

Fidelity of DNA ligation: a novel experimental approach based on the polymerisation of libraries of oligonucleotides

J. Nicholas Housby* and Edwin M. Southern

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received April 30, 1998; Revised June 8, 1998; Accepted August 4, 1998

ABSTRACT

Complete libraries of oligonucleotides were used as substrates for *Thermus thermophilus* DNA ligase, on a M13mp18 ssDNA template. A 17mer primer was used to start a polymerisation process. Ladders of ligation products were analysed by gel electrophoresis. Octa-, nona- and decanucleotide libraries were compared. Nonanucleotides were optimum for polymerisation and up to 15 monomers were ligated. The fidelity of incorporation was studied by sequencing 28 clones (2268 bases) of nonanucleotide polymers, 12 monomers in length. Of the ligated monomers, 79% were the correct complementary sequence. In a total of 57 (2.5%) mispaired bases, there was a strong bias to G.T, G.A, G.G and A.G mismatches. Of the mismatches, 86% were found to be purines on the incoming oligonucleotide, of which 71% were G. There is evidence for clustering of mismatches within specific 9mers and at specific positions within these 9mers. The most frequent mismatches were at the 5'-terminus of the oligonucleotide, followed by the central position. We suggest that sequence selection was imposed by the ligase and not just by base pairing interactions. The ligase directs polymerisation in the 3' to 5' direction which we propose is linked to its role in lagging strand DNA replication.

INTRODUCTION

DNA ligases catalyse the formation of phosphodiester bonds in nicked duplex DNA molecules or between oligonucleotides which are in duplex with a complementary strand. Ligases form part of the replication machinery and are involved in other biological processes, such as repair of DNA damage, in which ligases may impart a degree of error checking. Most studies of the DNA replication process have centred around the fidelity of DNA polymerases. It is important to understand all the mechanisms which ensure faithful copying of DNA sequence at replication. Oligonucleotide substrates have been used previously to study the effects of mispaired bases on the ligation reaction (1–8). A complete study requires rate measurement of oligonucleotides with all possible

mismatches at each position in the sequence, and the measurements should be repeated for a number of different sequences.

Previous work from this laboratory has shown that *Thermus thermophilus* (*Tth*) DNA ligase can discriminate mispaired bases up to nine bases away from the join (7). To study all possible substitutions at all nine positions would require 27 rate measurements for each 9mer sequence studied.

The new method presented in this paper involves ligation of a library of oligonucleotides; all sequences are present in the library. The result is a polymer of oligonucleotides. Quantitative analysis of the ladder provides information on relative rates of ligation for a number of different sequences, those represented in the template.

Sequence analysis of cloned 12mers gives a measure of the fidelity of the incorporation and of the relative rates of incorporation of all possible mismatches. This is equivalent to the information provided by 324 separate ligations with single pairs of oligonucleotides.

In this study, we confirm that the length of oligonucleotide required by *Tth* ligase is eight or more bases. Fully complementary nonanucleotides are incorporated four times more frequently than those with one or more base mismatches. This is a remarkable degree of selectivity, given that the library comprises 262 144 different nonanucleotide sequences, and for each molecule of fully complementary sequence there are 27 nonanucleotides in the library with a single base mismatch.

MATERIALS AND METHODS

Synthesis of oligonucleotide primers and libraries

All DNA synthesis was performed using an Applied Biosystems 392 DNA/RNA synthesiser. All phosphoramidite chemistry materials were purchased from Cruachem. Complete libraries of oligonucleotides were synthesised by programming the synthesiser to add equimolar amounts of the four synthons at $N-1$ positions, where N is equal to the length of oligonucleotide. The 3'-terminal base is an equal mixture of dA, dC, dG and dT controlled-pore-glass support (Cruachem). All libraries were chemically phosphorylated at the 5' position with Phosphor-ON (Cruachem) at the termination of library synthesis. Oligonucleotides were subjected to ammonolysis deprotection for 16 h at 55°C and

*To whom correspondence should be addressed. Tel: +44 1865 275224; Fax: +44 1865 275283; Email: housby@bioch.ox.ac.uk

vacuum dried using a SpeedVac SVC-100 (Savant) and used with no further purification.

Phosphorylation of oligonucleotides

The oligonucleotide used to prime the ligation reaction was M13P; 5'-GTAAAACGACGGCCAGT-3'. This primer was designed to hybridise adjacent to the multiple cloning site of M13mp18 ssDNA (New England BioLabs). M13P was phosphorylated in a 20 μ l reaction containing: 6 pmol of M13P, 1 μ l 0.5 M [γ - 32 P]ATP (Amersham International), 10 U Polynucleotide Kinase (Boehringer Mannheim) and the corresponding buffer and H₂O to 20 μ l. The reactions were incubated for 45 min at 37°C followed by 10 min at 65°C to denature the kinase. The volume of the reaction was increased to 75 μ l with H₂O and the unincorporated radiolabel removed by centrifugation (3000 r.p.m.) through G-25 Sephadex (Pharmacia) for 4 min. Non-radioactive 0.5 M ATP was substituted in the above reaction where necessary. The reactions were then stored at -20°C until needed.

Temperature optimum for library ligation

A library consisting of 262 144 individual nonanucleotides was incorporated into the following ligation reaction: 15 fmol of M13mp18 ssDNA (New England BioLabs), 15 fmol of each individual nonanucleotide in the library (3.9 nmol total), 60 fmol of γ - 32 P-labelled M13P oligonucleotide, 60 fmol of unlabelled M13P oligonucleotide, 0.5 \times buffer [10 mM Tris-HCl, pH 8.3, 25 mM KCl, 5 mM MgCl₂, 0.5 mM EDTA, 0.5 mM NAD⁺, 5 mM DTT, 0.25% Triton X-100 (Advanced Biotechnologies Ltd)], 50 U *Tth* DNA ligase (Advanced Biotechnologies Ltd), in a total volume of 10 μ l. Exactly the same conditions were used for ligation using octa- and decanucleotide libraries which consisted of 65 536 and 1 048 576 individual oligonucleotide species, respectively. Control samples were lacking in either *Tth* DNA ligase, nonanucleotide library or M13mp18 ssDNA template. All samples were heated to 95°C for 1 min, centrifuged for 10 s (3000 r.p.m.) and immediately transferred to an aluminium heating block for ligation at various temperatures: 44, 46, 48 and 50°C, and for different periods of time: 30, 60, 120 and 240 min. The reaction was stopped by adding 2 μ l of 5 \times formamide gel loading buffer and the samples were either kept on ice or stored at -20°C until needed. Samples were electrophoresed in a 15% polyacrylamide gel at a constant power setting of 60 W. The gel was fixed in 10% methanol, 5% acetic acid for 30 min prior to drying. The resulting gel was then subjected to autoradiography by Fuji medical X-ray film or exposed to a storage phosphor screen (Fuji STIII) overnight, in the dark. Phosphor screens were scanned by a PhosphorImager 400A (Molecular Dynamics) and digital images were transferred to a Sun-4 workstation for analysis using xvseq (J.K.Elder, unpublished data).

PCR amplification of fragment F12

A ligation reaction was performed as described above for 48 h at 46°C followed by 15% polyacrylamide gel electrophoresis. The gel was not fixed and was exposed to autoradiography overnight, using radioactive markers to position the gel exactly. The band corresponding to 12 nonanucleotide additions was removed from the gel. The gel slice was soaked in 100 μ l of H₂O for 10 h followed by boiling for 15 min. The paper debris was then pelleted by centrifugation at 14 000 r.p.m. in a microcentrifuge

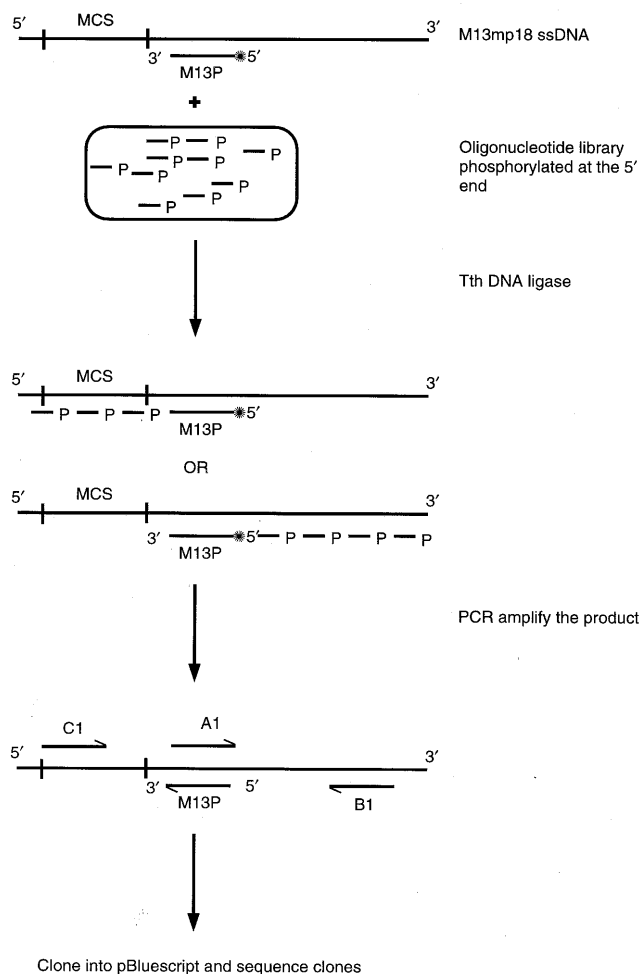


Figure 1. Ligation of library and cloning strategy. A schematic representation of the ligation of oligonucleotide libraries to M13mp18 ssDNA directed by the M13P primer. The 'star' on M13P represents the 32 P-radiolabel at the 5'-terminus. The ligation reaction is shown proceeding in both directions though experiments indicate that it proceeds only 3' to 5'. The band corresponding to the 12th nonanucleotide addition was eluted from a gel and PCR amplified using Vent DNA polymerase and the primers A1, B1, C1 and M13P, described in the text. The presence of PCR product indicated the 3' to 5' directionality of the ligation process. The PCR product was then cloned into pBluescript for sequence analysis to determine the type, position and number of mismatches generated during the ligation process.

and the supernatant removed to a sterile 1.5 ml Eppendorf tube. The DNA was precipitated using 10 μ l of sodium acetate and 450 μ l of ethanol, incubating at -80°C for at least 1 h. DNA was recovered by centrifuging in a microcentrifuge for 10 min at 14 000 r.p.m. and the supernatant discarded. The pellet was then washed in 200 μ l of 85% ethanol, air dried for 15 min and resuspended in 10 μ l of H₂O.

The four oligonucleotide primers used in PCR amplification were: A1 5'-AAGGGCGATCGGTGCGG-3', B1 5'-ACTGGC-CGTCGTTTTAC-3', C1 5'-TGAGCGGATAACAATTT-3' and M13P (see above). These primers were designed to amplify a DNA fragment, corresponding to the 12 nonanucleotide additions, in either of the possible directions for ligation (Fig. 1). The PCR reaction components consisted of 200 μ M dNTPs (Pharmacia),

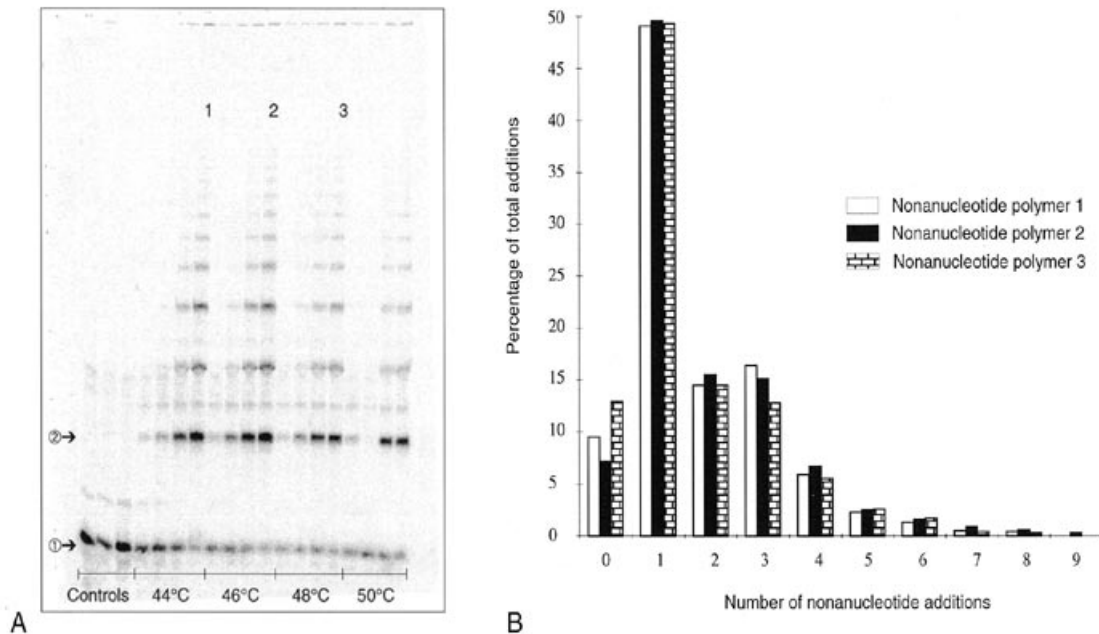


Figure 2. The effect of time and temperature upon ligation of nonanucleotides from a library. (A) A 15% polyacrylamide gel showing the effects of temperature and time of ligation on nonanucleotide polymer formation. The image is the result of a scan of a phosphor screen, using a PhosphorImager 400A (Molecular Dynamics), see text for details. Each of the three control samples is missing one constituent of the ligation reaction, from left to right; without *Tth* DNA ligase, M13mp18 ssDNA template and oligonucleotide library, respectively. For each temperature there are four time points, from left to right; 30, 60, 120 and 240 min. The arrow labelled 1 indicates the position of the M13P ³²P-labelled primer used to direct the ligation reaction. The arrow labelled 2 indicates the position of the first addition of nonanucleotide from the library. (B) The histogram shows the number of nonanucleotide additions, represented by the intensity of the corresponding bands on the gel (A), and is expressed as a percentage of the total intensity of all the bands. (A) 1, 2 and 3 indicate the ladders analysed, as described and represented in the histogram (B). Faint bands in positions off the ladder are of unknown origin and were ignored in this analysis.

0.4 μ M each primer, 2 U Vent DNA polymerase and the appropriate buffer (New England BioLabs), 4 μ l of fragment 12 template DNA, to a total volume of 50 μ l with H₂O. Cycling, using an MJ research PTC-200 Peltier Thermal Cycler, was; 92°C for 30 s, 46°C for 2 min and 72°C for 30 s, for a total of 30 cycles, followed by 72°C for 5 min. The sample (40 μ l) was run on a 1% agarose gel. The PCR amplified fragment was removed from the agarose with a razor blade and the DNA recovered using the QUIEXII DNA extraction kit (Quiagen). If, following PCR amplification and ethidium bromide agarose gel electrophoresis, there was barely any detectable PCR product visible, then a second round of amplification was performed with 1 μ l of the sample diluted 100-fold. As a control, to test the fidelity of PCR amplification and cloning, we used 250 ng of M13mp18 ssDNA in a typical PCR reaction as described above.

Cloning and sequencing of fragment F12

The vector pBluescript[®] SK (+/-) (Stratagene) was restriction endonuclease treated with *EcoRV* (Boehringer Mannheim) and phosphorylated as described above. Ligation of insert to vector was in a 1:1 ratio and incubation was at room temperature for 16 h. An aliquot of 2 μ l of the ligation reaction was transformed into INV α F' *Escherichia coli* cells (Invitrogen) and putative clones were selected, according to Invitrogen's instructions. Plasmid DNA was isolated (9) from each putative clone and sequenced using the T7 sequenase quick-denature plasmid sequencing kit (Amersham International)

with SK or KS oligonucleotide primers; SK 5'-CGCTCTAGAAC-TAGTGGATC-3', KS 5'-TCGAGGTCGACGGTATC-3', and using [α -³²P]dATP (Amersham International) as the radiolabel.

RESULTS

Ligation of an oligonucleotide library on M13mp18 ssDNA template

Our aim in this study was to compare the rates of ligation of fully complementary, phosphorylated oligonucleotides in the presence of all other sequences of the same length, and on a wide range of template sequences. The 'primer', M13P, a 17mer, was labelled with [γ -³²P]dATP and used to direct the ligation of oligonucleotides from the library. Analysis of the resulting ladder of oligonucleotides, by gel electrophoresis, allowed us to determine any effects on the rate of reaction that may result from specific sequences within the template. For example, if there were specific stalling points then the ladder would build up a specific band. The length of each ladder of oligonucleotides provides a sensitive measurement of the extent of ligation and enables optimisation of the reaction conditions: temperature, concentration of primer, template and library. Most studies of DNA ligase fidelity measure the joining together of two oligonucleotides on a longer template oligonucleotide (1-8). The method we used provides more information, as a large number of ligation reactions are studied simultaneously; no bias is introduced by differences in oligonucleotide concentration

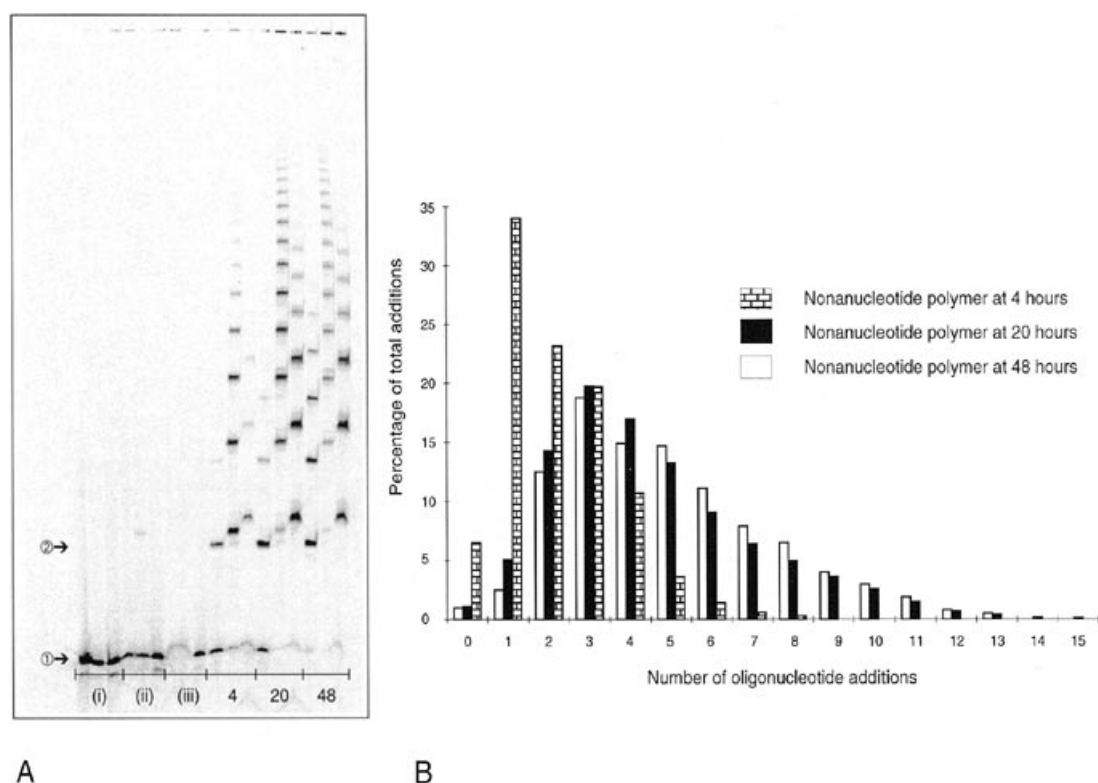


Figure 3. The effect of time and length of oligonucleotide on the production of octa- nona- and decanucleotide polymers. (A) A 15% polyacrylamide gel showing the effects of time and oligonucleotide length on polymer formation. The three control areas are labelled (i), (ii) and (iii) for octa- nona- and decanucleotides, respectively. Each of the control samples is missing one constituent of the ligation reaction, from left to right; without *Tth* DNA ligase, M13mp18 ssDNA template and oligonucleotide library, respectively. Ligations were carried out for 4, 20 and 48 h. From left to right for each time point, the lanes represent the ligations of octa-, nona- and decanucleotide libraries. The arrow labelled 1 indicates the position of the ^{32}P -labelled M13P primer used to initiate the ligation reaction. The arrow labelled 2 indicates the position of the first addition from the octanucleotide library. (B) The histogram shows the number of nonanucleotide additions, represented by the intensity of the corresponding bands on the gel (A), as a percentage of the total intensity of all the bands. Faint bands in positions off the ladder are of unknown origin and were ignored in this analysis.

(10) and it is the selectiveness of the ligase and/or the hybridisation reaction that determines which oligonucleotide will be added to the growing polymer.

Ligation products were analysed by electrophoresis on a 15% polyacrylamide gel. Ligations were carried out within the range 44–50°C; a temperature of 46°C gives the longest ladder of ligation products, up to nine additions from a nonanucleotide library, in 4 h (Fig. 2A). There was no ligation of monomers without DNA template. In an experiment using radiolabelled library, there was no ligation of monomers without the directing M13P 'primer' (data not shown). Individual band intensities were expressed as a percentage of the total intensity of all the bands (Fig. 2B). M13P (93%) was incorporated into oligomers at 4 h, increasing to 99% at longer ligation times.

For octa- and decanucleotide libraries, the optimum temperature for the formation of ligation ladders was also 46°C (data not shown), although the number of oligonucleotides incorporated was less than that for nonanucleotides at each point of a time course (Fig. 3A). The octa- and deca-nucleotide polymers increased in length over 48 h. The nonanucleotide library ligation reached a maximum at 20 h, and at this time point up to 15 nonanucleotide additions could be detected.

Sequence analysis of oligonucleotide polymers

The band corresponding to 12 additions of 9mers was selected by gel electrophoresis. PCR amplifications were directed to the strand made from ligated oligonucleotides in each of the possible directions of ligation (Fig. 1). Only primers A1 and B1 produced any detectable product implying that the ligation reaction proceeded to 12 additions in only one direction, 3' to 5'. The PCR product was cloned into pBluescript[®]. Twenty eight clones were sequenced (Figs 4 and 5). Only sequence data from 9 × 9mers is included for two reasons: (i) of the possible 12 additions, the last two are used for the PCR primer A1 and (ii) during the cloning procedure some deletions were created at the first addition. The control DNA clone proved to have no mismatches, as expected for high fidelity Vent polymerase (11). Fifty seven mismatches were detected among the 2268 bases sequenced from the 28 clones of ligated polymers.

We compared the observed frequency of mismatch at any one of the 81 positions in the polymer to that of random expectation (Fig. 6A). Mismatches were found at all nine positions of nonanucleotides. The largest number of mismatches occur at the 5'-terminus of the ligated oligonucleotide (Fig. 6B). The most

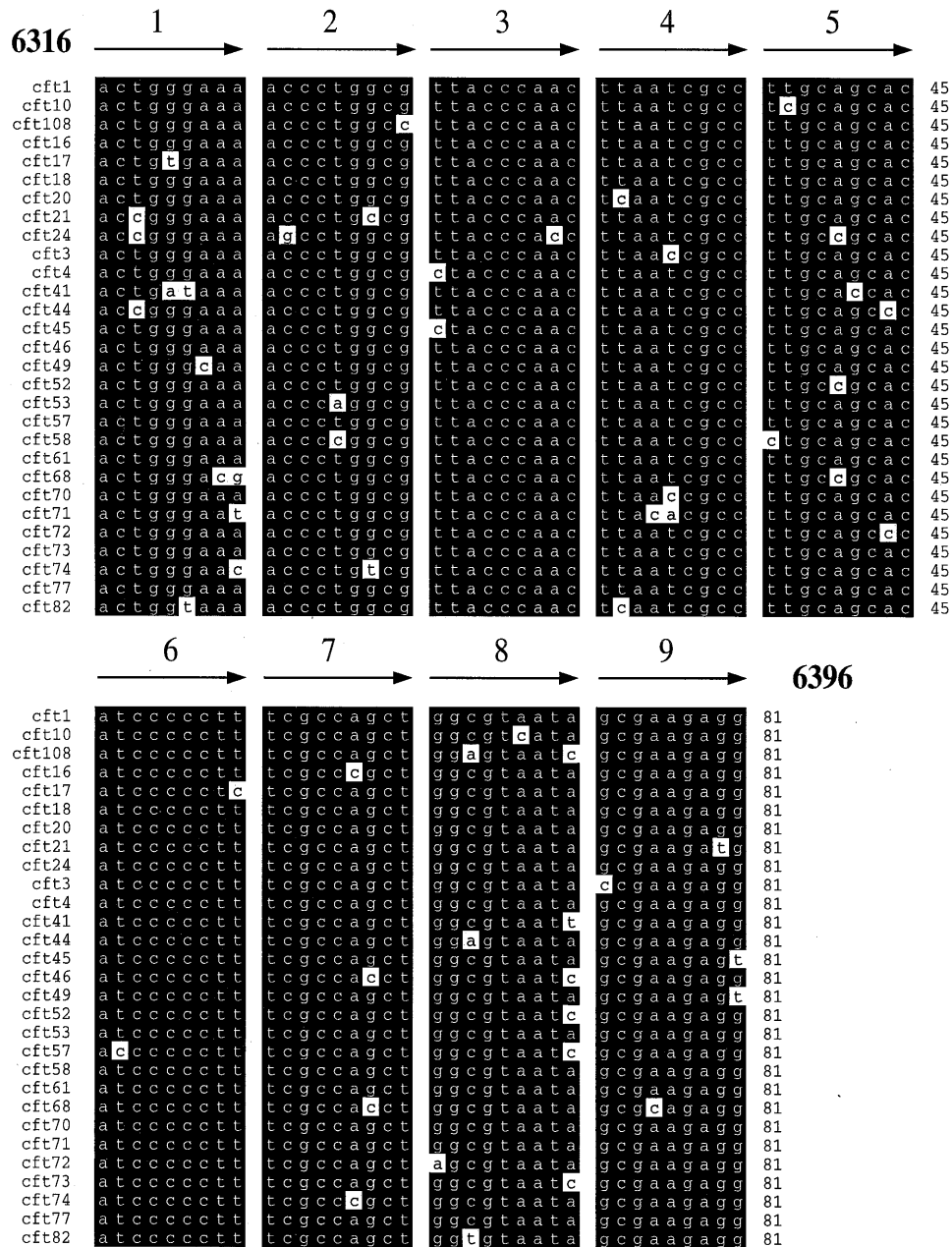


Figure 4. Sequence analysis of nonanucleotide polymers. The M13mp18 template sequence is shown between residues 6316 and 6396, in the 5' to 3' direction. The clones are all labelled with the prefix cft. cft1 is the sequence of the control to monitor the PCR reaction, as described in experimental procedures. The arrows indicate the direction and order of nonanucleotide addition. The shaded areas indicate complementary base pairing whilst the unshaded areas represent the position and type of mismatch.

frequent mismatch over all positions is G.A, which occurs 19 times (Fig. 5). At the 3'-terminus there is around one third the mismatch incorporation by comparison to the 5'-terminus. The most frequent mismatch at the 3'-terminus is G.T. There were a high number of mismatches at the central fifth position. Of all the nonanucleotides ligated, the correct one was ligated 79% of the time, which is surprising given the complexity of the library used.

DISCUSSION

Th has a very high ligation rate with large (>8mer) oligonucleotide substrates

The system we use to analyse ligase activity and specificity—polymerisation of a library of oligonucleotides initiated from a

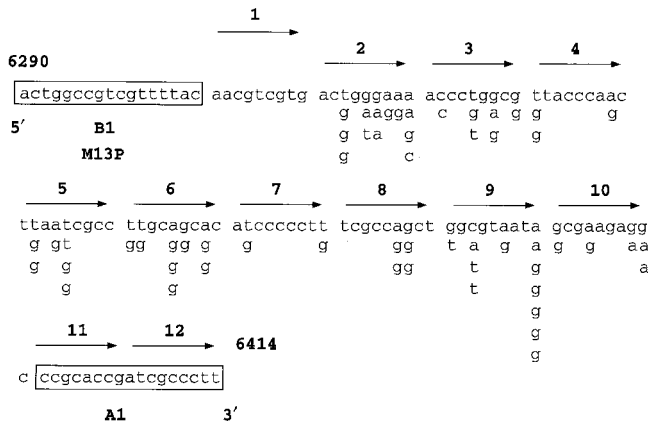


Figure 5. Distribution of mismatches generated during the ligation assay. The 5' to 3' (+)-template sequence of M13mp18 is shown here between residues 6290 and 6414. The primer used to initiate the ligation reaction, M13P, is depicted as the 17 bp boxed segment. B1 and A1 represent the primers for the PCR reaction used to clone this region of M13mp18. Nonnucleotides were ligated in the direction indicated by the arrows and are labelled 1–12. This indicates a directionality of 3' to 5' for the reaction as additions are to the 5'-terminus of M13P, see text for details. The PCR products of the 12th addition have been cloned and sequenced, see experimental procedures for details. The cloning produced some deletions in the area of the arrow labelled 1 and so data has been restricted to the second addition of nonanucleotide through to the 10th addition. Additions 11 and 12 are omitted as this sequence is used for PCR amplification. Below the template sequence is the mismatched base that has been incorporated with the nonanucleotide. In all, 57 mismatches were found in the 28 clones sequenced. Each mismatch is represented at the position and nonanucleotide in which it was found.

'primer' and a long template—differs significantly from the more usual system comprising pairs of oligonucleotides hybridised to a template. A number of questions can be addressed simultaneously and in a single experiment. Gel electrophoresis of the ligation products gives a 'ladder' of bands. The number of bands provides a sensitive measure of the rate of ligation which we used to compare the relative rates of ligation of oligonucleotides. The rate of incorporation of oligonucleotides is surprisingly high, given the low concentration of the oligonucleotides in the library. It is such that the primer 17mer, is incorporated completely into ligation products in a few hours, and most of the reaction products include more than one added oligonucleotide. Indeed, the longest product incorporates 15 nonanucleotides. Thus the method provides a more revealing analysis of ligase specificity than methods which examine ligation of pairs of oligonucleotides.

The library is constructed in a way that ensures that all sequences of a chosen length are represented in equal amounts, within narrow limits (10). This was confirmed by digestion of the library with phosphodiesterase and DNase I followed by treatment with phosphatase to obtain nucleosides. Analysis of the four nucleosides by HPLC showed similar ratios (data not shown).

The template may be envisaged as a series of linked oligomer substrates of differing sequence and composition. The relative rates of addition at each position can be measured from the intensities of the bands in the ladder. The measured distribution differs from what is expected if all oligonucleotides were added at the same rate (J.K.Elder, personal communication). Those added early are incorporated more quickly than those added later. It is possible that the effect is due to accumulated terminations on account of the

incorporation of mispaired oligonucleotides, discussed in more detail below, rather than an effect of specific base sequences. It is unlikely to be due to incomplete phosphorylation of oligonucleotides as different DNA ligases produce ladders with different lengths in this assay (J.N.Housby *et al.*, in preparation).

Mismatches on both sides of the join reduce ligation rate

The low extent of misincorporation shows that mismatched bases have a strong effect on ligation rate. Mismatched bases on both sides of the join appear to exert an effect (Fig. 4); the frequency of pairs of adjacent mismatched 9mers is lower than that expected from random distribution, suggesting that incorporation of a mismatch in one 9mer inhibits the incorporation of a mismatched 9mer in the next addition. That the ladder extends to 15 additions, with no obvious stalling points, indicates that the ligase can deal with a wide range of composition and sequence under the conditions used in these experiments.

Octanucleotide ladders did not extend as far as those for nonanucleotides which is thought to be attributable to the enzyme preference for longer oligonucleotides (12). Nonanucleotides produced the longest ladder and it is this length of oligonucleotide that we believe is optimal for the enzyme to exert its specificity for complementary incoming oligonucleotides. Decanucleotides extend beyond the region seen by the enzyme and may thus lead to the incorporation of oligonucleotides with mismatches at the terminal position, inhibiting ligation of the oligonucleotide at the next position almost completely (this study; 4,7). We are undertaking an analysis of the bases at the ends of each polymer and also DNA/protein binding studies to address these points.

Tth ligase discriminates between mismatched and fully complementary oligonucleotides

The selectivity of the reaction may be due to specificity of duplex formation or of the ligation reaction or, most likely, of both. Discrimination due to duplex formation alone is substantially lower than the discrimination seen in the present experiments (13). However, the reaction conditions used for duplex formation by hybridisation alone are considerably less stringent than those used with the enzyme. The mismatches in the ligated oligonucleotides are predominantly bases which form non-canonical base pairs: G.T, G.A, G.G and A.G (14). There is a significant bias (86%) towards purine mismatches on the incoming oligonucleotide. For example, there are 14 G.T mismatches compared to two T.G mismatches. This distribution indicates that it is the ligase rather than duplex formation which limits the incorporation of mispaired oligonucleotides, as there is no obvious reason why duplex formation should show such a bias.

The distribution of mismatches is non-random

A previous study into the fidelity of *Tth* DNA ligase (4), using a fluorescence based assay, detected only T.G and G.T mismatches at the 3' side of a nicked DNA. A bias to certain types of mismatch by DNA ligase has previously been described for Vaccinia virus (1), T4 DNA ligases (2,3) and DNA ligase I from *Saccharomyces cerevisiae* (5). Our study demonstrates strong biases in both the position and type of mismatched bases. There is some degree of clustering of mismatches to positions five and nine (Fig. 5). The clustering is partly due to a few 'hot-spots'; for example, there are five G.A and one A.A mismatches at the 9th position of

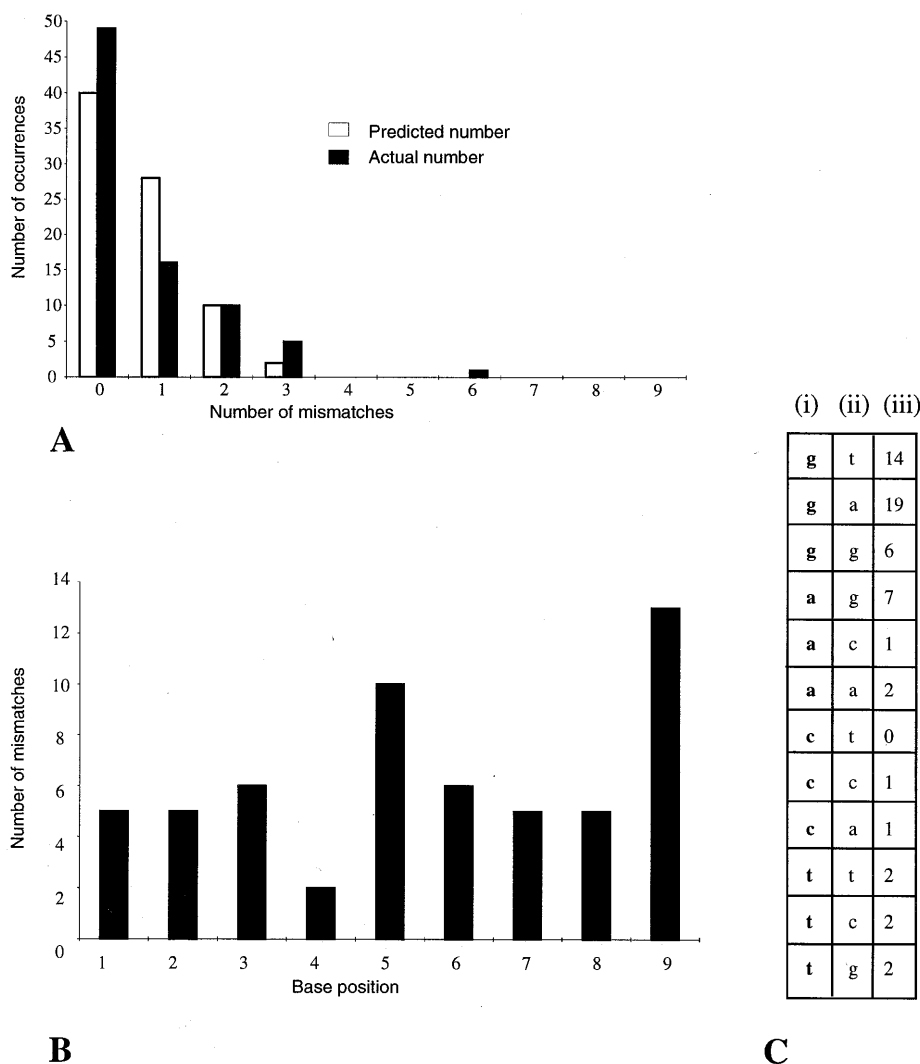


Figure 6. Analysis of mismatch base pairing by *Tth*. (A) The predicted number of mismatches occurring at any of the 81 positions in the polymer, assuming that the 57 mismatches were distributed randomly, is depicted in the histogram. The distribution is that corresponding to throwing 57 balls, at random, into 81 cells. The actual number is derived from the number of mismatch occurrences at any of the 81 positions within the 28 clones (see Fig. 5 for distribution analysis). The data indicates some bias towards clustering. (B) Cumulative distribution of mismatches at each of the positions in the nonanucleotides. The 3'-terminus of the nonanucleotide is equivalent to position 1 and the 5'-terminus is equivalent to position 9. (C) Frequency of the possible types of mismatch that could occur. (i) The base on the incoming nonanucleotide that pairs with (ii) the base on the M13mp18 template DNA. The total number of mismatches of each type is indicated in column (iii). Note the strong bias towards purines, especially G mispairings, on the incoming strand.

nonanucleotide 9 (Fig. 5); the fifth positions of nonanucleotide 5 and 6 each have three 'hits'. The random probability of having six mismatches at any one position is 5.5×10^{-3} (Fig. 6A), which highlights the non-randomness of the observed distribution.

Mispaired bases in the incoming nonanucleotide which cause the mismatch are strongly biased (86%) to purines. The bias towards purines is not explained by any evident feature in the target. Furthermore, 39 of the 57 mismatched bases were G (68%). We found no significant bias in the base composition of the nearest neighbours to the highest frequency mismatches in the relatively small sample. The high degree of mismatch at the fifth position is also unexpected. As it is known that mismatches at the middle of oligonucleotide duplexes have the greatest effect on duplex stability, this tolerance of mismatches at the fifth position must be due to the ligase.

A possible explanation for the bias is that the ligase could bind to the duplex in such a configuration as to impose strict specificity on the incoming oligonucleotides 3'-terminus whilst tolerating mismatches at the 5'-terminus and the central base. *In vivo*, the *Tth* DNA polymerase lacks a proof-reading function (4,15). Our observations support the view (4) that *Tth* DNA ligases may have evolved a high specificity for the 3'-terminus of DNA, whilst allowing endogenous 3' to 5' exonuclease activity to remove any incorrect base.

Biological implications

Because both library and 'primer' are phosphorylated, the ligation reaction could go in either a 5' or a 3' direction (Fig. 1), into or away from the multiple cloning site (MCS) of M13mp18.

PCR amplifications show that ligation to 12 nonanucleotides occurs only in the 3' to 5' direction. This was also shown by restriction enzyme digestion of ligation reactions, targeted to the MCS of M13mp18, when an octanucleotide library was used as substrate (unpublished data). This directionality makes sense in terms of the role of ligase in lagging strand DNA replication. Polymerisation of the lagging strand is primed by RNA primers and proceeds in a 5' to 3' direction. This results in the formation of 1000 bp Okazaki fragments which grow in a 3' to 5' direction. When a stretch between two fragments has been completed, the RNA primer is removed by the 5' to 3' exonuclease activity of DNA polymerase I, and the resulting gap is filled by its polymerase function using an adjacent Okazaki fragment as a primer. The resulting nick is then sealed by DNA ligase. Two conclusions can be drawn from this. First, as the process of DNA replication has to be highly discriminative in order to accomplish error free replication (16–18), DNA ligase may have evolved to be highly discriminative, sealing only correctly base paired DNA termini. Second, the direction of ligation seen in this study could reflect that seen in DNA replication, where the ligase joins incoming 3' ends of Okazaki fragments to the 5' end of one that has already been synthesised.

How the ligase interacts with its substrates to discriminate mismatched bases must have a structural basis. The crystal structure of the ATP dependent bacteriophage T7 DNA ligase (19) has been determined and reveals two domains. In between these two domains is a cleft that is likely to be the DNA binding region. The larger N-terminal domain has been shown to complex with ATP. The whole structure was found to be tightly folded and contains five conserved residues typical of a family of nucleotidyl transferases (see 20 for a review). It is these conserved residues that are thought to determine the specificity of DNA ligase for duplex DNA. Vaccinia DNA ligase has three structural domains as defined by partial proteolysis (21). The domain structure of vaccinia DNA ligase is similar to that of T7 DNA ligase bar the N-terminal domain of vaccinia ligase. The catalytic domains of both the T7 and the vaccinia DNA ligases are comparable in size. Although the crystal structure of NAD⁺ dependent DNA ligases has not yet been determined, we suggest that differing patterns of

discrimination by different ligases (e.g. *Tth*) may be due to variation in the domain structure and the number and position of conserved residues at the active site.

ACKNOWLEDGEMENTS

We thank Clare Pritchard for critically reading this manuscript; Martin Johnson for the construction of an aluminium heating block for ligations at various temperatures; John Elder for his statistical analysis for Figure 6; and Amersham International for supporting and funding this project.

REFERENCES

- Shuman,S. (1995) *Biochemistry*, **34**, 16138–16147.
- Harada,K. and Orgel,L.E. (1993) *Nucleic Acids Res.*, **21**, 2287–2291.
- Wu,D.Y. and Wallace,R.B. (1989) *Gene*, **76**, 245–254.
- Luo,J., Bergstrom,D.E. and Barany,F. (1996) *Nucleic Acids Res.*, **24**, 3071–3078.
- Tomkinson,A.E., Tappe,N.J. and Friedberg,E.C. (1992) *Biochemistry*, **31**, 11762–11771.
- Goffin,C., Bailly,V. and Verly,W.G. (1987) *Nucleic Acids Res.*, **15**, 8755–8771.
- Pritchard,C.E. and Southern,E.M. (1997) *Nucleic Acids Res.*, **25**, 3403–3407.
- Husain,I., Tomkinson,A.E., Burkhart,W.A., Moyer,M.B., Ramos,W., Mackey,Z., Besterman,J.M. and Chen,J. (1995) *J. Biol. Chem.*, **270**, 9683–9690.
- Evaluating and Isolating Synthetic Oligonucleotides—The Complete Guide. (1994) In *Applied Biosystems DNA/RNA Synthesiser Manual*. pp 2–4–2–5.
- Goode,B.L. and Feinstein,S.C. (1992) *Biotechniques*, **12**, 374–375.
- Cline,J., Braman,J.C. and Hogrefe,H.H. (1996) *Nucleic Acids Res.*, **24**, 3546–3551.
- Takahashi,M., Yamaguchi,E. and Uchida,T. (1984) *J. Biol. Chem.*, **259**, 10041–10047.
- Southern,E.M., Maskos,U. and Elder,J.K. (1992) *Genomics*, **13**, 1008–1017.
- Topal,M.D. and Fresco,J.R. (1976) *Nature*, **263**, 285–289.
- Lawyer,F.C., Stoffel,S., Saiki,R.K., Myambo,K., Drummond,R. and Gelfand,D.H. (1989) *J. Biol. Chem.*, **264**, 6427–6437.
- Loeb,L.A. and Kunkel,T.A. (1982) *Annu. Rev. Biochem.*, **52**, 429–457.
- Echols,H. and Goodman,M.F. (1991) *Annu. Rev. Biochem.*, **60**, 477–511.
- Kunkel,T.A. and Bebenek,K. (1988) *Biochimica Biophysica Acta*, **951**, 1–15.
- Subramanya,H.S., Doherty,A.J., Ashford,S.R. and Wigley,D.B. (1996) *Cell*, **85**, 607–615.
- Shuman,S. (1996) *Structure*, **4**, 653–656.
- Sekiguchi,J. and Shuman,S. (1997) *Nucleic Acids Res.*, **25**, 727–734.