spans of thousands of years, Le Mouël has demonstrated that the non-dipole field cannot disappear during the process of a geomagnetic reversal. On the other hand, Bloxham and Gubbins[33] have found evidence for flux diffusion in the South Atlantic region and have tentatively concluded that the frozen-flux hypothesis can be rejected with 95% confidence. However, as pointed out by the authors themselves, this conclusion depends critically on error estimates for the field at the core-mantle boundary which are difficult to assess, so that at present it cannot be considered as definitely established.

In this respect the observed continuity of the secular variation during the KS 06 reversal is to our best knowledge the first experimental evidence supporting Le Mouël's theoretical results. Moreover, the fact that the processes leading to the observed secular variation would persist unchanged during a reversal suggests that only the internal part of the core would be involved in the mechanism of reversal.

We thank Dr C. Langereis and Dr H. Zijderveld for many helpful discussions and comments; C. Jehanno for supplying the microprobe analyses; L. De Seze for help in the mathematical treatment by cubic splines; and C. Kissel and S. Sen for help in sampling. The director of the Institute of Geology and Mining Research (IGME) at Athens kindly provided the necessary permits. This work was partly supported by the French CNRS ATP Géologie et Géophysique des Océans and CNRS-INSU ATP-Noyau, This is CFR contribution no. 761.

Received 3 February; accepted 7 May 1986.

1. Hillhouse, J. W. & Cox, A. *Earth planet. Sci. Lett.* **29**, 51-64 (1976).
2. Williams, I. & Fuller, M. *J. geophys. Res.* **87**, 9408-9418 (1982).
3. Prévot, M., Mankinen, E. A., Grommé, C. S. & Coe, R. S. *Nature* **316**, 230-234 (1985).
4. Theyer, F., Herrero-Bervera, E. & Hsu, U. *J. geophys. Res.* **90**, 1963-1982 (1985).
5. Hoffman, K. A. *Earth. planet. Sci. Lett.* **44**, 7-17 (1979).
6. Kaiser, A. D. & Verosub, K. L. *Geophys. Res. Lett.* **12**, 777-780 (1985).
7. Clement, B. M. & Kent, D. V. *J. geophys. Res.* **89**, 1049-1055 (1984).
8. Prévot, M., Mankinen, E. A., Coe, R. S. & Grommé, C. S. *J. geophys. Res.* **90**, 10417-10448 (1985).
9. Liddicoat, J. C. *Phil. Trans. R. Soc.* A**306**, 121-128 (1982).
10. Valet, J.-P., Laj, C. & Langereis, C. G. *Nature* **304**, 330-332 (1983).
11. Valet, J.-P. & Laj, C. *Nature* **311**, 552-555 (1984).
12. Langereis, C. G., Zachariasse, W. J. & Zijderveld, J. D. A. *Mar. Micropaleont.* **8**, 261-282 (1983-84).
13. Harstra, R. L. *Geophys. J.R. astr. Soc.* **71**, 477-495 (1982).
14. Dankers, P. H. M. thesis, State Univ. Utrecht (1978).
15. Laj, C., Jamet, M., Sorel, D. & Valente, J.-P. *Tectonophysics* **86**, 45-67 (1982).
16. Langereis, C. G. thesis, State Univ. Utrecht (1984).
17. Jacobs, J. A. *Reversals of the Earth's Magnetic Field* (Hilger, Bristol, 1984).
18. Hoffman, K. A. *Rev. Geophys. Space Phys.* **21**, 614-620 (1983).
19. Hoffman, K. A. *J. geophys. Res.* **89**, 6285-6292 (1984).
20. Tucker, P. *Geophys. J.R. astr. Soc.* **63**, 149-163 (1980).
21. Tucker, P. in *Geomagnetism of Baked Clays and Recent Sediments* (eds Creer, K. M., Tucholka, P. & Barton, C. E.) 3 (Elsevier, Amsterdam, 1983).
22. Salloway, J. C. thesis, Univ. Edinburgh (1983).
23. Merrill, R. T. & McElhinny, M. W. (eds) *The Earth's Magnetic Field* (Academic, London, 1983).
24. Creer, K. M. & Tucholka, P. *Phil. Trans. R. Soc.* A**306**, 87-102 (1982).
25. Valet, J.-P. & Laj, C. *Earth planet. Sci. Lett.* **54**, 53-63 (1981).
26. Valet, J.-P. thesis, Univ. Paris XI (1985).
27. Herrero-Bervera, E. & Helsley, C. E. *J. geophys. Res.* **88**, 3506-3522 (1983).
28. Valet, J.-P., Laj, C. & Tucholka, P. *Nature* **316**, 217-218 (1985).
29. Grommé, C. S., Mankinen, E. A., Prévot, M. & Coe, R. S. *Nature* **318**, 487 (1985).
30. Valet, J.-P., Laj, C. & Tucholka, P. *Nature* **318**, 487-488 (1985).
31. Hoffman, K. A. *Nature* **294**, 67-68 (1981).
32. Le Mouël, J.-L. *Nature* **311**, 734-735 (1984).
33. Bloxham, J. & Gubbins, D. *Geophys. J.R. astr. Soc.* **84**, 139-152 (1986).

Cloning the gene for an inherited human disorder—chronic granulomatous disease—on the basis of its chromosomal location

Brigitte Royer-Pokora*, Louis M. Kunkel†, Anthony P. Monaco†, Sabra C. Goff*, Peter E. Newburger‡, Robert L. Baehner*, F. Sessions Cole§, John T. Curnutte‖ & Stuart H. Orkin*¶#

* Division of Hematology-Oncology, The Children's Hospital, Dana-Farber Cancer Institute, Department of Pediatrics, † Division of Genetics, The Children's Hospital, Department of Pediatrics and the Program in Neuroscience, § Division of Cell Biology, The Children's Hospital, Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA
‡ Division of Pediatric Hematology, Department of Pediatrics, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA
‖ Division of Pediatric Hematology, Department of Pediatrics, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA
¶ Howard Hughes Medical Institute, Children's Hospital, Boston, Massachusetts 02115, USA

The gene that is abnormal in the X-linked form of the phagocytic disorder chronic granulomatous disease has been cloned without reference to a specific protein by relying on its chromosomal map position. The transcript of the gene is expressed in the phagocytic lineage of haematopoietic cells and is absent or structurally abnormal in four patients with the disorder. The nucleotide sequence of complementary DNA clones predicts a polypeptide of at least 468 amino acids with no homology to proteins described previously.

MANY inherited disorders in man result from mutations in genes whose protein products are as yet unknown. For some disorders specific proteins may be suspected as candidates for the primary gene products, whereas in other situations no biochemical clues exist. Molecular cloning offers an approach to the analysis of disease gene loci which, in principle, may be executed without specific information regarding the nature of the proteins involved. The use of genetic linkage to establish the location of a disease gene within the chromosome complement has generally been considered an initial step in this analysis[1]. Major diseases of unknown aetiology such as Duchenne muscular dystrophy (DMD), Huntington's disease and cystic fibrosis, have been the focus of intensive study from this perspective[2-7]. It has been anticipated that extended genomic cloning[8], coupled with searches for transcribed regions, will provide a means of defining each disease locus in precise genetic terms. Here we describe the successful application of this general strategy to the identification and characterization of the gene involved in chronic granulomatous disease (CGD), an X-chromosome-linked disorder of phagocytic cells. Affected individuals have greatly impaired host defences against infection with a wide variety of microorganisms[9]. In contrast to the normal situation, phagocytes

To whom correspondence should be addressed.

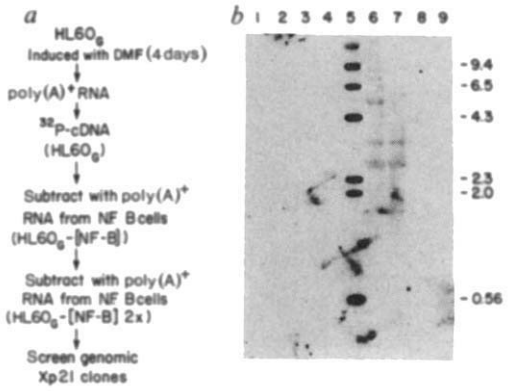


Fig. 1 Detection of a transcribed region of Xp21. *a*, Strategy for the preparation of an enriched cDNA probe. Granulocytic HL60 cells ($HL60_G$) were used as a source of mRNA and total ^{32}P-cDNA. Sequences in common between $HL60_G$ and the B-cell line of patient N.F. were removed by the method of Davis[32]. *b*, Southern blot hybridization[33] of the enriched cDNA probe with *Eco*RI + *Hin*dIII-digested DNAs of bacteriophage derived from Xp21. Bacteriophage clone designations: lane 1, 469-7A2; 2, 378-40A1; 3, 378-29A1; 4, 378-19A1; 6, 379A6; 7, 379π; 8, 55π; 9, 145-A1. These bacteriophage correspond to clones hybridizing with the pERT clones 469, 378, 379, 55 and 145 (ref. 15). Lane 5, *Hin*dIII-digested λ DNA as a marker. Hybridization to fragments of 2.5 and 3.3 kb was observed for the samples in lanes 6 and 7. These clones overlapped the same genomic region. Bands of greater size, visible in lane 6, are due to slightly incomplete digestion of phage 379-A6 DNA and not additional hybridizing fragments, as verified by subsequent direct analysis of the phage with cloned cDNA (not shown).

Methods. *a*, Total poly(A)$^+$ RNA was prepared from $HL60_G$ cells as described elsewhere[20]. ^{32}P-labelled cDNA was synthesized using random primers and AMV (avian myeloblastosis virus) reverse transcriptase[32] (specific activity of 1×10^8 c.p.m. μg^{-1}). cDNA was depleted of template RNA by base hydrolysis and hybridized in 0.5 M sodium phosphate buffer, 0.1% SDS, 5 mM EDTA in sealed glass capillaries at 69 °C to a R_0t of 2,700 with a 20-fold excess of poly(A)$^+$ RNA prepared from cultured B cells of patient N.F. Single-stranded cDNA was isolated in 0.12 M sodium phosphate on hydroxyapatite at 60 °C and rehybridized to N.F. poly(A)$^+$ RNA as described above[32]. *b*, Single-stranded cDNA (1×10^6 c.p.m.) from the second hydroxyapatite step was hybridized with the digested phage DNAs on a GeneScreen filter in a 1.0-ml solution containing 1 M NaCl, 2 × Denhardt's solution, 10% dextran sulphate, 1% SDS, 0.2 M sodium phosphate, 50 $\mu g\ ml^{-1}$ salmon sperm DNA, 50 $\mu g\ ml^{-1}$ oligo$(dA)_{18}$ for 60 h at 68 °C. The filter was washed in 1 × SSC, 1% SDS, and 0.5 × SSC, 1% SDS at the same temperature. Autoradiographic exposure was for 10 days with an intensifying screen.

(granulocytes, monocytes and eosinophils) isolated from the blood of CGD patients fail to generate superoxide and activated oxygen derivatives on ingestion of microbes. The precise lesion in the membrane-associated NADPH-oxidase system of these cells is unknown. Cell complementation[10] and family studies[9] have defined X-linked and autosomal recessive varieties of the disorder. In most families the disease is transmitted as an X-linked trait. Over the past two decades several proteins, including an unusual *b*-type cytochrome[11] and a flavoprotein[12], have been considered as potential candidates for the primary products of CGD loci, but difficulties in solubilizing membrane-associated components have impeded biochemical characterization of the oxidase system. A consistent finding in the X-linked variety, however, is the absence of the haem spectrum derived from cytochrome *b*[11]. Given the potential complexity of the oxidase system and the possible requirement for assembly of several components into an enzymatically active unit, this recognized deficiency in the cytochrome *b* spectrum may be secondary to the primary genetic abnormality.

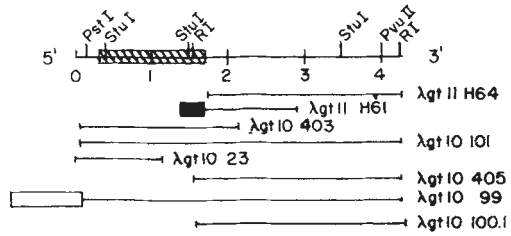


Fig. 2 cDNA clones spanning the 379-encoded RNA. Granulocytic HL60 cDNA libraries were constructed in λgt11 and λgt10 using procedures described elsewhere[18–20]. Initial screening with a subclone from the 379-A6 bacteriophage yielded λgt11 H64 and H61. The latter contained a segment absent from other overlapping clones (indicated by the solid box) and thought to represent intron sequences. Repeated screenings with additional cDNA subfragments of the λgt10 library yielded cDNAs spanning the entire mRNA, representative examples of which are shown. The open box at the 5′ end of clone 99 contained sequences unrelated to the 379 transcript. The cross-hatched box represents the major open reading frame of the mRNA, as defined by DNA sequence analysis (Fig. 6*a*).

In view of the problems inherent in conventional biochemical analysis, we have adopted a genetic approach to define the molecular basis of the X-linked form of CGD as a means of delineating at least one critical protein of the oxidase system.

General strategy

Four discrete steps are involved in our analysis: (1) The position of the CGD gene on the X chromosome (X-CGD) was mapped by both deletion and formal linkage analysis, described recently[13]. (2) Messenger RNA transcripts derived from the assigned chromosomal region were sought. (3) The relevance of one specific mRNA transcript to the disorder was investigated by studying affected patients. (4) The protein product encoded by the X-CGD gene was predicted from isolated complementary DNA clones. Taken together, the results provide persuasive genetic evidence that we have identified the X-CGD gene and characterized its mRNA transcript.

Expressed Xp21 sequences

From the study of two patients (B.B. and N.F.) with interstitial deletions of Xp21 and from linkage of X-CGD with two Xp21 DNA markers (p745 and pERT 84), the CGD locus has been assigned to Xp21.1 (refs 13–15, 42). This delimits the X-CGD locus to ~0.1% of the genome, or perhaps 3,000 kilobases (kb) of DNA. As part of chromosome walking experiments originally aimed at characterizing the DMD locus, a series of bacteriophage clones derived from seven regions of Xp21 that were absent from the DNA of patient B.B. were obtained[3,15], using pERT clones 55, 84, 87, 145, 378, 379 and 469 as probes (L.M.K. and A.P.M., unpublished data). (These bacteriophage clones span at most 10% of the DNA deleted in patient B.B.) In this manner, approximately 250 kb of Xp21 DNA was isolated.

The experiment outlined in Fig. 1*a* was performed to search for regions of these bacteriophage clones that are transcribed in phagocytic cells. Briefly, a Southern blot filter bearing restriction enzyme-digested bacteriophage DNAs was hybridized with a radioactive cDNA probe enriched for sequences expressed in phagocytic cells. As a source likely to contain transcripts of the CGD gene, we chose cultured human HL60 leukaemic cells[16] treated with dimethylformamide (DMF) for 4–7 days. Such treatment induces the NADPH-oxidase system together with most other components of granulocytic differentiation[17]. To remove constitutive transcripts, but not sequences derived from Xp21, HL60-cell-induced cDNA was enriched for Xp21 transcripts by subtractive hybridization twice with RNA from an Epstein–Barr virus (EBV)-transformed B-cell line of the

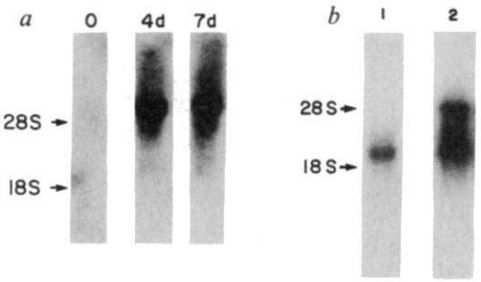


Fig. 3 Expression of 379 RNA in HL60 cells and monocytes. *a*, Total cell RNA (10 μg) isolated from uninduced (0) or dimethyl sulphoxide (DMSO)-induced (either 4 days or 7 days) human leukaemic HL60 cells was analysed with a fragment of 379 cDNA derived from the 3′-untranslated region. The uninduced cells are undifferentiated HL60 cells prepared by centrifugal elutriation[17]. *b*, Poly(A)$^+$ mRNA (~1 μg) from human hepatoma cells (HepG2; lane 1) or from cultured human monocytes (lane 2) was similarly examined using the 379 probe plus phosphoglycerate kinase (PGK) cDNA. The latter detects a 2-kb mRNA species in all somatic cells[34].

Methods. RNA was prepared by guanidine-HCl precipitation and phenol extraction (*a*) or by centrifugation through $CsCl_2$ following solubilization with guanidine isothiocyanate[35] (*b*). Samples were electrophoresed in 1% agarose formaldehyde gels[36], transferred to nitrocellulose, and hybridized with a 1.8-kb *Bgl*II–*Bgl*II restriction enzyme fragment of H64 cDNA (see Fig. 5*a*) labelled by random priming[37]. 28S and 18S denote the migration of ribosomal RNAs. In *b*, the filter was simultaneously hybridized with labelled insert prepared from the cDNA clone PGK-7e[34].

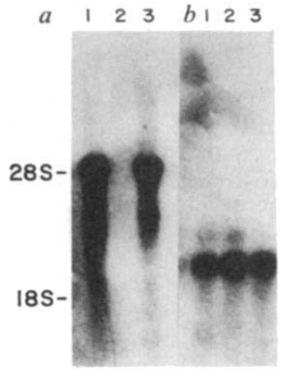


Fig. 4 Absence of 379 RNA from X-CGD monocytes. Total cell RNA (5 μg) prepared from the cultured monocytes of two X-CGD patients (lanes 1 and 2) and from a normal individual (lane 3) was examined for 379 sequences (*a*) or PGK sequences (*b*). In addition to the patient whose sample is shown in lane 2, two unrelated patients also had monocytes that were negative for 379 RNA. Peripheral blood monocytes were isolated and cultured as described previously[38]. Other methods were as given in Fig. 3 legend.

CGD/DMD patient N.F. (Fig. 1*a*). The subtracted radiolabelled cDNA, representative of ~500 individual mRNA sequences, was hybridized to a Southern blot of the Xp21 bacteriophage clones (Fig. 1*b*). Two overlapping clones, designated 379-A6 and 379-π, showed hybridization.

Isolation of cDNA clones

The results shown in Fig. 1*b* indicated that the bacteriophage clones originally isolated with pERT 379 contained sequences ultimately represented in the mRNA of induced HL60 cells. To locate the transcribed segment(s), non-repetitive DNA fragments of the 379-A6 insert were used as hybridization probes to Northern blots of the HL60 RNA used to generate the enriched cDNA. A probe consisting of three fragments (0.45, 0.5 and 1.8 kb) arising from a *Hin*dIII/*Bgl*II digest of a 5-kb *Hin*dIII fragment (designated 379-10) detected a 5-kb mRNA transcript (see below). cDNA clones corresponding to this mRNA species were isolated from 4-day-induced HL60 cDNA libraries constructed in bacteriophage λgt10 and λgt11 (refs 18–20) (Fig. 2). Repeated screening yielded cDNA clones spanning the entire mRNA transcript. Based on the frequency with which clones were obtained from the cDNA libraries, the abundance of the specific RNA in DMF-induced HL60 cells was ~0.02–0.05%. For purposes of discussion these are referred to as 379 cDNA and 379 RNA. Cloned cDNA was used to examine (1) the tissue distribution of the RNA, (2) its structure and expression in X-CGD, and (3) the nature of its predicted protein sequence.

Tissue distribution

The 379 cDNA detects a 5-kb RNA that is markedly induced in HL60 cells during granulocytic differentiation stimulated by DMF treatment (Fig. 3*a*). This RNA was also induced on treatment with either the tumour promoter TPA (12-*O*-tetradecanoyl phorbol-13-acetate) or vitamin D_3, agents that promote monocytic differentiation of HL60 cells in culture[21]. Moreover, the transcript was abundant in cultured normal human monocyte RNA (Fig. 3*b*) and absent from the hepatoma line HepG2 (Fig. 3*b*) as well as from total kidney, fibroblasts and HeLa cells (not shown). A low level of the transcript was detected in normal EBV-transformed B-cell lines (not shown), which have been reported to exhibit low levels of oxidase activity[22].

As the RNA transcript encoded by 379 sequences appeared to be a marker for the phagocytic lineage (granulocytes and monocytes) based on our limited survey of different cell types, it represented a suitable candidate for the product of the X-CGD locus.

Relevance of 379 transcript

Given that the X-CGD locus was mapped to ~0.1% of the genome, as many as 20 transcripts derived from this region might exist in an average cell (containing perhaps 10,000–20,000 mRNA species). This estimate does not take into account possible variation in the fraction of DNA transcribed in different regions of the genome. The finding of an Xp21 phagocyte-specific RNA cannot, by itself, be interpreted as strong evidence that it encodes the X-CGD product. To determine more specifically the potential relevance of the transcript to the disease, we analysed material from classical X-linked CGD patients by hybridization with the putative X-CGD cDNA. If the transcript were derived from the relevant disease gene locus, quantitative and/or qualitative RNA or DNA alterations would be anticipated in these genetic variants.

We first analysed RNA isolated from cultured human monocytes of both normal and CGD origin. Only CGD patients with the X-linked variety were examined. Their clinical and family histories were consistent with this classification, as were negative phagocyte NBT tests[9] and the absence of the cytochrome *b* spectrum. Figure 4 shows the monocyte RNA analysis of two such patients: in one patient (lane 2) no 379-specific RNA was detectable; in the other patient (lane 1) 379 RNA was apparently normal in size and abundance. Integrity of RNA samples was confirmed by hybridization with a constitutively expressed sequence (phosphoglycerate kinase) derived from a different X-chromosomal region (Fig. 4*b*). Monocyte RNA of two additional X-CGD patients examined also lacked 379-specific RNA (not shown). In the RNA-negative CGD patients, Southern blot analysis with cDNA probes spanning the entire gene revealed no DNA deletions or rearrangements (not shown). These genetic variants provide strong initial support for the relevance of the 379 transcript to the disease. Although unlikely, these findings might be explained by abnormal regulation of

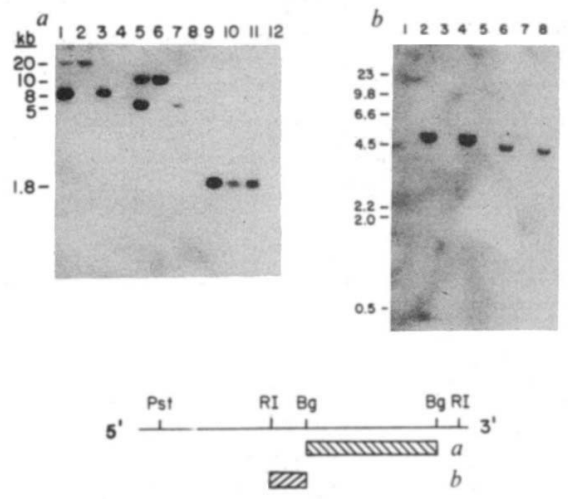


Fig. 5 Partial gene deletion in an X-CGD patient. Total cellular DNAs (5 μg per lane) were digested with *Bam*HI (*a*, lanes 1-4), *Hin*dIII (*a*, 5-8; *b*, 1-4) or *Bgl*II (*a*, 9-12; *b*, 5-8), electrophoresed, and hybridized with the regions of the 379 cDNA indicated at the bottom of the figure. *a*, DNAs from mother of patient J.W. (lanes 1, 5, 9), X-CGD patient J.W. (lanes 2, 6, 10) and Xp21 deletion patient N.F.[13] affected with CGD and DMD (lanes 4, 8, 12). Lanes 3, 7, 11, normal DNA. *b*, DNAs from patient J.W. (lanes 1, 5), mother of J.W. (lanes 2, 6) and patient N.F. (lanes 3, 7). Lanes 4, 8, normal DNA.

the 379 transcript due to a defect in a *trans*-acting factor produced by the authentic X-CGD locus.

Conclusive evidence regarding the role of the 379 transcript, though, was provided by an analysis of patient J.W. whose monocyte RNA appeared grossly normal in abundance and structure (Fig. 4*a*, lane 1). Analysis of DNA from patient J.W. with the 3′ half of the cDNA revealed a DNA alteration on digestion with some (for example, *Hin*dIII and *Bam*HI), but not all (for example, *Bgl*II), restriction enzymes (Fig. 5*a*, lanes 2, 6, 10). The patient's mother was heterozygous for this alteration (lanes 1, 5), consistent with her assignment as an obligate carrier. Various subregions of the full-length cDNA were used as probes to define the nature of the DNA alteration in J.W. Use of a 0.3-kb segment encompassing the middle of the cDNA revealed an interstitial deletion in the DNA (Fig. 5*b*, lanes 1, 5) and monocyte RNA (not shown). Most of the region between the central *Eco*RI and *Bgl*II restriction sites of the cDNA was absent. Sequences 3′ to the deleted segment were present in both the DNA and monocyte RNA (not shown). Although the precise boundaries of the interstitial deletion have yet to be defined, it overlaps the 3′ terminus of the large open reading frame (ORF) of the cDNA (see Fig. 2 and below). The failure of cDNA probes to hybridize with the DNA of the CGD/DMD patient N.F. established that the transcribed sequences were derived from Xp21 (Fig. 5*a*, lanes 4, 8, 12).

Thus, the 379-encoded transcript was abnormal in abundance or structure in the monocytes of all four classical X-linked CGD patients examined. These observations provide conclusive evidence that the transcript defines the product of the X-CGD locus. Therefore, we refer to it henceforth as the X-CGD RNA.

Analysis of genomic DNAs of 17 additional, independent X-CGD patients with the 3′ half of the cDNA revealed no detectable DNA deletions or alterations (not shown).

Sequence of X-CGD cDNA

cDNA clones spanning the 379-encoded RNA are shown in Fig. 2. The transcriptional orientation of the mRNA was established by hybridization of single-stranded SP6 RNA transcripts[23] to Northern blots (not shown). DNA sequences of cDNA clones were obtained from both strands either from restriction enzyme fragments subcloned into M13 vectors[24] or by progressive deletions as described by Dale *et al.*[25]; the assembled sequence is shown in Fig. 6*a*.

Although the estimated size of the RNA transcript from Northern blot analysis is nearly 5 kb, two findings indicate that the 4.27 kb of cDNA sequence that we have obtained encompasses the entire mRNA. First, the 3′ terminus is defined by clone 100.1 which contained a short oligo(dA) tract. Second, primer extension analysis revealed an extended product corresponding to the most 5′ extent of clone 23 (Fig. 6*b*). Although we cannot exclude the presence of a strong stop for reverse transcriptase, the failure to observe cDNAs with additional 5′ sequences probably indicates that the true 5′ terminus of the mRNA has been recognized. Genomic studies that are in progress should assign the precise transcription initiation site(s).

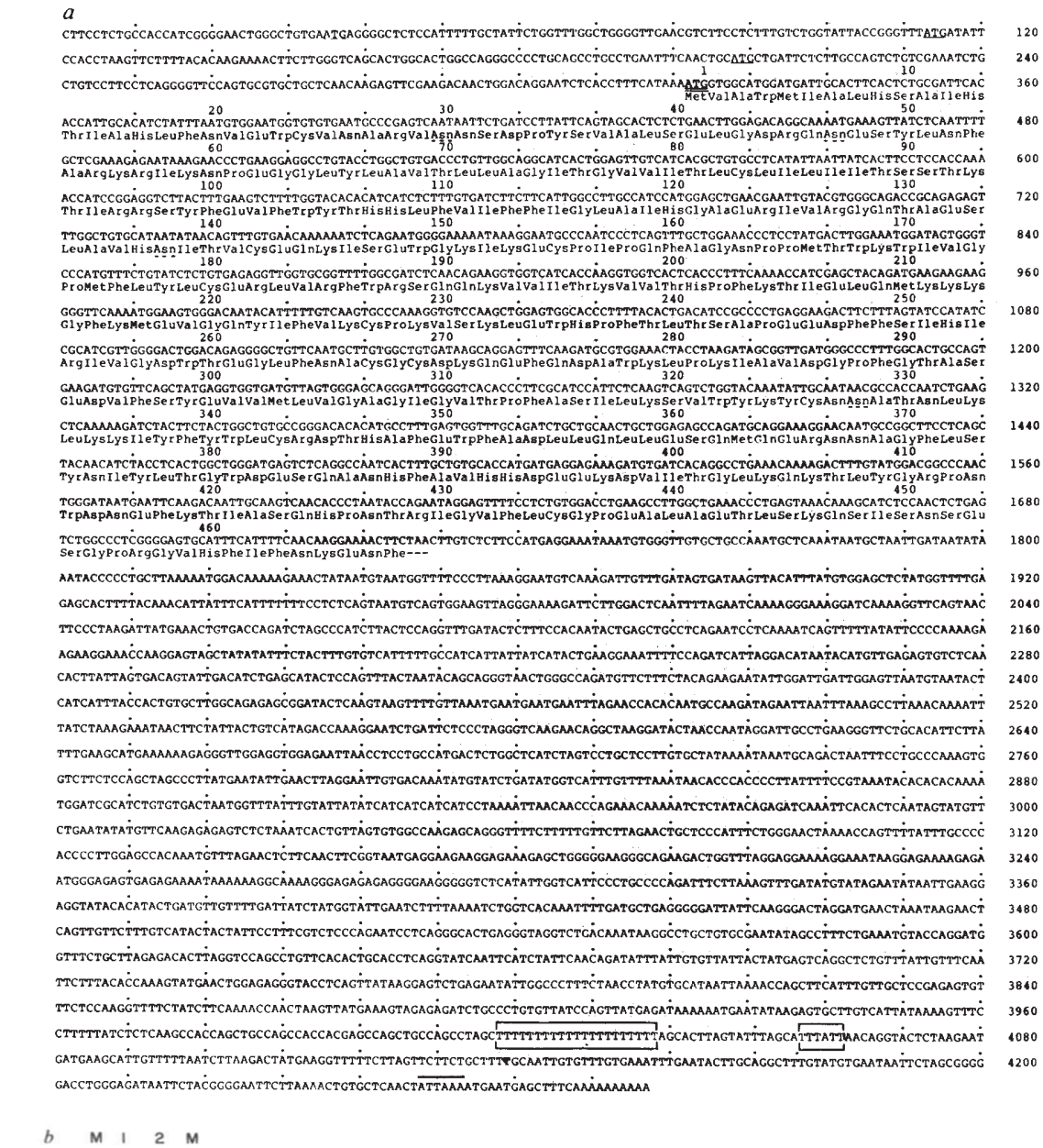
A single large open reading frame, extending from potential initiator ATGs at nucleotide 208 or 322 to a termination codon (TAA) at 1,726-1,728, is present. Additional ATGs, each followed closely by an in-frame termination codon, are located at positions 35 and 113. As the ATG at position 208 is in the same translational frame as that located at position 322, we cannot unambiguously assign the initiation codon. However, given that the ATG at 322 closely resembles the consensus sequence for functional initiator codons defined by Kozak[26], whereas that at 208 does not, it is highly likely that the former would be used for translation initiation *in vivo*. Whether the upstream ATGs have a role in regulating production of the X-CGD-encoded protein is unknown.

The 3′-untranslated region of the mRNA is 2.5 kb long. Although we have not analysed in detail the organization of the CGD gene in genomic DNA, comparison of the 379 bacteriophage subclones and cDNA clones indicated that the entire 3′-untranslated region is represented in a signal exon, which also contains a small coding segment 3′ to the central *Eco*RI site of the cDNA. A putative polyadenylation signal (ATTAAA) occurs 14 nucleotides before the poly(A) tract of clone 100.1. Several more typical potential polyadenylation signals (AATAAA)[27] are found elsewhere in the cDNA sequence but do not appear to be utilized in mRNA processing. A notable feature of the 3′-untranslated region is the presence of T_{21} (nucleotides 4,018-4,028), followed 19 nucleotides in the 3′ direction by TTTATT (nucleotides 4,057-4,062). On the complemetary strand this organization [AATAAA-19 nucleotides-A_{21}] resembles the 3′ end of a processed transcript[28,29]. It is possible that a processed sequence integrated at the 3′ end of the CGD gene. As the genomic DNA encoding this region has the identical sequence (not shown), this novel organization in the transcript is not the result of a cloning artefact.

The open reading frame extending from position 322 to 1,726 predicts a primary translation product of 468 amino acids (Fig. 6) with an estimated relative molecular mass (M_r) of 54,000 (54K). If the atypical initiator ATG at position 208 were used, a polypeptide with an additional 38 amino acids at the N-terminus would result. Four potential *N*-linked glycosylation sites of the canonical form Asn-X-(Thr/Ser) are present. The calculated *p*I of the predicted protein is 9.5. Figure 6*c* shows the distribution of charged residues and a hydropathicity profile of the predicted protein. One extensive hydrophobic segment is evident; this spans amino acids 65-92. Screening of the GENBANK and protein database revealed no significant homology of the cDNA or of its predicted protein to known sequences.

Conclusions

Based on genetic linkage data, chromosome walking within a defined region, and hybridization with an enriched cDNA probe, we have identified and characterized the genetic locus for an



inherited human disorder (chronic granulomatous disease) without reference to a specific protein product. The derivation of the 379 RNA transcript from Xp21, its high-level expression in the phagocytic lineage, and its derangement in patients with classical X-linked CGD provide the elements of the formal argument that the appropriate locus has been recognized.

Several aspects which contributed to the success of the transcript search are relevant to its potential application to other disorders. (1) Although only ~10% of Xp21 was represented in bacteriophage clones isolated with pERT clones[15], a portion of the X-CGD gene (notably its 3′-untranslated region) was fortuitously present in the collection. (2) The 3′-untranslated region of 2.5 kb provided a large, uninterrupted exon for hybridization with the enriched cDNA probe. Although almost any transcribed segments would, in principle, be detectable by hybridization of subtracted cDNA with cloned genomic DNAs, the strength of the observed signal will be greater for larger exons. (3) Previous knowledge of cells expressing the phagocytic oxidase system suggested a convenient source for the isolation of RNA used in preparation of the enriched probe. Specifically, the abundance of the 379-encoded RNA transcript in DMF-treated HL60 cells is appreciable (on the order of 0.02–0.05%). (4) The availability of an EBV-transformed cell line from patient N.F. with a deletion through the entire Xp21 region permitted removal of most constitutive transcripts from the enriched cDNA without depletion of Xp21 sequences. The extent to which our strategy may be applied to the isolation of transcripts for other disorders for which a chromosomal map position is established may well depend on these factors.

Although the oxidase system of phagocytes has been studied by conventional biochemical analyses for two decades, no clear view of the components involved or the specific protein abnormalities manifest in the major functional deficiency seen in CGD has emerged. Recently,, the findings of Segal and colleagues have implicated an unusual *b*-type cytochrome in the oxidase system[11,30,31]. As such, it represented a plausible candidate for the protein mutated in the X-linked form of CGD[11,30,31]. Our results, however, argue against this conclusion. First, the amino-acid composition of the protein predicted by our sequence does not resemble that reported for purified cytochrome *b*[30]. Second, the predicted protein encoded by the 379 RNA does not show significant homology to previously sequenced cytochromes of several origins. No evidence for a haem-binding region was evident from analysis of the predicted protein. Third, although its spectrum is absent in CGD granulocytes, the proposed cytochrome *b* protein has been reported to be present in extracts of neutrophils from affected individuals[31]. This finding is not in accord with the apparently frequent observation of RNA-negative X-CGD patients. Whether the predicted protein is related to the flavoprotein found in phagocytes[12] must await further biochemical characterization of the latter.

We propose that the 379-encoded protein, which we designate the X-CGD protein, is an essential component of the oxidase system of the phagocyte and probably interacts with other proteins to form an active complex. Among those components that may interact with the X-CGD protein are the membrane-associated cytochrome *b* and a flavoprotein. Spectral abnormalities of cytochrome *b* in CGD phagocytes are probably secondary to the absence (or abnormal structure) of the predicted X-CGD protein.

Characterization of the X-CGD protein *in vivo* and examination of proteins with which it associates should help to elucidate the organization and function of the oxidase system of the phagocyte. Analysis of the predicted sequence of the X-CGD protein has provided few clues, as yet, to the nature of the protein itself. The protein has an estimated M_r of a minimum 54K, and is probably basic. Modification at the four potential *N*-linked glycosylation sites may alter the apparent size of the observed protein *in vivo*. One marked hydrophobic region noted above (amino acids 65–92) may represent a transmembrane segment. Finally, the interstitial deletion found in patient J.W. may serve to locate an important functional domain at the carboxy-terminus of the protein, as the deletion is predicted to remove the last 50 amino acids (distal to the *Eco*RI site at nucleotide 1,571). Further characterization of the X-CGD protein and delineation of its functional domains will ultimately require the introduction and expression of the cDNA in phagocytic cells of X-CGD patients.

We thank Dr Temple Smith of the Dana-Farber Cancer Institute for assistance and advice in computer analysis of the sequence and Dr Mark Davis for advice on subtractive hybridization. S.H.O. is an Investigator of the Howard Hughes Medical Institute. P.E.N. was supported by NIH grant CA38325, by the Socha Memorial Fund for CGD research, and by an Established Investigatorship of the American Heart Association with funds contributed in part by its Massachusetts affiliate. This work was also supported in part by grants from the Muscular Dystrophy Association to L.M.K. and from the NIH to J.T.C. (AI-21320) and S.H.O. (HD18661).

Received 17 May; accepted 21 May 1986.

1. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. *Am. J. hum. Genet.* **32,** 314–331 (1980).
2. Davies, K. E. *et al. Nucleic Acids Res.* **11,** 2303–2312 (1983).
3. Monaco, A. P. *et al. Nature* **316,** 842–845 (1985).
4. Knowlton, R. G. *et al. Nature* **318,** 380–382 (1985).
5. White, R. *et al. Nature* **318,** 382–384 (1985).
6. Wainwright, B. J. *et al. Nature* **318,** 384–385 (1985).
7. Gusella, J. F. *et al. Nature* **306,** 234–238 (1983).
8. Collins, F. & Weissman, S. *Proc. natn. Acad. Sci. U.S.A.* **81,** 6812–6818 (1984).
9. Tauber, A. I., Borregaard, N., Simons, E. & Wright, J. *Medicine* **62,** 286–309 (1983).
10. Weening, R. S. *et al. J. clin. Invest.* **75,** 915–920 (1985).
11. Segal, A. W. *et al. New Engl. J. Med.* **308,** 245–251 (1983).
12. Markert, M., Glass, G. A. & Babior, B. M. *Proc. natn. Acad. Sci. U.S.A.* **82,** 3144–3148 (1985).
13. Baehner, R. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **83,** 3398–3401 (1986).
14. Francke, U. *et al. Am. J. hum. Genet.* **37,** 250–267 (1985).
15. Kunkel, L. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82,** 4778–4782 (1985).
16. Collins, S. J., Ruscetti, F. W., Gallagher, R. E. & Gallo, R. C. *Proc. natn. Acad. Sci. U.S.A.* **75,** 2458–2462 (1978).
17. Newburger, P. E. *et al. J. biol. Chem.* **259,** 3771–3776 (1984).
18. Young, R. A. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **80,** 1194–1198 (1983).

Fig. 6 (Left) DNA sequence of the X-CGD cDNA (*a*), primer extension analysis (*b*) and charge distribution and hydropathicity profile of predicted X-CGD protein (*c*). ***a***, Bacteriophage cDNA inserts were cloned in M13 derivatives[24] for sequence analysis by the methods of Sanger *et al.*[39] and Dale *et al.*[25]. All regions were sequenced on both DNA strands. Clones from which sequences were derived are shown in Fig. 2. H64 and H61 were sequenced in their entirety. The initial 0.1-kb displayed was assembled from clones 23, 101, 403 and 99. The 2 kb spanning the open reading frame was derived from clone 99. Overlaps with the 5′ regions of H64, H61 and 405 were examined. Based on this, and restriction mapping of genomic DNA, the intron sequences at the 5′ end of H61 were identified. The extreme 3′-terminal sequences were obtained from clone 100.1. The first four ATGs in the sequence are underlined, with the predicted initiator codon double-underlined. The presumed processed cDNA sequence near the 3′ end of the transcript is boxed. Canonical *N*-glycosylation sites are underscored with broken lines, and the presumptive poly(A) addition signal (ATTAAA) is overlined. ***b***, The position of the 5′ terminus of the X-CGD mRNA was assigned by primer extension analysis. A 70-nucleotide (nt) end-labelled primer was hybridized with yeast transfer RNA (10 μg, lane 1) or total monocyte (20 μg, lane 2) and extended with reverse transcriptase. Lanes labelled M, marker DNA. The observed product of 250 nt corresponds precisely to the 5′ extent of clone 23 (see Fig. 2). Use of a synthetic 30-mer primer (nucleotides 80–110) confirmed the position of the extended product and suggested, in addition, the existence of three possible mRNA start sites within about 10 nucleotides (not shown). ***c***, Distribution of charged residues and the hydropathicity profile of the predicted X-CGD protein. +, H and − refer to cationic (Arg, His, Lys), hydrophobic (Ala, Ile, Met, Trp, Val, Phe, Pro, Leu) and anionic (Glu, Ser, Tyr, Asp, Thr) amino acids, respectively. In the hydrophobicity profile, shown at the bottom, upward deflections indicate hydrophobic regions. **Methods.** *b*, The indicated 70-nt *Mst*II–*Pst*I restriction fragment was 5′-end-labelled with [γ-^{32}P]ATP using polynucleotide kinase[40], hybridized with RNA, and extended with AMV reverse transcriptase using methods described previously[41]. Reaction products were phenol-extracted, ethanol-precipitated and separated in an 8% urea/acrylamide gel. Marker DNAs were pBR322 digested with *Hin*fI. Autoradiography was performed overnight with an intensifying screen.

19. Gubler, U. & Hoffman, B. J. *Gene* **25**, 263-279 (1983).
20. Ginsburg, D. *et al. Science* **228**, 1401-1406 (1985).
21. Rovera, G., Santoli, D. & Damsky, C. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2779-2783 (1979).
22. Volkman, D. J., Buescher, E. S., Gallin, J. I. & Fauci, A. S. *J. Immun.* **133**, 3006-3009 (1984).
23. Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035-7056 (1984).
24. Messing, J. *Meth. Enzym.* **101**, 20-78 (1983).
25. Dale, R. M. K., McClure, B. A. & Houchins, J. P. *Plasmid* **13**, 31-40 (1985).
26. Kozak, M. *Nucleic Acids Res.* **12**, 857-872 (1984).
27. Proudfoot, N. J. & Brownless, G. G. *Nature* **263**, 211-214 (1976).
28. Nishioka, Y., Leder, A. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2806-2809 (1980).
29. Sharp, P. A. *Nature* **301**, 471-472 (1983).
30. Harper, A. M., Chaplin, M. F. & Segal, A. W. *Biochem. J.* **227**, 783-788 (1985).
31. Segal, A. W. *Lancet* i, 1378-1382 (1985).
32. Davis, M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2194-2198 (1984).
33. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
34. Michelson, A. M., Markham, A. F. & Orkin, S. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 472-476 (1983).
35. Chirwin, J. M., Przygyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5298 (1979).
36. Maniatis, T., Fritsch, E. & Sambrook, J. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
37. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).
38. Strunk, R. C., Whitehead, A. S. & Cole, F. S. *J. clin. Invest.* **76**, 985-990 (1985).
39. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
40. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-580 (1980).
41. Treisman, R., Proudfoot, N. J., Shander, M. & Maniatis, T. *Cell* **29**, 903-911 (1982).
42. Francke, U. *Cytogenet. Cell Genet.* **38**, 298-307 (1984).

LETTERS TO NATURE

A QSO with redshift 3.8 found on a UK Schmidt telescope IIIa-F prism plate

C. Hazard*†, R. G. McMahon*† & W. L. W. Sargent‡

* Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA
† Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
‡ Palomar Observatory, California Institute of Technology, Pasadena, California 91125, USA

High-redshift QSOs (quasi-stellar objects) are important because of the information they provide on the early history and evolution of the Universe; the brighter ones are particularly valuable as probes of the intervening material. During the past 20 years much effort has been devoted to trying to discover QSOs at redshifts $z > 3.5$, but progress has been so slow that it has been suggested that there is a cutoff in the QSO distribution at $z \simeq 3.5$ (ref. 1). However, we have already demonstrated[2-4] how unfiltered UK Schmidt telescope (UKST) IIIa-F low-dispersion objective prism plates can be used successfully in such searches up to at least $z \simeq 3.7$ (ref. 3). We report here the use of this technique to discover a QSO, 1208+1011, with a redshift $z = 3.80$. The highest redshift previously known was that of the radio-selected QSO, PKS2000−330 with $z = 3.78$ (ref. 5). In two UKST fields we have now discovered six QSOs with redshifts between 3.3 and 3.8, of which four have $z \geq 3.50$; including PKS2000−330, only four other QSOs with $z \geq 3.50$ are known over the whole sky[5-7,10]. Our success up to $z = 3.8$ indicates that redshifts >4 could soon be attained.

Up to $z \simeq 3.3$, searches for high-z QSOs by their Ly-α + N v emission are best carried out on IIIa-J prism plates, which have a red cutoff to their spectral response at ~5,200 Å. The IIIa-F plates, with a corresponding limit at 6,800 Å, extend the range above $z \simeq 3.3$ to a possible maximum of $z = 4.6$. However, to avoid confusion with cool stars our searches on these plates have been restricted to the interval $3.3 \leq z \leq 3.9$. With this restriction and using auxiliary IIIa-D and IIIa-J prism and direct J and R or I plates to check the spectral features and colours of the IIIa-F-selected objects we have shown how high-redshift QSOs can be selected with complete reliability[2,3]. For these prime selections we require that the reality of the emission line in the IIIa-F spectrum be confirmed by the presence of a corresponding emission line in the IIa-D spectrum and that its identification with Ly-α + N v be confirmed by the detection of O vi emission on the IIIa-J spectrum. We also look for a fall in the continuum level going from red to blue across Ly-α + N v and check for, but do not demand, a cutoff near the Lyman limit. When these conditions are met, the selections are no longer candidate high-redshift QSOs but certain QSOs with redshifts within about ±0.05 of those measured from the prism spectra; the accurate redshifts now available for our other high-z selections have allowed us to correct a slight systematic bias of +0.03 in our earlier prism redshifts. These prime selections are necessarily restricted to bright QSOs with strong O vi as well as strong Ly-α + N v emission and including 1208+1011 we have so far found only five objects[2-4] which could confidently be classified as QSOs with $z \geq 3.3$ from the prism data alone; a sixth, the broad absorption line (BAL) QSO 0105−2634 with $z = 3.50$, appeared a certain QSO from the prism data but was relegated to the 'candidate' category because of its unusual colours[4]. As among the numerous IIIa-F selected candidates there may be high-z QSOs with weak O vi, and because we are still refining our selection criteria, we have not confined our follow-up slit spectroscopy to the certain selections. However,

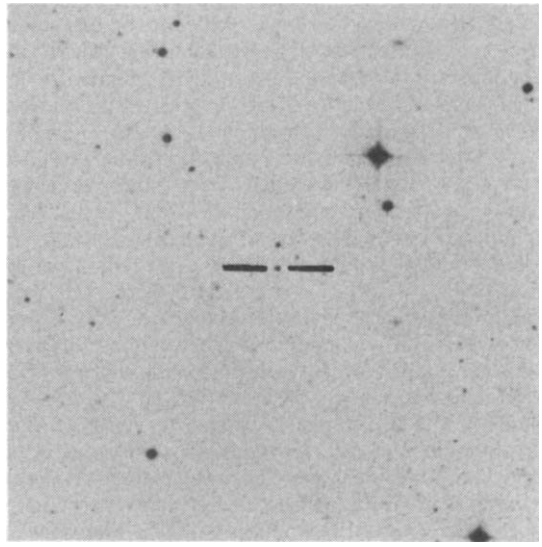


Fig. 1 Finding chart of 1208+1011, 8×8 arc min reproduced from a direct IIIa-J UKST plate (courtesy of the UKST unit). North is up, and east is to the left.

Table 1 Measured wavelengths of emission lines, with corresponding redshifts

Emulsion	Feature	Wavelength* (Å)	Identification	Redshift*
IIIa-F	Emission	5,860±50	Ly-α + N v	3.82±0.04
IIa-D	Emission	~5,950†	Ly-α + N v	~3.89
IIIa-J‡	Emission	4,980±80	O vi	3.82±0.08
			Adopted redshift	3.82±0.04

* The errors are maximum errors. The calibration curve for the IIIa-F spectra was normalized using Ly-α + N v in QSOs of known redshift, assuming the wavelength of the blend to be 1,216 Å. It automatically takes into account the N v contribution unless it differs markedly from the average.

† Ly-α + N v is too close to the red limit of the IIa-D spectrum to be accurately measured. The quoted wavelength is based on measurements relative to nearby stars.

‡ The IIIa-J spectrum dispersion is perpendicular to that of the IIIa-F spectrum, and is thus not confused with the nearby star.