A High-Fidelity Codon Set for the T4 DNA ligase-Catalyzed Polymerization of Modified Oligonucleotides

Yi Lei, Dehui Kong, Ryan Hili*
Department of Chemistry, University of Georgia, 140 Cedar Street, Athens, Georgia 30602-2556, United States

Supporting Information

Supporting Methods

General Information

Unless otherwise noted, water was purified with Milli-Q purification system. DNA oligonucleotides without amine modification were purchased from Integrated DNA Technologies. DNA oligonucleotides with amine modification were synthesized on a Bioautomation Mermade 12 synthesizer. All materials and reagents used for oligonucleotide synthesis were purchased from Glen Research. All oligonucleotides were synthesized and deprotected according to the manufacturer's protocols. Oligonucleotides were purified by reverse-phase high-pressure liquid chromatography (HPLC, Agilent 1260) using a C18 stationary phase (Eclipse-XDB C18, 5 μm, 9.4 x 200 mm) and an acetonitrile/100 mM triethylammonium acetate gradient. Oligonucleotide concentrations were quantitated by UV spectroscopy using a Nanodrop ND2000 spectrophotometer.

DNA sequences

The sequences below are written from 5'→3'. <Aam> = Amino-modifer C6 dA, <N>=A/T/C/G, <W>=A/T

Templates

TP(NNNNN)8P /5Phos/GA TTC GCC TGC CGT CGC ANN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNN NNC ACG TGG AGC TCG GAT CCT

TP(WNWNT)8P /5Phos/GA TTC GCC TGC CGT CGC AWN WNT WNW NTW NWN TWN WNT WNW NTW NWN TWN WNT WNW NTC ACG TGG AGC TCG GAT CCT

TP(NNNNT)8P /5Phos/GA TTC GCC TGC CGT CGC ANN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTC ACG TGG AGC TCG GAT CCT

TP(NNNTN)8P /5Phos/GA TTC GCC TGC CGT CGC ANN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNC ACG TGG AGC TCG GAT CCT

TP(NNTNN)8P /5Phos/GA TTC GCC TGC CGT CGC ANN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNC ACG TGG AGC TCG GAT CCT

TP(NTNNN)8P /5Phos/GA TTC GCC TGC CGT CGC ANT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNC ACG TGG AGC TCG GAT CCT

TP(TNNNN)8P /5Phos/GA TTC GCC TGC CGT CGC ATN NNN TNN NNT NNN NTN NNN TNN NNT NNN NTN NNN TNN NNC ACG TGG AGC TCG GAT CCT

Primers

PR5 /5Phos/GG ATC CGA GCT CCA CGT G

PR6 /5Phos/TG CGA CGG CAG GCG AAT CT

iTruS_i7_D701 CAA GCA GAA GAC GGC ATA CGA GAT ATT ACT CGG TGA CTG GAG TTC AG

iTruS_i7_D702 CAA GCA GAA GAC GGC ATA CGA GAT TCC GGA GAG TGA CTG GAG TTC AG

iTruS_i7_D703 CAA GCA GAA GAC GGC ATA CGA GAT CGC TCA TTG TGA CTG GAG TTC AG

iTruS_i7_D704 CAA GCA GAA GAC GGC ATA CGA GAT GAG ATT CCG TGA CTG GAG TTC AG

iTruS_i7_D705 CAA GCA GAA GAC GGC ATA CGA GAT ATT CAG AAG TGA CTG GAG TTC AG

iTruS_i7_D706 CAA GCA GAA GAC GGC ATA CGA GAT GAA TTC GTG TGA CTG GAG TTC AG

iTruS_i7_D707 CAA GCA GAA GAC GGC ATA CGA GAT CTG AAG CTG TGA CTG GAG TTC AG

iTruS_i7_D708 CAA GCA GAA GAC GGC ATA CGA GAT TAA TGC GCG TGA CTG GAG TTC AG

iTruS_i7_D709 CAA GCA GAA GAC GGC ATA CGA GAT CGG CTA TGG TGA CTG GAG TTC AG

iTruS_i7_D710 CAA GCA GAA GAC GGC ATA CGA GAT TCC GCG AAG TGA CTG GAG TTC AG

iTruS_i7_D711 CAA GCA GAA GAC GGC ATA CGA GAT TCT CGC GCG TGA CTG GAG TTC AG

iTruS_i7_D712 CAA GCA GAA GAC GGC ATA CGA GAT AGC GAT AGG TGA CTG GAG TTC AG

PRIMER B AAT GAT ACG GCG ACC ACC GAG

Adapters

Adapter A AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T

Adapter B /5phos/AC TGN NNN NNN NNN NNA GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC

Pentanucleotides

NNNNN /5Phos/NNNNN

ANWNW /5Phos/ANWNW

ANNNN /5Phos/ANNNN

NH_2-ANNNN/5Phos/<Ama>NNNN

NANNN /5Phos/NANNN

NH_2-NANNN /5Phos/N<Ama>NNN

NNANN /5Phos/NNANN

NH_2-NNANN /5Phos/NN<Ama>NN

NNNAN /5Phos/NNNAN

NH_2-NNNAN /5Phos/NNN<Ama>N

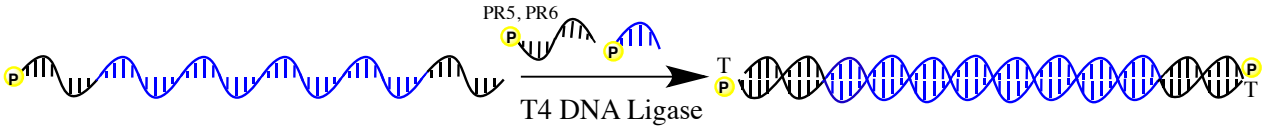
NNNNA /5Phos/NNNNA

NH_2-NNNNA /5Phos/NNNN<Ama>

Synthesis of amino-modified pentanucleotides

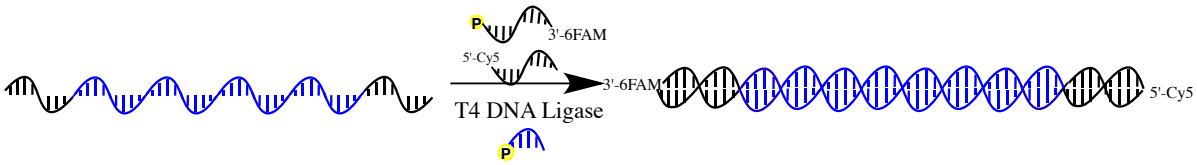
Pentanucleotides were synthesized on a Mermaid 12 DNA synthesizer using a DMT-ON protocol on a 1 µmol scale (1000 Å CPG column). Amine-modifier C6 dA (Glen Research 10-1089), dA+dC+dG+dT-CE Phosphoramindite (Glen Research 10-1000, 10-1010, 10-1020, 10-1030), Chemical Phosphorylation Reagent II (10-1901) were incorporated as specified by the manufacturer. Following synthesis, the oligonucleotide was cleaved from the resin by incubation at 65 °C in 500 µL of a 1:1 mixture of ammonium hydroxide and methylamine for 15 minutes. The cleaved resin was filtered away by filtration, and the oligonucleotide was concentrated under reduced pressure using a speedvac. The residue was then taken up into 100 µL of H2O, and purified using reverse-phase HPLC purification using a [10% acetonitrile in 0.1 M TEAA, pH 7] to [80% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was then incubated at room temperature in 500 µL of 80% aqueous acetic acid for 1 h to cleave the DMT group, and then frozen and lyophilized. The oligonucleotide was incubated in 500 µL 30% ammonium hydroxide at room temperature for 15 minutes to cleave the CPRII linker. Following deprotection, the oligonucleotide was concentrated under reduced pressure using a speedvac. The dried product was dissolved into 100 µL H2O and subjected to reverse-phase HPLC purification using a [10% acetonitrile in 0.1 M TEAA, pH 7] to [80% acetonitrile in 0.1 M TEAA, pH 7] solvent gradient with a column temperature of 45°C. The purified oligonucleotide was dissolved in water.

T4-DNA Ligase-mediate polymerization protocols



In a PCR tube was added DNA template (15 pmol in 1.5 µl water), PR5 (22.5 pmol in 2.25 µl water), PR6 (22.5 pmol in 2.25 µl water), 4µl NEBNext® Quick Ligation Reaction Buffer 5X, 7 µl water. The mix was heated to 90 °C for 2 minutes and then cooled to 25°C at the rate of 0.1 °C/s. In this PCR tube was then added pentanuclotides (480 pmol in 1 µl water), BSA (2µg in 1µl water) and 400 U T4 DNA ligase (New England Biolabs, M0202L). The polymerization was performed at 25 °C for 24 hours. The products were then purified with MinElute® PCR Purification Kit for adapter ligation.

Polymerization efficiency with fluorescently-labeled primers



In a PCR tube was added DNA template TP(NNNNT)8P (15 pmol in 1.5 µl water), 5' Cy5 modified PR5 (22.5 pmol in 2.25 µl water), 3' 36FAM modified PR6 (22.5pmol in 2.25µl water), 4 µl NEBNext® Quick Ligation Reaction Buffer 5×, 7µl water for the experimental reaction, and 8 µl water for the two control reactions. The mix was heated to 90 °C for 2 minutes and then cooled to 25 °C at the rate of 0.1 °C/s. In experimental reaction PCR tube was then added pentanucleotides ANNNN (480 pmol in 1 µl water), BSA (2 µg in 1 µl water) and 400 U T4 DNA ligase (New England Biolabs, M0202L). In one of the control reaction was added pentanucleotides ANNNN (480 pmol in 1 µl water) and BSA (2 µg in 1 µl water). The other control reaction was added 400 U T4 DNA ligase (New England Biolabs, M0202L) and BSA (2 µg in 1 µl water). The polymerization was performed at 25 °C for 24 hours. The products were then purified with MinElute® PCR Purification Kit. The gel was imaged using a Typhoon 9410 imager.

Forced misincorporation polymerization

DNA template

matched-3' TCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTACGCTGCCGTCCCCTTGGAC GGCAGCGTAGAGA

mismatch1-3' TCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTACGCTGCCGTCCCCTTGGAC GGCAGCGTAGACA

mismatch2-3' TCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTACGCTGCCGTCCCCTTGGAC GGCAGCGTAGTGA

mismatch3-3' TCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTACGCTGCCGTCCCCTTGGAC GGCAGCGTACAGA

mismatch4-3' TCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTACGCTGCCGTCCCCTTGGAC GGCAGCGTTGAGA
matched-5' /5Phos/AGAGATGCGACGGCAGGTTCCCCTGCCGTCGCATCTCTTCTCTTCTCTTCTCT TCTCTTCTCTTCTCTTCTCTC
mismatch1-5' /5Phos/ACAGATGCGACGGCAGGTTCCCCTGCCGTCGCATCTCTTCTCTTCTCTTCTC TTCTCTTCTCTTCTCTTCTCTC
mismatch2-5' /5Phos/ACTGATGCGACGGCAGGTTCCCCTGCCGTCGCATCTCTTCTCTTCTCTTCTC TTCTCTTCTCTTCTCTTCTCTC
mismatch3-5' /5Phos/ACAGATGCGACGGCAGGTTCCCCTGCCGTCGCATCTCTTCTCTTCTCTTCTC TTCTCTTCTCTTCTCTTCTCTC
mismatch4-5' /5Phos/ACAGTTGCGACGGCAGGTTCCCCTGCCGTCGCATCTCTTCTCTTCTCTTCTC TTCTCTTCTCTTCTCTTCTCTC
matchedB-3' TCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTACGCTGCCGTCCCCTTGGAC GGCAGCGT
mismatchedB1-3' TCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCAACGCTGCCGTCCCCTTG GACGGCAGCGT
mismatchedB2-3' TCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTGTACGCTGCCGTCCCCTTG GACGGCAGCGT
mismatchedB3-3' TCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCACTACGCTGCCGTCCCCTTG GACGGCAGCGT
mismatchedB4-3' TCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTGTCTACGCTGCCGTCCCCTTG GACGGCAGCGT
mismatchedB5-3' TCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCATCTCTACGCTGCCGTCCCCTTG GACGGCAGCGT
matchedB-5' /5Phos/TGCGACGGCAGGTTCCCCTGCCGTCGCATCTCTTCTCTTCTCTTCTCTTCTCT TCTCTTCTCTTCTCTC
mismatchedB1-5' /5Phos/TGCGACGGCAGGTTCCCCTGCCGTCGCAACTCTTCTCTTCTCTTCTC TTCTCTTCTCTTCTCTTCTCTC
mismatchedB2-5' /5Phos/TGCGACGGCAGGTTCCCCTGCCGTCGCATGTCTTCTCTTCTCTTCTC TTCTCTTCTCTTCTCTTCTCTC
mismatchedB3-5' /5Phos/TGCGACGGCAGGTTCCCCTGCCGTCGCATCACTTCTCTTCTCTTCTC TTCTCTTCTCTTCTCTTCTCTC
mismatchedB4-5' /5Phos/TGCGACGGCAGGTTCCCCTGCCGTCGCATCTGTTCTCTTCTCTTCTC TTCTCTTCTCTTCTCTTCTCTC
mismatchedB5-5' /5Phos/TGCGACGGCAGGTTCCCCTGCCGTCGCATTCATCTCTTCTCTTCTC TTCTCTTCTCTTCTCTTCTCTC

Pentanucleotides
P-AGAGA 5Phos/AGAGA

Protocol

In the PCR tube was added DNA template (15 pmol in 1.5 µl water), 4 µl NEBNext® Quick Ligation Reaction Buffer 5X, 11.5 µl water. The mix was heated to 90 °C for 2 minutes and then cooled to 25 °C at the rate of 0.1 °C/s. In all tubes were added BSA (2 µg in 1 µl water), 400 U T4 DNA ligase (New England Biolabs, M0202L) and P-AGAGA (480 pmol in 1 µl water).

The products were then purified with MinElute® PCR Purification Kit. PAGE analysis was then performed (15% Mini-PROTEAN® TBE-Urea Gel) at 55 °C for 90 minutes.

Single-nucleotide chain termination polymerization

DNA template
idelitymixA1-c-5' 5Phos/TG CGA CGG CAG GTT CCC CTG CCG TCG CAT CTC TAG AGT AGA GTA GAG TAG AGT AGA GTA GAG TAG AGT C

Pentanucleotides
P-ACTCT 5Phos/ACTCT
P-AGAGA 5Phos/AGAGA
TGAGA TGAGA
ACAGA ACAGA
AGTGA AGTGA
AGAGT AGAGT

Protocol

In 5 PCR tubes were added DNA template (15 pmol in 1.5 µl water), 4 µl NEBNext® Quick Ligation Reaction Buffer 5X, 11.5 µl water. The mix was heated to 90 °C for 2 minutes and then cooled to 25 °C at the rate of 0.1 °C/s. In all tubes were added BSA (2 µg in 1 µl water), 400 U T4 DNA ligase (New England Biolabs, M0202L) and P-ACTCT (420 pmol in 1 µl water). In the first tube was added P-AGAGA (60 pmol in 1 µl water), the second tube was added TGAGA (60 pmol in 1 µl water), the third tube was added TGAGA/ACAGA/AGTGA/AGACA/AGAGT mixture (60 pmol each, 300 pmol total in 5 µl water), the fourth tube was added mixture (120 pmol each, 600 pmol total in 5 µl water), and the fifth tube was added mixture (180 pmol each, 900 pmol total in 5 µl water). The polymerization was performed at 25 °C for 24 hours. The products were then purified with MinElute® PCR Purification Kit. PAGE analysis was then performed (15% Mini-PROTEAN® TBE-Urea Gel) at 55 °C for 90 minutes.

Restriction digestion of polymerization product

DNA template
TP(rand1)8P GAT TCG CCT GCC GTC GCA TCT CTT CTC TTC TCT TCT CTT CTC TTC TCT TCT CTT CTC TCA CGT GGA GCT CGG ATC C

Protocol
In a PCR tube polymerization was performed with TP(rand1)8P and ANNNN following the T4-DNA ligase-mediate polymerization protocol.
In another PCR tube was added DNA template TP(rand1)8P (15 pmol in 1.5 µl water), PR1 (22.5 pmol in 2.25 µl water), 5 µl 10× NEBuffer2 (New England Biolabs, M0212L), 37.5 µl water. The mix was heated to 90 °C for 2 minutes and then cooled to 37 °C at the rate of 0.1 °C/s. The tube was then added 5 U Klenow Fragment (3'→5' exo$^-$, New England Biolabs, M0212L) and dNTP (1.25 µl 10 mM stock solution, Thermo Scientific, R0192). The extension

was then performed at 37 °C for 1 hour. The products were then purified with MinElute® PCR Purification Kit for further digestion.
In the PCR tubes were added 10µl polymerized dsDNA and extended dsDNA separately, 5µl 10× CutSmart® Buffer (New England Biolabs, R0542S), 10 U BcoDI (New England Biolabs, R0542S) and 34 µl water. The mix was then incubated at 37 °C overnight. The digestion products were purified with Centri-Sep columns (Princeton Separations, CS-900). PAGE analysis was then performed (15% Mini-PROTEAN® TBE-Urea Gel) at 55 °C for 90 minutes.

Sanger sequencing of polymerized product

DNA template

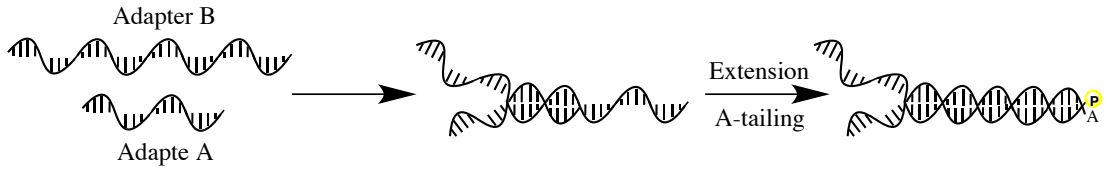
TP(rand1)8PS GAT TCG CCT GCC GTC GCA TCT CTT CTC TTC TCT TCT CTT CTC TTC TCT TCT CTT CTC TCA CGT GGA GCT CGG ATC C/iSp18/A ACA ACA ACA ACA A

Protocol

In a PCR tube polymerization was performed with TP(rand1)8PS and ANNNN following the T4-DNA ligase-mediate polymerization protocol.
In another PCR tube was added DNA template TP(rand1)8PS (15 pmol in 1.5 µl water), PR1 (22.5 pmol in 2.25 µl water), 5 µl 10× NEBuffer2 (New England Biolabs, M0212L), 37.5 µl water. The mix was heated to 90 °C for 2 minutes and then cooled to 37 °C at the rate of 0.1 °C/s. The tube was then added 5 U Klenow Fragment (3'→5' exo$^-$, New England Biolabs, M0212L) and dNTP (1.25 µl 10 mM stock solution, Thermo Scientific, R0192). The extension was then performed at 37 °C for 1 hour. The products were then purified with MinElute® PCR Purification Kit.
The polymerized strand and extended strand were then purified by gel purification (15% Mini-PROTEAN® TBE-Urea Gel). After stained with 0.5× SYBR® safe DNA gel stain (Life Technologies, S33100) for 30 minutes, the product bands were cut out and the nucleotides were eluted with 0.3 M NaCl at room temperature over night. The products were then purified with Centri-Sep columns (Princeton Separations, CS-900). The nucleotides were then amplified with 16-cycle and 25-cycle PCR for extended strand and polymerized strand separately. The PCR products were sent to Georgia Genomic Facility at UGA for SANGER sequencing.

T4-DNA Ligase-mediate adapter ligation protocols

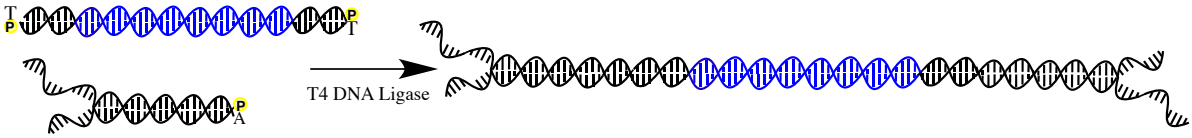
Adapter duplex synthesis



In a PCR tube was added 15 µl of 100 µM adapter A and 15 µl of 100 µM adapter B, then the tube was heated to 95 °C for 5 minutes, and cooled to room temperature over 1 hour. Then in this PCR tube was added 4 µl NEBuffer 2 10× (New England Biolabs, M0212L), 25 U Klenow Fragment (3'→5' exo$^-$, New England Biolabs, M0212L), 1 µl dNTP Mix (Thermo

Scientific, 10mM each). The extension was performed at 37 °C for 1 hour. The adapter duplex was purified with QIAquick Nucleotide Removal Kit, and then diluted in 30 µl water.
In a PCR tube was added 30µl purified adapter duplex, 5 µl NEBuffer 2 10× (New England Biolabs, M0212L), 25 U Klenow Fragment (3'→5' exo$^-$, New England Biolabs, M0212L), 5µl dATP (Thermo Scientific, 10mM), 5µl water. This PCR tube was incubated at 37 °C for 1 hour for A-tailing. Then product was purified with QIAquick Nucleotide Removal Kit, and then diluted in 30 µl water.

Adapter ligation



In a PCR tube was added polymerization products (12 pmol in 10 µl water), A-tailing adapter duplex (240 pmol), 10µl NEBNext® Quick Ligation Reaction Buffer 5x, BSA (5 µg in 2.5 µl water), 1000 U T4 DNA ligase (New England Biolabs, M0202L), water (make total volume 50 µl). Then the ligation was performed at 16 °C for 16 hours, resulting in approximately 40% conversion into desired double-ligated product. Gel purification was then performed. The products were diluted to 10 µl water.

PCR protocols

Each purified adapter ligation product was amplified with a different primer from iTrus_D701 to iTrus_D712. The combinations were shown below:
iTrus_D701+ TP(NNNNN)8P+ NNNNN
iTrus_D702+ TP(WNWNT)8P+ ANWNW
iTrus_D703+ TP(NNNNT)8P+ ANNNN
iTrus_D704+ TP(NNNNT)8P+ NH_2-ANNNN
iTrus_D705+ TP(NNNTN)8P+ NANNN
iTrus_D706+ TP(NNNTN)8P+ NH_2-NANNN
iTrus_D707+ TP(NNTNN)8P+ NNANN
iTrus_D708+ TP(NNTNN)8P+ NH_2-NNANN
iTrus_D709+ TP(NTNNN)8P+ NNNAN
iTrus_D710+ TP(NTNNN)8P+ NH_2-NNNAN
iTrus_D711+ TP(TNNNN)8P+ NNNNA
iTrus_D712+ TP(TNNNN)8P+ NH_2-NNNNA

In a PCR tube was added 1 µl purified adapter ligation product, 1.25 µl 10 µM Primer B, 1.25 µl 10 µM corresponding iTrus_D7XX primer, 9 µl water and 12.5 µl Q5® High-Fidelity 2× Master Mix (New England Biolabs). The tube was then transferred to a preheated themocylcler (98°C). The annealing temperature was 55°C for the first two cycles, and remained 71 °C for the rest of the cycles.
The PCR cycles were 20 for iTrus_D704 and iTrus_D712 since the corresponding adapter ligation yields were lower, and all the others were amplified for 10 cycles. The PCR products were then purified by gel purification.

In vitro selection

DNA template

POS2(rand1) /5BiosG/GA TTC GCC TGC CGT CGC AGC TCT TCC GTC ACG TAC CTT AGT TTG CCA TCG TCT CAC CTC ACG TGG AGC TCG GAT

Protocol

Polymerization was performed with the template mixture of POS2(rand1) :TP(NTNNN)8P= 1:100, following the following the T4-DNA ligase-mediate polymerization protocol with NNNAN pentanucleotides. The polymerized products were then incubated with 30µl streptavidin magnetic beads (Life Technologies, Dynabeads® MyOne™ Streptavidin C1), which were washed with 200 µl binding buffer (0.5 M NaCl, 200 mM Tris-HCl pH=7.5, 1 mM EDTA) 3 times before binding. After 30 minutes incubation at room temperature, TP(NTNNN)8P products were washed away with 200 µl binding buffer 3 times. The dsDNA was then separated with 40 µl freshly made 150 mM NaOH. After incubation at room temperature for 15 minutes, 4 µl 1.5 M HCl quenched the reaction. The liberated strand was then purified with Centri-Sep columns (Princeton Separations, CS-900). The products were then PCR amplified with 10 cycles. The products were then purified with MinElute® PCR Purification Kit for further digestion. In the PCR tubes were added 10 µl PCR amplified dsDNA, 5 µl 10× CutSmart® Buffer (New England Biolabs, R0542S), 10 U BcoDI (New England Biolabs, R0542S) and 34 µl water. The mix was then incubated at 37 °C overnight. The digestion products were purified with Centri-Sep columns (Princeton Separations, CS-900). PAGE analysis was then performed (15% Mini-PROTEAN® TBE-Urea Gel) at 55 °C for 90 minutes.

Supporting Results

T4 DNA-ligase catalyzed polymerization of tetranucleotides

The efficiency of polymerization decreased dramatically upon moving from a 64-codon unmodified tetranucleotide system to a 64-codon modified tetranucleotide system (Fig. S1). Upon evaluation of pentanucleotide codon systems, it was clear that larger sequence space and a greater number of modifications could be accommodated using a pentanucleotide codon set.

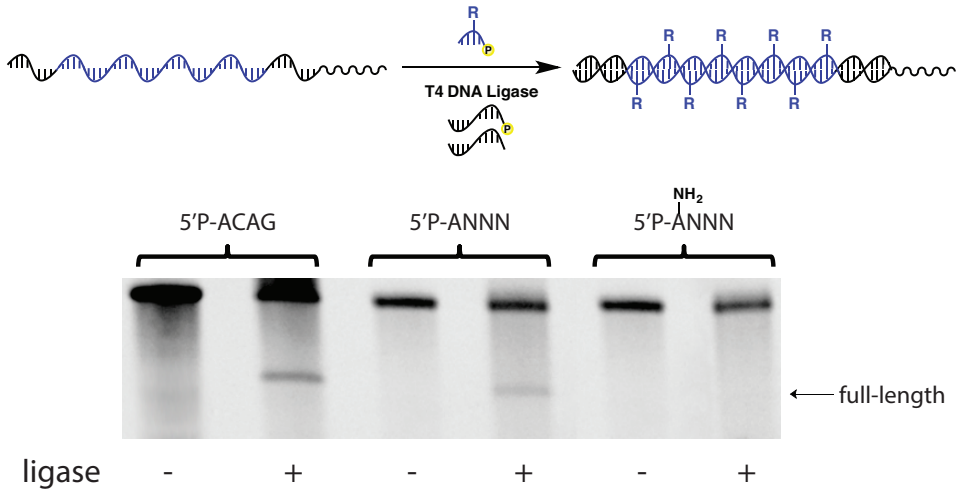


Fig. S1. Evaluation of tetramer libraries as substrates in the T4-DNA ligase-catalyzed polymerization of oligonucleotides.

Restriction enzyme digestion of polymerized product

After identifying that T4 DNA ligase was able to efficiently polymerize large libraries of pentanucleotides along a corresponding DNA template, we undertook preliminary studies to assess the fidelity of the process. We designed a DNA template that contained a BcoDI digest site, which would allow us to evaluate the fidelity of polymerization around the digestion site (Fig. S2). T4 DNA ligase was challenged to correctly incorporate the correct pentanucleotides from the 5'P-ANNNN library along the DNA template. As a positive control, Klenow fragment was used to generate the correct DNA duplex prior to restriction digest. We found that the product of the T4 DNA ligase-catalyzed polymerization reaction was efficiently cleaved by BcoI, suggesting that highly specific incorporation of pentanucleotides was occurring at the digest site.

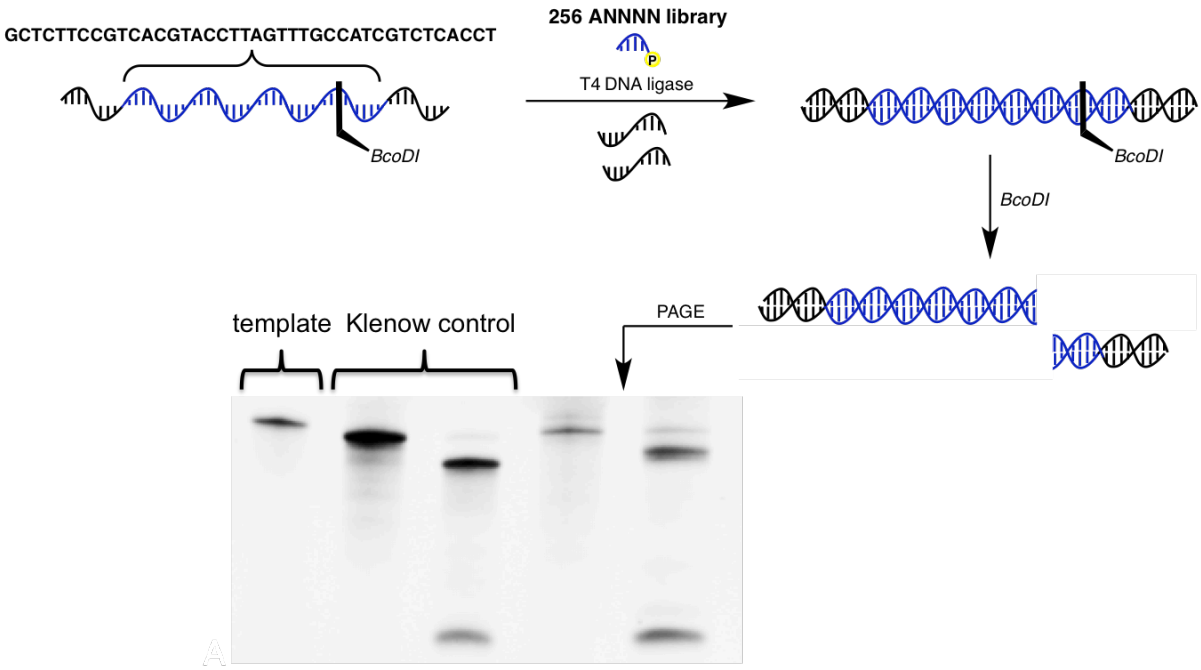


Fig. S2. Evaluation of sequence specificity by restriction enzyme digestion of synthesized duplex product.

Sanger sequencing of polymerization product

The high fidelity of T4 DNA ligase was further supported by Sanger sequencing of the polymerized strand following PAGE purification, which revealed a consensus sequence that was the reverse complement of the starting template strand (Figure S3). These findings suggest that the polymerization proceeds with high fidelity.

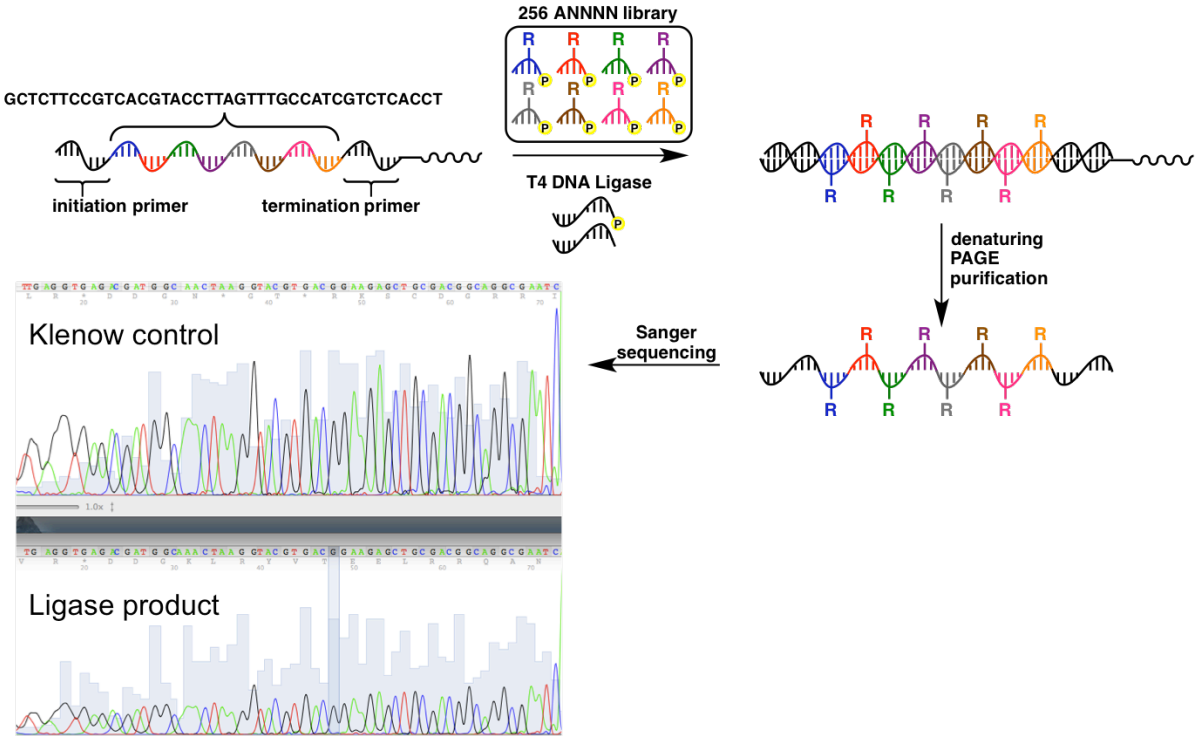


Figure S3. Sanger sequencing workflow for analyzing fidelity of polymerization.

Influence of modification site on polymerization efficiency

Using a 3'-hairpin with 5' primer binding site, the influence of nucleobase modification position on the polymerization efficiency was examined. For this experiment, the modification was the hexylamine on the dA nucleobase. The result in Figure S4 show that polymerization efficiency decreases sharply as the modification position moves from the 5'-end to the 3'-end.

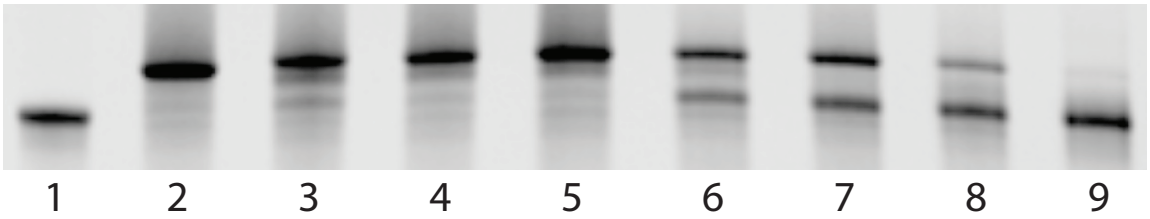


Figure S4. Influence of modification position. Lane 1: template; Lane 2: four equivalents/codon ANNNN unmodified control; Lane 3: four equivalents/codon A(NH_2)NNNN; Lane 4: eight equivalents/codon A(NH_2)NNNN; Lane 5: ten equivalents/codon A(NH_2)NNNN; Lane 6: four equivalents/codon NA(NH_2)NNN; Lane 7: four equivalents/codon NNA(NH_2)NN; Lane 8: four equivalents/codon NNNA(NH_2)N; Lane 9: four equivalents/codon NNNNA(NH_2)

Duplex DNA sequencing data

Note: Data is shown for 35 amol sequencing experiments of unmodified, aminohexylamine modified, and *N*-hexylacetamide modified pentanucleotides from the codon set NTNNN. Due to limited space, codon identities of misincorporation sites have been removed. All fastq files and processed sequencing data (.xls format) for all sequencing experiments are available upon request. Red colour indicates low fidelity pentanucleotides, while blue colour indicates high fidelity.

5'P-NNNAN – 35 amol PCR

pentamer	codon	observed	errors	error rate
CGGAA	TTCCG	6618	621	9.383499547
CCTAT	ATAGG	7342	1345	18.31925906
CCTAA	TTAGG	10321	1507	14.60129832
CCTAC	GTAGG	9743	1508	15.47777892
CCTAG	CTAGG	7082	712	10.05365716
CGCAC	GTGCG	8996	1732	19.25300133
GATAG	CTATC	5734	489	8.52807813
CGCAA	TTGCG	8891	1454	16.35361602
CGCAG	CTGCG	6166	934	15.14758352
GATAC	GTATC	7445	792	10.63801209
GATAA	TTATC	9592	1223	12.75020851
GATAT	ATATC	6819	1075	15.76477489
CGCAT	ATGCG	6805	1454	21.36664217
AGCAA	TTGCT	9291	2445	26.31578947
TTCAT	ATGAA	5734	1188	20.7185211
AGCAC	GTGCT	8316	2129	25.6012506
CCAAT	ATTGG	9320	2189	23.48712446
AGCAG	CTGCT	6164	1386	22.48539909
CCAAA	TTTGG	12906	2669	20.68030373
CCAAC	GTTGG	12171	2589	21.27187577
TTCAA	TTGAA	8289	1082	13.05344432
AGCAT	ATGCT	6604	2118	32.07147184
CCAAG	CTTGG	8709	1409	16.17866575
GAAAG	CTTTC	7909	745	9.419648502
GAAAC	GTTTC	10028	1142	11.38811328
GAAAA	TTTTC	12768	1674	13.11090226
AACAG	CTGTT	9140	1180	12.91028446
TATAT	ATATA	7958	1806	22.69414426
AACAC	GTGTT	12300	2196	17.85365854
AACAA	TTGTT	15059	2645	17.56424729
GAAAT	ATTTC	9255	1281	13.84116694

TATAG	CTATA	6658	685	10.28837489
AACAT	ATGTT	10965	2358	21.50478796
TATAA	TTATA	10975	1561	14.22323462
TATAC	GTATA	8656	1290	14.90295749
ACTAT	ATAGT	8443	1643	19.45990762
ACTAC	GTAGT	9372	1447	15.43960734
ACTAA	TTAGT	11463	1855	16.18250022
ACTAG	CTAGT	7186	992	13.80462009
CACAG	CTGTG	8675	1273	14.67435159
CACAA	TTGTG	13043	2727	20.90776662
CACAC	GTGTG	12175	2328	19.1211499
ATGAT	ATCAT	6960	1357	19.49712644
ATGAA	TTCAT	10040	1646	16.39442231
ATGAC	GTCAT	7963	1187	14.9064423
ATGAG	CTCAT	6185	680	10.99434115
CGTAT	ATACG	5707	951	16.66374628
CGTAG	CTACG	5015	629	12.54237288
CGTAA	TTACG	7854	930	11.84110008
CGTAC	GTACG	7047	921	13.06939123
GTAAT	ATTAC	7079	1362	19.24000565
GTAAG	CTTAC	6127	771	12.58364616
GTAAA	TTTAC	10019	1711	17.07755265
GTAAC	GTTAC	7845	1245	15.86998088
AGGAG	CTCCT	4829	736	15.24125078
AGGAA	TTCCT	7208	1381	19.15926748
AGGAC	GTCCT	6298	1195	18.97427755
TCTAC	GTAGA	6563	1126	17.15678805
TCTAA	TTAGA	7835	1055	13.46522017
TCTAG	CTAGA	5027	487	9.687686493
AGGAT	ATCCT	5341	1301	24.35873432
TCCAG	CTGGA	5766	999	17.32570239
CTCAG	CTGAG	6046	638	10.55243136
ACAAC	GTTGT	12414	2137	17.21443531
ACAAA	TTTGT	15684	2882	18.37541444
ACAAG	CTTGT	9767	1524	15.60356302
GTTAA	TTAAC	7753	955	12.31781246
GTTAC	GTAAC	6202	665	10.72234763
GTTAG	CTAAC	4708	258	5.480033985
ACAAT	ATTGT	11011	2217	20.13441104
GTTAT	ATAAC	5478	832	15.18802483
TTCAG	CTGAA	5180	491	9.478764479
TTCAC	GTGAA	7023	1104	15.71977787

AAGAC	GTCTT	9608	1250	13.00999167
CGAAG	CTTCG	6460	1190	18.42105263
AAGAG	CTCTT	7363	762	10.34904251
CGAAC	GTTCG	8668	1609	18.56252884
CGAAT	ATTCG	6957	1485	21.3454075
AAGAT	ATCTT	8498	1442	16.96869852
TCTAT	ATAGA	5558	1198	21.55451601
TACAT	ATGTA	8084	2231	27.5977239
TGTAG	CTACA	4504	640	14.20959147
TGTAC	GTACA	6004	1112	18.52098601
TGTAA	TTACA	7396	1275	17.23904813
TACAC	GTGTA	9484	1930	20.35006326
TACAA	TTGTA	11205	1995	17.80455154
TACAG	CTGTA	7098	927	13.06001691
TGTAT	ATACA	5266	1288	24.45879225
GCGAG	CTCGC	4857	227	4.673666873
CATAC	GTATG	9806	1442	14.70528248
CATAA	TTATG	11637	1958	16.82564235
GCGAC	GTCGC	6264	619	9.881864623
CATAG	CTATG	7338	788	10.73862088
GCGAA	TTCGC	6398	537	8.39324789
GCGAT	ATCGC	5005	518	10.34965035
CATAT	ATATG	8573	1734	20.22629185
TCCAT	ATGGA	6392	1755	27.45619524
CTCAT	ATGAG	6712	1470	21.90107271
ACGAT	ATCGT	7211	1384	19.19289974
CTCAC	GTGAG	8596	1358	15.7980456
CTCAA	TTGAG	9211	1459	15.83975681
TCCAA	TTGGA	8576	1562	18.2136194
TCCAC	GTGGA	7968	1855	23.28062249
ACGAA	TTCGT	9581	1655	17.27377101
ACGAC	GTCGT	8446	1434	16.97845134
ACGAG	CTCGT	6463	906	14.01825778
GGCAT	ATGCC	4982	902	18.10517864
GGCAG	CTGCC	4514	349	7.731501994
GGCAC	GTGCC	6718	1145	17.04376302
GGCAA	TTGCC	6677	960	14.37771454
TTTAC	GTAAA	7095	761	10.72586328
TTTAA	TTAAA	9119	805	8.827722338
TTTAG	CTAAA	5345	311	5.818521983
TTTAT	ATAAA	6237	1106	17.7328844
GACAT	ATGTC	7546	1577	20.89848927

AAAAT	ATTTT	14761	2228	15.09382833
GACAA	TTGTC	10120	1845	18.2312253
GACAC	GTGTC	9068	1411	15.56021173
GACAG	CTGTC	6479	784	12.10063281
AAAAA	TTTTT	20472	2545	12.43161391
AAAAC	GTTTT	15211	1719	11.30103215
AAAAG	CTTTT	12273	1118	9.109427198
AGAAT	ATTCT	8381	2182	26.03507935
CTAAG	CTTAG	7422	1049	14.1336567
CTAAA	TTTAG	12080	2327	19.26324503
CTAAC	GTTAG	10057	1811	18.00735806
CTAAT	ATTAG	8506	1939	22.79567364
AGAAG	CTTCT	7424	1477	19.89493534
GTGAT	ATCAC	4696	677	14.4165247
ATAAC	GTTAT	11898	2041	17.15414355
TGAAT	ATTCA	6984	1810	25.9163803
ATAAA	TTTAT	16069	3204	19.93901301
ATAAG	CTTAT	9405	1366	14.52418926
TGAAG	CTTCA	6098	981	16.08724172
TGAAC	GTTCA	7606	1684	22.14041546
ATAAT	ATTAT	11059	2289	20.69807397
TGAAA	TTTCA	9842	2037	20.6970128
GCCAC	GTGGC	9528	1654	17.35936188
GCCAA	TTGGC	9202	1512	16.43121061
GCCAG	CTGGC	6720	829	12.33630952
GCCAT	ATGGC	6888	1325	19.23635308
GTGAC	GTCAC	5911	512	8.661816951
GTGAA	TTCAC	6770	730	10.78286558
GTGAG	CTCAC	4364	170	3.895508708
TTAAT	ATTAA	8178	1658	20.2739056
AGAAC	GTTCT	9087	1953	21.49224166
AGAAA	TTTCT	11811	2706	22.91084582
TTAAC	GTTAA	9047	1247	13.78357467
TTAAA	TTTAA	12017	1469	12.22434884
TTAAG	CTTAA	6965	602	8.64321608
CTGAT	ATCAG	5904	1024	17.34417344
ACCAT	ATGGT	9421	2328	24.71075257
TCGAC	GTCGA	5677	992	17.47401797
TCGAG	CTCGA	4519	512	11.32994025
CTGAG	CTCAG	5407	316	5.844275939
CTGAC	GTCAG	7505	836	11.13924051
CTGAA	TTCAG	7983	1016	12.72704497

ACCAG	CTGGT	8726	1596	18.29016732
TCGAT	ATCGA	4709	1096	23.27458059
ACCAC	GTGGT	11899	2616	21.98504076
TAGAT	ATCTA	6148	1508	24.52830189
TAGAG	CTCTA	5387	674	12.511602
TAGAC	GTCTA	7169	1274	17.77095829
TAGAA	TTCTA	8709	1315	15.09932254
ACCAA	TTGGT	12899	2826	21.90867509
TGGAG	CTCCA	3866	632	16.34764615
TGGAC	GTCCA	5317	1049	19.72917058
TGGAT	ATCCA	4137	1127	27.24196277
CCCAG	CTGGG	9427	1651	17.51352498
GAGAT	ATCTC	5492	731	13.31026948
CGGAT	ATCCG	4793	647	13.49885249
CCCAC	GTGGG	13548	3637	26.84529082
CCCAA	TTGGG	12582	2740	21.77714195
CGGAG	CTCCG	4599	391	8.501848228
GAGAG	CTCTC	4960	431	8.689516129
CCCAT	ATGGG	9603	2322	24.17994377
CGGAC	GTCCG	6656	740	11.11778846
GAGAC	GTCTC	6825	672	9.846153846
GGGAT	ATCCC	3691	413	11.18937957
GGGAC	GTCCC	4904	436	8.890701468
GGGAA	TTCCC	5095	416	8.164867517
GGGAG	CTCCC	3602	139	3.85896724
TCGAA	TTCGA	6431	902	14.02581247
AAGAA	TTCTT	11776	1682	14.28328804
CGAAA	TTTCG	9707	1787	18.40939528
TAAAG	CTTTA	9085	952	10.47881123
TAAAA	TTTTA	15354	1999	13.01940862
TAAAC	GTTTA	11454	1749	15.26977475
TAAAT	ATTTA	10722	2356	21.9735124
GCTAG	CTAGC	5170	425	8.220502901
GCTAA	TTAGC	7697	886	11.5109783
GCTAC	GTAGC	6935	805	11.60778659
GCTAT	ATAGC	5708	842	14.75122635
AGTAT	ATACT	6238	1362	21.83392113
CAAAT	ATTTG	10883	2079	19.10318846
CAAAC	GTTTG	12681	1995	15.73219778
AGTAG	CTACT	5511	781	14.17165669
CAAAA	TTTTG	15191	2570	16.91791192
CAAAG	CTTTG	9755	1203	12.33213737

AGTAC	GTACT	7135	1281	17.95374912
AGTAA	TTACT	8988	1675	18.63595906
GGAAA	TTTCC	7522	1516	20.1542143
GGAAC	GTTCC	5908	1199	20.29451591
GGAAG	CTTCC	4904	861	17.55709625
GGAAT	ATTCC	5310	1070	20.15065913
ATTAT	ATAAT	8270	1574	19.03264813
ATTAC	GTAAT	8827	1259	14.26305653
ATTAA	TTAAT	12054	1957	16.2352746
ATTAG	CTAAT	6869	721	10.49643325
AATAA	TTATT	14945	2145	14.3526263
AATAC	GTATT	11263	1358	12.05717837
TTGAT	ATCAA	5169	949	18.35945057
AATAG	CTATT	8725	838	9.604584527
TTGAA	TTCAA	7505	734	9.780146569
TTGAC	GTCAA	6262	706	11.27435324
AATAT	ATATT	10679	1920	17.97921154
TTGAG	CTCAA	4720	273	5.783898305
GTCAG	CTGAC	4808	446	9.276206323
GTCAC	GTGAC	6674	993	14.8786335
GTCAA	TTGAC	7372	1120	15.19262073
CCGAT	ATCGG	6406	1053	16.43771464
GTCAT	ATGAC	5329	1045	19.60968287
CCGAC	GTCGG	8901	1294	14.53769239
CCGAA	TTCGG	8387	1028	12.2570645
CCGAG	CTCGG	6168	556	9.014267185
CACAT	ATGTG	9475	2267	23.92612137
CTTAT	ATAAG	6797	1205	17.72840959
CTTAG	CTAAG	5884	406	6.900067981
CTTAA	TTAAG	9312	1191	12.78994845
CTTAC	GTAAG	8019	967	12.05886021
CAGAA	TTCTG	9414	1469	15.60441895
GGTAT	ATACC	4359	609	13.97109429
GGTAA	TTACC	6116	702	11.47809026
GGTAC	GTACC	5126	606	11.8220835
GGTAG	CTACC	3819	236	6.179628175
TGCAA	TTGCA	7087	1683	23.74770707
TGCAC	GTGCA	6750	1601	23.71851852
TGCAG	CTGCA	4787	891	18.61290996
TGGAA	TTCCA	5835	939	16.09254499
TGCAT	ATGCA	5426	1670	30.77773682
ATCAT	ATGAT	7421	1730	23.31222207

ATCAG	CTGAT	6695	931	13.90589993
ATCAA	TTGAT	10733	2147	20.00372682
ATCAC	GTGAT	8579	1604	18.69681781
GCAAT	ATTGC	7180	1279	17.81337047
GCAAG	CTTGC	6827	822	12.04042771
GCAAA	TTTGC	9882	1563	15.81663631
GCAAC	GTTGC	8627	1299	15.057378
CAGAT	ATCTG	6934	1214	17.50793193
TCAAT	ATTGA	7235	1750	24.18797512
CAGAC	GTCTG	8739	1248	14.28081016
CAGAG	CTCTG	6392	708	11.07634543
TCAAC	GTTGA	8473	1536	18.12817184
TCAAA	TTTGA	10298	1629	15.81860555
TCAAG	CTTGA	6718	901	13.41172968
GAGAA	TTCTC	7643	908	11.88015177
		2058018	335813	
	ERROR RATE	**=**	**16.3173014**	

5'P-NNNA(NH2)N – 35 amol PCR

pentamer	codon	observed	errors	error rate
GAGAC	GTCTC	559	7	1.252236136
CCTAT	ATAGG	1131	15	1.326259947
CCTAA	TTAGG	1792	25	1.395089286
CCTAC	GTAGG	1383	21	1.518438178
CCTAG	CTAGG	973	15	1.541623844
CGCAC	GTGCG	1592	25	1.570351759
GATAG	CTATC	819	8	0.976800977
CGCAA	TTGCG	1478	21	1.420838972
CGCAG	CTGCG	813	9	1.10701107
GATAC	GTATC	1194	14	1.172529313
GATAA	TTATC	1488	23	1.545698925
GATAT	ATATC	1227	9	0.733496333
CGCAT	ATGCG	1271	13	1.02281668
AGCAA	TTGCT	1372	29	2.113702624
TTCAT	ATGAA	441	20	4.535147392
AGCAC	GTGCT	1327	42	3.165033911
CCAAT	ATTGG	1506	40	2.656042497
AGCAG	CTGCT	694	23	3.314121037
CCAAA	TTTGG	1937	25	1.290655653
CCAAC	GTTGG	1961	39	1.988781234
TTCAA	TTGAA	938	11	1.172707889
AGCAT	ATGCT	1100	34	3.090909091
CCAAG	CTTGG	1012	29	2.865612648
GAAAG	CTTTC	589	8	1.358234295
GAAAC	GTTTC	1185	20	1.687763713
GAAAA	TTTTC	1164	16	1.374570447
AACAG	CTGTT	871	18	2.066590126
TATAT	ATATA	299	9	3.010033445
AACAC	GTGTT	1653	43	2.601330913
AACAA	TTGTT	1778	26	1.46231721
GAAAT	ATTTC	1070	17	1.588785047
TATAG	CTATA	430	5	1.162790698
AACAT	ATGTT	1452	35	2.41046832
TATAA	TTATA	871	8	0.918484501
TATAC	GTATA	436	6	1.376146789
ACTAT	ATAGT	802	19	2.369077307
ACTAC	GTAGT	930	24	2.580645161
ACTAA	TTAGT	1330	25	1.879699248
ACTAG	CTAGT	654	17	2.599388379
CACAT	ATGTG	2090	51	2.440191388
CACAG	CTGTG	1231	26	2.112103981
CACAA	TTGTG	2305	39	1.69197397

CACAC	GTGTG	2480	47	1.89516129
ATGAT	ATCAT	668	9	1.347305389
ATGAA	TTCAT	959	22	2.294056309
ATGAC	GTCAT	870	25	2.873563218
ATGAG	CTCAT	528	9	1.704545455
CGTAT	ATACG	859	14	1.629802095
CGTAG	CTACG	681	10	1.468428781
CGTAA	TTACG	1266	22	1.737756714
CGTAC	GTACG	1038	16	1.541425819
GTAAT	ATTAC	1242	19	1.52979066
GTAAA	TTTAC	1489	34	2.283411686
GTAAC	GTTAC	1431	30	2.096436059
GTAAG	CTTAC	804	16	1.990049751
AGGAA	TTCCT	583	6	1.02915952
AGGAC	GTCCT	585	17	2.905982906
TCTAC	GTAGA	289	18	6.228373702
TCTAA	TTAGA	631	9	1.426307448
TCTAG	CTAGA	321	6	1.869158879
AGGAT	ATCCT	392	16	4.081632653
TCCAG	CTGGA	547	14	2.559414991
TCTAT	ATAGA	234	11	4.700854701
ACAAC	GTTGT	1266	20	1.579778831
ACAAA	TTTGT	1498	49	3.271028037
ACAAG	CTTGT	671	29	4.321907601
GTTAA	TTAAC	1577	17	1.077996195
GTTAC	GTAAC	1188	9	0.757575758
GGTAG	CTACC	503	15	2.982107356
AGGAG	CTCCT	317	6	1.892744479
GTTAG	CTAAC	879	6	0.682593857
ACAAT	ATTGT	1017	42	4.12979351
GTTAT	ATAAC	1208	17	1.407284768
TTCAG	CTGAA	526	7	1.330798479
TTCAC	GTGAA	601	22	3.660565724
AAGAC	GTCTT	487	16	3.285420945
CGAAG	CTTCG	625	16	2.56
CGAAA	TTTCG	1291	16	1.239349342
CGAAC	GTTCG	1326	26	1.960784314
CGAAT	ATTCG	974	16	1.642710472
TCCAA	TTGGA	949	32	3.371970495
AAGAT	ATCTT	334	9	2.694610778
TACAT	ATGTA	484	14	2.892561983
TGTAG	CTACA	442	9	2.036199095
TGTAC	GTACA	504	14	2.777777778
TGTAA	TTACA	715	8	1.118881119

TACAC	GTGTA	690	25	3.623188406
TACAA	TTGTA	991	18	1.816347124
TACAG	CTGTA	502	10	1.992031873
TGTAT	ATACA	429	13	3.03030303
GCGAG	CTCGC	267	2	0.74906367
CATAC	GTATG	1830	27	1.475409836
CATAA	TTATG	2128	27	1.268796992
GCGAC	GTCGC	477	10	2.096436059
CATAG	CTATG	1040	7	0.673076923
GCGAA	TTCGC	529	9	1.701323251
GCGAT	ATCGC	382	11	2.879581152
CATAT	ATATG	1640	18	1.097560976
TCCAT	ATGGA	385	36	9.350649351
CTCAT	ATGAG	1610	28	1.739130435
ACGAT	ATCGT	338	15	4.437869822
CTCAC	GTGAG	1817	31	1.706108971
CTCAA	TTGAG	1990	30	1.507537688
CTCAG	CTGAG	1208	15	1.241721854
TCCAC	GTGGA	645	69	10.69767442
ACGAA	TTCGT	553	13	2.350813743
ACGAC	GTCGT	405	20	4.938271605
ACGAG	CTCGT	209	11	5.263157895
GGCAT	ATGCC	920	13	1.413043478
GGCAG	CTGCC	706	17	2.407932011
GGCAC	GTGCC	1076	18	1.672862454
GGCAA	TTGCC	1168	25	2.140410959
TTTAC	GTAAA	337	4	1.18694362
TTTAA	TTAAA	692	6	0.867052023
TTTAG	CTAAA	350	1	0.285714286
TTTAT	ATAAA	297	5	1.683501684
GACAT	ATGTC	1415	28	1.978798587
AAAAT	ATTTT	928	18	1.939655172
GACAA	TTGTC	1716	42	2.447552448
GACAC	GTGTC	1707	33	1.933216169
GACAG	CTGTC	978	20	2.044989775
AAAAA	TTTTT	1245	21	1.686746988
AAAAC	GTTTT	1105	15	1.357466063
AAAAG	CTTTT	540	15	2.777777778
GTGAA	TTCAC	875	11	1.257142857
CTAAG	CTTAG	960	14	1.458333333
CTAAA	TTTAG	1645	27	1.641337386
CTAAC	GTTAG	1532	23	1.501305483
CTAAT	ATTAG	1347	28	2.078693393
AGAAA	TTTCT	1091	22	2.016498625

ATAAC	GTTAT	1555	16	1.028938907
TGAAT	ATTCA	549	13	2.367941712
ATAAA	TTTAT	1896	22	1.160337553
ATAAG	CTTAT	962	21	2.182952183
TGAAG	CTTCA	489	4	0.81799591
TGAAC	GTTCA	774	11	1.42118863
ATAAT	ATTAT	1363	25	1.834189288
TGAAA	TTTCA	989	18	1.820020222
GCCAC	GTGGC	1795	19	1.058495822
GCCAA	TTGGC	1962	49	2.49745158
GCCAG	CTGGC	1055	28	2.654028436
GCCAT	ATGGC	1379	20	1.450326323
GTGAC	GTCAC	874	10	1.14416476
AGAAT	ATTCT	848	20	2.358490566
GTGAG	CTCAC	515	9	1.747572816
TTAAT	ATTAA	302	6	1.986754967
AGAAG	CTTCT	471	7	1.486199575
AGAAC	GTTCT	1023	14	1.368523949
GTGAT	ATCAC	780	7	0.897435897
TCGAA	TTCGA	268	4	1.492537313
TTAAC	GTTAA	436	9	2.064220183
TTAAA	TTTAA	932	20	2.145922747
TTAAG	CTTAA	380	2	0.526315789
CTGAT	ATCAG	622	6	0.964630225
ACCAT	ATGGT	1193	47	3.939647946
TCGAC	GTCGA	168	12	7.142857143
TCGAG	CTCGA	162	5	3.086419753
CTGAG	CTCAG	498	6	1.204819277
CTGAC	GTCAG	855	12	1.403508772
CTGAA	TTCAG	926	9	0.971922246
ACCAG	CTGGT	905	19	2.099447514
TAGAT	ATCTA	114	9	7.894736842
ATCAA	TTGAT	1923	21	1.092043682
TAGAG	CTCTA	134	3	2.23880597
TAGAC	GTCTA	162	7	4.320987654
TAGAA	TTCTA	231	3	1.298701299
ACCAA	TTGGT	1771	38	2.145680407
TGGAG	CTCCA	192	0	0
TGGAA	TTCCA	411	9	2.189781022
TGGAC	GTCCA	284	9	3.169014085
ACCAC	GTGGT	1536	49	3.190104167
TGGAT	ATCCA	198	9	4.545454545
CCCAG	CTGGG	1529	27	1.765860039
GAGAT	ATCTC	419	4	0.954653938

CGGAT	ATCCG	450	10	2.222222222
CCCAC	GTGGG	2553	52	2.036819428
CCCAA	TTGGG	2595	40	1.541425819
CGGAG	CTCCG	361	8	2.216066482
GAGAG	CTCTC	303	5	1.650165017
CCCAT	ATGGG	1853	42	2.266594711
CGGAC	GTCCG	648	11	1.697530864
CGGAA	TTCCG	667	10	1.499250375
GGGAT	ATCCC	446	8	1.793721973
GGGAC	GTCCC	478	15	3.138075314
GGGAA	TTCCC	567	13	2.292768959
GGGAG	CTCCC	308	10	3.246753247
AAGAA	TTCTT	666	8	1.201201201
AAGAG	CTCTT	242	3	1.239669421
TAAAG	CTTTA	351	1	0.284900285
TAAAA	TTTTA	877	14	1.596351197
TAAAC	GTTTA	553	13	2.350813743
TAAAT	ATTTA	361	5	1.385041551
GCTAG	CTAGC	664	13	1.957831325
GCTAA	TTAGC	1153	26	2.25498699
GCTAC	GTAGC	881	18	2.043132804
GCTAT	ATAGC	828	16	1.93236715
AGTAT	ATACT	873	21	2.405498282
CAAAT	ATTTG	1573	24	1.52574698
CAAAC	GTTTG	1861	25	1.343363783
AGTAG	CTACT	568	10	1.76056338
CAAAA	TTTTG	1778	26	1.46231721
CAAAG	CTTTG	847	16	1.889020071
AGTAC	GTACT	933	12	1.286173633
AGTAA	TTACT	1198	26	2.170283806
GGAAA	TTTCC	819	27	3.296703297
GGAAC	GTTCC	674	14	2.077151335
GGAAG	CTTCC	452	8	1.769911504
GGAAT	ATTCC	716	16	2.234636872
ATTAT	ATAAT	1381	23	1.665459812
ATTAC	GTAAT	1422	23	1.617440225
ATTAA	TTAAT	1918	29	1.511991658
ATTAG	CTAAT	1023	12	1.173020528
AATAA	TTATT	1774	20	1.127395716
AATAC	GTATT	1350	27	2
TTGAT	ATCAA	149	4	2.684563758
AATAG	CTATT	880	14	1.590909091
TTGAA	TTCAA	350	4	1.142857143
TTGAC	GTCAA	220	2	0.909090909

AATAT	ATATT	1252	15	1.198083067
TTGAG	CTCAA	178	5	2.808988764
GTCAG	CTGAC	1003	17	1.694915254
GTCAC	GTGAC	1617	23	1.422387137
GTCAA	TTGAC	1756	33	1.879271071
CCGAT	ATCGG	524	9	1.717557252
GTCAT	ATGAC	1344	27	2.008928571
CCGAC	GTCGG	641	14	2.184087363
CCGAA	TTCGG	775	11	1.419354839
CCGAG	CTCGG	331	7	2.114803625
CTTAT	ATAAG	1233	14	1.135442011
CTTAG	CTAAG	1005	11	1.094527363
CTTAA	TTAAG	1903	17	0.893326327
CTTAC	GTAAG	1402	13	0.92724679
GGTAT	ATACC	686	12	1.749271137
GGTAA	TTACC	895	14	1.56424581
GGTAC	GTACC	687	10	1.455604076
TGCAA	TTGCA	807	13	1.610904585
TGCAC	GTGCA	720	34	4.722222222
TGCAG	CTGCA	531	8	1.506591337
TGCAT	ATGCA	548	25	4.562043796
ATCAT	ATGAT	1582	31	1.95954488
ATCAG	CTGAT	1072	24	2.23880597
ATCAC	GTGAT	1805	38	2.105263158
GCAAT	ATTGC	1096	24	2.189781022
GCAAG	CTTGC	711	18	2.53164557
GCAAA	TTTGC	1311	23	1.754385965
GCAAC	GTTGC	1325	33	2.490566038
TCGAT	ATCGA	149	12	8.053691275
CAGAT	ATCTG	565	13	2.300884956
TCAAT	ATTGA	385	13	3.376623377
CAGAA	TTCTG	931	21	2.255639098
CAGAC	GTCTG	847	11	1.298701299
CAGAG	CTCTG	444	14	3.153153153
TCAAC	GTTGA	546	19	3.47985348
TCAAA	TTTGA	974	19	1.950718686
TCAAG	CTTGA	503	9	1.789264414
GAGAA	TTCTC	577	13	2.253032929
		240177	4594	
	ERROR RATE	**=**	**1.912756009**	

5’P-NNNA(NHAc)N – 35 amol PCR

pentamer	codon	observed	errors	error rate
TAAAG	CTTTA	288	9	3.125
TTCAC	GTGAA	664	15	2.259036145
GAGAA	TTCTC	510	17	3.333333333
AGTAT	ATACT	824	22	2.669902913
CTAAG	CTTAG	522	13	2.490421456
CCTAT	ATAGG	962	19	1.975051975
CTAAA	TTTAG	1063	24	2.257761054
CAAAT	ATTTG	1462	27	1.846785226
GAGAC	GTCTC	603	8	1.326699834
TCCAG	CTGGA	729	14	1.920438957
CTAAT	ATTAG	934	31	3.319057816
AGTAG	CTACT	508	20	3.937007874
CAAAA	TTTTG	1123	17	1.513802315
CCTAC	GTAGG	1388	29	2.089337176
CAAAG	CTTTG	623	16	2.568218299
AGTAC	GTACT	800	22	2.75
CCTAG	CTAGG	980	13	1.326530612
TAGAC	GTCTA	225	7	3.111111111
CGCAC	GTGCG	1548	26	1.679586563
GATAG	CTATC	567	9	1.587301587
CGCAA	TTGCG	1522	32	2.102496715
CGCAG	CTGCG	776	25	3.221649485
GATAC	GTATC	824	12	1.45631068
GGAAG	CTTCC	344	8	2.325581395
GATAA	TTATC	994	10	1.006036217
GATAT	ATATC	810	14	1.728395062
GGAAT	ATTCC	554	10	1.805054152
CGCAT	ATGCG	1426	27	1.893408135
AGCAA	TTGCT	1072	40	3.731343284
TTCAT	ATGAA	580	21	3.620689655
AGCAC	GTGCT	1206	41	3.399668325
CGAAA	TTTCG	954	20	2.096436059
CCAAT	ATTGG	1426	33	2.314165498
CGAAC	GTTCG	1054	25	2.371916509
ATTAT	ATAAT	906	16	1.766004415
AAGAT	ATCTT	399	10	2.506265664
CCAAA	TTTGG	1615	41	2.53869969
CGAAT	ATTCG	956	23	2.405857741
TTCAG	CTGAA	436	5	1.146788991
ATAAA	TTTAT	1018	43	4.223968566
ATTAG	CTAAT	544	11	2.022058824
AGCAT	ATGCT	1081	39	3.607770583
CCAAG	CTTGG	902	30	3.32594235

AATAA	TTATT	888	21	2.364864865
AGTAA	TTACT	867	24	2.76816609
CTAAC	GTTAG	1041	20	1.921229587
TAGAA	TTCTA	273	3	1.098901099
GAAAG	CTTTC	352	4	1.136363636
ATAAC	GTTAT	1150	34	2.956521739
TGAAT	ATTCA	626	16	2.555910543
AATAC	GTATT	837	20	2.38948626
GAAAC	GTTTC	899	23	2.55839822
ATAAG	CTTAT	560	16	2.857142857
GAAAA	TTTTC	694	21	3.025936599
AATAG	CTATT	509	13	2.554027505
CCCAC	GTGGG	2559	89	3.477921063
TATAT	ATATA	363	11	3.03030303
AACAC	GTGTT	1464	43	2.93715847
AACAA	TTGTT	1371	49	3.574033552
TTGAA	TTCAA	430	9	2.093023256
TGAAG	CTTCA	410	8	1.951219512
TTGAC	GTCAA	380	6	1.578947368
GAAAT	ATTTC	758	20	2.638522427
AATAT	ATATT	857	13	1.516919487
TGAAC	GTTCA	832	16	1.923076923
ATAAT	ATTAT	949	30	3.161222339
TGAAA	TTTCA	772	20	2.590673575
TATAG	CTATA	277	5	1.805054152
TCTAT	ATAGA	346	10	2.89017341
TATAA	TTATA	527	14	2.65654649
TATAC	GTATA	414	9	2.173913043
GCCAC	GTGGC	1782	27	1.515151515
GTCAG	CTGAC	793	9	1.134930643
GCCAA	TTGGC	1891	34	1.797990481
GCCAG	CTGGC	1067	23	2.155576382
GTCAC	GTGAC	1374	25	1.819505095
TACAT	ATGTA	629	23	3.656597774
GTCAA	TTGAC	1463	21	1.435406699
TGTAG	CTACA	465	3	0.64516129
TGTAC	GTACA	639	16	2.503912363
TGTAA	TTACA	706	14	1.983002833
CCGAT	ATCGG	697	9	1.291248207
TACAC	GTGTA	841	25	2.972651605
GTCAT	ATGAC	1070	20	1.869158879
TACAA	TTGTA	892	21	2.35426009
TACAG	CTGTA	475	12	2.526315789
CCTAA	TTAGG	1586	22	1.387137453
CCGAC	GTCGG	988	25	2.530364372

CCGAA	TTCGG	955	18	1.884816754
TGTAT	ATACA	448	6	1.339285714
CCGAG	CTCGG	397	9	2.267002519
GTGAC	GTCAC	961	12	1.248699272
AGAAT	ATTCT	643	23	3.576982893
GTGAG	CTCAC	469	4	0.852878465
GCGAG	CTCGC	331	5	1.510574018
CATAC	GTATG	1307	28	2.142310635
CATAA	TTATG	1398	32	2.288984263
GCGAC	GTCGC	648	11	1.697530864
CATAG	CTATG	809	6	0.741656366
GCGAA	TTCGC	567	9	1.587301587
TTAAT	ATTAA	335	10	2.985074627
AGAAG	CTTCT	309	19	6.148867314
CGTAG	CTACG	698	20	2.865329513
AGAAC	GTTCT	666	26	3.903903904
AGAAA	TTTCT	670	40	5.970149254
TCGAA	TTCGA	470	12	2.553191489
TTAAC	GTTAA	473	15	3.171247357
TTAAA	TTTAA	619	12	1.938610662
GCGAT	ATCGC	468	11	2.35042735
TTAAG	CTTAA	279	6	2.150537634
CATAT	ATATG	1396	24	1.719197708
CTGAT	ATCAG	671	9	1.341281669
TTGAG	CTCAA	218	5	2.293577982
ACCAT	ATGGT	1552	35	2.255154639
TCGAC	GTCGA	389	11	2.827763496
TCGAG	CTCGA	236	3	1.271186441
CTGAG	CTCAG	458	6	1.310043668
CTGAC	GTCAG	979	10	1.02145046
CTGAA	TTCAG	845	14	1.656804734
ACCAG	CTGGT	1055	42	3.981042654
ACCAA	TTGGT	2031	63	3.101920236
ACCAC	GTGGT	1989	68	3.418803419
TCCAT	ATGGA	702	30	4.273504274
TAGAT	ATCTA	157	7	4.458598726
AACAT	ATGTT	1340	33	2.462686567
GGTAT	ATACC	569	6	1.054481547
CTCAT	ATGAG	1115	28	2.511210762
ACGAT	ATCGT	518	21	4.054054054
GGTAG	CTACC	520	4	0.769230769
GGTAA	TTACC	852	18	2.112676056
ACTAA	TTAGT	1231	37	3.005686434
TCCAA	TTGGA	1180	29	2.457627119
TCCAC	GTGGA	1007	48	4.766633565

ACGAA	TTCGT	678	22	3.244837758
ACGAC	GTCGT	756	17	2.248677249
GGAAC	GTTCC	449	15	3.340757238
ACGAG	CTCGT	312	15	4.807692308
TGCAA	TTGCA	747	20	2.677376171
TGCAC	GTGCA	890	31	3.483146067
TGCAG	CTGCA	607	17	2.800658979
CTTAT	ATAAG	822	7	0.851581509
GGCAT	ATGCC	826	23	2.784503632
CTCAG	CTGAG	788	18	2.284263959
TGCAT	ATGCA	756	24	3.174603175
GGCAG	CTGCC	542	9	1.660516605
GGCAC	GTGCC	1058	37	3.497164461
CTTAG	CTAAG	614	6	0.977198697
GGCAA	TTGCC	1071	15	1.400560224
CACAT	ATGTG	2012	43	2.137176938
AAGAA	TTCTT	518	20	3.861003861
CTTAA	TTAAG	1052	23	2.186311787
TGGAG	CTCCA	268	2	0.746268657
TGGAA	TTCCA	472	6	1.271186441
TGGAC	GTCCA	458	8	1.746724891
CACAG	CTGTG	1091	36	3.299725023
CACAA	TTGTG	2055	63	3.065693431
CTTAC	GTAAG	809	20	2.472187886
CACAC	GTGTG	2132	41	1.923076923
TGGAT	ATCCA	313	12	3.833865815
GACAC	GTGTC	1478	31	2.097428958
CCCAG	CTGGG	1444	53	3.670360111
GAGAT	ATCTC	431	9	2.088167053
CGGAT	ATCCG	629	13	2.066772655
ATGAT	ATCAT	790	15	1.898734177
CCCAA	TTGGG	2727	54	1.98019802
TTTAC	GTAAA	316	12	3.797468354
CAAAC	GTTTG	1340	35	2.611940299
TTTAA	TTAAA	493	8	1.622718053
TTTAG	CTAAA	270	7	2.592592593
CGGAG	CTCCG	325	6	1.846153846
ATGAA	TTCAT	938	22	2.345415778
GAGAG	CTCTC	283	9	3.180212014
ATGAC	GTCAT	1075	20	1.860465116
CGGAC	GTCCG	716	18	2.51396648
CGGAA	TTCCG	638	14	2.194357367
ATGAG	CTCAT	507	13	2.564102564
TTTAT	ATAAA	313	5	1.597444089
CGTAT	ATACG	1034	12	1.160541586

GACAT	ATGTC	1218	20	1.642036125
GGGAT	ATCCC	450	15	3.333333333
ATCAT	ATGAT	1412	26	1.841359773
AAAAT	ATTTT	702	18	2.564102564
GACAA	TTGTC	1399	27	1.929949964
TTGAT	ATCAA	247	11	4.453441296
GGGAA	TTCCC	499	11	2.204408818
GGGAC	GTCCC	520	14	2.692307692
GGGAG	CTCCC	294	6	2.040816327
CGTAA	TTACG	1270	29	2.283464567
GACAG	CTGTC	718	16	2.228412256
CGTAC	GTACG	1155	13	1.125541126
AAAAA	TTTTT	602	27	4.485049834
AACAG	CTGTT	695	22	3.165467626
AAAAC	GTTTT	761	14	1.839684625
ATCAG	CTGAT	768	28	3.645833333
ATCAA	TTGAT	1413	33	2.335456476
AAAAG	CTTTT	332	9	2.710843373
ATCAC	GTGAT	1461	36	2.464065708
AAGAC	GTCTT	496	11	2.217741935
ACTAT	ATAGT	861	17	1.974448316
GTAAT	ATTAC	950	30	3.157894737
GCAAT	ATTGC	969	16	1.651186791
GCCAT	ATGGC	1278	28	2.190923318
CGAAG	CTTCG	454	17	3.744493392
TCAAG	CTTGA	560	9	1.607142857
GTAAA	TTTAC	959	25	2.606882169
GTAAC	GTTAC	1080	33	3.055555556
GTAAG	CTTAC	552	13	2.355072464
AAGAG	CTCTT	230	2	0.869565217
GCAAG	CTTGC	613	19	3.099510604
GCAAA	TTTGC	1143	36	3.149606299
GCAAC	GTTGC	1249	25	2.001601281
AGGAG	CTCCT	232	14	6.034482759
AGGAA	TTCCT	482	18	3.734439834
AGCAG	CTGCT	542	23	4.243542435
AGGAC	GTCCT	613	12	1.957585644
CTCAC	GTGAG	1343	39	2.903946389
TCTAC	GTAGA	458	18	3.930131004
TCTAA	TTAGA	644	11	1.708074534
TAGAG	CTCTA	139	0	0
TCTAG	CTAGA	395	5	1.265822785
TAAAA	TTTTA	631	20	3.169572108
TAAAC	GTTTA	552	15	2.717391304
AGGAT	ATCCT	452	13	2.876106195

CTCAA	TTGAG	1453	39	2.684101858
TCGAT	ATCGA	255	10	3.921568627
TAAAT	ATTTA	457	14	3.06345733
ACTAG	CTAGT	649	30	4.622496148
ACAAC	GTTGT	1433	42	2.930914166
ACAAA	TTTGT	1320	44	3.333333333
ACAAG	CTTGT	690	28	4.057971014
CAGAT	ATCTG	904	16	1.769911504
GTTAA	TTAAC	1008	18	1.785714286
GTTAC	GTAAC	792	6	0.757575758
GCTAG	CTAGC	597	13	2.177554439
GCTAA	TTAGC	1029	13	1.263362488
GTTAG	CTAAC	589	7	1.188455008
GCTAC	GTAGC	774	19	2.454780362
CAGAA	TTCTG	934	16	1.713062099
CAGAC	GTCTG	1072	19	1.77238806
GTGAA	TTCAC	867	14	1.614763552
ATTAA	TTAAT	997	23	2.306920762
GGTAC	GTACC	595	9	1.512605042
CAGAG	CTCTG	455	10	2.197802198
ACAAT	ATTGT	1185	42	3.544303797
GCTAT	ATAGC	557	9	1.615798923
TCAAC	GTTGA	869	31	3.567318757
TCAAA	TTTGA	998	18	1.803607214
GTTAT	ATAAC	681	7	1.027900147
ATTAC	GTAAT	771	22	2.853437095
GGAAA	TTTCC	576	16	2.777777778
TCAAT	ATTGA	541	17	3.14232902
CCAAC	GTTGG	1931	62	3.210771621
CCCAT	ATGGG	2154	52	2.414113278
ACTAC	GTAGT	1034	24	2.321083172
TTCAA	TTGAA	849	22	2.591283863
GTGAT	ATCAC	690	11	1.594202899
		212757	**5148**	

ERROR RATE = 2.441537732

Codon frequency analysis

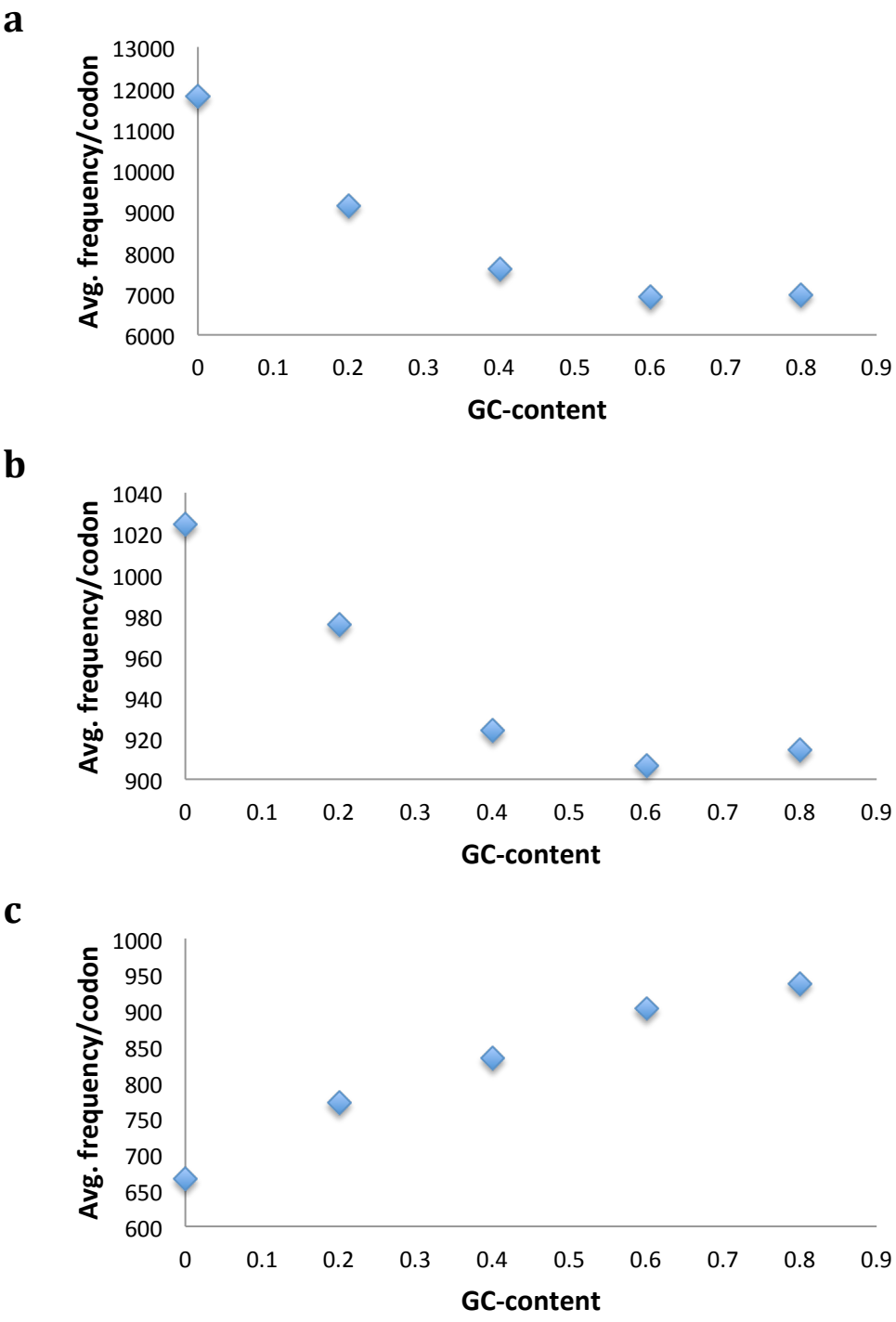


Figure S5. Influence of GC-content on codon bias during pentanucleotide polymerization using a) NNNAN b) NNNA(NH_2)N and c) NNNA(NHAc)N.