Published in final edited form as:

Org Biomol Chem. 2017 March 21; 15(11): 2349–2352. doi:10.1039/c6ob02792j.

Structure-Activity Relationships of the ATP cofactor in Ligase-Catalysed Oligonucleotide Polymerisations

Yi Lei and Ryan Hili^a

^aDepartment of Chemistry, University of Georgia, Athens, GA 30602, USA

Abstract

A T4 DNA ligase-catalysed oligonucleotide polymerisation process has been recently developed to enable the incorporation of multiple functional groups throughout a nucleic acid polymer. T4 DNA ligase requires ATP as a cofactor to catalyse phosphodiester bond formation during the polymer process. Herein, we describe the structure-activity relationship of ATP within the context of T4 DNA ligase-catalysed oligonucleotide polymerisation. Using high-throughput sequencing, we study not only the influence of ATP modification on polymerisation efficiency, but also on the fidelity and sequence bias of the polymerisation process.

Aptamers are sequence-defined, mono-dispersed nucleic acid polymers that can bind molecular targets.¹ The molecular recognition properties of this class of polymer can be evolved and tuned using *in vitro* selection technologies.^{2–3} Aptamers with high affinities and selectivities have been evolved for myriad targets, ranging from metal ions⁴ to mammalian cells.⁵ Notwithstanding their success as high affinity reagents, the dearth of chemical functionality present in natural nucleic acid polymers has limited their potential. This contrasts with the diverse side chains present in proteinogenic polymers. To address the functional group deficit of aptamers, researchers have explored ways to sequence-specifically incorporate diverse functionality either through nucleobase or sugar backbone modifications.^{6–7} Indeed, enhancing the functional group repertoire of nucleic acid polymers has yielded more favourable binding properties,⁸ and also expanded the utility of this class of polymer within the realm of catalysis.^{9–11}

The traditional method to incorporate modifications throughout a nucleic acid polymer has been to use a polymerase to incorporate modified nucleotide triphosphates.^{12–13} This approach has been successful at increasing the functional properties of nucleic acid polymers; however, it is inherently limited to a maximum of four unique modifications using a standard genetic code;¹⁴ current efforts to expand the genetic code can be used to expand the chemical diversity of the polymer.¹⁵ Recently, our lab developed Ligase-catalysed Oligonucleotide PolymERisation (LOOPER), which enables the sequence-defined incorporation of multiple unique instances of functional groups along a single-stranded DNA (ssDNA) polymer (Scheme 1).^{16–18} The method relies upon T4 DNA ligase to catalyse the DNA-templated polymerisation of modified 5'-phosphorylated oligonucleotides.

[†]Electronic Supplementary Information (ESI) available: Molecular characterisation, processed sequencing data, methods, and supplementary data.

LOOPER has been shown to proceed with excellent fidelities in a library context with codon sets as large as 256 members (codon set = NNNNT). Furthermore, the system can also tolerate a host of chemical modifications on the pentanucleotide building blocks, ranging from small molecules¹⁸ to peptide fragments.¹⁶ Pentanucleotides are ideal for this LOOPER, as they polymerise efficiently, yet require a primer to initiate the polymerisation process, thus defining the reading frame. Longer oligonucleotides also decrease the density of functional group modifications on the polymer product. Despite the efficiency and high fidelity of LOOPER, increased polymerisation fidelities are likely to benefit its application to the *in vitro* selection of modified nucleic acids by increasing enrichment factors and allowing access to larger reading frames.

We have previously shown that adjustment of certain LOOPER reaction parameters had no observable effect on polymerisation fidelity.¹⁷ For example, varying temperatures ranging from 10–30 °C did not produce significantly different polymerisation fidelities, presumably due to the temperatures being higher than the annealing temperatures of the pentanucleotide library. We sought to modify and evaluate other factors that might result in a more discerning polymerisation. To this end, we took aim at the T4 DNA ligase cofactor, ATP. Since the ATP cofactor is used by T4 DNA ligase to covalently activate the 5'-phosphate, it is thus intimately involved in the transition state of the rate-limiting phosphodiester bond formation step.¹⁹ We hypothesised that modifications of the ATP cofactor may disrupt ligation kinetics, resulting in a more stringent and higher fidelity ligation process.

Previous reports have shown that specific modifications can be tolerated on the ATP cofactor during the T4 DNA ligase-catalysed splint ligation. For example, 2'-deoxy modification on the ribose ring or 2-amino modifications on the adenine ring are accommodated by T4 DNA ligase, albeit with decreased levels of efficiency.²⁰ Interestingly, the 1-thiotriphosphate derivative of ATP was recently shown to modestly enhance substrate specificity during a single-ligation process.²¹ In order to explore the potential of enhancing LOOPER fidelity using different cofactors, we chose to examine the influence of chemical modifications at four specific regions of the ATP cofactor, namely modification at the triphosphate region; the 2' and 3' positions of the ribose ring; the Hoogsteen face of the adenine nucleobase; and the Watson-Crick face of the adenine nucleobase (Figure 1). Critical to exploring the hypothesis that a modified adenylated oligonucleotide is ligated with perturbed fidelity, we explored only those ATP modifications that are not lost upon 5'-adenylation of the oligonucleotide substrate.

Our selected library of modified ATP cofactors (Figure 2) was first examined for their efficiency during LOOPER. This was performed using a 3'-hairpin template library comprising eight-codon reading frame of the NNNNT codon set (library complexity = $256^8 = 10^{19}$) (Table 1). The modified 5'-phosphorylated pentanucleotide building block library was A*NNNN, where A* is the 5'-dA nucleotide modified with an aminohexylamine at position 8 of the adenine ring. This modification has been shown to be well-behaved during previous LOOPER and sequencing studies.^{17–18} While our primary goal was to increase the fidelity of LOOPER, the efficiency of the process would need to remain sufficiently high to ensure generation of full-length polymers. Our previous findings have shown that high cofactor concentration can have a detrimental impact of yield during LOOPER, due to re-

adenylation of dissociated ligase during the process. Since the ligase must be in its un-adenylated state to catalyse phosphodiester-bond formation, this effectively shuts down the polymerisation.¹³ However, T4 DNA ligase is likely to have an increased K_m for modified ATP cofactors. Thus, LOOPER was performed at both high (1 mM) and low (25 μ M) cofactor concentrations and the yields determined by denaturing PAGE analysis.

The majority of modified ATP cofactors were ineffective at promoting the LOOPER process. Contrary to the splint-ligation study, which showed that dATP was an effective cofactor for T4 DNA ligase, this was not true during LOOPER (Table 1, entry 3). In fact, we observed that any modification to the ribose ring at either the 2' or 3' positions resulted in very poor polymerisation efficiency (Table 1, entries 3–7, 21). Oxidation of ATP at position 8 of adenine results in shut-down of the polymerisation (Table 1, entry 8). Removal or substitution of the amine with an oxo group at position 6 of adenine was also not tolerated (Table 1, entries 13–14, 17–18). Of the tested cofactors, only five yielded measurable amounts of full-length product during LOOPER. 7-deaza ATP cofactor **8** resulted in only 5% yield of full-length product when used at 25 μ M, with higher concentrations failing to improve yield. 2-Amino ATP cofactor **9** gave a comparable yield (87%) to ATP, and operated best at 25 μ M. Cofactor **12**, which has a methylation of the amine at position 6 of ATP was also well-tolerated and gave good yield (70%). Interestingly, highly efficient polymerisation when using cofactor **12** was only observed at high cofactor concentration, suggesting that T4 DNA ligase exhibits an increased K_m with this cofactor. Noteworthy is that when the modifications of cofactors **9** and **12** are combined to yield cofactor **14**, the LOOPER process become very inefficient (Table 1, entry 18). 2-Cl ATP cofactor **15** was also effective at promoting the polymerization; however, unidentified additional higher weight products were also produced (Figure S5). Lastly, the 1-thiotriphosphate derivative **16** gave measurable yield of the full-length product.

We decided to analyse the fidelity of LOOPER using the cofactors that enabled sufficient yield of pure full-length polymer (cofactors **1**, **8**, **9**, **12**, **16**). To do this, we employed the high-throughput duplex DNA sequencing method that was recently developed in our lab.¹⁷ This method implements a randomised barcoding strategy of the LOOPER products in a library format, which enables the reading of both the template and the corresponding product polymer strand using next-generation sequencing.

Errors in the LOOPER process can then be readily identified allowing the fidelity and bias of each codon within the entire codon set to be calculated (Table 2). Fidelity was calculated as the percentage by which the correct pentanucleotide building block was incorporated across from its cognate codon. Codon bias was calculated as the standard deviation of codon enrichment, whereby codon enrichment is the frequency of a codon in the polymer product divided by its corresponding frequency in the original template library. A codon bias closer to zero denotes lower codon bias during the LOOPER process.

Using ATP as a cofactor, the original LOOPER process with an amine-modified ANNNN library proceeded at 95.1 % fidelity with a codon bias of 0.21. This level of fidelity is sufficient for short reading frames, but for extended reading frames, the portion of the polymer products with at least one error can become a significant fraction of the total

population. Consistent with our hypothesis, we observed that all modified cofactors examined improved the fidelity of LOOPER. While cofactors **8** and **16** provided the greatest fidelities, their low polymerisation efficiencies may preclude their use in preparative applications. Furthermore, both of these cofactors suffered from greater codon bias compared to the other tested cofactors; indeed, the codon bias observed with cofactor **16** led to strong de-enrichment of several codons, primarily codons with low GC-content (Figure 3b). All other cofactors gave low to moderate degrees of codon bias (Figure 3), with individual codon enrichments/de-enrichments being suitable for applications such as *in vitro* selection. We also examined the ability of LOOPER to accommodate larger hydrophobic modifications (benzoylated amino-modified dA) using the optimal set of modified ATP cofactors. While ATP (25 μ M) generated full-length product in 56% yield, cofactors **9** (25 μ M) and **12** (1 mM) gave 35% and 39% yield (Figure S3 and S4).

In conclusion, we have shown that specific ATP modifications can enhance the fidelity of T4 DNA ligase during LOOPER. Of those tested, cofactors **9** and **12** are the most promising ATP substitutes, as they gave the highest polymerisation yield with nucleobase modifications, lowest codon bias, and higher fidelity than unmodified ATP. Due to T4 DNA ligase catalysing self-adenylation, substrate adenylation, and phosphodiester formation, cofactor modifications are likely to have a complex effect on this catalytic process. Thus, further studies are needed to unravel the molecular basis behind the observed increase in fidelity using modified ATP cofactors. Notwithstanding, this improvement to fidelity is likely to have a marked impact on enrichment factors during the *in vitro* selection of modified nucleic acid polymers generated via LOOPER. Moreover, access to longer reading frames will enable longer polymers to be surveyed during selection experiments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health (R21CA207711) and the Office for the Vice President of Research, University of Georgia. We would like to thank the Georgia Genomics Facility for DNA sequencing and the PAMS core facility at the University of Georgia for their help in the characterisation of oligonucleotides.

References

1. Wilson DS, Szostak JW. *Annu Rev Biochem.* 1999; 68:611–647. [PubMed: 10872462]
2. Tuerk C, Gold L. *Science.* 1990; 249:505–510. [PubMed: 2200121]
3. Ellington AD, Szostak JW. *Nature.* 1990; 346:818–822. [PubMed: 1697402]
4. Qu H, Csordas AT, Wang J, Oh SS, Eisenstein MS, Soh HT. *ACS Nano.* 2016; 10:7558–7565. [PubMed: 27399153]
5. Fang X, Tan W. *Acc Chem Res.* 2010; 43:48–57. [PubMed: 19751057]
6. Kong D, Yeung W, Hili R. *ACS Comb Sci.* 2016; 18:355–370. [PubMed: 27275512]
7. Meek KN, Rangel AE, Heemstra JM. *Methods.* 2016; doi: 10.1016/j.ymeth.2016.03.008
8. Rohloff JC, Gelinas AD, Jarvis TC, Ochsner UA, Schneider DJ, Gold L, Janjic N. *Mol Ther Nucleic Acids.* 2014; 3:e201. [PubMed: 25291143]
9. Santoro SW, Joyce GF, Sakthivel K, Gramatikova S, Barbas CF. *J Am Chem Soc.* 2000; 122:2433–2439. [PubMed: 11543272]

10. Hollenstein M, Hipolito CJ, Lam CH, Perrin DM. ACS Comb Sci. 2013; 15:174–182. [PubMed: 23485334]
11. Zhou C, Avins JL, Klauser PC, Brandsen BM, Lee Y, Silverman SK. J Am Chem Soc. 2016; 138:2106–2109. [PubMed: 26854515]
12. Hocek M. J Org Chem. 2014; 79:9914–9921. [PubMed: 25321948]
13. Hottin A, Marx A. Acc Chem Res. 2016; 49:418–427. [PubMed: 26947566]
14. Jager S, Rasched G, Kornreich-Leshem H, Engeser M, Thum O, Famulok M. J Am Chem Soc. 2005; 127:15071–15082. [PubMed: 16248646]
15. Malyshev DA, Romesberg FE. Angew Chem, Int Ed. 2015; 54:11930–11944.
16. Guo C, Watkins CP, Hili R. J Am Chem Soc. 2015; 137:11191–11196. [PubMed: 26274786]
17. Lei Y, Kong D, Hili R. ACS Comb Sci. 2015; 17:716–721. [PubMed: 26513677]
18. Kong D, Lei Y, Yeung W, Hili R. Angew Chem Int Ed Engl. 2016; 55:13164–13168. [PubMed: 27633832]
19. Cherepanov AV, de Vries S. Eur J Biochem. 2003; 270:4315–4325. [PubMed: 14622296]
20. Kinoshita Y, Nishigaki K. J Biochem. 1997; 122:205–211. [PubMed: 9276690]
21. Kim J, Mrksich M. Nucl Acids Res. 2010; 38:e2. [PubMed: 19854942]

Figure 1.
Modification sites of the ATP cofactor.

Figure 2.
Library of ATP cofactor derivatives surveyed in LOOPER.

Figure 3. Bias analysis of NNNNT codon sets used during LOOPER with corresponding hexylamine-modified pentanucleotide libraries and various cofactors. a) cofactor **1** ; b) cofactor **8** ; c) cofactor **9** ; d) cofactor **12** ; e) cofactor **16**.

Scheme 1.

Synthesis of modified ssDNA polymers using T4 DNA ligase-catalyzed polymerisation of modified DNA oligonucleotides

Table 1

Efficiency of polymerisation using modified ATP cofactors

Entry	Cofactor	Cofactor conc. (μM) ^a	Yield (%) ^b
1	1	25	90
2	1	1000	60
3	2	25 or 1000	0
4	3	25 or 1000	0
5	4	25 or 1000	0
6	5	25 or 1000	0
7	6	25 or 1000	0
8	7	25 or 1000	0
9	8	25	5
10	8	1000	0
11	9	25	87
12	9	1000	48
13	10	25 or 1000	0
14	11	25 or 1000	0
15	12	25	0
16	12	1000	70
17	13	25 or 1000	0
18	14	25 or 1000	0
19	15	25	64
20	15	1000	37
21	16	25	10
22	16	1000	0
23	17	25 or 1000	0

^aOptimal cofactor concentration was determined by assessing the polymerisation efficiency by PAGE in the presence of either 25, 250, or 1000 μM of modified ATP cofactor.

^bYield of full-length product was determined by PAGE.

Table 2Fidelity and bias of polymerisation using modified ATP cofactors^a

Entry	Cofactor	Reads ^b	Bias ^c	Fidelity ^d
1	1	31,936	0.21	95.1 %
2	8	10,023	0.51	98.2 %
3	9	12,565	0.32	96.6 %
4	12	12,278	0.28	96.2 %
5	16	21,962	0.42	98.6 %

^aFidelity and bias data was determined by illumine sequencing on a MiSeq (Paired-end 150). 35 amols were used for PCR prior to sequencing to ensure sufficient number of duplex pairs.

^bNumber of codon reads from sequencing data.

^ccodon bias was calculated as $\sigma(\text{freq. polymer}/\text{freq. template})$.

^dFidelity was calculated for pentanucleotide incorporation; the fidelity is considerably higher if evaluated at the single nucleotide level.