Bioorganic & Medicinal Chemistry 24 (2016) 4177–4187

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

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

Synthesis of alanyl nucleobase amino acids and their incorporation into proteins

Poulami Talukder [a], Larisa M. Dedkova [a], Andrew D. Ellington [b], Petro Yakovchuk [a], Jaebum Lim [b], Eric V. Anslyn [b], Sidney M. Hecht [a,*]

[a] Biodesign Center for BioEnergetics and School of Molecular Sciences, Arizona State University, Tempe, AZ 85287, USA
[b] Department of Chemistry, University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Article history:
Received 25 April 2016
Revised 2 July 2016
Accepted 5 July 2016
Available online 6 July 2016

Keywords:
Nucleobases
DNA-protein interaction
N-Boc-L-serine β-lactone
DNA polymerase I

ABSTRACT

Proteins which bind to nucleic acids and regulate their structure and functions are numerous and exceptionally important. Such proteins employ a variety of strategies for recognition of the relevant structural elements in their nucleic acid substrates, some of which have been shown to involve rather subtle interactions which might have been difficult to design from first principles. In the present study, we have explored the preparation of proteins containing unnatural amino acids having nucleobase side chains. In principle, the introduction of multiple nucleobase amino acids into the nucleic acid binding domain of a protein should enable these modified proteins to interact with their nucleic acid substrates using Watson-Crick and other base pairing interactions. We describe the synthesis of five alanyl nucleobase amino acids protected in a fashion which enabled their attachment to a suppressor tRNA, and their incorporation into each of two proteins with acceptable efficiencies. The nucleobases studied included cytosine, uracil, thymine, adenine and guanine, i.e. the major nucleobase constituents of DNA and RNA. Dihydrofolate reductase was chosen as one model protein to enable direct comparison of the facility of incorporation of the nucleobase amino acids with numerous other unnatural amino acids studied previously. The Klenow fragment of DNA polymerase I was chosen as a representative DNA binding protein whose mode of action has been studied in detail.

1. Introduction

DNA-protein interaction plays an important role in the regulation of multiple aspects of DNA function including DNA replication, repair, recombination and transcription.[1] DNAs can assume a wide variety of secondary and tertiary structures, and individual DNA-binding proteins recognize specific DNA structures.[2–4] In many cases, a DNA-binding domain of protein involves a specific structured motif such as an α-helix or a β-sheet, and amino acids in the protein structure make specific contacts to the DNA backbone and nucleobases.[5] Notably, hydrogen bonding interactions between the amino acid residues in a protein and the nucleobases in a DNA can be important to increase the binding affinity and specificity of the protein for its DNA substrate.[5] Analogous interactions operate to mediate the binding of RNA binding proteins to their substrates.[6]

The appreciation that interactions between nucleobases are important for controlling nucleic acid structure and function, and that base-base recognition can be highly selective, led to the preparation and characterization of peptide nucleic acids (PNAs).[7] These oligomers contain nucleobases attached at defined positions to amide-linked backbones, and have been shown to bind surprisingly tightly and specifically to a variety of nucleic acid structures using Watson-Crick base pairing.[7] In spite of the presence of amide bonds in the PNA backbone, PNAs do not form protein-like tertiary structures. To obtain molecules capable of recognizing nucleobases in nucleic acids, but also capable of forming protein-like structures, peptides containing L-α-amino acids with a nucleobase side chain were designed by the Mihara laboratory.[8–10] They reported the use of L-amino-γ-nucleobase-butyric acids for the construction of a number of RNA binding peptides.

The specific nucleobase amino acids of interest in the present study are the "nucleo alanine" derivatives first described by the Diederichsen laboratory for the elaboration of alanyl-PNA

Abbreviations: APCI, atmospheric pressure chemical ionization; ESI, electrospray ionization; NVOC-Cl, 4,5-dimethoxy-2-nitrobenzyl chloroformate; TBA, tetra-*n*-butylammonium; DBU, 1,8-diazabicyclo 5.4.0 undec-7-ene.

* Corresponding author. Tel.: +1 480 965 6625; fax: +1 480 965 0038.
E-mail address: Sidney.Hecht@asu.edu (S.M. Hecht).

http://dx.doi.org/10.1016/j.bmc.2016.07.008
0968-0896/

chimeras.[11] As shown in Figure 1, alanyl nucleobase amino acids contain a single methylene group connecting the α-carbon atom and nucleobase, more closely analogous to the positioning of (hetero)cyclic functional groups in proteinogenic amino acids such as histidine, tryptophan, phenylalanine and tyrosine. The asymmetric syntheses of alanyl nucleobase amino acids (**A**–**E**) were achieved by nucleophilic ring opening of *N*-Boc-L-serine β-lactone. A site specific mutagenesis technique, which permits the insertion of unnatural amino acids into any predetermined position of a protein by the suppression of TAG or four base codons with misacylated-tRNAs, had not been used before for incorporation of amino acids having nucleobase side chains. Therefore, this strategy was attempted for the incorporation of alanyl nucleobase amino acids **A**–**E** into *E. coli* dihydrofolate reductase (DHFR) and also into the Klenow fragment of *E. coli* DNA polymerase I, a DNA-binding protein. In order to synthesize the requisite aminoacylated tRNAs for protein synthesis, the pdCpA derivatives of alanyl nucleobase amino acids were prepared.

2. Results

2.1. Synthesis of nucleobase amino acids and tRNA activation

The synthesis of the pdCpA derivative of amino acid **A** is outlined in Scheme 1. The synthesis of the aminoacylated pdCpA derivative of amino acid **A** (Figure 1) was accomplished starting from commercially available cytosine (Scheme 1). CBz protection of cytosine afforded compound **1** in 74% yield.[12] Compound **1** underwent coupling with *N*-Boc-L-serine β-lactone in the presence of DBU to afford N-protected cytosyl amino acid derivative **2** in 56% yield.[11]

Compound **2** was treated with $SOCl_2$ in methanol to form methyl ester **3** and then the CBz protecting group was removed by hydrogenolysis over Pd/C to afford **4**. Boc deprotection (CF_3-COOH in CH_2Cl_2) and subsequent NVOC protection of compound **4** afforded **5** in 35% overall yield for two steps.[11,13] *N*-protected methyl ester **5** was converted to the free acid by treatment with aq LiOH, and the latter was treated with chloroacetonitrile to afford the desired cyanomethyl ester **6** in 34% yield for two steps.[13,14] The key intermediate cyanomethyl ester was coupled with tris(tetrabutylammonium) salt of pdCpA[15] to give pdCpA ester **7** in 25% yield.

In order to introduce **A** as an amino acid constituent of proteins, the pdCpA derivative of this amino acid was used to activate a suppressor $tRNA_{CUA}$ (Scheme 2). Accordingly, aminoacylated dinucleotide **7** was ligated to an abbreviated $tRNA_{CUA}$-C_{OH} transcript via the agency of T4 RNA ligase in the presence of ATP to afford NVOC-aminoacyl-tRNA.[16] The activated tRNA was deprotected by UV irradiation to afford the free aminoacyl-tRNA.[13]

The synthesis of the aminoacylated pdCpA derivative of alanyl nucleobase amino acid **B** was accomplished starting from commercially available uracil (Scheme 3) which underwent coupling with *N*-Boc-L-serine β-lactone in the presence of DBU to afford compound **8** in 41% yield. Free carboxylic acid **8** was treated with $SOCl_2$ in methanol to form the methyl ester **9** in 75% yield. Boc deprotection of methyl ester **9** and subsequent NVOC protection afforded NVOC-carbamate **10** in 54% yield for two steps. Methyl ester **10** was subjected to saponification in the presence of aqueous LiOH to produce the free acid, which was activated as cyanomethyl ester **11** in 30% overall yield. Treatment of cyanomethyl ester **11** with the tris(tetrabutylammonium) salt of pdCpA[14] in anhydrous DMF afforded pdCpA ester **12** in 46% yield.

The synthesis of the aminoacylated pdCpA derivative of amino acid **C** (Figure 1) followed the same procedure as for amino acid **B**, and is outlined in Scheme 4. The synthesis of theaminoacylated pdCpA derivative of nucleobase amino acid **D** was accomplished starting from commercially available adenine (Scheme 5). *N*-Pentenoyl protection of adenine afforded compound **18**, the latter of which underwent coupling with *N*-Boc-L-serine β-lactone to afford compound **19** in 43% yield.[11,12] Boc deprotection of N^{α} and subsequent reprotection using 4-pentenoic acid succinimide ester in the presence of K_2CO_3 afforded *N*-dipentenoyl amide **20** in 41% yield for two steps.[14] Free carboxylic acid **20** was activated as cyanomethyl ester **21** in 31% yield. The pdCpA derivative of **D** (**22**) was then prepared from cyanomethyl ester **21** in 39% yield.

The synthesis of the aminoacylated pdCpA derivative of amino acid **E** (Figure 1) was accomplished starting from commercially available 2-amino-6-chloropurine which underwent coupling with *N*-Boc-L-serine β-lactone to afford compound **23** in 80% yield (Scheme 6).[17] Solvolysis of the purinyl chloride under acidic conditions and subsequent protection using 4-pentenoic acid succinimide ester in the presence of 1 N NaOH afforded *N*-protected acid **24** in

24% yield for two steps.[17] Compound **24** was activated as cyanomethyl ester **25** in 53% yield. The pdCpA derivative of **E** (**26**) was then prepared from cyanomethyl ester **25** in 70% yield.

The individual *N*-NVOC and *N*-pentenoyl protected aminoacylated pdCpA derivatives were ligated to a suppressor $tRNA_{CUA}$ lacking its 3′-terminal cytidine and adenosine residues ($tRNA_{CUA}$-C_{OH})[16] via the agency of T4 RNA ligase, as illustrated in Scheme 2 for nucleobase amino acid **A**. The success of the ligation reaction was verified by analysis on a polyacrylamide gel under acidic conditions (Figure 2). The NVOC protecting groups were then removed by exposure to high intensity UV light at 4 °C.[13] The *N*-pentenoyl protecting group was removed by treatment with aqueous iodine.[14] Removal of the protecting group was done immediately prior to the use of the misacylated tRNAs in protein synthesis.

2.2. Incorporation of nucleobase amino acids into proteins

The five aminoacyl-$tRNA_{CUA}$s obtained as described above were employed in an *in vitro* cell free transcription-translation system, which was programmed with one of two DNA plasmids. The first contained a TAG codon at the position corresponding to residue Val10 of DHFR. The incorporation of all the alanyl nucleobase amino acids into position 10 was evaluated via denaturing PAGE analysis and it was found that alanyl nucleobase amino acids (**A**–**E**) could suppress the UAG codon at position 10 of DHFR mRNA,

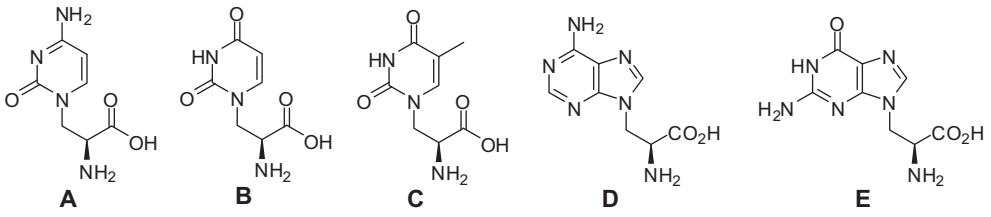

Figure 1. Series of alanyl nucleobase amino acids (**A**–**E**) synthesized for site-directed incorporation into DHFR and the Klenow fragment of *E. coli* polymerase I.