Nucleic Acids Research, 1999, Vol. 27, No. 12 2545–2553

Determinants of nucleotide sugar recognition in an archaeon DNA polymerase

Andrew F. Gardner and William E. Jack*

New England Biolabs Inc., 32 Tozer Road, Beverly, MA 01915, USA

Received January 15, 1999; Revised and Accepted April 26, 1999

DDBJ/EMBL/GenBank accession no. M74198

ABSTRACT

Vent DNA polymerase normally discriminates strongly against incorporation of ribonucleotides, 3'-deoxyribonucleotides (such as cordycepin) and 2',3'-dideoxyribonucleotides. To explore the basis for this discrimination we have generated a family of variants with point mutations of residues in conserved Regions II and III and assayed incorporation of nucleotides with modified sugars by these variants, all of which were created in an exonuclease-deficient form of the enzyme. A Y412V variant incorporates ribonucleotides at least 200-fold more efficiently than the wild-type enzyme, consistent with Y412 acting as a 'steric gate' to specifically exclude ribonucleotides. The most striking variants tested involved changes to A488, a residue predicted to be facing away from the nucleotide binding site. The pattern of relaxed specificity at this position roughly correlates with the size of the substituted amino acid sidechain and affects a variety of modified nucleotide sugars.

INTRODUCTION

DNA polymerases have evolved to efficiently and faithfully replicate DNA. Fidelity requires the sensing of nucleotide base complementarity, as well as additional structural features of the sugar and heterocyclic base. The determinants for this discrimination have been explored through X-ray crystallography and biochemical studies using modified substrates, with both native and genetically engineered polymerases. From these studies, a picture of sugar recognition has begun to emerge (1).

Almost all DNA polymerases appear to be members of a superfamily, based on sequence and structural similarities. Amino acid similarities allow the classification of most DNA polymerases into three families, A, B and C, according to similarities with *Escherichia coli* DNA polymerases I, II and III, respectively (2,3). Of these enzymes, the Family A members are the best studied (4).

The Family B DNA polymerases, including numerous archaeal DNA polymerases, human DNA polymerase α and phage DNA polymerases from T4, RB69 and ϕ29, have been less well characterized. Mutational data are available for human DNA polymerase α (reviewed in 5), ϕ29 DNA polymerase (reviewed in 6,7) and T4 DNA polymerase (reviewed in 8) and this information has confirmed the functional significance of the conserved regions identified by amino acid sequence comparisons. Recently, the crystal co-ordinates for RB69 phage DNA polymerase have become available (9), allowing extrapolations of the spatial positioning of conserved residues in other Family B DNA polymerases.

Mutational studies of amino acid residues in conserved Region III (10) are consistent with the involvement of these residues in nucleotide binding (11–13). Region III, later expanded to include the Family A DNA polymerases and renamed Motif B (14; Fig. 1), is characterized by the motif KX_nYG, where X is an amino acid residue and *n* has the value 7 or 6 for Family A or B DNA polymerases, respectively. Tabor and Richardson (15) found in the case of several Family A DNA polymerases that swapping Tyr and Phe residues within Motif B (*E.coli* Pol I F762, *Thermus aquaticus* Pol I F667 and T7 DNA polymerase Y526) makes a profound difference in the efficiency of ddNTP incorporation, with Tyr promoting ddNTP incorporation. Similarly, Gao *et al.* (16) have described reverse transcriptase variants from Motif A, a region adjacent to Motif B in the three-dimensional crystal structure, that are better able to incorporate NTPs; a property shared by an *E.coli* Pol I mutant recently described by Joyce and co-workers (17).

We have examined nucleotide insertion by the Family B DNA polymerase from the hyperthermophilic archaeon *Thermococcus litoralis* (Vent™ DNA polymerase; 18), assaying genetic variants to probe the determinants of sugar discrimination. Incorporation of ribo-, 2'- and 3'-deoxyribo- and 2',3'-dideoxyribonucleotides was examined using variants with amino acid substitutions in Regions II and III. The greatest relaxation in specificity arose from substitution of an Ala residue within Region III.

MATERIALS AND METHODS

DNA modifying reagents and DNAs

Restriction endonucleases, DNA modifying enzymes, $Vent_R$ (exo^-) DNA polymerase, thermostable inorganic pyrophosphatase, deoxyribonucleotides, dideoxyribonucleotides and phage DNAs were from New England Biolabs (Beverly, MA) and were used in accordance with the supplier's recommendations. Synthetic oligonucleotides were provided by New England Biolabs Organic Synthesis Division (Beverly, MA). 5-Bromo-2'-deoxyuridine 5'-triphosphate and cordycepin (3'-deoxyadenosine 5'-triphosphate) were obtained

*To whom correspondence should be addressed. Tel: +1 978 927 5054; Fax: +1 978 921 1350; Email: jack@neb.com

from Sigma (St Louis, MO) and ribonucleotides from Amersham-Pharmacia Biotech (Piscataway, NJ). Radioactively labeled nucleotides were from NEN Life Science Products (Boston, MA). Primed M13mp18 template refers to primer 1224 (New England Biolabs) annealed to single-stranded M13 DNA, as described previously (19). ^{32}P-primed M13mp18 refers to the same substrate created with a 5′-end-labeled primer 1224. An additional substrate for monitoring incorporation of NMP was created by digesting 0.5 mg λ DNA with 80 U of *Sau*3AI endonuclease at 37°C overnight, followed by heating at 65°C for 20 min to inactivate *Sau*3AI, to generate 5′-GATC overhangs. The concentration of this substrate refers to the number of *Sau*3AI ends in the reaction.

Generation of Vent DNA polymerase variants

Numbering of Vent DNA polymerase amino acid residues refers to the mature polymerase (i.e. without inteins; 18; GenBank accession no. M74198). Vent DNA polymerase variants were expressed from derivatives of the plasmid pALK1 (19) that encodes an exonuclease-deficient form of the polymerase due to the dual mutation D141A/E143A [as contained in Vent$_R$ (exo$^-$) DNA polymerase]. To simplify notation in this paper, this exonuclease-deficient form is described as wild-type and indicated mutations are those present in addition to D141A/E143A. Genes encoding the variants were created by cassette mutagenesis, replacing either an *Mfe*I–*Nhe*I DNA fragment (encompassing residues P476–K490) or an *Nhe*I–*Bsu*36I DNA fragment (encompassing residues N494–Y502). For Y412 variants, pALK1 was modified by making silent changes in codons S347 (TCG→TCT, destroys a *Xho*I site), G398 (GGT→GGC) and L399 (TTG→CTG, creates a *Stu*I site), L426 (CTT→CTC) and E427 (GAA→GAG, creates a *Xho*I site), resulting in a plasmid designated pCFA1. These silent changes allowed cassette mutagenesis of residues 400–425 via replacement of a *Stu*I–*Xho*I fragment with a synthetic duplex containing the desired changes. Variants were sequenced using an automated PE-ABI DNA sequencer, with Dye-Deoxy™ terminators (PE-ABI, Foster City, CA) to verify changes.

Expression and purification of Vent DNA polymerase variants

Plasmids coding for the desired amino acid changes were transformed into strains BL21(DE3) (20), WJ56 or ER2566 for expression of the variant polymerase. WJ56 and ER2566 are derivatives of BL21 in which the T7 RNA polymerase has been integrated into the bacterial chromosome within the *lacZ* coding sequence (W.E.Jack, J.F.Menin, M.Sibley and E.Raleigh, unpublished results). In this strain, T7 RNA polymerase expression is under the control of the inducible *lac* promoter. Expression in these three strains was found to be comparable and the strains were used interchangeably in these experiments. Cultures (4 l) were grown at 37°C in LB medium (10 g tryptone, 5 g yeast extract, 10 g NaCl, 1 g dextrose, 1 g $MgCl_2 \cdot 6H_2O$ per liter, pH adjusted to 7.2) containing 0.1 mg/ml ampicillin. When the density of the culture reached a Klett reading of 200, isopropyl β-D-thiogalactopyranoside was added to a final concentration of 0.4 mM and incubation at 37°C was continued overnight with shaking. The culture was centrifuged at 4000 r.p.m. for 30 min at 4°C in a Beckman JA 4.2 rotor and the supernatant discarded. All subsequent steps, except where indicated, were carried out at 4°C. The cell pellet (25 g) was suspended in 100 ml buffer A [10 mM KPO_4 (pH 7.0), 1 mM dithiothreitol, 0.1 mM EDTA, 10% v/v glycerol] containing 0.075 M NaCl, heated at 80°C for 20 min and then centrifuged for 15 min at 10 000 r.p.m. in a Beckman JA-14 rotor. The supernatant was passed through a 10 ml DEAE–Sepharose column (Amersham-Pharmacia Biotech) and the flow-through was immediately loaded onto a 10 ml heparin–Sepharose column (Amersham-Pharmacia Biotech), both columns having previously been equilibrated with buffer A containing 0.075 M NaCl. Vent DNA polymerase variants were eluted from the heparin column with a 100 ml 0.075–0.9 M NaCl gradient in buffer A. Fractions of 0.75 ml were collected and assayed for DNA polymerase activity using a previously described acid precipitation assay (19). Peak fractions were pooled and dialyzed against Vent Storage Buffer [0.1 M KCl, 0.01 M Tris–HCl (pH 7.6), 1 mM dithiothreitol, 0.1 mM EDTA, 0.1% v/v Triton X-100, 50% v/v glycerol] and stored at –20°C. Protein concentrations for specific activity determinations were calculated by comparing Coomassie Brilliant Blue stained gel samples to known amounts of Vent DNA polymerase run on the same gel. Gels were scanned and quantified using a Microtek ScanMaker III attached to a Macintosh 8100 computer running NIH Image v.1.59 software (National Institutes of Health, Bethesda, MD).

```
Motif B
Family A consensus               G K x h N F G V L Y G
Family B consensus                 K h x x N - S L Y G     G
Vent DNA polymerase        Q R A I K L L A N - S Y Y G Y M G
   Variants:               E K C   A       D   A F F   F
                           L   S   R             L S   L
                           N   L   N
                               I
                               F
                               V
-------------------------------------------------------------
Motif A
Family A consensus                   D h S x I E L R
   DNA polymerase I          I V S A D Y S Q I E L R I M
   Vent DNA polymerase       I I Y L D F R S L Y P S I I V
Family B consensus                   D h x S L Y P S
   M-MuLV RT                 L D L K D - - - A F F C L R L H
                                             A F
RT consensus                         D h - - G Y
```

Figure 1. DNA polymerase conserved sequences. Consensus amino acid residues are shown for Motif B. The corresponding Vent DNA polymerase sequence is shown (residues 486–498), along with a listing of point mutants described in this work. x, a non-conserved amino acid residue; h, a hydrophobic residue; –, a gap (23). For Motif A consensus amino acid residues are shown for DNA polymerase Families A and B and reverse transcriptases (RT), along with representatives of each class. Residues displayed are: *E.coli* DNA polymerase I, residues 701–714; Vent DNA polymerase, residues 403–417; Moloney murine leukemia virus (M-MuLV), residues 149–161. Vent DNA polymerase residue Y412 is underlined. E710 of DNA polymerase I and F155 of M-MuLV appear in bold.

Sequencing assays

DNA polymerases were assayed by adding 2 U to reactions containing 10 nM ^{32}P-primed M13mp18 in 14 µl of 0.01 M KCl, 0.01 M $(NH_4)_2SO_4$, 0.02 M Tris–HCl (pH 8.8), 5 mM $MgSO_4$, 0.2% (v/v) Triton X-100. A 3.2 µl aliquot of this reaction mixture was distributed to A, C, G and T sequencing reactions, each of which contained 3 µl of the nucleotide mixture listed in Table 1. After incubation for 15 min at 72°C the reaction was halted by adding 4 µl Stop/Loading Dye (0.3% xylene cyanol FF, 0.3% bromophenol blue, 0.37% EDTA, pH 7.0) and heating at 100°C for 3 min. Reaction

products were separated by electrophoresis on a 6% urea/polyacrylamide gel in Tris–borate EDTA (21) buffer at 45 W. The gel was fixed by soaking in 5% acetic acid/5% methanol and dried. Alternatively, samples were separated on a Quick Point DNA sequencing gel (Novex, San Diego, CA) and processed as recommended by the manufacturer. In both cases samples were visualized by autoradiography.

Table 1. Deoxy/dideoxy sequencing mix nucleotide concentrations (μM)

	A Mix	C Mix	G Mix	T Mix
ddATP	900	–	–	–
ddCTP	–	480	–	–
ddGTP	–	–	400	–
ddTTP	–	–	–	720
dATP	30	30	30	30
dCTP	100	37	100	100
dGTP	100	100	37	100
dTTP	100	100	100	33

Values are from the CircumVent™ Thermal Cycle Dideoxy Sequencing Kit Instruction Manual, v.2.0, Appendix E (New England Biolabs).

The relative ability to use the chain terminator ddATP or cordycepin was measured by completing sequencing reactions as described above (i.e. saturating enzyme conditions) using a terminator:dATP nucleotide mixture which contained 50 μM dATP and varying concentrations of terminator. Reactions were overlaid with a drop of sterile mineral oil and incubated at 72°C for 20 min. The reaction was terminated and reaction products visualized as described above.

Assays for NTP incorporation

The incorporation of NTPs versus dNTPs was determined in buffer containing 0.1 μM λ/*Sau*3AI ends, 0.01 M KCl, 0.02 M Tris–HCl (pH 8.8), 0.01 M $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 100 U/ml thermostable inorganic pyrophosphatase, 1 μM NTP or dNTP and 40 μCi/ml [α-^{32}P]CTP or dCTP, respectively (NTP and dNTP concentrations denote the amount of each nucleotide in the mix). The reaction was initiated by adding polymerase to a final concentration of 200 U/ml and incubated at 72°C for 20 min, at which time 20 μl aliquots were removed and the amount of acid-insoluble material quantified as described above. Parallel samples without added enzyme were also analyzed and used to correct for background radiation in the assay. The relative incorporation is represented as the ratio of incorporation of ribonucleotides to deoxyribonucleotides. The limit of detection in these assays was taken as a corrected value equal to the background radiation. When this threshold was not reached, the relative incorporation is listed as being less than the ratio defined by this limit of detection.

Incorporation into a primed single-stranded substrate was similar: 1 U of polymerase was added to 17.5 μl of 0.01 M KCl, 0.02 M Tris–HCl (pH 8.8), 0.01 M $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 15 nM primed M13mp18, 50 μM NTP (or dNTP), 40 μCi/ml [α-^{32}P]CTP (or dCTP), 100 U/ml thermostable inorganic pyrophosphatase and incubated at 72°C for 15 min, following which acid-precipitable radioactivity was determined. Counts were corrected for background radiation as described above.

The ability to synthesize ribonucleotide-substituted products was measured by primer extension reactions in the presence of dNTP and varying amounts of NTP. DNA sequencing ladders were generated by alkali cleavage and separated on a denaturing polyacrylamide gel (22). Specifically, 3 μl of 0.02 μM ^{32}P-primed M13mp18, 0.02 M KCl, 0.02 M $(NH_4)_2SO_4$, 0.04 M Tris–HCl (pH 8.8), 4 mM $MgSO_4$, 0.2% (v/v) Triton X-100, 200 U/ml thermostable inorganic pyrophosphatase, 0.15 U polymerase was added to 3 μl of each NTP:dNTP mix, giving a final concentration of 0.05 mM dNTP and varying concentrations of NTP. Reactions were overlaid with a drop of sterile mineral oil and incubated for 20 min at 72°C. The resulting ribonucleotide-substituted extension products were cleaved by adding KOH to 0.1 M and incubating at 100°C for 20 min. Reactions were terminated and products visualized as described above.

Metal ion optima

DNA polymerase activity was assayed in the presence of varying Mg^{2+} or Mn^{2+} by adding 0.1 vol of a 10× $MnSO_4$ or $MgSO_4$ stock to reactions containing 62 U/ml polymerase, 15 nM primed M13mp18, 0.01 M KCl, 0.01 M $(NH_4)_2SO_4$, 0.02 M Tris–HCl (pH 8.8), 0.1% Triton X-100, 0.05 mM dNTP, 40 μCi/ml [α-^{32}P]dCTP and 100 U/ml thermostable inorganic pyrophosphatase. Reactions were incubated at 72°C for 15 min. Acid-insoluble material was quantified as described above.

RESULTS

Generation and purification of variants

Prior studies on Family A and B DNA polymerases implicated amino acid residues in Region II and Region III as being important for dNTP utilization (23). Accordingly, we mutated residues in these regions, emphasizing changes that introduced or removed potential hydrogen bond interactions, such as A488S, N494D and S495A. In addition to these changes, the potential involvement of aromatic residues in ddNTP discrimination (24), dNTP binding (5,25) and template binding (25,26) led us to make additional alterations in Tyr residues in Vent DNA polymerase Region III. An additional Region II Tyr, Y412, whose analog is near the catalytic site in RB69 (9) and is predicted to affect ribonucleotide insertion (1,9), was also mutated. All variants were created in an exonuclease-deficient form of the polymerase to eliminate exonuclease proofreading, thus accentuating misincorporation by the enzyme.

Variants were expressed in *E.coli* and purified by heat treatment of the cell extract, a process that obliterated host DNA polymerase activity and led to the loss of >80% of the host proteins (19). Two further column purification steps yielded proteins that were ~80% pure, as judged by Coomassie stained SDS–PAGE. All purified variants were active as DNA polymerases, although several had specific activities at least 10-fold lower than the wild-type enzyme (K490N, N494D and

Table 2. Relative activities of Vent DNA polymerase variants using nucleotides with modified sugars

Enzyme	Specific activity[a]	Gel assays[b]			Ppt assay
		ddNTP	Cordycepin	ATP	NTP[c]
Vent (exo⁻)	1.0	1	1	1	<0.0083
Region II					
Y412V	1.0	2	2	200	0.018
Y412L	0.77	1	1	4	0.014
Y412F	0.75	1	1	1	<0.007
Region III					
Q486E	2.5	0.25			0.013
Q486L	1.3	0.25			0.005
Q486N	0.89	0.25			<0.005
R487K	0.20	n.d.			
A488C	1.2	4	2		0.014
A488S	0.89	7	10		0.011
A488L	0.16	12	34	40	0.071
A488I	0.14	10			0.035
A488F	0.12	15	6		0.047
A488V	0.12	4	2		0.020
K490A	0.14	*			<0.018
K490R	0.12	*			<0.016
K490N	0.03	*			<0.021
N494D	0.06	*			
S495A	0.44	*			
Y496F	1.4	1		1	0.0095
Y496L	0.88				0.011
Y497S	0.50	3	1	2	0.007
Y497F	0.03	*			<0.01
Y499L	2.0	5		10	0.018
Y499F	1.2	1		2	
A488C/Y499F	1.4	1	2		
A488L/Y499L	0.06	5		20	0.077

[a]Specific activity is that of the variant relative to the wild-type enzyme.

[b]Derived from gel assays. Numerical entries indicate that the given variant produced clear sequencing gel patterns, similar to the wild-type enzyme, and additionally refer to the stoichiometric ratio of modified nucleotide required in the wild-type relative to the variant to give a similar extent of incorporation. Higher numbers indicate increased incorporation by the variant relative to the wild-type enzyme. Entries designated * produced a less clear sequencing pattern marked by faint bands in incorrect positions. The low recovery of R487K did not allow a determination for this variant (n.d.).

[c]Derived from acid precipitation incorporation assays. Numbers give the ratio of incorporation in the presence of NTPs relative to incorporation in the presence of dNTPs. In cases where the limit of detection was not reached, < indicates the limit of detection (see Materials and Methods).

Y497F; Table 2). Thus, significant portions of the polymerase active site must remain in each of these point mutants, although the decline in specific activity indicates disruption of the site.

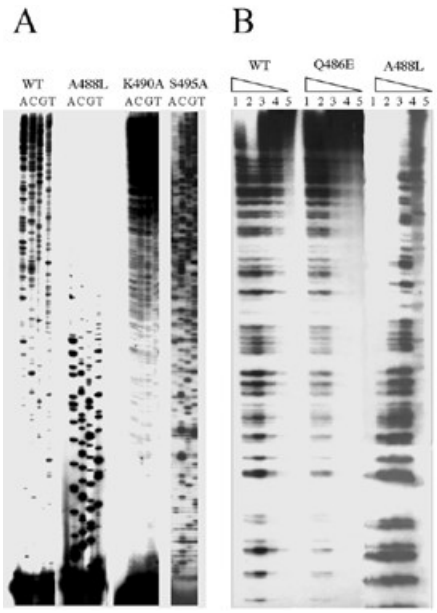

Figure 2. Sequencing gel assays for dideoxyribonucleotide incorporation. (**A**) Dideoxy sequencing reactions were performed as described in Materials and Methods. The polymerases used are representative of the classes observed: wild-type, A488L (shifted wild-type pattern), K490A and S495A (increased shadow bands). (**B**) Relative incorporation efficiencies were determined by titrations using dideoxyribonucleotides. Sequencing reactions were performed with the indicated variants using conditions described in Materials and Methods: [dATP], 0.05 mM; [ddATP]:[dATP], (lane 1) 300:1, (lane 2) 100:1, (lane 3) 33:1, (lane 4) 11:1 and (lane 5) 3.7:1.

Incorporation of dideoxyribonucleotides

Vent DNA polymerase variants were screened for the ability to accurately utilize 2',3' dideoxyribonucleotide (ddNTP) chain terminators by analyzing the banding patterns produced in sequencing reactions. In each case sufficient enzyme was added to completely extend the primer in the absence of ddNTPs. Patterns, therefore, reflect the ability to incorporate ddNTPs. Production of clear banding patterns without non-specific shadow bands signaled significant retention of base insertion fidelity and polymerization characteristics (Fig. 2A). A number of variants fit this criteria, yielding products similar to those obtained with the wild-type polymerase, including Y412L, Y412F, Y496F and Y499F (Table 2). As might be expected, this first class of variants displayed wild-type specific activities. A second group showed a less clear pattern marked by increased shadow band intensities, indicating occasional chain termination at incorrect positions (K490A, K490N, K490R, N494D, S495A and Y497F). These shadow bands could arise from pausing by the polymerase variants, perhaps due to an increased K_m for dNTPs, ddNTPs or both. The low specific activities noted for these variants may reflect such an increase (Table 2). Alternatively, extra bands could arise in part from misincorporation since not all shadow bands were abolished during a subsequent chase reaction involving addition of 0.25 mM unlabeled dNTPs (data not shown). A final set of variants displayed clear banding patterns, but a shift in the ability to incorporate dideoxyribonucleotides as demonstrated by the difference in average chain length of the banding patterns (Y412V, Y497S, Y499L and variants of Q486 and A488). This final pattern is that expected for variants able to sense nucleotide complementarity, but with altered sugar discrimination.

The relative ability to utilize ddNTPs was determined for this last group of variants. Sequencing assays were performed with varying ratios of ddATP to dATP. The concentration of ddATP required to give an equivalent banding pattern with the wild-type and variant enzymes was then compared (Fig. 2B). Using this method, a hierarchy of polymerase variants was inferred assuming lower amounts of ddATP indicated an increased relative ability to incorporate ddAMP: A488F > A488L > A488I > A488S ≈ Y499L ≈ A488V ≈ A488C ≈ Y497S ≈ Y412V > wild-type > Q486E ≈ Q486L ≈ Q486N (Table 2). The boost in incorporation with A488 variants was modest (at most 15-fold), but reproducible. Interestingly, the trend paralleled the size of the amino acid sidechain, with larger substituents being most tolerant towards ddNMP incorporation (F > L > I > S ≈ V ≈ C > A).

We tested the possibility that the most influential mutations would have additive effects by creating double mutants. Incorporation of ddNMPs by A488C/Y499F and A488L/Y499L paralleled that of Y499F and Y499L, respectively, even though the specific activity of A488L/Y499L more closely matched the A488L parent. Thus, no additive effects were seen with these specific variant combinations.

Cordycepin incorporation

Similar assays probed potential determinants of 3'-deoxyadenosine 5'-triphosphate (cordycepin) addition (Table 2). The pattern of cordycepin inhibition was similar to that of ddNTPs: A488L > A488S > A488F > Y412V ≈ A488V ≈ A488C > Y412L ≈ Y412F ≈ wild-type. The relative effects on incorporation were of the same magnitude as ddNMP incorporation (at most 34-fold). Comparison with the previous experiment indicated that cordycepin was not as readily incorporated as ddAMP.

Ribonucleotide incorporation

As a final permutation, we analyzed the ability of NTPs to act as substrates with the various polymerase variants. In this case, incorporation could be directly analyzed since the availability of the 3'-OH allowed chain elongation to proceed after ribonucleotide addition. We first examined incorporation of labeled CMP into the 5'-overhang created by *Sau*3AI digestion by comparing polymerization in the presence of NTPs or dNTPs. The relative incorporation of NMPs was clearly enhanced in a number of Vent DNA polymerase variants (e.g. A488L, A488F, A488I, A488V, Y412V and Y499L; Table 2). Since the GATC overhang required addition of GMP, AMP and UMP prior to labeled CMP, a terminal ribonucleotide can act as a substrate for ribonucleotide addition.

When the assay substrate was primed M13 DNA, these same variants appeared to be more able to incorporate NMPs than the wild-type enzyme, although the levels of incorporation were less than twice the background, limiting the confidence of such measures (data not shown). The cause of this limited incorporation was revealed by extending labeled primers on the M13 template in the presence of NTPs and separating the reaction products on a denaturing gel (Fig. 3). Such analysis revealed that wild-type and Y412V extension was sharply

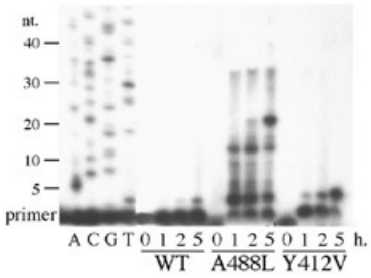

Figure 3. Ribonucleotide incorporation by A488L. Synthesis of RNA used 2 U of polymerase in a 40 µl reaction containing 0.01 µM ^{32}P-primed M13mp18, 0.1 mM NTP, 100 U/ml thermostable inorganic pyrophosphatase, 0.01 M KCl, 0.02 M Tris–HCl (pH 8.8), 0.01 M $(NH_4)SO_4$, 2 mM $MgSO_4$ and 0.1% Triton X-100. After incubation at 72°C for the number of hours indicated below each lane, 6 µl aliquots were removed and added to 14 µl Quick Point sample buffer (Novex) to stop the reaction. Reaction products were resolved alongside dideoxy sequencing reactions (lanes marked A, C, G and T) and detected as described in Materials and Methods.

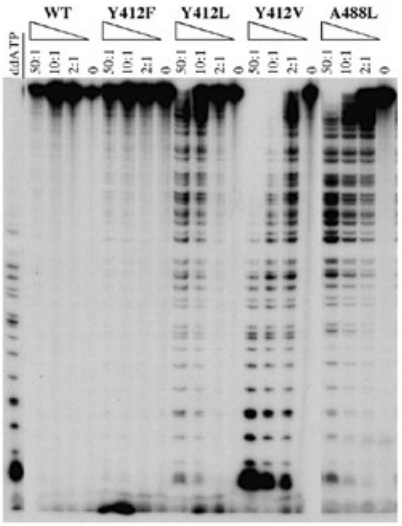

Figure 4. Gel assay of ribonucleotide incorporation. Indicated DNA polymerase variants (100 U/ml) were analyzed by alkali cleavage of primer extension products as described in Materials and Methods. Ratios indicate the concentration of ATP to dATP in the reaction. The first lane is a standard ddATP sequencing reaction.

limited after 2 nt, with the Y412V variant yielding a greater level of incorporation. By comparison, the A488L variant extended the primer at least 20 nt, although incorporation beyond that point dropped rapidly. The pattern of ribonucleotide incorporation suggests that only short tracts of ribonucleotides can be incorporated, echoing results obtained by Joyce and co-workers with variants of *E.coli* DNA polymerase I (17), although wild-type Vent DNA polymerase extensions were more limited than those observed for DNA polymerase I. This is the expected result if extended RNA/DNA hybrid structure, rather than polymerization *per se*, is the limiting factor in these reactions.

If the polymerase accommodates short, but not extended, tracts of RNA/DNA duplex, then it should be possible to replicate DNA with a dNTP pool doped with a limited amount of NTPs. This premise was tested by extending a labeled primer using a doped dNTP pool, visualizing the positions of ribonucleotide incorporation by hydrolyzing the phosphodiester linkages adjacent to ribonucleotide residues with alkali and separating the products on urea–acrylamide gels (Fig. 4). Banding patterns demonstrated the uniform incorporation of all four ribonucleotides and, once again, allowed deduction of a hierarchy of variants, this time with respects to ribonucleotide incorporation: Y412V > A488L > Y499L > Y412L ≈ Y412F ≈ Y496F ≈ Y497S ≈ Y499F ≈ wild-type. A comparison of concentrations required to give similar gel banding patterns revealed that ribonucleotides were less readily incorporated than either ddNMPs or cordycepin.

Acid precipitation assays measuring ribonucleotide addition to λ/*Sau*3AI-cut termini did not show as great a difference between the variants as seen in gel assays. Additionally, the hierarchy of ribonucleotide incorporation efficiency for the variants differed between these two assays. Since detection in the acid precipitation assay requires prior addition of GMP, AMP and UMP, this discrepancy likely reflects constraints imposed by the partial RNA/DNA duplex on addition of CMP (Fig. 3). Thus, these two assays measure different aspects of ribonucleotide addition.

Although incorporation of NMPs was vastly increased in selected variants, the four ribonucleotides were not equivalently incorporated. For example, incorporation of UMP by A488L was ~10-fold less efficient than incorporation of AMP. A similar bias against incorporation of dUMP by the wild-type enzyme (27) led us to test whether the absence of a 5-substituent in UTP accounted for the lack of dUTP utilization. When 5-bromouridine 5'-triphosphate was used in place of UTP, incorporation was similar to that observed when ATP was the substrate in both gel and acid precipitation assays (data not shown). Accordingly, the sensing mechanism for the 5-substituent of the pyrimidine base is still active in the A488L mutation. Lasken *et al.* (27) also noted that DNA templates containing uracil acted as strong competitive inhibitors for archaeal DNA polymerases, an observation that could account for the lack of extensive RNA synthesis on the M13 template. This seems unlikely since substitution of UTP by 5-bromouridine 5'-triphosphate did not increase the synthesis observed on the M13 template.

Metal ion utilization

Optimal Mg^{2+} and Mn^{2+} concentrations for A488L polymerization were slightly lower than those of the wild-type enzyme, differing more noticeably in the sharper optima of both A488L curves (Fig. 5). This similarity suggests that the metal-binding center is intact in A488L, as might be expected from the separation of the analogous residue and the metal-binding site in the RB69 crystal structure (N558; 9). Whether this narrowing of the metal optima reflects altered ability to bind to a substrate/metal chelate or metal-induced structural changes in the enzyme is unclear. Earlier studies with T7 DNA polymerase showed an increased utilization of ddNTP in the presence of Mn^{2+} when compared to Mg^{2+} (24). Both the wild-type and A488L variant showed only a slight increase in ddNMP incorporation when incubated in Mn^{2+} as opposed to Mg^{2+} (data not shown).

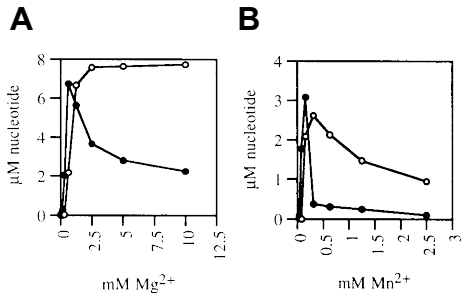

Figure 5. Metal ion optima for A488L. Nucleotide incorporation into a primed M13mp18 substrate was assayed by acid precipitation as described in Materials and Methods. Wild-type enzyme, open circles; A488L, closed circles. (**A**) Mg^{2+} titration; (**B**) Mn^{2+} titration.

DISCUSSION

We set out to create DNA polymerase variants that would more effectively incorporate ribonucleotides, 2',3'-dideoxyribonucleotides or 3'-deoxyribonucleotides. The availability of comparative sequence information, kinetic analysis of mutant enzymes and three-dimensional structural information from several other polymerases provided a framework on which to design such variants. Joyce earlier pointed out (1,9), on the basis of studies on Moloney murine leukemia virus reverse transcriptase (16) and *E.coli* DNA polymerase I (17), that the invariant Tyr residue in Region II of Family B DNA polymerases (Y412 in Vent DNA polymerase) could potentially play a role in 2'-deoxyribonucleotide/ribonucleotide discrimination. Results reported above with Y412 variants are entirely consistent with that prediction, with specific variants at this position affecting NMP, but not ddNMP, incorporation. The hydroxyl group of Tyr is not the major determinant for ribonucleotide exclusion as variant Y412F showed the same discrimination against NMP addition as the wild-type enzyme. Ribonucleotide incorporation was increased by a Y412L substitution and was even greater in the Y412V variant, reinforcing the view that the bulk of the larger amino acid group acts as a 'steric gate' to block access of the 2'-OH of the ribonucleotide sugar to the binding site. This steric clash could be directly with the 2'-OH or a metal ion chelated to this hydroxyl or via repositioning of neighboring residues in the nucleotide binding site.

The enhancement of ribonucleotide incorporation with Y412V in the gel assay is ~200-fold relative to the wild-type enzyme, similar to the 300-fold effect seen for F155V in Moloney murine leukemia virus reverse transcriptase (16) and the 800-fold effect observed with the E710A variant of *E.coli* DNA polymerase I (17). In the case of Vent DNA polymerase, this discrimination is specific for the 2'-OH, as dideoxyribonucleotides and cordycepin incorporation was only slightly enhanced with Y412V and not affected with the Y412L and Y412F variants.

Variants at the position analogous to Y412 have been constructed in ϕ29 DNA polymerase and human DNA polymerase α, both of which are Family B DNA polymerases. Unfortunately, in neither case was the incorporation of ribonucleotides analyzed. Dideoxyribonucleotide incorporation was evaluated for human DNA polymerase α, however, showing no effect with a Y865F variant and increased incorporation with a Y865S variant (28). Although these qualitative assays are not directly comparable with those presented here, these variants appear to have little effect on ddNMP incorporation, similar to the analogous Vent DNA polymerase Y412F and Y412V variants. Since Phe substitution at this point has little effect on ribonucleotide or dideoxyribonucleotide incorporation, the hydroxyl group of this conserved Tyr appears to play at best a minor role in discrimination of 2' or 3' sugar determinants. Despite this lack of sugar recognition, the K_m for dNTPs with both human DNA polymerase α Y865F and the analogous ϕ29 variant Y254F have increased ~10-fold (12), suggesting a role for this same Tyr hydroxyl group in nucleotide binding. Saturno *et al.* (29) argue that the magnitude of the K_m effect, together with a modest decrease in fidelity, establish the Y254 residue as being a key component in the selection strategy. This could occur, as Astatke *et al.* (17,30) pointed out, via chelation of a metal by this hydroxyl. Similar K_m effects may also be operative with these Vent DNA polymerase variants, but masked in these assays where nucleotide concentrations exceed the wild-type K_m (60 µM; 19) and polymerase concentrations are saturating. The combined data from several analogous DNA polymerases cast this conserved Region II Tyr in a role in both ribonucleotide exclusion and nucleotide binding.

While structural similarities in Region II were predictive for Vent DNA polymerase Y412 variant behavior, in other cases comparisons were not as instructive. We were initially impressed by the profound differences in ddNMP incorporation observed in Family A DNA polymerases following a single amino acid substitution in Motif B (15). Unfortunately, there was no obvious homolog of this F/Y residue in the analogous Region III of the Family B DNA polymerases, even though in both cases these amino acid residues form α-helices bordering the nucleotide binding site. In fact, mutations that replace Tyr with Phe in these regions of Vent DNA polymerase were either unchanged (Y412F, Y496F and Y499F) or somewhat decreased (Y497F) in the ability to incorporate dideoxyribonucleotides. A similar decrease in incorporation of ddNMPs was noted for an analogous substitution in ϕ29 DNA polymerase (Y390F in ϕ29 is analogous to Y497F in Vent DNA polymerase; 12). Curiously, Y497S retains wild-type characteristics, although it is not clear how Ser substitution compensates for the loss of Tyr. These Tyr substitutions do not have the dramatic effect towards dideoxyribonucleotide incorporation observed in Family A DNA polymerases. Thus, while some parallels can be drawn within the DNA polymerase superfamily, there are still difficulties in making precise extrapolations between Family A and B DNA polymerases.

In contrast to Y412 point mutations, substitutions of A488 affected both NTP and ddNTP utilization. Whereas increasing the size of the A488 sidechain facilitated insertion of both NMPs and ddNMPs, decreasing the size of the Y412 sidechain only enhanced addition of NMPs. The analogous A488 residue in the RB69 apo-enzyme crystal structure is part of an α-helix and faces away from the active site of the protein (Fig. 6), in contrast to the conserved Lys, Asn, Ser, Tyr and Gly residues in this motif, which face towards the proposed nucleotide binding site. Somewhat weaker global effects are also seen with Y499 variants, also predicted to be facing away from the nucleotide-binding site on the same α-helix. In this case,

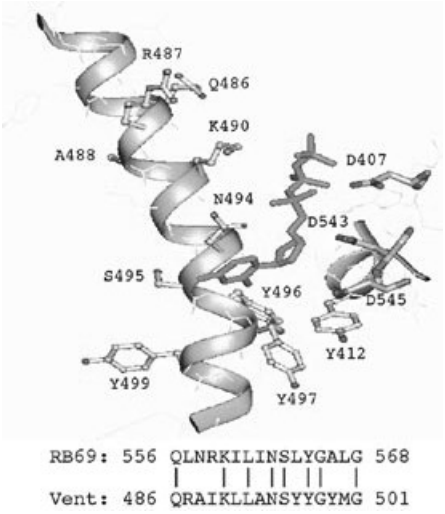

Figure 6. A model for Regions II and III of Vent DNA polymerase based on the crystal structure of RB69 DNA polymerase. Relevant coordinates from the crystal structure of RB69 DNA polymerase, including the modeled dCTP (Brookhaven PDB deposit 1WAH), were extracted and the displayed residues were mutated to the corresponding residues in Vent DNA polymerase, as indicated below the diagram. Individual sidechain orientations were energy minimized using a rotamer library within the program QUANTA (Molecular Simulations, San Diego, CA). Residues mutated in this study are indicated as ball-and-stick figures and are labeled. The substrate dCTP is shown in green. The catalytic triad of aspartate residues (not studied in this work) are D407, D543 and D545.

decreasing the bulk of the residue accentuated ddNMP and NMP incorporation. What indirect effects can account for greater tolerance for substituted nucleotides in these variants?

Several observations may be relevant. First, sugars differing at both the 2' and 3' position from the natural substrate are more effectively incorporated by the A488 variants. Preliminary experiments show that these variants are also more tolerant toward incorporation of nucleotides with substituted bases (A.F.Gardner, W.E.Jack and J.Killian, unpublished results). Thus, A488 variants appear to facilitate polymerization globally, rather than targeting specific nucleotide determinants. Second, unlike Y412 and analogous DNA polymerase I variants, the A488L variant allows for a more extensive tract of ribonucleotide polymerization, indicating increased tolerance for the inserted nucleotide and also for binding of the extended RNA/DNA product. Third, increased incorporation of modified nucleotides has a rough correlation with the bulk of the sidechain inserted in place of A488 or Y499. Assuming that Vent DNA polymerase adopts an α-helical conformation analogous to RB69 in Region III, one possible explanation for these observations is a lowered activation energy for a conformational change required prior to polymerization. This reduction may originate from the displacement of the helix in the direction required for catalysis due to the increased bulk of the substituted A488 sidechain or the decreased bulk of the Y499 residue. This movement may be similar to that seen in the T7 DNA replication complex, where the fingers subdomain is rotated inwards by ~41° towards the primer–template, bringing the O-helix (containing amino acid residues in Motif B) in closer proximity to the nucleotide-binding site than is observed in the apo-enzyme (25). Kiefer *et al.* (26) also note that in *Bst* DNA polymerase a conformational shift of the O-helix would be required to access the template base prior to addition. Alternatively, the substitutions may alter the flexibility of the helix, again facilitating the conformational change thought to be required for catalysis. Double A488/Y499 mutations do not accumulate the effect of the single mutations, suggesting that there are limits to this motion.

Kinetic studies of Family A DNA polymerases have demonstrated a rate-limiting step between nucleotide binding and polymerization (31), with a suggestion that this step may correspond to a conformational change. Astatke *et al.* (30) further argue that ddNTP discrimination occurs at this step. Perhaps the series of A488 variants facilitates this change and in so doing weakens some of the normal determinants of nucleotide selectivity. Since fidelity is coupled to this selectivity, A488 variants might be expected to exhibit lowered fidelity, a prediction that has not yet been tested. It may also be possible to create a second site revertant by decreasing the bulk of residues that contact A488L in the Vent DNA polymerase three-dimensional structure.

A unified picture of polymerization is emerging as a wider variety of polymerases are identified, their primary and secondary sequences determined and biochemical properties characterized. The first Family B-type DNA polymerase crystal structure, that of RB69 DNA polymerase, forecasts the structure of Vent DNA polymerase. However, significant differences in primary sequence are found between the two polymerases. Definitive answers to the roles of specific residues await a detailed structure for Vent DNA polymerase, preferably in a catalytically competent complex.

ACKNOWLEDGEMENTS

We are grateful to Jiming Wang for his insight into Family B DNA polymerase structure, in particular how A488 variants might influence polymerase selectivity. We thank Lorena Beese for discussions and sharing of unpublished data. Comments by Fran Perler, Christopher Noren, Richard Roberts, Paul Riggs and Huimin Kong on this manuscript were extremely useful and are gratefully acknowledged. J. Menin provided several of the constructs utilized here, B. Slatko, J. Ware, L. Mazzola and M. Ganatra provided DNA sequencing services and Chris Benoit assisted with polymerase isolation. We are particularly grateful for the ongoing support of Don Comb.

NOTE ADDED IN PROOF

Apo-enzyme crystal structures for two archeal hyperthermophilic Family B DNA polymerases are now available (L.Beese, personal communication; Hopfner,K.-P., Eichinger,A., Engh,R.A., Laue,F., Ankenbauer,W., Huber,R. and Angerer,B. (1999) Proc. Natl Acad. Sci. USA, 96, 3600–3605). In these structures Region II forms an α-helix, similar to the RB69 DNA polymerase structure, with residues corresponding to A488 and Y499 in Vent DNA polymerase on one face of the helix and conserved Lys, Asn and Ser residues facing the proposed active site on the opposite face.

REFERENCES

1. Joyce,C.M. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 1619–1622.
2. Jung,G., Leavitt,M.C., Shieh,J.-C. and Ito,J. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 8287–8291.
3. Ito,J. and Braithwaite,D.K. (1991) *Nucleic Acids Res.*, **19**, 4045–4057.
4. Brautigam,C.A. and Steitz,T.A. (1998) *Curr. Opin. Struct. Biol.*, **8**, 54–63.
5. Copeland,W.C., Dong,Q. and Wang,T.S.-F. (1995) *Methods Enzymol.*, **262**, 294–303.
6. Blanco,L. and Salas,M. (1995) *Methods Enzymol.*, **262**, 283–294.
7. Blanco,L. and Salas,M. (1996) *J. Biol. Chem.*, **271**, 8509–8512.
8. Reha-Krantz,L.J. (1995) *Methods Enzymol.*, **262**, 323–331.
9. Wang,J., Sattar,A.K.M.A., Wang,C.C., Karam,J.D., Konigsberg,W.H. and Steitz,T.A. (1997) *Cell*, **89**, 1087–1099.
10. Wong,S.W., Wahl,A.F., Yuan,P.M., Arai,N., Pearson,B.E., Arai,K., Korn,D., Hunkapiller,M.W. and Wang,T.S.-F. (1988) *EMBO J.*, **7**, 37–47.
11. Dong,Q. and Wang,T.S.-F. (1995) *J. Biol. Chem.*, **270**, 21563–21570.
12. Blasco,M.A., Lázaro,J.M., Bernad,A., Blanco,L. and Salas,M. (1992) *J. Biol. Chem.*, **267**, 19427–19434.
13. Blasco,M.A., Lázaro,J.M., Blanco,L. and Salas,M. (1993) *J. Biol. Chem.*, **268**, 16763–16770.
14. Delarue,M., Poch,O., Tordo,N., Moras,D. and Argos,P. (1990) *Protein Eng.*, **3**, 461–467.
15. Tabor,S. and Richardson,C.C. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 6339–6343.
16. Gao,E., Orlova,M., Georgiadis,M.M., Hendrickson,W.A. and Goff,S.P. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 407–411.
17. Astatke,M., Ng,K., Grindley,N.D.F. and Joyce,C. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 3402–3407.
18. Perler,F.B., Comb,D.G., Jack,W.E., Moran,L.S., Qiang,B., Kucera,R.B., Benner,J., Slatko,B.E., Nwankwo,D.O., Hempstead,S.K., Carlow,C.K.S. and Jannasch,H. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5577–5581.
19. Kong,H., Kucera,R.B. and Jack,W.E. (1993) *J. Biol. Chem.*, **268**, 1965–1975.
20. Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
21. Peacock,A.C. and Dingman,C.W. (1968) *Biochemistry*, **7**, 668–674.
22. Barnes,W.M. (1978) *J. Mol. Biol.*, **119**, 83–99.
23. Joyce,C.M. and Steitz,T.A. (1995) *J. Bacteriol.*, **177**, 6321–6329.
24. Tabor,S. and Richardson,C.C. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 4076–4080.
25. Doublié,S., Tabor,S., Long,A.M., Richardson,C.C. and Ellenberger,T. (1998) *Nature*, **391**, 251–258.
26. Kiefer,J.R., Mao,C., Braman,J.C. and Beese,L.S. (1998) *Nature*, **391**, 304–307.
27. LaskenR.S.,, Schuster,D.M. and Rashtchian,A. (1996) *J. Biol. Chem.*, **271**, 17692–17696.
28. Dong,Q., Copeland,W.C. and Wang,T.S.-F. (1993) *J. Biol. Chem.*, **268**, 24163–24174.
29. Saturno,J., Blanco,L., Salas,M. and Esteban,J.A. (1995) *J. Biol. Chem.*, **270**, 31235–31243.
30. Astatke,M., Grindley,N.D.F. and Joyce,C.M. (1998) *J. Mol. Biol.*, **278**, 147–165.
31. Kuchta,R.D., Benkovic,P. and Benkovic,S. (1988) *Biochemistry*, **27**, 6716–6725.