

DNA sequence of the site-specific recombination function *cin* of phage P7

Wolfgang Ritthaler and Dietmar Kamp

Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, FRG

Submitted April 22, 1988

Accession no.X07724

We have determined the nucleotide sequence of the *cin* gene of phage P7, which encodes a DNA inversion function with a predicted molecular weight of 21246 D. The sequence of one strand was determined by Maxam and Gilbert sequencing from SphI and KpnI sites in plasmid pMR1 (1,2) and is shown together with the derived aminoacid sequence of *cin*. As previously described (1,2,3), the location and orientation of the *cin* gene of P7 with respect to the adjacent invertible C segment are identical to that of the related phage P1, i.e. the COOH-terminal part overlaps the outer end of the inverted repeat (IR []) sequence, which is the substrate for the *cin*-promoted recombination. Comparison with the DNA sequence of P1 (4) shows that the *cin* genes are 96% identical, and only altered nucleotides in P1 are shown. Three exchanges with respect to the aminoacid sequence do not affect the consensus sequence of the family of DNA invertases (5), so that we assume that P7 Cin is structurally and functionally indistinguishable from P1 Cin. In contrast to the well conserved coding region, the 5' upstream sequence is more diverged except for a "10"-like sequence (boxed), which therefore could be part of the promoter used in both systems. These potential Pribnow boxes are located in the left arm of a symmetrical AT-rich sequence (<---->), which could well fit a recognition site for a regulatory protein of yet unknown origin and function.

1 P7 5'-CAGGAACCTATTCAAGATTGTTATTCCTGTCATAACGC..TTTGTCTTAAATTCGGAGACTCTAAAGAACACGCTTATAGTAAAAATC
P1 5'-GGAAACCGTAGTACTTTGATTTCCACACATCTGGTTTTTGTTAAATTCGGAGCTTGTCTATAAAACGTAAGTGGCAGAAAAAT

MetLeuIleGlyTyrValArgValSerThrAsnGluGlnAsnThrAlaLeuGlnArgAsnAlaLeuGluSerAlaGlyCysGluLeuIlePheGluAspLys
ATGCTGTCGGCATGTACCGCTATAACAAATGAAACAAAACACTGCCTTCCACGAGTGCCTTGAAAGCCGAGGTGAGCTAATTGAGGATA
A A C T C

LysAla
 192 AlaSerGlyLysGlnThrLysIleArgProGlyLeuLysLysValIleArgMetLeuSerArgGlyAspThrLeuValValThrPlysLeuAspArgLeuGly
 AGCGCAGCCGAAACAGACTGAACGCCCTGGCTAAAAAAGGTTTGGCTATGCTTCCAGGGTGAACCCCTGGCTGTATGGAAGTTAGATCGCTTGG
 A G G T AC C T A C

ArgSerMetArgHisLeuValValLeuValIGIusIluLeuArgAspArgGlyIleAsnPheArgSerLeuThrAspSerIleAspThrSerThrProMet
292 GCGCCAGCATGCCTCACTTGGTTGTGCTGGTGAAGAGGCTGCGTGCACAGAGGAATTATTCGGAGGCTACTGACTCTACATGCACCCAGTACACCAATG
A C C C

392 GlyArgPhePhePheHisValMetGlyAlaLeuAlaGluMetGluArgGluLeuIleVaIg1uArgThrArgAlaGlyLeuAspAlaAlaArgAlaGluGly
GGGCCTTTTCTTCACGTAATGGGGGCGCTGGCAGAAATGGAACGTGAGCTTATCGTTGAACGTACACGGCTGGACTTGATGCGACGCTCGCAGAAG

Lys
Arg1leGlyArgArgProLysTyrGlnGluGluThrTrpGlnGlnMetArgArgLeuLeuGluAsnGlyIleProProArgLysGlnValAlaIleIle
492 GTCGTATAGGTGGCCGTCGCCCTAAATACCAAGAACATGGCAGCAAATCGGCCATTGCTGGAGATGGCATCCCCCGTAAGCAGGTTGCAATCAT
G

592 TyrAspVa1aLaValSerThrLeuTyrLysLysPheProAla1aSerGlyGlnSer * *
 592 CTATGATGTGGCTGTTCCACGCTTATAAGAAGTTCCGGCTCGTCATTCAATCCTAACCTGGTTAA
A IR

REFERENCES

1. Schmucker, R., Ritthaler, W., Stern, B. and Kamp, D. (1986) *J. gen. Virol.* **67**, 1123-1133
 2. Kamp, D., Kardas, E., Ritthaler, W., Sandulache, R., Schmucker, R. and Stern, B. (1984) *Cold Spring Harbor Symp. Quant. Biol.* **49**, 301-311
 3. Iida, S., Hiestand-Nauer, R., Meyer, J. and Arber, W. (1985) *Virology* **143**, 347-351
 4. Hiestand-Nauer, R. and Iida, S. (1983) *EMBO J.* **2**, 1733-1740
 5. Plasterk, R.H.A. and van de Putte, P. (1984) *Biochim. Biophys. Acta* **782**, 111-119