Journal of Molecular Catalysis B: Enzymatic 78 (2012) 57–64

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Catalysis B: Enzymatic

journal homepage: www.elsevier.com/locate/molcatb

Reassignment of sense codons: Designing and docking of proline analogs for *Escherichia coli* prolyl-tRNA synthetase to expand the genetic code

Nadarajan Saravanan Prabhu[a,1], Niraikulam Ayyadurai[a,1], Kanagavel Deepankumar[a], Taeowan Chung[a], Dong Joon Lim[b], Hyungdon Yun[a,*]

[a] School of Biotechnology, Yeungnam University, Gyeongsan, South Korea
[b] School of Chemical Engineering, Yeungnam University, Gyeongsan, South Korea

ARTICLE INFO

Article history:
Received 31 January 2011
Received in revised form
13 December 2011
Accepted 13 February 2012
Available online 22 February 2012

Keywords:
Non-canonical amino acid
Proline
Prolyl-tRNA synthetase
Homology modeling
Virtual screening
Molecular docking

ABSTRACT

Amino acyl-tRNA synthetases (AARSs) play a vital role in protein synthesis by catalyzing the aminoacylation of tRNA with its cognate amino acid. More recently, the endogenous AARS has been reported to recognize the close structural analogs of its cognate amino acid and helps in the *in vitro* and *in vivo* incorporation of analogs into recombinant proteins. By exploiting this substrate promiscuity, a number of non-canonical amino acids were successfully incorporated into the recombinant proteins. However, the incorporation efficiency varies with the different structural analogs depending on their reactivity towards the tRNA synthetases, which is due to the interaction and accommodation in the active site. Here, to analyze the incorporation efficiency of different proline analogs and to predict the active site residues responsible for the recognition, we carried out molecular docking study with the modeled *Escherichia coli* prolyl-tRNA synthetase (*Ec*ProRS). We also mapped the binding mode for the reported, virtually generated proline analogs and compared it with the reported crystal structure. The reactivity of the reported analogs was correlated with the biochemical data with respect to their interaction and orientation in the active site, which demonstrates the role of active site residues for the recognition of proline analogs and some new substrates such as chloro, bromo and iodoproline for *Ec*ProRS. We also rationally designed a *Ec*ProRS mutant for desired proline analog and validated by docking simulation with 3D model structure.

1. Introduction

Translation machinery of cell requires 20 canonical amino acids to synthesize the cellular proteins. So far, the protein engineers have developed novel characteristic proteins by replacing the key amino acids of wild-type protein with the canonical amino acids. In recent years, the incorporation of non-canonical amino acid (NCAA) at the translational level with the help of codon reassignment opened a new door for the protein engineers and provided enormous amino acid analogs for translation, which helped in the expansion the proteome library of cell [1–4]. The incorporation of NCAA was successfully achieved by two different methods, namely residue-specific incorporation (sense codon reassignment) and site-specific incorporation (nonsense codon reassignment). In the former method, NCAAs were incorporated into protein by exploiting the substrate promiscuity of the endogenous amino acyl-tRNA synthetases (AARSs) of an auxotrophic cell in response to sense codon [5]. In the latter approach, NCAA incorporation was achieved with the help of evolved orthogonal tRNA synthetases, which are cognate for the blank codons (stop codon, four-base codon) and NCAA [3,6]. In both the approaches, incorporation efficiency of NCAA depends on the substrate specificity of the tRNA synthetases that in turn affect the protein expression. Moreover, these approaches provide the proteins with a wide variety of applications such that the substitution of canonical amino acid with NCAAs containing a highly electronegative atom (fluorine) or heavy atoms (selenium) helps in isomorphous replacement method during structural analysis of protein molecules. It also improves the protein function and enhances the characteristic features of protein such as thermal stability, structural stability, and spectral properties [2].

Proline is an important amino acid that plays a pivotal role in maintaining the structural stability of proteins [7]. Currently, more than 10 proline analogs were incorporated successfully into recombinant protein by exploiting substrate promiscuity of the *Escherichia coli* prolyl-RNA synthetase (*Ec*ProRS) through *in vivo* and *in vitro* approaches [8,9]. In these approaches, some proline

* Corresponding author at: School of Biotechnology, Yeungnam University, Gyeongsan, Gyeongbuk 712-749, South Korea. Tel.: +82 53 810 3044; fax: +82 53 810 4619.
E-mail address: hyungdon@ynu.ac.kr (H. Yun).
[1] These authors have equally contributed.

1381-1177/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.molcatb.2012.02.004

analogs were efficiently incorporated, and some were not incorporated successfully. For example, the *in vivo* incorporation of proline analogs, such as hydroxyproline, fluoroproline, thiaproline, showed a high yield of protein, whereas selenaproline and 3,4-phenylproline were not able to incorporate into the recombinant proteins [8,10–12]. This shows the discrimination of *Ec*ProRS towards the close structural analogs of proline during incorporation. Predicting the binding mode of the analogs with the *Ec*ProRS will provide the discrimination between substrate specificity of proline analogs [13,14]. By studying the structure of the active site of AARS, the substrate specificity of *Ec*ProRS can be expanded by generating a possible *Ec*ProRS mutant (orthogonal *Ec*ProRS) to introduce any analog into the protein structure in response to nonsense codon [6,15]. Among the proline analogs, the *Ec*ProRS show low incorporation rate for some analogs such as hydroxyproline, which is efficiently incorporated through *in vitro* translation but not through *in vivo* incorporation. This is because these analogs were not permeable for bacterial cell membrane during incorporation, whereas the permeability of this analog was improved by the addition of NaCl [10,16]. Thus, the incorporation of NCAA depends on the bioavailability and reactivity towards the AARS.

Molecular docking study is one way to predict the interaction between the substrate and active site of protein, which is extensively used for predicting inhibitors and substrates for a target enzyme with the help of available 3D structures [13,14,17]. In this study, we virtually designed the proline analogs with highly important functional groups and performed molecular docking analysis of these designed analogs and reported proline analogs with *Ec*ProRS using AutoDock4.2 [18]. We investigated the binding mode and the molecular recognition mechanism of the proline analogs by *Ec*ProRS that favoring the *in vivo* and *in vitro* incorporation of the proline analog into recombinant protein. Using this result, we performed a comparative study on the substrate specificity of *Ec*ProRS towards analogs in terms of binding energy and binding mode. We also analyzed the important active site residues that favor the interaction of proline analogs with *Ec*ProRS and residues that sterically affect incorporation of proline analogs. Further, to improve the binding efficiency of analogs such as pipecolic acid and azetidine-2-carboxylic acid, we rationally designed the ProRS and validated the mutant by docking it with respective analogs using the docking tool.

2. Materials and methods

2.1. Homology modeling and protein preparation for docking simulation

The protein sequence for wild-type *Ec*ProRS was retrieved from the National Center for Biotechnology Information (GenBank accession number: 147362) and the potential template was identified through BLASTP search against Protein Data Bank (PDB). A comparative protein structure modeling tool, MODELLER, was used for constructing the model, which implements modeling by satisfaction of spatial restraints. The three-dimensional structure for the protein sequence was constructed using the template *Enterococcus faecalis* ProRS structure (PDB ID-2J3L; resolution, 2.30 Å) co-crystallized with the prolyl-adenylate [7]. This template shares a high sequence similarity with the *Ec*ProRS and specifically with its active site residues. Before constructing the model, sequence alignment was done using ClustalX [19] and was compared with the results obtained from align2d tool of MODELLER and manually checked. Models were constructed using the structural template and sequence alignment. The constructed model was refined with energy minimization step of DS 2.1 (200 steps of steepest descent and followed conjugate gradient). The overall structural and stereo chemical qualities of the protein were assessed using PROCHECK, and WHAT IF [20,21].

2.2. Preparation of ligands

The co-crystallized proline, misbound cysteine and alanine in the active site of *Ec*ProRS of different species were retrieved (Table S1, Supplementary information) and energy was minimized with the help of Discovery Studio 2.1 (DS 2.1; Accelyrs Inc., CA, USA) for pharmacophoric search. A set of five featured common pharmacophoric search was done at an interatomic distance of 2.0 Å using the Catalyst module of DS 2.1. The generated pharmacophoric model was validated by analyzing the interaction between the reported structural complexes (Table S1, Supplementary information). Proline analogs were virtually generated by introducing interesting functional groups at 3′ and 4′ position of proline. A three-dimensional structural database of proline, its reported and virtually designed analogs was prepared using Build Database tool of DS 2.1. Using the generated common pharmacophore, a 3D search was performed for the generated 3D database. The resulted proline analogs were energy minimized using the DS 2.1 (Ligand minimization step: smart minimizer) and then exported for docking analysis.

2.3. Molecular docking simulation

Automated docking calculation was performed for the *Ec*ProRS model using AutoDock4.2 [18]. In the ligand (compound) preparation step, each ligand was treated with Kollmann charges and autodock atom types. Gasteiger charges and autodock atom types were assigned to the constructed model protein using autodock force field, and the nonpolar hydrogens were added and merged. The grid size was set to $40 \times 40 \times 40$ points with a grid spacing of 0.375 Å centered around the XYZ coordinate position of the nitrogen atom of Arg140. Affinity maps were calculated automatically for the given grid size around the defined active site. Lamarckian genetic algorithm was employed as search parameter, and for each compound 50 GA run was carried out. For each GA run, 25,000,000 (long) evaluations were performed for a population size of 150. The operator weight for crossover, mutation, and migration were set as the default parameter. The best docked conformation was selected based on the docking score and interactions with the active site residues. The docked complexes were exported in PDB format for further analysis.

2.4. Post docking analysis

The binding mode and interactions between the compounds and protein were analyzed using Autodock, PyMOL and compared with the available prolyl-adenylate co-crystallized structure (PDB ID-2J3L). To compare the affinity of proline analogs with the protein, the binding mode of the proline analogs was completely compared with the docked complex of natural substrate, proline. The binding mode of the docked complex is then compared with the experimental binding mode, and a root-mean-square distance (RMSD) between the two is calculated using PyMOL.

3. Results and discussion

The main goal of this work is to study the substrate specificity of *Ec*ProRS toward the proline analogs and to design virtual proline analogs that can be incorporated by the *Ec*ProRS. As the crystal structure for the *Ec*ProRS is not available, we developed a 3D model with the help of structural template and performed the docking analysis.