Nucleic Acids Research, 2024, **52**, D476–D482
https://doi.org/10.1093/nar/gkad1090
Advance access publication date: 20 November 2023
Database issue

iNClusive: a database collecting useful information on non-canonical amino acids and their incorporation into proteins for easier genetic code expansion implementation

Leon-Samuel Icking[1,†], Andreas Martin Riedlberger[1,†], Fabian Krause[1,‡], Jonas Widder[1,‡], Anne Smedegaard Frederiksen[1,‡], Fabian Stockert[1,‡], Michael Spädt[1,‡], Nikita Edel[1,‡], Daniel Armbruster[1,2], Giada Forlani[1,2,3], Selene Franchini[1], Paulina Kaas[4], Büşra Merve Kırpat Konak[1,2], Fabrice Krier[1,2], Maïwenn Lefebvre[1,2], Daniel Mazraeh[1,2], Jeremy Ranniger[1], Johanna Gerstenecker[1], Pia Gescher[1], Karsten Voigt[5], Pavel Salavei[1], Nicole Gensch[1], Barbara Di Ventura[1,2,*] and Mehmet Ali Öztürk[1,2,*]

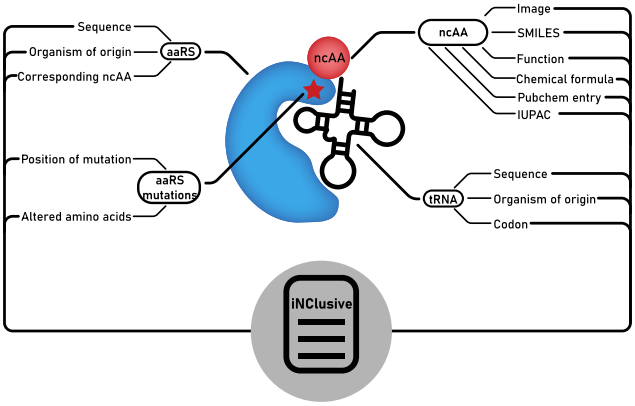
[1]Signalling Research Centres BIOSS and CIBSS, University of Freiburg, Schänzlestr. 18, 79104, Freiburg, Germany
[2]Institute of Biology II, Faculty of Biology, University of Freiburg, Schänzlestr. 1, 79104, Freiburg, Germany
[3]Spemann Graduate School of Biology and Medicine (SGBM), University of Freiburg, Albertstr. 19A, 79104, Freiburg, Germany
[4]Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Meyerhofstraße 1, 69117, Heidelberg, Germany
[5]Institute of Biology III, Faculty of Biology, University of Freiburg, Schänzlestr. 1, 79104, Freiburg, Germany

[*]To whom correspondence should be addressed. Tel: +49 761 203 2787; Email: mehmet.oeztuerk@bioss.uni-freiburg.de
Correspondence may also be addressed to Barbara Di Ventura. Tel: +49 761 203 2764; Email: barbara.diventura@bio.uni-freiburg.de
[†]The authors acknowledge that the first two are joint first authors.
[‡]Co-second authors.

Abstract

The incorporation of non-canonical amino acids (ncAAs) into proteins is a powerful technique used in various research fields. Genetic code expansion (GCE) is the most common way to achieve this: a specific codon is selected to be decoded by a dedicated tRNA orthogonal to the endogenous ones. In the past 30 years, great progress has been made to obtain novel tRNA synthetases (aaRSs) accepting a variety of ncAAs with distinct physicochemical properties, to develop robust *in vitro* assays or approaches for codon reassignment. This sparked the use of the technique, leading to the accumulation of publications, from which gathering all relevant information can appear daunting. Here we present iNClusive (https://non-canonical-aas.biologie.uni-freiburg.de/), a manually curated, extensive repository using standardized nomenclature that provides organized information on ncAAs successfully incorporated into target proteins as verified by mass spectrometry. Since we focused on tRNA synthetase-based tRNA loading, we provide the sequence of the tRNA and aaRS used for the incorporation. Derived from more than 687 peer-reviewed publications, it currently contains 2432 entries about 466 ncAAs, 569 protein targets, 500 aaRSs and 144 tRNAs. We foresee iNClusive will encourage more researchers to experiment with ncAA incorporation thus contributing to the further development of this exciting technique.

Graphical abstract



Received: August 14, 2023. **Revised:** October 27, 2023. **Editorial Decision:** October 28, 2023. **Accepted:** October 30, 2023

Introduction

In nature, 20 amino acids (AAs) are encoded by dedicated nucleotide triplets and are naturally incorporated into proteins during the process of translation by all organisms. Some methanogenic archaea and bacteria additionally incorporate pyrrolysine into proteins at the amber STOP codon (UAG) (1). Selenocysteine is also not encoded by a dedicated sense nucleotide triplet, but it can be incorporated into proteins at the opal STOP codon (UGA) in the presence of an *in-cis* structural element in the selenoprotein mRNA (2). These 22 AAs are referred to as canonical or proteinogenic to differentiate them from the non-proteinogenic or non-canonical AAs (ncAAs), which occur in cells, but are not used directly as building blocks by the ribosome. This does not mean that they do not appear in proteins at all; indeed, some canonical AAs can be turned into ncAAs thanks to post-translational modifications (e.g. proline can become hydroxyproline after hydroxylation and glutamic acid can become carboxyglutamic acid after carboxylation). Beyond their function as part of proteins, ncAAs have many important physiological roles, such as in neurotransmission (3), bacterial cell wall synthesis (4), and metabolism (5). ncAAs are also the substrates of non-ribosomal peptide synthetases, mega enzymes that assemble non-ribosomal peptides with them as well as canonical AAs (6). Apart from those occurring in nature, ncAAs can be chemically synthesized in the laboratory (7). In this case, one should rather speak of unnatural AAs; for simplicity, however, we will refer to all these non-proteinogenic AAs as ncAAs.

Given their unique physicochemical properties, ncAAs incorporated into proteins would potentially equip them with new functionalities, tremendously expanding the natural proteome. Already back in the 1950s, it has been shown that selenomethionine could be incorporated into proteins in place of methionine thanks to the natural substrate tolerance of the dedicated aminoacyl-tRNA synthetase (aaRS)(8). Since then, various strategies have been devised for residue-specific incorporation of ncAAs, involving the use of auxotrophic strains for a specific AA (9), *in vitro* chemical manipulation of the AA-loaded tRNAs (e.g. desulfurization of Cys-tRNA^{Cys} into Ala-tRNA^{Cys} using Raney nickel (10)), or *in vitro* chemical acylation of the tRNA with the ncAA (11). The exciting possibility to incorporate ncAAs at specific sites in proteins prompted scientists later on to develop a method beyond the residue-specific incorporation based on the transplantation of an aaRS-tRNA pair naturally assigned to a STOP codon from an organism into a distantly related one (12). This technology is called genetic code expansion (GCE) because effectively one additional amino acid gets incorporated into proteins in response to an anticodon (13–15). From those early days it has now reached a state mature enough to allow robust incorporation of ncAAs in living cells, be it bacteria, yeast or mammalian cells (16–19). While many exciting methods exist to expand/reprogram the genetic code for the site-specific incorporation of ncAAs into proteins, including the use of quadruplets of nucleotides (20) or artificial base pairs (21), currently the most commonly used is the Amber STOP codon suppression (22). This method lends itself well for usage in various organisms as well as *in vitro*, given the scarcity of this codon in the genome (23) and the existence of an aaRS-tRNA pair that naturally suppresses it, as mentioned earlier (1). Nonetheless, not only the translational STOP codon is encoded by three separate codons. The whole genetic code is degenerate, and each AA is encoded by several nucleotide triplets. Thus, codon reassignment could potentially take place without perturbing the natural proteome too much (however, one should keep in mind that codon usage plays non-negligible roles in the regulation of gene expression (24)). With the advent of genome engineering or *de novo* chromosome design, it is indeed becoming possible to also re-assign sense codons to ncAAs (25,26).

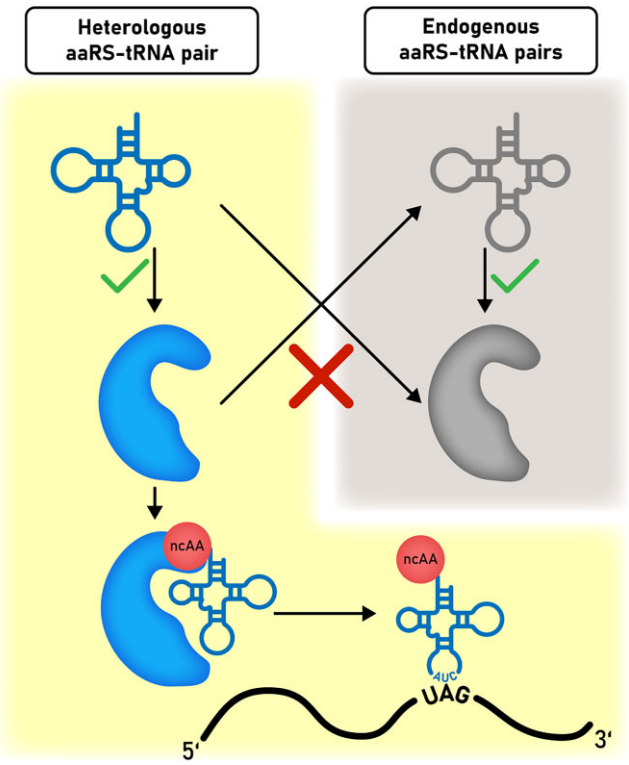


Figure 1. Scheme of the orthogonal translation system for the incorporation of ncAAs into target proteins. The heterologous aaRS and its cognate tRNA (blue) do not interact with the endogenous aaRSs and tRNAs (grey). The orthogonal aaRS recognizes the ncAA and loads it onto its cognate tRNA. The ncAA-loaded tRNA then interacts with endogenous ribosomes to allow for the translation of the target mRNA. In this example, the amber STOP codon (UAG) codes for the ncAA. aaRS, aminoacyl tRNA synthetase. ncAA, non-canonical amino acid. For simplicity only one endogenous aaRS–tRNA pair is shown (grey box).

Regardless of whether a STOP or sense codon is employed, to successfully incorporate the ncAA only into a target protein at a specific site during translation in a living cell, several requirements must be met (Figure 1). First, the aaRS-tRNA pair must at the same time function congruently and orthogonally to the translational machinery of the host cell. Indeed, the ncAA-loaded tRNA should still be able to interact with the ribosome; yet the heterologous aaRS and the endogenous aaRSs should not load endogenous canonical AAs onto the heterologous tRNA. Orthogonality usually also requires that the nucleotide triplet redirected to code for the ncAA be either not present in the genome anymore (27) – thus ensuring that the ncAA is not incorporated in other non-target proteins whose mRNAs contain that triplet–, or that the mRNA of the target protein be sequestered into a separate 'organelle' together with the aaRS-tRNA pair to keep it away from all other endogenous mRNAs (19). Second, the aaRS must recognize the ncAA and use it as a substrate. If this is not the case, it must either be mutated/evolved to do so or another

aaRS must be sought that has this property. The engineering of aaRSs with novel substrate specificities has become almost routine nowadays (28).

Despite the tremendous advancement of the field (29), a single resource gathering all information needed to effectively embrace this technique in the laboratory was still missing. We set ourselves the goal to create such a repository. We called it iNClusive (for 'inclusion of Non-Canonical AAs into proteins').

iNClusive database creation

Selection of publications

There are several assays that can be used to prove the successful incorporation of ncAAs into target proteins. Some are indirect –they show the 'effect' of having the ncAA in the target protein, for instance cross-linking to a fluorescent dye (30)–, while others offer direct proof of the presence of the ncAA in the polypeptide chain: mass spectrometry (MS), nuclear magnetic resonance (NMR) and X-ray crystallography. We decided to focus on publications in which MS has been applied to prove the presence of the ncAA into the target protein. Despite this choice excludes many interesting and valuable works, we had to find a compromise between exhaustiveness and manageability. Since the mining of the information from publications and public repositories (see below in Data mining and Discussion) was far from trivial and quite time-consuming, we had to restrict the number of publications to be analysed. Considering a direct proof superior to an indirect one, we opted for MS. NMR could have been just as good a choice (X-ray crystallography is still quite rare in this context).

Moreover, we decided to focus on publications in which the loading of the tRNA with the ncAA was done via a tRNA synthetase. Again, this selection has been done to limit the number of publications to consider. We are aware alternative methods exist (see Introduction) and our decision at this stage does not represent a stance in favour of the usage of tRNA synthetases.

To retrieve these publications, we used the 'Publish or Perish' software (31) to search Google Scholar for terms typical of studies involving ncAAs incorporated into proteins and MS as validation method (Figure 2). 'Publish or Perish' allows setting keywords and saving the bibliographic data of the papers as comma-separated value files, which can be later used for downstream analysis.

As there is no standardized nomenclature in the field (e.g. some researchers speak of ncAAs, while other of non-standard AAs; unnatural AAs might not be included in the term ncAAs etc.) and considering that the same term could be written differently in different journals (for instance noncanonical *vs* non-canonical), we used several combinations of search terms to retrieve the publications from which to extract the data (Table 1). Before processing the hits for the database, duplicates, as well as reviews, patents and master or doctoral theses were removed, leaving 687 peer-reviewed publications to manually analyse.

Table 1. Keyword combinations used to find publications on ncAA incorporation into proteins proven by mass spectrometry

Mass spectrometry search term	UAA	ncAA	nsAA
'Mass spectrometry'	1960	752	112
'Electrospray ionization'	50	8	0
'Electrospray ionisation'	6	1	0
'MALDI'	102	36	5
'MS/MS'	74	16	0
'LC MS'	145	37	9
'GC MS'	20	4	0
'HPLC MS'	16	2	0

Date of search: 22.09.2023. For each keyword search the term 'tRNA Synthetase' was additionally used. 'unnatural amino acid', 'non-canonical amino acid', and 'non-standard amino acid' are represented by UAA, ncAA, and nsAA, respectively. Different versions of 'non canonical' and 'non standard' ('noncanonical', 'non-standard') were not used because the 'Publish or Perish' software already replaces the blank during the search for a hyphen or eliminates it. Without removing duplications, 3355 hits were gathered. For all mass spectrometry search terms except for 'Mass spectrometry', the term 'Mass spectrometry' was excluded with '-'mass spectrometry" to limit the amount of duplicates.

Data mining

The publications were processed manually due to the information of interest being presented in different formats. For example, mutations in the aaRS may be presented in the text, in the figures, in the supplementary information, given in another publication cited within the publication being processed or omitted all together.

From each publication, we retrieved the following information:

1) abbreviation and name of ncAA
2) organism of origin and natural substrate of the aaRS used to incorporate the ncAA
3) mutations applied to the aaRS. Note that we decided to assign to each aaRS a unique name composed of the following information: abbreviation of the organism of origin; abbreviation of the natural substrate of the aaRS; the letters 'RS'; and, if applicable, the mutations carried by the synthetase. For example, Ec-MetRS is the aaRS from *E. coli* that naturally loads methionine on the corresponding tRNA
4) name of tRNA used for the incorporation. We decided to assign to each tRNA a name composed of three words: (a) abbreviation of the organism from which it was derived; (b) tRNA; (c) AA naturally transported by the tRNA. For example, 'Bs-tRNA Tyr' indicates a tRNA naturally found in *Bacillus subtilis*, which transports tyrosine
5) codon recognized by the tRNA
6) modifications made to the tRNA (if any)
7) protein in which the ncAA was incorporated
8) position of incorporation (if given)
9) organism in which incorporation was tested, if any (could be *in vitro*)
10) application for the ncAA (if given)
11) original publication in APA citation style
12) DOI link to the publication

We additionally collected the following information:

1) sequence of the aaRS. This was a very time-consuming and laborious task, because the sequences are typically not found in the publications themselves, but had to be retrieved elsewhere with information provided in the materials and methods section (e.g. plasmid name and number as deposited in a public repository such as Addgene (https://www.addgene.org/))

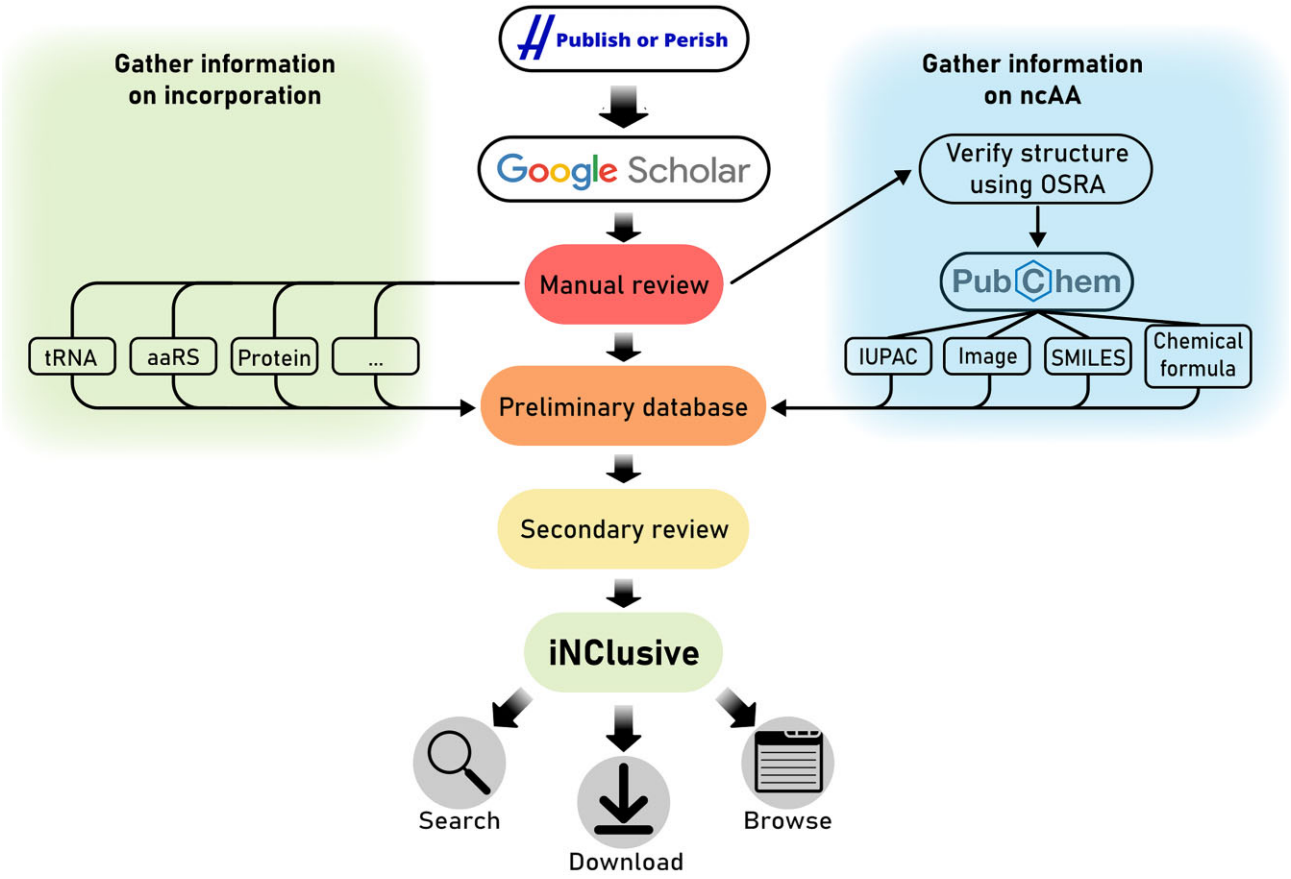


Figure 2. Scheme depicting the workflow used to generate the iNClusive database. The hits were generated using the 'Publish or Perish' software (31) upstream of Google Scholar. The publications were then manually reviewed and inserted into a preliminary database. The SMILES of the ncAAs were generated with the Optical Structure Recognition (OSRA) software (32) and compared with those given in the PubChem database (33). The retrieved information was confirmed manually by a different team member and only then added into the iNClusive database. A second review process done by a third distinct person was additionally carried out on 10% of randomly picked entries.

2) sequence of the tRNA. Also in this case, we had to read many publications and retrieve the sequence from other databases or repositories (e.g. GenBank (34) or Addgene)
3) molecular-input line-entry system (SMILES) and IUPAC name of the ncAA. To this aim, we uploaded images of the ncAAs into the 'Optical Structure Recognition A (OSRA)' software (32) and compared the generated SMILES with the entries of the PubChem database (33). Some of the ncAAs could unfortunately not be found on PubChem, but we have nevertheless provided the (SMILES) and the chemical formulas for these
4) AA closest in structure to the ncAA
5) comments (if applicable). These include, for example, Addgene links, information on flanking sequences on the tRNA, or specific conditions necessary for the incorporation of the ncAA

Figure 2 shows the entire workflow used to create iNClusive.

Website creation

The iNClusive website was built with React (https://react.dev/), a popular web framework, and utilises the Blueprint component library (https://blueprintjs.com/) for user interface design. We chose TypeScript (https://www.typescriptlang.org/) as the programming language because it provides type safety and enhanced code quality. The CSV data processing was handled by Papa Parse (https://www.papaparse.com/), ensuring efficient parsing. Additionally, the website incorporates structure images sourced from PubChem (https://pubchem.ncbi.nlm.nih.gov/).

Content and use

Currently, iNClusive has a total of 2432 distinct entries, with information on 466 different ncAAs that were introduced into 569 proteins by 500 different aaRSs, collected from 687 different publications. Modifications of the same protein at different positions by the same ncAA are grouped as a single entry, while modifications of the same protein with different ncAAs or incorporation of the same ncAA into different proteins appear as separate entries, even if the publication describing them is the same. Unique aaRS/tRNA pairs are calculated on the basis of the unique IDs we assigned to them. Users can explore the content online or export the dataset in a single file where the entries are given as comma-separated values (CSV). A search tool is available on the website allowing users to easily find and download specific information. As the database is quite extensive and not all columns can be visualized on the screen at the same time, we implemented the possibility to set up filters under 'Filter data'. With these filters it is possible to search either all or only certain columns for the desired terms. Several filters can also be applied at once,

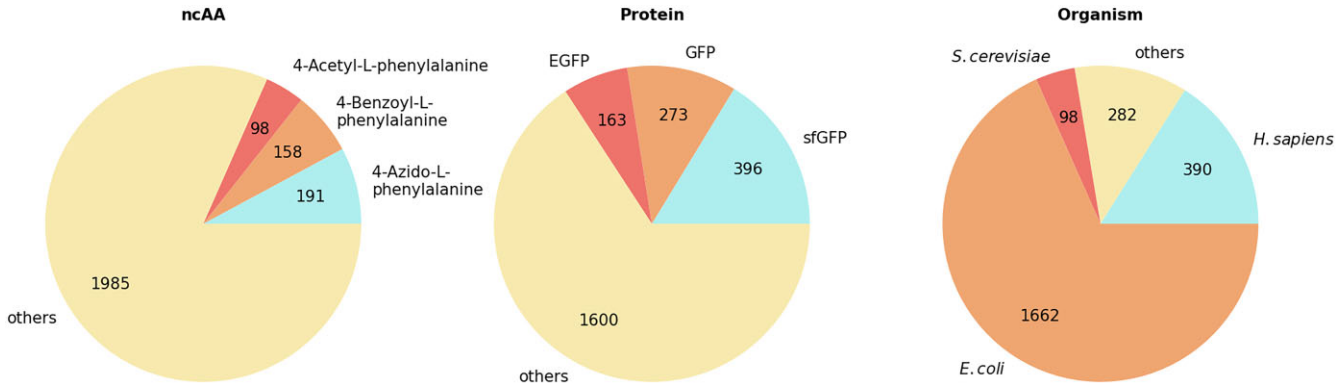


Figure 3. Pie charts showing the number of entries in the database for the indicated entities. For instance, the number 396 for sfGFP indicates that this protein has been modified 396 times. Two entries might differ only for the organism used for the experiment or for the aminoacyl tRNA synthetase employed. Similarly, the number 191 for 4-azido-l-phenylalanine does not mean this ncAA has been incorporated in 191 different target proteins. Entries are distinct if any of the other categories in the database are different, such as tRNA, protein, organism in which the incorporation was tested, etc.

for instance to find out how often a certain ncAA has been incorporated into proteins using a certain aaRS. It is also possible to hide certain columns using the 'Show/Hide columns' function. We have dedicated a page on the website to exemplary searches, which we hope will help the users navigate the database more confidently.

Example of data extraction from iNClusive

iNClusive allows users to readily get insights into different aspects of ncAA incorporation. As an example, we plotted pie charts showing the most used ncAAs (Figure 3, left panel), the most modified proteins (Figure 3, middle panel), and the most frequently adopted organisms (in this category we included also cell-free systems; Figure 3, right panel). We found that 4-Benzoyl-L-phenylalanine and 4-Azido-L-phenylalanine are the most used ncAAs. Perhaps not surprisingly, the superfolder green fluorescent protein (sfGFP) from *Aequorea victoria* is the most predominantly targeted protein for ncAA incorporation, given the possibility it offers to use fluorescence as read out of successful full protein translation. *Escherichia coli* is the most used organism for the incorporation.

Discussion

iNClusive is the first database entirely dedicated to ncAAs and the proteins that have been successfully modified with them. It offers users the possibility to access valuable information all at once, sparing them the need to search for it in various individual publications and/or on different web servers/repositories. Despite being dedicated to ncAAs, iNClusive is not ncAA-centered, meaning that the entries are not based on individual ncAAs. This might be confusing or questionable to some. We wanted to report about the different aspects of the research projects, such as organism and synthetase used for in the incorporation, target protein and positions, etc. While having entries corresponding strictly to ncAAs would simplify the database, important information would be lost or not searchable anymore. By selecting the columns to be displayed, users should be able to navigate the information and eventually find what they are looking for, even if it might take a bit more time (including own calculations beyond those automatically given by us) for certain ncAA-centered searches.

Importantly, iNClusive provides the sequences of the aaRSs and tRNAs needed for the incorporation, allowing readily performing the experiments. Having to invest time to retrieve the sequences (especially of mutated aaRSs, for which the mutations are described in various publications) might discourage researchers from actually using the technique in the laboratory.

Although we strived to report correct information in all the categories and in each entry of this database –reflected in our workflow including a first check on all entries and a second one on 10% of the entries–, some mistakes are bound to be present. This is mostly due to the vast amount of data extracted from different sources (often several papers and then public repositories) by different researchers at different career stages and with different backgrounds. We welcome the users to contact us with requests for amendments whenever inaccuracies or inconsistencies are found (email to: inclusive@bio.uni-freiburg.de).

Unfortunately, some entries in the database are incomplete (marked as 'not available') because the data were not provided in the publications, and it was not possible for us to unambiguously assign the missing information using other sources (e.g. GenBank). Moreover, if a typo was present in the original publication, the name of the ncAA might be inaccurately reported in iNClusive. Finally, publications in which mass spectrometry was wrongly indicated as mass spectroscopy have not been considered.

Currently, iNClusive contains only data retrieved from publications in which MS was used to prove the successful incorporation of the ncAA into the protein of interest and which relied on the use of an aaRS-tRNA pair. In the future, we will try to extend it to also include publications in which NMR was used. Moreover, we encourage users to suggest to us new entries for the database, which may either be from publications we inadvertently missed or future publications (email to: inclusive@bio.uni-freiburg.de). These will be incorporated in iNClusive provided they pass a quality check and adhere to our selection criteria (that is, they refer to works in which a tRNA synthetase was used and MS was applied to verify incorporation). Regardless of when such extension will be made, we aim to keep on updating the database every six months with new publications following the same criteria used so far. We hope iNClusive will prove a useful resource to the commu-

nity and will stimulate the adoption of ncAA incorporation by more researchers.

Data availability

The iNClusive database will be continuously hosted at: https://non-canonical-aas.biologie.uni-freiburg.de/

Acknowledgements

We thank Eyal Arbely for feedback and the anonymous reviewers for an extremely careful assessment of the database and their valuable suggestions. This project was conceived as part of a larger project presented at the internationally Genetically Engineered Machine (iGEM) competition in 2022. Therefore, we would like to thank all the sponsors and the collaborators who supported us.

Author contribution: L.-S.I.: creation of the initial database prototype; definition of criteria for the actual database; data collection and curation; writing of processing scripts for specific entries; project organization; writing of first manuscript draft. A.M.R.: data collection and curation; project organisation and implementation; contributing to the writing of the first draft. J.W., A.F., F.S., M.S., N.E.: main data collection and curation. B.M.K., F.K., M.L, G.F., D.A., D.M.: data collection. M.S.: preparation of graphical abstract and figures. F.K.: creation of website. K.V.: website maintenance. S.F., J.G., J.R., P.G., P.K., P.S., N.G.: critical input throughout the project. B.D.V., M.A.O.: supervision of the project. B.D.V.: writing of the final manuscript.

Funding

Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy through EXC2189 (CIBSS–Centre for Integrative Biological Signalling Studies, Project ID390939984); M.A.Ö. and B.D.V. received additionally funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program [ERC Co grant to B.D.V.; Grant agreement No. 101002044]. Funding for open access charge: EXC2189 (CIBSS–Centre for Integrative Biological Signalling Studies, Project ID 390939984).

Conflict of interest statement

None declared.

References

1. Hao,B., Gong,W., Ferguson,T.K., James,C.M., Krzycki,J.A. and Chan,M.K. (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science*, **296**, 1462–1466.
2. Böck,A., Forchhammer,K., Heider,J., Leinfelder,W., Sawers,G., Veprek,B. and Zinoni,F. (1991) Selenocysteine: the 21st amino acid. *Mol. Microbiol.*, **5**, 515–520.
3. Mody,I. (2009) GABAA receptor synaptic functions. In: *Encyclopedia of Neuroscience*. Elsevier, pp. 441–445.
4. Gupta,R., Gupta,N. and Bindal,S. (2021) Bacterial cell wall biosynthesis and inhibitors. In *Fundamentals of Bacterial Physiology and Metabolism*. Springer Singapore, Singapore, pp. 81–98.
5. Morris,S.M. (1992) Regulation of enzymes of urea and arginine synthesis. *Annu. Rev. Nutr.*, **12**, 81–101.
6. Payne,J.A.E., Schoppet,M., Hansen,M.H. and Cryle,M.J. (2017) Diversity of nature's assembly lines – recent discoveries in non-ribosomal peptide synthesis. *Mol. Biosyst.*, **13**, 9–22.
7. Kauer,J.C., Erickson-Viitanen,S., Wolfe,H.R. and DeGrado,W.F. (1986) p-benzoyl-L-phenylalanine, a new photoreactive amino acid. Photolabeling of calmodulin with a synthetic calmodulin-binding peptide. *J. Biol. Chem.*, **261**, 10695–10700.
8. Cohen,G.N. and Munier,R. (1956) [Incorporation of structural analogues of amino acids in bacterial proteins]. *Biochim. Biophys. Acta*, **21**, 592–593.
9. Johnson,J.A., Lu,Y.Y., Van Deventer,J.A. and Tirrell,D.A. (2010) Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Chem. Biol.*, **14**, 774–780.
10. Chapeville,F., Lipmann,F., Ehrenstein,G.V., Weisblum,B., Ray,W.J. and Benzer,S. (1962) On the role of soluble ribonucleic acid in coding for amino acids. *Proc. Natl. Acad. Sci. U.S.A.*, **48**, 1086–1092.
11. Noren,C.J., Anthony-Cahill,S.J., Griffith,M.C. and Schultz,P.G. (1989) A general method for site-specific incorporation of unnatural amino acids into proteins. *Science*, **244**, 182–188.
12. Wang,L., Brock,A., Herberich,B. and Schultz,P.G. (2001) Expanding the genetic code of *Escherichia coli*. *Science*, **292**, 498–500.
13. Manandhar,M., Chun,E. and Romesberg,F.E. (2021) Genetic code expansion: inception, development, commercialization. *J. Am. Chem. Soc.*, **143**, 4859–4878.
14. Young,D.D. and Schultz,P.G. (2018) Playing with the molecules of life. *ACS Chem. Biol.*, **13**, 854–870.
15. Chin,J.W. (2017) Expanding and reprogramming the genetic code. *Nature*, **550**, 53–60.
16. Wang,L. and Schultz,P.G. (2001) A general approach for the generation of orthogonal tRNAs. *Chem. Biol.*, **8**, 883–890.
17. Chin,J.W., Cropp,T.A., Anderson,J.C., Mukherji,M., Zhang,Z. and Schultz,P.G. (2003) An expanded eukaryotic genetic code. *Science*, **301**, 964–967.
18. Mukai,T., Kobayashi,T., Hino,N., Yanagisawa,T., Sakamoto,K. and Yokoyama,S. (2008) Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.*, **371**, 818–822.
19. Reinkemeier,C.D., Girona,G.E. and Lemke,E.A. (2019) Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. *Science*, **363**, eaaw2644.
20. Chen,Y., He,X., Ma,B., Liu,K., Gao,T., Niu,W. and Guo,J. (2022) Noncanonical amino acid mutagenesis in response to recoding signal-enhanced quadruplet codons. *Nucleic Acids Res.*, **50**, e94.
21. Lee,K.H., Hamashima,K., Kimoto,M. and Hirao,I. (2018) Genetic alphabet expansion biotechnology by creating unnatural base pairs. *Curr. Opin. Biotechnol.*, **51**, 8–15.
22. Brabham,R. and Fascione,M.A. (2017) Pyrrolysine amber stop-codon suppression: development and applications. *ChemBioChem*, **18**, 1973–1983.
23. Sharp,P.M., Cowe,E., Higgins,D.G., Shields,D.C., Wolfe,K.H. and Wright,F. (1988) Codon usage patterns in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res.*, **16**, 8207–8211.
24. Liu,Y. (2020) A code within the genetic code: codon usage regulates co-translational protein folding. *Cell Commun. Signal.*, **18**, 145.
25. Robertson,W.E., Funke,L.F.H., De La Torre,D., Fredens,J., Elliott,T.S., Spinck,M., Christova,Y., Cervettini,D., Böge,F.L., Liu,K.C., *et al.* (2021) Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science*, **372**, 1057–1062.
26. Mukai,T., Yamaguchi,A., Ohtake,K., Takahashi,M., Hayashi,A., Iraha,F., Kira,S., Yanagisawa,T., Yokoyama,S., Hoshi,H., *et al.*

(2015) Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. *Nucleic Acids Res.*, **43**, 8111–8122.
27. Lajoie,M.J., Rovner,A.J., Goodman,D.B., Aerni,H.-R., Haimovich,A.D., Kuznetsov,G., Mercer,J.A., Wang,H.H., Carr,P.A., Mosberg,J.A., *et al.* (2013) Genomically recoded organisms expand biological functions. *Science*, **342**, 357–360.
28. Krahn,N., Tharp,J.M., Crnković,A. and Söll,D. (2020) Engineering aminoacyl-tRNA synthetases for use in synthetic biology. In: *The Enzymes*. Elsevier, Vol. **48**, pp. 351–395.
29. Davis,L. and Chin,J.W. (2012) Designer proteins: applications of genetic code expansion in cell biology. *Nat. Rev. Mol. Cell Biol.*, **13**, 168–182.
30. Baskin,J.M. and Bertozzi,C.R. (2007) Bioorthogonal click chemistry: covalent labeling in living systems. *QSAR Comb. Sci.*, **26**, 1211–1219.
31. Harzing,A.W. (2007) Publish or perish software, https://harzing.com/resources/publish-or-perish.
32. Filippov,I.V. and Nicklaus,M.C. (2009) Optical structure recognition software to recover chemical information: OSRA, an open source solution. *J. Chem. Inf. Model.*, **49**, 740–743.
33. Kim,S., Chen,J., Cheng,T., Gindulyte,A., He,J., He,S., Li,Q., Shoemaker,B.A., Thiessen,P.A., Yu,B., *et al.* (2019) PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.*, **47**, D1102–D1109.
34. Benson,D.A., Cavanaugh,M., Clark,K., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2012) GenBank. *Nucleic Acids Res.*, **41**, D36–D42.

Received: August 14, 2023. **Revised:** October 27, 2023. **Editorial Decision:** October 28, 2023. **Accepted:** October 30, 2023