nature portfolio

Peer Review File

Phage therapy with nebulized cocktail BX004-A for chronic Pseudomonas aeruginosa infections in cystic fibrosis: a randomized first-in-human trial

Corresponding Author: Dr Iddo Weiner

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
I thank the authors for addressing my previous comments and for the additional information and the revisions made to the manuscript. However, I still have a few considerations.

From table S11, I deduce that the 7 patients in the BX004-A group were treated with a combination of nebulized phage cocktail BX004-A and one or more (8 antibiotics for 7 patients) inhaled antibiotics [aztreonam (n=1), tobramycin (n=4), or colistin (n=3)], while the 2 patients in the control group were treated with the antibiotic aztreonam only. It was not possible (unless I missed this) to link the clinical trial results (Fig. 6) to the exact antibiotic therapy that was administered (and to the antibiotic susceptibility tests), for each individual patient.

Table S13 shows variable sensitivities (resistant, but also susceptible and intermediate) of the bacterial isolates from the 7 patients of the BX004-A group to the aztreonam and tobramycin antibiotics, which were applied concomitantly with the phage cocktail. No information is given with regard to the sensitivities towards colistin, a potent last resort treatment for Pseudomonas aeruginosa infections. The bacterial isolates from the control group were all resistant to the applied antibiotic aztreonam.

The authors state “four genetically distinct strains were grown in the presence of BX004-A, antibiotics (aztreonam, colistin, or tobramycin), or both.” and “The results show no antagonism between BX004-A and any of the tested antibiotics (Fig. S6).”
In my opinion, Fig. S6 also suggest a possible additive or synergistic effect between BX004-A and the antibiotics aztreonam, colistin, or tobramycin.
Since the interaction between phages and antibiotics can also depend on the targeted bacterial isolate, it is a pity that these experiments were not performed using reference strains instead of the patients’ bacterial isolates (which were at hand).

Considering…
1) the possible synergistic or additive effect of BX004-A and the concomitantly applied antibiotics (at least in the 4 tested strains),
2) the in vitro antibiotic sensitivity test results, suggesting a certain susceptibility of some of the patients’ bacterial isolates to the concomitantly applied antibiotics aztreonam and tobramycin (no information with regard to colistin) in the BX004-A group, compared to aztreonam resistance only in the control group.
…in my opinion, it is more appropriate to state that the present clinical trial evaluated the microbiological efficacy of BUX004-A in combination with one or more antibiotics (aztreonam, tobramycin, or colistin) compared to the antibiotic aztreonam alone, rather than BX004-A compared to placebo (cf. the abstract).

Reviewer #3

(Remarks to the Author)
In re-considering my critiques in light of the authors responses, my opinion is as follows:

1. While I did not think that the Pa sequencing and expression analysis added much to the study, little harm is done by including all this data (except perhaps having the reader wade through somewhat extraneous information).

2. While it would be more rigorous to include a positive control for the biofilm studies (and many biofilm studies use controls), little harm is done by omitting this.

3. However, I do not think it reasonable to conclude that they demonstrative antimicrobial efficacy (even with the "preliminary" qualifier) of the Bx004 cocktail for the following reasons (these points were largely ignored in the response):

a. As noted, both placebo subjects were not taking CFTR modulators while ~1/2 of the treatment group subjects were. This is a major problem as modulators are more effective against CF infections than any previously developed treatment. As noted, their approach would be akin to a cancer trial in which half of the treatment group had a new therapy added on to an extremely-effective, revolutionary treatment but the placebo group had neither therapy. I note that this critical information is not mentioned in the text, but rather buried in supplementary table-11 (even in the revision).

b. The efficacy claims are largely due to the fact that the placebo group showed increases in Pa density in the absence of any treatment. Thus, the comparisons touted between the untreated and treated groups are in large part due to a curious, unexplained, and perhaps spurious worsening in the placebo group. It appears possible/likely that much of the differences between treatment and placebo groups were driven by a single outliers in both groups at day 4 and 15.

c. Making matters worse, even a quick look at table S-15 (where average Pa density changes are reported) reveals two important findings. First, the authors' pre-specified scheme plan shows a minimal efficacy signal, however the data actually reported in the main text (Fig 6b) was obtained by swapping out the Pa baseline value of one placebo patient, to a value more favorable for a positive signal. The authors explanation for why they made this change did not make sense. If the sample is poor (because the subject inappropriately took antibiotics) this would be reflected in both the CFU and Pa relative abundance), so the finding that these parameters move in the same direction in no way justifies the switch in baseline data.

Without this post-hoc maneuver the day 4 difference between treatment and placebo groups changed from 1.92 logs (favoring the treatment group) to only 0.39 logs. The day 15 difference between treatment and placebo groups changed from 2.68 logs (favoring the treatment group) to 1.15 logs. This post-hoc change (which produces the appearance of efficacy) is irregular and is not reported to the reader in the main text. I wondered if the academic authors of the paper were aware of this post-hoc change in baseline value data, and how the appearance of a positive signal of efficacy was enhanced by it.

Reviewer #4

(Remarks to the Author)
The authors have toned down the efficacy statements, but not the safety statements. With only 7 patients in the treatment arm, one cannot conclude that "these data demonstrate the excellent safety profile". I do not have the expertise to judge whether the SAE reported in Table S12 is indeed unrelated to the study drug.

Make clear in the abstract that this is a phase 1b trial. The sentence "Safety and microbiological efficacy of BX004-A were evaluated in a first-in-human double-blind placebo-controlled clinical trial" may suggest a phase 3 trial to the reader.

A few remaining comments:

1) Thanks for adding the p-values in figure 6. For 6B-D they are very useful. However, I don't see much relevance in giving p-values for Figure 6A. Could there be any explanation for having PFU larger than zero in the placebo group? Moreover, the assumptions for validity of the t-test are clearly violated.
I guess log(PFU+1) was used, otherwise log(PFU) would give -infinity. Please make that clear.

2) The values in Fig 1B have a very skewed distribution. Pearson correlation assumes a normal distribution. I suggest to use Spearman correlation. How many points are in the scatterplot? The labels suggest a different number (n=143 vs n=335)?

R.B.Geskus

Reviewer #5

(Remarks to the Author)

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
I would like to thank the authors for the answers to my questions and comments and for the subsequent clarifying additions to the manuscript.

I realize I made a mistake in my penultimate comment. I meant "..., it is a pity that these experiments were performed using reference strains instead of the patients' bacterial isolates."

The authors rightfully answered that they did use reference strains (e.g. PAO1) to test BX004-A - antibiotic interactions. I still feel that these test should also have been performed on the patients' isolates, as results not only depend on the combined phages and antibiotics, but also on the targeted bacterial isolates.
However, this is in principle covered by the general reference to potential confounders (including combination with antibiotics) added to the manuscript.

I have no additional comments.

Point-by-point responses to reviewers' comments

General notes from the authors:

First, we would like to deeply thank the reviewers for their time and for their professional comments. We are happy to update that we have addressed all the points and are happy to submit our revised manuscript.

Below please find all comments raised in the previous revision step, with our responses, which are written in green font and indented, to avoid confusion. In the submitted files we used WORD's built-in "track changes" to facilitate reviewing the changes compared to the previous version. In the supplemental tables file (attached in .xlsx format) we manually marked changes using green font color.

Beyond the revisions implemented based on the reviewer feedback below, we made two additional edits to the manuscript. Both changes are based on the recent QC / finalization of our clinical study report for the study:

Section in manuscript	Updated value	Previous value	Reason for change
Figure 6B, Placebo arm, Day 2 datapoint on curve; Table S15, post-hoc analysis, Placebo arm, Day 2 value	3.07	1.53	**Input from our statistician**: The previous value of 1.53 was the average of two values for CFU change from baseline: [3.07, 0], where the zero value was derived from a patient with a missing baseline CFU value (only a qualitative result was available at baseline, due to low sputum volume). Since the baseline CFU was missing, the D2 value for that patient was used as the baseline value (per the SAP) for change from baseline calculations, however the change from baseline at day 2 was computed as 0 (the D2 value was inadvertently subtracted from the same value). After clarification from the statistician who finalized the clinical study report, there would only be one value for CFU change from baseline at D2 (3.07).
Table S15, pre-specified analysis, BX004-A arm, Day 3b value	-1.06	-0.43	Human typing error found during QC (the post-hoc analysis only affected one of the patients in the placebo group; the BX004-A values remained identical between pre-specified and post-hoc analyses).

The above changes do not affect any of the interpretations presented in this manuscript, and do not trigger any changes in wording.

REVIEWER COMMENTS

Reviewer #2 (Remarks to the Author):

I thank the authors for addressing my previous comments and for the additional information and the revisions made to the manuscript. However, I still have a few considerations.

From table S11, I deduce that the 7 patients in the BX004-A group were treated with a combination of nebulized phage cocktail BX004-A and one or more (8 antibiotics for 7 patients) inhaled antibiotics [aztreonam (n=1), tobramycin (n=4), or colistin (n=3)], while the 2 patients in the control group were treated with the antibiotic aztreonam only. It was not possible (unless I missed this) to link the clinical trial results (Fig. 6) to the exact antibiotic therapy that was administered (and to the antibiotic susceptibility tests), for each individual patient.

> Please allow us to clarify that the study protocol required background use of one of 3 inhaled antibiotics (tobramycin, aztreonam, or colistin) which are standard of care in CF patients with *Pseudomonas aeruginosa*, in addition to study drug (BX004-A or placebo). One subject on BX004-A switched from aztreonam (Days 1 through 5) to colistin (Days 6 through 8) due to supply issues of the inhaled antibiotic, and resumed aztreonam Days 9 through 15; therefore, this subject is counted under aztreonam and colistin (hence 8 antibiotics for 7 patients). We added the following footnote to Table S11 to clarify: *"One subject switched from aztreonam (Days 1 through 5) to colistin (Days 6 through 8) due to supply and resumed aztreonam Days 9 through 15. Another subject switched from tobramycin to colistin on Day 9. All other subjects were on the same inhaled antibiotic from Days 1 through 15 (last sputum collection in Part 1)."*
>
> Subsequently, in the revised manuscript, we added Fig. S10 that shows the CFU dynamics within the treatment group, with color grouping based on either antibiotic type or CFTR modulator usage. We refer to this analysis from the main text:
>
> *"Additionally, we assessed the impact of two potential confounders in the BX004-A arm: type of inhaled antibiotic and use of a CFTR modulator. Neither subgroup analysis provided conclusive evidence of confounding effects (Fig. S10). However, given the small sample size of Part 1 of this first-in-human study, caution is warranted when interpreting efficacy-related outcomes. Therefore, both statistical significance and the impact of confounders should be monitored in larger studies."*

Table S13 shows variable sensitivities (resistant, but also susceptible and intermediate) of the bacterial isolates from the 7 patients of the BX004-A group to the aztreonam and tobramycin antibiotics, which were applied concomitantly with the phage cocktail. No information is given with regard to the sensitivities towards colistin, a potent last resort treatment for Pseudomonas aeruginosa infections. The bacterial isolates from the control group were all resistant to the applied antibiotic aztreonam.

Please allow us to clarify that while antimicrobial susceptibility testing informs the appropriate selection of *systemic* antibiotics for routine infections (by selecting an antibiotic to which the bacteria are susceptible, based on clinical breakpoints), it is not considered relevant for the selection of *inhaled* antibiotics in CF patients with chronic *Pseudomonas* pulmonary infection. Clinicians who treat this population often select inhaled antibiotics irrespective of in vitro susceptibility, guided by clinical response, rather than MICs. There is a disconnect between the efficacy of *inhaled* antibiotics and in vitro MICs because of the very high concentrations of antibiotics delivered locally in the lung which can overcome antibiotic resistance mechanisms. Currently, all FDA-recognized breakpoints or Susceptibility Test Interpretive Criteria (STIC) and conventional Clinical and Laboratory Standards Institute (CLSI) breakpoints refer only to *injected* or *orally* administered antimicrobials; there are no criteria for *inhaled* antibiotics. In a workshop on endpoints for CF clinical trials held by the European Medicines Agency, it was clearly stated that there is no scientific basis for establishing breakpoints for inhaled antibiotics, that patients should not be selected for CF clinical trials based on MICs at baseline, and that observations from clinical trials indicate that increasing MICs do not predict a lack of clinical response (*https://www.ema.europa.eu/en/documents/report/report-workshop-endpoints-cystic-fibrosis-clinical-trials_en.pdf*).

Antibiotic susceptibility testing in our study was done to assess if there was a significant change in antibiotic MIC after phage therapy, since other researchers have reported that *Pseudomonas* can become more susceptible to antibiotics after phage therapy. In our study there was no significant change in MICs that can be attributed to phage therapy.

As for colistin, there are no current breakpoints to determine if a *Pseudomonas* isolate is susceptible even to *systemic* colistin (breakpoints only exist to determine if *Pseudomonas* is resistant or intermediate). Therefore, the utility of this additional testing was considered low (especially given the limitation of MICs in guiding use of inhaled antibiotics, as noted above) and was not done in the microbiology laboratory.

For reference, we made a few corrections to the categorization of S/I/R in Table S13, noted in green font (using breakpoints for systemic antibiotics) based on the recent QC / finalization of our clinical study report for the study.

The authors state “four genetically distinct strains were grown in the presence of BX004-A, antibiotics (aztreonam, colistin, or tobramycin), or both.” and “The results show no antagonism between BX004-A and any of the tested antibiotics (Fig. S6).” In my opinion, Fig. S6 also suggest a possible additive or synergistic effect between BX004-A and the antibiotics aztreonam, colistin, or tobramycin. Since the interaction between phages and antibiotics can also depend on the targeted bacterial isolate, it is a pity that these experiments were not performed using reference strains instead of the patients’ bacterial isolates (which were at hand).

Please note that one of these strains is PAO1, the reference *P. aeruginosa* strain commonly used for basic research. Additionally, we published the

genomes of all these strains on NCBI (in the BioProject linked to this manuscript: PRJNA911168).

Considering…
1) the possible synergistic or additive effect of BX004-A and the concomitantly applied antibiotics (at least in the 4 tested strains),
2) the in vitro antibiotic sensitivity test results, suggesting a certain susceptibility of some of the patients' bacterial isolates to the concomitantly applied antibiotics aztreonam and tobramycin (no information with regard to colistin) in the BX004-A group, compared to aztreonam resistance only in the control group.

…in my opinion, it is more appropriate to state that the present clinical trial evaluated the microbiological efficacy of BX004-A in combination with one or more antibiotics (aztreonam, tobramycin, or colistin) compared to the antibiotic aztreonam alone, rather than BX004-A compared to placebo (cf. the abstract).

Please allow us to clarify that the study protocol evaluated BX004-A or placebo on top of standard of care inhaled antibiotics (tobramycin, aztreonam, or colistin, the 3 most commonly used antibiotics in CF patients). For the reasons stated earlier, the in vitro antibiotic sensitivity of *Pseudomonas* isolates to the standard of care inhaled antibiotics is not considered relevant for clinical efficacy in this population.

Reviewer #3 (Remarks to the Author):

In re-considering my critiques in light of the authors responses, my opinion is as follows:

1. While I did not think that the Pa sequencing and expression analysis added much to the study, little harm is done by including all this data (except perhaps having the reader wade through somewhat extraneous information).

2. While it would be more rigorous to include a positive control for the biofilm studies (and many biofilm studies use controls), little harm is done by omitting this.

3. However, I do not think it reasonable to conclude that they demonstrative antimicrobial efficacy (even with the "preliminary" qualifier) of the Bx004 cocktail for the following reasons (these points were largely ignored in the response):

a. As noted, both placebo subjects were not taking CFTR modulators while ~½ of the treatment group subjects were. This is a major problem as modulators are more effective against CF infections than any previously developed treatment. As noted, their approach would be akin to a cancer trial in which half of the treatment group had a new therapy added on to an extremely-effective, revolutionary treatment but the placebo group had neither therapy. I note that this critical information is not mentioned in the text, but rather buried in supplementary table-11 (even in the revision).

> We agree that CFTR modulators are an important covariate, and although randomization is one way to try to achieve even distributions of baseline covariates, its impact is more limited with small samples sizes as in Part 1 (n=7 on phage, n=2 on placebo). We updated the wording as noted below.

b. The efficacy claims are largely due to the fact that the placebo group showed increases in Pa density in the absence of any treatment. Thus, the comparisons touted between the untreated and treated groups are in large part due to a curious, unexplained, and perhaps spurious worsening in the placebo group. It appears possible/likely that much of the differences between treatment and placebo groups were driven by a single outliers in both groups at day 4 and 15.

c. Making matters worse, even a quick look at table S-15 (where average Pa density changes are reported) reveals two important findings. First, the authors' pre-specified scheme plan shows a minimal efficacy signal, however the data actually reported in the main text (Fig 6b) was obtained by swapping out the Pa baseline value of one placebo patient, to a value more favorable for a positive signal. The authors explanation for why they made this change did not make sense. If the sample is poor (because the subject inappropriately took antibiotics) this would be reflected in both the CFU and Pa relative abundance), so the finding that these parameters move in the same direction in no way justifies the switch in baseline data.

> Please allow us to clarify: one of the two placebo patients had taken their standard of care inhaled antibiotic within 60 minutes of sputum collection on Day 1, which the statistical analysis plan (SAP) pre-specified would lead to the Day 2 post-dose sputum sample being used as baseline CFU for calculations in any case where this applied. The reason that the SAP noted this was due to

a *potential* interference with an accurate assessment of sputum CFU level (due to presence of antibiotic possibly reducing CFU ex vivo in the sample), although this interference is not definitive or predictable; it is possible based on some observations. The directives to the sites to advise patients to delay inhaled antibiotic until after sputum collection were added during the course of the study when this was identified as a potential confounder, after this patient had enrolled and completed study drug (ie, there was no instruction at the time to prevent it from occurring in this patient, only a way to manage it programmatically via the SAP). This subject had a relatively low CFU measurement at Day 1 (6x10^4 CFU/g). Relative abundance of 16S ribosomal DNA, which is a molecular method that identifies DNA of both live and dead bacteria, showed low PsA levels in correlation with the CFU measurement. Based on this 16s analysis which showed an overall strong and significant correlation between *Pseudomonas* DNA relative abundance and Day 1 baseline CFU levels, the post-hoc analysis was conducted using the original Day 1 CFU level as the baseline value, as this analysis gives the most accurate scientific picture of the CFU trends based on all the available data. We updated the wording in the Table S15 footnote accordingly.

Without this post-hoc maneuver the day 4 difference between treatment and placebo groups changed from 1.92 logs (favoring the treatment group) to only 0.39 logs. The day 15 difference between treatment and placebo groups changed from 2.68 logs (favoring the treatment group) to 1.15 logs. This post-hoc change (which produces the appearance of efficacy) is irregular and is not reported to the reader in the main text. I wondered if the academic authors of the paper were aware of this post-hoc change in baseline value data, and how the appearance of a positive signal of efficacy was enhanced by it.

We appreciate this feedback and agree with the limitations imposed by the small sample size. Although we are internally excited by these results, we agree with the approach of toning down efficacy claims and have done so in this revised manuscript.

The abstract now reads:

"...BX004-A met the primary endpoints of safety and tolerability. Exploratory endpoints included pharmacokinetics and Pseudomonas aeruginosa sputum density reduction. Efficient phage delivery to the lower respiratory tract was observed, and a potential reduction in P. aeruginosa sputum burden was noted in the phage arm. However, due to the study's small sample size, definitive conclusions regarding efficacy are limited. These data pave the way toward further development of novel phage-based therapeutics in antibiotic-resistant pulmonary bacterial infections."

We added to this revised manuscript two additional subgroup analyses, where we can see CFU dynamics of patients taking different antibiotics and patients taking or not taking a CFTR modulator. We also added some language on the limitations in interpreting small studies:

"...Additionally, we assessed the impact of two potential confounders in the BX004-A arm: type of inhaled antibiotic and use of a CFTR modulator. Neither

subgroup analysis provided conclusive evidence of confounding effects (Fig. S10). However, given the small sample size of Part 1 of this first-in-human study, caution is warranted when interpreting efficacy-related outcomes. Therefore, both statistical significance and the impact of confounders should be monitored in larger studies."

Finally, we would like to emphasize that the data is accurate to the best of our ability. Given the scarcity of PK-PD data from placebo-controlled phage studies, we have exercised care to ensure high data quality and transparent reporting. This commitment ensures that our peers around the world have reliable access to the data, facilitating advancements in the field of phage therapy. We have been clear throughout the process on how we collected, analyzed, and visualized all the data presented in this study.

Reviewer #4 (Remarks to the Author):

The authors have toned down the efficacy statements, but not the safety statements. With only 7 patients in the treatment arm, one cannot conclude that "these data demonstrate the excellent safety profile". I do not have the expertise to judge whether the SAE reported in Table S12 is indeed unrelated to the study drug.

> We agree to tone down and updated the sentence: “these data demonstrate the *favorable* safety profile.”
>
> Additionally, the safety data in Table S15 were updated to include all data until the last 6-month safety follow-up. There were no premature discontinuations from study drug or study, no AEs of special interest, and no study drug-related AEs. The most common AEs (infective pulmonary exacerbation of CF, which occurred in 3 subjects and did not meet modified Fuchs’ criteria) occurred at least 52 days after last dose during the long-term safety follow-up and recovered / resolved.

Make clear in the abstract that this is a phase 1b trial. The sentence "Safety and microbiological efficacy of BX004-A were evaluated in a first-in-human double-blind placebo-controlled clinical trial" may suggest a phase 3 trial to the reader.

> Done; please allow us to clarify that the trial was a Phase 1b/2a trial divided into 2 Parts: Part 1 (presented here, with more intense safety monitoring, escalating doses, small sample size of 9 patients, and 7 days of study drug including 4 days of twice daily high dose phage) and Part 2 (to be presented in a future manuscript, with 10 days of twice daily high dose phage and larger sample size of 34 subjects). Although it may sound as if Part 1 was the Phase 1b portion and Part 2 was the Phase 2a portion, the study protocol did not categorize Part 1 and Part 2 in this manner; both Part 1 and Part 2 were parts of a Phase 1b/2a trial. Please see updated wording:
>
> “...*We evaluated BX004-A in Part 1 of a first-in-human double-blind placebo-controlled phase 1b/2a clinical trial, which included nine adult cystic fibrosis patients chronically infected with P. aeruginosa...”*

A few remaining comments:

1) Thanks for adding the p-values in figure 6. For 6B-D they are very useful. However, I don't see much relevance in giving p-values for Figure 6A. Could there be any explanation for having PFU larger than zero in the placebo group? Moreover, the assumptions for validity of the t-test are clearly violated.

> We agree. We removed the p-values from Figure 6A.

I guess log(PFU+1) was used, otherwise log(PFU) would give -infinity. Please make that clear.

> Our method was to compute the log on the native number (without adding +1), but instead to “manually fix” would be -inf cases to zero:

log10_PFU = 0 if total_PFU == 0 else np.log10(total_PFU)

We clarified in the SAP as follows: “For CFU and PFU values of 0, their log value will be substituted by 0, in order to include these values in the respective summary and analysis tables.”

In the legend of Figure 6 we address this: “*Y-axis values of zero represent no detectable phages in sample*”. In the relevant methods section (“Assessment of phage counts”) we added the above formula for clarity.

2) The values in Fig 1B have a very skewed distribution. Pearson correlation assumes a normal distribution. I suggest to use Spearman correlation. How many points are in the scatterplot? The labels suggest a different number (n=143 vs n=335)?

Thank you for this comment, we agree. We switched the metric to Spearman correlation and updated the correlation coefficient and p-value accordingly. We also clarified the values 143 and 335. These are the total sample size from which we calculate prevalence on each of the axes, in other words – each axis shows a % from a different total. We modified the axis labels to clarify this.

R.B.Geskus