UC Riverside
UC Riverside Previously Published Works

Title
Establishing the fundamental rules for genetic code expansion.

Permalink
https://escholarship.org/uc/item/48x4730k

Journal
Nature Chemistry, 15(7)

Authors
Sinha, Souvik
Ahsan, Mohd
Palermo, Giulia

Publication Date
2023-07-01

DOI
10.1038/s41557-023-01243-9

Peer reviewed

Establishing the fundamental rules for genetic code expansion

Souvik Sinha,

Department of Bioengineering, The University of California, Riverside, California 92521, United States

Mohd Ahsan,

Department of Bioengineering, The University of California, Riverside, California 92521, United States

Giulia Palermo

Department of Bioengineering and Department of Chemistry, The University of California, Riverside, California 92521, United States

Abstract

Genetic code expansion beyond α-amino acids is a major challenge, in which stitching non-natural building blocks within the ribosome is a critical barrier. Now, the molecular determinants for the efficient incorporation of non-natural amino acids within the ribosome have been unlocked, accelerating ribosomal synthesis.

The foundation of life depends on the coherent functioning of thousands of proteins, composed of a conserved set of twenty L-α-amino acids. The ribosomal machinery weaves around 64 triplet codons, which are decoded to amino acids in an instruction-directed manner to create proteins. Over the last 20 years, numerous efforts successfully expanded the genetic code to incorporate non-canonical amino acids into proteins, customizing diverse functionalities with tremendous impact on molecular biology and therapeutics.[1] The reactivity of non-native substrates at the peptidyl transferase centre (PTC) of the ribosome, which catalyses the peptide bond formation, varies significantly and affects the reaction yield. Hence, it is of the utmost importance to know the structural features that discriminate between reactive and non-reactive substrates at the PTC.

Now in *Nature Chemistry*, Abramyan and co-workers have implemented a structure-based computational workflow to identify promising non-L-α-amino acid substrates.[2] This approach identifies the conformational requirements for the catalysis at the PTC level, which addresses the main structural barriers that hamper the ribosome-promoted biosynthesis of diverse hetero-oligomers, and offers a tool for prioritizing amino acid substrates for *in vivo* and *in vitro* applications.

giulia.palermo@ucr.edu, @palermo_lab.

Peptide bond formation at the PTC occurs through a nucleophilic α-amino group (N_α) of one substrate at the A-site aminoacyl-tRNA, attacking the sp^2-hybridized carbonyl carbon (C_{sp2}) of the growing peptide chain at the P-site peptidyl-tRNA (Fig. 1a). The ribosome enhances the rate of peptide bond formation by lowering the activation entropy,[3] mainly by positioning the substrates to achieve a 'near-attack conformation' prone to nucleophilic attack. This motivated Abramyan and co-workers to investigate the geometric relation between the A- and P-site substrates and establish the structural features that are foundational for reactivity. To characterize these details, the researchers implemented a computational workflow that builds on high-resolution structures and uses enhanced molecular simulation methods to differentiate reactive and non-reactive substrates.

Despite the non-stop development of parallel algorithms and improved computational power, molecular simulations of the ribosome remain a grand challenge.[4] This is due to the large size of the system, the timescales associated with substrate binding, and the complexity of protein/RNA interactions. The researchers have overcome these challenges through a reduced ribosome model, holding a high-resolution description of the PTC (Fig. 1b). As previously resolved ribosome structures displayed poor density at the level of the catalytic centre,[5-7] cryo-electron microscopy (cryo-EM) was used to solve the *E. coli* ribosome at 2.1 Å resolution with well-defined α-amino acid monomers at the P- and A-sites.

This model was used for metadynamics simulations to define the conformational free energy landscape of multiple non-L-α-amino acid substrates, characterized by structural and stereochemical diversity, within the PTC. Metadynamics is an enhanced sampling method that enables the dynamic study of the free energy surface along a set of pre-defined degrees of freedom (i.e., collective variables), by applying an external and history-dependent bias potential.[8] By applying this approach, the conformational space of substrates at the catalytic centre were efficiently sampled, which was not possible via classical MD, and provided an accurate description of the structural and energetic requirements for catalysis.

The obtained free energy surfaces clearly differentiate non-reactive and reactive substrates (Fig. 2). Specifically, the simulations reveal that reactive substrates occupy a conformational space characterized by an A-site nucleophile to P-site carbonyl distance (N_α—C_{sp2} distance) < 4 Å and a Bürgi-Dunitz angle[9] for nucleophilic attack (α_{BD}) — the angle with optimal overlap between the HOMO and LUMO — of 76-115°. On the other hand, substrates whose free energy minima lie outside a region where the N_α—C_{sp2} distance is less than 4 Å, do not react despite acceptable α_{BD} for the nucleophilic attack. Several structurally related and unrelated substrates were evaluated, revealing the robustness of metadynamics simulations in exploring the configurational landscape in an exhaustive manner. This free-energy approach thereby successfully discriminates between reactive and non-reactive substrates even without a quantum mechanical description of the PTC. Considering the high computational cost of quantum mechanical simulations, Abramyan and co-workers offer a rapid, yet efficient computational approach that differentiates substrates for genetic code expansions in a manner that is coherent with their experimental reactivity.

In summary, building on high-resolution cryo-EM structures, a self-consistent picture of accessible geometries for the efficient incorporation of non-L-α-amino acid substrates in the

E. coli ribosome is now understood. The computational protocol implemented by Abramyan and co-workers addresses critical barriers to the ribosome-promoted biosynthesis of diverse hetero-oligomers, paving the way to accelerate their *in vivo* and *in vitro* ribosomal synthesis. For applications *in vivo*, this approach can help prioritize substrates that require orthogonal aminoacyl tRNA synthetase variants. For applications *in vitro*, it can identify substrates that are more likely to react within the catalytic core of wild-type ribosomes and those for which engineered ribosomes are needed. Taken together, this is a tremendous contribution toward genetic code expansion beyond L-α-amino acids, which holds a major impact in molecular biology and therapeutics.

References

1. Chin JW Nature 550, 53–60 (2017). [PubMed: 28980641]

2. Watson ZL et al. Nat. Chem (2023).

3. Sievers A, Beringer M, Rodnina MV & Wolfenden R Proc. Natl. Acad. Sci 101, 7897–7901 (2004). [PubMed: 15141076]

4. Sinha S, Pindi C, Ahsan M, Arantes PR & Palermo GJ Chem. Theory Comput 19, 1945–1964 (2023).

5. Polikanov YS, Steitz TA & Innis CA Nat. Struct. Mol. Biol 21, 787–793 (2014). [PubMed: 25132179]

6. Pichkur EB et al. RNA 26, 715–723 (2020). [PubMed: 32144191]

7. Watson ZL et al. Elife 9, (2020).

8. Laio A & Parrinello M Proc. Natl. Acad. Sci. U. S. A 99, 12562–6 (2002). [PubMed: 12271136]

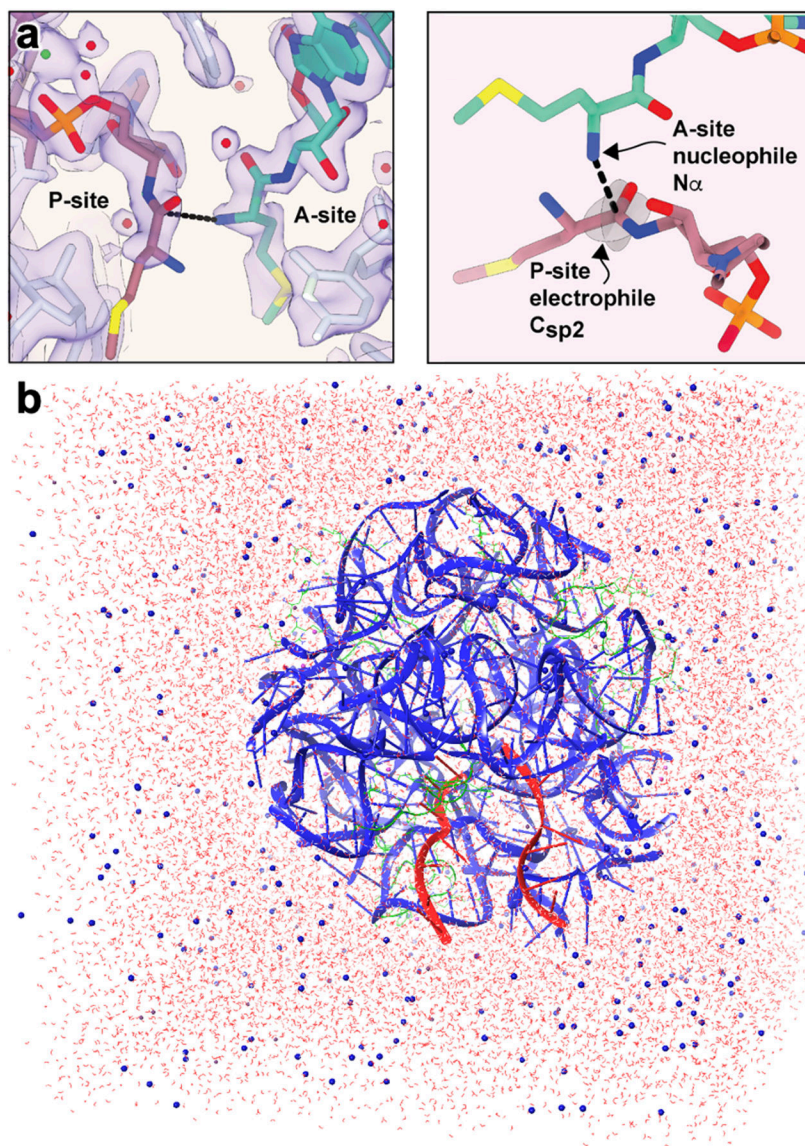9. Burgi HB, Dunitz JD, Lehn JM & Wipff G Tetrahedron 30, 1563–1572 (1974).

Fig. 1. Structure-based computational approach to identify reactive non-L-α-amino acid substrates of the ribosome.

a. Structural requirements for peptide bond formation. The A-site α-amino group (N_α) of the aminoacyl-tRNA is in line for the nucleophilic attack to the carbonyl carbon (C_{sp2}) of the P-site peptidyl RNA. **b.** Reduced model of the *E. Coli* ribosome (PDB: 8EMM)[2] used for molecular simulations. Protein residues are shown as green lines, the RNAs of the 50S subunit (blue) and the tRNAs (red) are shown as ribbons. Water molecules (red lines) and ions (blue spheres) are also shown. Figure adapted with permission from ref.[2], Springer Nature Ltd.
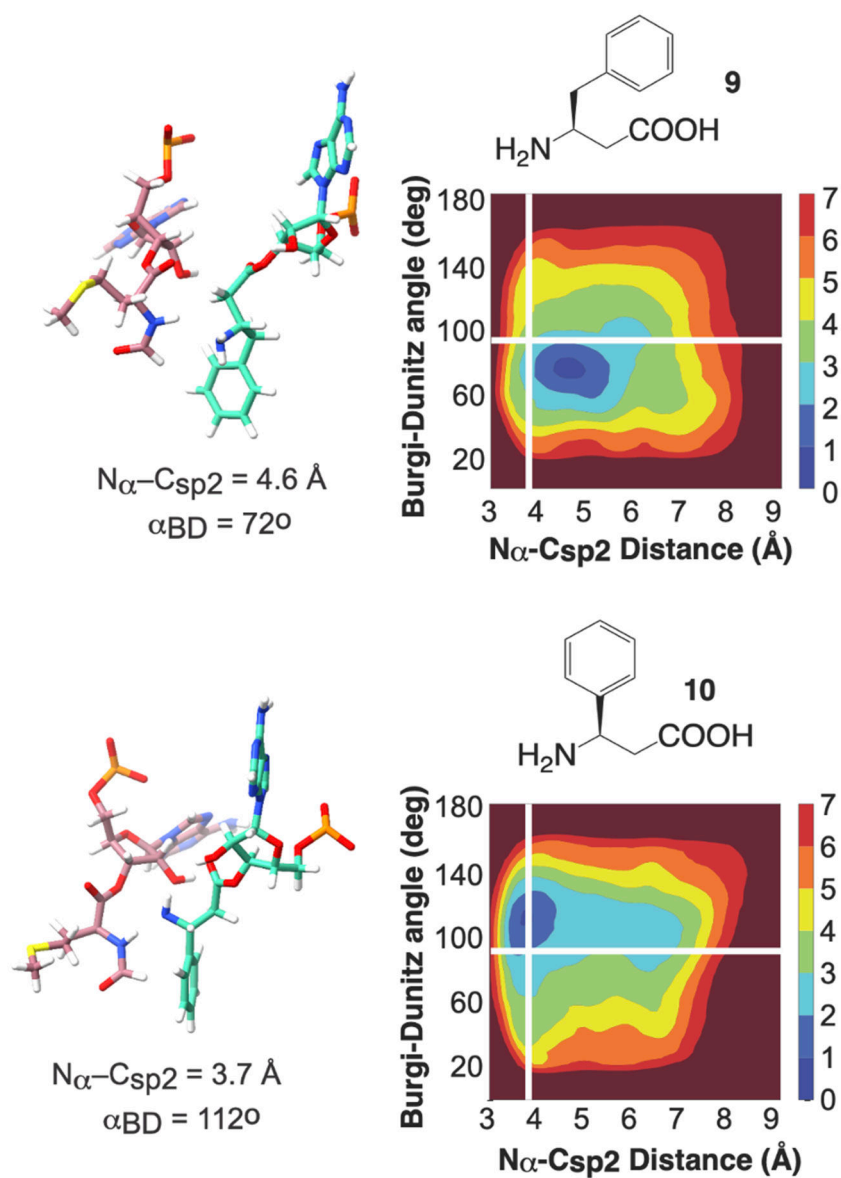
Fig. 2. Free energy surfaces differentiate non-reactive and reactive substrates.
Non-reactive substrates (top, compound **9**) find a minimum at an N_α—C_{sp2} distance greater than 4.6 Å and a low Bürgi-Dunitz angle α_{BD} of ~72°. For reactive substrates (bottom, compound **10**), the N_α—C_{sp2} distance is at or below 4 Å, and α_{BD} oscillates around 111°. Figure adapted with permission from ref.[2], Springer Nature Ltd.