

Nucleotide sequence of two genes from *Helicobacter pylori* encoding for urease subunits

Christopher L. Clayton*, Mark J. Pallen, Harry Kleanthous, Brendan W. Wren and Soad Tabaqchali

Department of Medical Microbiology, St Bartholomew's Hospital Medical College, West Smithfield, London EC1A 7BE, UK

Submitted December 8, 1989

EMBL accession no. X17079

We have previously described the cloning of genes encoding the immunodominant urease subunits of *Campylobacter pylori* (now *Helicobacter pylori*)¹. The nucleotide sequence of the cloned DNA contains two open reading frames, capable of encoding proteins of 26.657 kD (α -subunit) and 60.473 kD (β -subunit). The ORFs are separated by three nucleotides. The ATG start codons are preceded by Shine-Dalgarno consensus sequences. No promoter-like sequences have been identified. The coding regions have a 44% G+C content. The derived amino-acid sequences of the two ORFs are directly homologous to the single subunit urease of the jackbean² (57% sites identical), and a fragment of the soy bean urease³. The *H. pylori* urease contains a histidine-rich region which might contain the nickel-binding active site. The N-terminus of the β -subunit is homologous to the N-termini of the large subunits of the ureases of *Klebsiella aerogenes* (61% identity) and *Proteus mirabilis* (71% identity)⁴.

ACKNOWLEDGEMENTS

Research supported by the Wellcome Trust and Medical Research Council. Thanks to J. Keyte for oligonucleotides. Some of the above information is included in U.K. Patent Application No. 8928625.6

REFERENCES

1. Clayton, C.L., Wren, B.W., Mullany, P., Topping, A. and Tabaqchali, S. (1989) *Infect. Immun.* **57**, 623–629.
2. Takishima, K., Suga, T. and Mamiya, G. (1988) *Eur. J. Biochem.* **175**, 151–165.
3. Krueger, R.W., Holland, M.A., Chisholm, D. and Polacco, J.C. (1987) *Gene* **54**, 41–50.
4. Mobley, H.L.T. and Hausinger, R.P. (1989) *Microb. Rev.* **53**, 85–108.

```

GCTCCAGGTCACGGATCTCGTTATGCTTCACGAAAAACCTTTANGAATAG
GAGAATGAGATGAAGCTCACCCCAAAGAGTTAGACAGTTGATGCTCCACTAC
  H K L L T P K F L D R L M L H Y
GCTGGAGAAATAGCTAAAAGCCAAAGAAAAGGCATTAAGCTTAACCTATGTG
A G E L A K K R K E K G I K L N Y V
GAAGCCGCTAGTTGATGAGTCCCATATTAATGAGAGCCGACGCTGGTATAA
E A V R L I S A H I M E E A R R G K
AAGACTGCGCTGAATGATCCAAAGAGGGCCACTCTTTAAACCGATGAT
K T A A E L M Q E G R T L L K P D D
GTGATGGATGGCTGGCAGCATGATCCATGAGTGGGTATTGAGGCGATGTT
V M D G V A S M I H E V G I E A M F
CCTGATGGGCAAACTCGTAACCGTGCATACCCCTATTAGGCCAATGGTAAA
P D G T K L V T V H T P I E A N G K
TTAGTCCCTGGTGAAGTTGTTCTAAAAGATGAGACATCACTATCAACGAGGC
L V P G E L F L K N E D I T I N E G
AAAAAGCCGTTAGCGTAAAGTTCCCGCTGTGGCAGCAGCCGGTTCAAAATC
K K A V S V K V P P V G D R P V Q I
GGCTCACATCCATTTCTTGAAGTGAATAGATGCTTAGACTTTGACAGAGAA
G S H F H F F E V N R C L D F D R E
AAAACTTCCGTAAGCGTTAGACATTCGCGAGGGGACAGCGGTAAAGTTGAG
K T F G K R L D I A S G T A V R F E
CCTGGCAGAAAATCCGTTAGAGTTGATGACATTTGGCGGTAACGAGAATC
P G E E K S V E L I D I G G N R R I
TTTGGATTAAAGCGTTGGTTGATAGCAAGCCGATACGAAAACAAAATTT
F G F H A L V D R Q A D N E S K K I
GCTTTACAGAGCTTAAGAGCGTGGTTTTCATGGCCGTAAGAGGATGACAA
A L H R A K E R G F H G A K S D D N
TATGTAAGAACTAAGAGTANGAA ATGAAAAGATTAAGCAGAAAGAA
Y V K T I K E * H K K I S R K E
TATGCTTCATGATGGCCCTACTACAGGCGTAAAGTGAATGAGGATGATCA
Y A S M Y G P T T G A K V R L G D T
GACTTTGATCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG
D L I A E V E H D Y T I Y G E E L K
TTCCGCGCGGTAAGAACTTAAGAGAGGCAATGAGCCAACTCAACAACTAGC
F G G G K T L R E G M S Q S N N P S
AAAGAAGAACTGGATCTAATCATCACTAAGCTTTAATCGTGGATTACACCGGT
K E E L D L I T T N A L I V D Y T G
ATTTAAGAGCGGATATGGTATAAAGATGGCAAAATCGCTGGCATTGGTAAA
I Y K R D I G T R D G K T A G I G K
GGCGTAAACAGACAGCAGAGTGGTTAAAACAACTCTTAAGCTGGTCCCT
G G N K D T Q D G V K N N I S V G P
GCTACTGAAGCCTTAGCCGTTGAAGTTGATGTAACCTGCTGGTGGTATTGAC
A T E A L A G E G L I V T A G G I D
ACACACATCCACTTCATCCCCCAACAAATCCCTACAGCTTTTGGCAAGCGGT
T H I H F I S P Q I P T A F A S G
GTAAACACCATGATGGTGGCGGAACGCGCCCTGCTAGGCAATACCGCAAC
V T T M I G G G T G P A D G T N A T
ACTATCACCTAGGTAAGAAATTTAAATTCATGCTCAGAGCGGCTGAAGAA
T I T P G R R N L K F M L R A A E A

```

```

TATTCATGAACTTGGTTCTTGGTAAAGGTAACGCTTCTAANGATGCAAGC
Y S N M F G F L A K G N A S N D A S
TTAGCCGATCAAAATGAAGCTGGTGCGATTGGCCCTTAAAATCCAGAAAGCTGG
L A D Q I E A G A I G L K I H E D W
GFCACCACTCCTTTCGCAATCAATCGCTTAGATGTTGGGCAAAATACGAT
G T T P S A I N H A L D V A D K Y D
GTGCAAGTCCGATCCACAGACACTTGAATGAAGCCGGTTCGCTGGAAAGC
V Q V A I H T D T L N E A G C V E D
ACTATGGCAGCTATTCGCGGAGCCACTATGACACTTACCACACTGAAGCGCT
T M A A I A G R T N H T V H T E G A
GGCGGGGACAGCCCTCGATATTAATGAAGTGGCGGTAACACAACTA
G G G H A P D I I K V P G F H N I L
CCCGTTCAGTACCCCACTTCCCTTCCCGTGAATACAGAGCCGAGAC
P A S T N P T I P F T V N T E A H
ATGGACATGCTTGGTGGCCACACTTGGATAAAGCAATTAAGAAATGTC
M D M L M V C H H L D K S I K E D V
CAGTTCGCTGATCAAGGATTCGGCCCTCAAAACATTCGGCTGAAGACACTTGG
Q F A D S R I R P Q T I A A E D T L L
CATGACATGGGATTTCTCAATCACTAGTTCTGACTCTCAAGCCGATGGCCGT
H D M G I F S I T S D S Q A M G R
GTGGTGAAGTATCACTAGACTTGGCAAGCAGCTGACAAAATAAAGAA
V G E V I T R T W Q T A D K N K K E
TTTGGCGCTTGAAGAGAAAAGGCGCTAACGCAACTTCAGGATCAAAAGC
F G R L K E E K G D N D N F R I K R
TACTTGTCAAATACAGCTTAACCCAGCGATCGCTCATGGGATTAAGCGAT
Y L S K Y T I N P A I A H G I S E Y
GTGGTTCGTGAAGTGGCAAGTGGCTGACTGGTATTGGGATCCCGCA
V G S V E V G R V A D L V L M S P A
TTCTTGGTGAAGCCACATGATCAATCAAGCCGGTTCATCGCATTGAGT
F F G V K P N H L I K G G F I A L S
CAAAATGGTGAAGTGGCAAGCTTCTATCCCTACCCCAAGCAAGCTTATTACAGA
Q M G D A N A S I P T P Q P V Y R
GAAATGTCCGCTCACTAGTAAAGCAATACCTAAAAGACATCACTTTGAG
E M F A H H G K A K Y D A N I T F V
TCTCAAGCCGCTTATGCAAGGCAATTAAGAAAGATTAAGGCTTGAAGACAA
S Q A A Y D K G I K E E L G L E R Q
GTGTGGCGGTAAGAAATTCGAGAAATACCTAAAAGACATGCAATCAAC
V L P V K N C R N I T K K D M Q F N
GACACTACCGCTACATGAGTCAATTCGAAACTTACCATGTTCGTTGGAT
D T T A H I E V N S E T Y H V F V D
GGCAAGAGTAACTCAAGCCCAATAAAGTGAAGTGGCTTGGCAACTCTTTA
G K E V T L N Q P I K *
GCATTTCTAGGATTTTATGAGCAAGCTTCTTAACTCTAGTTTTAGCT
CTCTGATTTTTGTATCAAAAATTTGGGGCTTTTTTTGTTTTATTTTTG
TCAATTTACTATTTTCTTATGATAGCTAGCAGCAAGAAATGTTATGCTAAG
GTGGTTTTGTAAAATTTTAAATTCATGCTCAGAGCGGCTGAAGAA
TATTTGATATGTTGGATGTTTTAATCGCAATGGGATTTGGCGTTAAACA
AAGTGCATCTAGAG

```

* To whom correspondence should be addressed