28. Morishita, W., Kirov, S. A. & Alger, B. E. Evidence for metabotropic glutamate receptor activation in the induction of depolarization-induced suppression of inhibition in hippocampal CA1. *J. Neurosci.* **18,** 4870–4882 (1998).
29. Marr, D. A theory of cerebellar cortex. *J. Physiol. (Lond.)* **202,** 437–470 (1969).
30. Ito, M. in *The Cerebellum and Neural Control.* (Raven, New York, 1984).

Acknowledgements

This work was supported by The Wellcome Trust, the European Union and the Medical Research Council (Research Studentship to S.J.M.). We thank A. Momiyama and T. Takahashi for helpful discussion, J. Clements for providing Axograph and D. Attwell, M. Farrant, M. Häusser, B. Katz and T. Takahashi for comments on the manuscript.

Correspondence and requests for materials should be addressed to R.A.S. (e-mail: a.silver@ucl.ac.uk).

Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491

J. Parkhill*, M. Achtman†, K. D. James*, S. D. Bentley*, C. Churcher*, S. R. Klee†, G. Morelli†, D. Basham*, D. Brown*, T. Chillingworth*, R. M. Davies*, P. Davis*, K. Devlin*, T. Feltwell*, N. Hamlin*, S. Holroyd*, K. Jagels*, S. Leather*, S. Moule*, K. Mungall*, M. A. Quail*, M.-A. Rajandream*, K. M. Rutherford*, M. Simmonds*, J. Skelton*, S. Whitehead*, B. G. Spratt‡ & B. G. Barrell*

* *The Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*
† *Max-Planck Institut für moleculare Genetik, Ihnestraße 73, D-14195 Berlin, Germany*
‡ *Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3FY, UK*

***Neisseria meningitidis* causes bacterial meningitis and is therefore responsible for considerable morbidity and mortality in both the developed and the developing world. Meningococci are opportunistic pathogens that colonize the nasopharynges and oropharynges of asymptomatic carriers. For reasons that are still mostly unknown, they occasionally gain access to the blood, and subsequently to the cerebrospinal fluid, to cause septicaemia and meningitis. *N. meningitidis* strains are divided into a number of serogroups on the basis of the immunochemistry of their capsular polysaccharides; serogroup A strains are responsible for major epidemics and pandemics of meningococcal disease, and therefore most of the morbidity and mortality associated with this disease. Here we have determined the complete genome sequence of a serogroup A strain of *Neisseria meningitidis*, Z2491 (ref. 1). The sequence is 2,184,406 base pairs in length, with an overall G+C content of 51.8%, and contains 2,121 predicted coding sequences. The most notable feature of the genome is the presence of many hundreds of repetitive elements, ranging from short repeats, positioned either singly or in large multiple arrays, to insertion sequences and gene duplications of one kilobase or more. Many of these repeats appear to be involved in genome fluidity and antigenic variation in this important human pathogen.**

The genome encodes complete sets of enzymes for glycolysis (apart from *fruK* and *pfkA*), gluconeogenesis, the pentose-phosphate and Entner–Doudoroff pathways, the pyruvate dehydrogenase complex and the trichloroacetic acid cycle. In addition to the aerobic respiration genes, there are also both nitrite (*aniA*) and nitrate (NMA1886) reductases; some capability for fermentation is also apparent. *N. meningitidis* appears to be capable of *de novo* synthesis of most of the amino acids (with the exception of asparagine and methionine) and purine and pyrimidine nucleotides; however, the pathways for production of folic acid, molybdopterin, pantothenate and pyridoxine are incomplete. All of the aminoacyl transfer RNA synthetases are present except tRNAAsn. As *N. meningitidis* encodes both glutaminyl tRNA synthetase and the three-subunit Glu-tRNAGln transamidase, the latter may be responsible for the production of tRNAAsn by the transamidation of tRNAAsp[2].

In addition to the 2,121 predicted coding sequences (CDSs), there are 4 copies of a 16S-23S-5S ribosomal RNA operon, 58 tRNAs, 1 tmRNA (10Sa RNA) and the RNA component of RNAase P. The average gene length is 877 base pairs (bp), at a density of 1 per 1.03 kilobases (kb). The overall coding density is 82.9%, higher than *Rickettsia*[3], but lower than most other sequenced bacteria. Base 1 of the sequence was chosen to correspond roughly with 12 o'clock on the published map[1]; the origin of replication (as indicated by the bias towards G on the leading strand[4]) is around 247,600 bp. Intriguingly, although the GC bias is similar in both replichores, there is a clear bias (60.2%) of CDSs towards the leading strand in only one replichore (anticlockwise in Fig. 1), with no coding bias at all on the opposite replichore. The genome contains at least 56 pseudogenes, of which 17 are remnants of insertion sequence (IS) elements, and there are at least 5 complete or partial prophages (pnm1–5).

The G+C content of the genome is extremely variable, with at least 60 coding regions (nearly 5% in total) having a significantly lower G+C content, ranging in size from 224 bp to 11.3 kb and averaging 1.8 kb (see Supplementary Information). These regions may represent recently acquired DNA, and given the natural competence of Neisserial species[5], this is perhaps unsurprising. Although most genes in these regions have no database matches, or are conserved hypothetical proteins, nearly one-third encode

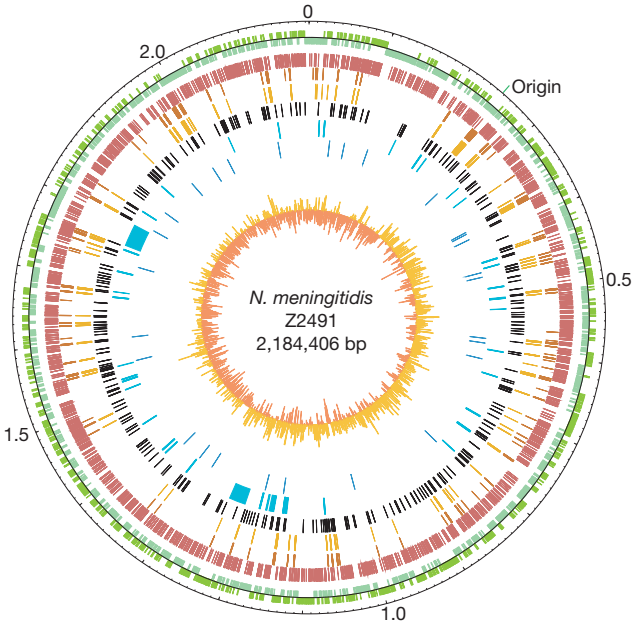

Figure 1 Circular representation of the *N. meningitidis* Z2471 genome. The concentric circles show, reading inwards: the scale in megabases, with the origin of replication indicated; predicted coding sequences clockwise (dark green) and anti-clockwise (light green); neisserial uptake sequences (red); dRS3 requences (dark orange); RS elements (light orange); dispersed repeats (Correia, ATR, REP2-5; black); IS elements and phage (narrow ticks and wide bars respectively; turquoise) and tandem repeats (dark blue). The inner histogram shows plot of (G-C)/(G+C) with values greater than zero in yellow and less than zero in orange. Figure generated with LASERGENE software (DNAStar).

proteins that are likely to be on the surface of the cell, or are responsible for the production of surface structures. Prominent among these are the well characterized CDSs of the serogroup A capsule cassette (*sacA-D*) which produce a serogroup-A-specific sugar residue in the capsular polysaccharide[6]; these are located in an extremely well defined region of only 29.7% G+C. Also notable are a single glycosyl-transferase (NMA0832), similar to capsule biosynthesis genes in other bacteria, and a pilin homologue (NMA0424). Some surface-exposed protein-encoding genes are each immediately followed by large regions of A+T-rich DNA encoding numerous small hypothetical proteins; these include the three *mafB* family genes and NMA1083. Eight A+T-rich regions encode DNA restriction/modification enzymes. Some caution is required in interpreting a bias in G+C composition as necessarily indicating recently acquired DNA; the largest region of A+T-rich DNA (41.7% G+C over 11.3 kb) encompasses much of the ribosomal protein operon.

One of the most striking and unique characteristics of the Z2491 genome is the abundance and diversity of repetitive DNA (Fig. 1). The most obvious example of this abundance is the Neisserial uptake sequence, which is involved in the recognition and uptake of DNA from the environment[7]. There are nearly two thousand copies of the 10-bp uptake sequence, which either occurs alone or in inverted repeats as part of a transcriptional terminator[7]. Other repetitive sequence elements in the *N. meningitidis* genome are concentrated within intergenic repeat arrays of 200–2,700 bp. These repeat arrays are composed of several different repeat types (Fig. 2 and Table 1), the most abundant of which are the 'neisserial intergenic mosaic elements' (NIMEs), which comprise repeat units of ~50–150 bp (RS elements), each flanked by 20-bp inverted repeats (dRS3 elements); there are between 1 and 60 copies of 117 different families of RS element.

Also present in the repeat arrays are larger units, which are also found in isolation, including the 'Correia elements' (CEs, 156-bp sequences bounded by 26-bp inverted repeats) that have been previously described in both *N. meningitidis* and *Neisseria gonorroheae*[8]. Almost one-third of the 257 CEs share a common 51-bp internal deletion, which presumably has not affected the ability of the element to be mobilized, and 29 have lost one or both of their terminal inverted repeats. Correia elements in repeat arrays are sometimes flanked by additional conserved sequences upstream and downstream (Fig. 2). A number of both complete and internally deleted CEs occur embedded within other sequence features, including genes, RS elements, other CEs and ISs. The pseudogenes NMA0059 (a modification methylase) and NMA0530 (a hypothetical protein) both appear to have been truncated or interrupted by CE insertion. Two repeat types are dispersed over the genome at low frequency, both alone and within repeat arrays: 19 copies of a 183-bp A+T-rich (30% G+C) repeat (ATR) whose ends form an imperfect 35-bp inverted repeat and 26 copies of a 120–150-bp repeat (REP 2). REP 2 occurs immediately upstream of 16 CDSs and contains a ribosome-binding-site-like conserved AAGGA motif within 5–13 bp of the predicted start codon. It seems likely that sequences within this repeat would exert some effect on the expression of these genes. The repeat arrays appear to be a target for the insertion of the mobile elements in the genome, with almost half of the 43 complete and partial copies of IS*1016*, IS*1106* and IS*1655* being integrated in or near to a repeat array.

Given the natural competence of neisserial species, the presence of such arrays could encourage sequence variation by acting as a target for a specific recombinase, thereby increasing the rate of horizontal gene transfer at the flanking locus, resulting in rapid allelic replacement or the creation of new mosaic alleles. A previously characterized example of enhanced recombinational exchange involves the rearrangements responsible for the modulation of *pilE* expression. PilE is the main structural pilin component and is expressed from the intact *pilE* gene. In *N. gonorrhoeae*, 'silent' pilin genes (*pilS*) exist, which contain only the 3′ region of the gene without a start codon or transcriptional signals[9]. Coding sequences from these silent regions can be recombined into the expression locus, in a *recA/Q/O*-dependent manner, changing the expressed PilE sequence[10]. Models put forward for this process suggest that the gene conversion may be intergenomic[11]. In the *N. meningitidis* Z2491 sequence, the *pilE* gene is surrounded by repeat arrays consisting of NIMEs and a CE (Fig. 3a). Immediately upstream of the *pilE* gene are eight *pilS* loci, each without an 5′ end and intimately bounded by arrays of NIMEs, suggesting that NIME sequences may be involved in *pilS/E* recombination. The *N. gonorrhoeae* Sma/Cla repeat, implicated in enhancing *pilE/S* recombination in *N. gonorrhoeae*[12], exists as a single copy in *N. meningitidis* Z2491, immediately downstream of the *pilE* gene.

Our analysis of the occurrence of repeat arrays suggests that antigenic variation mediated by gene conversion, similar to that of *pilE*, may occur widely in the *N. meningitidis* genome. In *N. meningitidis* Z2491, most shorter repeat arrays are associated with genes directly or indirectly related to cell-surface functions (including pilins, substrate-binding proteins and transporter components), and the larger repeat arrays are exclusively associated with such genes (Fig. 4; and Table 2, Supplementary Information). The

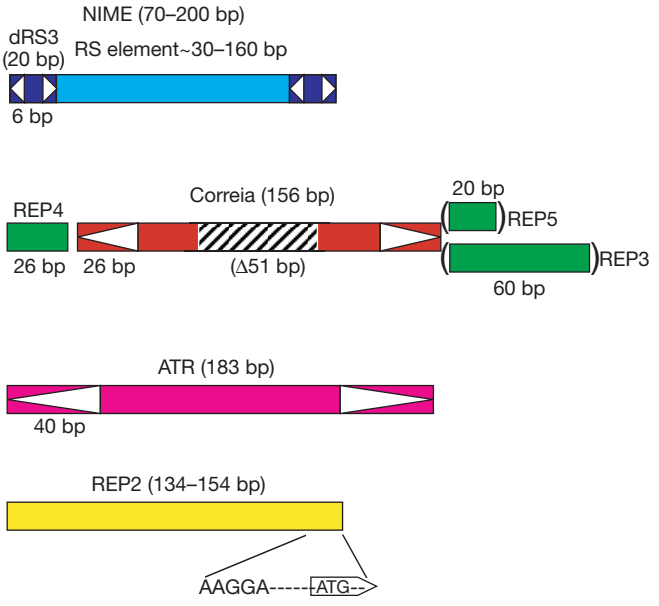

Figure 2 Types of *N. meningitidis* repeat. The name of each repeat type is indicated above the repeat. Inverted repeat sequences are represented by open triangles, and the internal deletion present in some Correia elements by a hatched box. The translational signals within REP2 are indicated below the repeat.

Table 1 Number of repeats by type in *N. meningitidis* Z2491

Type	Size (bp)	Frequency
DNA uptake sequence: gccgtctgaa	10	1,892
RS	24–161	681
dRS3: attcccnnnnnnnnngggaat	20	772
Correia (full)	150–159	173
Correia (internal deletion)	~104	84
Correia (partial)	37–145	29
ATR	183	19
REP 2	59–154	26
REP 3	60	13
REP 4	26	20
REP 5	20	9
IS*1016*	256–740	14 (including partial)
IS*1106*	263–1219	22 (including partial)
IS*1655*	1,074–1,257	7 (including partial)
Prophage	2,330–38,964	5

products of these genes are potential targets for the immune system, and antigenic variation of these proteins would be beneficial in evading the immune response. The genes encoding the outer membrane proteins PorA (NMA1642) and PorB (NMA0398) are flanked on both sides by arrays of repeats, as are *lbpAB* (ref. 13) (Fig. 3b), *tbpAB* (ref. 14) and *hpuAB* (ref. 15), which encode lactoferrin-, transferrin- and haemoglobin-/haptoglobin-binding proteins, respectively. The putative ferric enterobactin-binding protein genes *fetABI* are also associated with two such arrays. In contrast, the neisserial ferric binding protein operon *fbpABC* (ref. 16), which may not be exposed to the immune response, the TonB-dependent receptor NMA0577 and the haemoglobin receptor *hmbR* (ref. 17) are not associated with repeat elements (a poly-(G) tract is present in *hmbR*, potentially allowing phase-variable expression; see below). The *opaD*, *opaB*, *opaA* and *opcA* genes, which are involved in various interactions with the host cell, are all flanked on one or both sides by repeat arrays. Comparison of these regions from a range of *N. meningitidis* isolates has shown that the repeat arrays themselves are highly variable between diverse *Neisseriae* but may be identical in related strains even after more than 30 years of epidemic spread (G.M., unpublished data).

Repeat-mediated gene rearrangements are also apparent at other sites, although, in these cases, the repeats are local. One such site encompasses the *mafA* and *mafB* genes; *mafA* is predicted to be a lipoprotein and *mafB* to be secreted; both have been proposed to have adhesin activity in *N. gonorrhoeae* (S. Eickernjaeger *et al.*, personal communication; EMBL accession number AF142582). The *N. meningitidis* orthologue of *mafB* is immediately followed by a 6.5-kb region of low G+C DNA encoding mainly short genes of unknown function. Interspersed with these genes, and with a normal G+C content, are three repeats of 303 bp and one of 79 bp from *mafB*, each corresponding to the beginning of an open reading frame (ORF) that has no start codon. This arrangement suggests that these repeat sequences may be capable of recombination

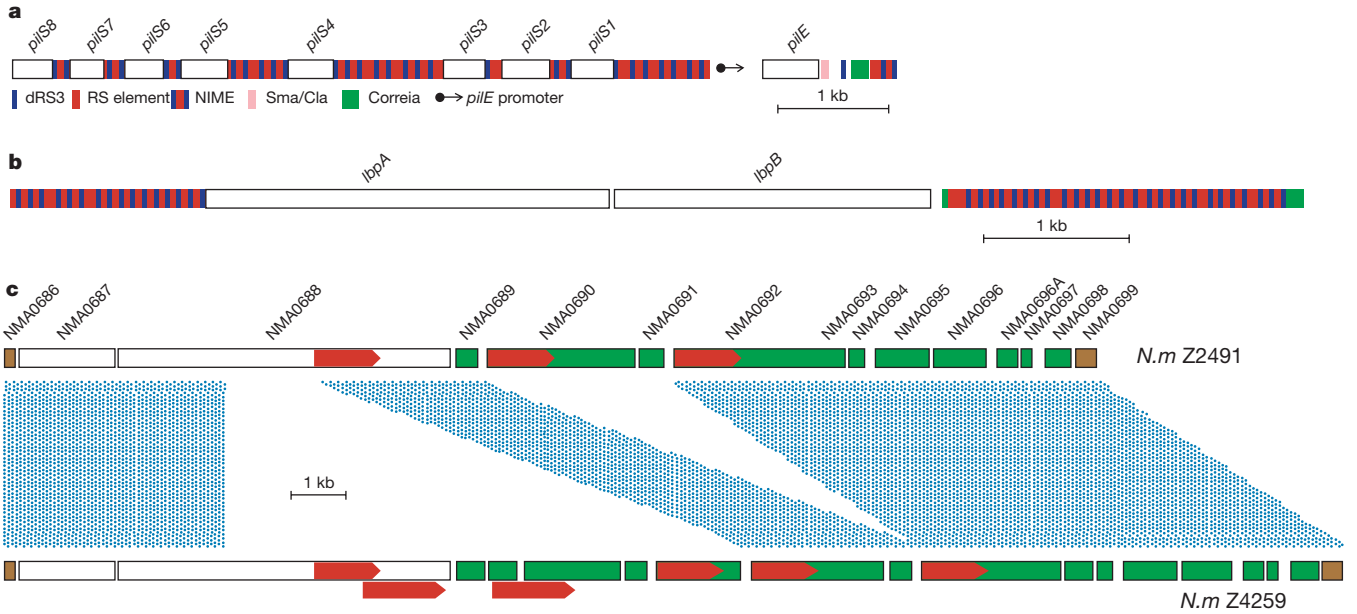

Figure 3 Structure of selected repeat regions. **a**, Repeats around the *pilE/S* locus. *pil* genes are indicated by open boxes and repeats by coloured boxes as described in the key. **b**, Large repeat arrays around the *lbpA* and *lbpB* genes. Repeats are coloured as in **a**. **c**, The filamentous haemagglutinin homologue gene sequences of *N. meningitidis* Z2491 and Z4259. Conserved sequences are indicated by light blue bars, and internal repeat regions by red arrows. The complex repeat structures have been simplified for clarity. Open boxes represent genes encoding surface-exposed proteins, and green boxes represent genes with no database matches. The brown boxes represent the two halves of an ABC transporter pseudogene.

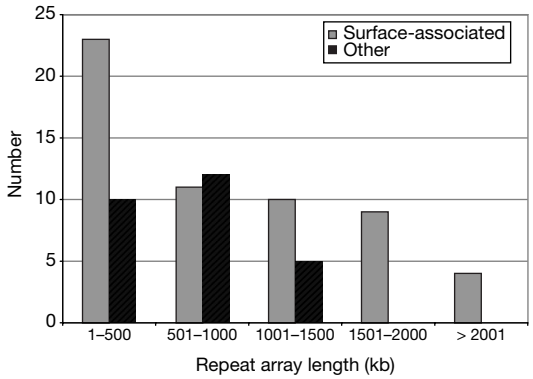

Figure 4 Graph showing the relationship between repeat array length and flanking gene function. For each category of repeat array length the number of flanking genes associated with surface structures is shown in grey, and the number of genes in all other categories is shown hatched.

with the *mafB* gene, leading to the attachment of different 3′ sequences. This interpretation is supported by the observation that the 3′ sequence of the gonococcal *mafB* orthologue is almost identical to the ORF following the first 303-bp repeat, and not to the 3′ sequence of the meningococcal *mafB* gene.

A second example of repeat-mediated rearrangement of the 3′ ends of genes encoding surface-exposed proteins is provided by NMA0688, a homologue[18] of the *Bordetella pertussis* filamentous haemagglutinin *fhaB* (ref. 19). The sequence downstream of NMA0688 contains complex repeated sequences from NMA0688 (Fig. 3c) and encodes several proteins of unknown function. Sequencing this region from *N. meningitidis* Z4259 (a serogroup C strain; EMBL accession number AJ391284) revealed a similar repeat structure, but with additional DNA inserted into the 3′ sequence of NMA0688 at one of the repeats. This has the effect of changing the 3′ end of the NMA0688 gene, as well as introducing additional genes, some containing similar repeats. The 3′ region of the Z2491 NMA0688 gene is within a downstream ORF in Z4259. It therefore seems likely that, like *mafB*, the 3′ end of this *fhaB* homologue can be altered by recombination with alternative 3′ ends. Intriguingly, the entire region consisting of the *fhaC* (NMA0687) and *fhaB* homologues, along with the downstream DNA, appears to have been inserted into the coding sequence of an ABC transporter pseudogene (NMA0686/NMA0699)

Tandem repeats of nucleotides in *Neisseria* are subject to lengthening or shortening during replication owing to a phenomenon called slipped-strand mispairing[20]. When these tandem repeats are present in coding sequences, this change in length can change the translation state of the gene, leading to an on/off switching of the gene product called 'phase variation'. The compete genome sequence reveals around 26 tandem repeats indicating potentially phase-variable genes (Table 3, Supplementary Information), many of which have been studied before. These repeats range from homopolymeric tracts of G or C nucleotides to di-, tetra- and penta-nucleotide repeats. As would be expected, most phase-variable genes are either surface-exposed or involved in the biosynthesis or modification of surface structures, which supports their suggested role in virulence and/or immune avoidance. Four of the tandem repeats are not directly within coding sequences, but are immediately upstream of candidate variable genes, and these may affect transcription from gene-specific promoters, as has been shown for *opcA* (ref. 21) and *porA* (ref. 22).

The analysis presented here supports the general operation of three mechanisms of repeat-mediated antigenic variation within the *N. meningitidis* genome: on/off switching and transcriptional modulation of gene expression by slipped-strand mispairing of short tandem repeats; intragenomic recombination of localized repeats leading to the use of different carboxy termini for surface-exposed proteins; and intergenomic gene conversion of specific surface-associated genes associated with large arrays of global repeats, mediated by the internalization of related DNA through the highly repetitive DNA uptake sequence. □

Methods

DNA was prepared from *N. meningitidis* Z2491 as described[23]. The DNA was fragmented by sonication, size-fractionated on an agarose gel, and two libraries were generated in pUC18 using size fractions ranging from 0.5 to 1.5 kb. Roughly 37,500 pUC clones were sequenced from both ends using Dye-terminator chemistry on ABI 373 and 377 sequencing machines; 58,269 reads were used to generate the final assembly, giving about a 10-fold coverage of the genome. Sequence assembly was accomplished using Phrap (P. Green, unpublished), and the sequencing was finished using GAP4 (ref. 24). The assembly was verified by genomic PCR reactions across all unbridged large repeats, in addition to 260 forward and reverse reads from a random library of 20–22 kb cloned in lambdaFixII (Stratagene), and 610 forward and reverse reads from 3 libraries of 9–13 kb cloned in pSP64 (Promega). The final assembly was checked against the published map using the positions of restriction sites, mapped genes and lambda clones. In the final assembly, less than 0.07% of the genome was covered by a single clone only, and less than 0.12% was unsequenced on both strands, or with complementary sequencing chemistries.

The DNA was compared with sequences in the EMBL database using BLASTN and BLASTX[25]. Transfer RNAs were predicted by tRNAscan-SE (ref. 26). Potential CDSs were predicted using ORPHEUS[27] and GLIMMER[28] (both trained on an initial ORF set generated by ORPHEUS), and the results were combined and checked manually. The predicted protein sequences were searched against a non-redundant protein database using WUBLASTP and FASTA. The complete six-frame translation was used to search PROSITE, and the predicted proteins were compared against the PFAM database of protein domain hidden Markov models[29]. The results of all these analyses were assembled together using the Artemis sequence viewer (K.M.R., unpublished) and used to inform a manual annotation of the sequence and predicted proteins. Annotation was based, wherever possible, on characterized proteins or genes. Repeat sequences were identified using HMMER (S. Eddy, personal communication) and the EMBOSS program 'profit' (http://www.sanger.ac.uk/Software/EMBOSS/), based on sequence alignments generated with ClustalW[30].

Received 1 March; accepted 8 March 2000.

1. Dempsey, J. A., Wallace, A. B. & Cannon, J. G. The physical map of the chromosome of a serogroup A strain of *Neisseria meningitidis* shows complex rearrangements relative to the chromosomes of the two mapped strains of the closely related species *N. gonorrhoeae*. *J. Bacteriol.* **177,** 6390–6400 (1995).
2. Curnow, A. W., Tumbula, D. L., Pelaschier, J. T., Min, B. & Soll, D. Glutamyl-tRNA (Gln) amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis. *Proc. Natl Acad. Sci. USA* **95,** 12838–12843 (1998).
3. Andersson, S. G. *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396,** 133–140 (1998).
4. Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13,** 660–665 (1996).
5. Smith, N. H., Holmes, E. C., Donovan, G. M., Carpenter, G. A. & Spratt, B. G. Networks and groups within the genus *Neisseria*: analysis of *argF*, *recA*, *rho*, and 16S rRNA sequences from human *Neisseria* species. *Mol. Biol. Evol.* **16,** 773–783 (1999).
6. Swartley, J. S. *et al.* Characterization of the gene cassette required for biosynthesis of the (α1→6)-linked *N*-acetyl-D-mannosamine-1-phosphate capsule of serogroup A *Neisseria meningitidis*. *J. Bacteriol.* **180,** 1533–1539 (1998).
7. Goodman, S. D. & Scocca, J. J. Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. *Proc. Natl Acad. Sci. USA* **85,** 6982–6986 (1988)
8. Correia, F. F., Inouye, S. & Inouye, M. A family of small repeated elements with some transposon-like properties in the genome of *Neisseria gonorrhoeae*. *J. Biol. Chem.* **263,** 12194–12198 (1988).
9. Haas, R. & Meyer, T. F. The repertoire of silent pilos genes in *Neisseria gonorrhoeae*: Evidence for gene conversion. *Cell* **44,** 107–115 (1986).
10. Mehr, I. J., & Seifert, H. S. Differential roles of homologous recombination pathways in *Neisseria gonorrhoeae* pilin antigenic variation, DNA transformation and DNA repair. *Mol. Microbiol.* **30,** 697–710 (1998).
11. Howell-Adams, B., Wainwright, L. A. & Seifert, H. S. The size and position of heterologous insertions in a silent locus differentially affect pilin recombination in *Neisseria gonorrhoeae*. *Mol. Microbiol.* **22,** 509–522 (1996).
12. Wainwright, L. A., Pritchard, K. H. & Seifert, H. S. A conserved DNA sequence is required for efficient gonococcal pilin antigenic variation. *Mol. Microbiol.* **13,** 75–87 (1994).
13. Pettersson, A., Prinz, T., Umar, A., van der Biezen, J. & Tommassen, J. Molecular characterization of LbpB, the second lactoferrin-binding protein of *Neisseria meningitidis*. *Mol. Microbiol.* **27,** 599–610 (1998).
14. Vonder Haar, R. A., Legrain, M., Kolbe, H. V. & Jacobs, E. Characterization of a highly structured domain in Tbp2 from *Neisseria meningitidis* involved in binding to human transferrin. *J. Bacteriol.* **176,** 6207–6213 (1994).
15. Lewis, L. A., Gray, E., Wang, Y. P., Roe, B. A., & Dyer, D. W. Molecular characterization of hpuAB, the haemoglobin-haptoglobin-utilization operon of *Neisseria meningitidis*. *Mol. Microbiol.* **23,** 737–749 (1997).
16. Adhikari, P., Berish, S. A., Nowalk, A. J., Veraldi, K. L., Morse, S. A. & Mietzner, T. A. The *fbpABC* locus of *Neisseria gonorrhoeae* functions in the periplasm-to-cytosol transport of iron. *J. Bacteriol.* **178,** 2145–2149 (1996).
17. Stojiljkovic, I. *et. al.* The *Neisseria meningitidis* haemoglobin receptor: its role in iron utilization and virulence. *Mol. Microbiol.* **15,** 531–541 (1995).
18. Klee, S. R. *et al.* Molecular and biological analysis of eight genetic islands that distinguish *Neisseria meningitidis* from the closely related pathogen *Neisseria gonorrhoeae*. *Infect. Immunity* (in the press).
19. Domenighini, M. *et al.* Genetic characterization of *Bordetella pertussis* filamentous haemagglutinin: a protein processed from an unusually large precursor. *Mol. Microbiol.* **4,** 787–800 (1990).
20. Henderson, I. R., Owen, P. & Nataro, J. P. Molecular switches–the ON and OFF of bacterial phase variation. *Mol. Microbiol.* **33,** 919–932 (1999).
21. Sarkari, J., Pandit, N., Moxon, E. R. & Achtman, M. Variable expression of the Opc outer membrane protein in *Neisseria meningitidis* is caused by size variation of a promoter containing poly-cytidine. *Mol. Microbiol.* **13,** 207–217 (1994).
22. van der Ende, A. *et al.* Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the-10 and-35 regions of the promoter. *J. Bacteriol.* **177,** 2475–2480 (1995).
23. Zhou, J. & Spratt, B. G. Sequence diversity within the *argF*, *fbp* and *recA* genes of natural isolates of *Neisseria meningitidis*: interspecies recombination within the *argF* gene. *Mol. Microbiol.* **6,** 2135–2146 (1992).
24. Bonfield, J. K., Smith, K. F. & Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res.* **23,** 4992–4999 (1995).
25. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
26. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25,** 955–964 (1997).
27. Frishman, D., Mironov, A., Mewes, H. W. & Gelfand, M. Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic Acids Res.* **26,** 2941–2947 (1998).
28. Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* **26,** 544–548 (1998).

29. Bateman, A. *et al.* Pfam 3.1: 1,313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res.* **27,** 260–262 (1999).
30. Higgins, D. G., Bleasby, A. J. & Fuchs, R. CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.* **8,** 189–191 (1992).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

This research was funded by The Wellcome Trust.

Correspondence and requests for materials should be addressed to J.P. (e-mail: parkhill@sanger.ac.uk). The complete sequence and annotation can be obtained from the EMBL database with the ID NMAZ2491 (accession number AL157959), and from our web pages (http://www.sanger.ac.uk/Projects/N_meningitidis).

The duration of antigen receptor signalling determines CD4$^+$ versus CD8$^+$ T-cell lineage fate

Koji Yasutomo*, Carolyn Doyle†, Lucio Miele‡ & Ronald N. Germain*

* *Lymphocyte Biology Section, Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA*
† *Department of Immunology, Duke University Medical Centre, Durham, North Carolina 27710, USA*
‡ *Cancer Immunology Program, Cardinal Bernardin Cancer Centre, Loyola University Medical Centre, Maywood, Illinois 60153, USA*

Signals elicited by binding of the T-cell antigen receptor and the CD4/CD8 co-receptor to major histocompatibility complex (MHC) molecules control the generation of CD4$^+$ (helper) or CD8$^+$ (cytotoxic) T cells from thymic precursors that initially express both co-receptor proteins[1]. These precursors have unique, clonally distributed T-cell receptors with unpredictable specificity for the self-MHC molecules involved in this differentiation process[2]. However, the mature T cells that emerge express only the CD4 (MHC class II-binding) or CD8 (MHC class I-binding) co-receptor that complements the MHC class-specificity of the T-cell receptor. How this matching of co-receptor-defined lineage and T-cell-receptor specificity is achieved remains unknown[1,3,4], as does whether signalling by the T-cell receptors, co-receptors and/or general cell-fate regulators such as Notch-1 (refs 5, 6) contributes to initial lineage choice, to subsequent differentiation processes or to both. Here we show that the CD4 versus CD8 lineage fate of immature thymocytes is controlled by the co-receptor-influenced duration of initial T-cell receptor-dependent signalling. Notch-1 does not appear to be essential for this fate determination, but it is selectively required for CD8$^+$ T-cell maturation after commitment directed by T-cell receptors. This indicates that the signals constraining CD4 versus CD8 lineage decisions are distinct from those that support subsequent differentiation events such as silencing of co-receptor loci.

The AND T-cell receptor (TCR) is specific for a pigeon cytochrome *c* peptide bound to the MHC class II molecule I-E^k (ref. 7). Transgenic thymocytes expressing this TCR are efficiently selected into the CD4$^+$ lineage in mice expressing wild-type I-A^b MHC class II molecules. In contrast, AND TCR transgenic mice expressing mutant I-A^b molecules that are defective in interaction with CD4 but not the TCR[8] generate CD8$^+$ but not CD4$^+$ mature cells[9]. To investigate how altering CD4 co-receptor binding controls the lineage fate of AND thymocytes, we took advantage of a modified two-stage reaggregate culture system[10] that allows controlled delivery of MHC-dependent and independent signals to thymocytes at distinct stages of maturation (Yasutomo *et al.*, manuscript in preparation). Immature CD69^{lo}CD4$^+$CD8$^+$ TCR transgenic thymocytes (double positive; DP) are incubated in dispersed culture with cells expressing the desired MHC molecule ligands, to initiate selection events (culture 1). This first TCR stimulation does not lead to silencing of the CD4 or CD8 locus in dispersed culture for up to three days, but it does rapidly induce upregulation of the activation marker CD69 that characterizes thymocytes undergoing maturation *in vivo*[1,11]. When the CD69$^{med/hi}$ thymocytes arising in culture 1 after 20 h are purified and reaggregated with either MHC-positive thymic stromal cells (TSC) or MHC-negative TSC in the presence of dendritic cells of the appropriate MHC type (culture 2), mature functional T cells expressing only a single co-receptor develop over the next 60 h. Experiments using this model showed that TCR–MHC molecule interactions are required not only in culture 1 to generate the CD69^{hi} thymocytes, but also in culture 2 to generate mature T cells. Using this approach, we asked whether the first, the second or both sets of TCR/co-receptor–MHC interaction events determined the differentiation of AND TCR transgenic T cells along the CD4 versus CD8 pathways.

Thymocytes from mice expressing only the AND TCR in the absence of MHC class II molecules (AND TCR transgenic RAG-2$^{-/-}$ MHC A$\beta^{-/-}$, referred to as AND throughout) were stimulated in culture 1 using dendritic cells from wild-type mice (WT-DC) or mutant (Mu-DC) I-A^b transgenic mice. When purified cells of the CD69$^+$CD4$^+$CD8$^+$ phenotype (CD69$^+$DP) generated by this stimulation were reaggregated with TSC from MHC class II$^{-/-}$ mice, no CD4$^+$ or CD8$^+$ mature T cells developed (Fig. 1a, b). Inclusion of

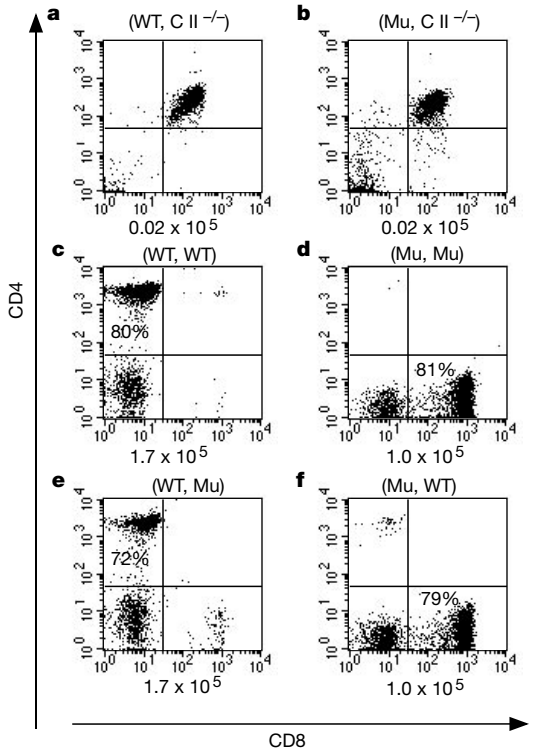

Figure 1 Role of co-receptor in CD4/CD8 lineage choice. Sorted CD69$^+$CD4$^+$CD8$^+$ cells obtained after stimulation by wild-type dendritic cells (WT-DC; **a**, **c**, **e**) or mutant dentritic cells (Mu-DC; **b**, **d**, **f**) were cultured with MHC class II$^{-/-}$ TSC (**a**, **b**) or MHC$^{-/-}$ TSC and wild-type dendritic cells (WT; **c**, **f**), or mutant dentritic cells (Mu; **d**, **e**) for 72 h in thymic reaggregates. The recovered cells from all cultures were stained with anti-CD4 or CD8 monoclonal antibodies and examined by flow cytometry. The percentage of cells with a CD4$^+$ or CD8$^+$ phenotype is given in the upper left or lower right quadrant of each panel and the absolute number of recovered viable thymocytes is given below each panel.