Applied Microbiology and Biotechnology

Novel thermostable amine transferases from hot spring metagenomes

--Manuscript Draft--

Manuscript Number:	AMAB-D-16-02618R1
Full Title:	Novel thermostable amine transferases from hot spring metagenomes
Article Type:	Original Article
Section/Category:	Biotechnologically relevant enzymes and proteins
Corresponding Author:	Daniela Monti Istituto di Chimica del Riconoscimento Molecolare, CNR ITALY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Istituto di Chimica del Riconoscimento Molecolare, CNR
Corresponding Author's Secondary Institution:	
First Author:	Erica Elisa Ferrandi
First Author Secondary Information:	
Order of Authors:	Erica Elisa Ferrandi
	Alessandra Previdi
	Ivan Bassanini
	Sergio Riva
	Xu Peng
	Daniela Monti
Order of Authors Secondary Information:	
Funding Information:	European Commission (PIEF-GA-2013-6222732) — Dr Erica Elisa Ferrandi
Abstract:	Hot spring metagenomes, prepared from samples collected at temperatures ranging from 55°C to 95°C, were submitted to an in silico screening aimed at the identification of novel amine transaminases (ATAs), valuable biocatalysts for the preparation of optically pure amines. Three novel (S)-selective ATAs, namely Is3-TA, It6-TA, and B3-TA, were discovered in the metagenome of samples collected from hot springs in Iceland and in Italy, cloned from the corresponding metagenomic DNAs and over-expressed in recombinant form in E. coli. Functional characterization of the novel ATAs demonstrated that they all possess a thermophilic character and are capable of performing amine transfer reactions using a broad range of donor and acceptor substrates, thus suggesting a good potential for practical synthetic applications. In particular, the enzyme B3-TA revealed to be exceptionally thermostable, retaining 85% of activity after 5 days incubation at 80°C and more than 40% after two weeks under the same condition. These results, which were in agreement with the estimation of an apparent melting temperature around 88°C, make B3-TA, to the best of our knowledge, the most thermostable natural ATA described to date. This biocatalyst showed also a good tolerance toward different water-miscible and water-immiscible organic solvents. A detailed inspection of the homology-based structural model of B3-TA showed that the overall active site architecture of mesophilic (S)-selective ATAs was mainly conserved in this hyperthermophilic homologue. Additionally, a subfamily of B3-TA-like transaminases, mostly uncharacterized and all from thermophilic microorganisms, was identified and analyzed in terms of phylogenetic relationships and sequence conservation.

Response to Reviewers:	Replies to Reviewers' comments: Reviewer #1: This paper describes the isolation and characterization of three w-TAs from different metagonomic libraries obtained from thermophilic enviroments. The authors have discovered a highly stable w-TAs with high similarity to a w-TA recently isolated from Thermomicrobium roseum. Among the three w-TAs herein presented, B3-TA seems to be the most promising for further synthetic application since it was a highly thermostable enzyme although its substrate scope for the assymetric amination of ketones was quite limited, the authors did not even show the activity towards aromatic ketones. Although the discovered w-TAs do not represent a major advance in the state of the art of the transaminases in terms of substrate scope, the high stability of B3-TA makes this enzyme a good candidate to engineer its binding pocket for futher synthetic application. The paper is clear and the experiments are well conducted. I only have one major concern for the publication of this work. The authors claim that these w-TAs are S-selective enzymes according to their sequence similarities with other already described w-TAs, however they did not do any analysis to verify that this enzyme is in fact S-selective and thus did not show the enantiomeric excess. I believe that this data is important to better evalute the potential of these enzyme in synthetic chemistry. Another important point that is not clear in the manuscript is what the 100% is to calculate the relative activity of the different w-TAs towards different substrates. The missing of these data has not allowed me evaluating the absolute activity of the new w-TA regarding to the well characterized Vf-TA. A: The enantioselectivity of the novel ATAs has been evaluated by testing their activity on a set of either (R) or (S) aromatic amines. A new figure (Fig. 4 in the revised manuscript) reports the results obtained in this experiment, while the raw data have been included in the Supplementary Materials (Table S8). Experimental conditions are reported at page 9, lines 10-15, and the results are commented at page 16, lines 1-6. Vf-TA has been included in this study, thus the same figure shows also the comparison of the specific activity (U/mg) of the four enzymes on these model substrates. The obtained results clearly confirm the (S)-selectivity inferred by the sequence analysis. As far as the activity toward other substrates concern, i.e., those shown in Table 1, the presented data are not "relative activities", but real conversion values of the respective reactions determined by GC or HPLC analysis. For a better graphical representation of the results, relative activities have been used only for the preparation of Fig. 2a and 2b (determination of pH and temperature optima). In these cases, the maximum activity (optimum) for each enzyme was set as 100%. Also the raw data (U/mg at the different pHs or temperatures) used to prepare these figures have been added to the Supplementary Materials (Tables S4 and S5). Finally, I have some minor comments that need to be adressed by the authors: 1) Do the authors have any explanation for the dramatic activity decay observed in the inactivation of B3-TA in Figure 2C? This results is very strange, I am not sure whether this result is due to enzyme inactivation or to an activity assay artifact. A: The thermostability profile shown by B3-TA is typical of iperthermophilic enzymes which are characterized by the capability of maintaining the native structure up to temperatures very close to that of their original environment. These enzymes usually show very quick transitions between the folded and unfolded states, that correspond to the active and inactive forms of the enzymes, respectively. As the metagenome where B3-TA was identified was collected in hot spring sediments at 85°C and the apparent melting temperature of this enzyme was estimated around 88°C, it is not surprising that the activity is lost between 80 and 90°C. According also to the suggestion of the other Reviewer (see below), we have additionally estimated the residual activity after a thermal treatment at 85°C under the same experimental conditions used for the other temperatures. About 12% of the starting activity was still detectable after 3 hours. This result is now shown in Fig. 2c and in Table S6 in the Supplementary Materials (including standard deviations for each measurement). Moreover, as an additional prove of the structural change occurring to B3-TA at temperatures over 80°C, we have added to Fig. S5 (overlap of CD spectra recorded at 20/80°C) also the overlap of the

20/95°C CD spectra.

2) Pag 16, line 19, I guess that the authors means K289 instead L289 because to form the aldimine with PLP cofactor is required a lysine residue which is well conserved in many enzymes doing amine transfers.

A: We apologize for this mistake that has been corrected.

Reviewer #2:

The manuscript describes the identification of novel transaminases from metagenomic libraries. The work is well structured, with a lot of experimental data and the three new transaminases identified are quite interesting to certify the sufficient innovation for publication.
There are some minor comments that would help the manuscript to be more solid, that I would like them to be addressed before publication:

- Page 7: Please provide the experimental data of the alignment. What parameters / algorithm did you use? With what criteria were these transaminases (Table S3) were selected from the query?

A: The bioinformatic search for new ATAs was performed by aligning query sequences shown in Table S3 with the metagenomic database sequences using the program LAST, which is mainly based on the algorithms described in Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. Genome Res. 2011 21(3):487-93. This reference has been added to the Materials and methods (page 7, lines 14-15). Default settings were used during the alignments and the query sequences were selected on the basis of the availability of experimental evidences of their aminotransferase activity as well as their applicability in chiral amines synthesis.

- I miss standard deviation in several experiments, such as the CD melting temperature data (page 13), as well as the data of Table 1. We cannot comment if the new enzymes are better or not, if we do not have an idea if the difference is significant or not. Also, in the long term stability of B3-TA (Fig. 2d) there doesn't seem to be any SD, are the measurements so accurate? In the other frames of Figure 2 there are some data with SD and other not, and I suppose that the deviation in the ones that is not seen was really low and cannot be depicted due to the size of the dot. Also, in the frame c, the authors should probably measure another point between 80 and 90ºC to avoid presenting from 100 % activity a total inactivation.

A: Standard deviation of data obtained in different experiments have been added or presented in a more suitable form than in the submitted manuscript. Specifically, SD of melting temperatures data have been added (page 13, lines 26-27); a note has been added to Table 1 indicating that the SD was on the average below 5%; error bars have been added to Fig. 2d, while Fig. 2a-c have been redrawn by reducing the width of lines and the size of points, so that the error bars indicating the SD are now clearly visible for each point. Additionally, all the data (including SD) used to prepare Fig. 2a-d are now available in the Supplementary Materials (Tables S4-S7).
Moreover, as mentioned in the reply to the other Reviewer, the result of the thermostability assay of B3-TA at 85°C has been added to Fig. 2c.

- Two of the sequences came from acidic environments (Table S2), still the transaminases' optimal pH is at pH 9-10. Any comment on that? Maybe in asymmetric synthesis mode the optimal would be in a more acidic pH, which would suggest that this is their physiological reaction in the organism that produces them?

A: For the acidophiles and alkalophiles, it is only the proteins exported from the cell that have to be stable and active under the extreme pHs of the growth environment, since the proteins inside the cell do not have to withstand these extreme conditions as

the intracellular pH is maintained at values close to neutrality by maintaining suitable proton gradients. For details, the following reviews provide a detailed description of pH homeostasis mechanisms in different extremophilic microorganisms: Baker-Austin C & Dopson M (2007) Life in acid: pH homeostasis in acidophiles. Trends Microbiol. 15, 165–171; Krulwich TA, Sachs G & Padan E (2011) Molecular aspects of bacterial pH sensing and homeostasis. Nat. Rev. Microbiol. 9, 330–43. As far as the novel ATAs described in this paper concern, sequence analysis of the respective contigs regions flanking the genes did not show the presence of typical signal peptides coding sequences for protein secretion, thus it is likely that these enzymes are located and active in the cytoplasm of the original microbial producers, and not exported in the external environment.

- All reactions in kinetic resolution were performed with (S)-MBA. Did any reactions were performed with (R)-MBA, to evaluate the enantioselectivity of the enzyme? Or in the asymmetric synthesis reaction, did the authors use a method that the L-alanine was seperated from the D-alanine? Did they see any traces of D-alanine in any of their reactions? A comment on the enantioselectivity of the novel enzymes would be really interesting. On this point, it is interesting that in Table 1 some amino donors are shown as optically pure, while for others are there is no designation of the chirality. Where they used as racemic mixtures?

A: As reported in the replies to the other Reviewer (see above), the enantioselectivity of the novel ATAs has been evaluated by testing their activity on a set of either (R) or (S) aromatic amines, including (R)-MBA (see Fig. 4 in the revised manuscript and Table S8 in the Supplementary Materials). As far as the formation of D-alanine concerns, the derivatization method used in our lab (OPA reagent) is suitable for detection only using reverse phase HPLC, while chiral HPLC column usually work in direct phase. However, we have performed some preliminary studies of application of the novel enzymes in amines synthesis and we have been able to observe the formation of the desired products when using L-alanine as amine donor, while no product was formed in the presence of D-alanine, this fact being consistent with the (S)-selectivity observed with the aromatic amines. The stereochemistry of the compounds listed in Table 1 has been corrected when needed.

- With what rational were the specific set of organic solvents selected? Are they often used as co-solvents in transamination reactions? The authors should comment on that. I would also like to see the effect of 5-10 % isopropylamine (IPA), as this is an interesting amine donor. The authors mention that the thermostability could help on the use of IPA as amine donor, as the produced acetone could be removed easily at elevated temperatures, but they do not show if the IPA is accepted as amine donor, or if the enzymes are stable at high concentrations of IPA; both of which are real challenge in transaminase engineering.

A: Water-miscible organic solvents are often used in TA-catalyzed reactions to enhance the solubility of substrates and thereby increase the productivity of the biotransformations. The most commonly used organic solvents in ω-TA reactions are alcohols, acetonitrile and DMSO, usually at 5–20% (v/v) concentration. Examples can be found in the references of the manuscript, e.g., Chen et al. 2016; Koszelewski et al. 2008b and 2009; Mathew et al. 2016a and 2016b; Midelfort et al, 2013; Savile et al. 2010. As far as isopropylamine concerns, all the novel enzymes accepted it as amine donor (see conversions in Table 1 and page 15, lines 25-26). We agree with the Reviewer about the challenge of enzyme stability at high isopropylamine concentration. In fact, both Is3-TA and It6-TA showed <5% residual activity after 3 h of incubation in the presence of 5% IPA. The best result has been again obtained with B3-TA that, after the same time, retained about 30% of the starting activity at 5% IPA and 7 % at 10% IPA. Further studies about the optimization of process conditions are currently running in our lab and the results will be reported in the future.

- From Figure 1, and if I compare the data with the paper from Steffen-Munsberg et al. 2015, the It6-TA seems more like a β-Alanine-Pyruvate transaminase, rather than a high activity amine transaminase. Did the authors test this activity? I would suggest this

experiment.

A: We thank very much the Reviewer for noticing the similarity of active site conserved residues in It6-TA and β-Alanine-Pyruvate transaminases that we missed to show and discuss in the submitted manuscript. Therefore, Figure 1 has been modified by including the fingerprint sequence of β-Ala:pyr TA and a comment has been added to the text discussing this figure (from page 11, line 31 to page 12, line 3). Then, as suggested by the Reviewer, we have tested the activity of the novels ATAs and Vf-TA toward this substrate as amine donor. As shown in Table 1, not only It6-TA, but also the other two novel ATAs showed a very good activity in the presence of β-alanine, while, in agreement with literature data, Vf-TA was not able to convert this amine. The obtained results have been commented in the Results (page 15, lines 27-28) and Discussion (page 21, lines 13-22) sections.

- In the results (page 14-15) and discussion I would focus more on the interesting differences of the new enzymes, which is only presented in the last 3-4 lines before Table 1. To me it is interesting to note that these enzymes accept acetaldehyde quite well as well as propionaldehyde. Especially for acetaldehyde, how the authors managed to keep the aldehyde in the reaction (as the acetaldehyde is quite volatile). For this reason the SD that I asked before is really important. Other substrates that are interesting to discuss are 2-phenylethylamine propylamine and 1,3-diaminopropane. It would be interesting to make a discussion to compare with the literature, if such conversions were reported before from other transaminases.

A: As far as the activity towards aldehydes concerns, the results obtained with the novel enzymes are basically consistent with the data available in the literature for other ATAs. In fact, it is not uncommon to observe good activity with aldehydes, in many cases even better than that observed with corresponding ketones. See for examples: Shin J-S, Kim B-G (2001) Comparison of the ω-transaminases from different microorganisms and application to production of chiral amines. Biosci Biotechnol Biochem 65:1782–1788; Cerioli L, Planchestainer M, Cassidy J, Tessaro D, Paradisi F (2015) Characterization of a novel amine transaminase from Halomonas elongata. J Mol Catal B Enzym 120:141–150, or among the references of the manuscript: Shin J, Kim B (2002); Chen et al 2016; Park et al.(2012). To keep the integrity and amount of substrates during the biotransformation course, all the reactions were performed in sealed vials at 30°C. As indicated in the footnote of Table 1, the conversions standard deviation were <5% with all the tested substrates. Concerning primary amines and diamines, there are not many information available in the literature as they have not been tested or ATAs don't use them. However, some of them could be cheap amine donors. The activity of a taurine–pyruvate transaminase from Geobacillus thermodenitrificans on 2-phenylethylamine was recently reported (ref. Chen et al. 2016 in this manuscript), while different diamines have been recently investigated as donors (see for example: Green AP et al., (2014) Angew Chemie - Int Ed 53:10714–10717; Gomm A et al., (2016) Chem - A Eur J 22:12692–12695; Galman JL et al., (2017) Green Chem 9:285–288). However, in the latter cases, the application of more specialized enzymes, i.e., putrescine transaminases, looks more promising at the moment.

- Page 16: If the authors used Vfl as template for the homology model, it is sure that the model with not differentiate from the template too much. It would be preferred to prepare a hybrid model, as the identity with Vfl is quite low and the accuracy of the model can be questionable in non-conserved areas.

A: For the obtainment of the homology model of B3-TA, a template search was performed with Blast and HHBlits against the SWISS-MODEL template library (SMTL). First, the target sequence was searched with BLAST against the primary amino acid sequence contained in the SMTL. A total of 116 templates were found. An initial HHblits profile has been built using the procedure outlined in Remmert, et al., 2011, followed by 1 iteration of HHblits against NR20. The obtained profile has then be searched against all profiles of the SMTL. A total of 719 templates were found. Thus, overall 835 templates were found. Model quality estimation was automatically performed by SWISS-MODEL to rank the templates. The Vf-TA structure (PDB: 4E3Q)

	chosen by us to build the model ranked at the 10th position of these 835 possible templates, the first nine positions of the ranking been occupied by (putative) transaminases with similarity close to that shown by 4E3Q (around 40%). The quality scores (QMEAN and GMQE) for the model based on 4E3Q indicate a high reliability. A detailed description of how the SWISS-MODEL server works is reported in ref. Biasini et al. 2014. More information about the method and a Table containing a list of the first thirty ranked templates have been added to Figure S7 in the Supplementary Materials. - Page 16, line 18: I suppose that the authors mean K289, as there cannot be internal aldimine with a leucine residue. A: We apologize for this mistake that has been corrected. - Page 21, lines 17-27: The enlargement of the binding pocket may be a challenged, but it is addressed in several publications that the authors also mention.For instance, Pavlidis et al. 2016 suggested a motif that can be easily transferable in (S)-selective transaminases, to enable the acceptance of bulky substrates. Can the authors comment how many of these suggested residues exist in the identified transaminases and what the overall similarity to 3FCR structure is? A: The overall similarity of the novel ATAs in the respect of the ATA from Silicibacter sp. (PDB: 3FCR) has been estimated by BLASTP analysis and resulted: 34% for It6-TA (92% query cover); 38% for Is3-TA (97% query cover); 40% for B3-TA (98% query cover). Thus, the novel ATAs are quite dissimilar from 3FCR in their sequences. However, it is worth noting that 3FCR ranked very close to 4E3Q (V. fluvialis ATA) in the SWISS MODEL template search carried out in the building of the 3D structural model of B3-TA (see the revised Fig. S7 in the Supplementary Materials). The residues of 3CFR targeted in Pavlidis et al. 2016, that allowed the obtainment of variants with improved acceptance of bulky substrates, i.e., Y59, Y87, Y152, and T231, resemble, for all the three novel enzymes, the same pattern of V. fluvialis ATA (Y59 is W, Y87 is F, Y152 is conserved, T231 is A, with the only exception of It6-TA, where is S). This is in accordance with the very limited acceptance of bulky substrates shown by the novel ATAs. To provide sufficient amount to the authors to improve the manuscript, I would suggest major revision, but I support the publication of an improved version of this manuscript.

Istituto di Chimica del Riconoscimento Molecolare, C. N. R.

Dr. Daniela Monti

Via Mario Bianco 9, 20131 Milano, Italy - Telephone: +39 02 285 000 25
Fax: +39 02 289 012 39 - E-mail: daniela.monti@icrm.cnr.it

Milano, February 20th 2017

Ref.: Ms. No. AMAB-D-16-02618
Novel thermostable amine transferases from hot spring metagenomes
Submitted to Applied Microbiology and Biotechnology

Dear Prof Schmidt-Dannert,

please find enclosed a revision of the manuscript titled “Novel thermostable amine transferases from hot spring metagenomes” by Erica Elisa Ferrandi, Alessandra Previdi, Ivan Bassanini, Sergio Riva, Xu Peng, and myself (corresponding authors: E. E. Ferrandi and myself), to be evaluated for publication in your Journal in the form of ‘Original Paper’.

We kindly acknowledge the Reviewers for careful reading and for the overall positive evaluation of our manuscript, as well as for providing us valuable suggestions to improve the presentation and completeness of our results.

According to their suggestions, we have performed some additional experimental work that has been added to the manuscript and tried to reply to all their questions and comments as in the following.

We really hope that in this form this manuscript will be suitable for publication in Applied Microbiology and Biotechnology.

Yours sincerely,
Daniela Monti
(on behalf of all the other Authors)

Replies to Reviewers' comments:

Reviewer #1:

This paper describes the isolation and characterization of three w-TAs from different metagonomic libraries obtained from thermophilic enviroments. The authors have discovered a highly stable w-TAs with high similarity to a w-TA recently isolated from Thermomicrobium roseum. Among the three w-TAs herein presented, B3-TA seems to be the most promising for further synthetic application since it was a highly thermostable enzyme although its substrate scope for the assymetric amination of ketones was quite limited, the authors did not even show the activity towards aromatic ketones. Although the discovered w-TAs do not represent a major advance in the state of the art of the transaminases in terms of substrate scope, the high stability of B3-TA makes this enzyme a good candidate to engineer its binding pocket for futher synthetic application. The paper is clear and the experiments are well conducted. I only have one major concern for the publication of this work.
The authors claim that these w-TAs are S-selective enzymes according to their sequence similarities with other already described w-TAs, however they did not do any analysis to verify that this enzyme is in fact S-selective and thus did not show the enantiomeric excess. I believe that this data is important to better evalute the potential of these enzyme in synthetic chemistry. Another important point that is not clear in the manuscript is what the 100% is to calculate the relative activity of the different w-TAs towards different substrates. The missing of these data has not allowed me evaluating the absolute activity of the new w-TA regarding to the well characterized Vf-TA.

A: The enantioselectivity of the novel ATAs has been evaluated by testing their activity on a set of either (*R*) or (*S*) aromatic amines. A new figure (Fig. 4 in the revised manuscript) reports the results obtained in this experiment, while the raw data have been included in the Supplementary Materials (Table S8). Experimental conditions are reported at page 9, lines 10-15, and the results are commented at page 16, lines 1-6. Vf-TA has been included in this study, thus the same figure shows also the comparison of the specific activity (U/mg) of the four enzymes on these model substrates. The obtained results clearly confirm the (*S*)-selectivity inferred by the sequence analysis. As far as the activity toward other substrates concern, i.e., those shown in Table 1, the presented data are not "relative activities", but real conversion values of the respective reactions determined by GC or HPLC analysis. For a better graphical representation of the results, relative activities have been used only for the preparation of Fig. 2a and 2b (determination of pH and temperature optima). In these cases, the maximum activity (optimum) for each enzyme was set as 100%. Also the raw data (U/mg at the different pHs or temperatures) used to prepare these figures have been added to the Supplementary Materials (Tables S4 and S5).

Finally, I have some minor comments that need to be adressed by the authors:
1) Do the authors have any explanation for the dramatic activity decay observed in the inactivation of B3-TA in Figure 2C? This results is very strange, I am not sure whether this result is due to enzyme inactivation or to an activity assay artifact.

A: The thermostability profile shown by B3-TA is typical of iperthermophilic enzymes which are characterized by the capability of maintaining the native structure up to temperatures very close to that of their original environment. These enzymes usually show very quick transitions between the

folded and unfolded states, that correspond to the active and inactive forms of the enzymes, respectively. As the metagenome where B3-TA was identified was collected in hot spring sediments at 85°C and the apparent melting temperature of this enzyme was estimated around 88°C, it is not surprising that the activity is lost between 80 and 90°C. According also to the suggestion of the other Reviewer (see below), we have additionally estimated the residual activity after a thermal treatment at 85°C under the same experimental conditions used for the other temperatures. About 12% of the starting activity was still detectable after 3 hours. This result is now shown in Fig. 2c and in Table S6 in the Supplementary Materials (including standard deviations for each measurement). Moreover, as an additional prove of the structural change occurring to B3-TA at temperatures over 80°C, we have added to Fig. S5 (overlap of CD spectra recorded at 20/80°C) also the overlap of the 20/95°C CD spectra.

2) Pag 16, line 19, I guess that the authors means K289 instead L289 because to form the aldimine with PLP cofactor is required a lysine residue which is well conserved in many enzymes doing amine transfers.

A: We apologize for this mistake that has been corrected.

Reviewer #2:

The manuscript describes the identification of novel transaminases from metagenomic libraries. The work is well structured, with a lot of experimental data and the three new transaminases identified are quite interesting to certify the sufficient innovation for publication.
There are some minor comments that would help the manuscript to be more solid, that I would like them to be addressed before publication:

- Page 7: Please provide the experimental data of the alignment. What parameters / algorithm did you use? With what criteria were these transaminases (Table S3) were selected from the query?

A: The bioinformatic search for new ATAs was performed by aligning query sequences shown in Table S3 with the metagenomic database sequences using the program LAST, which is mainly based on the algorithms described in Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. Genome Res. 2011 21(3):487-93. This reference has been added to the Materials and methods (page 7, lines 14-15). Default settings were used during the alignments and the query sequences were selected on the basis of the availability of experimental evidences of their aminotransferase activity as well as their applicability in chiral amines synthesis.

- I miss standard deviation in several experiments, such as the CD melting temperature data (page 13), as well as the data of Table 1. We cannot comment if the new enzymes are better or not, if we do not have an idea if the difference is significant or not. Also, in the long term stability of B3-TA (Fig. 2d) there doesn't seem to be any SD, are the measurements so accurate? In the other frames of Figure 2 there are some data with SD and other not, and I suppose that the deviation in the ones that is not seen was really low and cannot be depicted due to the size of the dot. Also, in the frame c, the authors should probably measure another point between 80 and 90ºC to avoid presenting from 100 % activity a total inactivation.

A: Standard deviation of data obtained in different experiments have been added or presented in a more suitable form than in the submitted manuscript. Specifically, SD of melting temperatures data have been added (page 13, lines 26-27); a note has been added to Table 1 indicating that the SD was on the average below 5%; error bars have been added to Fig. 2d, while Fig. 2a-c have been redrawn by reducing the width of lines and the size of points, so that the error bars indicating the SD are now clearly visible for each point. Additionally, all the data (including SD) used to prepare Fig. 2a-d are now available in the Supplementary Materials (Tables S4-S7).
Moreover, as mentioned in the reply to the other Reviewer, the result of the thermostability assay of B3-TA at 85°C has been added to Fig. 2c.

- Two of the sequences came from acidic environments (Table S2), still the transaminases' optimal pH is at pH 9-10. Any comment on that? Maybe in asymmetric synthesis mode the optimal would be in a more acidic pH, which would suggest that this is their physiological reaction in the organism that produces them?

A: For the acidophiles and alkalophiles, it is only the proteins exported from the cell that have to be stable and active under the extreme pHs of the growth environment, since the proteins inside the cell do not have to withstand these extreme conditions as the intracellular pH is maintained at values close to neutrality by maintaining suitable proton gradients. For details, the following reviews provide a detailed description of pH homeostasis mechanisms in different extremophilic microorganisms: Baker-Austin C & Dopson M (2007) Life in acid: pH homeostasis in acidophiles. *Trends Microbiol.* **15**, 165–171; Krulwich TA, Sachs G & Padan E (2011) Molecular aspects of bacterial pH sensing and homeostasis. *Nat. Rev. Microbiol.* **9**, 330–43. As far as the novel ATAs described in this paper concern, sequence analysis of the respective contigs regions flanking the genes did not show the presence of typical signal peptides coding sequences for protein secretion, thus it is likely that these enzymes are located and active in the cytoplasm of the original microbial producers, and not exported in the external environment.

- All reactions in kinetic resolution were performed with (S)-MBA. Did any reactions were performed with (R)-MBA, to evaluate the enantioselectivity of the enzyme? Or in the asymmetric synthesis reaction, did the authors use a method that the L-alanine was seperated from the D-alanine? Did they see any traces of D-alanine in any of their reactions? A comment on the enantioselectivity of the novel enzymes would be really interesting. On this point, it is interesting that in Table 1 some amino donors are shown as optically pure, while for others are there is no designation of the chirality. Where they used as racemic mixtures?

A: As reported in the replies to the other Reviewer (see above), the enantioselectivity of the novel ATAs has been evaluated by testing their activity on a set of either (*R*) or (*S*) aromatic amines, including (*R*)-MBA (see Fig. 4 in the revised manuscript and Table S8 in the Supplementary Materials). As far as the formation of D-alanine concerns, the derivatization method used in our lab (OPA reagent) is suitable for detection only using reverse phase HPLC, while chiral HPLC column usually work in direct phase. However, we have performed some preliminary studies of application of the novel enzymes in amines synthesis and we have been able to observe the formation of the desired products when using L-alanine as amine donor, while no product was formed in the presence of D-alanine, this fact being consistent with the (*S*)-selectivity observed with the aromatic amines. The stereochemistry of the compounds listed in Table 1 has been corrected when needed.

- With what rational were the specific set of organic solvents selected? Are they often used as co-solvents in transamination reactions? The authors should comment on that. I would also like to see the effect of 5-10 % isopropylamine (IPA), as this is an interesting amine donor. The authors mention that the thermostability could help on the use of IPA as amine donor, as the produced acetone could be removed easily at elevated temperatures, but they do not show if the IPA is accepted as amine donor, or if the enzymes are stable at high concentrations of IPA; both of which are real challenge in transaminase engineering.

A: Water-miscible organic solvents are often used in TA-catalyzed reactions to enhance the solubility of substrates and thereby increase the productivity of the biotransformations. The most commonly used organic solvents in ω-TA reactions are alcohols, acetonitrile and DMSO, usually at 5–20% (v/v) concentration. Examples can be found in the references of the manuscript, e.g., Chen et al. 2016; Koszelewski et al. 2008b and 2009; Mathew et al. 2016a and 2016b; Midelfort et al, 2013; Savile et al. 2010. As far as isopropylamine concerns, all the novel enzymes accepted it as amine donor (see conversions in Table 1 and page 15, lines 25-26). We agree with the Reviewer about the challenge of enzyme stability at high isopropylamine concentration. In fact, both Is3-TA and It6-TA showed <5% residual activity after 3 h of incubation in the presence of 5% IPA. The best result has been again obtained with B3-TA that, after the same time, retained about 30% of the starting activity at 5% IPA and 7 % at 10% IPA. Further studies about the optimization of process conditions are currently running in our lab and the results will be reported in the future.

- From Figure 1, and if I compare the data with the paper from Steffen-Munsberg et al. 2015, the It6-TA seems more like a β-Alanine-Pyruvate transaminase, rather than a high activity amine transaminase. Did the authors test this activity? I would suggest this experiment.

A: We thank very much the Reviewer for noticing the similarity of active site conserved residues in It6-TA and β-Alanine-Pyruvate transaminases that we missed to show and discuss in the submitted manuscript. Therefore, Figure 1 has been modified by including the fingerprint sequence of β-Ala:pyr TA and a comment has been added to the text discussing this figure (from page 11, line 31 to page 12, line 3). Then, as suggested by the Reviewer, we have tested the activity of the novels ATAs and Vf-TA toward this substrate as amine donor. As shown in Table 1, not only It6-TA, but also the other two novel ATAs showed a very good activity in the presence of β-alanine, while, in agreement with literature data, Vf-TA was not able to convert this amine. The obtained results have been commented in the Results (page 15, lines 27-28) and Discussion (page 21, lines 13-22) sections.

- In the results (page 14-15) and discussion I would focus more on the interesting differences of the new enzymes, which is only presented in the last 3-4 lines before Table 1. To me it is interesting to note that these enzymes accept acetaldehyde quite well as well as propionaldehyde. Especially for acetaldehyde, how the authors managed to keep the aldehyde in the reaction (as the acetaldehyde is quite volatile). For this reason the SD that I asked before is really important. Other substrates that are interesting to discuss are 2-phenylethylamine propylamine and 1,3-diaminopropane. It would be interesting to make a discussion to compare with the literature, if such conversions were reported before from other transaminases.

A: As far as the activity towards aldehydes concerns, the results obtained with the novel enzymes are basically consistent with the data available in the literature for other ATAs. In fact, it is not uncommon to observe good activity with aldehydes, in many cases even better than that observed with corresponding ketones. See for examples: Shin J-S, Kim B-G (2001) Comparison of the ω-transaminases from different microorganisms and application to production of chiral amines. *Biosci Biotechnol Biochem* 65:1782–1788; Cerioli L, Planchestainer M, Cassidy J, Tessaro D, Paradisi F (2015) Characterization of a novel amine transaminase from Halomonas elongata. *J Mol Catal B Enzym* 120:141–150, or among the references of the manuscript: Shin J, Kim B (2002); Chen et al 2016; Park et al.(2012). To keep the integrity and amount of substrates during the biotransformation course, all the reactions were performed in sealed vials at 30°C. As indicated in the footnote of Table 1, the conversions standard deviation were <5% with all the tested substrates. Concerning primary amines and diamines, there are not many information available in the literature as they have not been tested or ATAs don't use them. However, some of them could be cheap amine donors. The activity of a taurine–pyruvate transaminase from *Geobacillus thermodenitrificans* on 2-phenylethylamine was recently reported (ref. Chen et al. 2016 in this manuscript), while different diamines have been recently investigated as donors (see for example: Green AP et al., (2014) *Angew Chemie - Int Ed* 53:10714–10717; Gomm A et al., (2016) *Chem - A Eur J* 22:12692–12695; Galman JL et al., (2017) *Green Chem* 9:285–288). However, in the latter cases, the application of more specialized enzymes, i.e., putrescine transaminases, looks more promising at the moment.

- Page 16: If the authors used Vfl as template for the homology model, it is sure that the model with not differentiate from the template too much. It would be preferred to prepare a hybrid model, as the identity with Vfl is quite low and the accuracy of the model can be questionable in non-conserved areas.

A: For the obtainment of the homology model of B3-TA, a template search was performed with Blast and HHBlits against the SWISS-MODEL template library (SMTL). First, the target sequence was searched with BLAST against the primary amino acid sequence contained in the SMTL. A total of 116 templates were found. An initial HHblits profile has been built using the procedure outlined in Remmert, et al., 2011, followed by 1 iteration of HHblits against NR20. The obtained profile has then be searched against all profiles of the SMTL. A total of 719 templates were found. Thus, overall 835 templates were found. Model quality estimation was automatically performed by SWISS-MODEL to rank the templates. The Vf-TA structure (PDB: 4E3Q) chosen by us to build the model ranked at the 10th position of these 835 possible templates, the first nine positions of the ranking been occupied by (putative) transaminases with similarity close to that shown by 4E3Q (around 40%). The quality scores (QMEAN and GMQE) for the model based on 4E3Q indicate a high reliability. A detailed description of how the SWISS-MODEL server works is reported in ref. Biasini et al. 2014. More information about the method and a Table containing a list of the first thirty ranked templates have been added to Figure S7 in the Supplementary Materials.

- Page 16, line 18: I suppose that the authors mean K289, as there cannot be internal aldimine with a leucine residue.

A: We apologize for this mistake that has been corrected.

- Page 21, lines 17-27: The enlargement of the binding pocket may be a challenged, but it is addressed in several publications that the authors also mention.For instance, Pavlidis et al. 2016 suggested a motif that can be easily transferable in (S)-selective transaminases, to enable the acceptance of bulky substrates. Can the authors comment how many of these suggested residues exist in the identified transaminases and what the overall similarity to 3FCR structure is?

A: The overall similarity of the novel ATAs in the respect of the ATA from *Silicibacter* sp. (PDB: 3FCR) has been estimated by BLASTP analysis and resulted: 34% for It6-TA (92% query cover); 38% for Is3-TA (97% query cover); 40% for B3-TA (98% query cover). Thus, the novel ATAs are quite dissimilar from 3FCR in their sequences. However, it is worth noting that 3FCR ranked very close to 4E3Q (*V. fluvialis* ATA) in the SWISS MODEL template search carried out in the building of the 3D structural model of B3-TA (see the revised Fig. S7 in the Supplementary Materials). The residues of 3CFR targeted in Pavlidis et al. 2016, that allowed the obtainment of variants with improved acceptance of bulky substrates, i.e., Y59, Y87, Y152, and T231, resemble, for all the three novel enzymes, the same pattern of *V. fluvialis* ATA (Y59 is W, Y87 is F, Y152 is conserved, T231 is A, with the only exception of It6-TA, where is S). This is in accordance with the very limited acceptance of bulky substrates shown by the novel ATAs.

To provide sufficient amount to the authors to improve the manuscript, I would suggest major revision, but I support the publication of an improved version of this manuscript.

Erica Elisa Ferrandi,[*,1,2] Alessandra Previdi,[1] Ivan Bassanini,[1,3] Sergio Riva,[1] Xu Peng,[2] and Daniela Monti[*,1]

Novel thermostable amine transferases from hot spring metagenomes

[1] Istituto di Chimica del Riconoscimento Molecolare, C.N.R., Via Mario Bianco 9, 20131 Milano, Italy.

[2] University of Copenhagen, Ole Maaløesvej 5, 2200 Copenhagen N, Denmark.

[3] Università degli Studi di Milano, Dipartimento di Chimica, Via Golgi 19, 20133 Milano, Italy.

Corresponding authors

Daniela Monti

e-mail: daniela.monti@icrm.cnr.it; Fax: +39 02 28901239; Tel: +39 02 28500025

Erica Elisa Ferrandi

e-mail: erica.ferrandi@icrm.cnr.it; Fax: +39 02 28901239; Tel: +39 02 28500028

Abstract

Hot spring metagenomes, prepared from samples collected at temperatures ranging from 55°C to 95°C, were submitted to an *in silico* screening aimed at the identification of novel amine transaminases (ATAs), valuable biocatalysts for the preparation of optically pure amines.

Three novel (*S*)-selective ATAs, namely Is3-TA, It6-TA, and B3-TA, were discovered in the metagenome of samples collected from hot springs in Iceland and in Italy, cloned from the corresponding metagenomic DNAs and over-expressed in recombinant form in *E. coli*.

Functional characterization of the novel ATAs demonstrated that they all possess a thermophilic character and are capable of performing amine transfer reactions using a broad range of donor and acceptor substrates, thus suggesting a good potential for practical synthetic applications.

In particular, the enzyme B3-TA revealed to be exceptionally thermostable, retaining 85% of activity after 5 days incubation at 80°C and more than 40% after two weeks under the same condition. These results, which were in agreement with the estimation of an apparent melting temperature around 88°C, make B3-TA, to the best of our knowledge, the most thermostable natural ATA described to date. This biocatalyst showed also a good tolerance toward different water-miscible and water-immiscible organic solvents.

A detailed inspection of the homology-based structural model of B3-TA showed that the overall active site architecture of mesophilic (*S*)-selective ATAs was mainly conserved in this hyperthermophilic homologue. Additionally, a subfamily of B3-TA-like transaminases, mostly uncharacterized and all from thermophilic microorganisms, was identified and analyzed in terms of phylogenetic relationships and sequence conservation.

Keywords: Amine transferases, metagenomics, industrial biocatalysis, chiral amines, thermostability

Introduction

Transaminases (TAs, EC 2.6.1.x) are pyridoxal-5'-phosphate (PLP)-dependent enzymes capable of transfering an amino group between an amino donor and a prochiral ketone substrate, thus creating a novel stereogenic center in the transamination reaction. These enzymes are ubiquitous in nature and play a key role in the nitrogen metabolism (Koszelewski et al. 2010; Mathew and Yun 2012).

Generally, TAs are classified into α-TAs, representing the majority of TAs and converting only α-amino and α-keto acids, and ω-TAs, which can accept amino acids having a distal carboxylic acid group as substrates (Mathew and Yun 2012; Fuchs et al. 2015). Among ω-TAs, the class of amine transferases (ATAs) are of particular interest for biotechnological applications as they can also convert amines lacking a carboxylic acid group in addition to α-aminoacids and ω-aminoacids. The use of transaminases for biocatalytic applications has been largely investigated in the last years to provide new sustainable production processes for the preparation of drugs and chemical intermediates and, particularly, to avoid the current use of metal catalysts (i.e., Ru, Pd, Ni) at high pressure in chiral amine chemical syntheses (Nugent and El-Shazly 2010) .

Both (*S*)- and (*R*)-selective ATAs have been found in Nature and exploited for the kinetic resolution of racemic amines as well as for the synthesis of chiral amines starting from the corresponding ketones (Koszelewski et al. 2010; Monti et al. 2015; Steffen-Munsberg et al. 2015). For example, ATAs have been used for the stereoselective synthesis of (*R*)-4-phenylbutan-2-amine, a precursor of the anti-hypertensive dilevalol, and of (*R*)-1-(4-methoxyphenyl)ethylamine, an important building block for the biologically active compound formoterolo (Koszelewski et al. 2008a; Koszelewski et al. 2008b). In another example, both enantiomers of mexiletine [1-(2,6-dimethylphenoxy)-2-propanamine], a chiral antiarrhythmic agent, have been prepared by deracemization starting from the commercially available racemic amine using two different enantioselective ATAs (Koszelewski et al. 2009).

However, many challenges inherent to transaminase stability under industrial process conditions still remain to be tackled before developing large-scale processes. In fact, transamination processes on industrial scale often require the use of organic cosolvents, to increase substrate solubility (Savile et al. 2010), and high temperatures, for example for the stripping of volatile co-

products, i.e., acetone, to shift the equilibrium of the transamination reaction toward product formation (Martin et al. 2007).

At present the array of ATAs from natural sources that have been investigated for their synthetic application is still quite restricted, comprising around 20 different (*S*)-selective ATAs, mainly from bacterial strains (Koszelewski et al. 2010), and around 20 different (*R*)-selective ATAs, mainly from fungi (Höhne et al. 2010; Sayer et al. 2014). In a few cases, ATAs suitable for industrial application have been obtained by protein engineering of existing ATAs, the most remarkable example having been reported by Merck and Codexis on the development of a ATA-catalyzed process for the large-scale manufacture of the antidiabetic compound sitagliptin (Savile et al. 2010). In this case, a highly stable and active (*R*)-selective ATA was obtained after 11 rounds of directed evolution and successfully applied for the industrial-scale conversion of prositagliptin ketone to the optically pure sitagliptin.

Biocatalysts produced by thermophiles often show attractive inherent properties such as high stability in the presence of organic solvent, under operational conditions and during long-term storage (de Miguel Bouzas et al. 2006; Littlechild 2015). When this investigation was started, no naturally thermostable ATAs had been previously described. Recently, the search of transaminases in thermophiles genomes has revealed to be a promising tool to find novel enzymes that might be suitable for industrial applications. Specifically, a taurine:pyruvate TA showing highest activity at 65°C and good stability at up to 20% (v/v) MeOH concentration has been identified from *Geobacillus thermodenitrificans* (Chen et al. 2016), while a TA from *Sphaerobacter thermophilus*, with a strict substrate specificity towards β- and γ-amino acids, has proved to be stable when incubated for 1 h at temperature up to 60°C (Mathew et al. 2016b). Even more recently, a (*S*)-selective ATA with a broad substrate scope and showing a remarkable thermostability has been identified from *Thermomicrobium roseum* and successfully exploited in asymmetric synthesis and kinetic resolution processes at high temperatures (Mathew et al. 2016a).

As an alternative to selected genomes mining, we have herein exploited metagenomics, i.e. the study of the genetic material (metagenome) recovered directly from environmental samples (Lorenz and Schleper 2002; Fernández-Arrojo et al. 2010; Ferrer et al. 2016), to find new thermostable ATAs. The potential of this innovative approach for the identification of novel active and stable biocatalysts has been recently demonstrated with enzymes of different classes

(Wilson and Piel 2013; Ferrandi et al. 2015; DeCastro et al. 2016; Zarafeta et al. 2016). In this work, we report on the discovery and functional characterization of three new ATAs from metagenomic samples collected in hot springs at very high temperature (76-90°C), ideal environments for (hyper)thermophilic microorganisms.

Materials and methods

Chemicals

Amino donors, amino acceptors, PLP, tryptone, yeast extract, agarose were purchased from Sigma-Aldrich (St Louis, MO, USA). Isopropyl-β-D-thiogalactopyranoside (IPTG) was obtained from VWR (Radnor, PA, USA). All other reagents were of analytical grade and commercially available.

Analytical methods

GC-MS analyses were performed using a HP-5MS column (30 m × 0.25 mm × 0.25 μm, Agilent) on a Finnigan TRACE DSQ GC/MS instrument (ThermoQuest, San Jose, CA). The samples of amine derivatives were acetylated before injection as follows: transamination reactions (0.5 mL) catalyzed by TAs were recovered after 24 h, then reaction pH was adjusted to 11.0 by adding 6 M NaOH (50 µL) and products were extracted with ethyl acetate (350 µL). To 100 µL of organic phase dried over anhydrous Na_2SO_4, pyridine and acetic anhydride (15 µL each) were subsequently added and the mixture was kept overnight at room temperature before analysis. The following temperature program was employed: 60°C (1 min) / 6°C min-1/ 150°C (1 min) /12°C min-1/ 210°C (5 min).

HPLC analyses were conducted using a Jasco 880-PU pump equipped with a Jasco 875-UV/Vis detector and a Kinetex 5µm EVO C18 100Å 150 × 4.6 mm column (Phenomenex, Torrance, CA). L-Alanine derivatization was achieved by diluting 10 x the reaction samples (50 µL) in 1 mM HCl and then adding 50 µL of phthaldialdehyde (OPA) reagent (Sigma-Aldrich). The mobile phase consisted of MeOH : THF : 50 mM phosphoric acid, (20:20:960), pH adjusted to 7.5 with NaOH (eluent A), and MeOH : H_2O, (65:35) (eluent B). L-alanine-OPA derivative was eluted at a flow rate of 0.5 mL min^{-1} with the following gradient: t = 0 min, 80% A; t = 17 min,

70% A; t = 32 min, 20% A; t = 37 min, 20% A; t = 47 min, 80% A. Peaks were detected at 340 nm and calibration curves were prepared using standard solutions of L-alanine-OPA derivative.

CD spectra were recorded on a nitrogen flushed Jasco J-1100 spectropolarimeter (Easton, MD, USA) interfaced with a thermostatically controlled cell holder. CD analysis was performed with purified protein samples dialyzed overnight against 50 mM potassium phosphate (KP) buffer, pH 9.0, and diluted in degassed water (0.15 mg mL^{-1} final concentration for far-UV analysis, 1 mg mL^{-1} final concentration for near-UV analysis) in quartz cuvettes with 0.1 cm path length (far-UV analysis) or 1 cm path lenght (near-UV analysis). CD spectra were recorded in the range between 185 nm and 250 nm (far-UV analysis) or from 250 nm and 350 nm (near-UV analysis) at 20°C or 95°C.

For the determination of apparent T_M variation of CD signal at 220 nm (far-UV) or 300 nm (near -UV), the following temperature programs were used:

IS3-TA and B3-TA sample

20°C up to 65°C at 5°C/min_ data pitch each 2°C, hold 30''

65°C up to 90°C at 2.5°C/min_data pitch each 0.5°C, hold 30''

90°C up to 95°C at 5°C/min pitch data each 2°C, hold 30''

It6-TA sample

20°C up to 55°C at 5°C/min_ data pitch each 2°C, hold 30''

55°C up to 85°C at 2.5°C/min_data pitch each 0.5°C, hold 30''

85°C up to 95°C at 5°C/min pitch data each 2°C, hold 30''

General molecular biology techniques

Gene cloning into the pJet1.2 vector was performed with the CloneJET PCR Cloning Kit (Thermo Fischer Scientific, Waltham, MA, USA). Plasmid DNA was purified by using the E.Z.N.A. Plasmid Mini kit II (Omega/VWR). Plasmids were transformed in *E.coli* DH5α or *E.coli* Rosetta by using standard techniques (Sambrook et al. 1989). Gene cloning in the pETite C-His Kan vector and *E. coli* 10G chemically competent cell transformations were carried out with the Expresso T7 Cloning and Expression kit from Lucigen (Middleton, WI, USA). If not stated otherwise, standard PCR amplifications were carried out on 25 µL reaction mixtures containing plasmid DNA (5 ng) or metagenomic DNA (50-100 ng), primers (1 µM each), dNTPs (0.2 mM each), 1.25 U of *Pfu* DNA polymerase and 5 µL of the *Pfu* buffer with $MgCl_2$. All PCR

reagents were from Thermo Fischer Scientific. PCR conditions were as follows: 95 °C for 2 min, followed by 35 cycles at 95 °C for 30 s, 60°C for 30 s, 72 °C for 3 min and then 72 °C for 10 min. Amplified genes were purified from agarose gel (0.7 % (w/v)) using the GeneJet gel extraction kit (Thermo Fischer Scientific) before cloning.

Plasmid inserts were sequenced on both strands by Eurofins (Hamburg, Germany) using primers F7/R7 (for inserts in pJet vector) or F8/R8 (for inserts in pETite vector) (see Supplementary Material, Table S1).

In silico screening for novel ATAs and bioinformatics analysis

Environmental samples were collected in Italy, China and Iceland (see Supplementary Material, Table S2). DNA extraction from samples, DNA sequencing and database of sequences generation were carried out as described in (Menzel et al. 2015). Bioinformatic search for new ATAs was performed by aligning query sequences (see Supplementary Material, Table S3) with database sequences using the program LAST (http://last.cbrc.jp/) with default settings (Kielbasa et al. 2011). Phylogenetic trees were created using the Clustal Omega webserver (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al. 2011) and visualized using the iTOL webserver (http://itol2.embl.de/) (Letunic and Bork 2011). Homology modelling was performed through the SWISS-MODEL server (Biasini et al. 2014) and protein structures were inspected by using PYMOL (DeLano 2002).

Enzyme cloning, expression and purification

Is3-TA, It6-TA and B3-TA gene regions were amplified under standard PCR conditions (see above) using primers F1/R1, F2/R2, F3/R3 (see Supplementary Material, Table S1), respectively and the DNA extracted from the corresponding environmental samples as template. Amplified products were cloned in the pJet1.2 vector and the resulting plasmids pJetIs3, pJetIt6 and pJetB3 were transformed in *E. coli* DH5α. Is3-TA, It6-TA and B3-TA genes were subsequently amplified under standard PCR conditions using primers F4/R4, F5/R5, F6/R6 (see Supplementary Material, Table S1), respectively, and the corresponding pJet as template for the cloning into the pETite vector in frame with the C-term His-tag. Gene cloning in the pETite vector and resulting plasmid (pETiteIs3, pETiteIt6 and pETiteB3) transformation in *E. coli* Hi control 10 G was carried out according to the Expresso T7 Cloning and Expression kit manual.

The Vf-TA gene (GenBank AEA39183.1) was synthesized and cloned into the pUC57 vector by BaseClear (Leiden, The Netherlands), then cloned in the pETite vector as previously described. pETite plasmids containing the metagenomic TA genes were subsequently transformed in *E. coli* Rosetta and transformants were inoculated in of LB medium supplemented with 30 μg mL^{-1} kanamycin and 34 μg mL^{-1} chloramphenicol ($LB_{kan30cam34}$) (50 mL) and grown overnight at 37°C and 220 rpm. Pre-cultures were then inoculated in 0.5 L $LB_{kan30cam34}$ medium and kept at 37°C and 220 rpm till the cell density at 600 nm reached 0.5-1. Gene expression was induced by the addition of 1 mL of 1 M IPTG solution in water and the culture was shifted to 17°C and 220 rpm and grown for further 72 h. After recovery by centrifugation (5000 rpm for 30 min), cells were resuspended in 20 mL of wash buffer (20 mM KP buffer, pH 7.0, 500 mM NaCl, 20 mM imidazole) and disrupted by sonication. Soluble protein fraction was separated from cell debris by centrifugation (11000 rpm for 30 min) and clear lysates were checked for the presence of soluble protein by SDS-PAGE analysis (15% T, 2.6% C) according to the method of Laemmli.

As only in the case of Is3-TA the protein was expressed in soluble form, pETiteIt6 and pETiteB3 were further transformed in *E. coli* ArcticExpress (DE3) Competent Cells RIL (Agilent Technologies) according to the manufacturer instructions. Protein expression and cell lysis was carried as described above, with the only exception of lowering the cell culture growth temperature to 11°C after induction and keeping the cells growing for 96 h after induction.

For protein purification clear cell lysates containing soluble protein (as shown by SDS-PAGE analysis) were subsequently incubated with the Ni Sepharose 6 Fast Flow agarose resin (Ni-NTA) (GE-Healthcare, Italy) for 1 h at 4°C under mild shaking. The mixture was then loaded onto a glass column (10 x 110 mm), the resin was washed with 20 mL of wash buffer and His tagged proteins were eluted using a 3 step gradient (10 ml washing buffer containing 100 mM, 200 mM and 300 mM imidazole respectively). If not stated otherwise the purified proteins were dialyzed against 20 mM KP buffer, pH 9.0, 20% glycerol at 4°C for 3 h and stored at -80°C at 1 mg/mL concentration. Protein content was measured using the Bio-Rad Protein Assay according to the method of Bradford and protein purity was verified by SDS-PAGE analysis (15% T, 2.6% C). The molecular weight protein standard mixture from BioRad (Karlsruhe, Germany) was used as reference. Gels were stained for protein detection with Coomassie Brilliant Blue.

Enzyme assay and determination of enzymatic properties

Transaminase activity was assayed by spectrophotometrically measuring the formation of acetophenone at 245 nm (ε = 3.66 mM^{-1} cm^{-1})(Genz et al. 2016) and 20°C on a Jasco V-530 UV/VIS spectrophotometer (Easton, MD, USA) after adding 10-50 µL of purified TA to the enzyme assay solution (2.5 mM pyruvate, 2.5 mM (*S*)-α-methylbenzylamine ((*S*)-MBA) in 50 mM KP buffer, pH 8.0 and 0.25 % (v/v) DMSO in 0.5 mL total volume). One unit of activity is defined as the enzyme activity that produces 1 µmol of acetophenone per minute under the assay conditions described above.

To evaluate the enantioselectivity of ATAs-catalyzed reactions, the acetophenone assay was performed using (*R*)-α-methylbenzylamine as substrate under the same conditions described for (*S*)-MBA. The formation of propiophenone and butyrophenone (for both compounds ε = 3.91 mM^{-1} cm^{-1}) were determined under the same conditions at 242 nm using either (*S*)- or (*R*)-ethylbenzylamine and (*S*)- or (*R*)-phenylbutylamine, respectively, as substrates (Nobili et al. 2015; Genz et al. 2016).

Optimal pH was determined by performing the assay described above using either 50 mM KP buffer (pH 6.0-8.0) or 50 mM Tris-glycine buffer (pH 9.0-10.0). Optimal temperature was determined by heating the assay solution in cuvettes in a water bath in the range 20-90°C for 10 minutes before adding the enzyme. Results were compared to blanks.

To assay the thermostability of the enzymes, purified enzyme solutions (250 µL) were incubated at temperatures in the range of 20°C-90°C for 3 h, then samples (10-50 µL) were added to the enzyme assay solution for residual activity determination. In case of B3-TA long thermal stability was also evaluated by incubating the purified enzyme solution (1 mL) at 80°C and determining the residual activity at scheduled times during three weeks.

Tolerance to organic co-solvents was determined by incubating 100 µL of purified enzyme solutions in 50 mM KP buffer pH 8.0 containing 5 %, 10 %, or 20 % (v/v) of co-solvent (DMSO or acetonitrile or methanol or ethanol) (0.5 mL total volume), respectively, for 5 h at 25°C. After that time, residual activities were evaluated using the assay described above.

B3-TA stability in biphasic system was determined by incubating purified enzyme solution (100 µL) diluted in 50 mM KP buffer, pH 8.0 (0.5 mL total volume) with 0.5 mL of solvent (*tert*-butyl methyl ether, ethyl acetate, toluene, petroleum ether). The mixtures were maintened at

25°C under vigorous shaking and the residual activity was estimated after 5 and 24 h by spectrophotometric analysis.

Substrate spectrum of new ATAs

Acceptor spectrum of ATA homologues was determined at 30°C in 0.5 mL reaction mixture containing 20 mM KP buffer, pH 9.0, 10 mM acceptor, 10 mM (*S*)-MBA, 1 mM PLP, 0.5 mg of purified enzyme. Conversion of (*S*)-MBA (after acetylation) into acetophenone was determined after 24 h by GC-MS analysis (see analytical methods).

Donor spectrum of ATA homologues was determined at 30°C in 0.5 mL reaction mixture containing 20 mM KP buffer, pH 9.0, 10 mM donor, 10 mM pyruvate, 1 mM PLP, 0.5 mg of purified enzyme. Conversions were determined after 24 h by GC-MS analysis, or, alternatively, by estimating the formation of L-alanine (after OPA derivatization) through HPLC analysis (see analytical methods).

Nucleotide sequence accession numbers

The nucleotide sequences reported in this study have been deposited in GenBank, under the following accession numbers: B3-TA: KX505389, Is3-TA: KX505387, It6TA: KX505388.

Results

Identification of ATAs in hot spring metagenomes and sequence analysis

Amine transaminases (ATAs) were searched in a pool of metagenomic sequences from environmental samples collected in hot springs in Iceland, China and Italy (Supplementary Material, Table S2). This search was carried out using the program LAST (http://last.cbrc.jp/) to perform a multiple sequence alignment of the metagenomes and known ATA sequences, either (*S*)- or (*R*)-selective (Supplementary Material, Table S3).

Three sequences (B3-TA, Is3-TA and It6-TA) showing similarity with known (*S*)-selective ATAs were found in metagenomes collected in Iceland (B3-TA – hot spring at 85°C, pH 7.0; Is3-TA – hot spring at 90°C, pH 3.5) and in Italy (It6-TA – hot spring at 76°C, pH 3.5),

respectively, while only partial sequences were found in a sample collected in China (hot spring at 55°C, pH 7.0). No sequences similar to (*R*)-selective ATAs were found in the investigated metagenomes.

The gene sequences coding for B3-TA, Is3-TA, and It6-TA are 1365 bp, 1398 bp, and 1308 bp long, respectively, corresponding to proteins of 455, 466, and 436 amino acids, respectively. According to the InterPro database (https://www.ebi.ac.uk/interpro), these proteins can be classified as class III aminotransferases (IPR005814), which is the class that comprises most of the known (*S*)-selective ATAs.

A comparison of the three novel ATA sequences showed that they are quite different from each other, B3-TA and It6-TA sharing 36% identity, B3-TA and Is3-TA 43% identity, and Is3-TA and It6-TA 35% identity, respectively. Furthermore, the three novel protein sequences appeared more similar to the known (*S*)-selective ATAs used for the screening than to each other. Specifically, Is3-TA showed the highest homology with the ATA from *Chromobacterium violaceum* (61% identity), while It6-TA was more closely related to the ATAs from *Achromobacter denitrificans* and *Caulobacter crescentus* (59% and 61% identity, respectively). B3-TA demonstrated to be the most phylogenetically distant from the well-known mesophilic (*S*)-selective ATAs, showing only 44% identity with ATAs from *Ochrobactrum anthropi* and *Ruegeria pomeroyi*. A multiple sequence alignment among the novel ATAs and the ATA sequences used in the bioinformatic search and the corresponding phylogenetic tree are reported in Fig. S1 and Fig. S2 of the Supplementary Material, respectively.

To understand if the discovered proteins could be functionally active ATAs, we searched their sequences for transaminase conserved residues. This analysis allowed also a possible classification (suggesting reaction and substrate specificity) of transaminases according to the TA fingerprint residues as recently suggested by Steffen-Munsberg et al. (2015).

As shown in Fig. 1, Is3-TA and B3-TA contain all the most conserved residues (see residues in capital letters in the ATA - high activity - row) of the so-called "high activity" ATAs (Steffen-Munsberg et al. 2015), while It6-TA carries all the most conserved residues of high activity ATAs except one (A185 is replaced by a serine). A deeper analysis indicates that Is3-TA contains also all the fingerprint residues with low degree of conservation characteristic of high activity ATAs (see residues in lowercase letters in the ATA - high activity - row) with only one exception (F16), while B3-TA and It6-TA show none of these residues. Interestingly, an

excellent match is shown when comparing It6-TA fingerprint residues with the fingerprint residues typical of β-alanine:pyruvate transaminases (β-Ala:pyr TA, last row in Fig. 1), i.e. ATAs that could accept β-alanine as amino donor (Steffen-Munsberg et al. 2015).

<<Insert Figure 1 here>>

Cloning and recombinant expression of the ATA homologues

The genes coding for the novel ATAs were cloned into the pETite vector in frame with a C-term His-Tag sequence and the resulting plasmids were subsequently transformed and over-expressed in *E. coli* Rosetta. SDS-PAGE analysis showed that Is3-TA was successfully over-expressed in this host in soluble form and purification of this protein by Ni-NTA chromatography afforded 180 mg of pure protein from 1 L culture (Supplementary Material, Fig. S3).

On the contrary, It6-TA and B3-TA were over-expressed in the same host, but accumulated in the cells as inclusion bodies, as revealed by SDS-PAGE analysis (Supplementary Material, Fig. S3). As an alternative host, *E. coli* Arctic Express RIL that expresses two cold adapted chaperone proteins (Cpn10 and Cpn60) favoring heterologous target protein expression with proper folding at very low temperature (10-12°C), was tested. As expected, this approach was successful, both proteins being expressed in soluble form and good yields. Specifically, 25 mg of pure B3-TA and 37 mg of pure It6-TA were obtained from 1 L of the respective cultures after purification by Ni-NTA chromatography (Supplementary Material, Fig. S4).

Determination of functional properties of the novel ATAs

The functional characterization of the ATA homologues was started by evaluating the effect of different reaction conditions in the transamination reaction between the benchmark substrates (*S*)-α-methylbenzylamine ((*S*)-MBA) and pyruvate (Supplementary Material, Scheme S1).

The influence of pH on ATAs activity was evaluated at pHs ranging from 6.0 to 10. As shown in Fig. 2a, all the three enzymes displayed high levels of activity under alkalophilic conditions, this result being in agreement with the average pH optimum of 9.0 typical of ATAs (Park et al. 2012).

Moreover, to assess whether the new ATAs had a thermophilic character, their activity was evaluated at different temperatures (30-90°C, Fig. 2b). Actually, the three transaminases showed

optimal activity at high temperatures, in particular both It6-TA and Is3-TA presented optimal activity at temperatures around 50°C and retained >10% relative activity at 60°C and 70°C, respectively. Even more interestingly, the activity of B3-TA increased constantly with temperature up to 90°C, showing a typical hyperthermophilic behavior. Determination of its activity at higher temperature values was technically unfeasible.

Interesting results were obtained from the thermal stability studies, performed by incubating the three enzymes at different temperatures between 20°C and 90°C for 3 hours and estimating their residual activity by spectrophotometric analysis. In fact, it resulted that Is3-TA and It6-TA maintained about 40% of the starting activity at 40°C, while B3-TA remarkably retained 100% activity after 3 h thermal treatments between 20°C and 80°C (Fig. 2c).

Furthermore, B3-TA long-term thermal stability was also investigated by incubating the enzyme at 80°C for up to 3 weeks and estimating the residual activity by spectrophotometric analysis at scheduled times. Surprisingly B3-TA retained 85% activity after 5 days incubation at 80°C, and more than 40% after two weeks (Fig. 2d).

<<Insert Figure 2 here>>

Structural stability data obtained by CD analysis of the three proteins were in agreement with the activity data estimated at the different temperatures. In particular, the far UV CD spectra of B3-TA recorded at 20°C and 80°C were compared to get some information about any conformational changes of the enzyme secondary structure. As shown in Fig. S5, Supplementary Material, the resulting spectra could be almost overlapping, thus indicating no significant changes occurring in the B3-TA structure at the two tested temperature.

CD analysis was also employed to define the apparent melting temperatures (T_M) of the new ATAs. This analysis, carried out by monitoring thermal events at 220 nm, clearly defined an apparent T_M of 57.32 ± 0.55°C and 79.23 ± 0.31°C for It6-TA and Is3-TA, respectively. Remarkably, CD analysis of B3-TA defined an apparent T_M of 88.15 ± 1.17°C only when the thermal events were monitored in the near UV at 300 nm. In the far UV the thermal shift was not detectable, suggesting that the secondary structure was basically retained during the thermal gradient, while only the ternary structure was affected at very high temperatures.

Finally, the stability of the three enzymes was tested in the presence of 5-20% (v/v) organic co-solvents, i.e. methanol (MeOH), ethanol (EtOH), dimethyl sulfoxide (DMSO), and acetonitrile (ACN). The novel ATAs were incubated in the presence of the different organic co-solvents at 25°C and the residual activity was evaluated spectrophotometrically after 5 hours (Fig. 3). By comparing the residual activity of the samples incubated without organic co-solvents (control sample) to those incubated with organic co-solvents, it was possible to highlight that Is3-TA could be stabilized by the presence of 10-20% (v/v) EtOH and 5-10% (v/v) DMSO, while the presence of MeOH and ACN, especially at the higher concentrations, had a detrimental effect on its activity. On the other hand, the presence of any organic co-solvent did not generate any significant stabilization effect on It6-TA and B3-TA (data not shown). In both cases, the enzyme stability decreased by increasing co-solvents concentration. No residual activity was observed with It6-TA at 20% (v/v) EtOH and 10-20% (v/v) ACN. On the contrary, the residual activity of B3-TA under any condition tested (with the exception of 20% (v/v) EtOH), was always above 40%, thus indicating this enzyme as the most tolerant to organic co-solvents among the three new ATAs.

<<Insert Figure 3 here>>

Considering the superior perfomances shown by B3-TA, its stability in buffer:organic solvent biphasic systems was investigated as well. The enzymatic aqueous solution was incubated in the presence of either *tert*-butyl methyl ether, ethyl acetate, toluene, or petroleum ether (1:1 v/v) under vigorous shaking and the residual activity was evaluated after 5 and 24 h. As shown in Fig. S6, Supplementary Material, B3-TA was quite tolerant to toluene and petroleum ether, retaining more than 50% of its activity in the presence of these two organic solvents after 24 h incubation. It proved to be sensitive to *tert*-butyl methyl ether and ethyl acetate, retaining only 25% activity in the presence of these two organic solvents after 5 h incubation.

Substrate scope of the novel ATAs

To evaluate the substrate specificity of the new ATAs, several keto acids, ketones and aldehydes were considered as potential acceptors (amino donor: (*S*)-MBA), as well as a series of aliphatic and aromatic amines and amino acids were investigated as potential donors (amino acceptor:

pyruvate). Reaction conversions were estimated after 24 h and, to better evaluate the practical synthetic utility of the novel ATAs, these results were compared to those achieved using the well-known ATA from *Vibrio fluvialis* (Vf-TA) (Shin and Kim 2002; Koszelewski et al. 2010) obtained by recombinant expression from its synthetic gene.

As shown in Table 1, the novel ATAs and Vf-TA displayed a quite similar behavior in the transamination reactions using different keto acids and aliphatic ketones as amino acceptors. Specifically, concerning the use of keto acids as acceptors, it was possible to appreciate that an increase in the acceptor side chain length from pyruvate to 2-oxobutyrate resulted in higher conversions, while in the case of the presence of a second carboxylic group in the acceptor, such as the case of α-ketoglutarate, the observed substrate conversions significantly decreased. Furthermore, conversions were significantly lower also with methyl acetoacetate, where the carboxylic group of the keto-acid is is protected as methyl ester, and when using aliphatic ketones as amine acceptors.

On the other hand, in some cases the use of aldehydes in place of ketones resulted in a significant increase of conversions. In particular, quantitative conversions were obtained with Is3-TA, B3-TA, and Vf-TA when using glyoxylate, while acetaldehyde demonstrated to be a very good amino acceptor for B3-TA, as well as propionaldehyde for Is3-TA.

<<Insert Table 1 here>>

As far as the amino donor specificity concerns, the three novel ATAs showed a good activity toward phenyl-substituted secondary amines bearing either a methyl or an ethyl group, while significant decreases in conversions were observed in the case of the (*S*)-α-phenylbutylamine substrate. When testing the primary amine 2-phenylethylamine, the best performances were observed with B3-TA, while 1-aminoindan was well accepted by all the tested enzymes. Among non-aromatic amines, isopropylamine was the best donor for all the novel enzymes, lower conversions being observed when using aliphatic primary mono- and diamines as donors. Interestingly, all the novel ATAs converted β-alanine, which, in agreement with literature data (Steffen-Munsberg et al. 2015), was instead not accepted by Vf-TA. However, no detectable activities were measured in the presence of the more sterically hindered β-amino acids, such as DL-β-phenylalanine and DL-β-homoleucine.

The (*S*)-selectivity of the novel ATAs inferred by sequence analysis has been confirmed by determining spectrophotometrically their specific activity using as substrates either (*R*) or (*S*) aromatic amines, bearing a methyl, ethyl, or propyl side chain adjacent to the amine function (for details, see Supplementary Material, Table S8). Vf-TA has been included in this study for comparison. The results shown in Fig. 4 indicate a very strict (*S*)-selectivity for all the novel ATAs with very low or negligible activity toward the tested (*R*)-amines.

<<Insert Figure 4 here>>

Homology modelling of B3-TA protein structure

Given the potential shown by the highly thermostable B3-TA enzyme for synthetic applications and the lack of crystal structures of (hyper)thermophilic ATAs so far, a structural model of this protein was generated by using the SWISS-MODEL automated server (Biasini et al. 2014). A total of 835 templates were found when searching the B3-TA sequence with BLAST and HHBlits (Remmert et al. 2011) into the SWISS-MODEL template library. Based on global and per-residue models quality and available literature data, the *Vibrio fluvialis* 3D structure (PDB ID: 4E3Q), containing a bound PMP molecule in the active site (Midelfort et al. 2013; Genz et al. 2016) and showing 41% sequence identity with B3-TA, was chosen as template (for details, see Supplementary Material, Fig. S7). The resulted model predicted a homo-dimeric structure with an overall high similarity to Vf-TA (Fig. 5a), and allowed a detailed inspection of the active site pocket generated with the contribution of residues from both monomers (Fig. 5b).

<<Insert Figure 5 here>>

Essential residues for the catalytic activity of B3-TA, for example K289, that forms an internal aldimine with the cofactor PLP, and the amino acids (G321, F322, and T323) forming the so-called "phosphate binding cup" (Humble et al. 2012) and involved in the coordination of the PLP phosphate group, could be easily identified in the model (for clarity, not shown in Fig. 5). Also the "large binding pocket" showed a high degree of similarity in the respect of that of Vf-TA and of other (*S*)-specific ATAs (Fig. 5b). In fact, the most important residues in determining the

shape of this pocket, i.e., L60, W61, and the "flipping arginine" R416, were conserved. Instead, L417 in Vf-TA was replaced by T418 in B3-TA, but this change didn't seem to significantly alter the secondary structure of this region as well as the steric hindrance of the protein side chains in the large binding pocket, at least in the proposed structural model.

On the contrary, some major variations could be observed in the small binding pocket. In fact, while a generally hydrophobic environment was provided by the conserved F89* (the asterisk denotes that this residue is from the other subunit of the dimer), F90*, and Y154, V153 and F19 in Vf-TA were substituted by I157 and V23, respectively, in B3-TA (Fig. 5b). In particular, the F19/V23 substitution apparently increased the available space in the small binding pocket. At this regard, it is important to point out that this residue is always a phenylalanine not only in Vf-TA, but in all the so far characterized mesophilic (*S*)-selective ATAs (Supplementary Material, Fig. S1), with the only exception of *O. anthropi* ATA carrying a tyrosine, and contributes to the formation of a continuous π-π stacked shell together with the other aromatic residues of the pocket, i.e., Y150, F85*, and F86* in Vf-TA (Han et al. 2015; Genz et al. 2016). Moreover, recent attempts of widening the enzyme active site by replacing F19 in Vf-TA, as well as structurally equivalent position in other (*S*)-selective ATAs, with smaller hydrophobic amino acids were not successful, leading to the production of variants showing either extremely low activity or extremely low stability (Han et al. 2015; Nobili et al. 2015; Genz et al. 2016).

Structure-guided sequence analysis of B3-TA-like subfamily members

A BLAST analysis of the NCBI database showed that the closest enzyme (92% identity) to B3-TA was the recently characterized ATA from *Thermomicrobium roseum* (Mathew et al. 2016a). As also this enzyme shows a thermophilic character and a broad substrate scope (see Discussion for further details), we wondered whether other B3-TA-like ATAs could be identified on the basis of sequence similarities.

A deeper look at the results obtained from the BLAST alignment of B3-TA protein sequence clearly showed that the first 15 hits (92-50% identity) corresponded to sequences identified all from the genomes of thermophilic microorganisms belonging to the genera *Thermomicrobium*, *Thermorudis*, *Sphaerobacter*, *Nitrolancea*, *Meiothermus*, *Chloroflexus*, and *Roseiflexus* (Table 2). With the exception of *T. roseum* ATA, none of the other enzymes has been expressed nor

characterized to any extent so far, however all of them have been annotated as putative transaminases.

<<Insert Table 2 here>>

A phylogenetic analysis of the B3-TA-like transaminase subfamily clearly showed that it separates into four distinct clades (Fig. 6). In fact, consistently with the sequence identity results, the sequences from either *Chloroflexus*, *Roseiflexus* or *Meiothermus* species were much more closely related to each other than to those belonging to the other genera, as well as to B3-TA.

<<Insert Figure 6 here>>

Thanks to the overall good alignment of the B3-TA-like transaminase sequences (Supplementary Material, Fig. S8), the conservation of selected active site residues playing a key role in determining the shapes of the small and large binding pockets (Han et al. 2015; Nobili et al. 2015; Genz et al. 2016) was analyzed in detail. As shown in Table 3, the overall active site architecture is well conserved when compared to that of *Vibrio fluvialis* ATA (taken as a representative of mesophilic ATAs on the basis of literature data). The residue corresponding to Y150 in the small binding pocket was fully conserved in the B3-TA-like subfamily. Also V153 was either conserved or substituted by I as previously evidenced for B3-TA. F19 was always substituted by a less bulky, but still hydrophobic residue, either V, as in B3-TA, or L, while F85 was frequently replaced by a tyrosine residue.

<<Insert Table 3 here>>

Instead, the residue found at the position corresponding to F86 in Vf-TA was conserved in the closest homologues to B3-TA, but not in the sequenced obtained from the *Chloroflexus* genomes, where it was always a valine, and from the two *Roseiflexus* representatives, where it was the polar amino acid threonine.

Interestingly, the residues of the large binding pocket were almost completely conserved, including the "flipping arginine" R415. A certain variation was observed only at the position

corresponding to L417 in Vf-TA, that can be occupied by either polar amino acids, e.g., threonine, or hydrophobic ones, e.g., valine. However, a quite low level of amino acidic conservation at this position has been previously reported also among mesophilic ATAs (Genz et al. 2015).

Discussion

ATAs are valuable biocatalysts for the preparation of optically pure amines, being capable of performing reductive amination reactions using either an amino acid or simple aliphatic amines as amine donor. However, the number of available ATAs for industrial applications is still quite limited and this is especially true for (thermo)stable enzymes suitable to be used under harsh conditions such as at high temperature and/or in the presence of organic solvents.

In this work we searched for new (thermo)stable ATAs in the metagenomes of samples collected in hot terrestrial environments in different places in the world, i.e. in Er-Yuan, Yunnan province, China, in Krísuvík and Grensdalur, Iceland, and in the Solfatara Pozzuoli, Naples, Italy, at temperatures ranging from 55 and 95°C. The metagenomic samples were carefully chosen to provide a high global biodiversity, with different relative abundance of bacterial and archaeal representatives and communities structure (Menzel et al. 2015).

ATAs producers demonstrated to be relatively rare in these harsh environments - at least on a basis of homology search - only a few homologues of these enzymes being found in the large sequence information of the samples analyzed. Moreover, the sequences found in the Chinese sample Ch2-EY55S were only partials, possibly because the N50, i.e. the shortest sequence length at 50% of this metagenome, was only 828 nt, while ATA coding sequences are usually quite long (around 1300-1400 bp).

Furthermore, it must be remarked that only homologues to (*S*)-selective ATAs were identified. This result could be related to the fact that known (*S*)- and (*R*)-selective ATAs belong to completely different classes, PLP fold type I and PLP fold type IV, respectively, and show a rather different distribution in Nature. In fact, while the majority of (*S*)-selective ATAs have a bacterial origin, most of the (*R*)-selective ATAs described to date are produced by fungi (Höhne

et al. 2010; Sayer et al. 2014), and fungi were not represented in the microorganisms populations found in the environmental samples we analyzed (Menzel et al. 2015).

The three ATAs full sequences identified in Krísuvík, Grensdalur, and Naples samples, and named Is3-TA, B3-TA, and It6-TA, respectively, originated from hot springs showing temperatures above 76°C, thus suggesting a likely (hyper)thermophilic nature. None of them corresponded to already characterized enzymes, but BLAST analysis allows some speculations about their possible origin. For example, Is3-TA is almost identical to some putative transaminases (GenBank accession codes: CQR42634, ADG29592) from chemolithoautotrophic mesophilic betaproteobacteria belonging to the *Thiomonas* genus. Considering that also some thermophilic strains, such as the *T. islandicus* strain isolated from a hot spring in Grensdalur, Iceland (Vésteinsdóttir et al. 2011), belong to the same genus, we can guess that one of these strains could be the natural producer of Is3-TA. The high homology between B3-TA and the ATA from the thermophile *T. roseum* (Mathew et al. 2016a) suggests that this protein in Nature could be produced by a thermophile belonging to the genus *Thermomicrobium* or similar, while the similarity between It6-TA and a putative transaminase from *Acidothiobacillus caldus* (GenBank code: WP_014003823.1) is too low (78% identity) to draw conclusions about the natural producer of this protein.

The cloning and recombinant expression of the novel ATAs allowed their functional characterization. Concerning thermophilicity and thermostability studies, Is3-TA and It6-TA demonstrated to be moderate thermophilic enzymes, showing optimal activity at 50°C and apparent T_M of 79°C and 57°C, respectively, which were consistent with those reported in the literature for other characterized ATAs (Humble et al. 2012; Chen et al. 2016).

The most interesting results were obtained with B3-TA, which showed both exceptional thermophilicity and thermostability. In fact, this enzyme showed a constantly increasing activity at reaction temperatures up to 90°C and retained more than 40% of starting activity after two-weeks incubation at 80°C. These results, together with the estimated apparent T_M of 88°C, make B3-TA, to the best of our knowledge, the most thermostable natural ATA described to date. In fact, far less thermostability was shown for the ω-transaminases from *Sphaerobacter thermophilus* (Mathew et al. 2016b) and from *Geobacillus thermodenitrificans* (Chen et al. 2016). Even the highly similar *T. roseum* ATA, for which a comparable apparent T_M value of 87°C was estimated, loses completely its activity after 24 h at 60°C (Mathew et al. 2016a).

Since the reactions catalyzed by ATAs often require organic co-solvents for improving substrate/product solubility in aqueous reaction systems (Savile et al. 2010), the tolerance of the new ATAs toward organic co-solvents was tested as well. It resulted that the three ATAs can be used with most of the tested co-solvents with only few exceptions. Moreover, B3-TA was also tested in biphasic systems using water-immiscible solvents such as toluene or petroleum ether and showed satisfactory performances that match well the superior thermostability of this biocatalyst.

As far as the substrate scope of the novel ATAs concerns, the broad range of donor and acceptor substrates utilized by these novel biocatalysts, not very far from that shown by the widely employed *V. fluvialis* ATA, confirms their potential for synthetic applications. The three enzymes showed a markedly preferences for aromatic secondary amines bearing a small substituent on the other side, i.e., an ethyl group or less, thus indicating steric constraints in the small binding pocket. Unexpectedly, not only It6-TA, which displays the typical fingerprint of β-Ala:pyr TAs, but also Is3-TA and B3-TA showed a very good activity in the presence of β-alanine as amine donor, while, as reported in the literature (Steffen-Munsberg et al. 2015), Vf-TA was not active in the presence of this substrate. The role of the conserved residue S185 (present in It6-TA, but substituted by A185 in the other two novel ATAs, Fig. 1) is probably less important than previously supposed on the basis of sequences comparison (Steffen-Munsberg et al. 2015), or other substitutions might help the recognition of β-alanine by Is3-TA and B3-TA. Furthermore, it might be possible that, as suggested by Sayer et al. (2013) in the case of the β-Ala:pyr TA from *Pseudomonas aeruginosa*, the ability to transform β-alanine is correlated to a more rigid active site than that occurring in "high activity" ATAs. No detectable activities were measured in the presence of other β-amino acids, such as DL-β-phenylalanine and DL-β-homoleucine, thus confirming the preference of the novel enzymes for not sterically hindering substrates.

The outstanding stability of B3-TA toward different reaction conditions prompted us in further investigating about its catalytic features by constructing and analyzing a structural model. Interestingly, one of the best templates ranked by the SWISS-MODEL server was the X-ray structure of the mesophilic (T_M 60°C) *V. fluvialis* ATA (Genz et al. 2016) containing a bound PMP molecule in the active site and showing only 41% sequence identity with B3-TA. The model inspection allowed a clear identification of B3-TA active site and a comparison with the

mesophilic counterpart Vf-TA in the respect of key catalytic residues, as well as of those involved in shaping the substrate binding pockets.

Generally speaking, the overall active site architecture of Vf-TA was largely maintained in B3-TA. The most relevant difference was the V23 substitution of the corresponding F19 residue in the Vf-TA small binding pocket that apparently increases the available space in this region. However, this change did not result in an increase of activity toward bulky substrates, B3-TA showing quite strict steric constraints even in the presence of a propyl substituent as in the case of (*S*)-α-phenylbutylamine. This result is not totally surprising taken into account the difficulties encountered by different research group in broadening the substrate specificity of (*S*)-selective ATAs and, in particular, in enlarging the small binding pocket for the acceptance of bulky amines (Midelfort et al. 2013; Genz et al. 2015; Han et al. 2015; Genz et al. 2016; Pavlidis et al. 2016). It is noteworthy that the F19 position in Vf-TA has been frequently addressed by mutagenesis studies aimed at introducing smaller residues, but in all cases the changes resulted in either significantly less active or less soluble variants. Therefore, it is likely that other amino acidic substitutions in B3-TA allow to compensate for the lack of this aromatic residue in order to achieve an active and properly folded enzyme.

In conclusion, this work further demonstrates the potential of metagenomics in the search of relatively rare enzymatic activities such as ATAs in extreme environments. Moreover, the identification of a subfamily of B3-TA-like transaminases, mostly uncharacterized and all from thermophilic microorganisms, may pave the way for the discovery of other useful biocatalysts. In fact, sequence and phylogenetic analysis of the members of this subfamily suggested an overall conservation of the key features necessary for their catalytic activity in amine synthesis. At the same time, given their origin and some variations observed at the active site level, it can be foreseen that members of this subfamily could possess useful applicative features, such as high stability under harsh reaction conditions and different substrate specificities.

Acknowledgments

We warmly thank Dr. Marta Vanoni (ICRM-CNR) for valuable technical assistance in enzymes production and purification.

Compliance with ethical standards

Funding

This work has been carried out in the framework of the HOTRAM project, a FP7 Marie Curie IEF post-doctoral program (grant agreement PIEF-GA-2013-622732 to EEF).

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:252–258. doi: 10.1093/nar/gku340

Chen Y, Yi D, Jiang S, Wei D (2016) Identification of novel thermostable taurine–pyruvate transaminase from *Geobacillus thermodenitrificans* for chiral amine synthesis. Appl Microbiol Biotechnol 100:3101–3111. doi: 10.1007/s00253-015-7129-5

de Miguel Bouzas T, Barros-Velázquez J, Villa TG (2006) Industrial applications of hyperthermophilic enzymes: a review. Protein Pept Lett 13:645–651. doi: 10.2174/092986606777790548

DeCastro M-E, Rodríguez-Belmonte E, González-Siso M-I (2016) Metagenomics of thermophiles with a focus on discovery of novel thermozymes. Front Microbiol 7:1521. doi: 10.3389/fmicb.2016.01521

DeLano W.L. (2002) The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

Fernández-Arrojo L, Guazzaroni ME, López-Cortés N, Beloqui A, Ferrer M (2010) Metagenomic era for biocatalyst identification. Curr Opin Biotechnol 21:725–733. doi: 10.1016/j.copbio.2010.09.006

Ferrandi EE, Sayer C, Isupov MN, Annovazzi C, Marchesi C, Iacobone G, Peng X, Bonch-Osmolovskaya E, Wohlgemuth R, Littlechild JA, Monti D (2015) Discovery and characterization of thermophilic limonene-1,2-epoxide hydrolases from hot spring metagenomic libraries. FEBS J 282:2879–2894. doi: 10.1111/febs.13328

Ferrer M, Martinez-Martinez M, Bargiela R, Streit WR, Golyshina O V., Golyshin PN (2016) Estimating the success of enzyme bioprospecting through metagenomics: Current status and future trends. Microb Biotechnol 9:22–34. doi: 10.1111/1751-7915.12309

Fuchs M, Farnberger JE, Kroutil W (2015) The industrial age of biocatalytic transamination. European J Org Chem 6965–6982. doi: 10.1002/ejoc.201500852

Genz M, Melse O, Schmidt S, Vickers C, van den Bergh T, Joosten H-J, Dörr M, Bornscheuer U (2016) Engineering the amine transaminase from *Vibrio fluvialis* towards branched-chain substrates. ChemCatChem 8:1–5. doi: 10.1002/cctc.201601007

Genz M, Vickers C, van den Bergh T, Joosten HJ, Dörr M, Höhne M, Bornscheuer UT (2015) Alteration of the donor/acceptor spectrum of the (*S*)-amine transaminase from *Vibrio fluvialis*. Int J Mol Sci 16:26953–26963. doi: 10.3390/ijms161126007

Han SW, Park ES, Dong JY, Shin JS (2015) Expanding substrate specificity of ω-transaminase by rational remodeling of a large substrate-binding pocket. Adv Synth Catal 357:2712–2720. doi: 10.1002/adsc.201500239

Höhne M, Schätzle S, Jochens H, Robins K, Bornscheuer UT (2010) Rational assignment of key motifs for function guides in silico enzyme identification. Nat Chem Biol 6:807–813. doi: 10.1038/nchembio.447

Humble MS, Cassimjee KE, Håkansson M, Kimbung YR, Walse B, Abedi V, Federsel H-J, Berglund P, Logan DT (2012) Crystal structures of the *Chromobacterium violaceum* ω-transaminase reveal major structural rearrangements upon binding of coenzyme PLP. FEBS J 279:779–792. doi: 10.1111/j.1742-4658.2012.08468.x

Kielbasa SM, Wan R, Sato K, Horton P, Frith MC (2011) Adaptive seeds tame genomic sequence comparison. Genome Res 21:487–493. doi: 10.1101/gr.113985.110

Koszelewski D, Clay D, Rozzell D, Kroutil W (2009) Deracemisation of α-chiral primary amines by a one-pot, two-step cascade reaction catalysed by ω-transaminases. European J Org Chem 2289–2292. doi: 10.1002/ejoc.200801265

Koszelewski D, Lavandera I, Clay D, Guebitz GM, Rozzell D, Kroutil W (2008a) Formal asymmetric biocatalytic reductive amination. Angew Chemie - Int Ed 47:9337–9340. doi: 10.1002/anie.200803763

Koszelewski D, Lavandera I, Clay D, Rozzell D, Kroutil W (2008b) Asymmetric synthesis of optically pure pharmacologically relevant amines employing ω-transaminases. Adv Synth Catal 350:2761–2766. doi: 10.1002/adsc.200800496

Koszelewski D, Tauber K, Faber K, Kroutil W (2010) ω-Transaminases for the synthesis of non-racemic α-chiral primary amines. Trends Biotechnol 28:324–332. doi: 10.1016/j.tibtech.2010.03.003

Letunic I, Bork P (2011) Interactive Tree of Life v2: Online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39:475–478. doi: 10.1093/nar/gkr201

Littlechild JA (2015) Enzymes from extreme environments and their industrial applications. Front Bioeng Biotechnol 3:1–9. doi: 10.3389/fbioe.2015.00161

Lorenz P, Schleper C (2002) Metagenome - A challenging source of enzyme discovery. J Mol Catal B Enzym 19–20:13–19. doi: 10.1016/S1381-1177(02)00147-9

Martin AR, DiSanto R, Plotnikov I, Kamat S, Shonnard D, Pannuri S (2007) Improved activity and thermostability of (S)-aminotransferase by error-prone polymerase chain reaction for the production of a chiral amine. Biochem Eng J 37:246–255. doi: 10.1016/j.bej.2007.05.001

Mathew S, Deepankumar K, Shin G, Hong EY, Kim B-G, Chung T, Yun H (2016a) Identification of novel thermostable ω-transaminase and its application for enzymatic synthesis of chiral amines at high temperature. RSC Adv 6:69257–69260. doi: 10.1039/C6RA15110H

Mathew S, Nadarajan SP, Chung T, Park HH, Yun H (2016b) Biochemical characterization of thermostable ω-transaminase from *Sphaerobacter thermophilus* and its application for producing aromatic β- and γ-amino acids. Enzyme Microb Technol 87–88:52–60. doi: 10.1016/j.enzmictec.2016.02.013

Mathew S, Yun H (2012) ω-Transaminases for the production of optically pure amines and unnatural amino acids. Acs Catal 2:993–1001. doi: 10.1021/cs300116n

Menzel P, Gudbergsdóttir SR, Rike AG, Lin L, Zhang Q, Contursi P, Moracci M, Kristjansson JK, Bolduc B, Gavrilov S, Ravin N, Mardanov A, Bonch-Osmolovskaya E, Young M, Krogh A, Peng X (2015) Comparative metagenomics of eight geographically remote terrestrial hot springs. Microb Ecol 70:411–424. doi: 10.1007/s00248-015-0576-9

Midelfort KS, Kumar R, Han S, Karmilowicz MJ, McConnell K, Gehlhaar DK, Mistry A, Chang JS, Anderson M, Villalobos A, Minshull J, Govindarajan S, Wong JW (2013) Redesigning and characterizing the substrate specificity and activity of *Vibrio fluvialis* aminotransferase for the synthesis of imagabalin. Protein Eng Des Sel 26:25–33. doi: 10.1093/protein/gzs065

Monti D, Forchin MC, Crotti M, Parmeggiani F, Gatti FG, Brenna E, Riva S (2015) Cascade coupling of ene-reductases and ω-transaminases for the stereoselective synthesis of diastereomerically enriched amines. ChemCatChem 7:3106–3109. doi: 10.1002/cctc.201500424

Nobili A, Steffen-Munsberg F, Kohls H, Trentin I, Schulzke C, Höhne M, Bornscheuer UT (2015) Engineering the active site of the amine transaminase from *Vibrio fluvialis* for the asymmetric synthesis of aryl-alkyl amines and amino alcohols. ChemCatChem 7:757–760.

doi: 10.1002/cctc.201403010

Nugent TC, El-Shazly M (2010) Chiral amine synthesis - Recent developments and trends for enamide reduction, reductive amination, and imine reduction. Adv Synth Catal 352:753–819. doi: 10.1002/adsc.200900719

Park ES, Kim M, Shin JS (2012) Molecular determinants for substrate selectivity of ω-transaminases. Appl Microbiol Biotechnol 93:2425–2435. doi: 10.1007/s00253-011-3584-9

Pavlidis IV, Weiß MS, Genz M, Spurr P, Hanlon SP, Wirz B, Iding H, Bornscheuer UT (2016) Identification of (*S*)-selective transaminases for the asymmetric synthesis of bulky chiral amines. Nat Chem 1–7. doi: 10.1038/NCHEM.2578

Remmert M, Biegert A, Hauser A, Söding J (2011) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods 9:173–175. doi: 10.1038/nmeth.1818

Sambrook J, MacCallum P, Russell D (2001) Molecular cloning: a laboratory manual (third edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Savile CK, Janey JM, Mundorff EC, Moore JC, Tam S, Jarvis WR, Colbeck JC, Krebber A, Fleitz FJ, Brands J, Devine PN, Huisman GW, Hughes GJ (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. Science 329:305–309. doi: 10.1126/science.1188934

Sayer C, Isupov MN, Westlake A, Littlechild JA (2013) Structural studies of *Pseudomonas* and *Chromobacterium* ω-aminotransferases provide insights into their differing substrate specificity. Acta Crystallogr Sect D Biol Crystallogr 69:564–576. doi: 10.1107/S0907444912051670

Sayer C, Martinez-Torres RJ, Richter N, Isupov MN, Hailes HC, Littlechild JA, Ward JM (2014) The substrate specificity, enantioselectivity and structure of the (*R*)-selective amine: pyruvate transaminase from *Nectria haematococca*. FEBS J 281:2240–2253. doi: 10.1111/febs.12778

Shin J, Kim B (2002) Exploring the active site of amine : pyruvate aminotransferase on the basis of the substrate structure - reactivity relationship : how the enzyme controls substrate specificity and stereoselectivity. J Org Chem 67:2848–2853. doi: 10.1021/jo016115i

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality

protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. doi: 10.1038/msb.2011.75

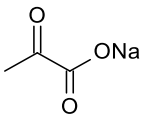
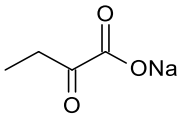
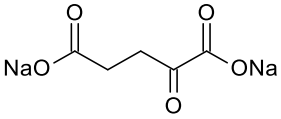
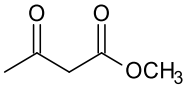
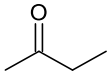
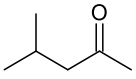
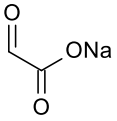
Steffen-Munsberg F, Vickers C, Kohls H, Land H, Mallin H, Nobili A, Skalden L, van den Bergh T, Joosten H-J, Berglund P, Höhne M, Bornscheuer UT (2015) Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications. Biotechnol Adv 33:566–604. doi: 10.1016/j.biotechadv.2014.12.012

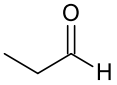
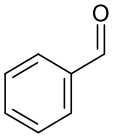
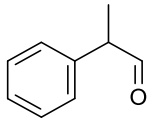
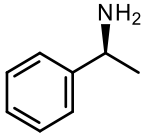
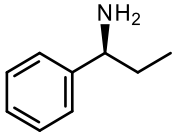
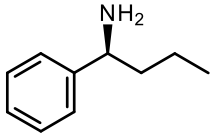
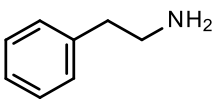
Vésteinsdóttir H, Reynisdóttir DB, Örlygsson J (2011) *Thiomonas islandica* sp. nov., a moderately thermophilic, hydrogen- and sulfur-oxidizing betaproteobacterium isolated from a hot spring. Int J Syst Evol Microbiol 61:132–137. doi: 10.1099/ijs.0.015511-0

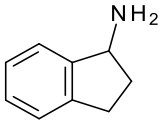
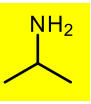
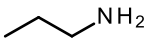
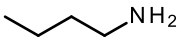
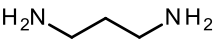
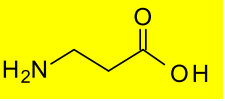
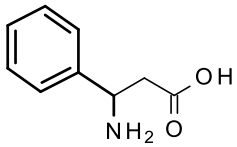
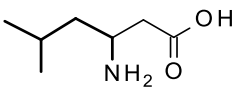
Wilson MC, Piel J (2013) Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. Chem Biol 20:636–647. doi: 10.1016/j.chembiol.2013.04.011

Zarafeta D, Kissas D, Sayer C, Gudbergsdottir SR, Ladoukakis E, Isupov MN, Chatziioannou A, Peng X, Littlechild JA, Skretas G, Kolisis FN (2016) Discovery and characterization of a thermostable and highly halotolerant GH5 cellulase from an icelandic hot spring isolate. PLoS One 11:e0146454. doi: 10.1371/journal.pone.0146454

Table 1 Substrate scope of the novel ATAs

Substrate	**Conversion (%)[a]**			
	Is3-TA	**It6-TA**	**B3-TA**	**Vf-TA**
Amino Acceptor				
Pyruvate 	48	45	51	50
2-Oxobutyrate 	>99	>99	>99	>99
α-Ketoglutarate 	8	2	4	8
Methyl acetoacetate 	16	6	5	15
Acetone 	7	2	8	5
2-Butanone 	3	2	5	11
Methyl isobutyl ketone 	1	1	5	10
Glyoxylate 	>99	62	>99	>99

Acetaldehyde 	6	6	68	18
Propionaldehyde 	74	10	33	34
Benzaldehyde 	3	10	10	15
2-Phenylpropionaldehyde 	16	3	16	19
Amino Donor				
(*S*)-α-Methylphenylamine 	50	41	43	>99
(*S*)-α-Ethylphenylamine 	63	46	23	>99
(*S*)-α-Phenylbutylamine 	2	3	1	>99
2-Phenylethylamine 	8	18	30	15

(*R*,*S*)-1-Aminoindan 	97	89	43	99
Isopropylamine 	11	27	13	28
Propylamine 	6.5	3.7	n.d.[b)]	1.2
Butylamine 	1.4	3.5	1	6.2
1,3-Diaminopropane 	0.7	3	2	n.d.
β-Alanine 	67	57	29	n.d.
DL-β-Phenylalanine 	n.d.	n.d.	n.d.	n.d.
DL-β-Homoleucine 	n.d.	n.d.	n.d.	n.d.

a) Reactions were performed in triplicate and conversions were determined after 24 h by GC/MS or HPLC analysis (see Materials and methods for details). Standard deviation was below 5%, no significant degradation of substrates/products or formation of by-products was observed; b) n.d.: below detection limit.

Table 2 NCBI Reference Sequence, source and similarity of B3-TA homologues

Reference Sequence	Organism	Identity
WP_015922033.1	*Thermomicrobium roseum*	92%
WP_038038660.1	*Thermorudis peleae*	80%
WP_012872904.1	*Sphaerobacter thermophiles*	72%
WP_008476568.1	*Nitrolancea hollandica*	65%
WP_027882148.1	*Meiothermus rufus*	66%
WP_013014879.1	*Meiothermus ruber*	64%
WP_027886436.1	*Meiothermus* sp.	66%
WP_013159295.1	*Meiothermus silvanus*	64%
WP_015941631.1	*Chloroflexus aggregans*	52%
WP_031459398.1	*Chloroflexus* sp. MS-G	50%
WP_044231836.1	*Chloroflexus* sp. Y-396-1	51%
WP_012119566.1	*Roseiflexus castenholzii*	50%
WP_012258069.1	*Chloroflexus* sp.	51%
WP_066785535.1	*Chloroflexus* sp. isl-2	51%
WP_011955192.1	*Roseiflexus* sp. RS-1	50%

Table 3 Amino acid variations in active site architecture among members of the B3-TA-like subfamily

Source	**Selected active site residues**											
	Small binding pocket					**Large binding pocket**						
	150	**153**	**19**	**85***	**86***	**56**	**57**	**415**	**417**	**228**	**259**	**258**
Vibrio fluvialis	**Y**	**V**	**F**	**F**	**F**	**L**	**W**	**R**	**L**	**A**	**I**	**V**
Thermorudis peleae	Y	I	V	F	F	L	W	R	T	A	I	V
Thermomicrobium roseum	Y	I	L	F	F	L	W	R	T	A	I	V
B3-TA	Y	I	V	F	F	L	W	R	T	A	I	V
Sphaerobacter thermophiles	Y	I	V	F	F	L	W	R	N	A	I	V
Nitrolancea hollandica	Y	I	V	Y	F	L	W	R	G	A	I	V
Meiothermus silvanus	Y	I	V	F	F	L	W	R	N	A	I	V
Meiothermus rufus	Y	I	V	F	F	L	W	R	N	A	I	V
Meiothermus ruber	Y	I	V	F	F	L	W	R	N	A	I	V
Meiothermus sp.	Y	I	V	F	F	L	W	R	N	A	I	V
Chloroflexus aggregans	Y	V	L	Y	V	L	W	R	V	A	I	V
Chloroflexus sp. MS-G	Y	V	L	Y	V	L	W	R	V	A	I	V
Chloroflexus sp. Y-396-1	Y	V	L	Y	V	L	W	R	V	A	I	V
Chloroflexus sp.	Y	V	L	Y	V	L	W	R	V	A	I	V
Chloroflexus sp. isl-2	Y	V	L	Y	V	L	W	R	V	A	I	V
Roseiflexus castenholzii	Y	V	L	Y	T	L	W	R	L	A	I	V
Roseiflexus sp. RS-1	Y	V	L	Y	T	L	W	R	L	A	I	V

Figure Captions

Fig. 1 Comparison between the occurrence of key residues in "high activity" ATAs, Is3-TA, B3-TA, It6-TA, and β-alanine:pyruvate transaminases (β-Ala:pyr TA) (Steffen-Munsberg et al. 2015). Conserved residues (more than 70% of the subset considered in (Steffen-Munsberg et al. 2015)) are shown in capital letters, while residues with low degree of conservation (30-70%) are shown in lower case letters. The colour code indicates the physicochemical properties of the residues. Numbering scheme is according to the numbering scheme of (Steffen-Munsberg et al. 2015), and the locations of corresponding amino acids are indicated in the alignment reported in Supplementary Material, Fig. S1

Fig. 2 Influence of pH (**a**) and temperature (**b**) on ATAs activity, influence of temperature (**c**) on ATAs stability, and long-term stability of B3-TA at 80°C (**d**). For details, see Supplementary Material, Tables S4-S7

Fig. 3 Influence of organic co-solvents on ATAs stability. Residual activity was estimated after 5 h of incubation at room temperature

Fig. 4 Enantioselectivity of the novel ATAs. Specific activities (U/mg) of ATAs and Vf-TA towards different (*S*) and (*R*) aromatic amines were determined spectrophotometrically as described in Nobili et al. (2015). For details, see Supplementary Material, Tables S8

Fig. 5 Simulated structure and active site model of B3-TA. (**a**) Overall structure alignment of B3-TA (*red*) and *V. fluvialis* ATA (PDB: 4E3Q, *green*). (**b**) Active pocket detail (B3-TA, *red*; Vf-TA, *green*; residues numbering: Vf-TA/B3-TA), PMP is colored yellow (color by element). The large binding pocket is indicated by the orange arc, while the light blue arc indicates the small binding pocket. Images generated in PYMOL (DeLano W.L. 2002)

Fig. 6 Phylogenetic analysis of B3-TA homologues identified in the NCBI database

Erica Elisa Ferrandi,[*,1,2] Alessandra Previdi,[1] Ivan Bassanini,[1,3] Sergio Riva,[1] Xu Peng,[2] and Daniela Monti[*,1]

Novel thermostable amine transferases from hot spring metagenomes

[1] Istituto di Chimica del Riconoscimento Molecolare, C.N.R., Via Mario Bianco 9, 20131 Milano, Italy.

[2] University of Copenhagen, Ole Maaløesvej 5, 2200 Copenhagen N, Denmark.

[3] Università degli Studi di Milano, Dipartimento di Chimica, Via Golgi 19, 20133 Milano, Italy.

Corresponding authors

Daniela Monti

e-mail: daniela.monti@icrm.cnr.it; Fax: +39 02 28901239; Tel: +39 02 28500025

Erica Elisa Ferrandi

e-mail: erica.ferrandi@icrm.cnr.it; Fax: +39 02 28901239; Tel: +39 02 28500028

Abstract

Hot spring metagenomes, prepared from samples collected at temperatures ranging from 55°C to 95°C, were submitted to an *in silico* screening aimed at the identification of novel amine transaminases (ATAs), valuable biocatalysts for the preparation of optically pure amines.

Three novel (*S*)-selective ATAs, namely Is3-TA, It6-TA, and B3-TA, were discovered in the metagenome of samples collected from hot springs in Iceland and in Italy, cloned from the corresponding metagenomic DNAs and over-expressed in recombinant form in *E. coli*.

Functional characterization of the novel ATAs demonstrated that they all possess a thermophilic character and are capable of performing amine transfer reactions using a broad range of donor and acceptor substrates, thus suggesting a good potential for practical synthetic applications.

In particular, the enzyme B3-TA revealed to be exceptionally thermostable, retaining 85% of activity after 5 days incubation at 80°C and more than 40% after two weeks under the same condition. These results, which were in agreement with the estimation of an apparent melting temperature around 88°C, make B3-TA, to the best of our knowledge, the most thermostable natural ATA described to date. This biocatalyst showed also a good tolerance toward different water-miscible and water-immiscible organic solvents.

A detailed inspection of the homology-based structural model of B3-TA showed that the overall active site architecture of mesophilic (*S*)-selective ATAs was mainly conserved in this hyperthermophilic homologue. Additionally, a subfamily of B3-TA-like transaminases, mostly uncharacterized and all from thermophilic microorganisms, was identified and analyzed in terms of phylogenetic relationships and sequence conservation.

Keywords: Amine transferases, metagenomics, industrial biocatalysis, chiral amines, thermostability

Introduction

Transaminases (TAs, EC 2.6.1.x) are pyridoxal-5'-phosphate (PLP)-dependent enzymes capable of transfering an amino group between an amino donor and a prochiral ketone substrate, thus creating a novel stereogenic center in the transamination reaction. These enzymes are ubiquitous in nature and play a key role in the nitrogen metabolism (Koszelewski et al. 2010; Mathew and Yun 2012).

Generally, TAs are classified into α-TAs, representing the majority of TAs and converting only α-amino and α-keto acids, and ω-TAs, which can accept amino acids having a distal carboxylic acid group as substrates (Mathew and Yun 2012; Fuchs et al. 2015). Among ω-TAs, the class of amine transferases (ATAs) are of particular interest for biotechnological applications as they can also convert amines lacking a carboxylic acid group in addition to α-aminoacids and ω-aminoacids. The use of transaminases for biocatalytic applications has been largely investigated in the last years to provide new sustainable production processes for the preparation of drugs and chemical intermediates and, particularly, to avoid the current use of metal catalysts (i.e., Ru, Pd, Ni) at high pressure in chiral amine chemical syntheses (Nugent and El-Shazly 2010) .

Both (*S*)- and (*R*)-selective ATAs have been found in Nature and exploited for the kinetic resolution of racemic amines as well as for the synthesis of chiral amines starting from the corresponding ketones (Koszelewski et al. 2010; Monti et al. 2015; Steffen-Munsberg et al. 2015). For example, ATAs have been used for the stereoselective synthesis of (*R*)-4-phenylbutan-2-amine, a precursor of the anti-hypertensive dilevalol, and of (*R*)-1-(4-methoxyphenyl)ethylamine, an important building block for the biologically active compound formoterolo (Koszelewski et al. 2008a; Koszelewski et al. 2008b). In another example, both enantiomers of mexiletine [1-(2,6-dimethylphenoxy)-2-propanamine], a chiral antiarrhythmic agent, have been prepared by deracemization starting from the commercially available racemic amine using two different enantioselective ATAs (Koszelewski et al. 2009).

However, many challenges inherent to transaminase stability under industrial process conditions still remain to be tackled before developing large-scale processes. In fact, transamination processes on industrial scale often require the use of organic cosolvents, to increase substrate solubility (Savile et al. 2010), and high temperatures, for example for the stripping of volatile co-

products, i.e., acetone, to shift the equilibrium of the transamination reaction toward product formation (Martin et al. 2007).

At present the array of ATAs from natural sources that have been investigated for their synthetic application is still quite restricted, comprising around 20 different (*S*)-selective ATAs, mainly from bacterial strains (Koszelewski et al. 2010), and around 20 different (*R*)-selective ATAs, mainly from fungi (Höhne et al. 2010; Sayer et al. 2014). In a few cases, ATAs suitable for industrial application have been obtained by protein engineering of existing ATAs, the most remarkable example having been reported by Merck and Codexis on the development of a ATA-catalyzed process for the large-scale manufacture of the antidiabetic compound sitagliptin (Savile et al. 2010). In this case, a highly stable and active (*R*)-selective ATA was obtained after 11 rounds of directed evolution and successfully applied for the industrial-scale conversion of prositagliptin ketone to the optically pure sitagliptin.

Biocatalysts produced by thermophiles often show attractive inherent properties such as high stability in the presence of organic solvent, under operational conditions and during long-term storage (de Miguel Bouzas et al. 2006; Littlechild 2015). When this investigation was started, no naturally thermostable ATAs had been previously described. Recently, the search of transaminases in thermophiles genomes has revealed to be a promising tool to find novel enzymes that might be suitable for industrial applications. Specifically, a taurine:pyruvate TA showing highest activity at 65°C and good stability at up to 20% (v/v) MeOH concentration has been identified from *Geobacillus thermodenitrificans* (Chen et al. 2016), while a TA from *Sphaerobacter thermophilus*, with a strict substrate specificity towards β- and γ-amino acids, has proved to be stable when incubated for 1 h at temperature up to 60°C (Mathew et al. 2016b). Even more recently, a (*S*)-selective ATA with a broad substrate scope and showing a remarkable thermostability has been identified from *Thermomicrobium roseum* and successfully exploited in asymmetric synthesis and kinetic resolution processes at high temperatures (Mathew et al. 2016a).

As an alternative to selected genomes mining, we have herein exploited metagenomics, i.e. the study of the genetic material (metagenome) recovered directly from environmental samples (Lorenz and Schleper 2002; Fernández-Arrojo et al. 2010; Ferrer et al. 2016), to find new thermostable ATAs. The potential of this innovative approach for the identification of novel active and stable biocatalysts has been recently demonstrated with enzymes of different classes

(Wilson and Piel 2013; Ferrandi et al. 2015; DeCastro et al. 2016; Zarafeta et al. 2016). In this work, we report on the discovery and functional characterization of three new ATAs from metagenomic samples collected in hot springs at very high temperature (76-90°C), ideal environments for (hyper)thermophilic microorganisms.

Materials and methods

Chemicals

Amino donors, amino acceptors, PLP, tryptone, yeast extract, agarose were purchased from Sigma-Aldrich (St Louis, MO, USA). Isopropyl-β-D-thiogalactopyranoside (IPTG) was obtained from VWR (Radnor, PA, USA). All other reagents were of analytical grade and commercially available.

Analytical methods

GC-MS analyses were performed using a HP-5MS column (30 m × 0.25 mm × 0.25 μm, Agilent) on a Finnigan TRACE DSQ GC/MS instrument (ThermoQuest, San Jose, CA). The samples of amine derivatives were acetylated before injection as follows: transamination reactions (0.5 mL) catalyzed by TAs were recovered after 24 h, then reaction pH was adjusted to 11.0 by adding 6 M NaOH (50 µL) and products were extracted with ethyl acetate (350 µL). To 100 µL of organic phase dried over anhydrous Na_2SO_4, pyridine and acetic anhydride (15 µL each) were subsequently added and the mixture was kept overnight at room temperature before analysis. The following temperature program was employed: 60°C (1 min) / 6°C min-1/ 150°C (1 min) /12°C min-1/ 210°C (5 min).

HPLC analyses were conducted using a Jasco 880-PU pump equipped with a Jasco 875-UV/Vis detector and a Kinetex 5µm EVO C18 100Å 150 × 4.6 mm column (Phenomenex, Torrance, CA). L-Alanine derivatization was achieved by diluting 10 x the reaction samples (50 µL) in 1 mM HCl and then adding 50 µL of phthaldialdehyde (OPA) reagent (Sigma-Aldrich). The mobile phase consisted of MeOH : THF : 50 mM phosphoric acid, (20:20:960), pH adjusted to 7.5 with NaOH (eluent A), and MeOH : H_2O, (65:35) (eluent B). L-alanine-OPA derivative was eluted at a flow rate of 0.5 mL min^{-1} with the following gradient: t = 0 min, 80% A; t = 17 min,

70% A; t = 32 min, 20% A; t = 37 min, 20% A; t = 47 min, 80% A. Peaks were detected at 340 nm and calibration curves were prepared using standard solutions of L-alanine-OPA derivative.

CD spectra were recorded on a nitrogen flushed Jasco J-1100 spectropolarimeter (Easton, MD, USA) interfaced with a thermostatically controlled cell holder. CD analysis was performed with purified protein samples dialyzed overnight against 50 mM potassium phosphate (KP) buffer, pH 9.0, and diluted in degassed water (0.15 mg mL^{-1} final concentration for far-UV analysis, 1 mg mL^{-1} final concentration for near-UV analysis) in quartz cuvettes with 0.1 cm path length (far-UV analysis) or 1 cm path lenght (near-UV analysis). CD spectra were recorded in the range between 185 nm and 250 nm (far-UV analysis) or from 250 nm and 350 nm (near-UV analysis) at 20°C or 95°C.

For the determination of apparent T_M variation of CD signal at 220 nm (far-UV) or 300 nm (near -UV), the following temperature programs were used:

IS3-TA and B3-TA sample

20°C up to 65°C at 5°C/min_ data pitch each 2°C, hold 30''

65°C up to 90°C at 2.5°C/min_data pitch each 0.5°C, hold 30''

90°C up to 95°C at 5°C/min pitch data each 2°C, hold 30''

It6-TA sample

20°C up to 55°C at 5°C/min_ data pitch each 2°C, hold 30''

55°C up to 85°C at 2.5°C/min_data pitch each 0.5°C, hold 30''

85°C up to 95°C at 5°C/min pitch data each 2°C, hold 30''

General molecular biology techniques

Gene cloning into the pJet1.2 vector was performed with the CloneJET PCR Cloning Kit (Thermo Fischer Scientific, Waltham, MA, USA). Plasmid DNA was purified by using the E.Z.N.A. Plasmid Mini kit II (Omega/VWR). Plasmids were transformed in *E.coli* DH5α or *E.coli* Rosetta by using standard techniques (Sambrook et al. 1989). Gene cloning in the pETite C-His Kan vector and *E. coli* 10G chemically competent cell transformations were carried out with the Expresso T7 Cloning and Expression kit from Lucigen (Middleton, WI, USA). If not stated otherwise, standard PCR amplifications were carried out on 25 µL reaction mixtures containing plasmid DNA (5 ng) or metagenomic DNA (50-100 ng), primers (1 µM each), dNTPs (0.2 mM each), 1.25 U of *Pfu* DNA polymerase and 5 µL of the *Pfu* buffer with $MgCl_2$. All PCR

reagents were from Thermo Fischer Scientific. PCR conditions were as follows: 95 °C for 2 min, followed by 35 cycles at 95 °C for 30 s, 60°C for 30 s, 72 °C for 3 min and then 72 °C for 10 min. Amplified genes were purified from agarose gel (0.7 % (w/v)) using the GeneJet gel extraction kit (Thermo Fischer Scientific) before cloning.

Plasmid inserts were sequenced on both strands by Eurofins (Hamburg, Germany) using primers F7/R7 (for inserts in pJet vector) or F8/R8 (for inserts in pETite vector) (see Supplementary Material, Table S1).

In silico screening for novel ATAs and bioinformatics analysis

Environmental samples were collected in Italy, China and Iceland (see Supplementary Material, Table S2). DNA extraction from samples, DNA sequencing and database of sequences generation were carried out as described in (Menzel et al. 2015). Bioinformatic search for new ATAs was performed by aligning query sequences (see Supplementary Material, Table S3) with database sequences using the program LAST (http://last.cbrc.jp/) with default settings (Kielbasa et al. 2011). Phylogenetic trees were created using the Clustal Omega webserver (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers et al. 2011) and visualized using the iTOL webserver (http://itol2.embl.de/) (Letunic and Bork 2011). Homology modelling was performed through the SWISS-MODEL server (Biasini et al. 2014) and protein structures were inspected by using PYMOL (DeLano 2002).

Enzyme cloning, expression and purification

Is3-TA, It6-TA and B3-TA gene regions were amplified under standard PCR conditions (see above) using primers F1/R1, F2/R2, F3/R3 (see Supplementary Material, Table S1), respectively and the DNA extracted from the corresponding environmental samples as template. Amplified products were cloned in the pJet1.2 vector and the resulting plasmids pJetIs3, pJetIt6 and pJetB3 were transformed in *E. coli* DH5α. Is3-TA, It6-TA and B3-TA genes were subsequently amplified under standard PCR conditions using primers F4/R4, F5/R5, F6/R6 (see Supplementary Material, Table S1), respectively, and the corresponding pJet as template for the cloning into the pETite vector in frame with the C-term His-tag. Gene cloning in the pETite vector and resulting plasmid (pETiteIs3, pETiteIt6 and pETiteB3) transformation in *E. coli* Hi control 10 G was carried out according to the Expresso T7 Cloning and Expression kit manual.

The Vf-TA gene (GenBank AEA39183.1) was synthesized and cloned into the pUC57 vector by BaseClear (Leiden, The Netherlands), then cloned in the pETite vector as previously described. pETite plasmids containing the metagenomic TA genes were subsequently transformed in *E. coli* Rosetta and transformants were inoculated in of LB medium supplemented with 30 μg mL^{-1} kanamycin and 34 μg mL^{-1} chloramphenicol ($LB_{kan30cam34}$) (50 mL) and grown overnight at 37°C and 220 rpm. Pre-cultures were then inoculated in 0.5 L $LB_{kan30cam34}$ medium and kept at 37°C and 220 rpm till the cell density at 600 nm reached 0.5-1. Gene expression was induced by the addition of 1 mL of 1 M IPTG solution in water and the culture was shifted to 17°C and 220 rpm and grown for further 72 h. After recovery by centrifugation (5000 rpm for 30 min), cells were resuspended in 20 mL of wash buffer (20 mM KP buffer, pH 7.0, 500 mM NaCl, 20 mM imidazole) and disrupted by sonication. Soluble protein fraction was separated from cell debris by centrifugation (11000 rpm for 30 min) and clear lysates were checked for the presence of soluble protein by SDS-PAGE analysis (15% T, 2.6% C) according to the method of Laemmli.

As only in the case of Is3-TA the protein was expressed in soluble form, pETiteIt6 and pETiteB3 were further transformed in *E. coli* ArcticExpress (DE3) Competent Cells RIL (Agilent Technologies) according to the manufacturer instructions. Protein expression and cell lysis was carried as described above, with the only exception of lowering the cell culture growth temperature to 11°C after induction and keeping the cells growing for 96 h after induction.

For protein purification clear cell lysates containing soluble protein (as shown by SDS-PAGE analysis) were subsequently incubated with the Ni Sepharose 6 Fast Flow agarose resin (Ni-NTA) (GE-Healthcare, Italy) for 1 h at 4°C under mild shaking. The mixture was then loaded onto a glass column (10 x 110 mm), the resin was washed with 20 mL of wash buffer and His tagged proteins were eluted using a 3 step gradient (10 ml washing buffer containing 100 mM, 200 mM and 300 mM imidazole respectively). If not stated otherwise the purified proteins were dialyzed against 20 mM KP buffer, pH 9.0, 20% glycerol at 4°C for 3 h and stored at -80°C at 1 mg/mL concentration. Protein content was measured using the Bio-Rad Protein Assay according to the method of Bradford and protein purity was verified by SDS-PAGE analysis (15% T, 2.6% C). The molecular weight protein standard mixture from BioRad (Karlsruhe, Germany) was used as reference. Gels were stained for protein detection with Coomassie Brilliant Blue.

Enzyme assay and determination of enzymatic properties

Transaminase activity was assayed by spectrophotometrically measuring the formation of acetophenone at 245 nm (ε = 3.66 mM^{-1} cm^{-1})(Genz et al. 2016) and 20°C on a Jasco V-530 UV/VIS spectrophotometer (Easton, MD, USA) after adding 10-50 µL of purified TA to the enzyme assay solution (2.5 mM pyruvate, 2.5 mM (*S*)-α-methylbenzylamine ((*S*)-MBA) in 50 mM KP buffer, pH 8.0 and 0.25 % (v/v) DMSO in 0.5 mL total volume). One unit of activity is defined as the enzyme activity that produces 1 µmol of acetophenone per minute under the assay conditions described above.

To evaluate the enantioselectivity of ATAs-catalyzed reactions, the acetophenone assay was performed using (*R*)-α-methylbenzylamine as substrate under the same conditions described for (*S*)-MBA. The formation of propiophenone and butyrophenone (for both compounds ε = 3.91 mM^{-1} cm^{-1}) were determined under the same conditions at 242 nm using either (*S*)- or (*R*)-ethylbenzylamine and (*S*)- or (*R*)-phenylbutylamine, respectively, as substrates (Nobili et al. 2015; Genz et al. 2016).

Optimal pH was determined by performing the assay described above using either 50 mM KP buffer (pH 6.0-8.0) or 50 mM Tris-glycine buffer (pH 9.0-10.0). Optimal temperature was determined by heating the assay solution in cuvettes in a water bath in the range 20-90°C for 10 minutes before adding the enzyme. Results were compared to blanks.

To assay the thermostability of the enzymes, purified enzyme solutions (250 µL) were incubated at temperatures in the range of 20°C-90°C for 3 h, then samples (10-50 µL) were added to the enzyme assay solution for residual activity determination. In case of B3-TA long thermal stability was also evaluated by incubating the purified enzyme solution (1 mL) at 80°C and determining the residual activity at scheduled times during three weeks.

Tolerance to organic co-solvents was determined by incubating 100 µL of purified enzyme solutions in 50 mM KP buffer pH 8.0 containing 5 %, 10 %, or 20 % (v/v) of co-solvent (DMSO or acetonitrile or methanol or ethanol) (0.5 mL total volume), respectively, for 5 h at 25°C. After that time, residual activities were evaluated using the assay described above.

B3-TA stability in biphasic system was determined by incubating purified enzyme solution (100 µL) diluted in 50 mM KP buffer, pH 8.0 (0.5 mL total volume) with 0.5 mL of solvent (*tert*-butyl methyl ether, ethyl acetate, toluene, petroleum ether). The mixtures were maintened at

25°C under vigorous shaking and the residual activity was estimated after 5 and 24 h by spectrophotometric analysis.

Substrate spectrum of new ATAs

Acceptor spectrum of ATA homologues was determined at 30°C in 0.5 mL reaction mixture containing 20 mM KP buffer, pH 9.0, 10 mM acceptor, 10 mM (*S*)-MBA, 1 mM PLP, 0.5 mg of purified enzyme. Conversion of (*S*)-MBA (after acetylation) into acetophenone was determined after 24 h by GC-MS analysis (see analytical methods).

Donor spectrum of ATA homologues was determined at 30°C in 0.5 mL reaction mixture containing 20 mM KP buffer, pH 9.0, 10 mM donor, 10 mM pyruvate, 1 mM PLP, 0.5 mg of purified enzyme. Conversions were determined after 24 h by GC-MS analysis, or, alternatively, by estimating the formation of L-alanine (after OPA derivatization) through HPLC analysis (see analytical methods).

Nucleotide sequence accession numbers

The nucleotide sequences reported in this study have been deposited in GenBank, under the following accession numbers: B3-TA: KX505389, Is3-TA: KX505387, It6TA: KX505388.

Results

Identification of ATAs in hot spring metagenomes and sequence analysis

Amine transaminases (ATAs) were searched in a pool of metagenomic sequences from environmental samples collected in hot springs in Iceland, China and Italy (Supplementary Material, Table S2). This search was carried out using the program LAST (http://last.cbrc.jp/) to perform a multiple sequence alignment of the metagenomes and known ATA sequences, either (*S*)- or (*R*)-selective (Supplementary Material, Table S3).

Three sequences (B3-TA, Is3-TA and It6-TA) showing similarity with known (*S*)-selective ATAs were found in metagenomes collected in Iceland (B3-TA – hot spring at 85°C, pH 7.0; Is3-TA – hot spring at 90°C, pH 3.5) and in Italy (It6-TA – hot spring at 76°C, pH 3.5),

respectively, while only partial sequences were found in a sample collected in China (hot spring at 55°C, pH 7.0). No sequences similar to (*R*)-selective ATAs were found in the investigated metagenomes.

The gene sequences coding for B3-TA, Is3-TA, and It6-TA are 1365 bp, 1398 bp, and 1308 bp long, respectively, corresponding to proteins of 455, 466, and 436 amino acids, respectively. According to the InterPro database (https://www.ebi.ac.uk/interpro), these proteins can be classified as class III aminotransferases (IPR005814), which is the class that comprises most of the known (*S*)-selective ATAs.

A comparison of the three novel ATA sequences showed that they are quite different from each other, B3-TA and It6-TA sharing 36% identity, B3-TA and Is3-TA 43% identity, and Is3-TA and It6-TA 35% identity, respectively. Furthermore, the three novel protein sequences appeared more similar to the known (*S*)-selective ATAs used for the screening than to each other. Specifically, Is3-TA showed the highest homology with the ATA from *Chromobacterium violaceum* (61% identity), while It6-TA was more closely related to the ATAs from *Achromobacter denitrificans* and *Caulobacter crescentus* (59% and 61% identity, respectively). B3-TA demonstrated to be the most phylogenetically distant from the well-known mesophilic (*S*)-selective ATAs, showing only 44% identity with ATAs from *Ochrobactrum anthropi* and *Ruegeria pomeroyi*. A multiple sequence alignment among the novel ATAs and the ATA sequences used in the bioinformatic search and the corresponding phylogenetic tree are reported in Fig. S1 and Fig. S2 of the Supplementary Material, respectively.

To understand if the discovered proteins could be functionally active ATAs, we searched their sequences for transaminase conserved residues. This analysis allowed also a possible classification (suggesting reaction and substrate specificity) of transaminases according to the TA fingerprint residues as recently suggested by Steffen-Munsberg et al. (2015).

As shown in Fig. 1, Is3-TA and B3-TA contain all the most conserved residues (see residues in capital letters in the ATA - high activity - row) of the so-called "high activity" ATAs (Steffen-Munsberg et al. 2015), while It6-TA carries all the most conserved residues of high activity ATAs except one (A185 is replaced by a serine). A deeper analysis indicates that Is3-TA contains also all the fingerprint residues with low degree of conservation characteristic of high activity ATAs (see residues in lowercase letters in the ATA - high activity - row) with only one exception (F16), while B3-TA and It6-TA show none of these residues. Interestingly, an

excellent match is shown when comparing It6-TA fingerprint residues with the fingerprint residues typical of β-alanine:pyruvate transaminases (β-Ala:pyr TA, last row in Fig. 1), i.e. ATAs that could accept β-alanine as amino donor (Steffen-Munsberg et al. 2015).

<<Insert Figure 1 here>>

Cloning and recombinant expression of the ATA homologues

The genes coding for the novel ATAs were cloned into the pETite vector in frame with a C-term His-Tag sequence and the resulting plasmids were subsequently transformed and over-expressed in *E. coli* Rosetta. SDS-PAGE analysis showed that Is3-TA was successfully over-expressed in this host in soluble form and purification of this protein by Ni-NTA chromatography afforded 180 mg of pure protein from 1 L culture (Supplementary Material, Fig. S3).

On the contrary, It6-TA and B3-TA were over-expressed in the same host, but accumulated in the cells as inclusion bodies, as revealed by SDS-PAGE analysis (Supplementary Material, Fig. S3). As an alternative host, *E. coli* Arctic Express RIL that expresses two cold adapted chaperone proteins (Cpn10 and Cpn60) favoring heterologous target protein expression with proper folding at very low temperature (10-12°C), was tested. As expected, this approach was successful, both proteins being expressed in soluble form and good yields. Specifically, 25 mg of pure B3-TA and 37 mg of pure It6-TA were obtained from 1 L of the respective cultures after purification by Ni-NTA chromatography (Supplementary Material, Fig. S4).

Determination of functional properties of the novel ATAs

The functional characterization of the ATA homologues was started by evaluating the effect of different reaction conditions in the transamination reaction between the benchmark substrates (*S*)-α-methylbenzylamine ((*S*)-MBA) and pyruvate (Supplementary Material, Scheme S1).

The influence of pH on ATAs activity was evaluated at pHs ranging from 6.0 to 10. As shown in Fig. 2a, all the three enzymes displayed high levels of activity under alkalophilic conditions, this result being in agreement with the average pH optimum of 9.0 typical of ATAs (Park et al. 2012).

Moreover, to assess whether the new ATAs had a thermophilic character, their activity was evaluated at different temperatures (30-90°C, Fig. 2b). Actually, the three transaminases showed

optimal activity at high temperatures, in particular both It6-TA and Is3-TA presented optimal activity at temperatures around 50°C and retained >10% relative activity at 60°C and 70°C, respectively. Even more interestingly, the activity of B3-TA increased constantly with temperature up to 90°C, showing a typical hyperthermophilic behavior. Determination of its activity at higher temperature values was technically unfeasible.

Interesting results were obtained from the thermal stability studies, performed by incubating the three enzymes at different temperatures between 20°C and 90°C for 3 hours and estimating their residual activity by spectrophotometric analysis. In fact, it resulted that Is3-TA and It6-TA maintained about 40% of the starting activity at 40°C, while B3-TA remarkably retained 100% activity after 3 h thermal treatments between 20°C and 80°C (Fig. 2c).

Furthermore, B3-TA long-term thermal stability was also investigated by incubating the enzyme at 80°C for up to 3 weeks and estimating the residual activity by spectrophotometric analysis at scheduled times. Surprisingly B3-TA retained 85% activity after 5 days incubation at 80°C, and more than 40% after two weeks (Fig. 2d).

<<Insert Figure 2 here>>

Structural stability data obtained by CD analysis of the three proteins were in agreement with the activity data estimated at the different temperatures. In particular, the far UV CD spectra of B3-TA recorded at 20°C and 80°C were compared to get some information about any conformational changes of the enzyme secondary structure. As shown in Fig. S5, Supplementary Material, the resulting spectra could be almost overlapping, thus indicating no significant changes occurring in the B3-TA structure at the two tested temperature.

CD analysis was also employed to define the apparent melting temperatures (T_M) of the new ATAs. This analysis, carried out by monitoring thermal events at 220 nm, clearly defined an apparent T_M of 57.32 ± 0.55°C and 79.23 ± 0.31°C for It6-TA and Is3-TA, respectively. Remarkably, CD analysis of B3-TA defined an apparent T_M of 88.15 ± 1.17°C only when the thermal events were monitored in the near UV at 300 nm. In the far UV the thermal shift was not detectable, suggesting that the secondary structure was basically retained during the thermal gradient, while only the ternary structure was affected at very high temperatures.

Finally, the stability of the three enzymes was tested in the presence of 5-20% (v/v) organic co-solvents, i.e. methanol (MeOH), ethanol (EtOH), dimethyl sulfoxide (DMSO), and acetonitrile (ACN). The novel ATAs were incubated in the presence of the different organic co-solvents at 25°C and the residual activity was evaluated spectrophotometrically after 5 hours (Fig. 3). By comparing the residual activity of the samples incubated without organic co-solvents (control sample) to those incubated with organic co-solvents, it was possible to highlight that Is3-TA could be stabilized by the presence of 10-20% (v/v) EtOH and 5-10% (v/v) DMSO, while the presence of MeOH and ACN, especially at the higher concentrations, had a detrimental effect on its activity. On the other hand, the presence of any organic co-solvent did not generate any significant stabilization effect on It6-TA and B3-TA (data not shown). In both cases, the enzyme stability decreased by increasing co-solvents concentration. No residual activity was observed with It6-TA at 20% (v/v) EtOH and 10-20% (v/v) ACN. On the contrary, the residual activity of B3-TA under any condition tested (with the exception of 20% (v/v) EtOH), was always above 40%, thus indicating this enzyme as the most tolerant to organic co-solvents among the three new ATAs.

<<Insert Figure 3 here>>

Considering the superior perfomances shown by B3-TA, its stability in buffer:organic solvent biphasic systems was investigated as well. The enzymatic aqueous solution was incubated in the presence of either *tert*-butyl methyl ether, ethyl acetate, toluene, or petroleum ether (1:1 v/v) under vigorous shaking and the residual activity was evaluated after 5 and 24 h. As shown in Fig. S6, Supplementary Material, B3-TA was quite tolerant to toluene and petroleum ether, retaining more than 50% of its activity in the presence of these two organic solvents after 24 h incubation. It proved to be sensitive to *tert*-butyl methyl ether and ethyl acetate, retaining only 25% activity in the presence of these two organic solvents after 5 h incubation.

Substrate scope of the novel ATAs

To evaluate the substrate specificity of the new ATAs, several keto acids, ketones and aldehydes were considered as potential acceptors (amino donor: (*S*)-MBA), as well as a series of aliphatic and aromatic amines and amino acids were investigated as potential donors (amino acceptor:

pyruvate). Reaction conversions were estimated after 24 h and, to better evaluate the practical synthetic utility of the novel ATAs, these results were compared to those achieved using the well-known ATA from *Vibrio fluvialis* (Vf-TA) (Shin and Kim 2002; Koszelewski et al. 2010) obtained by recombinant expression from its synthetic gene.

As shown in Table 1, the novel ATAs and Vf-TA displayed a quite similar behavior in the transamination reactions using different keto acids and aliphatic ketones as amino acceptors. Specifically, concerning the use of keto acids as acceptors, it was possible to appreciate that an increase in the acceptor side chain length from pyruvate to 2-oxobutyrate resulted in higher conversions, while in the case of the presence of a second carboxylic group in the acceptor, such as the case of α-ketoglutarate, the observed substrate conversions significantly decreased. Furthermore, conversions were significantly lower also with methyl acetoacetate, where the carboxylic group of the keto-acid is is protected as methyl ester, and when using aliphatic ketones as amine acceptors.

On the other hand, in some cases the use of aldehydes in place of ketones resulted in a significant increase of conversions. In particular, quantitative conversions were obtained with Is3-TA, B3-TA, and Vf-TA when using glyoxylate, while acetaldehyde demonstrated to be a very good amino acceptor for B3-TA, as well as propionaldehyde for Is3-TA.

<<Insert Table 1 here>>

As far as the amino donor specificity concerns, the three novel ATAs showed a good activity toward phenyl-substituted secondary amines bearing either a methyl or an ethyl group, while significant decreases in conversions were observed in the case of the (*S*)-α-phenylbutylamine substrate. When testing the primary amine 2-phenylethylamine, the best performances were observed with B3-TA, while 1-aminoindan was well accepted by all the tested enzymes. Among non-aromatic amines, isopropylamine was the best donor for all the novel enzymes, lower conversions being observed when using aliphatic primary mono- and diamines as donors. Interestingly, all the novel ATAs converted β-alanine, which, in agreement with literature data (Steffen-Munsberg et al. 2015), was instead not accepted by Vf-TA. However, no detectable activities were measured in the presence of the more sterically hindered β-amino acids, such as DL-β-phenylalanine and DL-β-homoleucine.

The (*S*)-selectivity of the novel ATAs inferred by sequence analysis has been confirmed by determining spectrophotometrically their specific activity using as substrates either (*R*) or (*S*) aromatic amines, bearing a methyl, ethyl, or propyl side chain adjacent to the amine function (for details, see Supplementary Material, Table S8). Vf-TA has been included in this study for comparison. The results shown in Fig. 4 indicate a very strict (*S*)-selectivity for all the novel ATAs with very low or negligible activity toward the tested (*R*)-amines.

<<Insert Figure 4 here>>

Homology modelling of B3-TA protein structure

Given the potential shown by the highly thermostable B3-TA enzyme for synthetic applications and the lack of crystal structures of (hyper)thermophilic ATAs so far, a structural model of this protein was generated by using the SWISS-MODEL automated server (Biasini et al. 2014). A total of 835 templates were found when searching the B3-TA sequence with BLAST and HHBlits (Remmert et al. 2011) into the SWISS-MODEL template library. Based on global and per-residue models quality and available literature data, the *Vibrio fluvialis* 3D structure (PDB ID: 4E3Q), containing a bound PMP molecule in the active site (Midelfort et al. 2013; Genz et al. 2016) and showing 41% sequence identity with B3-TA, was chosen as template (for details, see Supplementary Material, Fig. S7). The resulted model predicted a homo-dimeric structure with an overall high similarity to Vf-TA (Fig. 5a), and allowed a detailed inspection of the active site pocket generated with the contribution of residues from both monomers (Fig. 5b).

<<Insert Figure 5 here>>

Essential residues for the catalytic activity of B3-TA, for example K289, that forms an internal aldimine with the cofactor PLP, and the amino acids (G321, F322, and T323) forming the so-called "phosphate binding cup" (Humble et al. 2012) and involved in the coordination of the PLP phosphate group, could be easily identified in the model (for clarity, not shown in Fig. 5). Also the "large binding pocket" showed a high degree of similarity in the respect of that of Vf-TA and of other (*S*)-specific ATAs (Fig. 5b). In fact, the most important residues in determining the

shape of this pocket, i.e., L60, W61, and the "flipping arginine" R416, were conserved. Instead, L417 in Vf-TA was replaced by T418 in B3-TA, but this change didn't seem to significantly alter the secondary structure of this region as well as the steric hindrance of the protein side chains in the large binding pocket, at least in the proposed structural model.

On the contrary, some major variations could be observed in the small binding pocket. In fact, while a generally hydrophobic environment was provided by the conserved F89* (the asterisk denotes that this residue is from the other subunit of the dimer), F90*, and Y154, V153 and F19 in Vf-TA were substituted by I157 and V23, respectively, in B3-TA (Fig. 5b). In particular, the F19/V23 substitution apparently increased the available space in the small binding pocket. At this regard, it is important to point out that this residue is always a phenylalanine not only in Vf-TA, but in all the so far characterized mesophilic (*S*)-selective ATAs (Supplementary Material, Fig. S1), with the only exception of *O. anthropi* ATA carrying a tyrosine, and contributes to the formation of a continuous π-π stacked shell together with the other aromatic residues of the pocket, i.e., Y150, F85*, and F86* in Vf-TA (Han et al. 2015; Genz et al. 2016). Moreover, recent attempts of widening the enzyme active site by replacing F19 in Vf-TA, as well as structurally equivalent position in other (*S*)-selective ATAs, with smaller hydrophobic amino acids were not successful, leading to the production of variants showing either extremely low activity or extremely low stability (Han et al. 2015; Nobili et al. 2015; Genz et al. 2016).

Structure-guided sequence analysis of B3-TA-like subfamily members

A BLAST analysis of the NCBI database showed that the closest enzyme (92% identity) to B3-TA was the recently characterized ATA from *Thermomicrobium roseum* (Mathew et al. 2016a). As also this enzyme shows a thermophilic character and a broad substrate scope (see Discussion for further details), we wondered whether other B3-TA-like ATAs could be identified on the basis of sequence similarities.

A deeper look at the results obtained from the BLAST alignment of B3-TA protein sequence clearly showed that the first 15 hits (92-50% identity) corresponded to sequences identified all from the genomes of thermophilic microorganisms belonging to the genera *Thermomicrobium*, *Thermorudis*, *Sphaerobacter*, *Nitrolancea*, *Meiothermus*, *Chloroflexus*, and *Roseiflexus* (Table 2). With the exception of *T. roseum* ATA, none of the other enzymes has been expressed nor

characterized to any extent so far, however all of them have been annotated as putative transaminases.

<<Insert Table 2 here>>

A phylogenetic analysis of the B3-TA-like transaminase subfamily clearly showed that it separates into four distinct clades (Fig. 6). In fact, consistently with the sequence identity results, the sequences from either *Chloroflexus*, *Roseiflexus* or *Meiothermus* species were much more closely related to each other than to those belonging to the other genera, as well as to B3-TA.

<<Insert Figure 6 here>>

Thanks to the overall good alignment of the B3-TA-like transaminase sequences (Supplementary Material, Fig. S8), the conservation of selected active site residues playing a key role in determining the shapes of the small and large binding pockets (Han et al. 2015; Nobili et al. 2015; Genz et al. 2016) was analyzed in detail. As shown in Table 3, the overall active site architecture is well conserved when compared to that of *Vibrio fluvialis* ATA (taken as a representative of mesophilic ATAs on the basis of literature data). The residue corresponding to Y150 in the small binding pocket was fully conserved in the B3-TA-like subfamily. Also V153 was either conserved or substituted by I as previously evidenced for B3-TA. F19 was always substituted by a less bulky, but still hydrophobic residue, either V, as in B3-TA, or L, while F85 was frequently replaced by a tyrosine residue.

<<Insert Table 3 here>>

Instead, the residue found at the position corresponding to F86 in Vf-TA was conserved in the closest homologues to B3-TA, but not in the sequenced obtained from the *Chloroflexus* genomes, where it was always a valine, and from the two *Roseiflexus* representatives, where it was the polar amino acid threonine.

Interestingly, the residues of the large binding pocket were almost completely conserved, including the "flipping arginine" R415. A certain variation was observed only at the position

corresponding to L417 in Vf-TA, that can be occupied by either polar amino acids, e.g., threonine, or hydrophobic ones, e.g., valine. However, a quite low level of amino acidic conservation at this position has been previously reported also among mesophilic ATAs (Genz et al. 2015).

Discussion

ATAs are valuable biocatalysts for the preparation of optically pure amines, being capable of performing reductive amination reactions using either an amino acid or simple aliphatic amines as amine donor. However, the number of available ATAs for industrial applications is still quite limited and this is especially true for (thermo)stable enzymes suitable to be used under harsh conditions such as at high temperature and/or in the presence of organic solvents.

In this work we searched for new (thermo)stable ATAs in the metagenomes of samples collected in hot terrestrial environments in different places in the world, i.e. in Er-Yuan, Yunnan province, China, in Krísuvík and Grensdalur, Iceland, and in the Solfatara Pozzuoli, Naples, Italy, at temperatures ranging from 55 and 95°C. The metagenomic samples were carefully chosen to provide a high global biodiversity, with different relative abundance of bacterial and archaeal representatives and communities structure (Menzel et al. 2015).

ATAs producers demonstrated to be relatively rare in these harsh environments - at least on a basis of homology search - only a few homologues of these enzymes being found in the large sequence information of the samples analyzed. Moreover, the sequences found in the Chinese sample Ch2-EY55S were only partials, possibly because the N50, i.e. the shortest sequence length at 50% of this metagenome, was only 828 nt, while ATA coding sequences are usually quite long (around 1300-1400 bp).

Furthermore, it must be remarked that only homologues to (*S*)-selective ATAs were identified. This result could be related to the fact that known (*S*)- and (*R*)-selective ATAs belong to completely different classes, PLP fold type I and PLP fold type IV, respectively, and show a rather different distribution in Nature. In fact, while the majority of (*S*)-selective ATAs have a bacterial origin, most of the (*R*)-selective ATAs described to date are produced by fungi (Höhne

et al. 2010; Sayer et al. 2014), and fungi were not represented in the microorganisms populations found in the environmental samples we analyzed (Menzel et al. 2015).

The three ATAs full sequences identified in Krísuvík, Grensdalur, and Naples samples, and named Is3-TA, B3-TA, and It6-TA, respectively, originated from hot springs showing temperatures above 76°C, thus suggesting a likely (hyper)thermophilic nature. None of them corresponded to already characterized enzymes, but BLAST analysis allows some speculations about their possible origin. For example, Is3-TA is almost identical to some putative transaminases (GenBank accession codes: CQR42634, ADG29592) from chemolithoautotrophic mesophilic betaproteobacteria belonging to the *Thiomonas* genus. Considering that also some thermophilic strains, such as the *T. islandicus* strain isolated from a hot spring in Grensdalur, Iceland (Vésteinsdóttir et al. 2011), belong to the same genus, we can guess that one of these strains could be the natural producer of Is3-TA. The high homology between B3-TA and the ATA from the thermophile *T. roseum* (Mathew et al. 2016a) suggests that this protein in Nature could be produced by a thermophile belonging to the genus *Thermomicrobium* or similar, while the similarity between It6-TA and a putative transaminase from *Acidothiobacillus caldus* (GenBank code: WP_014003823.1) is too low (78% identity) to draw conclusions about the natural producer of this protein.

The cloning and recombinant expression of the novel ATAs allowed their functional characterization. Concerning thermophilicity and thermostability studies, Is3-TA and It6-TA demonstrated to be moderate thermophilic enzymes, showing optimal activity at 50°C and apparent T_M of 79°C and 57°C, respectively, which were consistent with those reported in the literature for other characterized ATAs (Humble et al. 2012; Chen et al. 2016).

The most interesting results were obtained with B3-TA, which showed both exceptional thermophilicity and thermostability. In fact, this enzyme showed a constantly increasing activity at reaction temperatures up to 90°C and retained more than 40% of starting activity after two-weeks incubation at 80°C. These results, together with the estimated apparent T_M of 88°C, make B3-TA, to the best of our knowledge, the most thermostable natural ATA described to date. In fact, far less thermostability was shown for the ω-transaminases from *Sphaerobacter thermophilus* (Mathew et al. 2016b) and from *Geobacillus thermodenitrificans* (Chen et al. 2016). Even the highly similar *T. roseum* ATA, for which a comparable apparent T_M value of 87°C was estimated, loses completely its activity after 24 h at 60°C (Mathew et al. 2016a).

Since the reactions catalyzed by ATAs often require organic co-solvents for improving substrate/product solubility in aqueous reaction systems (Savile et al. 2010), the tolerance of the new ATAs toward organic co-solvents was tested as well. It resulted that the three ATAs can be used with most of the tested co-solvents with only few exceptions. Moreover, B3-TA was also tested in biphasic systems using water-immiscible solvents such as toluene or petroleum ether and showed satisfactory performances that match well the superior thermostability of this biocatalyst.

As far as the substrate scope of the novel ATAs concerns, the broad range of donor and acceptor substrates utilized by these novel biocatalysts, not very far from that shown by the widely employed *V. fluvialis* ATA, confirms their potential for synthetic applications. The three enzymes showed a markedly preferences for aromatic secondary amines bearing a small substituent on the other side, i.e., an ethyl group or less, thus indicating steric constraints in the small binding pocket. Unexpectedly, not only It6-TA, which displays the typical fingerprint of β-Ala:pyr TAs, but also Is3-TA and B3-TA showed a very good activity in the presence of β-alanine as amine donor, while, as reported in the literature (Steffen-Munsberg et al. 2015), Vf-TA was not active in the presence of this substrate. The role of the conserved residue S185 (present in It6-TA, but substituted by A185 in the other two novel ATAs, Fig. 1) is probably less important than previously supposed on the basis of sequences comparison (Steffen-Munsberg et al. 2015), or other substitutions might help the recognition of β-alanine by Is3-TA and B3-TA. Furthermore, it might be possible that, as suggested by Sayer et al. (2013) in the case of the β-Ala:pyr TA from *Pseudomonas aeruginosa*, the ability to transform β-alanine is correlated to a more rigid active site than that occurring in "high activity" ATAs. No detectable activities were measured in the presence of other β-amino acids, such as DL-β-phenylalanine and DL-β-homoleucine, thus confirming the preference of the novel enzymes for not sterically hindering substrates.

The outstanding stability of B3-TA toward different reaction conditions prompted us in further investigating about its catalytic features by constructing and analyzing a structural model. Interestingly, one of the best templates ranked by the SWISS-MODEL server was the X-ray structure of the mesophilic (T_M 60°C) *V. fluvialis* ATA (Genz et al. 2016) containing a bound PMP molecule in the active site and showing only 41% sequence identity with B3-TA. The model inspection allowed a clear identification of B3-TA active site and a comparison with the

mesophilic counterpart Vf-TA in the respect of key catalytic residues, as well as of those involved in shaping the substrate binding pockets.

Generally speaking, the overall active site architecture of Vf-TA was largely maintained in B3-TA. The most relevant difference was the V23 substitution of the corresponding F19 residue in the Vf-TA small binding pocket that apparently increases the available space in this region. However, this change did not result in an increase of activity toward bulky substrates, B3-TA showing quite strict steric constraints even in the presence of a propyl substituent as in the case of (*S*)-α-phenylbutylamine. This result is not totally surprising taken into account the difficulties encountered by different research group in broadening the substrate specificity of (*S*)-selective ATAs and, in particular, in enlarging the small binding pocket for the acceptance of bulky amines (Midelfort et al. 2013; Genz et al. 2015; Han et al. 2015; Genz et al. 2016; Pavlidis et al. 2016). It is noteworthy that the F19 position in Vf-TA has been frequently addressed by mutagenesis studies aimed at introducing smaller residues, but in all cases the changes resulted in either significantly less active or less soluble variants. Therefore, it is likely that other amino acidic substitutions in B3-TA allow to compensate for the lack of this aromatic residue in order to achieve an active and properly folded enzyme.

In conclusion, this work further demonstrates the potential of metagenomics in the search of relatively rare enzymatic activities such as ATAs in extreme environments. Moreover, the identification of a subfamily of B3-TA-like transaminases, mostly uncharacterized and all from thermophilic microorganisms, may pave the way for the discovery of other useful biocatalysts. In fact, sequence and phylogenetic analysis of the members of this subfamily suggested an overall conservation of the key features necessary for their catalytic activity in amine synthesis. At the same time, given their origin and some variations observed at the active site level, it can be foreseen that members of this subfamily could possess useful applicative features, such as high stability under harsh reaction conditions and different substrate specificities.

Acknowledgments

We warmly thank Dr. Marta Vanoni (ICRM-CNR) for valuable technical assistance in enzymes production and purification.

Compliance with ethical standards

Funding

This work has been carried out in the framework of the HOTRAM project, a FP7 Marie Curie IEF post-doctoral program (grant agreement PIEF-GA-2013-622732 to EEF).

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:252–258. doi: 10.1093/nar/gku340

Chen Y, Yi D, Jiang S, Wei D (2016) Identification of novel thermostable taurine–pyruvate transaminase from *Geobacillus thermodenitrificans* for chiral amine synthesis. Appl Microbiol Biotechnol 100:3101–3111. doi: 10.1007/s00253-015-7129-5

de Miguel Bouzas T, Barros-Velázquez J, Villa TG (2006) Industrial applications of hyperthermophilic enzymes: a review. Protein Pept Lett 13:645–651. doi: 10.2174/092986606777790548

DeCastro M-E, Rodríguez-Belmonte E, González-Siso M-I (2016) Metagenomics of thermophiles with a focus on discovery of novel thermozymes. Front Microbiol 7:1521. doi: 10.3389/fmicb.2016.01521

DeLano W.L. (2002) The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

Fernández-Arrojo L, Guazzaroni ME, López-Cortés N, Beloqui A, Ferrer M (2010) Metagenomic era for biocatalyst identification. Curr Opin Biotechnol 21:725–733. doi: 10.1016/j.copbio.2010.09.006

Ferrandi EE, Sayer C, Isupov MN, Annovazzi C, Marchesi C, Iacobone G, Peng X, Bonch-Osmolovskaya E, Wohlgemuth R, Littlechild JA, Monti D (2015) Discovery and characterization of thermophilic limonene-1,2-epoxide hydrolases from hot spring metagenomic libraries. FEBS J 282:2879–2894. doi: 10.1111/febs.13328

Ferrer M, Martinez-Martinez M, Bargiela R, Streit WR, Golyshina O V., Golyshin PN (2016) Estimating the success of enzyme bioprospecting through metagenomics: Current status and future trends. Microb Biotechnol 9:22–34. doi: 10.1111/1751-7915.12309

Fuchs M, Farnberger JE, Kroutil W (2015) The industrial age of biocatalytic transamination. European J Org Chem 6965–6982. doi: 10.1002/ejoc.201500852

Genz M, Melse O, Schmidt S, Vickers C, van den Bergh T, Joosten H-J, Dörr M, Bornscheuer U (2016) Engineering the amine transaminase from *Vibrio fluvialis* towards branched-chain substrates. ChemCatChem 8:1–5. doi: 10.1002/cctc.201601007

Genz M, Vickers C, van den Bergh T, Joosten HJ, Dörr M, Höhne M, Bornscheuer UT (2015) Alteration of the donor/acceptor spectrum of the (*S*)-amine transaminase from *Vibrio fluvialis*. Int J Mol Sci 16:26953–26963. doi: 10.3390/ijms161126007

Han SW, Park ES, Dong JY, Shin JS (2015) Expanding substrate specificity of ω-transaminase by rational remodeling of a large substrate-binding pocket. Adv Synth Catal 357:2712–2720. doi: 10.1002/adsc.201500239

Höhne M, Schätzle S, Jochens H, Robins K, Bornscheuer UT (2010) Rational assignment of key motifs for function guides in silico enzyme identification. Nat Chem Biol 6:807–813. doi: 10.1038/nchembio.447

Humble MS, Cassimjee KE, Håkansson M, Kimbung YR, Walse B, Abedi V, Federsel H-J, Berglund P, Logan DT (2012) Crystal structures of the *Chromobacterium violaceum* ω-transaminase reveal major structural rearrangements upon binding of coenzyme PLP. FEBS J 279:779–792. doi: 10.1111/j.1742-4658.2012.08468.x

Kielbasa SM, Wan R, Sato K, Horton P, Frith MC (2011) Adaptive seeds tame genomic sequence comparison. Genome Res 21:487–493. doi: 10.1101/gr.113985.110

Koszelewski D, Clay D, Rozzell D, Kroutil W (2009) Deracemisation of α-chiral primary amines by a one-pot, two-step cascade reaction catalysed by ω-transaminases. European J Org Chem 2289–2292. doi: 10.1002/ejoc.200801265

Koszelewski D, Lavandera I, Clay D, Guebitz GM, Rozzell D, Kroutil W (2008a) Formal asymmetric biocatalytic reductive amination. Angew Chemie - Int Ed 47:9337–9340. doi: 10.1002/anie.200803763

Koszelewski D, Lavandera I, Clay D, Rozzell D, Kroutil W (2008b) Asymmetric synthesis of optically pure pharmacologically relevant amines employing ω-transaminases. Adv Synth Catal 350:2761–2766. doi: 10.1002/adsc.200800496

Koszelewski D, Tauber K, Faber K, Kroutil W (2010) ω-Transaminases for the synthesis of non-racemic α-chiral primary amines. Trends Biotechnol 28:324–332. doi: 10.1016/j.tibtech.2010.03.003

Letunic I, Bork P (2011) Interactive Tree of Life v2: Online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39:475–478. doi: 10.1093/nar/gkr201

Littlechild JA (2015) Enzymes from extreme environments and their industrial applications. Front Bioeng Biotechnol 3:1–9. doi: 10.3389/fbioe.2015.00161

Lorenz P, Schleper C (2002) Metagenome - A challenging source of enzyme discovery. J Mol Catal B Enzym 19–20:13–19. doi: 10.1016/S1381-1177(02)00147-9

Martin AR, DiSanto R, Plotnikov I, Kamat S, Shonnard D, Pannuri S (2007) Improved activity and thermostability of (S)-aminotransferase by error-prone polymerase chain reaction for the production of a chiral amine. Biochem Eng J 37:246–255. doi: 10.1016/j.bej.2007.05.001

Mathew S, Deepankumar K, Shin G, Hong EY, Kim B-G, Chung T, Yun H (2016a) Identification of novel thermostable ω-transaminase and its application for enzymatic synthesis of chiral amines at high temperature. RSC Adv 6:69257–69260. doi: 10.1039/C6RA15110H

Mathew S, Nadarajan SP, Chung T, Park HH, Yun H (2016b) Biochemical characterization of thermostable ω-transaminase from *Sphaerobacter thermophilus* and its application for producing aromatic β- and γ-amino acids. Enzyme Microb Technol 87–88:52–60. doi: 10.1016/j.enzmictec.2016.02.013

Mathew S, Yun H (2012) ω-Transaminases for the production of optically pure amines and unnatural amino acids. Acs Catal 2:993–1001. doi: 10.1021/cs300116n

Menzel P, Gudbergsdóttir SR, Rike AG, Lin L, Zhang Q, Contursi P, Moracci M, Kristjansson JK, Bolduc B, Gavrilov S, Ravin N, Mardanov A, Bonch-Osmolovskaya E, Young M, Krogh A, Peng X (2015) Comparative metagenomics of eight geographically remote terrestrial hot springs. Microb Ecol 70:411–424. doi: 10.1007/s00248-015-0576-9

Midelfort KS, Kumar R, Han S, Karmilowicz MJ, McConnell K, Gehlhaar DK, Mistry A, Chang JS, Anderson M, Villalobos A, Minshull J, Govindarajan S, Wong JW (2013) Redesigning and characterizing the substrate specificity and activity of *Vibrio fluvialis* aminotransferase for the synthesis of imagabalin. Protein Eng Des Sel 26:25–33. doi: 10.1093/protein/gzs065

Monti D, Forchin MC, Crotti M, Parmeggiani F, Gatti FG, Brenna E, Riva S (2015) Cascade coupling of ene-reductases and ω-transaminases for the stereoselective synthesis of diastereomerically enriched amines. ChemCatChem 7:3106–3109. doi: 10.1002/cctc.201500424

Nobili A, Steffen-Munsberg F, Kohls H, Trentin I, Schulzke C, Höhne M, Bornscheuer UT (2015) Engineering the active site of the amine transaminase from *Vibrio fluvialis* for the asymmetric synthesis of aryl-alkyl amines and amino alcohols. ChemCatChem 7:757–760.

doi: 10.1002/cctc.201403010

Nugent TC, El-Shazly M (2010) Chiral amine synthesis - Recent developments and trends for enamide reduction, reductive amination, and imine reduction. Adv Synth Catal 352:753–819. doi: 10.1002/adsc.200900719

Park ES, Kim M, Shin JS (2012) Molecular determinants for substrate selectivity of ω-transaminases. Appl Microbiol Biotechnol 93:2425–2435. doi: 10.1007/s00253-011-3584-9

Pavlidis IV, Weiß MS, Genz M, Spurr P, Hanlon SP, Wirz B, Iding H, Bornscheuer UT (2016) Identification of (*S*)-selective transaminases for the asymmetric synthesis of bulky chiral amines. Nat Chem 1–7. doi: 10.1038/NCHEM.2578

Remmert M, Biegert A, Hauser A, Söding J (2011) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods 9:173–175. doi: 10.1038/nmeth.1818

Sambrook J, MacCallum P, Russell D (2001) Molecular cloning: a laboratory manual (third edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Savile CK, Janey JM, Mundorff EC, Moore JC, Tam S, Jarvis WR, Colbeck JC, Krebber A, Fleitz FJ, Brands J, Devine PN, Huisman GW, Hughes GJ (2010) Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. Science 329:305–309. doi: 10.1126/science.1188934

Sayer C, Isupov MN, Westlake A, Littlechild JA (2013) Structural studies of *Pseudomonas* and *Chromobacterium* ω-aminotransferases provide insights into their differing substrate specificity. Acta Crystallogr Sect D Biol Crystallogr 69:564–576. doi: 10.1107/S0907444912051670

Sayer C, Martinez-Torres RJ, Richter N, Isupov MN, Hailes HC, Littlechild JA, Ward JM (2014) The substrate specificity, enantioselectivity and structure of the (*R*)-selective amine: pyruvate transaminase from *Nectria haematococca*. FEBS J 281:2240–2253. doi: 10.1111/febs.12778

Shin J, Kim B (2002) Exploring the active site of amine : pyruvate aminotransferase on the basis of the substrate structure - reactivity relationship : how the enzyme controls substrate specificity and stereoselectivity. J Org Chem 67:2848–2853. doi: 10.1021/jo016115i

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality

protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. doi: 10.1038/msb.2011.75

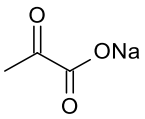
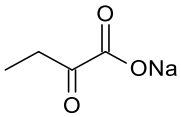
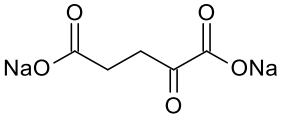
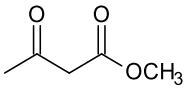
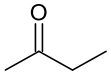
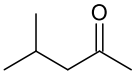
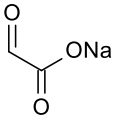
Steffen-Munsberg F, Vickers C, Kohls H, Land H, Mallin H, Nobili A, Skalden L, van den Bergh T, Joosten H-J, Berglund P, Höhne M, Bornscheuer UT (2015) Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications. Biotechnol Adv 33:566–604. doi: 10.1016/j.biotechadv.2014.12.012

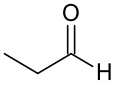
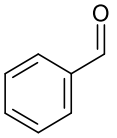
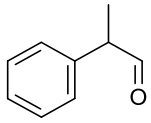
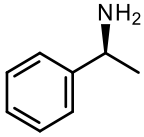
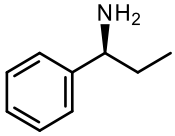
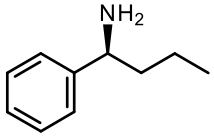
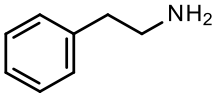
Vésteinsdóttir H, Reynisdóttir DB, Örlygsson J (2011) *Thiomonas islandica* sp. nov., a moderately thermophilic, hydrogen- and sulfur-oxidizing betaproteobacterium isolated from a hot spring. Int J Syst Evol Microbiol 61:132–137. doi: 10.1099/ijs.0.015511-0

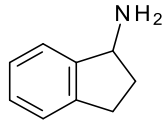
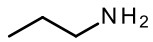
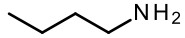
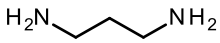
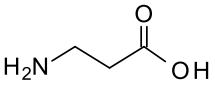
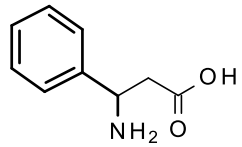
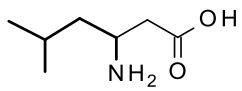
Wilson MC, Piel J (2013) Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. Chem Biol 20:636–647. doi: 10.1016/j.chembiol.2013.04.011

Zarafeta D, Kissas D, Sayer C, Gudbergsdottir SR, Ladoukakis E, Isupov MN, Chatziioannou A, Peng X, Littlechild JA, Skretas G, Kolisis FN (2016) Discovery and characterization of a thermostable and highly halotolerant GH5 cellulase from an icelandic hot spring isolate. PLoS One 11:e0146454. doi: 10.1371/journal.pone.0146454

Table 1 Substrate scope of the novel ATAs

Substrate	**Conversion (%)[a]**			
	Is3-TA	**It6-TA**	**B3-TA**	**Vf-TA**
Amino Acceptor				
Pyruvate 	48	45	51	50
2-Oxobutyrate 	>99	>99	>99	>99
α-Ketoglutarate 	8	2	4	8
Methyl acetoacetate 	16	6	5	15
Acetone 	7	2	8	5
2-Butanone 	3	2	5	11
Methyl isobutyl ketone 	1	1	5	10
Glyoxylate 	>99	62	>99	>99

Acetaldehyde 	6	6	68	18
Propionaldehyde 	74	10	33	34
Benzaldehyde 	3	10	10	15
2-Phenylpropionaldehyde 	16	3	16	19
Amino Donor				
(*S*)-α-Methylphenylamine 	50	41	43	>99
(*S*)-α-Ethylphenylamine 	63	46	23	>99
(*S*)-α-Phenylbutylamine 	2	3	1	>99
2-Phenylethylamine 	8	18	30	15

(*R*,*S*)-1-Aminoindan 	97	89	43	99
Isopropylamine 	11	27	13	28
Propylamine 	6.5	3.7	n.d.[b)]	1.2
Butylamine 	1.4	3.5	1	6.2
1,3-Diaminopropane 	0.7	3	2	n.d.
β-Alanine 	67	57	29	n.d.
DL-β-Phenylalanine 	n.d.	n.d.	n.d.	n.d.
DL-β-Homoleucine 	n.d.	n.d.	n.d.	n.d.

a) Reactions were performed in triplicate and conversions were determined after 24 h by GC/MS or HPLC analysis (see Materials and methods for details). Standard deviation was below 5%, no significant degradation of substrates/products or formation of by-products was observed; b) n.d.: below detection limit.

Table 2 NCBI Reference Sequence, source and similarity of B3-TA homologues

Reference Sequence	Organism	Identity
WP_015922033.1	*Thermomicrobium roseum*	92%
WP_038038660.1	*Thermorudis peleae*	80%
WP_012872904.1	*Sphaerobacter thermophiles*	72%
WP_008476568.1	*Nitrolancea hollandica*	65%
WP_027882148.1	*Meiothermus rufus*	66%
WP_013014879.1	*Meiothermus ruber*	64%
WP_027886436.1	*Meiothermus* sp.	66%
WP_013159295.1	*Meiothermus silvanus*	64%
WP_015941631.1	*Chloroflexus aggregans*	52%
WP_031459398.1	*Chloroflexus* sp. MS-G	50%
WP_044231836.1	*Chloroflexus* sp. Y-396-1	51%
WP_012119566.1	*Roseiflexus castenholzii*	50%
WP_012258069.1	*Chloroflexus* sp.	51%
WP_066785535.1	*Chloroflexus* sp. isl-2	51%
WP_011955192.1	*Roseiflexus* sp. RS-1	50%

Table 3 Amino acid variations in active site architecture among members of the B3-TA-like subfamily

Source	Selected active site residues											
	Small binding pocket					Large binding pocket						
	150	**153**	**19**	**85***	**86***	**56**	**57**	**415**	**417**	**228**	**259**	**258**
Vibrio fluvialis	**Y**	**V**	**F**	**F**	**F**	**L**	**W**	**R**	**L**	**A**	**I**	**V**
Thermorudis peleae	Y	I	V	F	F	L	W	R	T	A	I	V
Thermomicrobium roseum	Y	I	L	F	F	L	W	R	T	A	I	V
B3-TA	Y	I	V	F	F	L	W	R	T	A	I	V
Sphaerobacter thermophiles	Y	I	V	F	F	L	W	R	N	A	I	V
Nitrolancea hollandica	Y	I	V	Y	F	L	W	R	G	A	I	V
Meiothermus silvanus	Y	I	V	F	F	L	W	R	N	A	I	V
Meiothermus rufus	Y	I	V	F	F	L	W	R	N	A	I	V
Meiothermus ruber	Y	I	V	F	F	L	W	R	N	A	I	V
Meiothermus sp.	Y	I	V	F	F	L	W	R	N	A	I	V
Chloroflexus aggregans	Y	V	L	Y	V	L	W	R	V	A	I	V
Chloroflexus sp. MS-G	Y	V	L	Y	V	L	W	R	V	A	I	V
Chloroflexus sp. Y-396-1	Y	V	L	Y	V	L	W	R	V	A	I	V
Chloroflexus sp.	Y	V	L	Y	V	L	W	R	V	A	I	V
Chloroflexus sp. isl-2	Y	V	L	Y	V	L	W	R	V	A	I	V
Roseiflexus castenholzii	Y	V	L	Y	T	L	W	R	L	A	I	V
Roseiflexus sp. RS-1	Y	V	L	Y	T	L	W	R	L	A	I	V

Figure Captions

Fig. 1 Comparison between the occurrence of key residues in "high activity" ATAs, Is3-TA, B3-TA, It6-TA, and β-alanine:pyruvate transaminases (β-Ala:pyr TA) (Steffen-Munsberg et al. 2015). Conserved residues (more than 70% of the subset considered in (Steffen-Munsberg et al. 2015)) are shown in capital letters, while residues with low degree of conservation (30-70%) are shown in lower case letters. The colour code indicates the physicochemical properties of the residues. Numbering scheme is according to the numbering scheme of (Steffen-Munsberg et al. 2015), and the locations of corresponding amino acids are indicated in the alignment reported in Supplementary Material, Fig. S1

Fig. 2 Influence of pH (**a**) and temperature (**b**) on ATAs activity, influence of temperature (**c**) on ATAs stability, and long-term stability of B3-TA at 80°C (**d**). For details, see Supplementary Material, Tables S4-S7

Fig. 3 Influence of organic co-solvents on ATAs stability. Residual activity was estimated after 5 h of incubation at room temperature

Fig. 4 Enantioselectivity of the novel ATAs. Specific activities (U/mg) of ATAs and Vf-TA towards different (*S*) and (*R*) aromatic amines were determined spectrophotometrically as described in Nobili et al. (2015). For details, see Supplementary Material, Tables S8

Fig. 5 Simulated structure and active site model of B3-TA. (**a**) Overall structure alignment of B3-TA (*red*) and *V. fluvialis* ATA (PDB: 4E3Q, *green*). (**b**) Active pocket detail (B3-TA, *red*; Vf-TA, *green*; residues numbering: Vf-TA/B3-TA), PMP is colored yellow (color by element). The large binding pocket is indicated by the orange arc, while the light blue arc indicates the small binding pocket. Images generated in PYMOL (DeLano W.L. 2002)

Fig. 6 Phylogenetic analysis of B3-TA homologues identified in the NCBI database

Figure 1

Enzyme activity	*Fingerprint residues*												
	16	*45*	*46*	*47*	*132*	*145*	*185*	*216*	*267*	*269*	*346*	*348*	*353*
ATA (High activity)	f	G	L	W	s	m	A	I	n	G	R	v	i
Is3-TA	L	G	L	W	S	M	A	I	N	G	R	V	I
B3-TA	L	G	L	W	I	Y	A	I	M	G	R	T	I
It6-TA	F	G	L	W	V	N	S	I	F	G	R	T	I
β-Ala:pyr TA	f	G	L	W	V	N	S	I	f	G	R	t	a

Figure 2

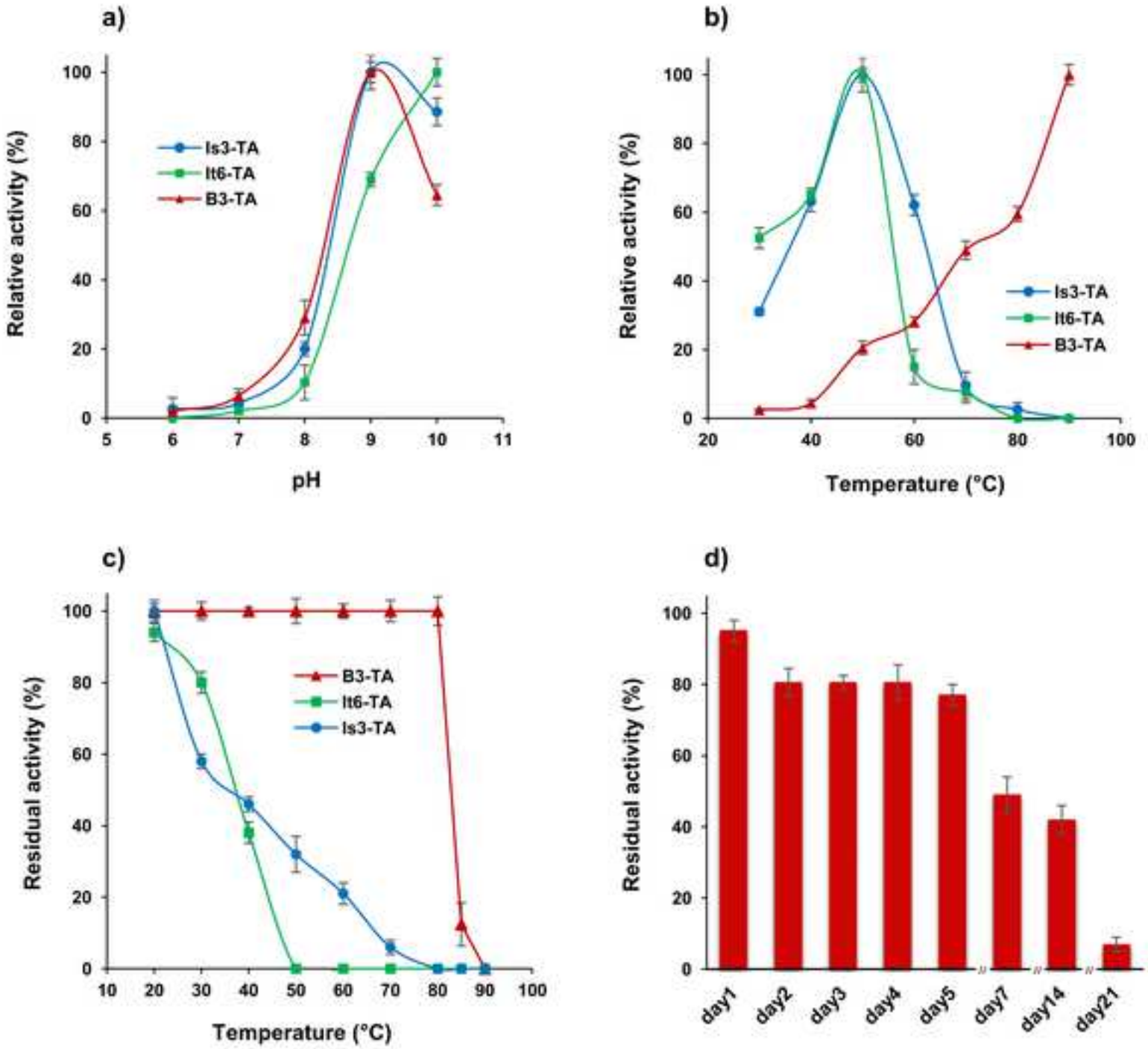


Figure 3

Click here to download Figure Figure 3_revised.tif

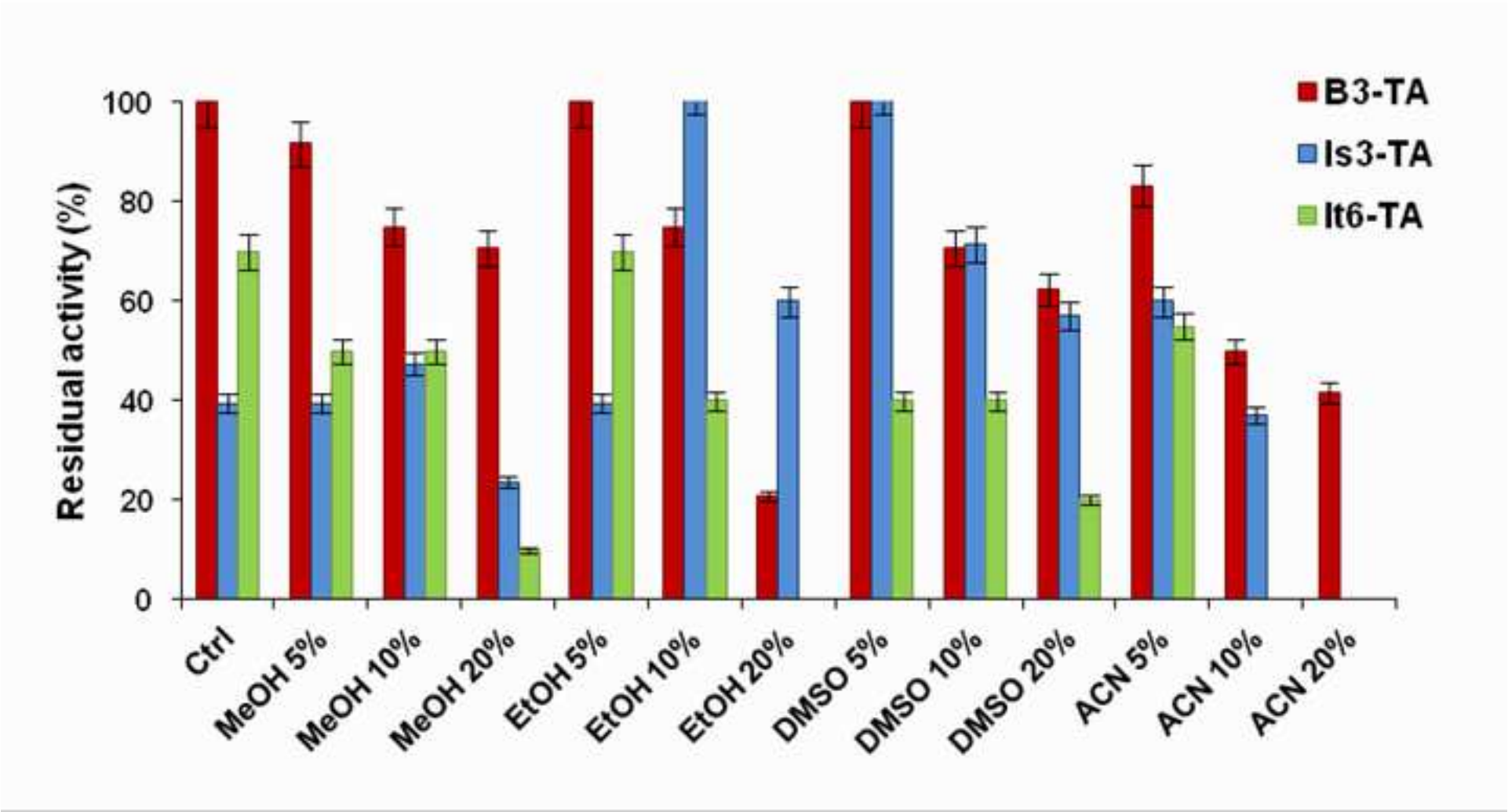


Figure 4

Click here to download Figure Figure 4_revised.tif

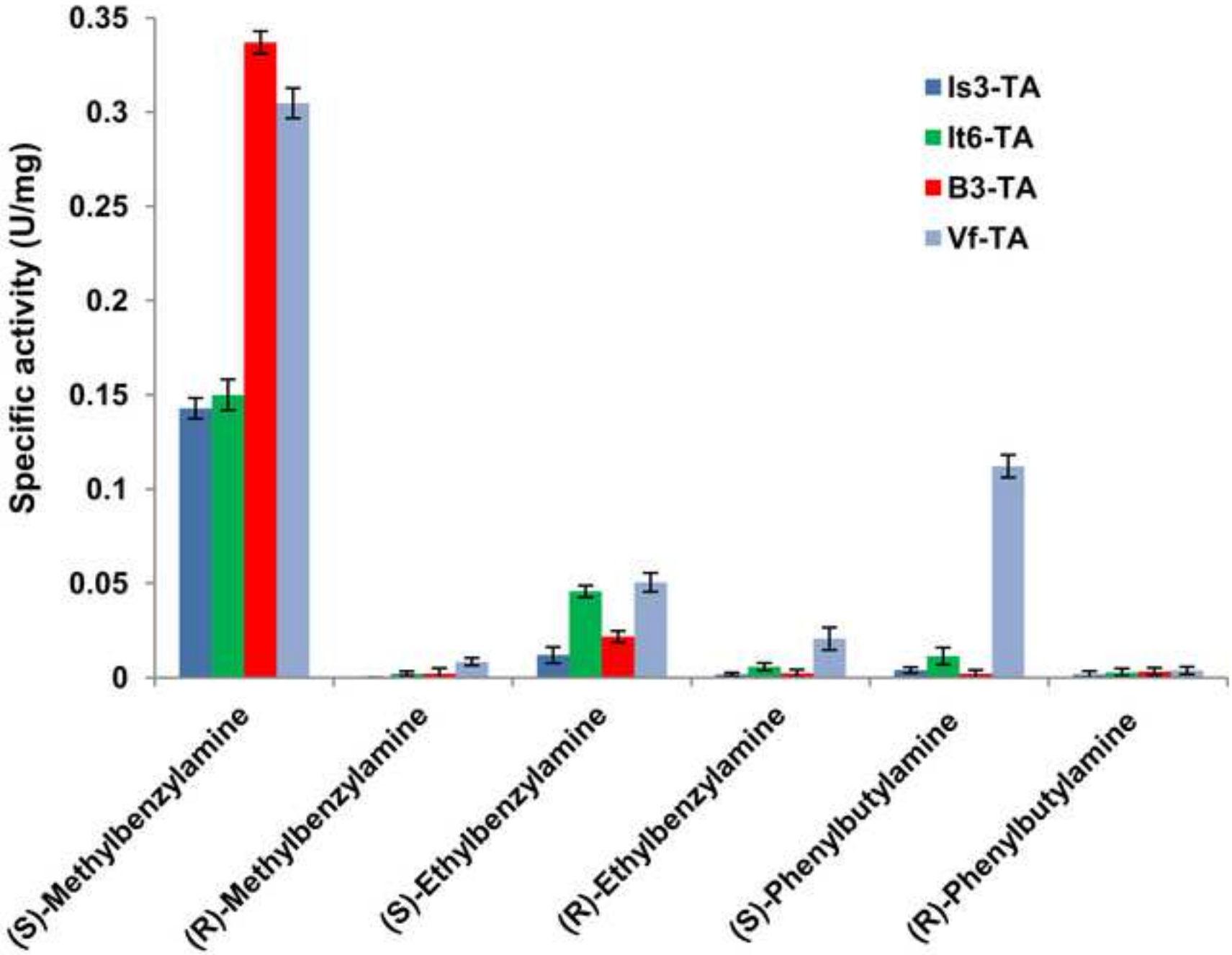


Figure 5

Click here to download Figure Figure 5_revised.tif

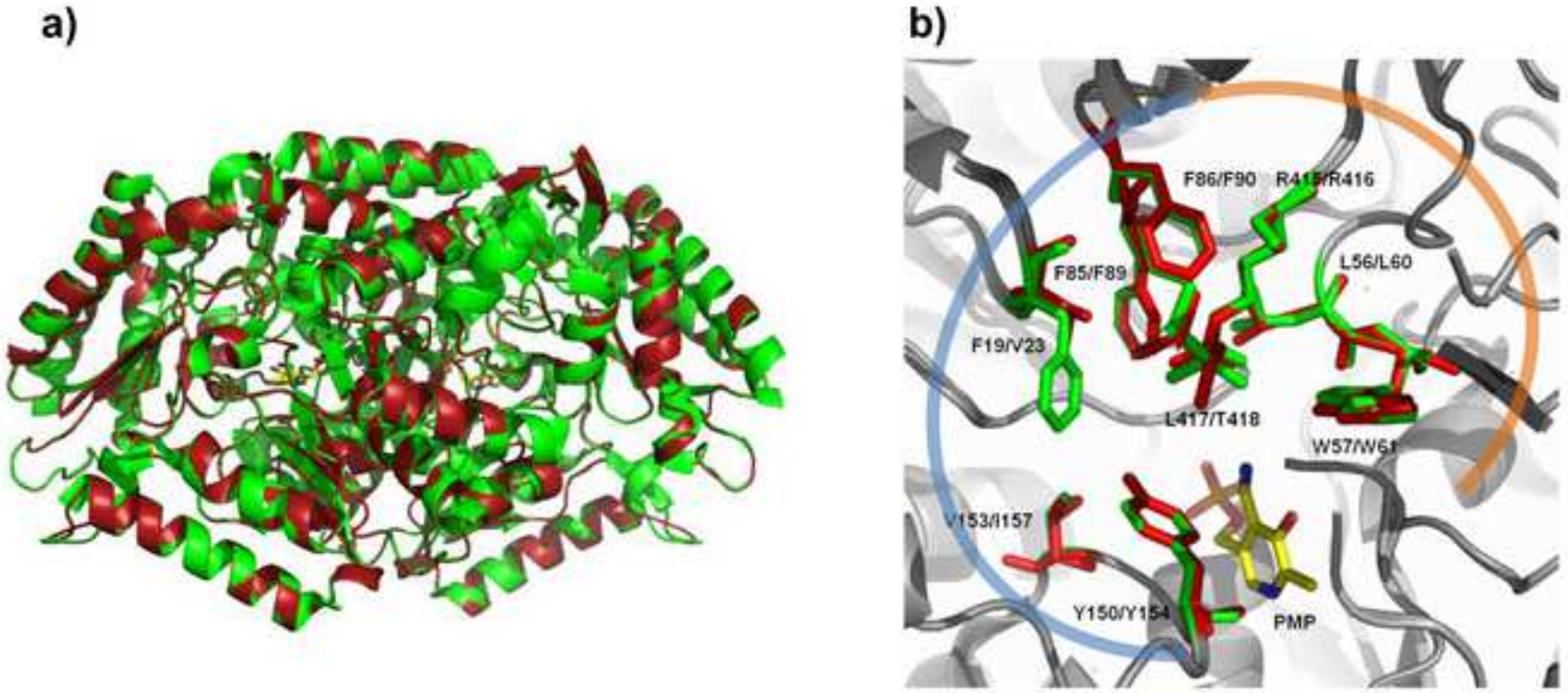


Figure 6

Click here to download Figure Figure 6_revised2.tif