

Binary Genetic Cassettes for Selecting XNA-Templated DNA Synthesis In Vivo.**

Valérie Pezo, FengWu Liu, Mikhail Abramov, Mathy Froeyen, Piet Herdewijn* and Philippe Marlière

Information transfer between natural nucleic acids (DNA and RNA) and xenobiotic nucleic acids (XNA) is rapidly gaining momentum for extending the range of chemical constitutions and the format of molecular evolution accessible to living organisms.^[1-3] Artificial coding by nucleic acid analogues previously focussed on structural alterations of base-pairs in order to expand the alphabet of genetic messages.^[4-7] Studies were mostly conducted *ex vivo* and few experiments have succeeded *in vivo* thus far. Kool and collaborators demonstrated that size-expanded nucleobases can serve as template for DNA synthesis in *E. coli*.^[8] Whole genome substitution of thymine for 5-chlorouracil could be performed through automated evolution of *E. coli*.^[9]

Conveying genetic information to DNA from an XNA with a chemically deviant backbone is amenable to tight metabolic selection, as demonstrated for HNA using the thymidylate synthase screen in *Escherichia coli*.^[10] We have now shown that various combinations of only the two bases guanine and thymine can be used to encode the active site of thymidylate synthase. This finding was exploited to simplify the synthesis of XNA to be assayed as templates for DNA biosynthesis *in vivo*, by halving the alphabet needed for this purpose. It could thus be demonstrated that CeNA can serve *in vivo* as template, mobilizing a limited effort of chemical synthesis. Further simplification of the binary system to uracil and hypoxanthine enabled to reprogram *E. coli* with templates simultaneously bearing non-canonical bases and a non-canonical

backbone, namely AraNA and HNA.

A functional *thyA* gene encoding thymidylate synthase is absolutely required by *E. coli* cells to grow in nutrient medium devoid of thymine or thymidine (TLM, thymidine-less medium).^[11] We took advantage of this selection scheme for constructing a plasmid carrying a defective *thyA* gene in which the six codons specifying the active site around the cysteine at position 146 have been deleted, leaving a gap when digested with the restriction enzymes *NheI* and *NsiI*.^[10] Mosaic DNA oligonucleotides in which several of the six codons are carried by an XNA backbone can be tested for informational transfer simply by selecting for active *thyA* genes following transformation of the *thyA* deficient strain G929 with heteroduplex ligation products (Figure 1). Up to 6 contiguous HNA nucleotides were found to serve as a short template for *E. coli* replication enzymes.^[10]

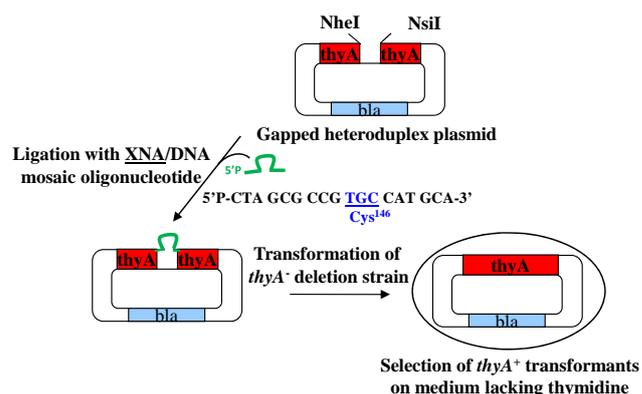


Figure 1. Selection screen to identify artificial XNA oligomers capable of templating DNA synthesis *in vivo*.

The first goal of the present study was to compact the coding set for specifying the active site of thymidylate synthase to two letters, namely T and G, or their pairing equivalents U and I (hypoxanthine). In this way, we hoped to reduce the burden of organic synthesis for evaluating the coding potential of various XNA structures *in vivo*. Although the four A,G,C,T monomers are available in the case of CeNA, HNA, araNA, GNA, quicker exploration of other XNA candidates should result from binary coding of amino acids at a selectable site. The choice of hypoxanthine (I) has the advantage over G that no base protection is required during oligonucleotide synthesis, just as for T and U.^[12] In addition, encoding an enzyme in a two-letter rather than four-letter code poses a functional challenge of 'protein simplification', as investigated by Hilvert and collaborators.^[13, 14]

The catalytic residue Cys146 of *ThyA*, which is known not to be replaceable by any other amino acid, can be encoded by the triplet TGT. The five other amino acids that can be encoded with binary G/T triplets, Phe, Leu, Trp, Val and Gly, belong to a small subset of aliphatic and aromatic residues.

[*] FengWu Liu, Mikhail Abramov, Mathy Froeyen, Piet Herdewijn
Laboratory for Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven
Minderbroedersstraat 10, 3000 Leuven, (Belgium)
E-mail: Piet.Herdewijn@rega.kuleuven.be

Valérie Pezo, Piet Herdewijn, Philippe Marlière
ISSB, Génopole, Genavenir6, Equipe Xénome, 5 rue Henri Desbruères, 91030 Evry Cedex, France

Valérie Pezo
CEA, DSV, IG, Genoscope, 2 rue Gaston Crémieux, 91057 Evry Cedex, France

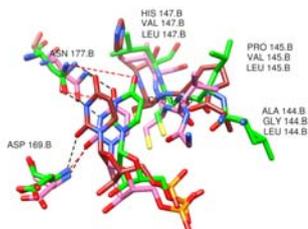
[**] Authors are indebted to G Schepers for the synthesis of the oligonucleotides, to Dr J Rozenski for mass spec. analysis, and to Lan N'Guyen and Mélanie Crouzard for technical assistance. We thank KU Leuven (GOA, IDO), FWO (Vlaanderen) and Genoscope (Genopole, Evry) for financial support (ATIGE, EA 4527, UPS 3509). The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) /ERC Grant agreement n° ERC-2012-ADG_20120216/320683

Supporting information for this article is available on the WWW under <http://www.angewandte.org> or from the author.

Table 1. Statistical composition of G/T binary DNA cassettes encoding the active site of thymidylate synthase in the *E. coli* thyA gene.

	Oligonucleotides						Number of oligo-nucleotides in the library	Number of bla ⁺ colonies x10 ⁴	Number of thyA ⁺ colonies	Ratio thyA ⁺ /bla ⁺
	Leu ¹⁴³ Ala ¹⁴⁴ Pro ¹⁴⁵ Cys ¹⁴⁶ His ¹⁴⁷ Ala ¹⁴⁸									
Wild-type	5'P-CTA	GCG	CCG	TGC	CAT	GCA-3'		63	245000	0.38
Library K7	5'P-CTA	GKK	KKK	TGT	KKT	GCA-3'	128	97	24	0.000027
Library K10	5'P-CTA	GKK	KKK	KKK	KKT	GCA-3'	1024	16	3	0.000018

K=T or G

**Figure 2.** Superposition of thymidylate active sites from wild-type (WT) *E. coli* (in brown), active mutant A144G-P145V-C146-H147V (in pink) and inactive mutant A144L-P145L-C146-H147L (in green). The wild-type structure corresponds to the B subunit of the homodimer in the X-ray structure. Representative active and inactive mutant structures were inferred from the 3 ns MD trajectory to highlight the hydrogen bond (HB) changes with the uracil moiety of the dUMP substrate shown at the bottom of the figure. The methylenetetrahydrofolate cosubstrate is not shown. The HB distances UMP903.O2/Asp169.N (3.0, 2.9, 4.0 Å for WT, active and inactive mutant), UMP903.O4/ Asp177.ND2 (3.0, 2.9, 5.2 Å for WT, active mutant and inactive mutant) and UMP903.N3/ Asp177.OD1 (3.0, 2.9, 4.3 Å for WT, active mutant and inactive mutant) are indicated by dashed lines. HBs in the inactive mutant are red. Structural comparison thus points to the loss of these HBs in the inactive mutant. Figure generated using the Chimera software [26].

We resorted to a combinatorial library of synthetic oligonucleotides for assessing whether a stretch of G's or T's could be extended upstream and downstream of the Cys codon without losing the function of the ThyA enzyme.

Table 1 shows the combinatorial cassettes (K7 and K10) which were used in the selection process. The cassette K7 consisted of 7 nucleotides (K stands for T or G) covering 3 amino acid codons, corresponding to a library of 128 different sequences. The cassette K10 consisted of 10 nucleotides covering 4 amino acid codons, corresponding to a library of 1024 sequences. Independent ligation of the two oligonucleotide libraries K7 and K10 with the gapped plasmid heteroduplex (Figure 1) followed by transformation of the thyA deleted strain G929 and growth on TLM plates resulted in growing colonies in both cases.

The DNA and protein sequences corresponding to such colonies with functional thyA genes are given in Table 2A and 2B. As expected, only cysteine (encoded by TGT) was found at the catalytically site 146 (Cassette K10 – Table1 and Table 2).

Despite the high conservation throughout evolution of the Ala-Pro-Cys-His sequence in the active site of thymidylate synthase [11], Val (GTT or GTG), Cys (TGT) and Gly (GGT or GGG) were found to be acceptable replacements at positions upstream and downstream of Cys 146 (Table 2). The larger G/T-encoded amino acids Phe, Leu and Trp were found neither downstream nor upstream. Active sites

further simplified were constructed separately by ligating appropriate oligonucleotides to the gapped plasmid heteroduplex. It was thus found that the sequences Cys-Cys-Cys-Cys and Val-Val-Cys-Val led to an active site, whereas the sequences Gly-Gly-Cys-Gly and Leu-Leu-Cys-Leu led to inactive enzymes.

Table 2. Plasmid sequences from thymidine prototrophic transformants of *E. coli* originating from G/T DNA binary cassettes. A) DNA sequences B) protein sequences

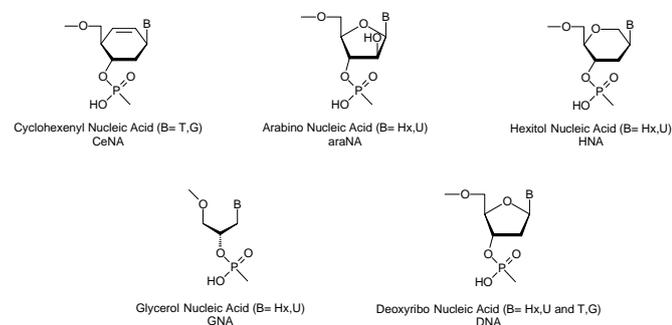
A									Number of thyA ⁺ colonies
WT	Leu ¹⁴³ Ala ¹⁴⁴ Pro ¹⁴⁵ Cys ¹⁴⁶ His ¹⁴⁷ Ala ¹⁴⁸								
K7	5'P-	CTA	GKK	KKK	TGT	KKT	GCA	-3'	
		CTA	GGT	GTT	TGT	GTT	GCA		2
		CTA	GGT	GTT	TGT	TGT	GCA		1
		CTA	GGG	GTT	TGT	TGT	GCA		1
		CTA	GGG	GTG	TGT	TGT	GCA		2
		CTA	GTT	GTT	TGT	GGT	GCA		3
		CTA	GTT	TGT	TGT	GGT	GCA		2
		CTA	GTT	TGT	TGT	TGT	GCA		1
		CTA	GTT	TGT	TGT	TGT	GCA		1
		CTA	GTG	TGT	TGT	GGT	GCA		3
		CTA	GTG	TGT	TGT	GTT	GCA		2
		CTA	GTG	GTG	TGT	GGT	GCA		1
		CTA	GTG	GTT	TGT	GGT	GCA		1
		CTA	GTG	GTT	TGT	TGT	GCA		1
		CTA	GTG	GGT	TGT	GTT	GCA		1
K10	5'P-	CTA	GKK	KKK	KKK	KKT	GCA	-3'	
		CTA	GTG	TGT	TGT	GTT	GCA		2
		CTA	GGT	GTG	TGT	TGT	GCA		1
B									
WT	Ala ¹⁴⁴ Pro ¹⁴⁵ Cys ¹⁴⁶ His ¹⁴⁷				Number of clones				
K7	Gly	Val	Cys	Val	4				
	Gly	Val	Cys	Cys	2				
	Val	Val	Cys	Gly	5				
	Val	Cys	Cys	Gly	5				
	Val	Cys	Cys	Cys	1				
	Val	Cys	Cys	Val	2				
	Val	Val	Cys	Cys	1				
	Val	Gly	Cys	Val	1				
K10	Val	Cys	Cys	Val	1				
	Gly	Val	Cys	Cys	1				

Molecular modelling was conducted to interpret these surprising genetic findings by structural analysis. Nine functional mutants (Suppl. Section, Table I) and one defective mutant (Leu-Leu-Cys-Leu) of thymidylate synthase were investigated using molecular dynamics over a simulation time of 3 ns^[15] (for experimental data, see supplementary section). No difference was observed between the distance of the catalytic Cys146 of the C6 atom of dUMP between the active and inactive mutants (attack by the sulphur atom of Cys146 to this C6 atom is the first step in the catalytic process

Table 3: Templating by G/T binary CeNA (cyclohexene nucleic acid) for restoring an active thymidylate synthase gene in *E. coli*.

	Oligonucleotides						Number of CeNA residues	Number of bla ⁺ colonies x10 ⁴	Number of thyA ⁺ colonies	Ratio thyA ⁺ /bla ⁺
	Leu ¹⁴³ Ala ¹⁴⁴ Pro ¹⁴⁵ Cys ¹⁴⁶ His ¹⁴⁷ Ala ¹⁴⁸									
Wild-type	CTA	GCG	CCG	TGC	CAT	GCA	0	634	2450000	0.38
Cys 146 deleted	CTA	GCG	CCG	ΔΔΔ	CAT	GCA	0	285	0	0
DNA/CeNA mosaic template	CTA	GTT	GTT	TGT	GGT	GCA	3	1273	2920000	0.23
	CTA	GTT	GTT	TGT	GGT	GCA	6	805	490000	0.06
	CTA	GTT	GTT	TGT	GGT	GCA	6	1207	700000	0.058
	CTA	GTT	GTT	TGT	GGT	GCA	9	1000	3600	0.00036
	CTA	GTT	GTT	TGT	GGT	GCA	12	611	900	0.00014
	CTA	GTG	GTG	TGT	GGT	GCA	12	556	370	0.000066

leading to dTMP). However, a clear difference was seen between the hydrogen bonding networks of dUMP in the active sites of the functional and defective mutants, suggesting that the substrate critically loses bonding strength in the latter case (Figure 2).

**Figure 3.** Structure of sugar- and base-modified XNA tested as binary genetic cassettes

The second goal of the present study was to assay the templating properties of chemically simplified XNA oligonucleotides. We call transliteration the process catalysed by DNA polymerases and other enzymes of the replication machinery through which a stretch of XNA acts as template in the condensation of deoxynucleoside triphosphates (dNTPs) to form a DNA sequence complementary to the XNA. Once formed that DNA sequence undergoes canonical replication in subsequent cell generation.

The structure of the hexitol nucleic acid (HNA) and the other chemically modified nucleic acids that were used in this study are depicted in Figure 3. Detailed description of the synthesis and analysis of the nucleoside [16-19] and of the oligonucleotide [20-22] analogues can be found in the supplementary section.

As the dodecamer GTT-GTT-TGT-GGT was the most abundantly selected sequence among the combinatorial libraries, we used this binary sequence for testing the coding potential of cyclohexenyl nucleic acid (CeNA), i.e. a polynucleotide analogue whose backbone bears little resemblance to the oxygenated five-membered ring of ribose and deoxyribose. Yet, CeNA is still able to hybridize in a sequence selective manner with DNA as well as RNA and the triphosphates of CeNA nucleosides can serve as substrate of several DNA polymerases.^[23] It is shown in Table 2 that substituting one (TGT), two (GTT-TGT or TGT-GGT), three (GTT TGT-GGT) or four (GTT-GTT-TGT-GGT) codons in the active site of ThyA by the binary CeNA corresponding G/T sequences led to the recovery of thyA⁺ colonies with a decreasing efficiency. This

experiment provided a proof of principle validating the use of binary code for systematically assessing informational transfer by xenobiotic heredity supports.

We then tested whether the binary G/T cassettes can be replaced by a binary I/U XNA cassettes. Indeed, when synthesizing sugar analogues of nucleic acids, the nucleoside with the G base is invariably the most difficult to obtain. Furthermore, the presence of multiple G residues in DNA and XNA oligonucleotides frequently hinders purification, rendering the protection of the G base (as well as the C and A base) mandatory for oligonucleotide synthesis. Therefore, we investigated a further chemical simplification of the G/T binary code to a I/U binary code, hypoxanthine serving as guanine surrogate for pairing with cytosine, and uracil serving as thymine surrogate for pairing with adenine. It should be noted that both these simplified purine and pyrimidine are actively repaired when present in cellular DNA by dedicated enzymes hydrolysing hypoxanthine or uracil deoxynucleotides.^[15, 24]

The same experiments as performed with binary G/T CeNA cassettes were repeated using binary I/U sequences borne on the DNA backbone and on the XNA backbones HNA^[25], araNA^[19] and presumably due to degradation through the action of the N-glycosylases Ung and Fpg. The GNA/DNA mosaic templates did not yield any thyA⁺ colony, even when only one GNA codon (UIU coding for Cys146) was present. Apparently the glycerol backbone structure was too alien for being used as template by the DNA replication machinery of *E. coli*. Implementation of I/U binary XNA sequences in the two other XNA templates arabinofuranosyl nucleic acid (AraNA) and hexitol nucleic acid (HNA) succeeded in restoring a functional thyA gene (Table 3). The yields of prototrophs were found to be similar for AraNA and HNA as for G/T binary CeNA sequences. The yield of thymidine prototrophic colonies decreased steeply with the length of the XNA stretch in the mosaic template. For each chemically modified template assayed, 16 thyA⁺ colonies growing on TLM were subjected to DNA sequencing of the Cys 146 region of ThyA. This analysis revealed that the genetic message conveyed under the form of binary G/T sequences in CeNA and of binary I/U sequences in AraNA and HNA was correctly copied into DNA in each of the scrutinized cases. Faithful transmission of hereditary information can thus be accomplished from templates lacking a canonical backbone and canonical bases at the same time. This augurs well for the propagation of a third type of nucleic acid aside from DNA and RNA in bacterial cells.

A tight genetic selection screen based on the restoration of thymidylate synthase active site was elaborated for assaying the use of short XNA stretches carrying only the two bases T and G, as

Table 4: Templating by I/U binary XNAs for restoring an active thymidylate synthase gene in *E. coli*. (A) I/U arabino-nucleic acid (B) I/U hexitol nucleic acid.

		Oligonucleotides					Number of XNA residues	Number of bla+ colonies	Number of thyA+ colonies	Ratio thyA+/bla+	
		Leu ¹⁴³	Ala ¹⁴⁴	Pro ¹⁴⁵	Cys ¹⁴⁶	His ¹⁴⁷	Ala ¹⁴⁸				
A	DNA/	CTA	GTT	GTT	TGT	GGT	GCA	0	150.5	124000	0.08
	arabino I/U	CTA	GTT	GTT	UGU	GGT	GCA	2	294	180000	0.06
	mosaic	CTA	GTT	GTT	UIU	GGT	GCA	3	215	112000	0.05
	template	CTA	GTT	IUU	UIU	GGT	GCA	6	260	6300	0.002
		CTA	GTT	GTT	UIU	IIU	GCA	6	0.5	20	0.004
		CTA	GTG	IUU	UIU	IIU	GCA	9	11.5	220	0.002
		CTA	IUU	IUU	UIU	GGT	GCA	9	12.5	140	0.001
B	DNA/	CTA	GTT	GTT	TGT	GGT	GCA	0	108	130000	0.12
	hexitol I/U	CTA	GTT	GTT	TIT	GGT	GCA	1	142	88800	0.06
	mosaic	CTA	GTT	GTT	UGU	GGT	GCA	2	57.5	45000	0.08
	template	CTA	GTT	GTT	UIU	GGT	GCA	3	393	151200	0.04
		CTA	GTT	IUU	UIU	GGT	GCA	6	42	19500	0.05
		CTA	GTT	GTT	TGT	GGT	GCA	0	1739	480000	0.02
		CTA	GTT	GTT	UIU	IIU	GCA	6	124	19300	0.01
		CTA	GTG	IUU	UIU	IIU	GCA	9	11.5	140	0.0012
		CTA	IUU	IUU	UIU	GGT	GCA	9	141	170	0.0001
		CTA	IUU	IUU	UIU	IIU	GCA	12	97	100	0.0001

templates for DNA biosynthesis in the model bacterium *Escherichia coli*. This screen allowed us to demonstrate that hereditary information can be conveyed to DNA under the form of CeNA, AraNA, HNA messages but not GNA. Moreover, simultaneously altering the structure of the nucleobase and the backbone sugar did not abolish templating of DNA biosynthesis, as exemplified by faithful reading of dodecameric hexitol stretches bearing hypoxanthine and uracil. This study incidentally demonstrated that various amino acid sequences are tolerated in the active site region of thymidylate synthase. These versatile G/T or I/U binary cassettes selectable in the thyA gene of *E. coli* will now be applied to improve XNA-dependent DNA polymerase activity *in vivo*.

Experimental Section

Synthesis, spectroscopic characterization of monomers and oligonucleotides, molecular modelling and biological testing are described in the supplementary information.

Received: ((will be filled in by the editorial staff))

Published online on ((will be filled in by the editorial staff))

Keywords: xeno-nucleic acid · thymidylate synthetase · hypoxanthine uracil

- [1] P. Herdewijn, P. Marliere, *Chem. Biodivers.* 2009, 6, 791-808.
- [2] V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S. Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn, P. Holliger, *Science* 2012, 336, 341-344.
- [3] V. B. Pinheiro, P. Holliger, *Curr. Opin. Chem. Biol.* 2012, 16, 245-252.
- [4] J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* 1990, 343, 33-37.
- [5] J. C. Morales, E. T. Kool, *Nat. Struct. Biol.* 1998, 5, 950-954.
- [6] D. L. McMinn, A. K. Ogawa, Y. Q. Wu, J. Q. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* 1999, 121, 11585-11586.

- [7] T. Ohtsuki, M. Kimoto, M. Ishikawa, T. Mitsui, I. Hirao, S. Yokoyama, *Proc Natl Acad Sci U S A* 2001, 98, 4922-4925.
- [8] A. T. Krueger, L. W. Peterson, J. Chelliserry, D. J. Kleinbaum, E. T. Kool, *J Am Chem Soc* 2011, 133, 18447-18451.
- [9] P. Marliere, J. Patrouix, V. Döring, P. Herdewijn, S. Tricot, S. Cruveiller, M. Bouzon, R. Mutzel, *Angew. Chem.* 2011, 50, 7109-7114.
- [10] S. Pochet, P. A. Kaminski, A. Van Aerschot, P. Herdewijn, P. Marliere, *C. R. Biol.* 2003, 326, 1175-1184.
- [11] B. Lemeignan, P. Sonigo, P. Marliere, *J Mol Biol* 1993, 231, 161-166.
- [12] F. Seela, X. Ming, *Helv. Chim. Acta* 2008, 91, 1181-1200.
- [13] K. U. Walter, K. Vamvaca, D. Hilvert, *J Biol Chem* 2005, 280, 37742-37746.
- [14] M. M. Muller, J. R. Allison, N. Hongdilokkul, L. Gaillon, P. Kast, W. F. van Gunsteren, P. Marliere, D. Hilvert, *PLoS Genet* 2013, 9, e1003187.
- [15] B. K. Tye, I. R. Lehman, *J Mol Biol* 1977, 117, 293-306.
- [16] B. De Bouvere, L. Kerremans, J. Rozenski, G. Janssen, A. van Aerschot, P. Claes, R. Busson, P. Herdewijn, *Liebigs Ann* 1997, 1997, 1453-1461.
- [17] J. Wang, J. Morral, C. Hendrix, P. Herdewijn, *The Journal of Organic Chemistry* 2001, 66, 8478-8482.
- [18] M. K. Schlegel, E. Meggers, *The Journal of Organic Chemistry* 2009, 74, 4615-4618.
- [19] A. M. Noronha, C. J. Wilds, C.-N. Lok, K. Viazovkina, D. Arion, M. A. Parniak, M. J. Damha, *Biochemistry (Mosc)*. 2000, 39, 7050-7062.
- [20] C. Hendrix, H. Rosemeyer, I. Verheggen, F. Seela, A. Van Aerschot, P. Herdewijn, *Chem-Eur J* 1997, 3, 110-120.
- [21] J. Wang, B. Verbeure, I. Luyten, E. Lescrier, M. Froeyen, C. Hendrix, H. Rosemeyer, F. Seela, A. Van Aerschot, P. Herdewijn, *J. Am. Chem. Soc.* 2000, 122, 8595-8602.
- [22] L. Zhang, A. E. Peritz, P. J. Carroll, E. Meggers, *Synthesis* 2006, 2006, 645-653.
- [23] V. Kempeneers, M. Renders, M. Froeyen, P. Herdewijn, *Nucleic Acids Res.* 2005, 33, 3828-3836.
- [24] B. Weiss, *Mutat. Res.* 2001, 461, 301-309.
- [25] P. Herdewijn, *Chem. Biodivers.* 2010, 7, 1-59.
- [26] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* 2004, 25, 1605-1612.

Supplementary Information

*Binary Genetic Cassettes for Selecting XNA-Templated DNA Synthesis In Vivo***

Valérie Pezo, FengWu Liu, Mikhail Abramov, Mathy Froeyen, Piet Herdewijn and Philippe Marlière*

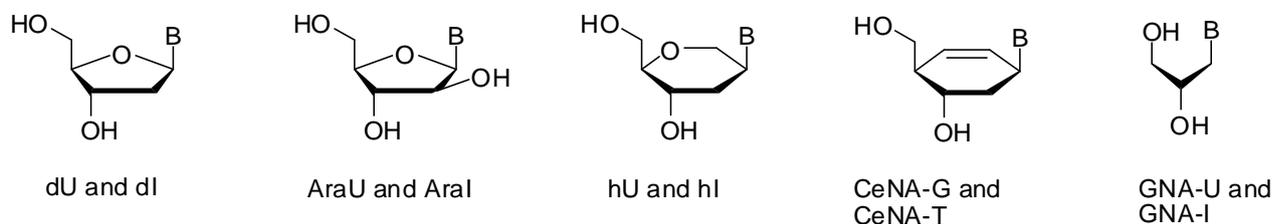
Table of Contents

Synthesis of phosphoramidite compounds	S1
Synthesis and characterization of oligonucleotides	S4
Molecular modelling of active/inactive mutants of E. coli ThyA	S6
Propagation of oligonucleotides in vivo	S14
References	S16

Synthesis of phosphoramidite compounds

Deoxyribose (dU and dI) modified oligonucleotides were obtained by using commercially available amidites (Glen Research). The araI phosphoramidite was synthesized as O^{5'}-DMTr O^{2'}-Ac derivative. The hI phosphoramidite was synthesized as O^{5'}-DMT derivative. The chemical synthesis of araU and hU modified phosphoramidites is previously described^{[1],[2]} as well as the

synthesis of the cyclohexenyl phosphoramidites^[3] and the synthesis of the phosphoramidites for GNA synthesis.^[4]



SS Figure 1. Structure of investigated chemical modifications

General. For all reactions, analytical grade solvents were used. All moisture-sensitive reactions were carried out in oven-dried glassware under N₂. Reagents and solvents were provided by Acros, Fluka, or Pharma Waldhof. TLC: Precoated aluminum sheets (Fluka silica gel/TLC cards, 254 nm); the spots were visualized with UV light. Column chromatography (CC): ICN silica gel 63–200 60 P. ³¹P-NMR spectra: a Bruker Avance 300-MHz, or a Bruker Avance 500-MHz spectrometer. Exact mass measurements were performed on a quadrupole time-of-flight mass spectrometer (Q-Tof-2, Micromass, Manchester, UK) equipped with a standard electrospray-ionization (ESI) interface; samples were infused in i-PrOH/H₂O 1 : 1 at 3 cm³/min.

5'-O-dimethoxytrityl-2'-O-acetyl-ara-inosine, 3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite. To a colorless solution of ara-inosine (1.0 g, 3.73 mmol) in pyridine (25 mL) DMTrCl (1.56 g, 4.47 mmol) was added in one portion at RT. The reaction mixture was stirred for 12 h and turn to yellow-orange color. When the starting material has disappeared the reaction mixture was cooled in an ice bath, methanol (1 mL) was added, the reaction mixture was concentrated and coevaporated twice with toluene. The residue was dissolved in DCM, washed

with H₂O, dried over Na₂SO₄, and purified by column chromatography on silica gel to yield 5'-*O*-dimethoxytrityl-ara-inosine (1.75g, 82%). This compound (1.0 g, 1.75 mmol) was dissolved in DCM (10 mL) and cooled in an ice-bath. DIPEA (1.5 mL, 8.76 mmol) and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (0.58 mL, 2.6 mmol) were added. The reaction solution was stirred for 30 min at RT. Upon completion the reaction mixture was concentrated and coevaporated twice with toluene. The crude material was purified by column chromatography on silica to yield 5'-*O*-dimethoxytrityl-ara-inosine, 3'-[(2-cyanoethyl)-(*N,N*-diisopropyl)]-phosphoramidite (0.75g, 55%). The isolated compound (0.5 g, 0.670 mmol) was dissolved in DCM (10 mL), 4-(dimethylamino)pyridine (0.033 g, 0.268 mmol), DIPEA (0.935 ml, 5.36 mmol), and acetic anhydride (0.253 ml, 2.68 mmol) were added dropwise. After stirring for 45 min, DCM (50 mL) was added, the reaction mixture was washed with H₂O, dried over Na₂SO₄, and purified by column chromatography on silica gel to yield 2'-*O*-acetyl-5'-*O*-dimethoxytrityl-ara-inosine, 3'-[(2-cyanoethyl)-(*N,N*-diisopropyl)]-phosphoramidite (0.33 g, 62%). ³¹P NMR (CDCl₃, 25 °C): δ = -151.0. HRMS calcd for C₄₂H₄₉N₆O₉P, [MH⁺] 813.3377, found 813.3401.

1',5'-Anhydro-5'-*O*-dimethoxytrityl-2',3'-dideoxy-2'-hypoxanthinyl-ara-D-hexitol, 3'-[(2-cyanoethyl)-(*N,N*-diisopropyl)]-phosphoramidite. To a colorless solution of 1',5'-anhydro-2',3'-dideoxy-2'-hypoxanthinyl-ara-D-hexitol^[5] (0.35 g, 1.3 mmol) in pyridine (10 mL) DMTrCl

(0.6 g, 1.7 mmol) was added in one portion at RT. The reaction mixture was stirred for 3 h and turn to yellow-orange color. When the starting material has disappeared the reaction mixture was cooled in an ice bath, methanol (1 mL) was added, the reaction mixture was concentrated and coevaporated twice with toluene. The residue was dissolved in DCM, washed with H₂O, dried over Na₂SO₄, and purified by column chromatography on silica gel to yield 1',5'-anhydro-5'-*O*-

dimethoxytrityl-2',3'-dideoxy-2'-hypoxanthinyl-ara-D-hexitol (0.32 g, 43%). This compound (0.32 g, 0.56 mmol) was dissolved in DCM (5 mL) and cooled in an ice-bath. DIPEA (0.73 mL, 4.2 mmol) and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (0.31 mL, 1.4 mmol) were added. The reaction solution was stirred for 30 min at RT. Upon completion the reaction mixture was concentrated and coevaporated twice with toluene. The crude material was purified by column chromatography on silica to yield 5'-O-dimethoxytrityl-2'-hypoxanthinyl-ara-D-hexitolinosine, 3'-[(2-cyanoethyl)-(*N,N*-diisopropyl)]-phosphoramidite (0.29g, 68%). ³¹P NMR (CDCl₃, 25 °C): δ = 151.0. HRMS calcd for C₄₁H₄₉N₆O₇P, [MH⁺] 769.3479, found 769.3476.

Synthesis and characterization of oligonucleotides

The phosphoramidites were incorporated into the DNA sequences through solid-phase DNA synthesis^[6] on an automated RNA synthesizer (for structures of chemically modified nucleotide monomers see SS Figure 1). For a standard DNA synthesis cycle (1 μmol scale), O^{5'}-DMTr protected DNA phosphoramidites for ultramild DNA synthesis (Glen Research) and common reagents were used and the stepwise coupling yield of all monomers was >99%. For incorporation of modified nucleotides, a coupling time of 10 min was used. Following standard

deprotection, purification and workup, the composition and purity (>90%) of the resulting The synthesis of CeNA oligonucleotides^[3] and GNA oligonucleotides^[7] have been described before.

SS Table 1. Deoxyribose, arabinose, hexitol and glycerol modified nucleic acids (U: uracil base, I: hypoxanthine base) and MS analysis.

<i>Deoxyribose modified nucleic acids (5' → 3')</i>								<i>Calcd</i>	<i>Found</i>
2951	²⁻ O ₃ P	CTA	GTT	GTT	TIT	GGT	GCA	5613.9	5613.9
2952	²⁻ O ₃ P	CTA	GTT	GTT	UGU	GGT	GCA	5600.9	5601.0
2953	²⁻ O ₃ P	CTA	GTT	GTT	UIU	GGT	GCA	5585.9	5586.0
2954	²⁻ O ₃ P	CTA	GTT	IUU	UIU	GGT	GCA	5542.8	5542.9
2955	²⁻ O ₃ P	CTA	GTT	IUU	UIU	IUU	GCA	5498.8	5498.9
2956	²⁻ O ₃ P	CTA	IUU	IUU	UIU	IUU	GCA	5455.7	5455.9
2961	²⁻ O ₃ P	CTA	GTT	GTT	UIU	IUU	GCA	5541.8	5541.8
2962	²⁻ O ₃ P	CTA	IUU	IUU	UIU	GGT	GCA	5499.8	5499.8
<i>Arabinose modified nucleic acids (5' → 3')</i>								<i>Calcd</i>	<i>Found</i>
2969	²⁻ O ₃ P	CTA	GTT	GTT	TIT	GGT	GCA	5629.9	5630.0
3004	²⁻ O ₃ P	CTA	GTT	GTT	UGU	GGT	GCA	5632.9	5632.7
3005	²⁻ O ₃ P	CTA	GTT	GTT	UIU	GGT	GCA	5633.8	5633.8
3006	²⁻ O ₃ P	CTA	GTT	IUU	UIU	GGT	GCA	5638.8	5638.8
3007	²⁻ O ₃ P	CTA	GTT	IUU	UIU	IUU	GCA	5642.7	5643.0
3011	²⁻ O ₃ P	CTA	GTT	GTT	UIU	IUU	GCA	5637.8	5638.3
3012	²⁻ O ₃ P	CTA	IUU	IUU	UIU	GGT	GCA	5643.7	5643.8

<i>Hexitol modified nucleic acids (5' → 3')</i>								<i>Calcd</i>	<i>Found</i>
2982	² -O ₃ P	CTA	GTT	GTT	TIT	GGT	GCA	5627.9	5627.7
2983	² -O ₃ P	CTA	GTT	GTT	UGU	GGT	GCA	5628.9	5628.7
2984	² -O ₃ P	CTA	GTT	GTT	UIU	GGT	GCA	5627.9	5627.9
2985	² -O ₃ P	CTA	GTT	IUU	UIU	GGT	GCA	5626.9	5627.0
2986	² -O ₃ P	CTA	GTT	IUU	UIU	IUU	GCA	5624.9	5625.1
2987	² -O ₃ P	CTA	IUU	IUU	UIU	IUU	GCA	5623.9	5623.9
2992	² -O ₃ P	CTA	GTT	GTT	UIU	IUU	GCA	5625.9	5626.3
2993	² -O ₃ P	CTA	IUU	IUU	UIU	GGT	GCA	5625.9	5626.3
<i>Glycerol modified nucleic acids (5' → 3')</i>								<i>Calcd</i>	<i>Found</i>
2694	² -O ₃ P	CTA	GTT	GTT	UIU	GGG	GCA	5484.8	5484.9
2695	² -O ₃ P	CTA	IUU	IUU	UIU	III	GCA	4975.6	4975.7
2696	² -O ₃ P	CTA	GTT	IUU	UIU	GGG	GCA	5315.8	5315.9
2697	² -O ₃ P	CTA	IUI	IUI	UIU	IUU	GCA	4999.6	4999.8

Modified nucleosides are shown in bold

Molecular modelling of active/inactive mutants of E Coli ThyA

Eight mutants of the E coli Thymidylate Synthase (SS table II) were created starting from pdb entry code 2g8o using the edpdb software.^[8] Amber simulation files were created for those mutants (SS Table II) and the wildtype enzyme having a dimer structure of the enzyme together with twice the substrate dUMP and the cofactor analogue CB3717. Antechamber and the RESP method as described elsewhere were used for the preparation of the CB3717 and dUMP molecules respectively.^{[9],[10]} Then MD simulations were started for wildtype and 8 mutants

using the Sander program from the Amber 9 software using a similar protocol as described elsewhere.^[10] Total simulation time was 3 ns for all systems.

Molecular Dynamics (MD) trajectories of 3 ns were obtained for wildtype and all mutant enzymes. However only the B subunit in the dimer is stable during the 3 ns of molecular dynamics simulation (SS Figure 2). Therefore we will only consider this subunit in the further discussion.

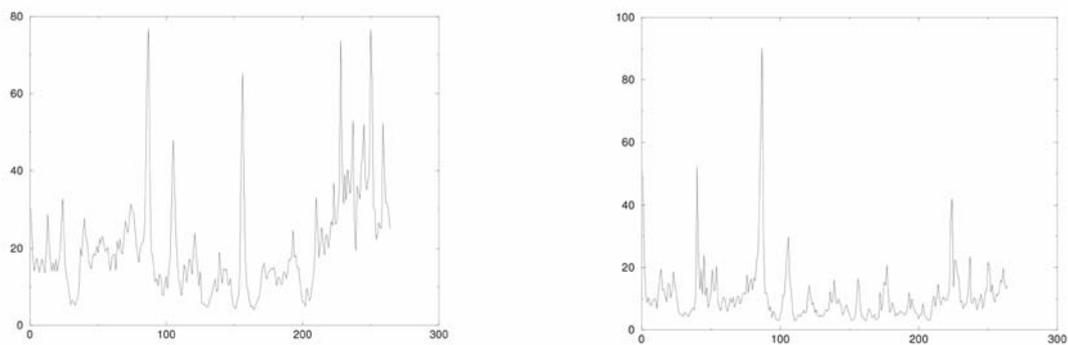
As a measure for the activity, we chose the distance of the catalytic cysteine-146 (SG atom) to the C6 atom in the nucleic base or dUMP. During the catalysis this negatively charged cysteine sulfur atom (SG) is believed to attack the C6 atom of dUMP and gets covalently bound to it. In the x-ray structure the SG-C6 distance is 1.90 and 1.79 Å in the A and B monomer respectively. In SS Figure 2 this distance is plotted as function of the Molecular Dynamics simulation time for monomer B. It can be seen that the distance is between 3 and 4 Å in the wildtype simulation. Sometimes the distance jumps to a higher value. Examination of the molecular structures learns that this is caused by rotation of the cysteine-146 side chain (SS Figure 4). When the cysteine-146 sidechain enters into another rotameric state, the distance between atom SG and atom C6 of dUMP changes accordingly.

Mutant A144L-P145L-C146-H147L is experimentally found to be inactive (SS Table 2). However from the MD simulation no specific changes were observed in the A144L-P145L-C146-H147L loop region; the C146 cysteine SG is still in a possible reacting distance from the C6 atom of UMP (SS Figure 3, SS Figure 5).

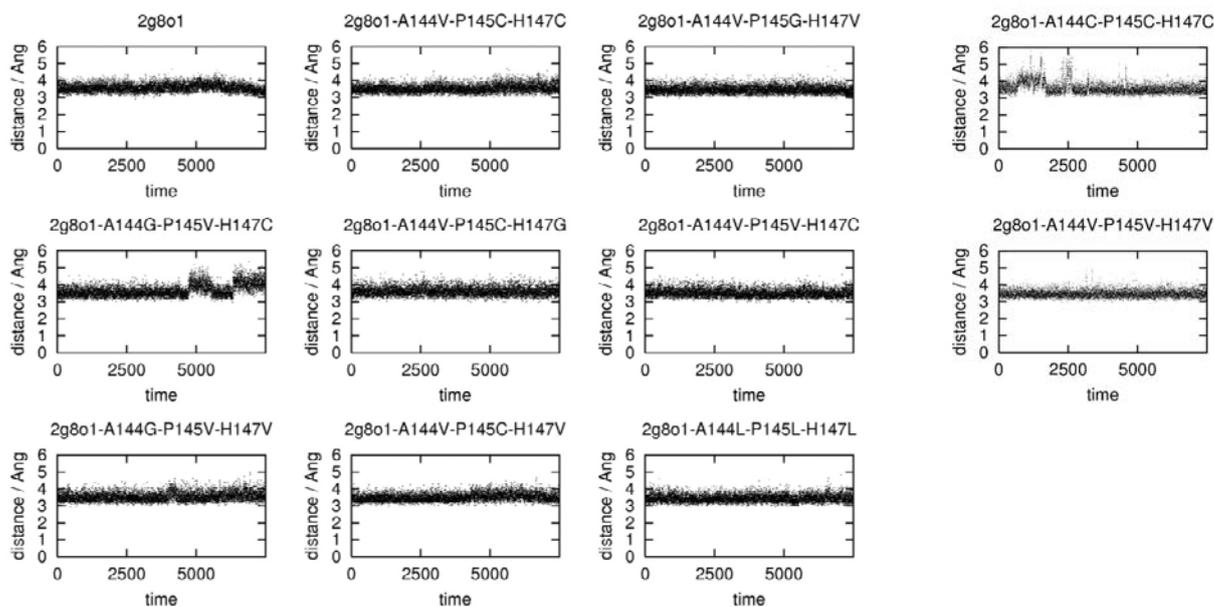
A closer investigation of the hydrogen bonding interactions of the dUMP with the enzyme in the x-ray structure reveals that there are 3 hydrogen bonds UMP903.O2/Asp169.N, UMP903.N3/Asp177.OD1 and UMP903.O4/Asp177.ND2. SS Table III shows the occupancy of these hydrogen bonds during the different MD simulations. In the triple Leucine mutant these hydrogen bonds are disrupted (Figure 2, main text). An explanation for this can be that Leu147 interacts with the base (SS Figure 6 and SS Figure 7) but also with residue Asp177 (SS Figure 8). These interactions may cause a change in the orientation of the base disturbing the hydrogen bond network, resulting in a less stable binding of dUMP in the active site (Figure 2, main text).

A144	P145	C146	H147	wt, reference
A144V	P145C	C146	H147G	Active
A144G	P145V	C146	H147C	Active
A144V	P145C	C146	H147V	Active
A144G	P145V	C146	H147V	Active
A144V	P145G	C146	H147V	Active
A144V	P145C	C146	H147C	Active
A144V	P145V	C146	H147C	Active
A144C	P145C	C146	H147C	Active
A144V	P145V	C146	H147V	Active
A144L	P145L	C146	H147L	Inactive

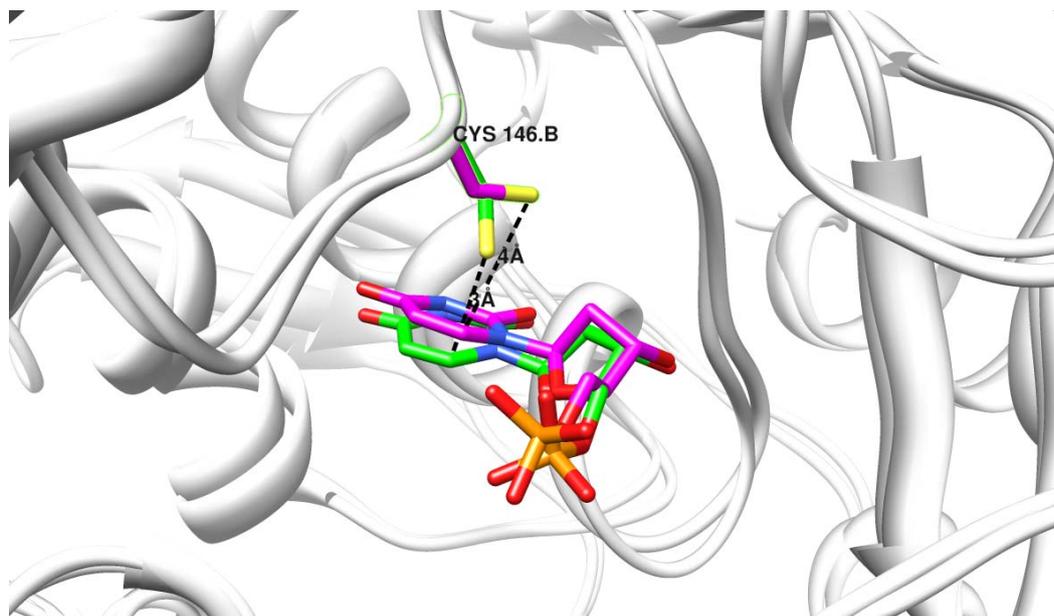
SS Table II. List of mutants used in this study. E Coli numbering, reference structure = wildtype; pdb entry 2g8o.



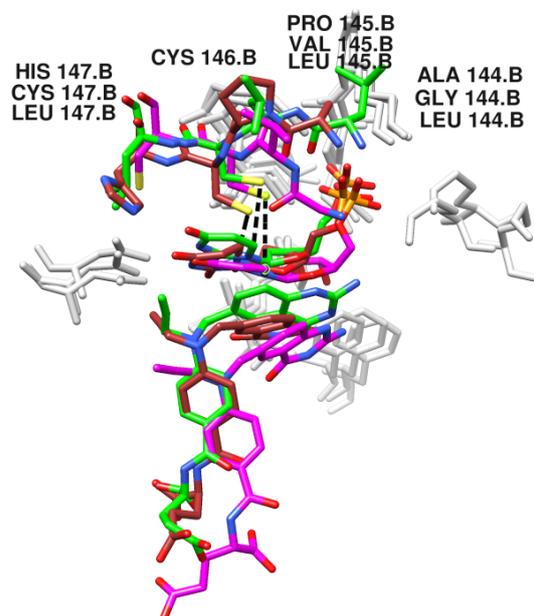
SS Figure 2. Temperature factor plot of the 3 ns simulation for the 264 residues in the A (left) and B (right) monomers.



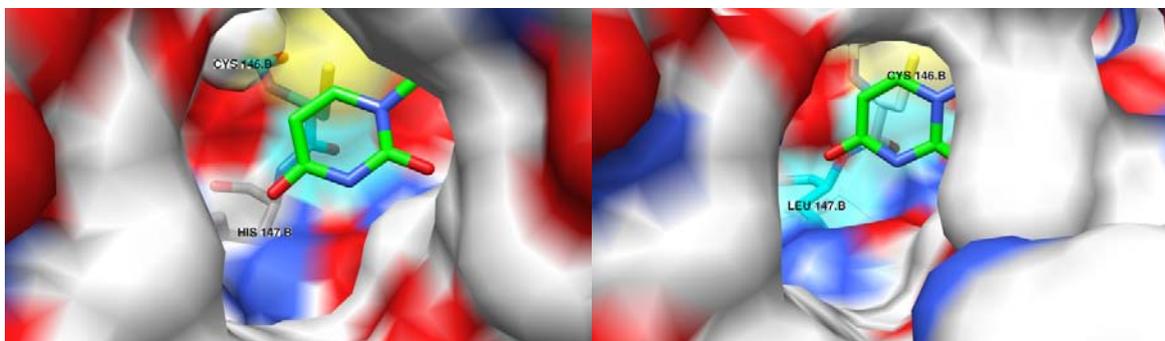
SS Figure 3. Distance of Cysteine-146B.SG atom to dUMP(B).C6 in the second monomer B during the 3 ns simulation of wild-type and 10 mutant structures.



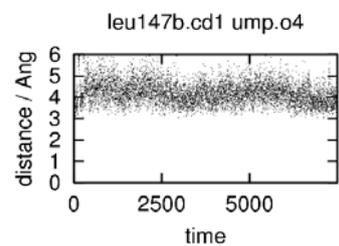
SS Figure 4. Superposition of 2 structures from the wildtype MD trajectory using Dali^[11] showing 2 different possible rotameric states of the cysteine-146 causing the change in distance from atom SG to dUMP atom C6 shown in SS Fig. 3. Figure generated using Chimera.^[12]



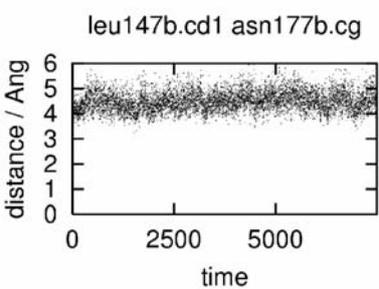
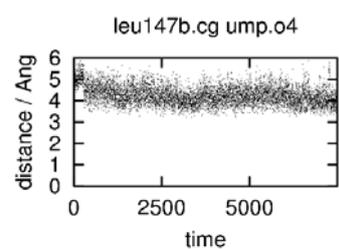
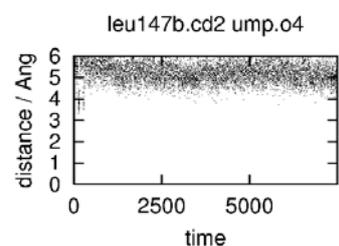
SS Figure 5. Superposition of the wt x-ray (brown carbons), the active A144G-P145V-C146-H147C (pink carbons) and the inactive mutant A144L-P145L-C146-H147L (green carbon atoms) structures from the MD trajectory. The UMP is in the middle with the cofactor in the bottom. For the inactive A144L-P145L-C146-H147L mutant, no significant difference in the position of the loop residues is observed after 3 ns of simulation time. Figure generated using Chimera.^[12]



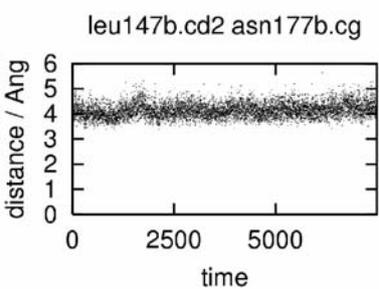
SS Figure 6. Close-up of the UMP binding site in the x-ray structure (left) and in the triple L mutant (right). The hydrophobic (non-aromatic) atoms of amino acids in the binding pocket are colored cyan. In the triple Leu mutant there are more contacts between this hydrophobic surface and the UMP polar base atoms. Figure generated using Chimera.^[12]



SS Figure 7. Distance of Leu-147B.CD1, CD2, and CG atoms to dUMP(B).O4 during the 3 ns simulation of the triple L mutant.



SS Figure 8. Distance of leu147b.CD1, CD2 atoms to asn177b.CG during the 3 ns simulation of the triple L mutant.



	X	WT	1	2	3	L	5	6	7	8	9	10
:903B@O2 :169B@H :169B@N	100	100	84	97	99	7	90	96	99	98	81	43
:903B@O4 :177B@HD21/HD22 :177B@ND2	100	53/11	85	99	99	-	94	100	100	97	91	48
:903B@N3:903B@H3 :177B@OD1	100	48	91	81	100	2	99	96	4	100	99	92

SS Table III. Hydrogen bond % occupied for selected donor, acceptor pairs during the 3 ns MD simulations. X=xray, WT=wildtype, 1=A144C-P145C-H147C, 2=A144G-P145V-H147C, 3=A144G-P145V-H147V, L=A144L-P145L-H147L, 5=A144V-P145C-H147C, 6=A144V-P145C-H147G, 7=A144V-P145C-H147V, 8=A144V-P145G-H147V, 9=A144V-P145V-H147C, 10= A144V-P145V-H147V. In the triple L mutant all 3 H-bonds are lost.

Templating of DNA synthesis by XNA-oligonucleotides in vivo

The gapped heteroduplex plasmid used for assaying in vivo XNA templates was constructed by digestion, denaturation and hybridization of the pAK1 and pAK2 plasmids.^[13] pAk-1 is a pTZ18R plasmid derivative which carries a deleted allele of the E coli thyA gene in which the 18 bp spanning from the NheI to the NsiI restriction site and encoding the active site around catalytic Cys146 has been replaced with a rifampicin resistance gene arr-2. PAK-2 was obtained from pAK 1 by ligating a 5' phosphorylated octonucleotide with sequence CTAGTGCA between the NheI and NsiI digested sites. This results in a deletion of 10 nucleotides in the thyA gene inactivating thymidylate synthase activity.

Formation of the gapped heteroduplex was performed by mixing equimolar amounts (50 ng) of pAK1, cleaved by NheI and NsiI and then purified on agarose gel, and of pAK2, linearized with EcoRI, in 10mM Tris-HCl (pH7.5) 100mM NaCl buffer, denaturing for 5 min at 85°C and cooling down to room temperature over 1 h. The reaction mix was dialyzed 30 min on nitrocellulose filter 0.05µm (Millipore) against water.

Each 18-mer oligonucleotide (10 pmol) to be assayed was hybridized with the gapped heteroduplex prepared as described above (0.02 pmol) in annealing buffer (50mM TrisHCl, 10mM MgCl₂, 10mM DTT, 25µg ml⁻¹ pH 8.0) by heat denaturation for 5 min at 85°C followed by 1 h at room temperature. Ligation was performed in the annealing buffer supplemented with 1mM ATP and 5Units of T4 DNA ligase, by incubating the mixtures overnight at 16°C. Ligation products were dialyzed on nitrocellulose filter for 30 min and used for transforming electrocompetent cells of the thyA-deleted strain G929. After incubating 1h at 37°C, transformed cells were plated on solid nutrient medium Mueller Hinton containing ampicilline (100 µg ml⁻¹) as well as on the same solid medium supplemented with 300µM thymidine.

SS REFERENCES:

- [1] M. J. Damha, N. Usman, K. K. Ogilvie, *Can J Chem* **1989**, *67*, 831-839.
- [2] C. Hendrix, H. Rosemeyer, I. Verheggen, F. Seela, A. VanAerschot, P. Herdewijn, *Chem-Eur J* **1997**, *3*, 110-120.
- [3] J. Wang, B. Verbeure, I. Luyten, E. Lescrinier, M. Froeyen, C. Hendrix, H. Rosemeyer, F. Seela, A. Van Aerschot, P. Herdewijn, *J. Am. Chem. Soc.* **2000**, *122*, 8595-8602.
- [4] M. K. Schlegel, E. Meggers, *The Journal of organic chemistry* **2009**, *74*, 4615-4618.
- [5] I. Verheggen, A. Van Aerschot, L. Van Meervelt, J. Rozenski, L. Wiebe, R. Snoeck, G. Andrei, J. Balzarini, P. Claes, E. De Clercq, et al., *J Med Chem* **1995**, *38*, 826-835.
- [6] S. L. Beaucage, M. H. Caruthers, *Curr Protoc Nucleic Acid Chem* **2001**, *Chapter 3*, Unit 3 3.
- [7] L. L. Zhang, A. E. Peritz, P. J. Carroll, E. Meggers, *Synthesis-Stuttgart* **2006**, 645-653.
- [8] X. J. Zhang, B. W. Matthews, *J. Appl. Crystallogr.* **1995**, *28*, 624-630.
- [9] J. M. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model.* **2006**, *25*, 247-260.
- [10] O. Adelfinskaya, M. Terrazas, M. Froeyen, P. Marliere, K. Nauwelaerts, P. Herdewijn, *Nucleic Acids Res.* **2007**, *35*, 5060-5072.
- [11] L. Holm, J. Park, *Bioinformatics* **2000**, *16*, 566-567.
- [12] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605-1612.
- [13] S. Pochet, P. A. Kaminski, A. Van Aerschot, P. Herdewijn, P. Marliere, *C. R. Biol.* **2003**, *326*, 1175-1184.