Reviewers' comments:

Reviewer #1 (Remarks to the Author):

In this paper, Perez-Jimenez and coworkers reconstructed a Precambrian endoglucanase that exhibits higher activity towards soluble CMC, Avice1 and lignocellulosic materials as compared to the modern Thermotoga maritima TmEG and Bacillus subtilis BsEG. They claimed that this ancestrally reconstructed enzyme could be used and further engineered for industrial applications.

Although previous studies have been published related to ancestral enzyme reconstruction with improved thermostability, activity, and/or promiscuity, this paper is definitively of relevance and of interest to other researchers working in the field of endoglucanases and enzyme design.

The authors claim that the resurrected Precambrian endoglucanase is highly efficient, however the reported kcat, KM and kcat/KM values for instance are not that different from the ones obtained for the thermophilic TmEG. The major differences are found in the degradation of cardboard, newspaper, and softwood, although the conversion of LFCA-EG alone is not that different to TmEG and BsEG. The most important finding is that LFCA_EG incorporated into a mini-cellulosome enhanced its activity in contrast to the other two enzymes. Do the authors have an explanation for such differences? I believe the title of the paper could be changed to better represent their findings.

The obtained X-ray structure for the ancestral enzyme is quite similar to the one from the modern enzymes as shown in Figure 5, thus the static structures are not able to explain the difference in activity. The analysis of the conformational dynamics of the enzymes in the presence of tetrasaccharide was therefore performed. The results shown in Figure 6 monitoring the catalytic distances indicate that BsEG enzyme should exhibit lower catalytic efficiencies with respect to the other two enzymes, which is in line with the kcat/KM values reported in Table 1. However, the dnuc distance is substantially shorter in TmEG than in LFCA_EG:
1. It is mentioned that the main difference comes from a long loop that forms a clamp for the substrate. They also mention that "This loop may require large amplitude motions that decrease ligand flexibility". How the conformational dynamics of this loop differ among the different enzymes, but especially between Tm_EG and LFCA_EG? Have the authors analysed for instance RMSF and B-factors of these regions in the X-rays? How is the loop dynamics affected by substrate binding?
2. "BsEG has a more opened and hence more accessible cavity", "This occurs despite having a similarly accessible binding site as that of BS-EG (talking about LFCA_EG)" The active site volume and access tunnels could also be better analysed through POVME or CAVER simulations and compared among the different systems studied.
3. "We speculate that the ability to retain the substrate together with the enhanced substrate dynamics in the opened cavity may certainly contribute to the increased activity of the ancestral enzyme with respect to its extant counterparts even for crystalline substrates". The ability to retain the substrate and the enhanced substrate dynamics seem contradictory, also I would expect that more flexibility especially in exploring catalytic distances would be translated into a lower catalytic activity.
4. The authors mention that 84 mutations occurred from ancestral LFCA EG to the modern descendants. Where are these mutations located? How many of them are located closed to the active site (1st, 2nd sphere), and are there any that may impact active site volume and loop conformational dynamics? A more detailed description of the structural and dynamical differences could be commented.

More details should be provided in the SI regarding oligosaccharide parametrization and sugar ring conformation.

Two replicas of 500 ns of MD simulations were performed, however in Figure 6b only one of the replicas is shown. I believe all replicas should be included at least in SI.

Additionally, the quality of Figure 6 could be substantially enhanced: in 6a there are some sticks displayed that do not provide any relevant data (it is hard to see anything), the same happens in 6c which would be more relevant to show

Reviewer #2 (Remarks to the Author):

The authors have reconstructed an ancestral endoglucanase from ~2.8 billion years ago and characterized its properties, including an x-ray crystal structure. The properties differ from modern enzymes in having ~2-fold higher activity (one comparison enzyme) and processivity (one comparison enzyme) and higher thermal stability, but not higher than that of thermophiles. Analysis of this structure suggests that the balance between accessibility and affinity of the substrate is the origin of the higher affinity.

The work is done carefully, but the claims are exaggerated and not supported the results.

Abstract lacks numbers and is therefore misleading. The claim of 'high efficiency' turns out to be a factor of two as compared to modern bacterial enzyme. Similarly, the claim of 'exceptional properties' and 'remarkable synergy' must be quantified and compared.

Like many ancestral enzymes, the ancestral endoglucanase is more stable than modern counterparts from mesophiles, but not more stable than those from thermophiles (Fig 2d).

The ancestral endoglucanase is 1.5-2-fold more active than the comparison Bacillus enzyme and the Thermotoga enzyme. This result is a fortuitous finding. There is no reason to expect that another ancestral reconstruction will also give a more active enzyme, although the author suggest that this is so when they suggest ancestral reconstruction as a protein design strategy. There is no support for the notion that ancestral sequence reconstruction is a protein design strategy.

Minor.
line 141. T. maritima is suggested to be a very ancient organism. All organisms on the planet today are modern organisms. Rewording needed.

line 178 The processive behavior of the ancestral endoglucanse is also interesting, but again, likely fortitous. Some modern hydrolases in GH5 family are processive; why is it 'surprising' that the ancestral enzyme is processive? line 361 states 'modern counterparts from family GH5 do not display exoglucanase activity.', which contradicts the earlier statement that some GH5 enzymes are processive.

Reviewer #3 (Remarks to the Author):
See attached.

Reviewer 3 attached file

The manuscript presents an interesting example of reconstruction of an ancestral endoglucanase enzyme. Based on the predicted sequence of an ancestral enzyme, the authors expressed it and investigated its activity, stability and catalytic efficiency. The crystal structure of the resurrected endoglucanase solved 1.45 Å revealed the conserved fold typical of present-day endonucleases belonging to GH5 subfamily. The structure of ancestral endoglucanase was then used for MD simulations in order to get a deeper understanding of enhanced catalytic efficiency of the enzyme compared to its modern-day counterparts. The present work was logically executed. However, there are some major issues that should be addressed before publication.

1. The reconstruction of the ancestral enzymes is widely reported in enzyme/antigen/other proteins (JACS, 2013, 135, 2899; PNAS, 2012, 109, 2966; eLife, 2014, 3, e02304; JACS, 2016, 138, 1046; PNAS, 2016, 114, 3897; PNAS, 2009, 106, 21631; Nat. Chem. Biol. 2016, 12, 944). The method used in the present manuscript for construction and analysis of ancestral enzymes is not novel enough to guarantee publication. Thus, the authors should explain the innovation point of their work, providing the differences in comparison with previous works.

2. Based on the MD simulations of ancestral and two extant endoglucanases performed in the presence of model tetrasaccharide substrate, the authors hypothesize that a greater dynamics and more opened/accessible active site is the structural base of enhanced catalytic efficiency of the ancestral endoglucanase. However, the activity of ancestral enzymes was tested towards different substrates. I would like to see a short discussion about the fact that the authors used a different substrate(s) for experiments and MD analysis. Why the authors did not test the activity of the ancestral endoglucanase also towards the model tetrasaccharide substrate.

3. Steady-state kinetic parameters presented in Table 1 do not correspond to the kinetic profiles presented in Figure 2c. Based on the results in Table 1, ancestral endoglucanase should exhibit the highest catalytic efficiency from the three tested enzymes, but from Figure 2c it seems that the highest catalytic efficiency possesses the present-day enzyme Tm_EG. The author should explain this discrepancy.

4. The authors determined kinetic parameters by linearized fitting the measured data by the mathematical model describing the Michaelis-Menten model. However, the data presented in Figure 2c, particularly the data measured for Tm_EG seems to be better fitted by the mathematic model describing substrate inhibition and also cooperative substrate binding. Did the authors try to analyze the measured kinetic data also by mathematical models describing the other kinetic models then the simplest Michaelis-Menten model?

5. In Table 1, also the catalytic efficiency (k_{cat}/K_m) should be presented with the errors.

Reviewer #1

We are delighted to read that this reviewer considers that our paper "*is definitively of relevance and of interest to other researchers working in the field of endoglucanases and enzyme design*". She/he has comments that we address in the following lines. We thank the reviewer for these comments as they have helped to make our work more comprehensive.

The authors claim that the resurrected Precambrian endoglucanase is highly efficient, however the reported kcat, KM and kcat/KM values for instance are not that different from the ones obtained for the thermophilic TmEG. The major differences are found in the degradation of cardboard, newspaper, and softwood, although the conversion of LFCA-EG alone is not that different to TmEG and BsEG. The most important finding is that LFCA_EG incorporated into a mini-cellulosome enhanced its activity in contrast to the other two enzymes. Do the authors have an explanation for such differences? I believe the title of the paper could be changed to better represent their findings.

The referee appreciates that the degradation of lignocellulosic materials is quite different for the ancestral enzyme as compared to Tm_EG and Bs_EG. The substrates used, cardboard, newspaper and softwood contain cellulose plus other component such as hemicellulose and lignin and given the recalcitrant nature of lignocellulosic materials, cellulose is not readily accessible by an endocellulase without degradation of the other components. Therefore, the activity of the three endocellulases alone is lower. Once cellulose is made accessible by the other enzymes, laccase and xylanase, the activity increases but the ancestral endocellulase shows higher activity. This is the truly significant result here. We have further clarified this in pages 5 and 6 of the manuscript.

In the case of the integration into the cellulosome, we do not have a precise answer to that, but we believe that it is again an effect of the higher activity of the ancestral enzyme as compared to the natural cellulase. This is especially true when the CBM or the scaffoldin are incorporated in the ancestral cellulase. The higher activity is displayed in figure 4c, 4d and 4e. We admit that we are also surprised by the well integration of the ancestral cellulase into the cellulosome which reflects that the enhancement of the activity by this complex structure also happens with the ancestral enzyme.

We have followed the reviewer advise and highlight the enhanced activity towards biomass hydrolysis in the title.

The obtained X-ray structure for the ancestral enzyme is quite similar to the one from the modern enzymes as shown in Figure 5, thus the static structures are not able to explain the difference in activity. The analysis of the conformational dynamics of the enzymes in the presence of tetrasaccharide was therefore performed. The results shown in Figure 6 monitoring the catalytic distances indicate that BsEG enzyme should exhibit lower catalytic efficiencies with respect to the other two enzymes, which is in line with the kcat/KM values reported in Table 1. However, the dnuc distance is substantially shorter in TmEG than in LFCA_EG:

We agree with the referee that, based on the apparent affinity for the substrate (as monitored by dnuc and dAB), the Tm_EG variant should have the strongest activity. From our interpretation of the experimental evidence, we argue that there are additional factors at play that make LFCA_EG perform best. That is where the accessibility argument comes in (see also response to point 2 below). Tm_EG seems to have an enhanced affinity for the substrate, but that is in return for a more buried active site as compared with that of Bs_EG and LFCA_EG. On the other hand, LFCA_EG exhibits a behaviour that encompasses the best of both worlds: it has a very exposed accessible

active site compared to Tm_EG, while it has an enhanced ability to retain substrate relative to Bs_EG. In the revised version of the paper, we have attempted to clarify this point by reorganizing the whole computational results section so that we gradually introduce the different arguments.

1. It is mentioned that the main difference comes from a long loop that forms a clamp for the substrate. They also mention that "This loop may require large amplitude motions that decrease ligand flexibility". How the conformational dynamics of this loop differ among the different enzymes, but especially between Tm_EG and LFCA_EG? Have the authors analysed for instance RMSF and B-factors of these regions in the X-rays? How is the loop dynamics affected by substrate binding?

We thank the referee for this insightful comment. Based on the longer loop in Tm_EG we speculated that, for this variant, large amplitude motions could be required that are not present in the Bs_EG or LFCA_EG enzymes. The smaller mobility of the substrate in the bound state for Tm_EG can be inferred from the time series data presented in Fig. 6. However, in our revision, we have carried on additional analysis of the fluctuations and the role of the extended loop in the thermophile, which we include in a new supporting figure (Supp. Fig. 12).

First, we looked at the MD simulation results as suggested by the referee and, in particular, to the RMSF observed in the simulation runs. In figure R1 below we show the RMSF values for the loop regions for the three proteins. There are qualitative differences between the longer loop of Tm_EG and the loops in Bs_EG and the LFCA_EG enzymes. However, these fluctuations in the presence of substrate do not inform of the large amplitude motions that would be required for Tm_EG to accommodate the substrate or how the loop dynamics are coupled to substrate binding.

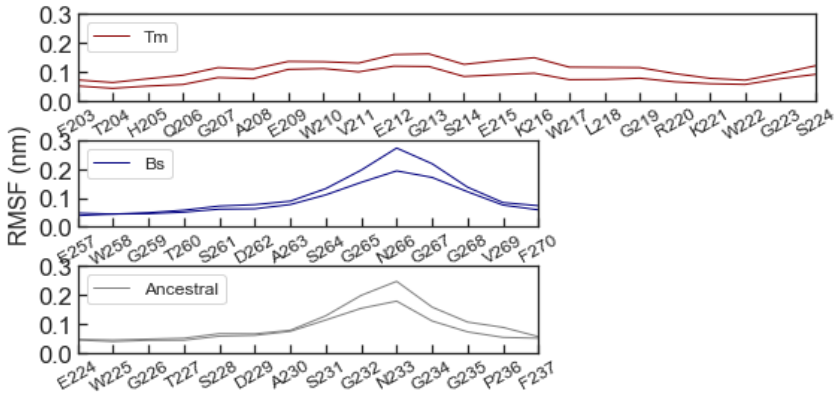


Figure R1. RMSF values for the three enzymes

In order to capture the magnitude of fluctuations that may occur in slow time-scales relative to MD, we have resorted instead to elastic network models. These models are useful for predicting functional conformational changes relevant to substrate binding. We now include the results from an anisotropic network model (ANM) derived from the alpha-carbon trace of the experimental 3D structures of three systems using the ProDy package (see Supplementary Figure 12). The ANMs reveal drastic differences between the Tm_EG loop and those in the LFCA_EF and Bs_EG enzymes. The slowest dynamical modes in the ANM for Tm_EG reveal large fluctuations localized in the loop region. These correspond to conformational rearrangements that may well correlate with substrate binding, as they correspond to opening and closing motions involving precisely the loop region. In the case of Bs_EG and LFCA_EG the slowest modes affect the protein much more globally.

Hence, this analysis is consistent with the interpretation that the longer loop in Tm_EG is involved in large amplitude motions, which may be involved in substrate binding and release.

2. “BsEG has a more opened and hence more accessible cavity”, “This occurs despite having a similarly accessible binding site as that of BS-EG (talking about LFCA_EG)” The active site volume and access tunnels could also be better analysed through POVME or CAVER simulations and compared among the different systems studied.

Again, we would want to thank the referee for this useful suggestion. While in the original version of the paper we were making the statement that the cavity in Bs_EG (and LFCA_EG) is more accessible than in Tm_EG, we were not providing evidence to support this. We have carried out new analysis which are included in the revised version of the paper as Supplementary Figure 11.

Specifically, we have followed the referee's advice and tested different tools for the characterization of the cavities in the three enzymes. We report results obtained with the CASTp 3.0 server, a popular tool that allows for the identification of pockets and cavities in proteins. The calculated cavities for the active sites of the enzymes are substantially different. The active site cavity of Tm has a much-enlarged surface due to its greater burial, while those of Bs_EG and LFCA_EG are much more exposed and have smaller surfaces areas.

3. “We speculate that the ability to retain the substrate together with the enhanced substrate dynamics in the opened cavity may certainly contribute to the increased activity of the ancestral enzyme with respect to its extant counterparts even for crystalline substrates”. The ability to retain the substrate and the enhanced substrate dynamics seem contradictory, also I would expect that more flexibility especially in exploring catalytic distances would be translated into a lower catalytic activity.

We understand the reviewer concern. When we say ability to retain substrate, we mean the ability to keep the substrate within the vicinity of the catalytic residues in the active site. This is evidenced in the simulations by the measurement of the distance between the nucleophile glutamic residue E224 and the glycosidic oxygen in the substrate. This distance is compatible with catalysis, but it has more fluctuations than in the case of the Tm_EG, and this what we mean by dynamics. We suggest that these enhanced dynamics of the substrate within the active site of the ancestral enzyme might be a reason for enhanced activity because it would allow for a faster achievement of a preferred conformation prior to the reaction. This is just a hypothesis. As the reviewer knows, enzyme and substrate dynamics during catalysis represent still a heated debate in the enzyme catalysis community and no unified explanation exists regarding the exact contribution of dynamics during catalysis. One such view is that enhanced protein and substrate dynamics may be crucial for catalysis[1,2], which is in line with our hypothesis. In addition, the role of evolution has been already recognized as a driving factor in the interplay between protein dynamics and function[3].

4. The authors mention that 84 mutations occurred from ancestral LFCA EG to the modern descendants. Where are these mutations located? How many of them are located closed to the active site (1st, 2nd sphere), and are there any that may impact active site volume and loop conformational dynamics? A more detailed description of the structural and dynamical differences could be commented.

We have now introduced and alignment an also a structural comparison between LFCA_EG and Bs_EG, to clearly show where the differences are in sequence and where are in the structure (New Supplementary Figure 2 and 9). In the comparative alignment between the two sequences, we can see that the mutations are distributed all over the sequence, with the catalytically relevant residues conserved (E136 and E224). The structural alignment shows that the differences mainly affects the loops and alpha-helixes rather than the active site itself. We believe that this is in line with our results on the computational simulations where dynamics and mobility are key elements.

More details should be provided in the SI regarding oligosaccharide parametrization and sugar ring conformation.

This is another useful point from the reviewer and we now have introduced further details in the Methods section.

Two replicas of 500 ns of MD simulations were performed, however in Figure 6b only one of the replicas is shown. I believe all replicas should be included at least in SI.

We have now introduced the second replica in the SI (Supp Fig 10).

Additionally, the quality of Figure 6 could be substantially enhanced: in 6a there are some sticks displayed that do not provide any relevant data (it is hard to see anything), the same happens in 6c which would be more relevant to show

We have followed this advice and have redesigned figure 6 altering colors and aesthetic elements to make it easier to follow. In particular, we have left only the heavy atoms from oligosaccharide and key residues. We have also matched the colors in Figure 5 with the color code in the rest of the figures.

Reviewer #2

We are happy to see that the reviewer appreciates the care taken in our work but have raised concerns about our claims. We provide answer to his/her comments as follow (reviewer´s comments in blue and answers in black):

Abstract lacks numbers and is therefore misleading. The claim of 'high efficiency' turns out to be a factor of two as compared to modern bacterial enzyme. Similarly, the claim of 'exceptional properties' and 'remarkable synergy' must be quantified and compared.

The reviewer is totally correct that wording must be more precise in the abstract. We have made an effort to rewrite the abstract to make it more accurate avoiding constructions mostly derived from the excitement in our work. Nevertheless, we would like to point out that is quite difficult to bring numbers to the abstract that summarize the main body of our work because we have tested a total of four enzymes in different combinations/constructs and under many different conditions. When we say *high efficiency* we mainly refer to the ability of the ancestral enzyme to work in a broad range of conditions and with many substrates, able to double the activity of the modern enzymes in many instances. These are what we find *exceptional properties*. We have clarified this in the abstract.

Like many ancestral enzymes, the ancestral endoglucanase is more stable than modern counterparts from mesophiles, but not more stable than those from thermophiles (Fig 2d).

As pointed out by the reviewer, the ancestral enzyme is not more stable than *T. maritima*. However, we need to clarify here that *T. maritima* is a hyperthermophile organism that grows at temperatures up to 90 °C. Hyperthermophiles are a class of thermophiles that live at temperatures above 80 °C. This temperature is detrimental to most life forms on Earth, and only *T.maritima* and other two bacteria, *Aquifex aeolicus* and *Geothermobacteriem ferrireducens*, are known to resist such temperature. Other species able to live under such conditions belong to the Archaea domain. We consider quite remarkable that our ancestral endoglucanase displays similar activity than that of T. *maritima* at temperatures above 80 °C (Fig 2a) which is a temperature only accessible to hyperthermophiles. Our enzyme also retains about half of the activity than that of *T. maritima* upon incubation at 80 °C. We can safely say that our ancestral enzyme displays a near-to-hyperthermophile phenotype.

We must remark here that ancestral enzymes are not necessarily hyperthermophiles, in fact, ancient life from the Archean eon has been described as thermophile. Ancestral proteins resurrected from that time have been shown to be highly resistant to temperature and our enzyme is no exception. This is included in the discussion part of the manuscript.

The ancestral endoglucanase is 1.5-2-fold more active than the comparison Bacillus enzyme and the Thermotoga enzyme. This result is a fortuitous finding. There is no reason to expect that another ancestral reconstruction will also give a more active enzyme, although the author suggest that this is so when they suggest ancestral reconstruction as a protein design strategy. There is no support for the notion that ancestral sequence reconstruction is a protein design strategy.

We agree with the reviewer that Ancestral Sequence Reconstruction is not *per se* a protein design strategy, and we have reworded our statement to make it clear. But it is also true that ancestral enzymes and proteins have displayed properties that, for the most part, are unattainable with current protein design techniques. Ancestral proteins have shown to be thermostable, pH tolerant, display chemical promiscuity, more activity, broad specificity, and other features that make them stand up versus their modern counterparts. In some instances, many of these improvements are well beyond what current design protocols have achieved so far. Some of the properties of the ancestral enzymes can be explained under the light of evolution, but other properties remain to be explained. Many examples in the literature exist. In **Table R1** below we have collected some prominent examples in the literature. Therefore, even though the Ancestral Sequence Reconstruction technique cannot predict *a priori* the features of the enzymes, if we have a vast and diverse collection of modern sequences in our phylogeny, we will be able to reach ancestral enzymes that supposedly lived under harsh conditions. We know this because the tree can be dated, and geological records have shown the environmental conditions of the past times. In this sense, we use the power of evolution to find these variants that can be potentially further improved by using other protein design techniques such as directed evolution.

In this work, our main goal is to show that the improved features make the ancestral endocellulases an attractive example for possible biotechnological application of the reconstruction technique. For instance, we know that the ancestral cellulase LFCA_EG can produce crystalline nanocellulose, something that we have proven hard for other modern endocellulases (work in progress). To the best of our knowledge, our work represents the first report in which an enzyme is improved in most of the aspects that are relevant from a biotechnological point of view, avoiding long screening processes and in one single shot. We do this by probing the ability of the ancestral enzymes to work under harsh conditions that resemble those of industrial settings. As such, ASR emerges as a top-notch method to unveil enzymes with industrial potential with the possibility of being combined with directed evolution aimed at designing more robust biocatalysts.

Table R1. Examples of resurrected proteins with significant improvement/changes in properties

Protein/Enzyme	Reference	Improved/altered properties
Elongation Factor	Gaucher et al, 2008[4]	Thermal stability
Thioredoxin	Perez-Jimenez et al, 2011[5]	Thermal and pH stability and activity
B-lactamase	Risso et al, 2013[6]	Thermal stability and chemical promiscuity
Coagulation Factor VIII	Zakas et al, 2017[7]	Activity, stability and immunoreactivity
Uricase	Kratzer et al, 2014[8]	Activity and immunoreactivity
LeuB	Hobbs et al, 2012[9]	Thermal stability and activity
Titin	Manteca et al, 2017[10]	Mechanical Stability
Cellulases	Present manuscript	Thermal and pH stability, activity, specificity, chemical promiscuity, synergy

Nevertheless, we decided to further probe the ability of ASR to provide functional cellulases and we reconstructed two other sequences belonging to the nodes LACA (Last Actinobacteria Common Ancestor) and LCCA (Last Clostridia Common Ancestor). The genes were cloned and tested in the lab and, as the reviewer can see, the new ancestral cellulases display good activity operating well in the temperature range 30-90 °C and pH range 4-10. Nevertheless, the maximum level of activity still corresponds to LFCA_EG. In general, the three ancestral endocellulases perform better than the extant Bs_EG. At high temperatures LFCA_EG and LCCA_EG perform similar than the extremophile enzyme from T maritima. We went even further and plotted the relative activity of the ancestral enzymes and query Bs_EG at 50 °C against the evolutionary time. The plot is now included in the new figure 2 and shows a clear decreasing trend of the activity over time. We were surprised to see that this trend runs parallel to the cooling trend of the seawater temperature over a 3.5 Billion-years time span, which suggests a tight relationship between enzyme activity, stability and environmental temperature. This trend seems to be general in ancestral enzymes stability, but we probe it here even for activity itself.

We have included these new experiments in the new Figure 2 and in sequence information in the supplementary material. We hope now that the reviewer sees these results as a probe of the accuracy of tree, the solidity of technique, and agrees that our experiments cannot be simply attributed to luckiness or fortuitousness.

Minor.

line 141. T. maritima is suggested to be a very ancient organism. All organisms on the planet today are modern organisms. Rewording needed.

The reviewer is correct in that T. maritima is a modern organism, but it has been suggested that thermophilicity is an ancient phenotype because it is mostly present in deep branches of bacterial an archaeal phylogenies[11]. Thus, organisms displaying thermophilicity or even more so hyperthermophilicity, are likely to be reminiscence of ancient forms of life. However, it is also true that hyperthermophilicty in T. maritima might have been the results of lateral gene transfer[12], thus we agree that this might be somehow speculative and have decided to remove the statement.

line 178 The processive behavior of the ancestral endoglucanse is also interesting, but again, likely fortitous. Some modern hydrolases in GH5 family are processive; why is it 'surprising' that the

ancestral enzyme is processive? line 361 states 'modern counterparts from family GH5 do not display exoglucanase activity.', which contradicts the earlier statement that some GH5 enzymes are processive.

The reviewer is correct in that some GH5 family enzymes are processive, but they have not been included in our phylogenetic analysis and reconstruction. Therefore, we believe that our results are somehow surprising, because we can capture a processive phenotype from sequences that are non-processive.

Processivity is a feature commonly found in exoglucanases. Some modern endoglucanases have been found to have processive activity but most of them belong to the GH9 family. As the reviewer points out, it is true that a few GH5 endoglucanases have been shown to have processive activity as well[12,13], but they are not included in our analysis. Also, we see processive activity in the ancestral endoglucanase even in the absence of CBM. We believe that this finding reflects chemical promiscuity in the ancestral enzymes. Chemical promiscuity has been suggested in ancestral enzymes but very few examples exist, and we consider this processive activity a good example. We have now included a clarification in our manuscript (page 5)

Reviewer #3

We thank this reviewer and her/his appreciation that our work is logically executed. Her/his comments clearly served us to improve the paper. In what follows, we reply to the comments:

1. The reconstruction of the ancestral enzymes is widely reported in enzyme/antigen/other proteins (JACS, 2013, 135, 2899; PNAS, 2012, 109, 2966; eLife, 2014, 3, e02304; JACS, 2016, 138, 1046; PNAS, 2016, 114, 3897; PNAS, 2009, 106, 21631; Nat. Chem. Biol. 2016, 12, 944). The method used in the present manuscript for construction and analysis of ancestral enzymes is not novel enough to guarantee publication. Thus, the authors should explain the innovation point of their work, providing the differences in comparison with previous works.

We understand the reviewer concern which we consider a very valid point. It is true that examples in the literature exist of reconstructed enzymes in the past years, but it is also true that most of the published work aim to prove an evolutionary idea/hypothesis or a specific feature and do not deeply explore the possibility of using ancestral enzymes with a biotechnological perspective. The examples provided by the reviewer are very well selected and published in top journals, but they also provide good examples in the mentioned direction, they represent studies that attend to probe an evolutionary driven aspect and not much about exploring the applicability of the reconstructed proteins. Although the possibility of using ancestral enzymes in biotechnology has been certainly pointed out before[6,14,15], it has not been proved to the extend we now do in our work. This is the main difference of our work with respect to other ancestral sequence reconstruction studies. We use this technique to improve an example of an enzyme relevant in biotechnology, and we focus on most of the aspects that are interesting in a possible industrial setting, i.e., thermostability, pH tolerance, broad substrate usage, chemical promiscuity and synergy in enzyme cocktails. To the best of our knowledge, this is the first report of an ancestral reconstruction that clearly aims towards a real application of ancestral enzymes in the broadest sense of the word.

Focusing on evolution is perfectly reasonable and, for the most part, it has been the main goal in most reconstructed proteins. In our work we certainly care about possible evolutionary implications, but we focus our interest on a purely practical utilization of the enzymes and the technique, and this

is what makes our work different and novel. We have now introduced a statement in the discussion that highlights this.

2. Based on the MD simulations of ancestral and two extant endoglucanases performed in the presence of model tetrasaccharide substrate, the authors hypothesize that a greater dynamics and more opened/accessible active site is the structural base of enhanced catalytic efficiency of the ancestral endoglucanase. However, the activity of ancestral enzymes was tested towards different substrates. I would like to see a short discussion about the fact that the authors used a different substrate(s) for experiments and MD analysis. Why the authors did not test the activity of the ancestral endoglucanase also towards the model tetrasaccharide substrate.

This is an interesting point. We have chosen a tetra-saccharide because it is a minimal substrate that can be easily adapted for computational calculations rendering a realistic situation as it covers most of the active site of the enzyme. It is common practice to use small polysaccharides composed of a just few sugar rings for computational docking calculations[13,16,17]. The main advantage of the small substrate is that it limits the number of degrees of freedom at least for one of the molecules involved in the ligand binding process, which keeps the sampling problem under reasonable control. Also, we would like to point out that the main purpose of the computational analysis was to analyse the dynamics of the sugar rings prior to the reaction itself nearby the active site, which can be easily estimated with the tetra-saccharide substrate. Therefore, we do not see necessary to perform the actual experimental reaction. Nevertheless, in our assay it is not unlikely that a substrate with just a few sugar rings may be present in solution as we are able to detect glucose and cellobiose.

3. Steady-state kinetic parameters presented in Table 1 do not correspond to the kinetic profiles presented in Figure 2c. Based on the results in Table 1, ancestral endoglucanase should exhibit the highest catalytic efficiency from the three tested enzymes, but from Figure 2c it seems that the highest catalytic efficiency possesses the present-day enzyme Tm_EG. The author should explain this discrepancy.

From the plot in Figure 2c we can determine K_M and V_{max} using a simple Michaelis Menten model, which can be used to determine the turnover rate, k_{cat}, and the catalytic efficiency described as k_{cat}/K_M. These are the parameters reported in table 1. We understand that the reviewer sees the Tm_EG kinetic profile reaching a higher V_{max} in Figure 2c, but k_{cat} is determined as $V_{max}/[E]$. The higher V_{max} (k_{cat}*[E]) is simply the result of a higher concentration of Tm_EG but is not a reflexion of a higher efficiency. From the parameter obtained, the higher efficiency corresponds to LFCA_EG. Thus, we see no discrepancy. We hope is now clearer, but we are happy to provide any further information needed by the reviewer to see our conclusions.

4. The authors determined kinetic parameters by linearized fitting the measured data by the mathematical model describing the Michaelis-Menten model. However, the data presented in Figure 2c, particularly the data measured for Tm_EG seems to be better fitted by the mathematic model describing substrate inhibition and also cooperative substrate binding. Did the authors try to analyze the measured kinetic data also by mathematical models describing the other kinetic models then the simplest Michaelis-Menten model?

This comment certainly provides an opportunity to test a different model. We have followed the reviewer advise and have fitted the data to a substrate inhibition kinetic model as well as to cooperative substrate binding model. For better accuracy we have introduced more data points at low and high concentrations where the models are more sensitive. In figure R2, we summarize the results obtained for the three models tested, Michalis-Menten, substrate inhibition and competitive

inhibition. We have used a classical substrate inhibition model in which $V_0=V_{max}*[S]/(K_M+[S]*(1+[S]K_i))$, where K_i is the dissociation constant. This model is characterized for the progressive decay of V_0 with substrate concentration after reaching a maximum. In Figure R2 we observed that our data set does not fit well to the substrate inhibition model, specially at high substrate concentrations. In the case of cooperative binding inhibition $V_0=V_{max}*[S]^h/(K_M{}^h+[S]^h)$, where *h* is the Hill coefficient that determines the cooperativity, we have measured *h* to be around 1, thus questioning the possibility of positive substrate cooperativity, virtually rendering a Michaelis-Menten fitting.

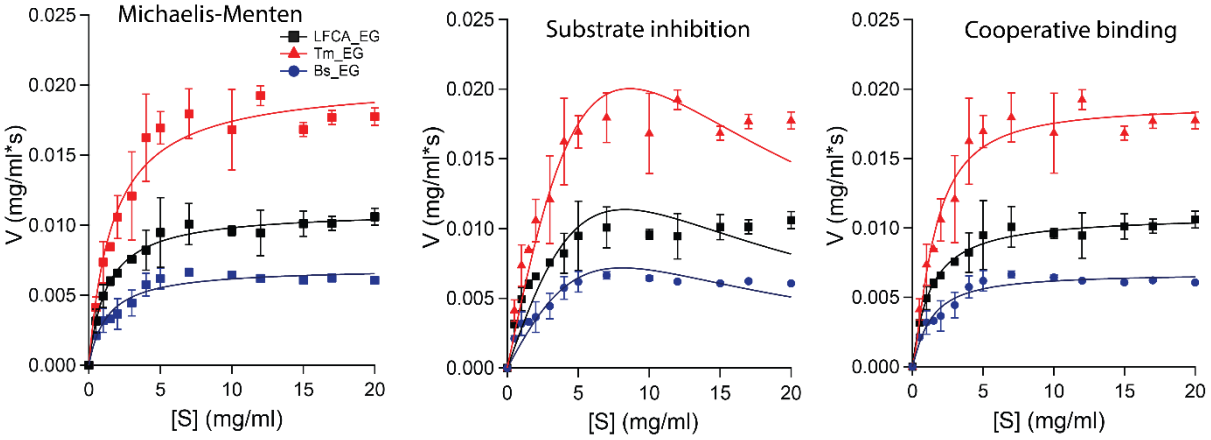


Figure R2. Fittings for the three models tested.

These fittings are in line with most previous studies that have reported endocellulases to follow a traditional Michaelis-Menten kinetics. We have kept our original Michaelis-Menten fitting including the new data point.

5. In Table 1, also the catalytic efficiency (kcat/Km) should be presented with the errors.

We have now included the errors in Table 1.

References

1 Mercadante, D., Melton, L. D., Jameson, G. B., Williams, M. A. & De Simone, A. Substrate dynamics in enzyme action: rotations of monosaccharide subunits in the binding groove are essential for pectin methylesterase processivity. *Biophysical journal* **104**, 1731-1739, doi:10.1016/j.bpj.2013.02.049 (2013).

2 Otten, R. *et al.* Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nat Commun* **9**, 1314, doi:10.1038/s41467-018-03562-9 (2018).

3 Campbell, E. *et al.* The role of protein dynamics in the evolution of new enzyme function. *Nature chemical biology* **12**, 944-950, doi:10.1038/nchembio.2175 (2016).

4 Gaucher, E. A., Govindarajan, S. & Ganesh, O. K. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. *Nature* **451**, 704-707, doi:10.1038/nature06510 (2008).

5 Perez-Jimenez, R. *et al.* Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. *Nature structural & molecular biology* **18**, 592-596, doi:10.1038/nsmb.2020 (2011).

6 Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A. & Sanchez-Ruiz, J. M. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases. *J Am Chem Soc* **135**, 2899-2902, doi:10.1021/ja311630a (2013).

7 Zakas, P. M. *et al.* Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction. *Nat Biotechnol* **35**, 35-37, doi:10.1038/nbt.3677 (2017).
8 Kratzer, J. T. *et al.* Evolutionary history and metabolic insights of ancient mammalian uricases. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 3763-3768, doi:10.1073/pnas.1320393111 (2014).
9 Hobbs, J. K. *et al.* On the origin and evolution of thermophily: reconstruction of functional precambrian enzymes from ancestors of Bacillus. *Molecular biology and evolution* **29**, 825-835, doi:10.1093/molbev/msr253 (2012).
10 Manteca, A. *et al.* Mechanochemical evolution of the giant muscle protein titin as inferred from resurrected proteins. *Nature structural & molecular biology* **24**, 652-657, doi:10.1038/nsmb.3426 (2017).
11 Farrah, J., Saharay, M., Georgiannos, S. N., Scurr, J. H. & Smith, P. D. Variable venous anatomy of the popliteal fossa demonstrated by duplex scanning. *Dermatol Surg* **24**, 901-903 (1998).
12 Zheng, F. & Ding, S. Processivity and enzymatic mode of a glycoside hydrolase family 5 endoglucanase from Volvariella volvacea. *Appl Environ Microbiol* **79**, 989-996, doi:10.1128/AEM.02725-12 (2013).
13 Wu, B. *et al.* Processivity and enzymatic mechanism of a multifunctional family 5 endoglucanase from Bacillus subtilis BS-5 with potential applications in the saccharification of cellulosic substrates. *Biotechnology for biofuels* **11**, 20, doi:10.1186/s13068-018-1022-2 (2018).
14 Alcalde, M. When directed evolution met ancestral enzyme resurrection. *Microb Biotechnol* **10**, 22-24, doi:10.1111/1751-7915.12452 (2017).
15 Risso, V. A., Gavira, J. A. & Sanchez-Ruiz, J. M. Thermostable and promiscuous Precambrian proteins. *Environ Microbiol* **16**, 1485-1489 (2014).
16 Bu, L., Crowley, M. F., Himmel, M. E. & Beckham, G. T. Computational investigation of the pH dependence of loop flexibility and catalytic function in glycoside hydrolases. *The Journal of biological chemistry* **288**, 12175-12186, doi:10.1074/jbc.M113.462465 (2013).
17 Saharay, M., Guo, H. & Smith, J. C. Catalytic mechanism of cellulose degradation by a cellobiohydrolase, CelS. *PloS one* **5**, e12947, doi:10.1371/journal.pone.0012947 (2010).

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

In this new revised version of the manuscript, the authors have successfully addressed most of my original concerns. The new provided version of the manuscript, especially the computational part of it, has been substantially improved and many of the statements that were not demonstrated before are now supported by additional computational analysis. I, therefore, recommend publication of the paper in Communications Chemistry journal.

The only minor point that the authors did not comment is whether they have analysed the B-factors of their available X-ray structures (especially how they differ in the loop region).

Reviewer #2 (Remarks to the Author):

The authors have carefully addressed the reviewer concerns in their revised manuscript. The end of the discussion still contains overly broad statements.
In the penultimate paragraph, the authors state that ancestral reconstruction will yield the stability, higher activity, unusual reactivity. There is good reason to expect that reconstructed ancestral enzyme from times that the earth was warmer will tolerate higher temperatures. The other properties are fortuitous. An ancestral reconstruction of another enzyme may not yield higher activity, broad pH range, activity at lower temperature and so on.
In the last paragraph, the authors claim that there are no protein engineering methods to optimize enzymes toward multiple criteria. This is not true. For example, Codexis has optimized numerous enzymes for industrial use using protein engineering. Their approach is time consuming and tedious, but to say there are no methods is incorrect.

Reviewer #3 (Remarks to the Author):

All my comments were addressed. Additions to the manuscript and supporting information are clear and complete. The revisions that were made significantly add depth to the study and clarify its impact. I recommend acceptance for publication in Communications Chemistry.

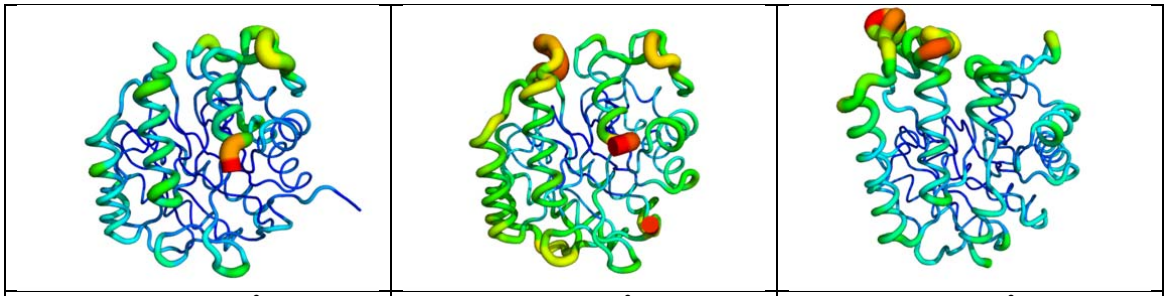
Response to reviewers

Reviewer #1 (Remarks to the Author):

In this new revised version of the manuscript, the authors have successfully addressed most of my original concerns. The new provided version of the manuscript, especially the computational part of it, has been substantially improved and many of the statements that were not demonstrated before are now supported by additional computational analysis. I, therefore, recommend publication of the paper in Communications Chemistry journal.

The only minor point that the authors did not comment is whether they have analysed the B-factors of their available X-ray structures (especially how they differ in the loop region).

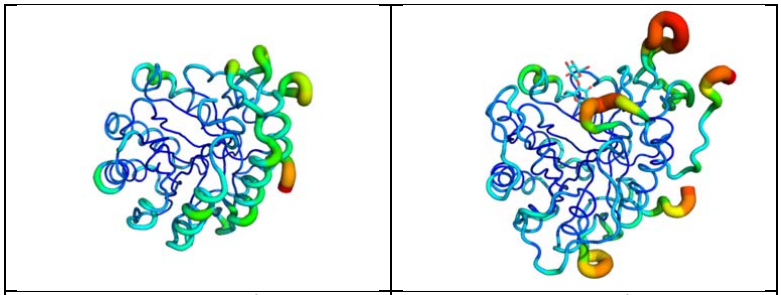
We are happy to see that the reviewer is satisfied with the revision and appreciate our effort to improve the work. We apologize for overseeing this point in our previous round. Thermal motion B-factor is also affected by crystal quality and accounts for the effect of the crystals packing, e.g. space group, etc. Therefore, this type of analysis is more appropriate when crystals are obtained in the same space group. Nevertheless, we have analysed B-factor with special focus on the loop regions. As shown in figure R1, there is a coincidence of loops with high B-factor for the LFCA, Bs and Tm structural models.



LFCA (1.45 Å)	Bs (1.97 Å)	Tm (2.2 Å)

Figure R1. 3D structural models of LFCA, Bs and Tm as B-factor putty representation generated with Pymol. Values in brackets correspond to the resolution limit. Note that the tubular spline which traces the backbone is also graded in color from lowest (blue)→highest(red) B-factor values.

In addition, if we compare LFCA_EG with the structure of alkaline cellulase from *Bacillus sp* bound to cellobiose (PDB ID. 1G0C), there are no significant differences in the B-factors distribution (Figure R2). From Fig.R2 it is clear that the length of the loops contributes more to increases the B-factor than the presence of the substrate.



LFCA (1.45 Å)	Bs (1.90 Å)

Figure R2. 3D structural models of LFCA and Bs as B-factor putty representation. Cellobiose in the Bs model is represented with stick. Values in brackets correspond to the resolution limit.

Reviewer #2 (Remarks to the Author):

The authors have carefully addressed the reviewer concerns in their revised manuscript. The end of the discussion still contains overly broad statements. In the penultimate paragraph, the authors state that ancestral reconstruction will yield the stability, higher activity, unusual reactivity. There is good reason to expect that reconstructed ancestral enzyme from times that the earth was warmer will tolerate higher temperatures. The other properties are fortuitous. An ancestral reconstruction of another enzyme may not yield higher activity, broad pH range, activity at lower temperature and so on.

We understand the reviewer concern, but it is our experience with ancestral enzymes that features such as stability, pH, activity, expression level or promiscuity tend to be also affected in the resurrected enzymes; and in some cases, there is an evolutionary explanation. For instance, it is known that oceans pH was below 6 until the raise of oxygen. Chemical promiscuity has also been suggested for ancestral enzymes which has been described as generalists versus specialists. Higher activity towards different substrates might be associated with less specificity in less evolved enzymes. Modern enzymes are likely better for their substrates in their modern host. It is true that many of these explanations are still hypotheses and need further testing, but we are fortunate that many of the modified features that we measure make these enzymes potentially suitable for biotechnology.

In the last paragraph, the authors claim that there are no protein engineering methods to optimize enzymes toward multiple criteria. This is not true. For example, Codexis has optimized numerous enzymes for industrial use using protein engineering. Their approach is time consuming and tedious, but to say there are no methods is incorrect.

We see the reviewer point and we have reworded the paragraph. Nevertheless, Codexis uses directed evolution and HT for screening which are time consuming and expensive, but still are the standard in the industry. We want to stress here that, with the required expertise, the reconstruction of ancestral enzymes can be achieved in a fraction of the time required to evolve an enzyme, and no screenings are needed. Nevertheless, combining the two techniques is perhaps a good solution.