Published in final edited form as:

Bioorg Med Chem. 2016 September 15; 24(18): 4177–4187. doi:10.1016/j.bmc.2016.07.008.

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, and Sidney M. Hecht^{a,*}

^aBiodesign Center for BioEnergetics and School of Molecular Sciences, Arizona State University, Tempe, AZ 85287, USA

^bDepartment of Chemistry, University of Texas at Austin, Austin, TX 78712, USA

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 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.

Graphical abstract

Keywords

Nucleobases; DNA-protein interaction; *N*-Boc-L-serine β -lactone; DNA polymerase I

*Corresponding author. Tel: +1 480 965 6625; fax: +1 480 965 0038. Sidney.Hecht@asu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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 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 SOCl₂ 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 (CF₃COOH in CH₂Cl₂) 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 pdCpA¹⁵ to give a 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}-COH 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 SOCl₂ 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 over for 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 pdCpA¹⁴ 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 the aminoacylated 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 K₂CO₃ 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}-COH)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 codons 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, leading to full length DHFR synthesis with suppression yields ranging from 6 to 18% as compared to wild-type DHFR (Figure 3A and Table 1).

The same aminoacyl-tRNAs were further employed in an *in vitro* cell free transcription-translation system programmed with a plasmid containing the gene for the Klenow fragment of DNA polymerase I; this gene contained a TAG codon at the position corresponding to position 484 of the Klenow fragment. As shown in Figure 3B and summarized in Table 1, the suppression efficiencies, relative to the Klenow fragment synthesized from the wild-type gene, were 6 to 15%.

The suppression yields obtained in replicate experiments are summarized in Table 1. For both proteins studied, incorporation of the pyrimidine amino acids (**A**, **B** and **C**) were found to be in the range 11 – 18%, while the purine amino acids (**D** and **E**) exhibited suppression yields from 6 – 8%. For any given nucleobase amino acid, the suppression yields were quite similar for the two proteins studied. Relative suppression yields tend to be quite reproducible from one set of experiments to another, such that it should be possible to introduce amino acids **A** – **E** into many other proteins in workable yields. In comparison, absolute protein yields can vary significantly from one set of experiments to another, especially if the several constituents of the protein biosynthesizing reaction are from different preparations. In this context it may be noted that the suppression yield obtained with phenylalanyl-tRNA_{CUA} (~30%) was toward the lower end of the range that we have observed over a period of time.

The most likely cause of this is the specific S-30 preparation employed for the current experiments. If this proves to be correct, higher incorporation yields for **A** – **E** can probably be achieved with other S-30 preparations.

3. Discussion

The nucleobase amino acid derivatives **A** – **E** were all prepared by an initial nucleophilic ring opening of *N*-Boc-L-serine β -lactone with a pyrimidine or purine derivative in the presence of DBU, in analogy with the work of Roviello et al.¹¹ In the case of uracil and thymine, the unprotected nucleobases were employed, while in the cases of cytosine and adenine, simple protection of the exocyclic amino groups afforded compounds which provided the amino acid derivatives. In the case of guanine, the synthesis began with the use of 2-amino-6-chloropurine, and the initial amino acid derivative (**23**, Scheme 6) was converted to the respective guanine derivative by solvolysis under strongly acidic conditions. The activation of suppressor tRNAs with unnatural amino acids is achieved routinely using NVOC¹⁸ or pentenoyl^{19,20} protection of N ^{α} of the aminoacyl moiety, either of which can be removed under conditions which does not cause concomitant deacylation of the amino acid attached to the tRNA as an activated ester. In the present study the NVOC protecting group was employed for the three pyrimidine amino acids. However, it was found that the syntheses of the two purine amino acids could be carried out more conveniently using the *N*-pentenoyl protecting group. Further, the activation of the adenine amino acid **D** as its pdCpA ester required protection of the exocyclic N group of adenine; this was accomplished using a second pentenoyl protecting group, enabling both groups to be removed simultaneously by treatment with aqueous iodine following ligation to the abbreviated suppressor tRNA-COH. A few attempts were made to utilize tRNAs with nucleobase amino acids containing protecting groups on the nucleobase; this was done in an effort to increase the suppression yields. In no case was this strategy successful (data not shown).

As documented in Figure 3 and Table 1, all five nucleobase amino acid analogues were incorporated into the two proteins studied. Based on earlier studies from our laboratory utilizing DHFR^{16,21-25} and the Klenow fragment of DNA polymerase I,²⁶ the present suppression yields should be entirely sufficient to allow routine biochemical studies to be carried out on the modified enzymes. While the suppression yields obtained with any unnatural amino acid cannot be predicted with certainty, in general small lipophilic amino acids are incorporated with greater facility than large or polar species. In the present case, all five modified nucleobase amino acids are reasonably polar so the suppression yields observed were quite reasonable. Given the effect of amino acid side chain size on incorporation efficiency, it was to be expected that **A** – **C** were incorporated with greater efficiency than **D** or **E**.

The prospects for productive use of the alanyl nucleobase amino acids described here may be judged by the experience of the Mihara laboratory. These workers have prepared nucleobase amino acids having one more bridging methylene group in the side chains than do compounds **A** – **E**, and have incorporated them into peptides by chemical synthesis to enable RNA binding studies.^{8-10,28,29} For example, when three nucleobase amino acid residues were included within arginine-rich α -helical peptides targeted to a hairpin RNA

derived from P22 phage, the affinities varied with the nature and location of the nucleobases employed, thus illustrating the potential for structure dependent RNA targeting.¹⁰ In a second study, the synthetic peptides were used to screen T7 phage libraries. A number of nucleic acid-related proteins were identified in this fashion. Interestingly, other types of proteins were identified as well.²⁷ Additionally, introduction of the nucleobase amino acids within the dimerization domain of HIV-1 protease enhanced affinity of the individual subunits, but lowered the catalytic activity of the formed dimer to some extent.²⁸ In the aggregate, these studies suggest numerous applications which may be anticipated for proteins containing nucleobase amino acids. While the preparation of peptides containing nucleobase amino acids can be addressed using existing methods of chemical synthesis, the routine preparation of proteins by chemical synthesis represents a significantly greater challenge. In such cases, the use of ribosomal synthesis, carried out *in vitro* or *in vivo*, represents a more realistic alternative.

While the design of peptides having unique affinities for specific nucleic acid targets is difficult enough, the complexity of protein–nucleic acid interactions reflected in myriad X-ray crystal structures²⁹⁻³¹ underscores the extraordinary challenge of attempting to alter the selectivity of such interactions by any simple predictive alteration of protein structure. In comparison, the detailed information provided by such structures might provide a useful starting point for introducing nucleobase amino acids into proteins at sites potentially capable of forming new H-bonds with specific nucleosides in target DNA and RNA structures. Enhanced selectivity of gene expression and nucleic acid maturation, for example, might be more readily achieved by this approach. Clearly, numerous other applications can be envisioned, including directed protein-protein interactions and the assembly of supramolecular arrays.

4. Conclusions

Five alanyl nucleobase amino acids have been synthesized and their aminoacyl-tRNA derivatives were used for incorporation of each of the five analogues into a predetermined of two proteins by suppression of a UAG codon. The protein sites modified were position 10 of *E. coli* DHFR and 484 of the Klenow fragment of *E. coli* DNA polymerase I. Pyrimidine-based alanyl nucleobase amino acids (**A** – **C**) exhibited better suppression efficiencies than did purine-based alanyl nucleobase amino acids (**D** and **E**). These findings enable the elaboration of proteins containing nucleobase amino acids as protein constituents at predetermined positions. It is anticipated that such “nucleoproteins” whose normal function includes DNA or RNA binding can be targeted selectively to particular regions of nucleic acids via nucleobase interactions presently associated with nucleic acids.

5. Experimental section

5.1 General experimental procedures

All experiments requiring anhydrous conditions were conducted in flame-dried glassware fitted with rubber septa under a positive pressure of dry nitrogen. Reactions were performed at room temperature unless otherwise indicated. Analytical thin layer chromatography (TLC) was performed using glass plates pre-coated with silica gel (0.25 mm, 60 Å pore size,

230-400 mesh, Silicycle) impregnated with a fluorescent indicator (254 nm). TLC plates were visualized by exposure to ultraviolet light (UV). Flash column chromatography was performed employing silica gel (60 Å pore size, 40-63 µm, standard grade, Silicycle). An acetone cooling bath was adjusted to the appropriate temperature by the addition of small portions of dry ice.

¹H NMR and ¹³C NMR spectra were recorded on Varian INOVA 400 (400 MHz) or Varian INOVA 500 (500 MHz) spectrometers at 25 °C. Proton chemical shifts are expressed in parts per million (ppm, δ scale) and are referenced to residual protium in the NMR solvent (CDCl₃, DMSO-*d*₆ or CD₃OD). Splitting patterns are designated as follows: s, singlet; br s, broad singlet; d, doublet; dd, doublet of doublets; t, triplet; q, quartet; m, multiplet. High resolution mass spectra were obtained at the Arizona State University CLAS High Resolution Mass Spectrometry Facility or the Michigan State University Mass Spectrometry Facility. HPLC purification was performed with a Waters 600 pump coupled with a Varian ProStar 340 detector and a Grace Econosil C₁₈ column (250 × 10 mm, 5 µm). The tetra-*n*-butylammonium (TBA) salt of pdCpA^X was prepared using Dowex 50W×8, 200-400 mesh, activated in its TBA form.

The chemicals used for synthesis were purchased from Aldrich Chemical Co., Sigma Chemical Co. or Combi Blocks. THF was distilled under argon from sodium-benzophenone ketyl and CH₂Cl₂ was distilled under argon from calcium hydride. Ni-NTA agarose was obtained from Qiagen Inc. DNA oligonucleotides were purchased from Integrated DNA Technologies. DEAE-Sepharose, ammonium persulfate, acrylamide, *N,N'*-methylene-bis-acrylamide, acetic acid, potassium glutamate, ammonium acetate, dithiothreitol, magnesium acetate, phospho(enol)pyruvate, *Escherichia coli* tRNA, isopropyl β-D-thiogalactopyranoside (IPTG), ATP, GTP, CTP, UTP, cAMP, amino acids, rifampicin, and formamide were obtained from Sigma-Aldrich. Tris and SDS were purchased from Bio-Rad Laboratories (Hercules, CA). [³⁵S]-methionine (1000 Ci/mmol, 10 µCi/µL) was purchased from PerkinElmer Inc. Protease inhibitor (complete, EDTA-free) was obtained from Boehringer Mannheim Corp. T4 RNA ligase and T4 polynucleotide kinase were purchased from New England Biolabs Inc.

Phosphorimager analysis was performed using an Amersham Biosciences Storm 820 equipped with ImageQuant version 5.2 software from Molecular Dynamics.

5.2. Synthesis of alanyl nucleobase aminoacyl-pdCpA derivatives

5.2.1. Methyl (S)-3-(4-(Benzyloxycarbonylamino)-2-oxopyrimidin-1(2H)-yl)-2-(tert butoxycarbonylamino)propionate (3)—To a stirred suspension of 1.44 g (36.0 mmol) of NaH (60% in mineral oil) in 50 ml of DMF at 0 °C was added 1.00 g (9.01 mmol) of cytosine. The reaction mixture was stirred at 0 °C for 1 h, then 1.35 mL (9.50 mmol) of benzyl chloroformate was added. The reaction mixture was stirred at room temperature for 14 h, then diluted with 100 mL of H₂O and ice. After neutralization with 5 N HCl, a colorless precipitate formed, and was filtered, washed with five 50-mL portions of H₂O and dried to give **1** as a colorless powder: yield 1.71 g (74%); mass spectrum (APCI), *m/z* 246.0890 (M+H)⁺ (C₁₂H₁₂N₃O₃ requires *m/z* 246.0879). To a stirred suspension containing 250 mg (1.02 mmol) of compound **1** in 5 mL of DMSO was added 0.18 mL (183 mg, 1.20

mmol) of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU). Within 15 min, a solution containing 220 mg (1.20 mmol) of *N*-(*tert*-butoxycarbonyl)-L-serine β -lactone¹¹ in 5 mL of DMSO was added. The reaction mixture was stirred at room temperature for 2 h under an argon atmosphere, then diluted with 10 mL of 0.5 N HCl and extracted with two 20-mL portions of EtOAc. The combined organic phase was dried (MgSO₄) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10 $\times$ 2 cm). Elution with 5:1 ethyl acetate–methanol gave **2** as a colorless solid: yield 246 mg (56%); silica gel TLC *R*_f 0.31 (5:1 ethyl acetate/methanol). To a cooled (0 °C) solution containing 246 mg (0.60 mmol) of **2** in 5 mL of anhydrous MeOH was added dropwise 0.04 mL (67.7 mg, 0.60 mmol) of SOCl₂. The reaction mixture was allowed to warm slowly to room temperature with stirring for 2 h, then diluted with 50 mL of brine and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO₄) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10 $\times$ 2 cm). Elution with 10:1 ethyl acetate/methanol gave **3** as a colorless solid: yield 165 mg (65%); silica gel TLC *R*_f 0.72 (5:1 ethyl acetate/methanol); ¹H NMR (CDCl₃) δ 1.35 (s, 9H), 3.64 (s, 3H), 4.01–4.60 (m, 3H), 5.19 (s, 2H), 6.29 (br s, 1H), 7.17 (s, 1H), 7.31–7.37 (m, 4H), 7.55 (s, 1H), 7.75 (s, 1H) and 8.70 (br s, 1H); ¹³C NMR (CDCl₃) δ 28.2, 50.8, 52.1, 52.8, 67.9, 80.5, 94.8, 128.3, 128.58, 128.63, 135.1, 149.7, 150.5, 152.5, 155.5, 162.9 and 170.4; mass spectrum (APCI), *m/z* 447.1890 (M+H)⁺ (C₂₁H₂₇N₄O₇ requires *m/z* 447.1880).

5.2.2. Methyl (S)-3-(4-Amino-2-oxopyrimidin-1(2H)-yl)-2-(tert-butoxycarbonylamino)propionate (4)—

To a solution containing 164 mg (0.36 mmol) of **3** in 5 mL of MeOH was added a catalytic amount of 10% Pd/C and the reaction was maintained under 1 atm of H₂ (g) overnight. The catalyst was removed by filtration through a pad of Celite and the filtrate was concentrated under diminished pressure. The crude product was purified by chromatography on a silica gel column (10 $\times$ 2 cm). Elution with 3:1 ethyl acetate–methanol gave **4** as a colorless solid: yield 98.0 mg (85%); silica gel TLC *R*_f 0.11 (5:1 ethyl acetate/methanol); ¹H NMR (CD₃OD) δ 1.37 (s, 9H), 3.71–3.73 (m, 4H), 4.34–4.59 (m, 2H), 5.83 (d, 1H, *J* = 7.2 Hz) and 7.44 (d, 1H, *J* = 7.2 Hz); ¹³C NMR (CD₃OD) δ 28.6, 53.2, 53.6, 80.9, 95.8, 147.9, 148.2, 157.6, 158.7, 167.9 and 172.0; mass spectrum (APCI), *m/z* 313.1506 (M+H)⁺ (C₁₃H₂₁N₄O₅ requires *m/z* 313.1512).

5.2.3. Methyl (S)-3-(4-Amino-2-oxopyrimidin-1(2H)-yl)-2-((4,5-dimethoxy-2-nitrobenzyloxy)carbonylamino)propionate (5)—

To a stirred solution containing 97.0 mg (0.31 mmol) of **4** in 2 mL of anhydrous CH₂Cl₂ was added 0.24 mL (353 mg, 3.10 mmol) of trifluoroacetic acid. The reaction mixture was stirred at room temperature for 2 h and then concentrated under diminished pressure. To a stirred solution containing the Boc-deprotected product in 2 mL of 1:1 dioxane/water was added 129 mg (0.93 mmol) of K₂CO₃ followed by 85.5 mg (0.31 mmol) of NVOC-Cl. The reaction mixture was stirred at room temperature for 14 h under argon, then diluted with 50 mL of brine and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO₄) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10 $\times$ 2 cm). Elution with 5:1 ethyl acetate/methanol gave **5** as a light yellow solid: yield 49.2 mg (35% for two steps); silica gel TLC *R*_f 0.25 (5:1 ethyl acetate/

methanol); ^1H NMR (CD_3OD) δ 3.71 (s, 3H), 3.84 (s, 3H), 3.90 (s, 3H), 4.35-4.39 (m, 1H), 4.77-4.81 (m, 2H), 5.31 (ABq, 2H, $J = 10.8$ Hz), 5.83 (d, 1H, $J = 7.2$ Hz), 7.04 (s, 1H), 7.48 (d, 2H, $J = 7.2$ Hz) and 7.58 (s, 1H); mass spectrum (APCI), m/z 452.1413 ($\text{M}+\text{H}^+$) ($\text{C}_{18}\text{H}_{22}\text{N}_5\text{O}_9$ requires m/z 452.1417).

5.2.4. Cyanomethyl (S)-3-(4-Amino-2-oxopyrimidin-1(2H)-yl)-2-((4,5-dimethoxy-2-nitrobenzyloxy)carbonylamino)propionate (6)—To a stirred solution containing 14.0 mg (0.03 mmol) of compound **5** in 1 mL of 1:3:1 water/THF/methanol was added 90.0 μL (0.09 mmol) of 1 N LiOH. The reaction mixture was stirred at room temperature for 2 h, and then concentrated under diminished pressure. The residue was dissolved in 1 mL of anhydrous DMF under argon and 25.0 mg (0.30 mmol) of NaHCO_3 was added followed by 4.0 μL (5.0 mg, 0.06 mmol) of chloroacetonitrile. The reaction mixture was stirred at 23 $^\circ\text{C}$ overnight, then diluted with 20 mL of brine and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10×2 cm). Elution with 5:1 ethyl acetate/methanol gave **6** as a light yellow solid: yield 5.01 mg (34% for two steps); silica gel TLC R_f 0.26 (5:1 ethyl acetate/methanol); ^1H NMR (CD_3OD) δ 3.89 (s, 3H), 3.94-3.96 (m, 4H), 4.36-4.41 (m, 1H), 4.71-4.75 (m, 1H), 4.92 (s, 2H), 5.38-5.48 (m, 2H), 5.84 (br s, 1H), 7.12 (s, 1H), 7.48 (d, 2H, $J = 7.6$ Hz) and 7.71 (s, 1H); mass spectrum (APCI), m/z 477.1377 ($\text{M}+\text{H}^+$) ($\text{C}_{19}\text{H}_{21}\text{N}_6\text{O}_9$ requires m/z 477.1370).

5.2.5. ((S)-3-(4-Amino-2-oxopyrimidin-1(2H)-yl) 2-((4,5-dimethoxy-2-nitrobenzyloxy)carbonylamino)propionic Acid pdCpA Ester (7)—To a solution containing 5.20 mg (4.00 μmol) of pdCpA tetrabutylammonium salt in 100 μL of 9:1 anhydrous DMF/triethylamine was added 10.0 mg (21.0 μmol) of compound **6**. The reaction mixture was sonicated for 5 h. The reaction mixture was purified by HPLC on a C_{18} reversed phase column (250×10 mm) using a linear gradient of 99:1 $\rightarrow$ 1:99 50 mM ammonium acetate, pH 4.5/acetonitrile. The retention time of the desired product was 20.1 min. The fractions containing the product were lyophilized to afford **7** as a colorless solid: yield 1.01 mg (25%); mass spectrum (ESI), m/z 1054.2106 ($\text{M}-\text{H}^-$) ($\text{C}_{36}\text{H}_{42}\text{N}_{13}\text{O}_{21}\text{P}_2$ requires m/z 1054.2093).

5.2.6. Methyl (S)-2-(tert-Butoxycarbonyl)-3-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)propionate (9)—To a stirred suspension containing 250 mg (2.23 mmol) of uracil in 5 mL of DMSO was added 0.40 mL (407 mg, 2.68 mmol) of DBU. Within 15 min, a solution containing 491 mg (2.68 mmol) of *N*-(tert-butoxycarbonyl)-L-serine β -lactone in 5 mL of DMSO was added. The reaction mixture was stirred at room temperature for 2 h under argon, then diluted with 10 mL of 0.5 N HCl and extracted with two 20-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10×2 cm). Elution with 5:1 ethyl acetate/methanol gave **8** as a colorless solid: yield 273 mg (41%); silica gel TLC R_f 0.25 (5:1 ethyl acetate/methanol). To a cooled (0 $^\circ\text{C}$) solution containing 246 mg (0.91 mmol) of **8** in 5 mL of anhydrous MeOH was added dropwise 0.07 mL (71.2 mg, 0.60 mmol) of SOCl_2 . The reaction mixture was allowed to

warm slowly to room temperature with stirring for 2 h, and was then diluted with 50 mL of brine and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10×2 cm). Elution with 10:1 ethyl acetate/methanol gave **9** as a colorless solid: yield 214 mg (75%); silica gel TLC R_f 0.81 (5:1 ethyl acetate/methanol); ^1H NMR (CDCl_3) δ 1.36 (s, 9H), 3.74 (s, 3H), 4.20–4.49 (m, 3H), 5.64 (s, 1H), 5.73 (s, 1H), 7.17 (s, 1H) and 10.10 (s, 1H); ^{13}C NMR (CDCl_3) δ 28.3, 49.7, 52.4, 53.1, 80.8, 102.2, 145.2, 151.4, 155.4, 164.2 and 170.6; mass spectrum (APCI), m/z 314.1345 ($\text{M}+\text{H}^+$) ($\text{C}_{13}\text{H}_{20}\text{N}_3\text{O}_6$ requires m/z 314.1352).

5.2.7. Methyl (S)-2-((4,5-Dimethoxy-2-nitrobenzyloxy)carbonylamino)-3-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)propionate (10)—To a stirred solution containing 94.0 mg (0.30 mmol) of **9** in 2 mL of anhydrous CH_2Cl_2 was added 0.23 mL (342 mg, 3.00 mmol) of trifluoroacetic acid. The reaction mixture was stirred at room temperature for 2 h and then concentrated under diminished pressure. To a stirred solution containing the Boc-protected product in 2 mL of 1:1 dioxane/water was added 124 mg (0.90 mmol) of K_2CO_3 followed by 82.7 mg (0.30 mmol) of NVOC-Cl. The reaction mixture was stirred at room temperature for 14 h under argon, then diluted with 50 mL of brine and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10×2 cm). Elution with 10:1 ethyl acetate/methanol gave **10** as a light yellow solid: yield 73.3 mg (54% for two steps); silica gel TLC R_f 0.60 (10:1 ethyl acetate/methanol); ^1H NMR ($\text{DMSO}-d_6$) δ 3.66 (s, 3H), 3.73–3.78 (m, 1H), 3.87 (s, 3H), 3.91 (s, 3H), 4.19–4.23 (m, 1H), 4.43–4.45 (m, 1H), 5.34 (ABq, 2H, $J=14.5$ Hz), 5.46 (d, 1H, $J=7.5$ Hz), 7.15 (s, 1H), 7.48 (d, 1H, $J=8.0$ Hz), 7.70 (s, 1H) and 8.10 (d, 1H, $J=8.5$ Hz); ^{13}C NMR ($\text{DMSO}-d_6$) δ 48.0, 52.1, 52.3, 56.0, 56.1, 62.8, 100.7, 108.1, 110.5, 127.3, 139.2, 145.8, 147.7, 150.9, 153.3, 155.5, 163.5 and 170.0; mass spectrum (APCI), m/z 453.1259 ($\text{M}+\text{H}^+$) ($\text{C}_{18}\text{H}_{21}\text{N}_4\text{O}_{10}$ requires m/z 453.1257).

5.2.8. Cyanomethyl (S)-2-((4,5-Dimethoxy-2-nitrobenzyloxy)carbonylamino)-3-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)propionate (11)—To a stirred solution containing 22.7 mg (0.05 mmol) of **10** in 1 mL of 1:3:1 water/THF/methanol was added 150 μL (0.15 mmol) of 1 N LiOH. The reaction mixture was stirred at room temperature for 2 h, and then concentrated under diminished pressure. The residue was redissolved in 1 mL of anhydrous DMF and 25.0 mg (0.30 mmol) of NaHCO_3 was added followed by 10.0 μL (11.3 mg, 0.15 mmol) of chloroacetonitrile. The reaction mixture was stirred at 23 $^\circ\text{C}$ overnight, then diluted with 20 mL of brine and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10×2 cm). Elution with 10:1 ethyl acetate/methanol gave the desired product **11** as a yellow solid: yield 7.17 mg (30% for two steps); silica gel TLC R_f 0.75 (5:1 ethyl acetate/methanol); ^1H NMR (CD_3OD) δ 3.91 (s, 3H), 3.99 (s, 3H), 4.33–4.36 (m, 2H), 4.67–4.71 (m, 1H), 4.87–4.93 (m, 2H), 5.41–5.51 (m, 2H), 5.67 (d, 1H, $J=7.6$ Hz), 7.13–7.15 (m, 1H), 7.33–7.43 (m, 1H), 7.73 (s, 1H) and 7.97 (s, 1H); mass spectrum (APCI), m/z 478.1213 ($\text{M}+\text{H}^+$) ($\text{C}_{19}\text{H}_{20}\text{N}_5\text{O}_{10}$ requires m/z 478.1210).

5.2.9. ((S)-2-((4,5-Dimethoxy-2-nitrobenzyloxy)carbonylamino)-3-(2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)propionic Acid pdCpA Ester (12)—To a solution containing 5.20 mg (4.00 μmol) of pdCpA tetrabutylammonium salt in 100 μL of 9:1 anhydrous DMF/triethylamine was added 10.0 mg (21.0 μmol) of **11**. The reaction mixture was sonicated for 5 h. The reaction mixture was purified by HPLC on a C_{18} reversed phase column (250 $\times$ 10 mm) using a linear gradient of 99:1 $\rightarrow$ 1:99 50 mM aq ammonium acetate, pH 4.5/acetonitrile. The retention time of the desired product was 18.6 min. The fractions containing the product were lyophilized to afford **12** as a colorless solid: yield 2.02 mg (46%); mass spectrum (ESI), m/z 1055.1943 (M-H^-) ($\text{C}_{36}\text{H}_{41}\text{N}_{12}\text{O}_{22}\text{P}_2$ requires m/z 1055.1934).

5.2.10. Methyl (S)-2-(tert-Butoxycarbonylamino)-3-(2,4-dioxo-3,4-dihydro-5-methylpyrimidin-1(2H)-yl)propionate (14)—To a stirred suspension containing 280 mg (2.23 mmol) of thymine in 5 mL of DMSO was added 0.40 mL (407 mg, 2.68 mmol) of DBU. Within 15 min, a solution containing 491 mg (2.68 mmol) of *N*-(tert-butoxycarbonyl)-L-serine β -lactone in 5 mL of DMSO was added. The reaction mixture was stirred at room temperature for 2 h under argon, then diluted with 10 mL of 0.5 N HCl and extracted with two 20-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10 $\times$ 2 cm). Elution with 10:1 ethyl acetate/methanol gave **13** as a colorless solid: yield 230 mg (33%); silica gel TLC R_f 0.30 (5:1 ethyl acetate/methanol). To a cooled (0 $^\circ\text{C}$) solution containing 230 mg (0.73 mmol) of **13** in 5 mL of anhydrous MeOH was added dropwise 0.05 mL (87.4 mg, 0.73 mmol) of SOCl_2 . The reaction mixture was allowed to warm slowly to room temperature with stirring for 2 h, then diluted with 50 mL of brine and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10 $\times$ 2 cm). Elution with 10:1 ethyl acetate/methanol gave **14** as a colorless solid: yield 156 mg (65%); silica gel TLC R_f 0.81 (5:1 ethyl acetate/methanol); ^1H NMR (CDCl_3) δ 1.41 (s, 9H), 1.88 (s, 3H), 3.78 (s, 3H), 4.03-4.20 (m, 2H), 4.47-4.50 (m, 1H), 5.55 (s, 1H), 6.98 (s, 1H) and 9.34 (br s, 1H); ^{13}C NMR (CD_3OD) δ 12.4, 28.6, 50.0, 50.3, 53.1, 80.9, 110.7, 143.0, 152.6, 157.2, 166.5 and 171.4; mass spectrum (APCI), m/z 328.1500 (M+H^+) ($\text{C}_{14}\text{H}_{22}\text{N}_3\text{O}_6$ requires m/z 328.1509).

5.2.11. Methyl (S)-2-((4,5-Dimethoxy-2-nitrobenzyloxy)carbonylamino)-3-(2,4-dioxo-3,4-dihydro-5-methylpyrimidin-1(2H)-yl)propionate (15)—To a stirred solution containing 49.2 mg (0.15 mmol) of **14** in 1.5 mL of anhydrous CH_2Cl_2 was added 0.12 mL (171 mg, 1.50 mmol) of trifluoroacetic acid. The reaction mixture was stirred at room temperature for 2 h and then concentrated under diminished pressure. To a stirred solution containing the Boc-protected product in 2 mL of 1:1 dioxane/water was added 124 mg (0.90 mmol) of K_2CO_3 followed by 41.4 mg (0.15 mmol) of NVOC-Cl. The reaction mixture was stirred at room temperature for 14 h under argon, then diluted with 50 mL of brine and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10 $\times$ 2 cm). Elution with 10:1 ethyl acetate/methanol gave **15** as a light yellow solid: yield 32.7 mg (47% for two steps); silica gel TLC

R_f 0.75 (5:1 ethyl acetate/methanol); ^1H NMR (CDCl_3) δ 1.88 (s, 3H), 3.78 (s, 3H), 3.95 (s, 3H), 3.99 (s, 3H), 4.13-4.15 (m, 1H), 4.34-4.37 (m, 1H), 4.57-4.59 (m, 1H), 5.49 (s, 2H), 6.03 (br s, 1H), 6.99 (s, 2H), 7.69 (s, 1H) and 9.08 (br s, 1H); ^{13}C NMR (CDCl_3) δ 12.3, 49.2, 53.2, 53.4, 56.5, 56.6, 64.3, 108.3, 110.4, 111.2, 127.5, 139.8, 140.9, 148.3, 151.7, 153.7, 155.7, 164.4 and 170.0; mass spectrum (APCI), m/z 467.1411 ($\text{M}+\text{H}^+$) ($\text{C}_{19}\text{H}_{23}\text{N}_4\text{O}_{10}$ requires m/z 467.1414).

5.2.12. Cyanomethyl (S)-2-((4,5-Dimethoxy-2-nitrobenzyloxy)carbonylamino)-3-(2,4-dioxo-3,4-dihydro-5-methylpyrimidin-1(2H)-yl)propionate (16)—To a stirred solution containing 23.3 mg (0.05 mmol) of **15** in 1 mL of 1:3:1 water/THF/methanol was added 150 μL (0.15 mmol) of 1 N LiOH. The reaction mixture was stirred at room temperature for 2 h, and then concentrated under diminished pressure. The residue was redissolved in 1 mL of anhydrous DMF and 25.0 mg (0.30 mmol) of NaHCO_3 was added followed by 10.0 μL (11.3 mg, 0.15 mmol) of chloroacetonitrile. The reaction mixture was stirred at 23 °C overnight, then diluted with 20 mL of brine and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10 $\times$ 2 cm). Elution with 10:1 ethyl acetate/methanol gave the desired product **16** as a yellow solid: yield 8.59 mg (35% for two steps); silica gel TLC R_f 0.77 (5:1 ethyl acetate/methanol); ^1H NMR (CD_3OD) δ 1.83 (s, 3H), 3.91 (s, 3H), 3.96 (s, 3H), 4.28-4.34 (m, 2H), 4.70-4.73 (m, 1H), 4.92 (s, 2H), 5.44 (ABq, 2H, $J = 14.8$ Hz), 7.10 (s, 2H), 7.26 (s, 1H) and 7.70 (s, 1H); ^{13}C NMR ($\text{DMSO}-d_6$) δ 11.8, 47.7, 49.8, 51.9, 56.0, 56.2, 62.9, 108.1, 108.4, 110.6, 115.5, 127.1, 139.2, 141.5, 147.7, 150.9, 153.3, 155.5, 164.1 and 168.9; mass spectrum (APCI), m/z 492.1363 ($\text{M}+\text{H}^+$) ($\text{C}_{20}\text{H}_{22}\text{N}_5\text{O}_{10}$ requires m/z 492.1366).

5.2.13. ((S)-2-((4,5-Dimethoxy-2-nitrobenzyloxy)carbonylamino)-3-(2,4-dioxo-3,4-dihydro-5-methylpyrimidin-1(2H)-yl)propionic Acid pdCpA Ester (17)—To a solution containing 5.20 mg (4.00 μmol) of pdCpA tetrabutylammonium salt in 100 μL of 9:1 anhydrous DMF/triethylamine was added 10.3 mg (21.0 μmol) of **16**. The reaction mixture was sonicated for 5 h. The reaction mixture was purified by HPLC on a C_{18} reversed phase column (250 $\times$ 10 mm) using a linear gradient of 99:1 $\rightarrow$ 1:99 50 mM aq ammonium acetate, pH 4.5/acetonitrile. The retention time of the desired product was 20.3 min. The fractions containing the product were lyophilized to afford **17** as a colorless solid: yield 1.81 mg (45%); mass spectrum (ESI), m/z 1069.2089 ($\text{M}-\text{H}^-$) ($\text{C}_{37}\text{H}_{43}\text{N}_{12}\text{O}_{22}\text{P}_2$ requires m/z 1069.2090).

5.2.14. N-(9H-Purin-6-yl)pent-4-enamide (18)—To a stirred solution containing 1.44 g (36.0 mmol) of NaH (60% in mineral oil) in 30 mL of anhydrous DMF at 0 °C was added 1.22 g (9.00 mmol) of adenine. The reaction mixture was stirred at 0 °C for 1 h under argon and then 1.87 g (9.50 mmol) of pentenoic acid NHS ester was added. The reaction mixture was stirred at room temperature overnight under argon, diluted with 150 mL of brine and extracted with two 100-mL portions of EtOAc. The organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10 $\times$ 4 cm). Elution with 10:1 ethyl acetate/methanol gave the desired product **18** as a yellow solid: yield 1.08 g (55%); silica gel TLC R_f 0.50 (10:1 ethyl acetate/

methanol); ^1H NMR (DMSO- d_6) δ 2.47-2.49 (m, 2H), 2.64-2.28 (m, 2H), 5.00 (dd, 1H, J = 10.4 and 1.6 Hz), 5.11 (dd, 1H, J = 17.2 and 2.0 Hz), 5.86-5.96 (m, 3H), 8.39 (s, 1H) and 8.60 (s, 1H); ^{13}C NMR (DMSO- d_6) δ 28.5, 34.5, 112.9, 115.4, 137.1, 143.7, 145.7, 151.1, 161.9 and 172.5; mass spectrum (APCI), m/z 218.1037 ($\text{M}+\text{H}^+$) ($\text{C}_{10}\text{H}_{12}\text{N}_5\text{O}$ requires m/z 218.1042).

5.2.15. (S)-2-(*tert*-Butoxycarbonylamino)-3-(6-Pent-4-enamido-9H-purin-9-yl)propionic acid (19**)**—

To a stirred suspension containing 217 mg (1.00 mmol) of compound **18** in 5 mL of DMSO was added 0.18 mL (183 mg, 1.20 mmol) of DBU. Within 15 min, a solution containing 220 mg (1.20 mmol) of *N*-(*tert*-butoxycarbonyl)-L-serine β -lactone in 5 mL of DMSO was added. The reaction mixture was stirred at room temperature for 12 h under argon, then diluted with 10 mL of 0.5 N HCl and extracted with two 20-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10 $\times$ 2 cm). Elution with 5:1 ethyl acetate/methanol gave **19** as a colorless solid: yield 174 mg (43%); silica gel TLC R_f 0.20 (9:1:0.05 EtOAc/MeOH/HOAc); ^1H NMR (3:1 $\text{CD}_3\text{OD}/\text{CDCl}_3$) δ 1.30 (s, 9H), 2.47-2.49 (m, 2H), 2.64-2.28 (m, 2H), 2.64-2.28 (m, 2H), 4.51-4.57 (m, 2H), 5.00 (dd, 1H, J = 10.4 and 1.6 Hz), 5.11 (dd, 1H, J = 17.2 and 2.0 Hz), 5.90-5.95 (m, 2H), 8.25 (s, 1H) and 8.64 (s, 1H); ^{13}C NMR (3:1 $\text{CD}_3\text{OD}/\text{CDCl}_3$) δ 28.5, 29.9, 37.4, 46.1, 54.8, 80.7, 116.0, 123.1, 129.2, 129.8, 137.9, 145.7, 150.2, 152.9, 157.2 and 173.6; mass spectrum (APCI), m/z 405.1880 ($\text{M}+\text{H}^+$) ($\text{C}_{18}\text{H}_{25}\text{N}_6\text{O}_5$ requires m/z 405.1886).

5.2.16. Cyanomethyl (S)-2-Pent-4-enamido-3-(6-pent-4-enamido-9H-purin-9-yl)propionate (21**)**—

To a stirred solution containing 125 mg (0.31 mmol) of **19** in 2 mL of anhydrous CH_2Cl_2 was added 0.24 mL (353 mg, 3.10 mmol) of trifluoroacetic acid. The reaction mixture was stirred at room temperature for 2 h and then concentrated under diminished pressure. To a stirred solution containing the Boc-protected product in 2 mL of 1:1 dioxane/water was added 129 mg (0.93 mmol) of K_2CO_3 followed by 61.1 mg (0.31 mmol) of pentenoic acid NHS ester. The reaction mixture was stirred at room temperature for 14 h under argon, then diluted with 20 mL of 0.5 N HCl and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10 $\times$ 2 cm). Elution with 5:1 ethyl acetate/methanol gave **20** as a light yellow solid: yield 49.1 mg (41% for two steps); silica gel TLC R_f 0.10 (9:1:0.05 EtOAc/MeOH/HOAc); mass spectrum (APCI), m/z 387.1781 ($\text{M}+\text{H}^+$) ($\text{C}_{18}\text{H}_{23}\text{N}_6\text{O}_4$ requires m/z 387.1781). To a stirred solution containing 54.0 mg (0.14 mmol) of **20** in 1 mL of anhydrous DMF was added 44.0 mg (0.42 mmol) of NaHCO_3 followed by 26.0 μL (32.0 mg, 0.42 mmol) of chloroacetonitrile. The reaction mixture was stirred at 23 $^\circ\text{C}$ for 16 h, then diluted with 20 mL of satd aq NaHCO_3 and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10 $\times$ 2 cm). Elution with ethyl acetate gave **21** as a light yellow solid: yield 18.4 mg (31%); silica gel TLC R_f 0.6 (1:10 methanol/ethyl acetate); ^1H NMR (CD_3OD) δ 2.19-2.21 (m, 4H), 2.48-2.50 (m, 2H), 2.73-2.76 (m, 2H), 3.77 (s, 2H), 4.64-4.75 (m, 2H), 4.99-5.03 (m, 2H), 5.10-5.15 (m, 4H), 5.68-5.69 (m, 1H), 5.90-5.97 (m, 1H), 8.26 (s, 1H) and 8.63 (s, 1H); ^{13}C NMR (CD_3OD) δ

30.2, 30.5, 35.9, 37.5, 45.0, 45.4, 50.7, 53.2, 116.0, 123.7, 138.0, 138.2, 145.9, 150.5, 153.2, 153.7, 169.6, 171.0, 173.8 and 175.4; mass spectrum (APCI), m/z 426.1881 (M+H)⁺ (C₂₀H₂₄N₇O₄ requires m/z 426.1890).

5.2.17. (S)-2-Pent-4-enamido-3-(6-pent-4-enamido-9H-purin-9-yl)propionic Acid pdCpA Ester (22)—To a solution containing 5.20 mg (4.00 μmol) of pdCpA tetrabutylammonium salt in 100 μL of 9:1 anhydrous DMF/triethylamine was added 8.76 mg (21.0 μmol) of **21**. The reaction mixture was sonicated for 5 h. The reaction mixture was purified by HPLC on a C₁₈ reversed phase column (250 × 10 mm) using a linear gradient of 99:1 → 1:99 50 mM aq ammonium acetate, pH 4.5/acetonitrile. The retention time of the desired product was 20.2 min. The fractions containing the product were lyophilized to afford **22** as a colorless solid: yield 1.50 mg (39%); mass spectrum (MALDI), m/z 1003.22 (M-H)[−] (theoretical m/z 1003.26).

5.2.18. (S)-3-(2-Amino-6-chloro-9H-purin-9-yl)-2-(tert-butoxycarbonylamino)propionic acid (23)—To a suspension containing 2.05 g (12.1 mmol) of 2-amino-6-chloropurine in 5 mL of DMSO was added 1.50 mL (1.56 g, 10.2 mmol) of DBU over a period of 15 min. Within 15 min, a solution containing 1.74 g (9.30 mmol) of *N*-(tert-butoxycarbonyl)-L-serine β-lactone dissolved in 5 mL of DMSO was added dropwise and the reaction mixture was stirred for additional 3 h. The reaction was stopped by adding HOAc (585 μL, 10.2 mmol) and the solvent was removed under diminished pressure. Purification of the crude product was done by flash chromatography on a silica gel column (10 × 2 cm) using a gradient of 9:1 EtOAc/MeOH to 9:1:0.05 EtOAc/MeOH/HOAc. After co-evaporation with portions of toluene, the product was obtained as a white solid: yield 2.60 g (80%); *R*_f 0.40 (9:1:0.05 EtOAc/MeOH/HOAc); ¹H NMR (DMSO-*d*₆) δ 1.24 (s, 9H), 4.15–4.45 (m, 3H), 6.69 (s, 1H), 6.89 (s, 2H), 7.89 (s, 1H) and 8.04 (s, 1H); ¹³C NMR (DMSO-*d*₆) δ 27.9, 43.6, 52.6, 78.3, 123.2, 143.3, 149.1, 154.1, 155.1, 159.7 and 171.2; mass spectrum (APCI), m/z 357.1083 (M+H)⁺ (C₁₃H₁₈N₆O₄Cl requires m/z 357.1078).

5.2.19. Cyanomethyl (S)-3-(2-Amino-6-oxo-1,6-dihydropurin-9-yl)-2-pent-4-enamidopropionate (25)—(*S*)-3-(2-Amino-6-chloro-9H-purin-9-yl)-2-(tert-butoxycarbonyl)propionic acid (**23**) (386 mg, 0.75 mmol) was dissolved in 10 mL of 3:1 TFA/H₂O and the reaction mixture was stirred at room temperature for 48 h. Toluene (~50 mL) was added and the solvent was concentrated. After drying *in vacuo* a white solid was obtained in quantitative yield; *R*_f 0.25 (70:30:3:0.3 CHCl₃/MeOH/H₂O/HOAc). The crude solid residue was dissolved in 20 mL of 1:1:2 H₂O/1 N NaOH/dioxane (pH 9) and the reaction mixture was stirred at room temperature for 4 h. The reaction mixture was cooled to 0 °C and 148 mg (0.31 mmol) of pentenoic acid NHS ester was added. The reaction mixture was stirred at room temperature for 14 h under argon, then the solution was acidified to pH 6.5 by adding 1 N HCl. After concentration of the solvent, purification was effected by flash chromatography on a silica gel column (10 × 2 cm) using a gradient of 9:1 EtOAc/MeOH to 9:1:0.05 EtOAc/MeOH/HOAc. After co-evaporation with portions of toluene, **24** was obtained as a white solid: yield 83.3 mg (24%). To a stirred solution containing 9.60 mg (0.03 mmol) of **24** in 1 mL of anhydrous DMF was added 25.0 mg (0.30 mmol) of NaHCO₃

followed by 4.0 μL (5.0 mg, 0.06 mmol) of chloroacetonitrile. The reaction mixture was stirred at 23 $^{\circ}\text{C}$ overnight under argon, then diluted with 20 mL of brine and extracted with two 50-mL portions of EtOAc. The combined organic phase was dried (MgSO_4) and concentrated under diminished pressure. The residue was purified by chromatography on a silica gel column (10×2 cm). Elution with ethyl acetate gave **25** as a yellow solid: yield 5.72 mg (53%); silica gel TLC R_f 0.10 (5:1 ethyl acetate/methanol); ^1H NMR ($\text{DMSO}-d_6$) δ 2.18 (s, 4H), 3.37 (s, 1H), 4.29–4.36 (m, 2H), 4.66–4.69 (m, 1H), 4.89–4.95 (m, 1H), 5.02 (s, 2H), 5.75 (s, 1H), 6.85 (s, 2H), 7.64 (s, 1H), 8.79 (s, 1H) and 10.96 (s, 1H); ^{13}C NMR ($\text{DMSO}-d_6$) δ 28.8, 33.9, 42.4, 49.7, 51.4, 115.0, 115.5, 116.3, 137.2, 137.6, 151.1, 153.9, 156.6, 168.7 and 172.0; mass spectrum (APCI), m/z 360.1425 ($\text{M}+\text{H}$) $^+$ ($\text{C}_{15}\text{H}_{18}\text{N}_7\text{O}_4$ requires m/z 360.1420).

5.2.20. (S)-3-(2-Amino-6-oxo-1,6-dihydropurin-9-yl)-2-pent-4-enamidopropionic Acid pdCpA Ester (26)—To a solution containing 5.20 mg (4.00 μmol) of pdCpA tetrabutylammonium salt in 100 μL of 9:1 anhydrous DMF/triethylamine was added 7.54 mg (21.0 μmol) of **25**. The reaction mixture was sonicated for 5 h. The reaction mixture was purified by HPLC on a C_{18} reversed phase column (250×10 mm) using a linear gradient of 99:1 $\rightarrow$ 1:99 50 mM aq ammonium acetate, pH 4.5/acetonitrile. The retention time of the desired product was 17.3 min. The fractions containing the product were lyophilized to afford **26** as a colorless solid: yield 2.50 mg (70%); mass spectrum (MALDI), m/z 937.44 ($\text{M}-\text{H}$) $^-$ (theoretical m/z 937.21).

5.3. Biochemical experiments

5.3.1. Preparation of aminoacyl-tRNA_{CUAS}—The activation of suppressor tRNA_{CUAS} was carried out as described previously.^{19,20} Briefly, a 100- μL reaction mixture of 100 mM Na Hepes, pH 7.5, contained 1.0 mM ATP, 15 mM MgCl_2 , 100 μg of suppressor tRNA_{CUA-COH}, 0.5 A_{260} unit of *N*-pentenoyl-protected aminoacyl-pdCpA, 15% DMSO, and 100 units of T4 RNA ligase. The reaction mixture was incubated at 37 $^{\circ}\text{C}$ for 1.5 h and quenched by the addition of 0.1 vol of 3 M NaOAc, pH 5.2. The *N*-protected aminoacylated tRNA was precipitated with 3 vol of cold ethanol. The efficiency of ligation was estimated by 8% polyacrylamide–7 M urea gel electrophoresis (pH 5.0).³² The *N*-pentenoyl-protected aminoacyl-tRNA_{CUAS} were deprotected by treatment with aqueous I_2 , typically 5 mM I_2 at 25 $^{\circ}\text{C}$ for 30 min.^{19,20} The NVOC protecting group was removed by exposure to high intensity UV light at 4 $^{\circ}\text{C}$ for 3 min.¹⁶ The solution was centrifuged, and the supernatant was adjusted to 0.3 M NaOAc and then treated with 3 vol of cold ethanol to precipitate the aminoacyl-tRNA. The tRNA pellet was collected by centrifugation, washed with 70% aq EtOH, air dried and dissolved in 10 μL of RNase free water.

5.3.2. Plasmids employed for protein expression—The DNA plasmids encoding *Escherichia coli* dihydrofolate reductase (DHFR), and having a TAG codon at the position corresponding to position 10 in DHFR,¹⁶ and encoding the Klenow fragment of *E. coli* DNA polymerase I,²⁶ and having a TAG codon at position 484, were prepared as described previously.

5.3.3. *In vitro* protein translation—Protein translation reactions were carried out in 15 μL of incubation mixture containing 0.3 $\mu\text{L}/\mu\text{L}$ of S-30 system, 170 ng/ μL of plasmid, 35 mM Tris acetate, pH 7.4, 190 mM potassium glutamate, 30 mM ammonium acetate, 2 mM dithiothreitol, 0.2 mg/mL total *E. coli* tRNA, 3.5% PEG 6000, 20 $\mu\text{g}/\text{mL}$ folinic acid, 20 mM ATP and GTP, 5 mM CTP and UTP, 100 μM amino acids mixture, 0.5 $\mu\text{Ci}/\mu\text{L}$ of ^{35}S -methionine and 1 $\mu\text{g}/\text{mL}$ rifampicin. In the case of plasmids having a gene with a TAG codon, an activated suppressor tRNA was added to a concentration of 1.5 $\mu\text{g}/\mu\text{L}$. Reactions were carried out at 37 °C for 1 h and were terminated by chilling on ice. Aliquots from *in vitro* translation mixtures were analyzed by SDS-PAGE followed by quantification of the radioactive bands by phosphorimager analysis (for 15- μL reaction mixtures).

Acknowledgements

We thank Prof. Stephen Benkovic, The Pennsylvania State University, for the expression plasmid containing the gene for the Klenow fragment of *E. coli* DNA polymerase I. This study was supported by the National Institutes of Health Research Grant GM103861, awarded by the National Institute of General Medical Sciences.

Abbreviations

APCI	atmospheric pressure chemical ionization
ESI	electrospray ionization
NVOC-Cl	4,5-dimethoxy-2-nitrobenzyl chloroformate
TBA	tetra- <i>n</i> -butylammonium
DBU	1,8-diazabicyclo 5.4.0 undec-7-ene

References

1. Dey B, Thukral S, Krishnan S, Chakrobarty M, Gupta S, Manghani C, Rani V. *Mol. Cell. Biochem. Biochem.* 2012; 65:279.
2. Travers, A. *DNA-Protein Interactions*. Chapman and Hall; London: 1993.
3. Pabo CO, Sauer RT. *Annu. Rev. Biochem.* 1984; 53:293. [PubMed: 6236744]
4. Dickerson RE. *Sci. Am.* 1983; 249:94.
5. Strauch MA. *Protein-DNA Complexes: Specific*. Encyclopedia of Life Sciences. 2001
6. Varani G, Nagai K. *Annu. Rev. Biophys. Biomol. Struct.* 1998; 27:407. [PubMed: 9646873]
7. Nielsen PE, Egholm M, Berg RH, Buchardt O. *Science*. 1991; 254:1497. [PubMed: 1962210]
8. Matsumura S, Takahashi T, Ueno A, Mihara H. *Chem. Eur. J.* 2003; 9:4829. [PubMed: 14566891]
9. Takahashi T, Hamasaki K, Ueno A, Mihara H. *Bioorg. Med. Chem.* 2001; 9:991. [PubMed: 11354682]
10. Miyanishi H, Takahashi T, Mihara H. *Bioconjugate Chem.* 2004; 15:694.
11. Roviello GN, Gröschel S, Pedone C, Diederichsen U. *Amino Acids*. 2010; 38:1301. [PubMed: 19629638]
12. Peifer M, Giacomo FD, Schandl M, Vasella A. *Helv. Chim. Acta.* 2009; 92:1134.
13. Talukder P, Chen S, Roy B, Yakovchuk P, Spiering MM, Alam MP, Madathil MM, Bhattacharya C, Benkovic SJ, Hecht SM. *Biochemistry*. 2015; 54:7457. [PubMed: 26618501]
14. Chen S, Fahmi NE, Wang L, Bhattacharya C, Benkovic SJ, Hecht SM. *J. Am. Chem. Soc.* 2013; 135:12924. [PubMed: 23941571]

15. Robertson SA, Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG. *Nucleic Acids Res.* 1989; 17:9649. [PubMed: 2602139]
16. Karginov VA, Mamaev SV, An H, Van Cleve MD, Hecht SM, Komatsoulis GA, Abelson JN. *J. Am. Chem. Soc.* 1997; 119:8166.
17. Diederichsen U, Weicherdingab D, Diezemann N. *Org. Biomol. Chem.* 2005; 3:1058. [PubMed: 15750649]
18. Robertson SA, Ellman JA, Schultz PG. *J. Am. Chem. Soc.* 1991; 113:2722.
19. Lodder M, Golovine S, Hecht SM. *J. Org. Chem.* 1997; 62:778.
20. Lodder M, Golovine S, Laikhter AL, Karginov VA, Hecht SM. *J. Org. Chem.* 1998; 63:794. [PubMed: 11672075]
21. Short GF III, Golovine SY, Hecht SM. *Biochemistry.* 1999; 38:8808. [PubMed: 10393557]
22. Eisenhauer BM, Hecht SM. *Biochemistry.* 2002; 41:11472. [PubMed: 12234190]
23. Dedkova LM, Fahmi NE, Golovine SY, Hecht SM, S M. *J. Am. Chem. Soc.* 2003; 125:6616. [PubMed: 12769555]
24. Fahmi NE, Dedkova L, Wang B, Golovine S, Hecht SM. *J. Am. Chem. Soc.* 2007; 129:3586–3597. [PubMed: 17338522]
25. Dedkova LM, Fahmi NE, Paul R, del Rosario M, Zhang L, Chen S, Feder G, Hecht SM. *Biochemistry.* 2012; 51:401. [PubMed: 22145951]
26. Talukder P, Chen P, Roy S, Yakovchuk B, Spiering P, Alam MM, Madathil MP, Bhattacharya MMC, Benkovic SJ, Hecht SJ, M. S. *Biochemistry.* 2015; 54:7457. [PubMed: 26618501]
27. Watanabe S, Tomizaki K, Takahashi T, Usui K, Kajikawa K, Mihara H. *Peptide Sci.* 2007; 88:131.
28. Takahashi T, Yana D, Mihara H, H. *ChemBioChem.* 2006; 7:729. [PubMed: 16550624]
29. Albright RA, Mossing MC, Matthews BW. *Protein Sci.* 1998; 7:1485. [PubMed: 9684880]
30. Tsai FTF, Sigler PB. *EMBO J.* 2000; 19:25. [PubMed: 10619841]
31. Garvie CW, Wolberger C. *Mol. Cell.* 2001; 8:937. [PubMed: 11741530]
32. Varshney U, Lee CP, RajBhandary UL. *J. Biol. Chem.* 1991; 266:24712. [PubMed: 1761566]

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.

Figure 2.

8% polyacrylamide–7 M urea gel electrophoresis of aminoacyl-tRNA samples. Lane 1, abbreviated tRNA_{CUA}-COH (74 nucleotides); lanes 2-6, ligation products of abbreviated tRNA_{CUA}-COH incubated with pdCpA derivatives activated with amino acids **A**, **B**, **C**, **D** and **E**, respectively (76 nucleotides).

Figure 3.

(A) Analysis of products of *in vitro* translation of DHFR from wild-type (lane 1) and modified mRNA (UAG codon in protein position 10) (lanes 2-8) in the absence (lane 2) and in the presence of nucleobase aminoacyl-tRNA_{CUA}S (A, B, C, D and E in lanes 3-7, respectively), and phenylalanyl-tRNA_{CUA} (lane 8) by 15% SDS-polyacrylamide gel electrophoresis. (B) Analysis of products of *in vitro* translation of the Klenow fragment of DNA polymerase I from wild-type (lane 1) and modified mRNA (UAG codon in protein position 484) (lanes 2-8) in the absence (lane 2) and in the presence of nucleobase aminoacyl-tRNA_{CUA}S (A, B, C, D and E in lanes 3-7, respectively), and phenylalanyl-tRNA_{CUA} (lane 8) by 10% SDS-polyacrylamide gel electrophoresis. The arrows denote full length protein bands; the lower bands are truncation products resulting from failure to suppress the UAG codon. The absence of lower bands in panel A reflects the small size of the truncated product (UAG codon at position 10).

Scheme 1.

Synthetic route employed for the preparation of protected amino acid **1** and its (NVOC-protected) aminoacyl-pdCpA derivative.

Scheme 2. Preparation of suppressor tRNA_{CUA} activated with amino acid A.

Scheme 3.

Synthesis of nucleobase amino acid **B** and its corresponding aminoacyl-pdCpA derivative.

Scheme 4. Synthesis of nucleobase amino acid **C** and its corresponding aminoacyl-pdCpA derivative.

Scheme 5.

Synthesis of nucleobase amino acid **D** and its corresponding aminoacyl-pdCpA derivative.

Scheme 6.

Synthesis of nucleobase amino acid **E** and its corresponding aminoacyl-pdCpA derivative.

Table 1

Expression Yields of Full Length DHFRs and Klenow Fragments of DNA Polymerase I, Modified at Positions 10 and 484, Respectively.

Amino acids	Yields of modified proteins relative to wild-type, %	
	DHFR (position 10)	Klenow fragment (position 484)
A	17 ± 1	15 ± 3
B	18 ± 2	14 ± 3
C	13 ± 3	11 ± 2
D	6 ± 2	6 ± 3
E	8 ± 1	7 ± 2
Phe (positive control)	30 ± 9	27 ± 8
– (negative control)	0.2 ± 0.1	0.3 ± 0.1