Nucleic Acids Research, Vol. 19, No. 15 4045–4057

Compilation and alignment of DNA polymerase sequences

Junetsu Ito and Dan K.Braithwaite
Department of Microbiology and Immunology, College of Medicine, University of Arizona Health Sciences Center, Tucson, AZ 85724, USA

Received May 29, 1991; Revised and Accepted July 9, 1991

INTRODUCTION

More than 40 different DNA polymerases, including some putative DNA polymerase sequences deduced from nucleotide sequence data, have recently been reported (1–39). The amino acid sequences of these DNA polymerases have been aligned and partial homologous regions identified by many investigators (2–4,9,10,12–25,27–36,42–51). Based on the segmental amino acid sequence similarities, DNA polymerases have been classified into two major groups; *E. coli* DNA polymerase I-Type and eukaryotic DNA polymerase α-Type (14,44,47,48,51), or family A DNA polymerases and family B DNA polymerases (4,9,50). As the number of DNA polymerase sequences increases, the classification of DNA polymerases becomes increasingly ambiguous. For example, DNA polymerase delta of yeast was shown to have amino acid sequence similarity to the α-Type DNA polymerases (17). It has become necessary to establish a unified classification of DNA polymerases. Here we propose to classify DNA polymerases into families A, B, and C (Figure 1: A, B, and C), according to the amino acid sequence homologies with *E. coli* DNA polymerases I, II, and III, respectively. As new and different prokaryotic and eukaryotic DNA polymerases are identified, the number of families can easily be expanded by using additional letters of the alphabet (i.e., D, E, etc.).

The bacterium *E. coli* (strain K12) contains three distinct DNA polymerases I, II, and III (52). *E. coli* DNA polymerase I, the first DNA polymerase discovered, is specified by the *pol*A gene (52). *E. coli* DNA polymerase II, encoded by the *pol*B gene, was recently sequenced and found to be identical to the *din*A gene, a DNA damage inducible gene whose expression is regulated by the SOS system in *E. coli* (8,53). Amino acid sequence alignment shows that *E. coli* DNA polymerase II has significant homology with family B (α-Type) DNA polymerases (8,53,54).

E. coli DNA polymerase III is a multisubunit enzyme encoded by various *dna* genes (55); the DNA polymerizing α-subunit encoded by the *pol*C (*dna*E) gene (56) and the 3′→5′ exonuclease performing ε-subunit encoded by the *dna*Q gene (57). The α-subunit of *E. coli* DNA polymerase III exhibits an extensive homology with the corresponding α-subunit of *Salmonella typhimurium* DNA polymerase III (35); and both show significant homology to *Bacillus subtilis* DNA polymerase III, a single-polypeptide encoded by the *pol*C gene (36).

In summary, family A DNA polymerases are named for their homology to the product of the *pol*A gene encoding *E. coli* DNA polymerase I; family B DNA polymerases are named for their homology to the product of the *pol*B gene encoding *E. coli* DNA polymerase II; and family C DNA polymerases are named for their homology to the product of the *pol*C gene encoding *E. coli* DNA polymerase III.

The eukaryotic DNA polymerase β, the smallest known DNA polymerase, does not have homology with those of any of the DNA polymerase families described above. Instead, DNA polymerase β has homology with terminal transferases (37). This β group we will call family X (Figure 1D). The classification and original reference(s) for the amino acid sequences of each DNA polymerase are shown in Table 1.

All of the family A DNA polymerases, except for yeast mitochondrial DNA polymerase I, are prokaryotic and are very sensitive to dideoxynucleotide inhibitors, and therefore are useful enzymes for DNA sequencing by the chain-termination method (58). The family A DNA polymerases are resistant to aphidicolin. The family B DNA polymerases are quite extensive in number and variety. Most of the family B DNA polymerases, if not all, are sensitive to aphidicolin and relatively resistant to dideoxynucleotide inhibitors. Most of the family B DNA polymerases, except for pAI2 (33) and yeast DNA polymerase II (16), contain the highly conserved amino acid sequence motif **YGDTD**, which has been suggested to form part of the dNTP binding site. Amino acid substitutions in this conserved sequence resulted in defects in the DNA polymerase activity without affecting the 3′→5′ exonuclease activity (59,60,61). The family C DNA polymerases are major bacterial replicative DNA polymerases which do not have appreciable homology with those of family A and B DNA polymerases. *B. subtilis* DNA polymerase III is a single polypeptide that is highly sensitive to hydroxyphenylazouracil (62). It is anticipated that the number of sequenced family C DNA polymerases will increase rapidly, since all of the aerobic bacteria may contain a member of this family of DNA polymerases.

SEQUENCE ALIGNMENT

The 37 complete DNA polymerase sequences and 3 complete terminal deoxynucleotidyltransferase (TDT) sequences are listed in 4 groups; the family A DNA polymerases, the family B DNA polymerases, the family C DNA polymerases, and family X DNA polymerases (including TDTs). In order to limit the space needed for the alignment, we omitted DNA polymerase sequences that are very similar to the prototype DNA polymerase. The DNA polymerases not shown include: herpes virus type-2 (63),

adenovirus type-5 (64), bacteriophage T3 (65), and bacteriophage PZA (66).

ACCURACY OF SEQUENCE DATA

Whenever a sequence ambiguity existed in a published sequence, we contacted the authors to obtain the updated sequence information. We found that a few published amino acid sequences differ at one or more positions from their GenBank/EMBL entry. Again, we have communicated with the primary author to confirm the correct sequences.

The multiple alignment of the amino acid sequences was obtained by a series of pairwise alignments combined and adjusted by eye into larger and larger subsets of similar sequences. The process of combining and adjusting by eye was aided by modified versions of the MOTIF program (67) and the ALIGN program (68). The GAP and BESTFIT programs, from UWGCG (University of Wisconsin Genetic Computer Group) (69), initially generated the pairwise alignments, adjusted for maximum alignment that allowed for a considerable number of gaps. We then compressed these alignments by eye to give a more contiguous alignment. The alignment of the sequences for optimal similarity is straightforward in the areas of relatively conserved structure, but is much more arbitrary in the more varied sequence areas. The alignment of the varied areas should therefore be regarded as less than optimal in view of the difficulties concerned with multiple alignments in these areas.

Finally, we invite further correction from readers, and welcome suggested revisions and alternative alignments.

ACKNOWLEDGEMENTS

This work was supported by grant GM28013 from the National Institutes of Health and by grant NP-704 from the American Cancer Society.

REFERENCES

1. Joyce, C.M., Kelley, W.S. and Grindley N.D.F. (1982) *J. Biol. Chem.* **257**, 1958–1964.
2. Lopez, P., Martinez, S., Diaz, A., Espinosa, M. and Lacks, S.A. (1989) *J. Biol. Chem.* **264**, 4255–4263.
3. Lawyer, F.C., Stoffel, S., Saiki, R.K., Myambo, K., Drummond, R. and Gelfand, D.H. (1989) *J. Biol. Chem.* **264**, 6427–6437.
4. Leavitt, M.C. and Ito, J. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4465–4469.
5. Dunn, J.J. and Studier, F.W. (1983) *J. Mol. Biol.* **166**, 477–535.
6. Rådén, B. and Rutberg, L. (1984) *J. Virol.* **52**, 9–15.
7. Foury, F. (1989) *J. Biol. Chem.* **264**, 20552–20560.
8. Iwasaki, H., Ishino, Y., Toh, H., Nakata, A. and Shinagawa, H. (1991) *Mol. Gen. Genet.* **226**, 24–33.
9. Jung, G., Leavitt, M.C., Hsieh, J.-C. and Ito, J. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8287–8291.
10. Savilahti, H. and Bamford D.H. (1987) *Gene* **57**, 121–130.
11. Yoshikawa, H. and Ito, J. (1982) *Gene* **17**, 323–335.
12. Matsumoto, K., Takano, H., Kim, C.I. and Hirokawa, H. (1989) *Gene* **84**, 247–255.
13. Spicer, E.K., Rush, J., Fung, C., Reha-Krantz, L.J., Karam, J.D. and Konigsberg, W.H. (1988) *J. Biol. Chem.* **263**, 7478–7486.
14. Wong, S.W., Wahl, A.F., Yuan, P.-M., Arai, N., Pearson, B.E., Arai, K.-i., Korn, D., Hunkapiller, M.W. and Wang, T.S.-F. (1988) *EMBO J.* **7**, 37–47.
15. Pizzagalli, A., Valsasnini, P., Plevani, P. and Lucchini, G. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3772–3776.
16. Morrison, A., Araki, H., Clark, A.B., Hamatake, R.K. and Sugino, A. (1990) *Cell* **62**, 1143–1151.
17. Boulet, A., Simon, M., Faye, G., Bauer, G.A. and Burgers, P.M.J. (1989) *EMBO J.* **8**, 1849–1854.
18. Morrison, A., Christensen, R.B., Alley, J., Beck, A.K., Bernstine, E.G., Lemontt, J.F. and Lawrence, C.W. (1989) *J. Bacteriol.* **171**, 5659–5667.
19. Gibbs, J.S., Chiou, H.C., Hall, J.D., Mount, D.W., Retondo, M.J., Weller, S.K. and Coen, D.M. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7969–7973.
20. Kouzarides, T., Bankier, A.T., Satchwell, S.C., Weston, K., Tomlinson, P. and Barrell, B.G. (1987) *J. Virol.* **61**, 125–133.
21. Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Séguin, C., Tuffnell, P.S. and Barrell, B.G. (1984) *Nature* **310**, 207–211.
22. Davison, A.J. and Scott, J.E. (1986) *J. Gen. Virol.* **67**, 1759–1816.
23. Binns, M.M., Stenzler, L., Tomley, F.M., Campbell, J. and Boursnell, M.E.G. (1987) *Nucl. Acids Res.* **15**, 6563–6573.
24. Earl, P.L., Jones, E.V. and Moss, B. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3659–3663.
25. Tomalski, M.D., Wu, J. and Miller, L.K. (1988) *Virology* **167**, 591–600.
26. Gingeras, T.R., Sciaky, D., Gelinas, R.E., Bing-Dong, J., Yen, C.E., Kelly, M.M., Bullock, P.A., Parsons, B.L., O'Neill, K.E. and Roberts, R.J. (1982) *J. Biol. Chem.* **257**, 13475–13491.
27. Engler, J.A., Hoppe, M.S. and van Bree, M.P. (1983) *Gene* **21**, 145–159.
28. Shu, L., Hong, J.S., Wei, Y.-f. and Engler, J.A. (1986) *Gene* **46**, 187–195.
29. Paillard, M., Sederoff, R.R. and Levings, C.S. III (1985) *EMBO J.* **4**, 1125–1128.
30. Stark, M.J.R., Mileham, A.J., Romanos, M.A. and Boyd, A. (1984) *Nucl. Acids Res.* **12**, 6011–6030.
31. Tommasino, M., Ricci, S. and Galeotti, C.L. (1988) *Nucleic Acids Res.* **16**, 5863–5878.
32. Oeser, B. and Tudzynski, P. (1989) *Mol. Gen. Genet.* **217**, 132–140.
33. Kempken, F., Meinhardt, F. and Esser, K. (1989) *Mol. Gen. Genet.* **218**, 523–530.
34. Tomasiewicz, H.G. and McHenry, C.S. (1987) *J. Bacteriol.* **169**, 5735–5744.
35. Lancy, E.D., Lifsics, M.R., Munson, P. and Maurer, R. (1989) *J. Bacteriol.* **171**, 5581–5586.
36. Hammond, R.A., Barnes, M.H., Mack, S.L., Mitchener, J.A. and Brown, N.C. (1991) *Gene* **98**, 29–36.
37. Matsukage, A., Nishikawa, K., Ooi, T., Seto, Y. and Yamaguchi, M. (1987) *J. Biol. Chem.* **262**, 8960–8962.
38. Abbotts, J., SenGupta, D.N., Zmudzka, B., Widen, S.G., Notario, V. and Wilson, S.H. (1988) *Biochemistry* **27**, 901–909.
39. SenGupta, D.N., Zmudzka, B.Z., Kumar, P., Cobianchi, F., Skowronski, J. and Wilson, S.H. (1986) *Biochem. Biophys. Res. Comm.* **136**, 341–347.
40. Peterson, R.C., Cheung, L.C., Mattaliano, R.J., White, S.T., Chang, L.M.S. and Bollum F.J. (1985) *J. Biol. Chem.* **260**, 10495–19502.
41. Koiwai, O., Yokota, T., Kageyama, T., Hirose, T., Yoshida, S. and Arai, K.-i. (1986) *Nucl. Acids Res.* **14**, 5777–5792.
42. Argos, P., Tucker, A.D. and Phillipson, L. (1986) *Virology* **149**, 208–216.
43. Larder, B.A., Kemp, S.D. and Darby, G. (1987) *EMBO J.* **6**, 160–175.
44. Hall, J.D. (1988) *Trends Genet.* **4**, 42–46.
45. Reha-Krantz, L.J. (1988) *J. Mol. Biol.* **202**, 711–724.
46. Bernad, A., Zaballos, A., Salas, M. and Blanco, L. (1987) *EMBO J.* **6**, 4219–4225.
47. Wang, T.S.-F., Wong, S.W. and Korn, D. (1989) *FASEB* **3**, 14–21.
48. Bernad, A., Blanco, L., Lazaro, J.M., Martin, G. and Salas, M. (1989) *Cell* **59**, 219–228.
49. Polesky, A.H., Steitz, T.A., Grindley, N.D.F. and Joyce, C.M. (1990) *J. Biol. Chem.* **24**, 14579–14591.
50. Ito, J. and Braithwaite, D.K. (1990) *Nucl. Acids Res.* **18**, 6716.
51. Blanco, L., Bernad, A. and Salas, M. (1991) *Nucl. Acids Res.* **19**, 955.
52. Kornberg, A. (1974) DNA replication. W.H. Freeman and Co., San Francisco.
53. Bonner, C.A., Hays, S., McEntee, K. and Goodman M.F. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7663–7667.
54. Chen, H., Lawrence, C.B., Bryan, S.K. and Moses, R.E. (1990) *Nucl. Acids Res.* **18**, 7185–7186.
55. McHenry, C.S.(1985) *Mol. Cell. Biochem.* **66**, 71–85.
56. Shepard, D., Oberfelder, R.W., Welch, M.W. and McHenry, C.S. (1984) *J. Bacteriol.* **158**, 455–459.
57. Scheuermann, R., Tam, S., Burgers, P.M.J., Lu, C. and Echols, H. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7085–7089.
58. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.
59. Dorsky, D.I. and Crumpacker, C.S. (1990) *J. Virol.* **64**, 1394–1397.
60. Bernad, A., Lázaro, J.M., Salas, M. and Blanco, L. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4610–4614.

61. Jung, G., Leavitt, M.C., Schultz, M. and Ito, J. (1990) *Biochem. Biophys. Res. Comm.* **170**, 1294–1300.
62. Neville, M.N. and Brown, N.C. (1972) *Nature New Biol.* **240**, 80–82.
63. Tsurumi, T., Maeno, K. and Nishiyama, Y. (1987) *Gene* **52**, 129–137.
64. Dekker, B.M.M. and Van Ormondt, H. (1984) *Gene* **27**, 115–120.
65. Beck, P.J., Gonzalez, S., Ward, C.L. and Molineux, I.J. (1989) *J. Mol. Biol.* **210**, 687–701.
66. Paces, V., Vlcek, C., Urbanek, P. and Hostomsky, Z. (1985) *Gene* **38**, 45–56.
67. Smith, H.O., Annau, T.M. and Chandrasegaran, S. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 826–830.
68. Doolittle, R.F.and Feng, D.-F. (1990) In Doolittle, R.F. (ed.), Methods in Enzymol. – Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences. Academic Press, New York, Vol. 183, pp. 659–669.
69. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucl. Acids Res.* **12**, 387–395.

Classification of DNA polymerases

			References
A.	**Family A DNA polymerases**		
	1.	**Bacterial DNA polymerases**	
		a) *E. coli* DNA polymerase I	(1)
		b) *Streptococcus pneumoniae* DNA polymerase I	(2)
		c) *Thermus aquaticus* DNA polymerase I	(3)
	2.	**Bacteriophage DNA polymerases**	
		a) T5 DNA polymerase	(4)
		b) T7 DNA polymerase	(5)
		c) Spo2 DNA polymerase	(6)
	3.	**Mitochondrial DNA polymerase**	
		Yeast mitochondrial DNA polymerase (MIP1)	(7)
B.	**Family B DNA polymerases**		
	1.	**Bacterial DNA polymerases**	
		E. coli DNA polymerase II	(8)
	2.	**Bacteriophage DNA polymerases**	
		a) PRD1 DNA polymerase*	(9,10)
		b) ϕ29 DNA polymerase*	(11)
		c) M2 DNA polymerase*	(12)
		d) T4 DNA polymerase	(13)
	3.	**Eukaryotic DNA polymerases**	
		a) Human DNA polymerase alpha	(14)
		b) Yeast DNA polymerase I	(15)
		c) Yeast DNA polymerase II	(16)
		d) Yeast DNA polymerase III (delta)	(17)
		e) Yeast DNA polymerase Rev3	(18)
	4.	**Viral DNA polymerases**	
		a) Herpes-1 DNA polymerase	(19)
		b) Human cytomegalovirus DNA polymerase	(20)
		c) Epstein-Barr virus DNA polymerase	(21)
		d) Varicella-Zoster virus DNA polymerase	(22)
		e) Fowlpox virus DNA polymerase	(23)
		f) Vaccinia virus DNA polymerase	(24)
		g) Autographa californica nuclear polyhedrosis virus (AcMNPV) DNA polymerase	(25)
		h) Adenovirus-2 DNA polymerase*	(26)
		i) Adenovirus-7 DNA polymerase*	(27)
		j) Adenovirus-12 DNA polymerase*	(28)
	5.	**Eukaryotic linear DNA plasmid encoded DNA polymerases**	
		a) S-1 maize mitochondrial DNA polymerase*	(29)
		b) *Kluyveromyces lactis* plasmid pGKL1 DNA polymerase*	(30)
		c) *Kluyveromyces lactis* plasmid pGKL2 DNA polymerase*	(31)
		d) *Claviceps purpurea* plasmid pCLK1 DNA polymerase*	(32)
		e) *Ascobolus immersus* plasmid pAI2 DNA polymerase*	(33)
C.	**Family C DNA polymerases**		
		Bacterial replicative DNA polymerases	
		a) *E. coli* DNA polymerase III α subunit	(34)
		b) *Salmonella typhimurium* DNA polymerase III α subunit	(35)
		c) *Bacillus subtilis* DNA polymerase III	(36)
D.	**Family X DNA polymerases**		
		a) Rat DNA polymerase β	(37)
		b) Human DNA polymerase β	(38,39)
		c) Human terminal deoxynucleotidyltransferase (TdT)	(40)
		d) Bovine terminal deoxynucleotidyltransferase (TdT)	(41)
		e) Mouse terminal deoxynucleotidyltransferase (TdT)	(41)

Table 1. The main families and subclassifications of DNA polymerases. Those DNA polymerases marked with a star (*) are protein-primed DNA polymerases.

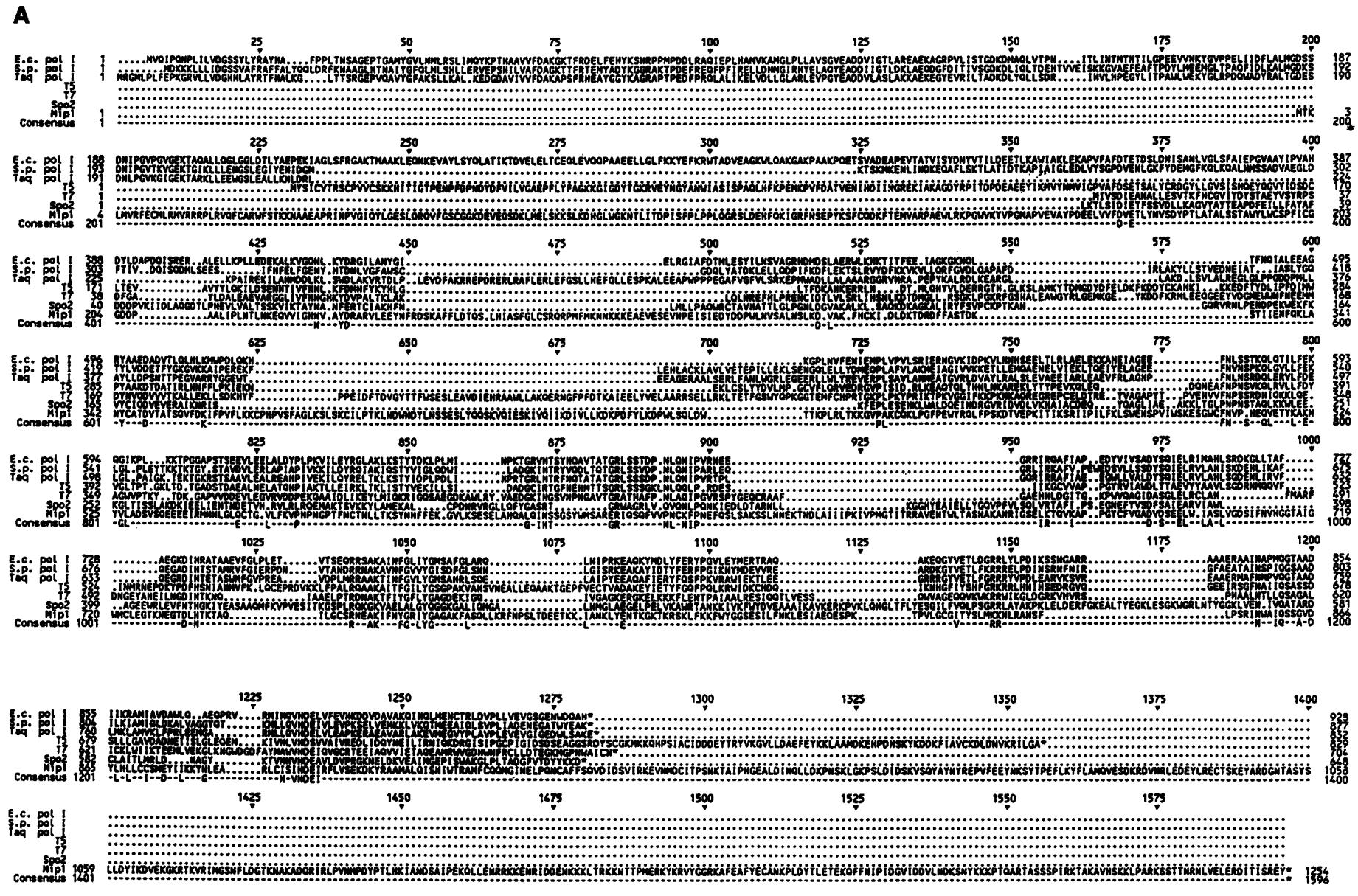
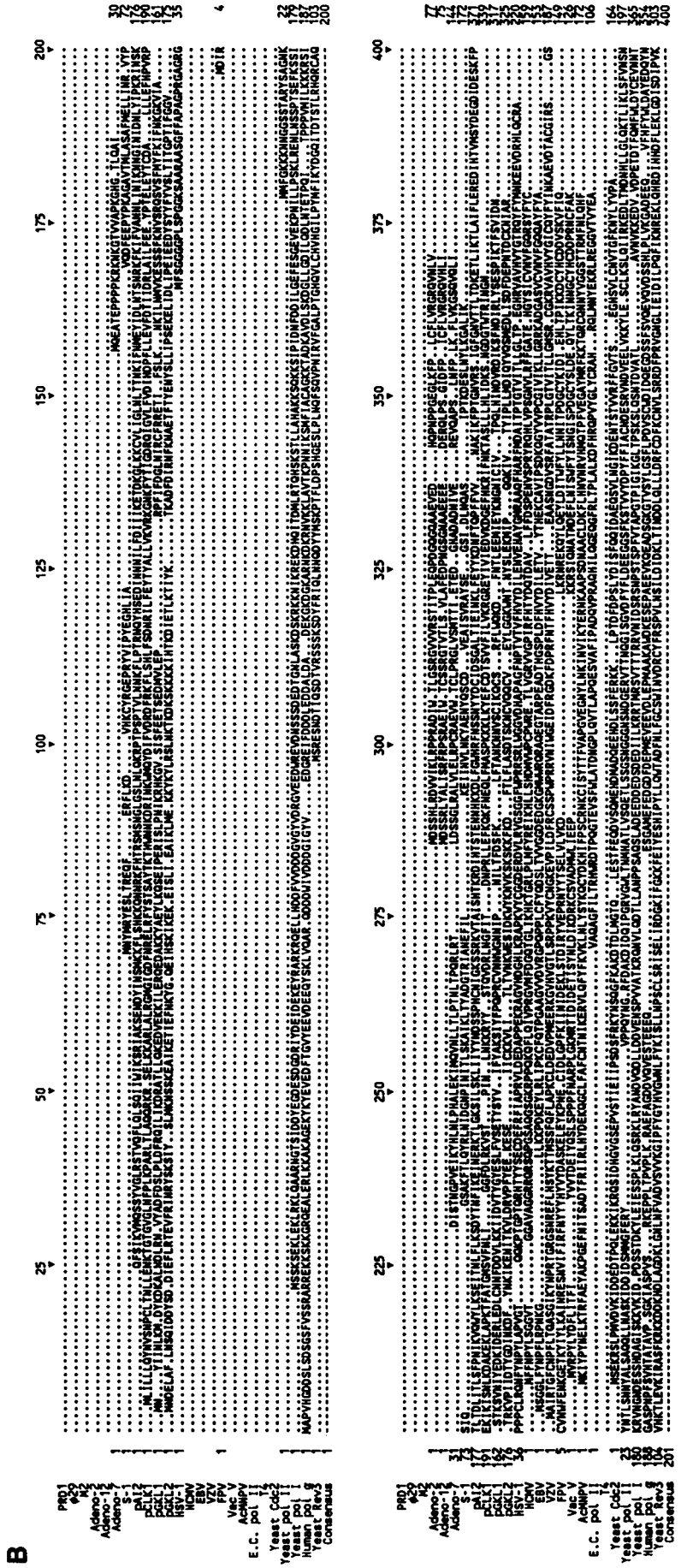
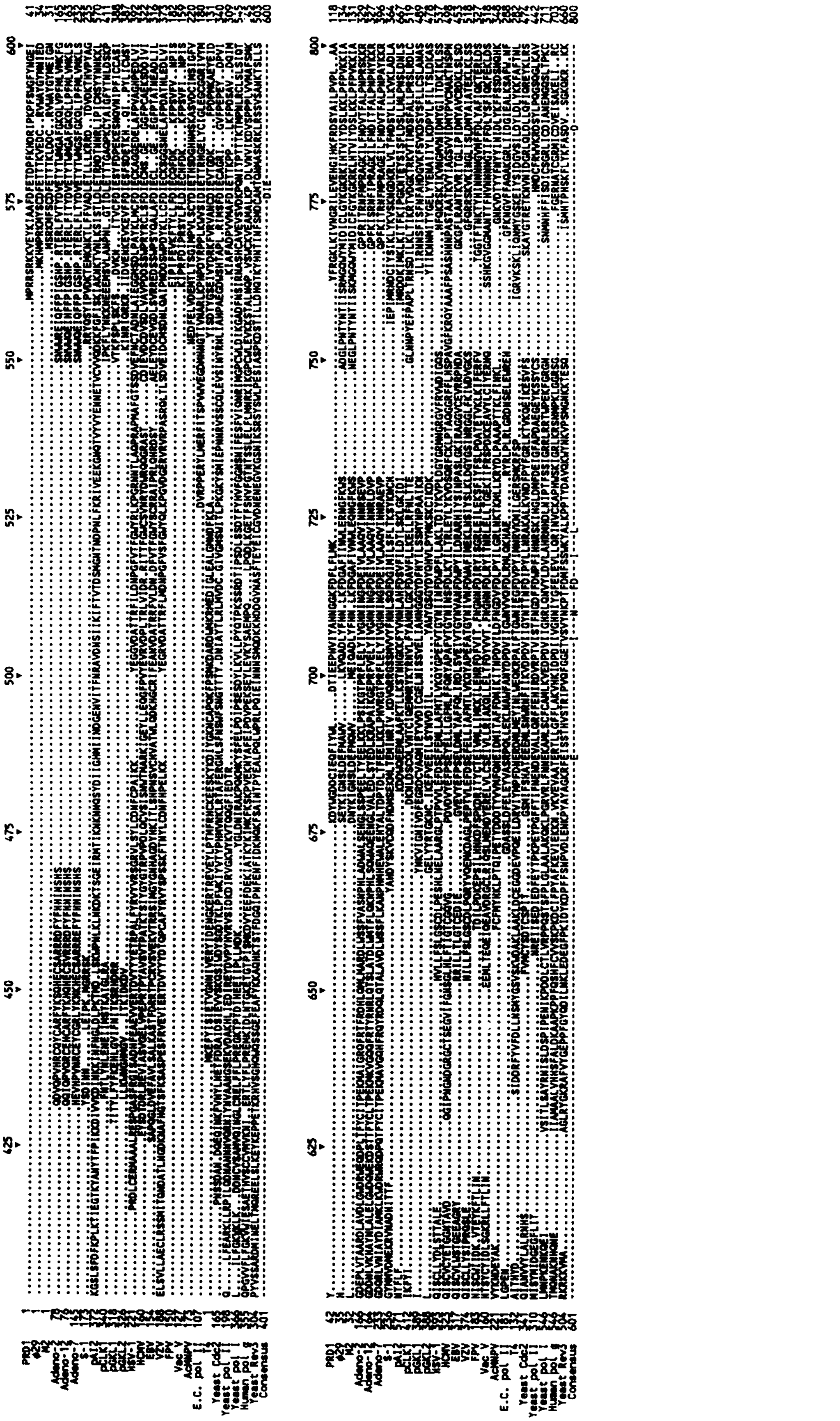
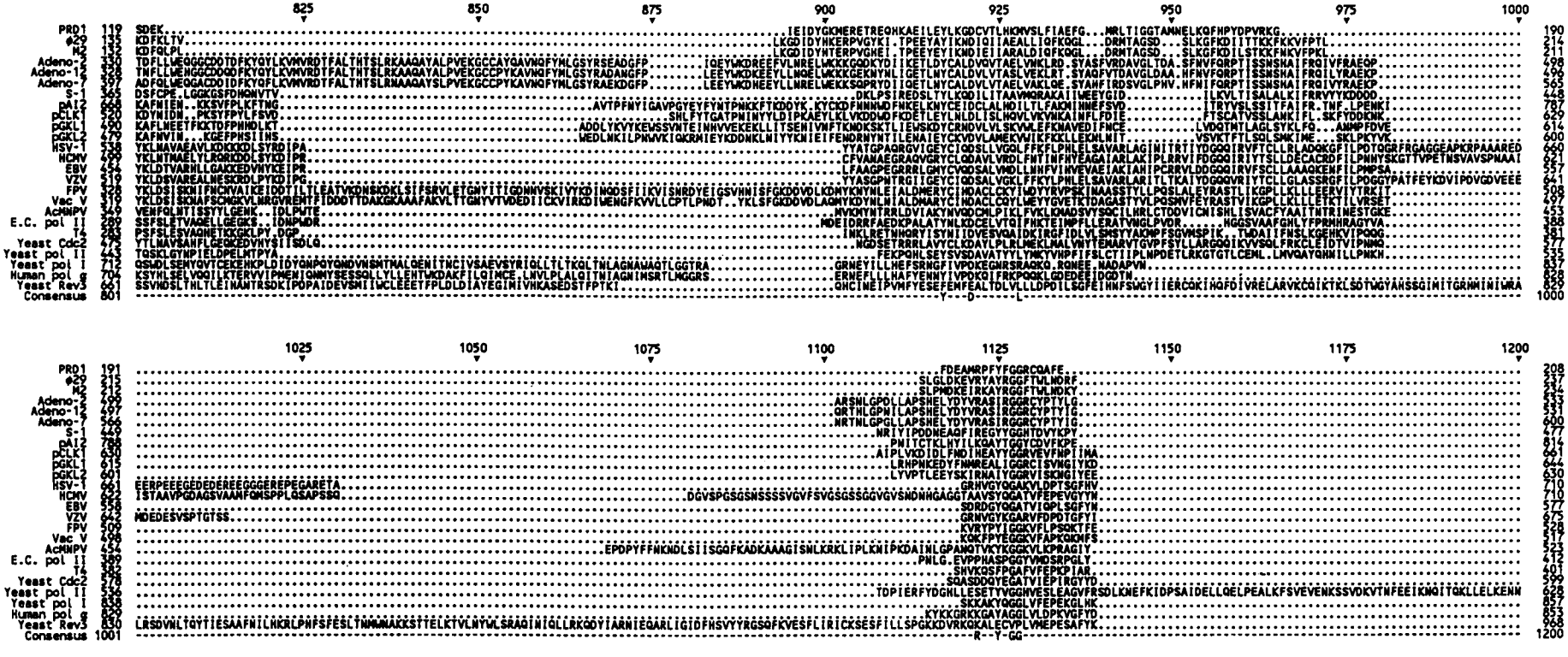
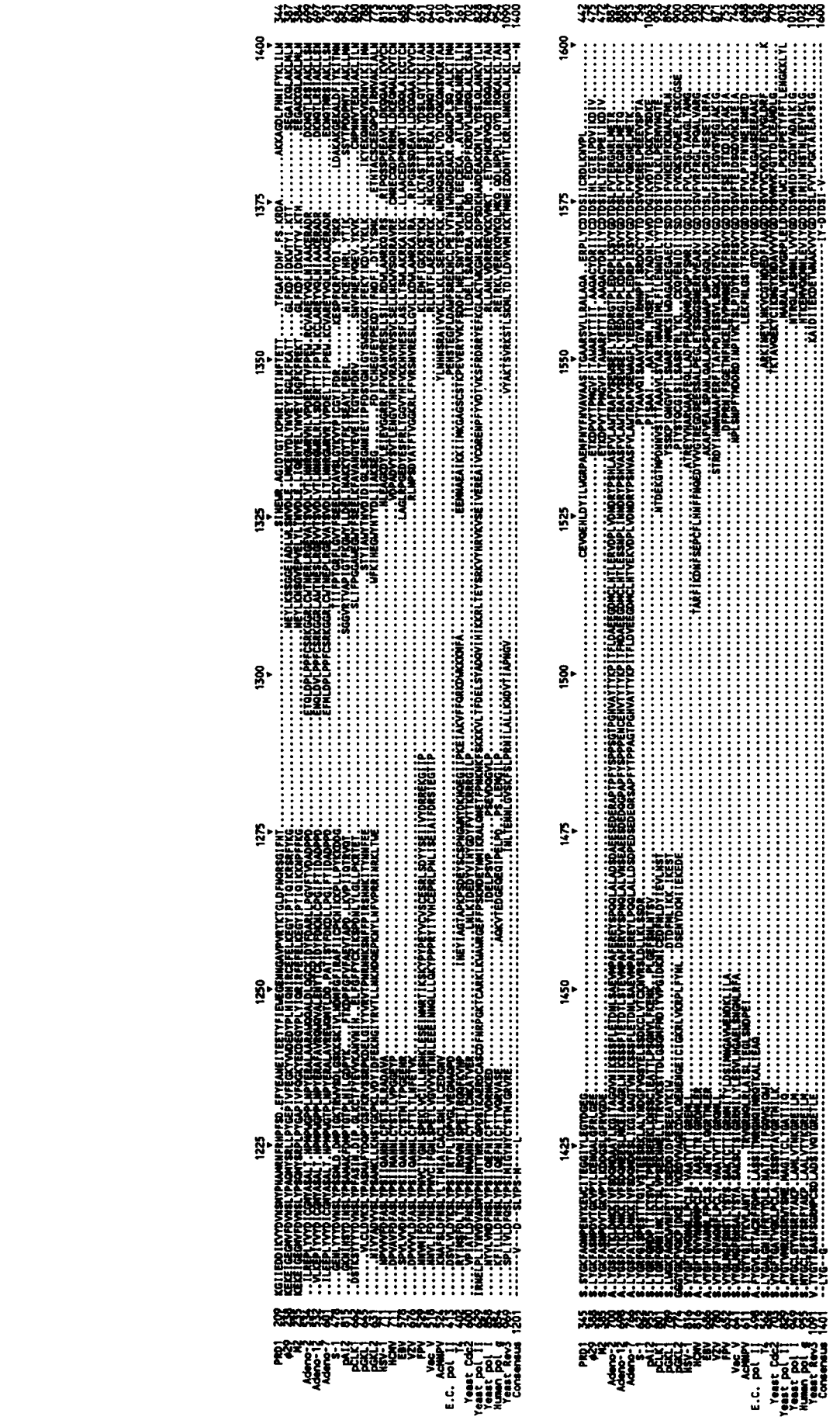


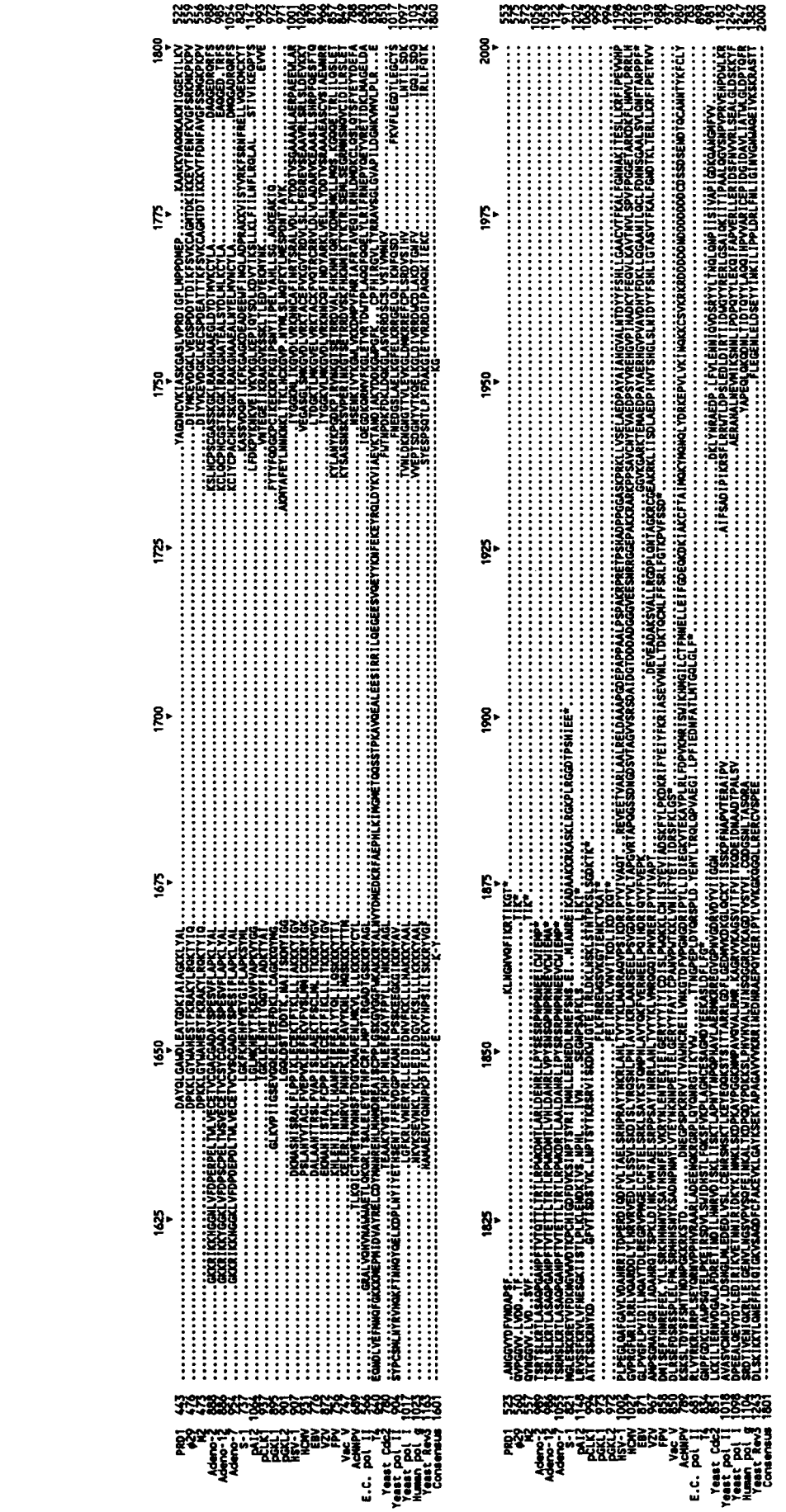
Figure 1A. Family A DNA polymerases—*E. coli* DNA pol. I (E.c. pol I)(1), *Streptococcus pneumoniae* DNA pol. I (S.p. pol I)(2), *Thermus aquaticus* DNA pol. I (Taq pol I)(3), T5 DNA pol. (4), T7 DNA pol. (5), Spo2 DNA pol. (6), and yeast mitochondrial DNA pol. (MIP1)(7).

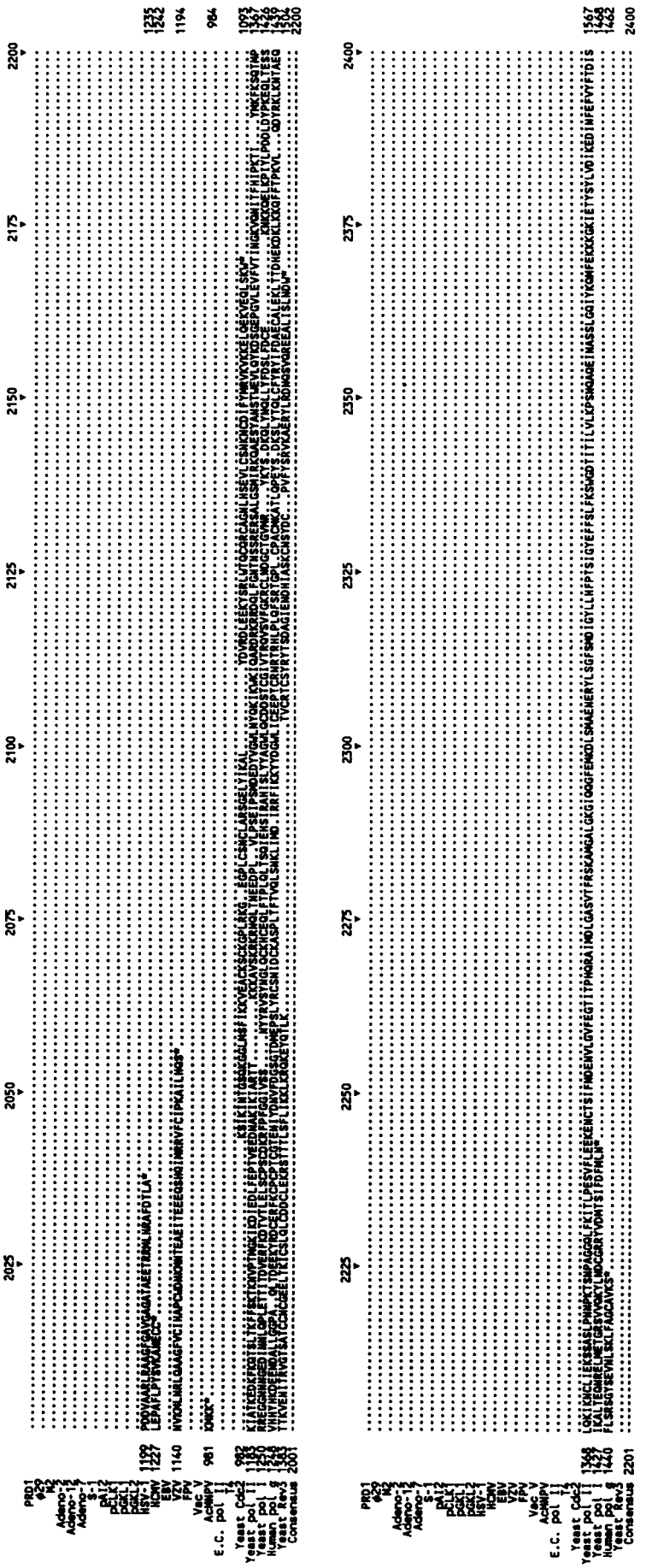

B
PRD1
P29
M2
Adeno-2
Adeno-12
Adeno-7
S-1
pAI2
pGKL1
pGKL2
HSV-1
HCMV
EBV
VZV
FPV
Vac V
AcNPV
E.C. pol II
T4
Yeast Cdc2
Yeast pol II
Yeast pol I
Human pol II
Yeast Rev3
Consensus

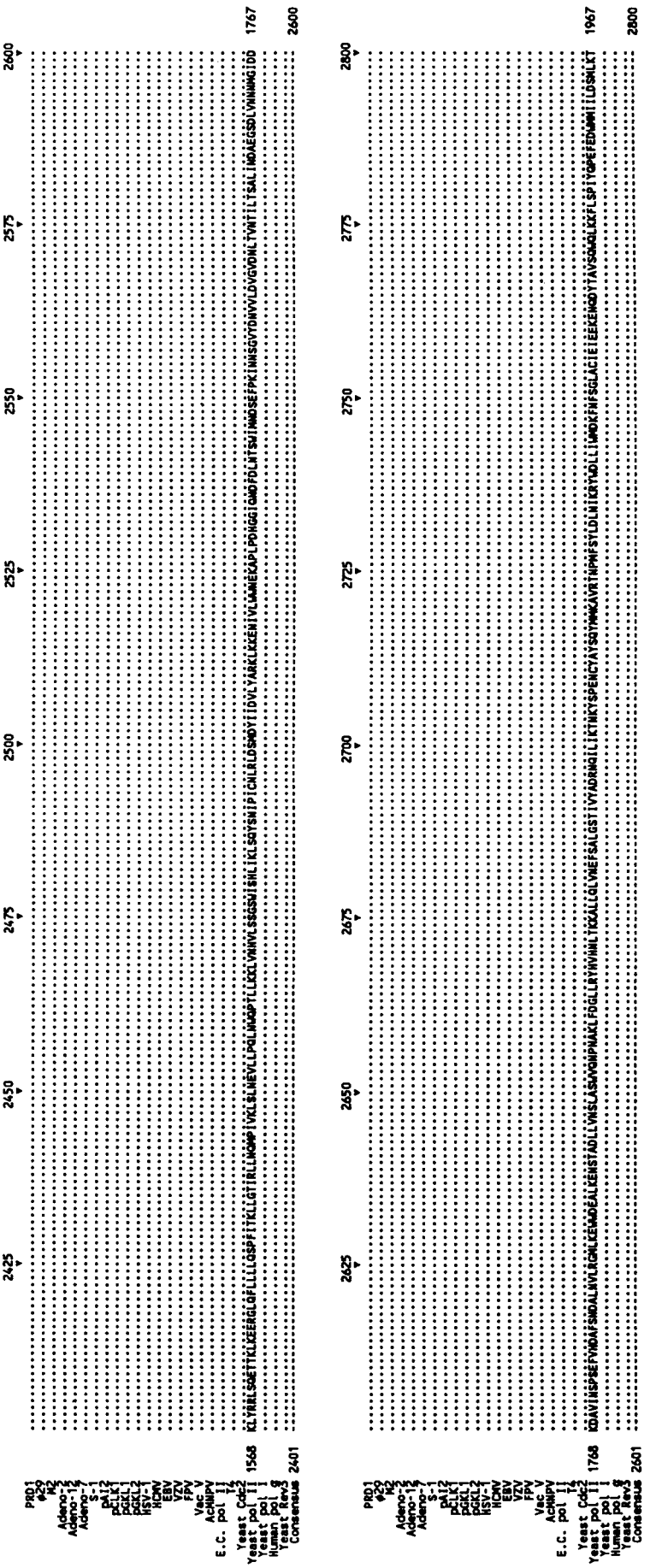



825
850
875
900
925
950
975
1000
PRD1 119 SDEK 190
φ29 135 KDFKLTV 214
M2 132 KDFQLPL 211
Adeno-2 330 TDFLLVEGGGCDDTDFKYQYLKVMVRDTFALTHTSLRKAAQAYALPVEKGCCAYGAVNGFYMLGSYRSEADGFP IQEYVKDREEFVLNRELWKKKGQDKYDIIKETLDYCALDVGVTAELVWKLRD.SYASFVRDAVGLTDA.SFNVFGRPTISSNSHAIFRGIVFRAEGP 498
Adeno-12 328 TNFLLVENGGCDDQDFKYQYLKVMVRDTFALTHTSLRKAAQAYALPVEKGCCPYKAVNGFYMLGSYRADANGFP LEEYVKDKEEYLLNQELWKKKGEKNYNLIGETLNYCALDVLVTASLVEKLRT.SYAQFVTDAVGLDAA.HFNVFGRPTISSNSHAIFRGILYRAEKP 499
Adeno-7 307 ADFQLVEQGACDDIDFKYQFLKVMVRDTFALTHTSLRNAAQAYSLPVEKGCCPYKAVNGFYMLGSYRAEKDGFP LEEYVKDHEEYLLNRELWKKKSQPRYDIIQETLNYCALDVLVTAELVAKLQE.SYAHFIRDSVGLPHV.HFNIFQRPTISSNSHAIFRGIVYRAEKP 565
S-1 365 DSFCPE.LGGKGSFDHQNVTV DKLPSIREDSLTYLKQDILITAAVMQRAKAIIWEEYGID ILKVLTISALALKIFRRVYYKDDQD 448
pAI2 668 KAFNIEN..KKSVFPLKFTNG AVTPFNYIGAVPGYEYFYNTPNKKFTKDDYK.KYCKDFNNHWDFNKELKNYCEIDCLALHQILTLFAKMIHNEFSVD ITRYVSLSSITFAIFR.TNF.LPEMKI 787
pCLK1 630 KDYNIDN..PKSYFPYLFSVD SHLFYTGATPNINYYLDIPKAEYLKLVKDDWDFKDETLEYLNLDLISLHQVLVKVMKAINFLFDIE FTSCATVSSLANKIFL.SKFYDDKNK 629
pGKL1 490 KAFLMEETFKCTDFPHWNDLKT ADDLYKVTKEVSSVNTEINHVVVEKEKLLITSEMIVNFTKNDKSKTLIEVSKDYCRNDVLVLSKVWLEFKNAVEDIFNCE LVDQTNTLAGLSYKLFG...AWWPFDVE 614
pGKL2 479 KAFNVIN...KGEFPHSIINS WEDLNKILPNWVKIQKRAIEYKDDNKLWIYYKNIEIFENDRNYNTILENAIEYCKVDVLAMEKVWIKFKCLLEKNLNIT VSVKTFTLSQLSWKIME...SKLPKYVK 600
HSV-1 538 YKLMAVAEAVLKDKCKKDLSYRDIPA YYATGPAQRGVIGEYCIQDSLLVGQLFFKFLPHLELSAVARLAGINITRTIYDGGQIRVFTCLLRLADGKGFILPDTGGRFRGAGGEAPKRPAAARED 660
HCMV 499 YKLNTMAELYLRQRCDDLSYKDIPR CFVANAEGRAQVGRYCLQDAVLVRDLFMTINFHYEAGAIARLAKIPLRRVIFDGGDIRIYTSLLDECACRDFIPLMNYSKGTTVPETNSVAVSPNAAI 621
EBV 454 YKLDTVARHLLGAKKEDVNYKEIPR LFAAGPEGRRRLGMYCVQDSALVMDLLMNFVIHVEVAEIAKIAHIPCRRVLDDGDQIRVFSCLLAAAQKENFILPMPSA 557
VZV 519 YKLDSVAREALNESKRDLPYKDIPG YYASGPNTRGIIGEYCIQDSALVGKLFFKYLPHLELSAVARLARITLTKAIYDGDGDVRIYTCLLGLASSRGFILPDGGYPATFEYKDVIPDVGDVEEE 641
FPV 528 YKLDSISKNIFNCMVAICEIDDTILTLEATVKDNSKDKLSIFSRVLETGMYITTGDNMVSKIVYKDINGDSFIIKVISNRDYEIGSVHNISFGKDDVDLKDHYKNYNLEIALDMERYCIHDACLCKYIWDYYRVPSKINAASSTYLLPQSLALEYRASTLIKGPLLKLLLEERVIYTRKIT 708
Vac V 519 YKLDSISKNAFSCMGKVLNRGVREMTFIDDDTTDAKGKAAAFAKVLTTGNYVTVDEDIICKVIRKDIWENGFKVVLLCPTLPNDT..YKLSFGKDDVDLAGMYKDYNLWIALDMARYCINDACLCGYLWEYYGVETKTDAGASTYVLPQSMVFEYRASTVIKGPLLKLLLETKTILVRSET 697
AcMNPV 549 VENFQLMTISSTYTLGENK..IDLPUTE NVKDYNTRRLDVIAKYNVQDCMLPIKLFVKLKNADSVYSQCILNRLCTDDVICNISHLISVACFYAAITNTRINESTGKE 453
E.C. pol II 289 SSFSLETVAQELLGEGKCE..IDNPLMDR MDEIDRRFAEDKPALATYNLKDCELVTQIFNKTEIMPFLLERATVNGLPVDR...HGGSVAAFGHLYFPRNHIRAGTVA 388
T4 283 PSFSLESVAQNETKSGKLPY.DGP IHKLRETNHQRYISYNIDVEEVQALDKIRGFIDLYLSNSYTYAKMPFSGVMSPIK..TVDAIIFNSLKQEHKCVIPQQG 381
Yeast Cdc2 475 YTLMAVEAWFLGEGKEDVNYSITSDLG NGDSETRRRLAVYCLKDAYLPLRLMEKLMALVNYTENARVTGVPFSYLLARGGQIKVVSQLFRKCLEIDTVIPNMQ 577
Yeast pol II 443 TQSKLGTYNPIELDPELMTPYA FEKPQHLSEYSVSDAVATYTYLYMKTVHPFIFSLCTIIPLNPDETLRKGTGTLCEML.LMVQAYGHMILLPNKH 535
Yeast pol I 712 QSWDLSEMYQVTCEKEHKPLDIDYQNPGYGNDVNSMTMALQENITNCIVSAEVSYRIQLLTLTKQLTNLAGMAWAQTLGGTRA GRNEYILLHEFSRNGFIVPDKEGNRSRAQKQ.RQNEE.NADAPVN 837
Human pol g 704 KSYHLSELVQQILKTERVVVIPHENIQNMYSESSQLLYLLENTVKDAKFILQIMCE.LNVLPLALQITNIAGNIMSRTLMGGRS ERNEFLLLHAFYENNYIVPDKGIFRKPQQKLGDEDEEIDGDTN 828
Yeast Rev3 661 SSVNDSLTHLTLEINANTRSDKIPDPAIDEVSMIIWCLEEETFPLDLDIAYEGINIVHKASEDSTFPTKI GHCIMEIPVMFYESEFEMFEALTDLVLLLDPDILSGFEIHNFSWGYIIERCQKINQFDIVRELARVKCQIKTKLSQTWGYAHSSGIMITGRNMINIWRA 829
Consensus 801 Y--D-----L 1000
1025
1050
1075
1100
1125
1150
1175
1200
PRD1 191 FDEAMRPFYFGGRCQAFE 208
φ29 215 SLGLDKEVRYAYRGGFTWLMDRF 237
M2 212 SLPMDKEIRKAYRGGFTWLMDKY 234
Adeno-2 499 ARSNLGPDLLAPSHELYDYVRASIRGGRCYPTYLG 533
Adeno-12 497 QRTHLGPNILAPSHELYDYVRASIRGGRCYPTYIG 531
Adeno-7 566 NRTNLGPGLLAPSHELYDYVRASIRGGRCYPTYIG 600
S-1 449 NRIYIPDDNEAGFIREGYYGGHTDVYKPY 477
pAI2 788 PNITCTKLHYILKGAYTGGYCDVFKPE 814
pCLK1 630 AIPLVKDIDLFNDIHEAYTGGRVEVFNPIIMA 661
pGKL1 615 LRHPNKEDYFNWREALIGGRCISVNGIYKD 644
pGKL2 601 LYVPTLEEYSKIRNAIYGGRVISKNGIYEE 630
HSV-1 661 GRNVGYQGAKVLDPTSGFHV 710
HCMV 622 ISTAAVPGDAGSVAANFGMSPPLQSAPSSQ DGVSPGSGSNSSSSVGVFSVGSGSSGGVGVSNDNHGAGGTAAVSYQGATVFEPEVGYTM 710
EBV 558 SDRDGYQGATVIGPLSGFYM 577
VZV 642 MDEDESVSPPTGTSS GRNVGYKGARVFDPDTGFYI 675
FPV 509 KVRYPYIGGKVFLPSQKTFE 528
Vac V 498 KQKFPYEGGKVFAPKQKQKMFS 517
AcMNPV 454 EPDPYFFNKNDLSIISGQFKADKAAAGISNLKRKLIPLKNIPKDAINLGPANQTVKYKGGKVLKPRAGIY 523
E.C. pol II 389 PNLG.EVPPHASPGGYVMDSRPGLY 412
T4 382 SNVKQSFPGAFVFEPKPIAR 401
Yeast Cdc2 578 SQASDDQQYEGATVIEPINGYYD 599
Yeast pol II 536 TDPIERFYDGNLLESETTYGGHVESLEAGVFRSDLKNEFKIDPSAIDELLQELPEALKFSVEVENKSSVDKVTNFEEIKMGITQKLLELKENN 628
Yeast pol I 838 SKKAKYQGGLVFEPEKGLHK 857
Human pol g 829 KYKKGRKKGAYAGGLVLDPKVGFYD 853
Yeast Rev3 830 LRSDVWLTGYTIESAAFNILKKRLPNFSFESLTNMAMAKKSTTELKTVLNYWLSRAGINIGLLRKCGYIARNIEGARLIGIDFHSVYYRGSGFKVESFLIRICKSESFILLSPGKKDVRKGKALECVPLVNEPESAFYK 968
Consensus 1001 R--Y-GG 1200








PRD1
φ29
M2
Adeno-2
Adeno-12
Adeno-7
S-1
pAI2
pGKL1
pGKL2
HSV-1
HCMV
EBV
VZV
FPV
Vac V
AcNMPV
E.C. pol I
T4
Yeast Cdc2
Yeast pol II
Yeast pol I
Human pol β
Yeast Rev3
Consensus
2425
2450
2475
2500
2525
2550
2575
2600
1568
1767
2401
2600
2625
2650
2675
2700
2725
2750
2775
2800
1768
1967
2601
2800

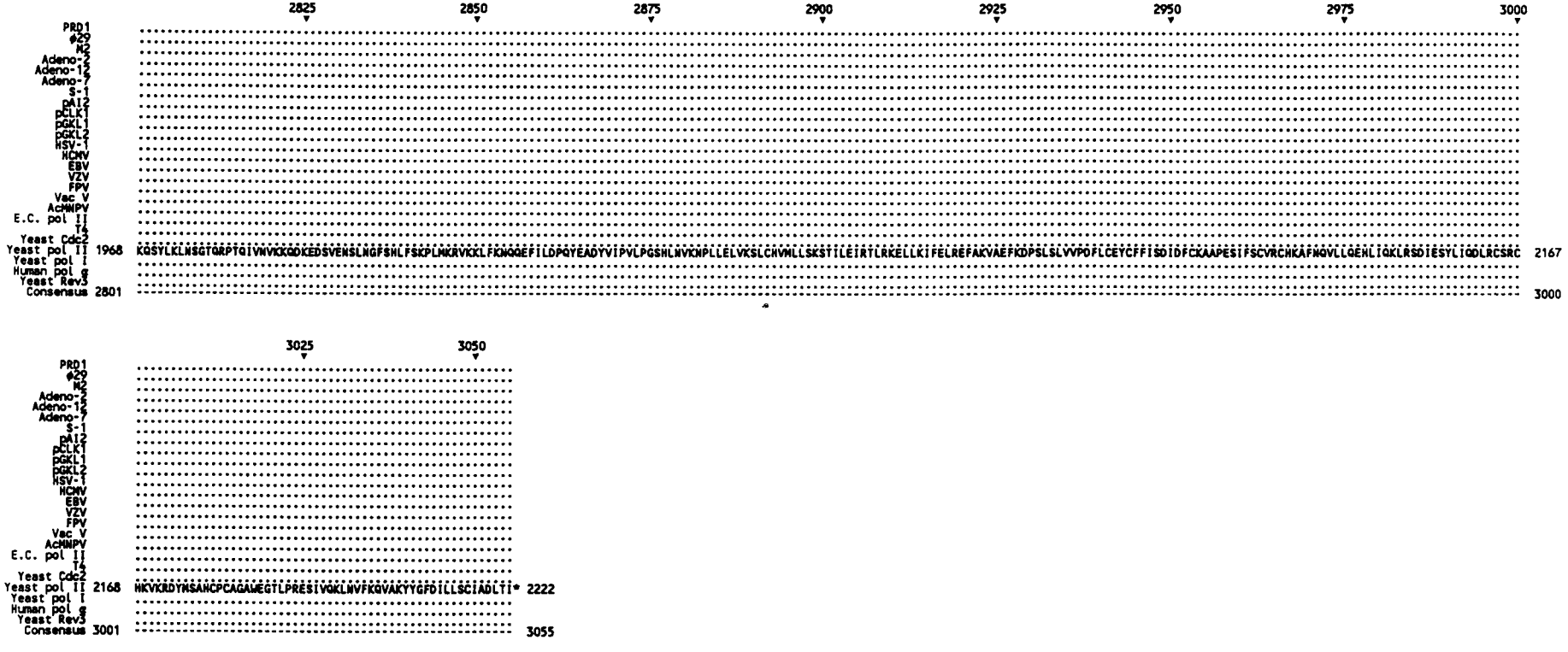


Figure 1B. Family B DNA polymerases—PRD1 DNA pol. (9,10), φ29 DNA pol. (11), M2 DNA pol. (12), Adenovirus type-2 DNA pol. (Adeno-2)(26), Adenovirus type-12 DNA pol. (Adeno-12)(28), Adenovirus type-7 DNA pol. (Adeno-7)(27), S-1 maize mitochondrial DNA pol. (S1)(29), *Ascobolus immersus* plasmid DNA pol. (pAI2)(33), *Claviceps purpurea* plasmid DNA pol. (pCLK1)(32), *Kluyveromyces lactis* plasmid DNA pol. (pGKL1)(30), *kluyveromyces lactis* plasmid DNA pol. (pGKL2)(31), herpes simplex type-1 DNA pol. (HSV-1)(19), Human cytomegalovirus DNA pol. (HCMV)(20), Epstein-Barr virus DNA pol. (EBV)(21), Varicella-zoster virus DNA pol. (VZV)(22), Fowlpox virus DNA pol. (FPV)(23), Vaccinia virus DNA pol. (VacV)(24), Autographa californica nuclear polyhedrosis virus DNA pol. (AcMNPV)(25), *E. coli* DNA pol. II (E.c. pol II)(8), T4 DNA pol. (13), Yeast DNA pol. III (CDC2)(17), Yeast DNA pol. II (16), Yeast DNA pol. I (15), Human DNA pol. α (14), and Yeast Rev3 DNA pol.(18).

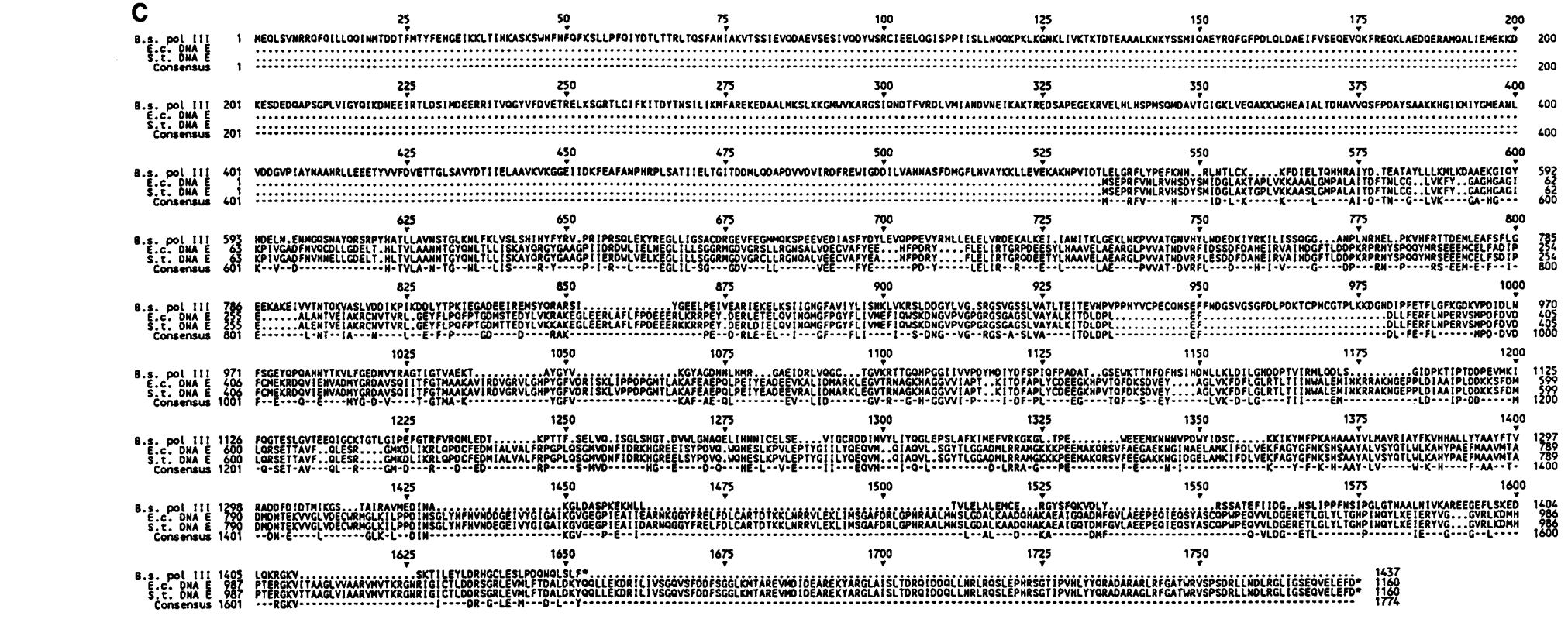


Figure 1C. Family C DNA polymerases—*Bacillus subtilis* DNA pol. III (B.s. pol III)(36), *E. coli* DNA pol. III α-subunit (E.c. DNA E)(34), and *Salmonella typhimurium* DNA pol. III α-subunit (S.t. DNA E)(35).

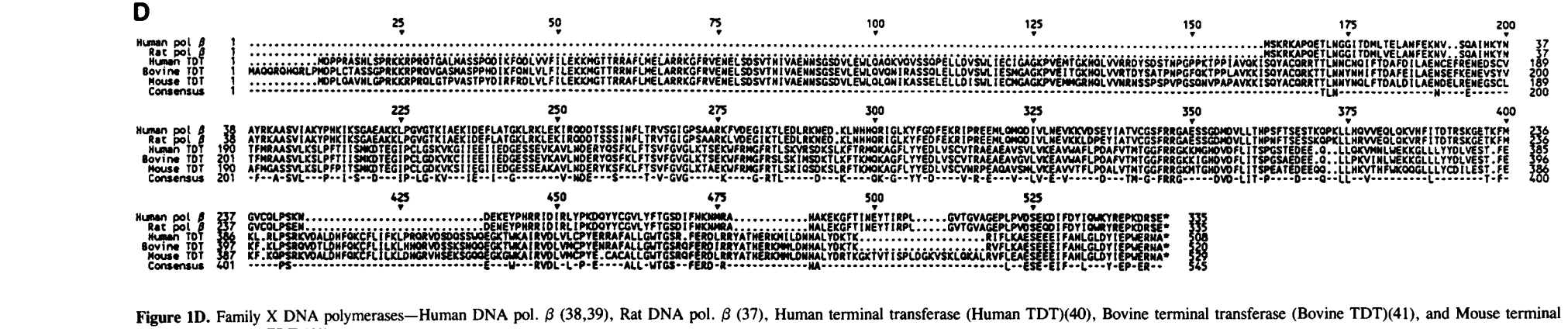


Figure 1D. Family X DNA polymerases—Human DNA pol. β (38,39), Rat DNA pol. β (37), Human terminal transferase (Human TDT)(40), Bovine terminal transferase (Bovine TDT)(41), and Mouse terminal transferase (Mouse TDT)(41).