REV}]</td>
<td>1</td>
<td>4 nts internal loop</td>
</tr>
<tr>
<td>Pr-38</td>
<td>133</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCCCGAACAGGGACUUUAAGUGAAAGUAGGGAAGAAUACGAGGAAACACAGCAGAGCUGUG [M_{REV}]</td>
<td>1</td>
<td>1 nt crossing of two helical stems</td>
</tr>
<tr>
<td>Pr-39</td>
<td>77</td>
<td>$[H_{FOR}]$CUAGACCGGUGUAAAAAUCUCUAG CAGUGGCGCCCGAACAGG [M_{REV}]</td>
<td>1</td>
<td>1 nt from 4 nts terminal loop</td>
</tr>
<tr>
<td>Pr-40</td>
<td>73</td>
<td>$[H_{FOR}]$CUAGACCGGUGUACGAAAGGAAGACGAGAGUCUGUG [M_{REV}]</td>
<td>2</td>
<td>1 nt bulge</td>
</tr>
<tr>
<td>Pr-41</td>
<td>109</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCCCGAACAGGGACUUUAACUCG GAGGAGAACACGAGCAGGUG [M_{REV}]</td>
<td>1</td>
<td>AA is substituted by GG in M2 region</td>
</tr>
<tr>
<td>Pr-42</td>
<td>119</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCCCGAACAGGGAAACAGGAGGGAAGGAAAGAACAGCAGAGCAGUGUG [M_{REV}]</td>
<td>1</td>
<td>3 nt crossing of two helical stems</td>
</tr>
<tr>
<td>Pr-43</td>
<td>74</td>
<td>$[H_{FOR}]$CUAGACCGGUGUAAAAAUCUCUAG CAGACGGTACUGUGCGCUGUG [M_{REV}]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pr-44</td>
<td>42</td>
<td>$[H_{FOR}]$GCUGUG [M_{REV}]</td>
<td>5</td>
<td>4 nts terminal loop</td>
</tr>
<tr>
<td>Pr-45</td>
<td>43</td>
<td>$[H_{FOR}]$ CUAGACG [M_{REV}]</td>
<td>4</td>
<td>1 nt from 2 nts internal loop</td>
</tr>
<tr>
<td>Pr</td>
<td>55</td>
<td>[H_{FOR}] [H_{FOR}] G [M_{REV}]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----</td>
<td>---------------------------------</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Pr-47</td>
<td>57</td>
<td>[H_{FOR}] CUAGACGGUGUAGGUGUGUG [M_{REV}]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pr-48</td>
<td>61</td>
<td>[H_{FOR}] CUAGACGGUGUAACUCUCUAGCA [M_{REV}]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Between two 1nt bulge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr-49</td>
<td>73</td>
<td>[H_{FOR}] [H_{FOR}] [H_{FOR}] G [M_{REV}]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pr-50</td>
<td>71</td>
<td>[H_{FOR}] CUAGACGGUGUAAAAUCUCUAGCAGUGCUUG [M_{REV}]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3 nts RNA bulge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1) Only RNA sequences between regions corresponding to primers H_{FOR} and M_{REV} are shown.

2) Number of colonies, detected by bacterial colony PCR and confirmed via nucleotide sequencing of plasmids; correspond to four experimental series: A, B, C, F (see legend for Fig. 2 for keys)

3) RNA regions originating from HIV-RNA are shown in red, from M2-RNA—in blue. Template-binding regions are underlined; sequences corresponding to primers are shown as [H_{FOR}] and [M_{REV}]. RNA nucleotides within bulge loops are not underlined, nucleotides in internal loops and mismatches are shown in italic. External insertions are highlighted with yellow.
Table 2. Summary of mean backbone RMSd and average inter-atomic distances ($d$) between the ligation points (i.e. 2’-P or 3’-P atom of 3’ cytidine (HIV-1 RNA) and the O5’ atom of 5’ uridine (M2-RNA)) for 53-mer, 38-mer, 33-mer and 27-mer model RNAs.

<table>
<thead>
<tr>
<th>Length of model</th>
<th>RMSd (backbone)</th>
<th>RMSd (loop)</th>
<th>Average inter-atomic distance ($d$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>53nt</td>
<td>5.998 Å ± 1.732</td>
<td>2.240 Å ± 0.392</td>
<td>7.410 Å ± 2.610</td>
</tr>
<tr>
<td>38nt</td>
<td>8.515 Å ± 2.037</td>
<td>1.544 Å ± 0.255</td>
<td>9.930 Å ± 5.576</td>
</tr>
<tr>
<td>33nt</td>
<td>5.708 Å ± 1.410</td>
<td>1.479 Å ± 0.252</td>
<td>6.308 Å ± 2.469</td>
</tr>
<tr>
<td>27nt</td>
<td>4.253 Å ± 0.814</td>
<td>1.606 Å ± 0.291</td>
<td>6.537 Å ± 2.329</td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Figure 1. General scheme of RNA recombination pathways in the presence and in the absence of ON-template. Fragments corresponding to HIV-RNA are shown in red, M2-RNA – in blue, ON-template – in black.

Figure 2. Products of RT-PCR with 5'-32P-Mrev primer, 10% denaturing PAAG. RNAs marked in caption were used in relevant recombination reactions. In “+ template” series recombination was performed in the presence of ON-template, “- template” – in the absence of ON-template. Lanes “1”: positive control PCR with primers Mfor/Mrev (25 PCR cycles). Lanes “2”, “3”, “4”: PCR with primers Hfor/Mrev, with cDNA dilutions 1:1, 1:10\(^{-3}\), 1:10\(^{-6}\), respectively (28 PCR cycles). In “E” and “F” series we used non-incubated mixtures of RNA and template for RT-PCR reactions. K1: negative control of RT-PCR with primers Hfor/Mrev, with reverse transcription performed in the absence of any RNA. K2, K3: negative controls of PCR in the absence of cDNA with primers Hfor/Mrev and Mfor/Mrev (25 PCR cycles). L1: dsDNA ladder. Positions of supposed products of template-directed ligation are shown with arrows (X-1 through X-9).

Figure 3. The most stable secondary structures of Pr-1 (A and B) (G0) and Pr-21 (D) (group G3). C. Common domain of products Pr-10 through Pr-19 (group G2). N and N’ indicate different nucleotides within ligating RNA dangling ends, n and m – number of nucleotides (see Table 1 for details). HIV-RNA fragment is shown in red, M2-RNA fragment – in blue. Ligating ends are marked as 3’ and 5’. E. Ligation site of Pr-42 (group G5). D. Variants of ligation site of G2 products. Dotted lines connect ligating nucleotides.

Figure 4. Two dimensional representations of the proposed structures of: A. HIV-1 A-rich hairpin loop. The nucleotides in red represent the naturally occurring nucleotide bases, those added by the authors to aid transcription are indicated in black. Insertion (Ins) shows the NMR-based 3D structure of HIV-1 RNA A-rich hairpin loop (PDB: 1BVJ, taken from (Puglisi EV and Puglisi JD 1998)). B. Model RNA structures of varying stem length (L-R: 53-mer, 38-mer, 33-mer, 27-mer). The nucleotides in red represent those belonging to the 3’ recombination partner and those in blue, correspond to the 5’ recombination partner. The dashed line encloses the part of the models corresponding to the experimentally derived structure from NMR (shown as insertion in A).

Figure 5. 3D structures of the selected RNA models systems 53nt (A), 38nt (B), 33nt (C) and 27nt (D), each representing the average structure from the 100 ns Molecular Dynamics run for the designated models. The red and blue nucleotides highlight the key ligation points between 3’-C of HIV-1 RNA fragment and 5’-U of M2-RNA fragment, respectively. The
rectangular insertions represent the enlarged 3D structures of the ligation sites showing the minimum distance between the ligation points (i.e. (HIV-1 RNA) 3’C > $\textbf{P}$ ... $\textbf{O}$5’: U5’ (M2-RNA)), measured over the 100 ns Molecular Dynamics simulations for each model.

**Figure 6.** RMSd over 100ns simulation of the backbone atoms of each of our four model RNAs, from the original structure (red): (A) 53-mer; (B) 38-mer; (C) 33-mer and (D) 27-mer. RMSd of the backbone atoms in the loop portion of each model from the NMR structure are shown in black.

**Figure 7.** Distance between the potential ligation partners taken as the distance d between the 2’- or 3’-P atom of 3’cytidine (HIV-1 RNA) and the O5’ atom of 5’uridine (M2-RNA) in each of the four model RNAs, over 100 ns simulation: (A) 53-mer; (B) 38-mer; (C) 33-mer and (D) 27-mer. The dashed read line represents the distance limit for ligation to occur. For each model RNA, P ... O5’ reached distance within 5Å at multiple points during 100 ns Molecular Dynamics simulations, totalling between 20% and 30% of the total simulation time in each case.

**Figure 8.** Opportunities for ligation, forming a 3’, 5’- linked product. Co-incidence of distance P-O5’ between 3 - 5Å and angle O5’-P-O2’ between 150-180˚, for each of our 4 model RNAs: (A) 53-mer; (B) 38-mer; (C) 33-mer and (D) 27-mer.

**Figure 9.** Opportunities for ligation, forming a 2’, 5’- linked product. Co-incidence of distance P-O5’ between 3 - 5Å and angle O5’-P-O3’ between 150-180˚, for each of our 4 model RNAs: (A) 53-mer; (B) 38-mer; (C) 33-mer and (D) 27-mer.
Staroseletz Yaroslav

**Non-enzymatic Recombination of RNA: Ligation in Loops**

Staroseletz Yaroslav¹*, Sergey Nechaev¹*, Elena Bichenkova², Richard A Bryce², Catherine Watson², Valentin Vlassov¹ and Marina Zenkova¹

¹Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of Russian Academy of Sciences, 8 Lavrentiev Avenue, Novosibirsk, Russia. 630090.

²Division of Pharmacy and Optometry, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Oxford Road, Manchester, M13 9PT, UK

*These persons contributed equally to this work

Reprint requests to: Marina Zenkova, Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of Russian Academy of Sciences, 8 Lavrentiev Avenue, Novosibirsk 630090, Russia, e-mail: marzen@niboch.nsc.ru.

**Keywords:** non-enzymatic RNA ligation, RNA world, origin of life, molecular dynamics.
Staroseletz Yaroslav

Abstract

Background

While the RNA world hypothesis is widely accepted, it is still far from complete: the existence of self-replicating ribozyme, consisting of potentially hundreds of nucleotides, is a core assumption for the majority of RNA world models. The appearance of such long RNA molecules under prebiotic conditions is not self-evident. Recombination seems to be a plausible way of creating RNA diversity, resulting in the appearance of functional RNAs, capable of self-replicating.

Methods

We report here on the study of recombination process modeled with two 96 nts RNA fragments. Detection of recombination products was performed with RT-PCR followed by TA-cloning and Sanger sequencing.

Results

A wide range of recombinant products was detected. We found that (i) the most efficient ligation was observed for RNA species forming bulges or internal loops, with ligation partners located within the loop; (ii) a strong preference was observed for formation of a few types of major products with a large variety of minor products; (iii) ligation could occur with participation of either 2',3'-cyclophosphate or 5'-ppp; (iv) the presence of key reaction components, i.e. 5’ppp-RNAs, enabled the formation of additional types of product; (v) molecular dynamics simulations of one of the most abundant products suggests that the ligation results in a preferable formation of 2'-5' rather than 3'-5'-linkages.

Conclusions

The study demonstrates regularities of new RNA molecules formation with non-enzymatic recombination process.

General significance

Our findings provide new data supporting the RNA World hypothesis and show the way of new RNA sequences emergence under prebiotic conditions.

1. Introduction

The RNA world hypothesis states that systems based on an RNA genome and RNA catalysts preceded current forms of life [1,2]. Several model systems of the RNA world have been described in detail [3–6]. The central notion of these models is an autoreplicase enzyme, an RNA molecule that can catalyze its own construction through template-directed polymerization of nucleotides. The plausibility of this theory was confirmed by several experimental findings. The strongest argument for the theory’s accuracy is the creation of artificial ribozymes capable
of synthesizing a spectrum of RNA sequences, like aptamers, ribozymes [7] and tRNA [8]. Moreover, amplification of short RNA templates in an RNA-catalyzed form of the PCR was shown for one type of ribozyme [8]. Other classes of in vitro selected ribozymes display such prebiotically important activities as nucleoside synthesis, triphosphorylation and polymerization of nucleoside triphosphates [9,10]. Dozens of recently created ribozymes with a wide variety of catalytic activities provide promise that a self-replicating ribozyme will be discovered.

Nevertheless, these RNA World models display various shortcomings [5,11]; in particular, the means of origin of such systems is unclear. One could envisage the following route to development of the RNA world: nucleotide synthesis from inorganic substances, formation of an RNA pool and then appearance of functional RNAs in this pool. Two possible ways could lead to the formation of the RNA pool: polymerization of nucleotides or recombination of RNA fragments. Polymerization of activated nucleotides produces RNA fragments not longer than 50 nucleotides [12,13], or the same order of magnitude [14] whereas the size of a presumable RNA replicase ribozyme is some hundreds of nucleotides in length [4,7–10]. The gap between short abiotic RNA fragments and long catalytically proficient molecules could readily be met by spontaneous RNA recombination occurring via SN2-mechanism through formation of a 2’,3’-cyclophosphate [15,16]. Distribution of lengths of RNA fragments formed by polymerization remains exponential with or without recombination but only if recombination occurs at equal rates at all points [17]. If the recombination occurs not randomly but at certain number of “hot points”, it can shift distribution of lengths, making possible formation of fragments long enough to show a capacity for catalysis.

Since the first demonstrations of the RNA recombination phenomenon using poliovirus model in the 1960s [18,19], a great number of studies have been devoted to RNA recombination, reviewed elsewhere [16,20,21]. The most widely accepted model of RNA recombination is a ‘copy choice’ recombination. In this process, the RNA polymerase that mediates viral replication switches from one template to another during synthesis, while remaining bound to the nascent nucleic acid chain [21]. Non-replicative RNA recombination has also been demonstrated experimentally [22–24]. It can be explained by transesterification reactions, not necessarily involving an enzymatic activity. Some of the crossover loci are clustered, thus indirectly indicating the involvement of structures similar to known ribozymes in the recombination reaction [22]. The co-transfection of replicating and nonreplicating rubella virus RNA transcripts containing nonoverlapping deletions did restore the infectious virus [23].

A series of studies published by Lehman’s group revealed fascinating phenomena of RNA recombination. In particular, it was demonstrated that four special RNA fragments of 40-
60 nucleotides can self-assemble with the formation of the Azoarcus intron I [25]. The latter, in turn, catalyzes RNA oligonucleotide recombination. There are some constraints concerning the sequence of the substrates; nevertheless, hammerhead ribozyme and ligase ribozyme were obtained from nonfunctional RNA fragments with Azoarcus group I intron as a catalyst [26,27]. These results show that entire catalytic RNA motifs could be built up from smaller RNAs through a series of energy neutral reactions [26]. These data are in accord with the model suggested by Briones [28]: in this model, a fraction of simple RNA fragments, notably hairpin-like ones, are obtained by random polymerization on mineral surfaces; these fragments could display RNA ligase activity and catalyze the assembly of larger, eventually functional RNA molecules [28].

All these studies indicate that RNA recombination was likely to play an important role in the early stages of evolution. But the open question is, could RNA recombination occur without the involvement of an external complex catalyst (either enzyme or ribozyme) and lead to the increase of RNA length and complexity. The major step in this direction was made in the works of Chetverin. Nonhomologous recombination between fragments of a viral RNA in a cell-free system was demonstrated using the molecular colonies technique [29,30].

In our earlier works with chimeric DNA/RNA oligonucleotides, the chemical mechanism for a new phosphodiester bond formation was investigated. It was shown that the reaction intermediates possessed a 2',3'-cyclophosphate and a 5'-OH group, and that the ligation proceeded through the mechanism of intermolecular transesterification [31–33]. In a previous paper, we reported studies of the consecutive cleavage/ligation reaction of two 96nt long dephosphorylated RNA fragments, catalyzed by Mg$^{2+}$. Similarly to [31–33] the reaction was performed in the presence of oligodeoxyribonucleotide template (ON), which guided the recombination within selected RNA sites (5'-OH RNAs, ON+). The products of recombination were amplified, isolated and their nucleotide sequences were determined. Based on the Sanger sequencing data, some preliminary conclusions on the characteristic properties and mechanisms of recombination events were made [34].

In this study, we examined reaction routes and range of recombinant products in other settings and compared range of novel recombinant products with those obtained under previously investigated conditions (5'-OH-RNAs, ON+). Firstly, we studied the system behavior in the absence of the template (5'-OH-RNAs, ON-) and found that free spontaneous (template-independent) recombination in our system also existed and was efficient. Then, we used 5’-ppp-RNAs instead of 5’-OH-RNAs both in the absence (5’-ppp-RNAs, ON-) and in the presence of the template (5’-ppp-RNAs, ON+). This resulted not only in appearance of new products of
recombination, but surprisingly influenced the ratio between the products. Finally, we examined
the most abundant recombination product in order to evaluate the probability of formation either
2'-5' or 3'-5'-linkages using molecular dynamics simulations.

Rationale behind the choice of RNA fragment is as follows. Appearance of 96-nts RNA fragments under prebiotic conditions is not proven yet however it is plausible similar to recombination-mediated formation of 198-nts Azoarcus intron I ribozyme from four 40-60-nts fragments [25]. As mentioned above random recombination is not very useful in formation of long RNA structures so the concern is to find short RNA structures (up to 50 nts according to [12]) prone to recombination preferably occurring at certain sites resulting in formation of long RNA sequences. Prevalence of such structures is low however it gets higher the longer RNA fragment formed [35]. Thus it seems reasonable to “shift” the limit of the search space from 50 nts to, for instance, 100 nts.

2. Results

2.1. Model system and reaction conditions.

Non-enzymatic reaction of RNA recombination was investigated with the use of two 96nt RNA fragments of viral RNAs as substrates (96 nts fragments of influenza virus M2 RNA and HIV-1 RNA, here and after M2-RNA and HIV-RNA, respectively) [34]. We hypothesize here that non-enzymatic recombination of two RNA fragments occurs through preliminary cleavage of substrates into shorter molecules, which are expected to be more eligible for ligation than intact substrates. In our reaction system, cleavage and ligation pathways are not spatially separated and occur, without any specific control, in one reaction system under constant thermal and chemical conditions. However, we will separate all the reactions into elementary steps in our further discussion to clarify the mechanisms of non-enzymatic processes.

In the first step, RNA molecules undergo cleavage in the presence of magnesium ions, known to catalyze these reactions through several mechanisms. Any RNA molecules contain "fragile" sites, whose location is predetermined by the RNA structure. The RNA substrates used in this study were investigated earlier with a view to locate "fragile" sites, which are particularly unstable under cleavage conditions. Location of these sites was recently discussed [34]. As a result of the cleavage step, a set of molecules comprising 2',3'-cyclophosphate at the 3'-end is formed. All of these molecules can be regarded as potential substrates of the ligation reaction, occurring through the mechanism of intermolecular transesterification and yielding a new phosphodiester bond between the reacting RNA molecules. The described mechanism of the transesterification reaction is generally acknowledged as a primary one, as far as 2',3'-cyclophosphate is essentially a self-activated form of a substrate, and should be considered as the
form of activated substrate in the absence of other chemical agents, such as BrCN, EDAC, etc [31,33,36].

Bearing in mind that these processes are random to a certain extent, ligation proceeding through the pathway described above should lead to the production of a large number of new molecules. Previously we limited the freedom of RNA substrates to react with random RNA "partner" with the use of oligodeoxyribonucleotide (ON-template) as a template for bringing together two certain RNA fragments, formed at the cleavage step, and the use of dephosphorylated RNAs (5'-OH-RNAs, ON+) [34]. In this study we used these conditions as a control of reproducibility of cleavage/ligation events and identified recombinant products emerged in other conditions (5'-OH-RNAs, ON-, 5'-ppp-RNAs, ON+, 5'-ppp-RNAs, ON-).

At the preliminary step formation of RNA/DNA duplexes was examined via incubation of M2- and HIV-RNAs in the presence of 5'-'32P-ON-template at 37˚C under two conditions: for 2 h in the absence of magnesium and for 72 h in its presence (Supplemental Fig. S1).

ON-template/M2-RNA duplex is readily formed upon incubation for 2 h (product A), while HIV-RNA is not prone to form this heteroduplex. Nevertheless, prolonged incubation (72 h) leads to formation of several duplexes of template with HIV-RNA fragments of different lengths (the major product is marked as "C"). This indicates the more facile formation of duplexes between template and partially cleaved RNA, experiencing less spatial complications for approaching the template. In the lanes corresponding to M2-RNA incubated 72 h with the template, one extra major product (B) is formed in addition to that detected upon 2 h incubation. In addition, a number of minor products, corresponding to a wide range of RNA lengths, are observed between bands “A” and “B” (Supplemental Fig. S1).

2.2 Reaction scheme

The suggested scheme of processes leading to the formation of recombinant products is depicted in Fig. 1. As it was shown previously, upon RNA incubation in the presence of Mg2+, a pool of activated substrates bearing 2',3'-cyclophosphosphate is formed (Fig. 1 1(a), 2(a), 2(b)). These substrates can be either free or complexed with the ON-template. Their self-assembling (root B) or association with the ON-template (root A) can lead to the formation of different complexes, i.e. those which are ready for ligation (Fig. 1 3(a)), those which are undergoing further cleavage in single-stranded regions formed within the complex with a template (Fig. 1 3(b)), and those which are self-assembling independently from the ON-template (Fig. 1 3(c), 3(d)). When the appropriate spatial complex for ligation is attained, the 5'-OH group of a 5'-substrate “attacks” the 2',3'-cyclophosphosphate of a 3'-substrate, with formation of a new phosphodiester bond (Fig. 1 4(b), 5(a,b)).
We note here that RNA complexes and recombinant products shown in Fig. 1 are limited to those where HIV-RNA is a 3’-substrate, and M2-RNA is a 5’-substrate. Actually, different RNA molecules can emerge in the reaction system, but an approach applied to detect recombinant products allows us to amplify only those formed by the pairs of substrates discussed above. The method of detection was described in detail in our previous work [34], and here only its basic principles are discussed. In the first stage, a pool of RNA molecules obtained after recombination is used as a mixture of templates for a reverse transcription reaction. Reverse transcription is performed with a use of sequence-specific Mrev primer. In the second stage, we ran PCR from the obtained pool of cDNAs using primers Hfor and Mrev (H – for HIV-RNA, and M – for M2-RNA specific primers). From these two stages, we obtain a mixture of PCR products corresponding to recombinant RNAs that saved the regions corresponding to Hfor and Mrev primers in their structures and were formed by HIV-RNA fragment as a 3’-substrate and M2-RNA as a 5’-substrate.

2.3 Detection of recombinant products: RT-PCR

In PCR, we used 5’-32P-labelled Mrev primer to ensure detection of all the amplicons via radioautographing of gels. Fig. 2 shows an autoradiogram of gel after separation of PCR products obtained from the recombinant RNA products incubated under different recombination conditions. Those conditions involved (i) use of 5’-triphosphorylated and 5’-OH-RNAs; (ii) 30 and 100 ml total reaction volume; (iii) in the presence (A1, A2, C, E) and in the absence (B, D, F) of ON-template; and (iv) PCR matching to non-incubated recombination reaction mixtures (E, F). We used a partially degraded at purine sites amplicon of M2-RNA, synthesized with Mfor and 5’-32P-Mrev primers, as a ladder.

cDNA was used in dilutions 1:1, 1:10³, 1:10⁶, and PCR included 28 cycles of amplification. For a positive control, we run 25 cycles of PCR with primers Mfor/Mrev and cDNA mixture in a dilution 1:10³. Negative controls involved reactions under three conditions: (1) PCR with primers Hfor/Mrev, with reverse transcription performed in the absence of any RNA; (2) PCR in the absence of cDNA, with primers Mfor/Mrev; and (3) PCR in the absence of cDNA, with primers Mfor/Mrev. We separated radioactive amplicons in 10% denaturing PAAG. Titles of the experimental series used in the capture of Fig. 2 will be used further for marking relevant experimental models.

As shown on Fig. 2, ligation products can be divided into two major types: (1) those formed under all reactions conditions (X-9 group of products) and (2) those formed only in the presence of one of the key components of the reaction system. These components include ON-template (X-1, X-2, and X-5) or triphosphate group at the 5’-terminus of M2-RNA (X-1 through
Staroseletz Yaroslav

X-4). It is noteworthy that products of the first type are formed in different proportions depending on specific reaction conditions, though this dependence is not strict. Lengths of products X-5, X-6, X-7 and X-8 are very close to the lengths of supposed products of template-dependent ligation. However, at least some of them seem to be formed in the absence of template as well, indicating the obvious limit of using electrophoretic methods for the explanation of sophisticated recombination events. Comparison of series A1 and A2 shows that the reaction is insensitive to the total volume increase, and 30 µl total volume used in most reactions is sufficient for the formation of recombinant products.

Comparative analysis of products in series A1 and B shows that X-6 disappears when the ON-template is excluded from the reaction system. A similar observation can be made for products X-1 and X-2; however correlation with the presence of template is less evident in the latter case.

Series C generates the only major product, X-6, and thus indicates the high selectivity of reaction toward formation of this specific product. Remarkably, the only difference between series A1 and C was the use of 5'-OH-RNA in series C. Therefore, in this group we exclude the participation of one of the reaction “components”, 5'- triphosphate group, and that leads to a sharp reduction in the diversity of the products, particularly so of the longer fraction. In summary, these data indicate that X-6 ligation occurs with participation of 2’,3’-cyclophosphate as an active group, and ligation is active only in the presence of ON-template (compare C, B and D series in Fig. 2).

The results of all negative controls are satisfactory; therefore formation of significant amounts of non-specific products in reverse transcription reaction and PCR is not observed.

2.4 Molecular cloning

The data presented in Fig. 2 indicate that the variety of products obtained under different reaction conditions is very large, and the electrophoretic method does not allow us to elucidate the mechanism and patterns of formation of recombinant products. In addition, it is not sensitive enough to detect the majority of products in the reaction mixture. Therefore we decided to sequence PCR products and to align the obtained sequences with those of initial RNA substrates. However, that required preliminary isolation of each PCR product from the mixture of amplicons. To achieve this goal, we used the TA cloning procedure, described previously [34] in detail and involving cloning of plasmids containing T-vector part (bearing sticky ddT-ends) and amplicon-part (bearing sticky mononucleotide 3'-A-ends). After transformation of competent E.coli cells with a pool of plasmids containing different types of TA-inserts and plating at Amp-containing LB agar, we obtained a set of colonies, each consisting of bacterial cells with a
specific type of plasmid. Colonies were screened for TA insert size via bacterial colony PCR with primers M13/pUC_for and M13/pUC_rev, directed to a permanent plasmid region outside of the insert area. A representative electrophoresis after separation of PCR products is shown in Supplemental Fig. S2. We divided all the products into groups on the basis of their lengths; from one to twelve colonies within each group were selected for plasmid sequencing. PCR products from 4 reaction groups were subjected to this procedure: A1, B, C and F (see Fig. 2 for keys). Group F was chosen as a negative control, group C, described in our previous work, as a positive one. In total, plasmids in 379 colonies were analyzed via bacterial colony PCR (81 colony in group A, 134 in B, 106 in C, and 58 in F), and the number of sequenced colonies was 35, 57, 27 and 19, respectively. After alignment with parent RNAs, 50 unique products were detected. The Sanger sequencing data are summarized in Table 1; the nucleotide sequence of each product in the area between primers H_for and M_rev and statistics of occurrence of each product in all the experimental series were determined.

2.5 Classification of recombinant products

The data presented in Table 1 demonstrate that the diversity of products arising from RNA recombination is very wide. Indeed, some products can be regarded as major ones (e.g. Pr-1, Pr-4, Pr-13, Pr-23); others were detected only in a few bacterial colonies (Pr-6, Pr-14, Pr-15, etc.); the rest are unique ones and were found only in a single colony (Pr-2, Pr-10, Pr-11). Each product can be also found in one or several experimental series (A, B, C and F). The first straightforward criterion for a product’s classification is the requirement of template for its formation. Group G1 (template-dependent, Pr-2 through Pr-9) unites products of ligation within bulges and internal loops, formed in a partially complementary complex of RNAs with ON-template. In this case, the template facilitated an approach of ligation of RNA fragments, but the ligation occurred in single-stranded regions of RNAs, bulged out of the heteroduplex in the elements of imperfect secondary structure of RNA. A large number of detected products were formed outside of the complementary complexes of RNAs with ON-template as a result of template-independent ligation (all other groups, Pr-10 through Pr-50). Herein and thereafter "template-independent" ligation means ligation occurring without the participation of ON-template, while this template may be present in the reaction mixture and/or some regions of RNA substrates may serve as templates for self-ligation. Pr-1 stands apart from all other products, since it can be formed as a result of both template-dependent and template-independent ligation.

In order to analyze the range of template-independent products, we classify them according to their distribution in experimental series (A, B, C and F). However, some products of
template-independent reaction were extremely similar to each other (with a difference in a few nucleotides in length and sequence) and were presented by too large number of colonies to be referred to as a random convergence of RNA fragments. Therefore, those products (Pr-10 through Pr-19) were attributed to a particular group G2. It contains products of template-independent ligation in RNA bulges and internal loops, formed within self-assembled structures of RNA containing double-stranded regions. Group G3 is a mixture of products with moderate frequency and encountered mainly in 3 or 4 experimental series, except Pr-23, that was found only in the F series. However, as it is one of the major products, we separate it from rare products of the G5 group. The G4 group contains products formed with participation of 5’-ppp and G5, combining residually all other rare products found only in one experimental series. All groups of recombinant products are described in detail below.

2.6 Product Pr-1

Pr-1 is a unique product, readily formed in series A and C in the presence of the ON-template, and its formation was described in detail in [34]. However, as summarized in Table 1, twice as much of relevant colonies are formed in the series "B", where no template is used. Moreover, Pr-1 is a major product in that experimental series. Therefore we assume that this product attains a very stable tertiary structure, which may facilitate ligation of the reacting RNA substrates. Two alternative secondary structures (Fig. 3 A,B) were found for this product with RNAfold [37]. In the first one, ligating 3’ and 5’ dangling ends of HIV-RNA and M2-RNA (1 nt and 3 nts, respectively) are located in the end of 10-pairs long RNA stem, turned into stem loop when a new phosphodiester bond is formed (Fig. 3 A). Another structure is an RNA stem-loop, formed by two RNA fragments. Within this complex, one 5’-nucleotide of a 5’-substrate is dangling, and ligation is accompanied by the formation of 1 nt bulge in the ligation site. This situation is similar to that in Pr-2, but in the current case a fragment of HIV-RNA serves both as a substrate and as a template for assembling of ligation intermediary complex.

2.7 G1 products

Group G1 includes 7 types of product: Pr-2 through Pr-9. All of them are formed exclusively in the presence of the ON-template, and ligation occurs within bulge loops of a size one (Pr-2, Pr-3), three (Pr-4, Pr-5) and six (Pr-6) nucleotides, as well as within 2 and 3 nts internal loops, formed by the RNAs and ON-template (Pr-7 and Pr-8, respectively). Most of these products were described and discussed by us earlier [34]. The main unexpected result from the analysis of this group is the strongest preference for formation of one type of product. Pr-4 was found in 101 colonies, 87 out of them belonging to the C series, 14 belonging to the A series, whereas all other products were found at best in 5 colonies. Thus ligation may occur in
different structures like 1- or 6-nts bulges either 2- or 3-nts internal loops (possibly, in some other structures, which have not been detected in our research) but the most favorable one is 3-nts bulge with 1- and 2-nts dangling ends, for 3’ and 5’ ligating fragments, respectively.

2.8 G2 products

Products of group G2 (Pr-10 through Pr-19, Fig. 3 C) have the largest degree of reciprocal similarity in comparison to products from other groups. Lengths of the relevant amplicons vary in the range of 104-110 nucleotides, and all the modelled secondary structures contain the same and very stable prolonged domain, shown in Fig. 3 C. That domain is formed mostly by a fragment of HIV-RNA, and a 7 bp duplex with M2-RNA fragment is formed at its 5'-end. Remarkably, the secondary structure of the initial HIV-RNA, used as a substrate, remains unchanged in this region, and hence this fragment serves as a feedstock for a template and as a 3'-side substrate for ligation at the same time. Therefore, it is reasonable that products formed in this sort of complex are present in all the experimental series, except a negative control one. The absence of any products from this group in the control series F, which refers to non-incubated recombination reaction mixtures, denotes a time requirement for the formation of those products. As highlighted above, cleavage of initial RNAs is a process occurring at moderate rates, and the lack of products in the control experiment argues that cleavage and, consequently, simplification and adjustment of secondary structure, is essential for the formation of stable intermediates.

Depending on the product type, dangling 3' and 5' RNA ends at the ligation site of G2 products have different lengths, and their sequences are provided in Table 1, along with a total number of colonies referring to each product in the experimental series. As shown in the table, the 3'-end can be 0, 1, 3, 4, 5, or 6 nucleotides long, while 5'-end is normally shorter and consists at most of two nucleotides. In the case of products Pr-11, Pr-13 and Pr-19, the 5'-end is fixed by the complementary interactions with an opposite strand, and in Pr-11, it is one nucleotide shorter than in Pr-13 and Pr-19.

The preference for the product formation here is striking, although not so strongly marked as in the G1 group. Between the major product (Pr-13, 39 colonies) and the unique products (Pr-10 and Pr-11, single colony each) several products were found with moderate frequency. The distribution of products between experimental series is also interesting. G2 products are most present in series B (46%, 61 colonies out of 134), twice more than in series A (23%, 19 colonies out of 81) and poorly represented in C series (4 out of 106). Therefore, the presence of template shifts the product formation process to the template-dependent products, decreasing the percentage of G2 products in series A. The low number of G2 products in the C series (dephosphorylated RNAs) is harder to explain. All explanations of this phenomenon serve
not more than speculation; however, it’s possible that formation of G2 products involves several stages, including formation of intermediate products like Pr-25 or Pr-26, which demand 5’-ppp-M2-RNA. Thus, by excluding 5’-ppp from reaction system, we can prevent formation of G2 products.

2.9 Molecular modelling of G2 products

Our primary goal here was to investigate the structural rules and molecular mechanisms of non-enzymatic RNA recombination using computational methods. To achieve this we used unrestrained Molecular Dynamics simulations focusing on behaviour around the ligation site for Pr-13 and three shortened models in explicit aqueous solvent (Fig. 4 for two dimentional representations and Fig. 5 for 3D structures). The choice of the model was due to the fact that Pr-13 is the most abundant product among template-independent ones, and the further development of our study intends to be concentrated mainly on template-independent recombination. All models preserve the same ligation site and differ in distal stem and loop regions, thus showing the influence of those regions on the geometry of the ligating groups. The root-mean-squared deviation (RMSD) in coordinates of backbone atoms, with respect to the initial modelled structure, was calculated over the time-course of the simulation. We also evaluate the backbone RMSD of the loop region of each model, relative to the NMR structure of the HIV-1 A-rich hairpin loop (Fig. 6).

The four model RNAs demonstrated relatively large fluctuations in backbone RMSD from its corresponding starting structure (Fig. 6): for example for the 53nt Pr-13 product, the average RMSD value is 6.0 ± 1.7 Å over the course of the 100 ns simulation (Table 2). The fluctuations observed here are suggestive of a high degree of flexibility within the models. The average backbone RMSD for the loop region, however, is considerably lower, ranging from 1.5 to 2.2 Å across the four models (Table 2). This indicates that the loop region is rather stable, agreeing well with the NMR model of [38], and that the mobility arises from the stem part of the structures.

To evaluate the likelihood of opportunities for recombination events within each model, the distance between the potential ligation partners was measured at each time point (Fig. 7). This was taken as the distance \(d\) between the 2’,3’ P atom of 3’cytidine (HIV-1 RNA) and the O5’ atom of 5’uridine (M2-RNA). For Pr-13, the average interatomic distance over 100 ns simulation is 7.41 ± 2.61 Å (Table 2). While the mean distance between the terminal atoms of the proposed ligation partners is relatively high, there is a large amount of fluctuation in this distance over the course of 100 ns. As such, there are multiple occurrences during the trajectory where this 2’,3’ P···O5’ distance comes within the required 3 - 5 Å distance for ligation to be
possible. Sampling of distances below 4 Å is evident from the time series of \( d \) for all four RNA models (Fig. 7).

Another consideration for the likelihood of non-enzymatic ligation is the angle at which the ligation partners approach one another. The requirement for non-enzymatic cleavage is that the attacking and leaving groups should occupy axial positions around the central phosphorus atom, equating to a O2’-P-O5’ angle of approximately 150 - 180°. As non-enzymatic ligation is thought to occur as a reverse of the cleavage reaction, proceeding through the same 5-membered transition state, it should follow that the 150 - 180° angle of incidence is also an important factor here. The attacking group, in this case, would be O5’ of the 5’-terminal U residue of M2-RNA and the leaving group may be either O2’ (resulting in 3’,5’-linked product) or O3’ (resulting in 2’,5’-linked product) of the 3’-terminal C residue of HIV-1 RNA. The angle of approach between the attacking and leaving groups was measured over the course of 100ns molecular modeling run. Assuming a distance of 3 - 5 Å and an angle of 150 - 180° to be necessary for spontaneous ligation to occur, pairwise comparison of the possible opportunities for ligation, forming either the natural 3’,5’-linked product (Fig. 8) or 2’,5’-linked product (Fig. 9), was performed for each model. Once again, it is apparent that all four models sample in this reactive region, both for formation of the natural 3’,5’-linked product (Fig. 8) and 2’,5’-linked product (Fig. 9). However, it appears from the 100 ns trajectories that in general the reactive region of the 2’,5’-linked product is more frequently visited relative to the 3’,5’-linked product (see Discussion below).

2.10 G3 products

G3 group is a mix of five different products. The most stable secondary structure of the Pr-21 is shown in Fig. 3 D. It is similar to that found for Pr-1 in Fig. 3 A, but in the case of Pr-21, the 3’ dangling end is 5 nucleotides long, and the 5’ end is paired with the HIV-RNA strand. Pr-24 probably has a similar structure, but its 3’ dangling end is 3 nucleotides long. Pr-22 may be just an PCR artifact, considering that it was found 8 times in the control series versus 7 times in three other series. Pr-23 was formed just by forward and reverse primers without any odd nucleotides in between, is presented by 26 colonies in series F and is not found in any other series. Probably it is formed under PCR template deficient conditions.

2.11 G4 products

These products appear as a result of ligation with the 5’- end of the initial M2 RNA, and require the presence of 5’-phosphate for recombination to occur. Although the statistics of occurrence of these products is very poor (5 colonies in case of Pr-26 and 1 colony for each of the remaining ones), it can be suggested that in this case another activated group, 5’-triphosphate,
Staroseletz Yaroslav

is involved in the reaction, instead of 2',3'-cyclophosphate; this provides an alternative pathway of recombination reaction in the current experiments.

### 2.12 G5 products

Group G5 combines all other products characterized by rarity of occurrence on the one hand and occurrence in only one experimental series on the other. Therefore it is difficult to make firm conclusions on the analysis of this group of products. For example Pr-39, Pr-43 and Pr-50 were found only in A series (in the presence of ON), but it is difficult to rationalise how the template directs formation of these products. Pr-44 - Pr-49 were detected only in the F series, but it is reasonable to suppose that they can be formed in the other series.

RNAFold analysis of G5 products showed great diversity of their secondary structures. The structures of some products look quite reasonable; for example, ligation occurs in 1- (Pr-40) and 3-nts (Pr-50) bulges or internal loops (Pr-35, Pr-37). The common secondary structures and the structures of the ligation sites of these molecules are similar to ones of abundant products from other groups and thus appearance of such products is not surprising. The appearance of many other G5 products especially long ones like Pr-38 and Pr-42 is more difficult to explain in the light of their secondary structures. Ligation occurs with 1- (Pr-38) or 3-nts (Pr-42) 3'-dangling end between two helical stems and one can imagine a large degree of freedom of ligation partners (Fig. 3 E). On the one hand, this structure does not preclude ligation; however, the ligation partners are out of effective distance for interaction almost all the time. However distant tertiary interactions could influence this situation considerably.

### 2.13 RNA cleavage patterns

To confirm the hypothesis of preferable synthesis of certain products due to the primary cleavage of RNA substrates in a limited number of sites, we studied the kinetics of both HIV and M2 RNA cleavage in the presence of magnesium ions under conditions reproducing the non-enzymatic recombination reaction, with the only exclusion that the two RNAs were treated separately. Both RNAs were labeled with 5'-32P-pCp and incubated in the presence and in the absence of the ON-template for 0, 2, 5, 24, 48 and 72 h. Mixtures of cleavage products were separated in 15% denaturing PAAG. Autoradiogram of the gel is shown in Supplemental Fig. S3. The figure shows that both RNAs have their own prominent "fragile" sites, and the cleavage depth at those sites rises as the incubation progresses. Such sites and their location in RNAs (counting both from 3' and 5'-ends) are shown on the sides of Supplemental Fig. S3. It is also remarkable that cleavage at several sites occurs mostly or only in the presence of ON-template: after nucleotides 33-43 and 45-63 in HIV-RNA and after 41-47 in M2-RNA. All fragile sites
Staroseletz Yaroslav

were juxtaposed with secondary structures of RNAs, and the locations of those areas are shown in Supplemental Fig. S4.

As shown in Supplemental Fig. S4, the upper stem of HIV-RNA, formed by nucleotides 24-60, is subjected to cleavage only when is partially unfold by the ON-template. Given that the template is absent from the reaction mixture (series B in Fig. 2), the stem remains almost intact and can lead to formation of products referring to group G2. Minor stems (formed by nucleotides 62-72 and 73-88) are cleaved under any conditions, and it supports the data provided in [39]. M2-RNA contains fewer numbers of template-sensitive "fragile" sites than HIV-RNA, they are located in the area 64-51 nts and most probably undergo cleavage due to unfolding of 54-71 nts stem by ON-template. Other sites of cleavage are located mostly in single-stranded regions of RNA, susceptible to rupture and generally coinciding with data demonstrated earlier [40] for the same RNA fragment, cleaved by artificial nucleases.

3 Discussion

3.1 RNA ligation in loops.

Recombination is a prebiotically plausible route to the emergence of a repertoire of functional RNA species. Whether it was the principal route is debatable but it is likely that it contributed to the development of the RNA world. In this work, we aimed to explore the free recombination process by modelling it with a simple system which included two RNA fragments and Mg$^{2+}$ as a catalyst. Here we discuss the observed features of this process.

There are several factors influencing the range of recombinant products formed. The most important factor is the selection of “building blocks” or RNA modules, capable of self-assembly. Predisposition to self-assembly is an inherent feature of RNA, enabling it to form multi-motif functional complexes (ribozymes). Self-assembly of a Group I intron from four inactive oligonucleotide fragments through RNA-directed recombination events was demonstrated in [27,41].

Selection of modules (“building blocks”) for the creation of recombinant products listed in Table 1 occurs both at cleavage and ligation steps of recombination reaction. As we have shown previously [34] and demonstrate in the current work (Supplemental Fig. S3 and S4), location of “fragile sites” is the major criterion for selection of building blocks at the cleavage step. However, an ease of dissociation of RNA fragments from each other and competition between fragments for the binding site determines availability of this blocks for the ligation reaction. For example, M2 fragment in Fig. 3 C competes with “leaving” HIV fragment and “wins” due to formation of longer duplex. When available, blocks are selected again, and those which form the most stable intermediate complex due to formation of a duplex with a template
or some kind of intramolecular duplex will benefit. The last selection stage is the ability to ligate fragments by placing the reacting groups (2',3'-cyclophosphate and 5'-OH) to achieve the optimal orientation, so-called in-line conformation [42]. Eventually, the success of RNA fragments in proceeding through all of these stages determines the relative yield of the various recombination products.

Disposition of reacting groups in the optimal orientation depends on an RNA structure. As we have shown in our previous work [34] and demonstrate in the current study, RNA ligation occurs mostly in the internal loops, bulge-loops and the end of stem structures, turned into hairpin upon ligation. Different kinds of loops are inherent structural motifs of almost all known long RNAs, and their presence gives a definite advantage to RNA over DNA in terms of ability to possess diverse catalytic functions. Loops, stem-loops and pseudo knots are critical elements for every known complex ribozyme [43–45], and certainly would be featured in a self-replicating ribozyme. Therefore, it is possible that under prebiotic conditions, RNA ligation resulted in the appearance of stem-loop structures capable of further ligation and rearrangement with formation of catalytically active RNA molecules.

Non-enzymatic ligation in RNA loops was discussed in the works of David Bartel [46]. Seven families of RNA ligases were isolated from a pool of random RNA sequences. The authors have reported the generation of 3 classes of ribozymes, in accordance with the ligation mechanism involved in their formation. Formation of class II ligases was the closest to the one we have demonstrated in the current study. A surprising feature of the class II catalytic motif is that the continuous base pairing across the ligation junction is replaced by an internal loop involving the ligator and these loops were formed by tandem GA mismatches, which were shown earlier to be unexpectedly strong stabilizers of loop regions within Watson-Crick helices [47]. Another interesting feature is that the 3’ OH of the ligating fragment was 60 times more active in GA tandem mismatch bearing loops in terms of formation of a new phosphodiester bond than 2’-OH, although the 2’-hydroxyl is well known to be at least 6 times more reactive than 3’-OH toward activated phosphate esters in an unconstrained aqueous environment [48,49]. However, exclusion of a hairpin structure located in the ligator 2-8 nucleotides away from the internal loop in question resulted in a rapid increase of rates of 2’,5’ linkage formation over the initial levels [46]. Hence the prevalence of 2’,5’ or 3’,5’ bond formation upon non-enzymatic ligation in loops does not depend entirely on the type, size and nucleotide composition of a loop. Reciprocal orientation of ligating groups is highly influenced by the interactions with other motifs of the same or other RNAs, and therefore general universal predictions on type of the bond seem to be senseless without a consideration of specific spatial setting of every single
structure. In the current study, we did not aim to investigate the type of newly formed phosphodiester bond experimentally, because our product detection scheme was based on the analysis of sequences of RT-PCR products, while the information about the linkage type is lost in transition from RNA to DNA. However, this information was evaluated via molecular dynamics simulation, albeit for only one product. Based on the structure of product Pr-13 (Fig. 3 C), which is 53nt in length, we also created 3 shortened models, 38-, 33-, 27-nts long (Fig. 4 and 6). Using a simultaneous co-incidence of a O2’-P-O5’ bond angle of 150 - 180˚ and a distance between reacting atoms of 3 – 5 Å as requirements for spontaneous ligation to occur, we found that 2’,5’- linked products are generally predicted as favored, based on the sampling of these 100 ns simulations (Fig. 8, 9). The 33nt model showed the strongest preference for 2’,5’-linked products over 3’,5’ ones whereas 53nt and 27nt models demonstrated only a slight preference for 2’,5’-linked products. The 38nt model in contrast is favorable for 3’,5’- linked products to be formed (Fig. 8, 9). Taking into account the identity in sequences at the ligation site of the four models, it appears that distal regions of the RNA molecule influence the structure of the ligation site.

Another feature of spatial architecture in ligation sites, discussed in [46], was less relevant to the current work, but nevertheless is notable. Class I ligases required fixing one ligation substrate in a fully complementary complex with a distal part of a ribozyme, while the second substrate partner had four nucleotide long dangling end, flexible enough to access the first substrate and to accomplish ligation reaction [46]. These data correspond well with the mechanism of formation of products of G2 group (see Fig. 3 C), where the major product, Pr-13, is formed in a similar complex, with the only difference that the 5’ rather than 3’ end is fixed in a Watson-Crick helix, and the dangling end is 3 nts long.

Ligation in terminal loops was demonstrated in direct experiments in [50] using a model system consisting of two synthetic 24mers ribooligonucleotides G24 and C24 and variants thereof. The authors showed that incubation of 5′P-G24-3′OH and 5′OH-C24-3′OH in water results in formation of two products: C24G24 and C24G. 3′OH group of C24 performs nucleophilic attack at either terminal 5’-phosphate of G24 (ligation) or internucleotide phosphate 5′GpG23-3′ (recombination). The second reaction is not similar to recombination process described here because it involves only one transesterification act without intermediate cyclic phosphate formation whereas in our model recombination involves two independent transesterification acts namely cleavage leading to the formation of cyclic phosphate and ligation when cyclic phosphate acts as activated phosphate form. The major difference between the models is the use of 5’-phosphate (in contrast to 2’,3’-cyclophosphate...
or 5′-ppp). The point is that 5′-phosphate is not activated form of phosphate so the ligation reaction in model system described in [50] is not energy neutral.

Authors assume that ligation occurs in 5-nts loop that is consistent with our results. However it is rather a suggestion based on computations then an established fact and we are entitled to suppose that ligation in this model system may also occur for instance in 4-nts or shorter loops.

The question whether 3′-OH group or 2′-OH acts as nucleophile is also beyond the scope of Italian’ group paper [49] though the authors assume 3′-OH group as nucleophile by default. As it was shown earlier [46,51] and here this question couldn’t be solved in general terms. Each structure provides benefits for the choice of either 3′-OH or 2′-OH or both in the certain ratio.

Non-enzymatic recombination attracted the attention of progenitors of the RNA world hypothesis after the finding that this kind of RNA rearrangement is a critical step in life cycles of a number of viroids [52,53]. Even bacterial tRNAs that contain self-splicing group I introns have their sequences interrupted at the same anticodonic location, giving strong evidence to the notion that tRNAs originated in several pieces and needed to be assembled through recombination events [15].

Another stimulus that provoked the study in this direction was the gradual understanding that ribozyme with replication activity will unlikely be found in nature and that, in turn, led to the conclusion that long catalytically active RNA molecules (ribozymes) had to evolve from short RNA fragments through the reactions of cleavage and intermolecular ligation (named collectively “recombination”) [28].

There is one especially attractive feature of recombination: it can be performed in an energy-neutral fashion, when the number and type of chemical bonds are unchanged in the reaction. Hence only a slight perturbation of the environment is enough to disturb the equilibrium and to drive the reaction substantially in either direction [15]. Certainly there must be local and global environmental conditions, i.e. divalent metal ions, temperature, pH or surfaces, which could force this reaction back or forward.

In spite of the high current interest in ribozymes, not many studies were done for the investigation of features on non-enzymatic ligation. Due to low yields of these reactions, model systems described in literature are limited to the use of polyN substrates or templates [54], short and chimeric oligonucleotides [55], use of templates and preformed substrate/template duplexes. Intuitively, these limitations are used to guide the reaction to the synthesis of one major product. However, as we have shown in this work, the information is lost when these kinds of restrictions
are applied. Given the opportunity to evolve in any energetically allowed direction, reaction of RNA recombination yields a number of surprising products whose formation was not predicted and therefore was inhibited in the previous restricted experimental schemes. Specifically, we speculate that ligation of RNAs in loops is more preferable than within fully complementary complexes with a template.

3.2 Similarities between spontaneous recombination and ribozyme’s cleavage and ligation reactions.

We have discussed RNA ligase activity only \textit{in cis} conformational arrangements. However we cannot exclude that some of the recombinant products are formed as a result of \textit{in trans} action of some catalytically active RNA elements spontaneously formed in a pool of RNA upon incubation.

The possibility of evolving complex and highly active RNA ligases has been documented \textit{in vitro}, using as the starting material pools of random RNA sequences [46], natural ribozymes such as group I introns [56], or previously evolved RNA ligases [57]. Those experiments have shown that the ligation and exchange of RNA structural domains through recombination can be used to engineer new functional RNAs [28,58,59].

The probability of hairpin ribozyme-like structures presence in a pool of random RNA fragments (40 nts length) was shown to be relatively high and its probable participation in recombination event should be taken into account. It was hypothesized that stable hairpin structures could have been endowed with a certain degree of primordial RNA ligase activity [28,35].

In the early 1990s it was shown that group I ribozymes in Tetrahymena and sunY broken into 3 parts can assemble in \textit{in trans} complexes without a covalent contiguous RNA chain [60–63]. Such trans complexes could perform ligation and splicing reactions on exogenous RNA substrates and could thereby promote the assembly of fragments complementary to the ribozymes themselves [41,60,61].

Computer simulations performed by Manrubia et al. show that upon hairpin-catalyzed ligation of two preformed RNA modules, up to 1% of the assembled molecules retain the previous modular structure of their building blocks [64]. That is what we observe in case of G2 group of products (see Fig. 3 C) that possess unchanged structural fragment of HIV and therefore are abundant in the reaction mixture under any reaction conditions.

3.3 Cleavage and ligation rates.

We have discussed above that recombination reactions that we performed in our study consist of two major steps: cleavage and ligation. Although they are considered as reciprocal
reactions in chemical terms, and theoretically equilibrium between them should be achieved, this
does not happen in the reaction system we use. The reason is that RNAs have multiple cleavage
and ligation sites, and cleavage at one site may be followed by ligation at the other. Generally
speaking, products of ligation reaction are different from the substrates of cleavage reaction.
Hence equilibrium is never attained, and ongoing cleavage/ligation events yield in continuous
enriching the pool of RNA with new molecules. RNA recombination is a dynamic process, and
any complementary complex can undergo dissociation, leading to fragments’ association with
new RNA molecules and so forth. Release of RNA from complex with template or with another
RNA molecule has two effects. Firstly, secondary structure of RNA is being changed and thus
location of “fragile” sites changes as well. Secondly, single-stranded RNA becomes very
sensitive to cleaving conditions (i.e. presence of divalent metal ions and high pH). Therefore
yields of recombination products depend not only on the ease of formation of ligation substrates
and the probability of their convergence, but also on the stability of double-stranded complexes
that preserve them from rapid degradation. Stability of these complexes, however, may also
inhibit ligation reaction as a result of a need for diffusion of the ligated product off the template
[41]. However, some RNAs ligating in cis possess an ability to adapt to the “structural”
conditions of reaction and perform self-ligation in two alternative complexes. It was shown that
the use of simple self-replicating systems of the form A + B → T, where A and B are the
substrates that bind to a complementary template T, and become joined to form a molecule that
is identical to the template [65]. Selected substrates A and B were designed in such a way that
they could ligate with or without a template. Adjustment was made by varying the length of
complementary stem formed between each of RNA substrates and RNA template. Evidently,
conditions can be found when one or another mechanism will be preferable [65].

Similar observations were made for Pr-1 in our study. It was demonstrated to be formed
at relatively high yields (in comparison to other products) both in the presence and in the absence
of template. Surprisingly, yields were almost the same under different conditions (26% of all
clones in series A and 28% in series B). Hence, fragments forming these products adjust to the
reaction conditions. Although the exact structures of intermediates remain unknown, a
suggestion can be made that even in the presence of template self-ligation (i.e. template
independent ligation) is preferable (see Fig.4 A and B for suggested intermediate structures).

As we have highlighted earlier, direction of reaction is determined by some general
reaction conditions, i.e. temperature, pH, metal ion (catalyst), etc. The raise of pH (from 7.0 to
9.5) leads to the increase of yield of products of cleavage/ligation reaction, but the degradation of
RNA at alkaline pH is also increased [31]. As for temperature formation of long RNA molecules was registered within 20-37°C [31].

Due to complicacy of the elaborated system, we did not expect obtaining of high yields of products in the current study. That determined the choice of indirect detection method which, in turn, made the yields impossible to be measured 37°C, pH 8.0 for 72 h. Reaction conditions chosen (37°C, pH 8.0 for 72 h) were suggested to ensure that both cleavage and ligation reaction can take place in the system, that dissociation of relatively short complexes is possible and that RNA substrates are not totally degraded during the reaction.

3.4 Similarities between spontaneous recombination and splicing.

As we have highlighted above, numerous “checkpoints” necessary for completion of ligation yield in the presence of major and minor products in the reaction mixture after 72 h of incubation. Obviously, major products are of everyone’s primary interest as long as they reflect the most important pathways that take place in the reaction. Nevertheless some minor products seem very curious and inspire us for continuing research.

In the RNA pool, there are several groups of products with (possibly) similar secondary structures that differ in several nucleotides at the ligation site (see Table 1). In addition to the hypothesis proposed earlier (formation of RNA fragments of slightly different lengths at the cleavage stage), a new suggestion about the possible inclusion of splicing events in the formation of shorter products from longer ones can be made now.

33 products of 50 detected in the pool of RNAs after recombination reaction were referred to the group G5 that unites all minor ligation products with undefined intermediate complexes. However, two products, Pr-47 and Pr-28, deserve a special attention. When compared to the sequences of products Pr-3 and Pr-25 respectively, one can see that the first two products differ from the latest two by the presence of one (Pr-47) or two (Pr-28) guanines at the ligation site, right between the HIV and M2 sections. Noteworthy, 5’ terminus of M2 is not cleaved prior to ligation in this case, i.e. M2 fragment ligates with its intact 5’-end and odd G nucleotides do not originate from 5’ end of M2. However, they are not present in the HIV sequence as well, thus proving that two ligation events occurred in these products, and exterior nucleotides were inserted as linkers from the outside. Guanines are known to play a key role in Group I Intron splicing, where they facilitate degradation of exon-intron phosphodiester bond and thus relieve 3’-OH group of exon to attack the bond between intron-Exon II [66]. Although that mechanism does not imply presence of guanine in a processed RNA, the “mediator” role that guanine plays in ligation can be relevant to our study. It should be noted that all four discussed products are minor and were detected only in one bacterial colony each. Hence, no
conclusions about preferences of formation and relative yields can be made with this poor statistics. Nevertheless, there is one more interesting point to be discussed in relevance to these products. Pr-3 and Pr-25 (without G) are formed in A series only (template-dependent recombination). Pr-47 and Pr-28 (with G inserts) are detected only in F series which is essentially a control group with no incubation under recombination conditions. However all these products are formed only in reactions where 5’-ppp-M2 RNA was used. The first suggestion is that they are formed due to a completely different mechanism where not 2’3’-cyclophosphate, but 5’-triphosphate acts as an activated form of phosphate that ensures the energy supply for transesterification reaction. This group was shown to be able to perform this role in a number of works [67–69]. The second suggestion is that products with G inserts actually appear during reverse transcription reaction performed at 42 C. If this is the case, 5’-triphosphate might attack 3’-OH group of free dGTP (that is present at high concentrations in reverse transcription reaction) and thus extend the length of M2 RNA in one deoxyribonucleotide, with 5’-triphosphate at the end. This pathway can be undergone several times, as it probably happens in case of Pr-28 (GG insert). Finally, 5’-triphosphate of terminal guanine of extended M2 fragment attacks relevant site of HIV RNA (which is ironically a “fragile” site as one can see from Supplemental Fig. S3) and completes recombination mission. Again, this happens only in F series (non-incubated mixture), not in A series (incubated for 72 h). Possible explanation for that can be found in the fact that in recombination reaction presence of free GTP is very unlikely, and the only option for M2 RNA is to attack HIV RNA via with its nascent 5’-triphosphate. Those M2 RNA fragments that did not choose this route will most probably lost their terminal 5’-triphosphate as a result of hydrolysis or cleavage of one or several nucleotides at 5’ termini, resulting in the loss of ability to accept dGTP in reverse transcription reaction.

Formation of non-specific RNA recombination products in reverse transcription reaction is obviously limited to those described above, as far as 2’, 3’-cyclophosphates can accept no more than one dNTP due to the lack of 2’-OH in dNTPs, and 5’-triphosphates, as shown above, are present in the mixture in negligible quantities. However, the described pathways, if true, can be suggested as a demonstration of first steps to the development of elementary splicing machineries, which were supported by enzymes later and finally became ubiquitous in the modern biological world.

4 Conclusion

We investigated the recombination process between two 96 nucleotides (nts) RNA molecules under certain conditions (pH 8.0, Mg^{2+}, 37˚C, 3 days incubation) and found several
remarkable regularities, namely, preference for formation of a few types of major product with a large variety of minor products, switching reaction routes when adding such components as 5’ppp-RNAs and ON-template, and preference for formation of 2’-5’- rather than 3’-5’-linkages. We are sure that it is a very important but a tiny step on a really uncharted territory of non-enzymatic RNA recombination.

The understanding of the molecular mechanisms of non-enzymatic recombination of RNAs may not only explain some aspects of the origin of life, but also provide a vital clarification of enormous structural and functional diversity of RNAs. More importantly, recombination involving RNA sequence exchange and rearrangements between and within RNA molecules are believed to contribute to the evolution and diversity of RNA viruses and might be responsible for the production of defective interfering RNAs. The insight into the fundamental principles of molecular interactions and regulatory mechanisms of RNA recombination and the understanding at the molecular level of the key factors promoting these reactions may shed the light on the origin of rapid, uncontrollable mutations of RNA viruses, which may facilitate production of new viral strains.

5 Materials and methods

5.1 Enzymes and chemicals

[γ-32P]-ATP and 32pCp (specific activity >3000 Ci/mmol) were from Biosan Co. (Russia). Enzymes: calf intestinal alkaline phosphatase and endonuclease Fok I were purchased from “SibEnzyme”, Russia. T7 RNA polymerase, Taq DNA polymerase, M-MuLV DNA polymerase were produced in this Institute. T4 polynucleotide kinase and T4 RNA Ligase were from Fermentas. TA cloning was proceeded using “InstAclone™ PCR Cloning Kit” (Fermentas). Plasmid DNA was isolated with “GenElute™ Plasmid Miniprep Kit” (Sigma). BigDye v.1.1 sequencing mixture was from Applied Biosystems. All the chemicals were of analytical or ACS grade, and autoclaved Milli-Q™ water was used for all procedures. M2-RNA of influenza virus was synthesized by T7 transcription from plasmid pSVK3M2, this Institute collection. HIV-1 fragment was obtained via T7 transcription of plasmid pHIV-2, kindly provided by Prof. H.J. Gross (University of Wurzburg, Germany).

5.2 Oligonucleotides

Oligonucleotides were synthesized using standard automated solid-phase methods and purified by reverse-phase liquid chromatography in this Institute. Primers were as follows: M2-96for – ACAAGCTTTAATACGACTCAGTG (T7 promoter region is underlined), M2-96rev – CGAGACAAATGACTGTCGTCAGC, Mfor –
CTACGGAAGGAGTACCTGA, \( M_{\text{rev}} \) – ATGACTGTCGTCAGCATC, \( H_{\text{for}} \) –
GAGATCCCTCAGACCACT, M13/pUC_for – GTAAAACGACGGCCAGT, M13/pUC_rev –
CAGGAAACAGCTATGAC. ON-template – CGATC
CAGACGGCGGTAGCTCAGA (overhangs are underlined).

5.3 5′- and 3′-labeling

Primer \( M_{\text{rev}} \) and ON-template were 5′-end labeled with \(^{32}\)P using T4 polynucleotide
kinase and \([\gamma^{–32}\text{P}]-\text{ATP}\) according to the manufacturer’s manual as described previously [34].
Radioactive RNAs were prepared by labeling with \([5′^{–32}\text{P}]\text{Cytidine 3′,5′-bisphosphate (}^{32}\text{pCp})\)
using T4 RNA ligase according to the manufacturer’s manual as described previously [70].
Labeled oligonucleotides and RNAs were purified in 8% denaturing polyacrylamide gel
dPAAG).

5.4 Plasmid linearization

Plasmid pHIV-2 was linearized with Fok I endonuclease according to the manufacturer’s
manual. After extraction with TE-phenol-chloroform (1:1) and chloroform-isoamyl alcohol
(24:1) DNA from the aqueous phase was precipitated with ethanol and dissolved in water.

5.5 Amplification of M2-DNA fragment

M2-DNA fragment was amplified from the plasmid pSVK3M2 using primers M2-96\(_{\text{for}}\)
and M2-96\(_{\text{rev}}\) as described in [33]. Amplification product (121 bp) was purified by phenol
extraction (TE-phenol : chloroform (1:1)) and precipitated from the aqueous phase with ethanol.
Dissolved in water precipitates were used as templates for T7 transcription.

5.6 T7 transcription

T7 transcription for the synthesis of M2-RNA and HIV-RNA from M2-DNA (obtained
by amplification of pSVK3M2 plasmid fragment) and linearized plasmid pHIV-2, respectively,
was performed as described in [33]. Both RNA transcripts were purification in 8% native PAAG
and were eluted from the gel with 3 mL of elution solution (0.6 M \( \text{NH}_4\text{Ac} \) (pH 5.0), 0.1 mM
EDTA, 0.1% SDS) on a shaker at 4°C for 4 h. Eluted RNA was precipitated with ethanol and
dissolved in water.

5.7 Dephosphorylation of RNA

A reaction mixture (total volume 50 \( \mu\text{L} \)) containing 8 \( \mu\text{g} \) of M2-RNA or HIV-RNA, 50
mM bis(tris(oxymethyl)methylamino)propane HCl buffer (pH 8.0), 2% formamide, 0.2% SDS,
1.25 mM DTT and 1.5 u of Calf intestinal alkaline phosphatase (CIP) was incubated at 37°C for
45 minutes. Then 1.5 u of CIP was added again to the reaction mixture and it was additionally
incubated under the same conditions. After incubation, mixture was subjected to twofold
extraction with phenol- chloroform (1:1), then chloroform-isoamyl alcohol (24:1). RNA was
precipitated from the aqueous phase by 95% ethanol, washed twice with 75% ethanol, dissolved in water and stored at -20°C.

5.8 Non-enzymatic cleavage/ligation reaction

Reaction mixtures (total volumes 30 µL) containing bis(tris(oxy)methyl)methylamino)propane HCl buffer (pH 8.0), 0.5 µM dephosphorylated or 5'-ppp-bearing M2-RNA and HIV-RNA, 1.0 µM of the ON-template (in template-directed reactions) and 5 mM MgCl₂, were incubated at 37°C for 3 days. After incubation, EDTA was added to the ligation mixtures in a concentration equal to that of MgCl₂. Volumes of all probes were adjusted to 100 µL with water, and RNA was precipitated with ethanol in the presence of 0.3 M NaAc and 10 µg of glycogen. Precipitates were washed with ethanol, dissolved in 15 µL of water and used in reverse transcription reaction.

5.9 Reverse transcription and PCR of ligation products

The DNA template for PCR amplification was produced from the ligation mixture of recombinant RNA in reverse transcription reaction. A reaction mixture with an initial volume of 12 µL, containing 1.25 µM of specific reverse primer and 1 µL of mixture of products of the cleavage/ligation reaction, was subjected to denaturation at 70°C for 5 min and then chilled on ice for 3 min. After that NTPs of a final concentration 1 mM and M-MuLV buffer were added, volume was adjusted to 18 µL with water, and mixtures were incubated at 37°C for 5 min. Finally, 20 u of M-MuLV polymerase were added and the mixture of a final volume 20 µL was incubated at 42°C for 60 min. The reaction was terminated by heating at 70°C for 10 min. Obtained cDNA was directly used in PCR. PCR was performed from reverse transcription (RT) products using primers H_for and M_rev. In a positive control (to confirm M2-RNA presence in the reaction mixture) primers M_for and M_rev were used. PCR mixtures contained 67 mM Tris-HCl (pH 8.0), 16 mM (NH₄)₂SO₄, 1.5 mM MgCl₂, 0.01% Tween-20, 0.25 mM dNTPs, 0.8 µM each PCR primer, 1 µL of RT-product in dilutions 1:1, 1:10³, 1:10⁶ and 2 u of Taq DNA polymerase in a total volume of 20 µL. PCR was started with 3 min of preincubation at 95°C and followed by 25 cycles of amplification (1 min, 94°C; 1 min, 56°C; 1 min, 72°C). Final cycle contained prolonged elongation step (72°C, 10 min) for complete formation of sticky A-ends. PCR products were used directly for TA cloning. In parallel, we run PCR with 5'-³²P-labeled M_rev primers for visualization purposes. In this case aliquots of PCR products were mixed with loading buffer and applied on 10% dPAAG electrophoresis. Visualization was accomplished by exposing dried gels on FX Imaging Screen, followed by screen scanning on Molecular Imager FX-PRO Plus (Bio-Rad) and images processing in Quantity One v.4.2.3 Software (Bio-Rad). Amplification products were identified by electrophoresis on 10% dPAAG using DNA ladder.
obtained via chemical cleavage of M2-80 radioactive PCR product (synthesized using primers M_for and 5'-32P-M_rev, product length 80 bp) at adenine and guanine sites. Reaction mixture of a volume 30 µL, containing 10 µL of 5'-32P-labeled M2-80, and 20 µL of 3% diphenylamine solution in formic acid, was incubated at 37°C for 5 and 10 minutes. At the end of incubation, 100 µL of water was added and reaction mixtures were subjected to twofold extraction with 300 µL of diethyl ether. DNA from the aqueous phase was precipitated with 10 volumes of 2% LiClO₄ solution in acetone, pellets were washed with acetone, dissolved in water and then added to the loading solution.

5.10 TA cloning

The PCR products were cloned using the "InsTAclone™ PCR Cloning Kit" (Fermentas), according to the manual provided by the supplier similar to [33].

5.11 Bacterial colony PCR

Fragments of individual bacterial colonies were dissolved in 20 µL of PCR master mix, containing 67 mM Tris-HCl (pH 8.0), 16 mM (NH₄)₂SO₄, 1.5 mM MgCl₂, 0.01% Tween-20, 0.2 mM dNTPs, 0.3 µM of primers M13/pUC_for and M13/pUC_rev, and 2 u of Taq DNA polymerase. PCR was started with 3 min of preincubation at 95°C and followed by 27 cycles of amplification (30 s, 94°C; 30 s, 51°C; 1 min, 72°C). PCR products were analyzed in 8% dPAAG and visualized by staining of gels with ethidium bromide.

5.12 DNA sequencing

Sequencing of constructs possessing the DNA inserts was performed via Sanger sequencing reaction with mixture BigDye v.1.1 (Applied Biosystems) followed by further analysis of products at automatic gel analyzer ABI 3130xl (Applied Biosystems). Detailed protocol is published in [33]. Sequences were determined in Sequence Scanner Software v. 1.0 (Applied Biosystems) and aligned with initial RNA sequences using Vector NTI v.10.0.1 Software (Invitrogen).

5.13 Molecular Modelling

The loop region of each of the selected models was based on the 1.92 Å resolution NMR structure of the HIV-1 A-rich hairpin loop [38] (PDB code 1BVJ). Nucleotides G156 – C172 of the original experimental structure correspond to nucleotides G18 – C34 of our full-length 53-mer model. The extended stem region of the 53nt model was created by manipulation of a standard double helix, generated via the make-na web server [71]. Nucleotides were then excised as appropriate from the stem opposite the predicted single-stranded regions and the single nucleotide bulge (C38); this bulge was modelled as a mirror image of the bulged nucleotide (C158) found in the experimental structure [38]. The 38nt, 33nt and 27nt models were
constructed via removal of the appropriate number of nucleotides from the stem region of the original 53nt model. As much of the experimentally derived geometry as possible was retained at each stage. Models were constructed and manipulated using Yasara [72].

The 3’ terminal cytidine 2’,3’-cyclic phosphate residue of the HIV-1 recombination partner was modelled from based on the geometry of a standard cytidine residue, using Yasara [72]. Geometry optimisation was performed and electrostatic potentials calculated at the HF/6-31G* level, using Gaussian 09 [73]. Restrained electrostatic potential (RESP) charges were fitted using the antechamber module of AmberTools 13 [74] and the general AMBER force field (GAFF) [75] used for non-electrostatic parameters.

Simulations were carried out using the ff99 force field with bsc0 [76] and XOl3 [77] modifications for RNA. A box of TIP3P water molecules [78], extending 12 Å in all directions beyond the solute, was used. Potassium ions (radius of 1.705 Å and well depth of 0.19368 kcal/mol) and chloride ions (radius of 2.513 Å and well depth of 0.03559 kcal/mol) [79] were added, to a final K⁺ concentration of 200 mM.

To equilibrate each system, energy minimisation was followed by 20 ps MD in the NVT ensemble, to heat the system to 100 K, with subsequent 200 ps MD in NPT with 25 kcal/(mol.Å²) positional restraints on solute atoms, while the system was heated from 100 to 300 K. Short (100 ps) simulations with gradually decreasing restraints of 20, 10, 5, 2 and 1 kcal/(mol.Å²) applied to the solute were followed by 150 ps with positional restraints of 0.5 kcal/(mol.Å²) and 150 ps unrestrained MD. Production simulations of 100 ns for each system were performed using the pmemd module of AMBER 12 [74]. The particle mesh Ewald (PME) method [80] (was used to treat long-range electrostatic interactions. The SHAKE algorithm [81] was used to constrain covalent bonds involving hydrogen atoms, along with a 2 fs time step. Simulations were carried out in the NPT ensemble, with periodic boundary conditions and a temperature of 300 K maintained using a Langevin thermostat [82].

**Acknowledgements**

This research has been supported by the Russian Science Foundation (Grant No. 14-44-00068) and Presidium of RAS research program ‘Molecular and Cellular Biology’ 6.1.
References


Staroseletz Yaroslav


Staroseletz Yaroslav


### TABLES

**Table 1.** Nucleotide sequences of products of HIV-RNA and M2-RNA recombination in the regions between H<sub>for</sub> and M<sub>rev</sub> PCR primer binding sites.

<table>
<thead>
<tr>
<th>Group</th>
<th>Name</th>
<th>Amp. length</th>
<th>Sequence (between primers)</th>
<th>Number of colonies in experimental series</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>G0</td>
<td>Pr-1</td>
<td>55</td>
<td>[H&lt;sub&gt;FOR&lt;/sub&gt;] CUAGACGGUGAUGCUGUG [M&lt;sub&gt;REV&lt;/sub&gt;]</td>
<td>21</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Pr-2</td>
<td>56</td>
<td>[H&lt;sub&gt;FOR&lt;/sub&gt;] CUAGACGGUGAUGCUGUG [M&lt;sub&gt;REV&lt;/sub&gt;]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr-3</td>
<td>56</td>
<td>[H&lt;sub&gt;FOR&lt;/sub&gt;] CUAGACGGUGAUGCUGUG [M&lt;sub&gt;REV&lt;/sub&gt;]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr-4</td>
<td>58</td>
<td>[H&lt;sub&gt;FOR&lt;/sub&gt;] CUAGACGGUGAUGCUGUG [M&lt;sub&gt;REV&lt;/sub&gt;]</td>
<td>14</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Pr-5</td>
<td>76</td>
<td>[H&lt;sub&gt;FOR&lt;/sub&gt;] [H&lt;sub&gt;FOR&lt;/sub&gt;] CUAGACGGUGAUGCUGUG [M&lt;sub&gt;REV&lt;/sub&gt;]</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>Pr-6</td>
<td>61</td>
<td>[H&lt;sub&gt;FOR&lt;/sub&gt;] CUAGACGGUGAUGCAGAUGCUGUG [M&lt;sub&gt;REV&lt;/sub&gt;]</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr-7</td>
<td>55</td>
<td>[H&lt;sub&gt;FOR&lt;/sub&gt;] CUAGACGGAGAUGCUGUG [M&lt;sub&gt;REV&lt;/sub&gt;]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr-8</td>
<td>54</td>
<td>[H&lt;sub&gt;FOR&lt;/sub&gt;] CUAGACGCUUGAUGCUGUG [M&lt;sub&gt;REV&lt;/sub&gt;]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pr-9</td>
<td>58</td>
<td>[H&lt;sub&gt;FOR&lt;/sub&gt;] CUAGACGGUGAUGCUGUG [M&lt;sub&gt;REV&lt;/sub&gt;]</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Pr-10</td>
<td>104</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG</td>
<td>GAAACACGCAGAGUCUGUG $[M_{REV}]$</td>
<td>1</td>
<td>/A</td>
</tr>
<tr>
<td>Pr-11</td>
<td>105</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG</td>
<td>GGAACACGCAGAGUCUGUG $[M_{REV}]$</td>
<td>1</td>
<td>CGC/-1</td>
</tr>
<tr>
<td>Pr-12</td>
<td>105</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG</td>
<td>GGAACACGCAGAGUCUGUG $[M_{REV}]$</td>
<td>3</td>
<td>/UA</td>
</tr>
<tr>
<td>Pr-13</td>
<td>106</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG</td>
<td>GGAACACGCAGAGUCUGUG $[M_{REV}]$</td>
<td>12 27</td>
<td>CGC/-</td>
</tr>
<tr>
<td>Pr-14</td>
<td>107</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG</td>
<td>GGAACACGCAGAGUCUGUG $[M_{REV}]$</td>
<td>4 2</td>
<td>CGC/A</td>
</tr>
<tr>
<td>Pr-15</td>
<td>108</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG</td>
<td>GGAACACGCAGAGUCUGUG $[M_{REV}]$</td>
<td>5 1</td>
<td>CGCC/A</td>
</tr>
<tr>
<td>Pr-16</td>
<td>109</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG</td>
<td>GGAACACGCAGAGUCUGUG $[M_{REV}]$</td>
<td>10</td>
<td>CGCCCC/A</td>
</tr>
<tr>
<td>Pr-17</td>
<td>110</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG</td>
<td>GGAACACGCAGAGUCUGUG $[M_{REV}]$</td>
<td>1</td>
<td>CGCCCG/A</td>
</tr>
<tr>
<td>Pr-18</td>
<td>105</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG</td>
<td>GGAACACGCAGAGUCUGUG $[M_{REV}]$</td>
<td>4</td>
<td>C/A</td>
</tr>
<tr>
<td>Pr-19</td>
<td>104</td>
<td>$[H_{FOR}]$CUAGACGGUGUAAAAAUCUCUAG</td>
<td>GGAACACGCAGAGUCUGUG $[M_{REV}]$</td>
<td>7 6</td>
<td>C/-</td>
</tr>
<tr>
<td>Pr-20</td>
<td>37</td>
<td>$[H_{FOR}]$G $[M_{REV}]$</td>
<td></td>
<td>4 1</td>
<td>1</td>
</tr>
<tr>
<td>Pr-21</td>
<td>42</td>
<td>$[H_{FOR}]$CUAGAC $[M_{REV}]$</td>
<td></td>
<td>4 6</td>
<td>2</td>
</tr>
</tbody>
</table>

5 nts terminal loop
<table>
<thead>
<tr>
<th>Pr</th>
<th>seq</th>
<th>length</th>
<th>nts</th>
<th>Loops</th>
<th>Additional Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>[H\text{FOR}] AU {M_{REV}}</td>
<td>38</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>[H\text{FOR}] M_{REV}</td>
<td>36</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>[H\text{FOR}] CUA M_{REV}</td>
<td>39</td>
<td>2</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>[H\text{FOR}]CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCCCGCAACAGGGACUUUAAA GUGAAAGUAACAGGGGGCCUUCUACG GAAGGAGUACCUGAGUCAUGAGGGAAGAAGAUACGAAAGGAACAGCA GUG M_{REV}</td>
<td>170</td>
<td>1</td>
<td>8 nts internal loop</td>
<td>Ligation with 5'-M2-RNA end</td>
</tr>
<tr>
<td>26</td>
<td>[H\text{FOR}]CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCCCGCAACAGGGACUUUAAA GUGAAAGUAACAGGGGGCCUUCUACG GAAGGAGUACCUGAGUCAUGAGGGAAGAAGAUACGAAAGGAACAGCA GUG M_{REV}</td>
<td>169</td>
<td>2</td>
<td>7 nts internal loop</td>
<td>Ligation with 5'-M2-RNA end</td>
</tr>
<tr>
<td>27</td>
<td>[H\text{FOR}]CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCCCGCAACAGGGACUUUAAA GUGAAAGUAACAGGGGGCCUUCUACG GAAGGAGUACCUGAGUCAUGAGGGAAGAAGAUACGAAAGGAACAGCA GUG M_{REV}</td>
<td>169</td>
<td>2</td>
<td>8 nts internal loop</td>
<td>Ligation with 5'-M2-RNA end</td>
</tr>
<tr>
<td>28</td>
<td>[H\text{FOR}]CUAGACGGUGUAAAAAUCUCUAG CAGUGGCGCCCGCAACAGGGACUUUAAA GUGAAAGUAACAGGGGGCCUUCUACG GAAGGAGUACCUGAGUCAUGAGGGAAGAAGAUACGAAAGGAACAGCA GUG M_{REV}</td>
<td>172</td>
<td>1</td>
<td>10 nts internal loop</td>
<td>Ligation with 5'-M2-RNA end</td>
</tr>
<tr>
<td>29</td>
<td>[H\text{FOR}] GAU M_{REV}</td>
<td>39</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>[H\text{FOR}] GUG M_{REV}</td>
<td>39</td>
<td>1</td>
<td></td>
<td>1 nt from 3 nts internal loop</td>
</tr>
<tr>
<td>31</td>
<td>[H\text{FOR}] CU GCUGUG M_{REV}</td>
<td>44</td>
<td>4</td>
<td></td>
<td>6 nts terminal loop</td>
</tr>
<tr>
<td>32</td>
<td>[H\text{FOR}] CU AGACCGU M_{REV}</td>
<td>45</td>
<td>1</td>
<td></td>
<td>1 nt bulge</td>
</tr>
<tr>
<td>Pr-33</td>
<td>55</td>
<td>([H_{FOR}] [H_{FOR}] A [M_{REV}])</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----</td>
<td>---------------------------------</td>
<td>----</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Pr-34  | 63 | \([H_{FOR}]CUAGACGGUGUAAAAAACUCUGCUG [M_{REV}]\) | 3 | 1 nt from 6 nts internal loop  
| Pr-35  | 75 | \([H_{FOR}]CUAGACGGUGUAAAAAACUCUAGCAGUGGCGGCCGCUGUG [M_{REV}]\) | 3 | 6 nts assimetric internal loop  
| Pr-36  | 87 | \([H_{FOR}]CUACCUGAGUCUAGAGGGAAGAUAUCGAAAGGAACAGCAGAGCU GUG [M_{REV}]\) | 2 | 5 nts terminal loop  
| Pr-37  | 91 | \([H_{FOR}]CUAGACGGUGUAAAAAACUCUAGCAGUGUACGAAAGGAACAGCAGAGCU GUG [M_{REV}]\) | 1 | 4 nts internal loop  
| Pr-38  | 133 | \([H_{FOR}]CUAGACGGUGUAAAAAACUCUAGCAGUGGCGGCCGCACCAGGACUUUUAG GUGAAAGUAUGAGGAAGAAUAUCGAAAGGAACAGCAGAGGCGUGUG [M_{REV}]\) | 1 | 1 nt crossing of two helical stems  
| Pr-39  | 77 | \([H_{FOR}]CUAGACGGUGUAAAAAACUCUAGCAGUGGCGCCCGAACCAGG [M_{REV}]\) | 1 | 1 nt from 4 nts terminal loop  
| Pr-40  | 73 | \([H_{FOR}]CUAGACGGUGUAAAAAACUCUAGCAGUGGCGCCCGAACCAGG [M_{REV}]\) | 2 | 1 nt bulge  
| Pr-41  | 109 | \([H_{FOR}]CUAGACGGUGUAAAAAACUCUAGCAGUGGCGCCCGAACCAGGACUUUACUC GAGGAACAGCAGAGGACUGUGUGCUGUG [M_{REV}]\) | 1 | AA is substituted by GG in M2 region  
| Pr-42  | 119 | \([H_{FOR}]CUAGACGGUGUAAAAAACUCUAGCAGUGGCGCCCGAACCAGGACUGAGGGAAGAAAUACGAAAGGAACCAGCAGAGGACUGUGU [M_{REV}]\) | 1 | 3 nt crossing of two helical stems  
| Pr-43  | 74 | \([H_{FOR}]CUAGACGGUGUAAAAAACUCUAGCAGACGTTAGUGUCUGUG [M_{REV}]\) | 1 |  
| Pr-44  | 42 | \([H_{FOR}] GCUGUG [M_{REV}]\) | 5 | 4 nts terminal loop  
| Pr-45  | 43 | \([H_{FOR}] CUAGACG [M_{REV}]\) | 4 | 1 nt from 2 nts internal loop  

Staroseletz Yaroslav
<table>
<thead>
<tr>
<th>Primer</th>
<th>3' Tailing Region</th>
<th>5' Tailing Region</th>
<th>Number of Bulges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pr-46</td>
<td>([H_{FOR}][H_{FOR}])G ([M_{REV}])</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Pr-47</td>
<td>([H_{FOR}]CUAGACGGUGUAGGUGCUGUG) ([M_{REV}])</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Pr-48</td>
<td>([H_{FOR}]CUAGACGGUGUAAAAAUCCUCUAG) ([M_{REV}])</td>
<td>Between two 1nt bulges</td>
<td></td>
</tr>
<tr>
<td>Pr-49</td>
<td>([H_{FOR}][H_{FOR}][H_{FOR}])G ([M_{REV}])</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Pr-50</td>
<td>([H_{FOR}]CUAGACGGUGUAAAAAUCCUCUAG) ([M_{REV}])</td>
<td>3 nts RNA bulge</td>
<td></td>
</tr>
</tbody>
</table>

1) Only RNA sequences between regions corresponding to primers H\(_{FOR}\) and M\(_{REV}\) are shown.

2) Number of colonies, detected by bacterial colony PCR and confirmed via nucleotide sequencing of plasmids; correspond to four experimental series: A, B, C, F (see legend for Fig. 2 for keys)

3) RNA regions originating from HIV-RNA are shown in red, from M2-RNA—in blue. Template-binding regions are underlined; sequences corresponding to primers are shown as \([H_{FOR}]\) and \([M_{REV}]\). RNA nucleotides within bulge loops are not underlined, nucleotides in internal loops and mismatches are shown in italic. External insertions are highlighted with yellow.
Table 2. Summary of mean backbone RMSd and average inter-atomic distances (d) between the ligation points (i.e. 2’-P or 3’-P atom of 3’cytidine (HIV-1 RNA) and the O5’ atom of 5’uridine (M2-RNA)) for 53-mer, 38-mer, 33-mer and 27-mer model RNAs.

<table>
<thead>
<tr>
<th>Length of model</th>
<th>RMSd (backbone)</th>
<th>RMSd (loop)</th>
<th>Average inter-atomic distance (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>53nt</td>
<td>5.998 Å ± 1.732</td>
<td>2.240 Å ± 0.392</td>
<td>7.410 Å ± 2.610</td>
</tr>
<tr>
<td>38nt</td>
<td>8.515 Å ± 2.037</td>
<td>1.544 Å ± 0.255</td>
<td>9.930 Å ± 5.576</td>
</tr>
<tr>
<td>33nt</td>
<td>5.708 Å ± 1.410</td>
<td>1.479 Å ± 0.252</td>
<td>6.308 Å ± 2.469</td>
</tr>
<tr>
<td>27nt</td>
<td>4.253 Å ± 0.814</td>
<td>1.606 Å ± 0.291</td>
<td>6.537 Å ± 2.329</td>
</tr>
</tbody>
</table>
**FIGURE LEGENDS**

**Figure 1.** General scheme of RNA recombination pathways in the presence and in the absence of ON-template. Fragments corresponding to HIV-RNA are shown in red, M2-RNA – in blue, ON-template – in black.

**Figure 2.** Products of RT-PCR with 5'-32P-M\textsuperscript{rev} primer, 10% denaturing PAAG. RNAs marked in caption were used in relevant recombination reactions. In “+ template” series recombination was performed in the presence of ON-template, “- template” – in the absence of ON-template. Lanes “1”: positive control PCR with primers M\textsubscript{for}/M\textsubscript{rev} (25 PCR cycles). Lanes “2”, “3”, “4”: PCR with primers H\textsubscript{for}/M\textsubscript{rev}, with cDNA dilutions 1:1, 1:10\textsuperscript{-3}, 1:10\textsuperscript{-6}, respectively (28 PCR cycles). In “E” and “F” series we used non-incubated mixtures of RNA and template for RT-PCR reactions. K1: negative control of RT-PCR with primers H\textsubscript{for}/M\textsubscript{rev}, with reverse transcription performed in the absence of any RNA. K2, K3: negative controls of PCR in the absence of cDNA with primers H\textsubscript{for}/M\textsubscript{rev} and M\textsubscript{for}/M\textsubscript{rev} (25 PCR cycles). L1: dsDNA ladder. Positions of supposed products of template-directed ligation are shown with arrows (X-1 through X-9).

**Figure 3.** The most stable secondary structures of Pr-1 (A and B) (G0) and Pr-21 (D) (group G3). C. Common domain of products Pr-10 through Pr-19 (group G2). N and N’ indicate different nucleotides within ligating RNA dangling ends, n and m – number of nucleotides (see Table 1 for details). HIV-RNA fragment is shown in red, M2-RNA fragment – in blue. Ligating ends are marked as 3’ and 5’. E. Ligation site of Pr-42 (group G5). D. Variants of ligation site of G2 products. Dotted lines connect ligating nucleotides.

**Figure 4.** Two dimensional representations of the proposed structures of: A. HIV-1 A-rich hairpin loop. The nucleotides in red represent the naturally occurring nucleotide bases, those added by the authors to aid transcription are indicated in black. Insertion (Ins) shows the NMR-based 3D structure of HIV-1 RNA A-rich hairpin loop (PDB: 1BVJ, taken from (Puglisi EV and Puglisi JD 1998)). B. Model RNA structures of varying stem length (L-R: 53-mer, 38-mer, 33-mer, 27-mer). The nucleotides in red represent those belonging to the 3’ recombination partner and those in blue, correspond to the 5’ recombination partner. The dashed line encloses the part of the models corresponding to the experimentally derived structure from NMR (shown as insertion in A).

**Figure 5.** 3D structures of the selected RNA models systems 53nt (A), 38nt (B), 33nt (C) and 27nt (D), each representing the average structure from the 100 ns Molecular Dynamics run for the designated models. The red and blue nucleotides highlight the key ligation points between 3’-C of HIV-1 RNA fragment and 5’-U of M2-RNA fragment, respectively. The
Staroseletz Yaroslav

rectangular insertions represent the enlarged 3D structures of the ligation sites showing the minimum distance between the ligation points (i.e. (HIV-1 RNA) 3’C > P ... O5’: U5’ (M2-RNA)), measured over the 100 ns Molecular Dynamics simulations for each model.

**Figure 6.** RMSd over 100ns simulation of the backbone atoms of each of our four model RNAs, from the original structure (red): (A) 53-mer; (B) 38-mer; (C) 33-mer and (D) 27-mer. RMSd of the backbone atoms in the loop portion of each model from the NMR structure are shown in black.

**Figure 7.** Distance between the potential ligation partners taken as the distance d between the 2’- or 3’-P atom of 3’cytidine (HIV-1 RNA) and the O5’ atom of 5’uridine (M2-RNA) in each of the four model RNAs, over 100 ns simulation: (A) 53-mer; (B) 38-mer; (C) 33-mer and (D) 27-mer. The dashed red line represents the distance limit for ligation to occur. For each model RNA, P ... O5’ reached distance within 5Å at multiple points during 100 ns Molecular Dynamics simulations, totalling between 20% and 30% of the total simulation time in each case.

**Figure 8.** Opportunities for ligation, forming a 3’, 5’- linked product. Co-incidence of distance P-O5’ between 3 - 5Å and angle O5’-P-O2’ between 150-180˚, for each of our 4 model RNAs: (A) 53-mer; (B) 38-mer; (C) 33-mer and (D) 27-mer.

**Figure 9.** Opportunities for ligation, forming a 2’, 5’- linked product. Co-incidence of distance P-O5’ between 3 - 5Å and angle O5’-P-O3’ between 150-180˚, for each of our 4 model RNAs: (A) 53-mer; (B) 38-mer; (C) 33-mer and (D) 27-mer.
Figure 1

Route A

5'OH or 5'-pppRNAs + ON-template

1(a) Hybridization (fast)

2(a) Cleavage (moderate)

Assembly

3(a) Ligation

4(a) Cleavage

5(a) Ligation

Route B

- ON-template 5'OH or 5'-pppRNAs

2(b) Cleavage (moderate)

Products of template-dependent RNA recombination (described earlier)

Products of spontaneous RNA recombination (this study)
**Supplemental Fig. S1.** Investigation of RNA/ON-template duplexes formation. Autoradiogram of 12% native PAAG. **A.** Incubation for 2 h at 37˚C in the absence of Mg\(^{2+}\). **B.** Incubation for 72 h at 37˚C in the presence of 10 mM Mg\(^{2+}\). Oligodeoxyribonucleotides and RNAs used in reactions are marked in the capture. Major duplexes are highlighted with letters A, B and C, arrow with "ON" symbol shows a location of free 5‘-32P-ON-template.
**Supplemental Fig. S2.** Representative electrophoresis showing products of RNA recombination in 15 colonies from the group "B". Lanes 1-15: products of bacterial colony PCR, amplified with the use of primers M13/pUC_for and M13/pUC_rev. P – positive control (PCR from circular plasmid without TA-insert), K – negative control (PCR without template), L – double-stranded DNA ladder.
**Supplemental Fig. S3.** Patterns of 5'-32pCp-labelled HIV-RNA and M2-RNA cleavage in the presence of Mg\(^{2+}\) for 0, 2, 5, 42, 48 and 72 hours (see the capture). Autoradiogram of 15% dPAAG. Incubation was performed both in the presence and in the absence of the ON-template. L – imidazole ladder. Location of "fragile" sites is marked with positions of correspondent nucleotides in RNA chain: on the left for HIV-RNA and on the right for M2-RNA. Numbers in the parantheses show the location of the same "fragile" sites counting from 5'-ends.
Supplemental Fig. S4. Location of "fragile" sites in HIV-RNA and M2-RNA. Black arrows show sites susceptible to cleavage under any conditions, red arrows – primarily or solely in the presence of the ON-template. Lengths of arrows reflect the depth of cleavage. Regions complementary to the ON-template are shown in red.
*Conflict of Interest form
Click here to download Conflict of Interest form: coi_disclosure_1.pdf