Amino acid sequence changes in the haemagglutinin of A/Hong Kong (H3N2) influenza virus during the period 1968–77

W. G. Laver, G. M. Air, T. A. Dopheide* & C. W. Ward*

Department of Microbiology, John Curtin School of Medical Research, Australian National University, PO Box 334, Canberra City, ACT, 2601, Australia

Haemagglutinin molecules from nine strains of A/Hong Kong/68 (H3N2) influenza virus, isolated between 1968 and 1977, were examined for changes in amino acid sequences. At least 18 changes, 9 of which were located precisely, occurred in the soluble tryptic peptides of the large haemagglutinin polypeptide (HA1) during this period. These peptides contained 262 residues (82% of HA1). In HA2, only two changes in 129 residues (58% of HA2) were detected. Sequential changes at a particular locus were not found; and as far as we can tell, once an amino acid changed, it did not change again in any subsequent variant examined.

INFLUENZA persists as an uncontrollable infection in man largely because of the high frequency with which antigenic changes occur in the surface proteins of the virus. Both influenza A and B viruses have two distinct surface antigens, the haemagglutinin (HA) and neuraminidase (NA). These may undergo a gradual change (antigenic drift) under increasing antibody prevalence in the population. Significant drift in the HA antigen is often associated with influenza epidemics of varying impact which occur without any regular time intervals. Occasionally, the HA and NA antigens of influenza A virus may abruptly undergo a complete change (antigenic shift) and bring about an influenza pandemic. Two recent influenza pandemics occurred, in 1957, when both the HA and the NA antigens changed ('Asian 'flu'), and in 1968, when only the HA antigen changed ('Hong Kong 'flu'). Influenza B viruses undergo antigenic drift, but are not known to undergo antigenic shift and have not been associated with pandemics.

Antigenic drift in type A influenza virus is believed to occur by the selection, in an immune population, of mutant virus particles with altered antigenic determinants on the HA and NA proteins. Peptide maps have shown that antigenic shift is associated with major differences in the amino acid sequence of the HA, and drift with minor differences[1].

Antigenic drift in the Hong Kong subtype

Since the appearance of the Hong Kong (H3N2) subtype of influenza type A virus in man in 1968, drift has occurred, resulting in the appearance of new variants, several of which are listed in Table 1. The original A/Hong Kong/68, and some of these variants, such as A/England/42/72, A/Port Chalmers/1/73, A/Victoria/1/75 and A/Texas/1/77, became widespread, causing epidemics in many countries, but most of the others were associated with only minor outbreaks and quickly disappeared[2]. Antigenic relationships between the viruses listed in Table 1 have been described previously[3,4].

In this study, we have examined the HAs of these nine Hong Kong variants for amino acid sequence changes, by matching the compositions of their soluble tryptic peptides with the known amino acid sequence of the homologous peptides from the Hong Kong variant, A/Memphis/102/72 (ref. 5).

The HA monomer is coded by a single, monocistronic segment of RNA[6,7] and is synthesised as a single polypeptide chain which is then normally cleaved by proteolytic enzymes to give two chains, HA1 and HA2, held together by disulphide bonds[8]. These two chains can be readily separated by column chromatography or density gradient centrifugation after denaturation and disruption of the disulphide bonds. HA1 is the N-terminal portion of the HA[9], has a molecular weight of 47,000 (ref. 10) and contains most of the carbohydrate[10,11]. HA2 is the C-terminal portion of the HA[9], has a molecular weight of approximately 30,000 (ref. 10), carries very little carbohydrate[10,11] and contains the hydrophobic tail by which the HA is attached to the viral membrane[9,12,13].

Sequence changes in HA1

Maps of the soluble tryptic peptides from HA1 of A/Hong Kong/68 and A/Memphis/102/72 viruses showed several differences[14]. Therefore, a detailed study of the changes occurring in nine different Hong Kong variants (Table 1) was carried out.

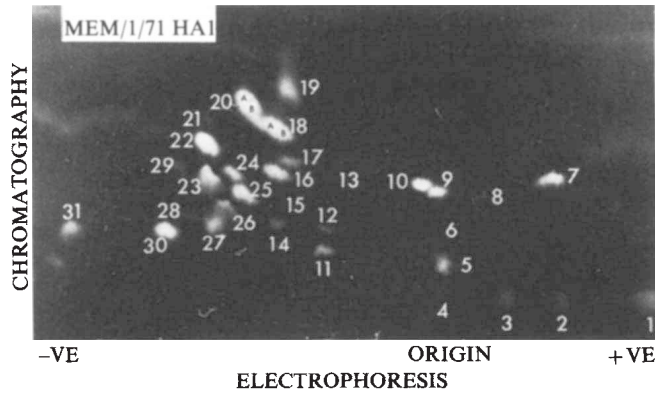

Fig. 1 Map of the soluble tryptic peptides from *S*-carboxymethylated HA1 of A/Memphis/1/71 influenza virus. Virus was grown in the allantoic sac of 11-d-old chick embryos and purified as described previously[22]. Haemagglutinin was isolated by electrophoresis of SDS-disrupted virus on cellulose acetate strips[19]. Heavy (HA1) and light (HA2) chains were separated by centrifugation on a guanidine hydrochloride–dithiothreitol density gradient[8]. The isolated chains were *S*-carboxymethylated as previously described[20], dialysed and precipitated with 3 volumes of ethanol. Digestion with trypsin (TPCK treated; Worthington), and mapping of the peptides soluble at *p*H 6.5 were as described elsewhere[14]. Peptides were located by staining with fluorescamine (10 μg per ml) in acetone containing 0.5% pyridine[21], eluted from the maps with 6 M HCl, hydrolysed in sealed, evacuated tubes for 22 h at 105 °C and analysed on a Beckman 119CL amino acid analyser. Tryptophan was qualitatively determined by staining duplicate maps with Ehrlich reagent. These methods were followed to prepare maps of HA1 and HA2 from the H3N2 viruses listed in Table 1. These viruses were grown as recombinants with the neuraminidase from A/Bel/42 (HON1).

* CSIRO, Division of Protein Chemistry, 343 Royal Parade, Parkville, Victoria 3052, Australia.

0028–0836/80/050454–04$01.00

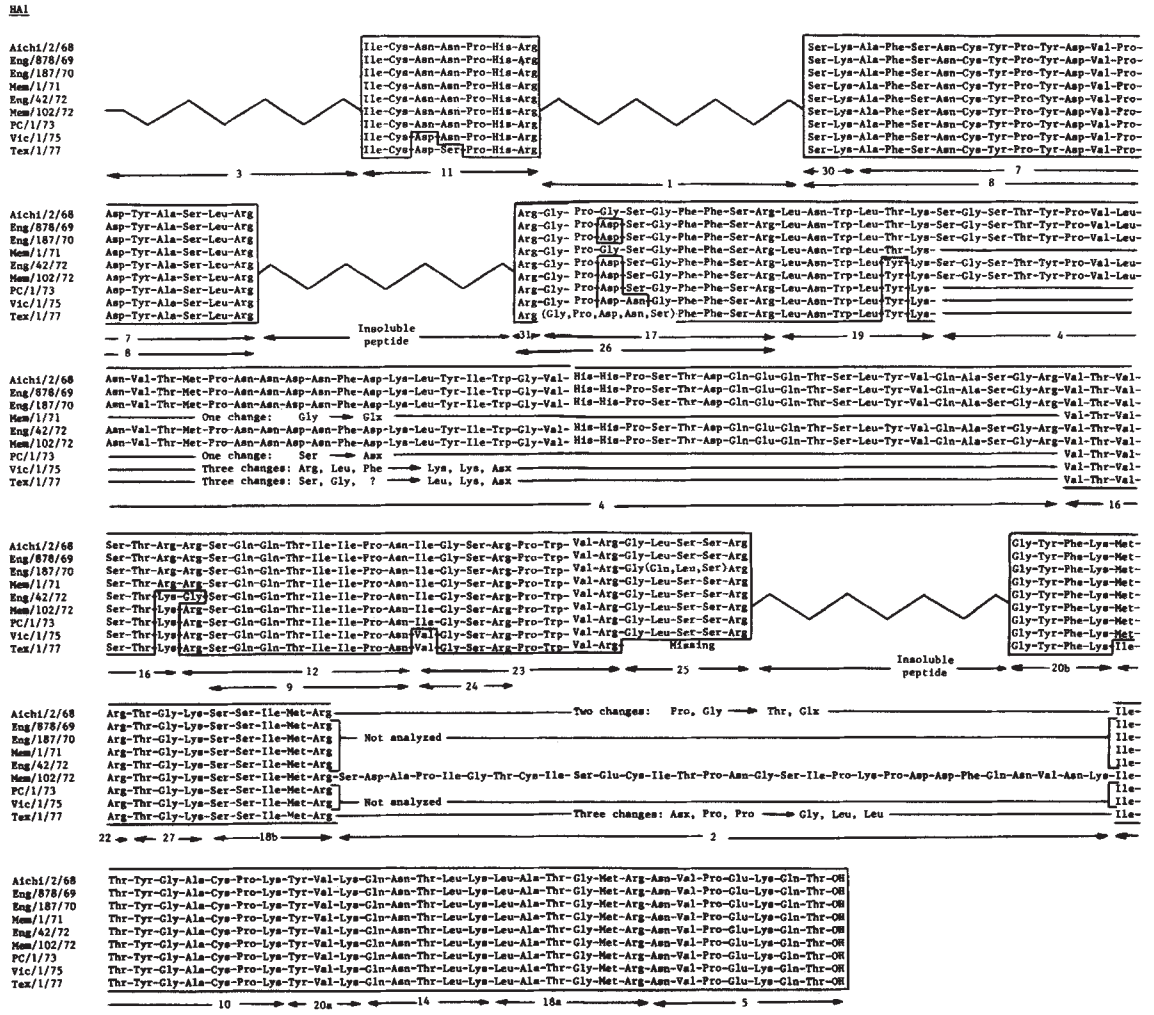

Fig. 2 Comparative partial amino acid sequences of the HA heavy chains (HA1) from nine different Hong Kong influenza A variants isolated between 1968 and 1977. The amino acid sequences were determined for the A/Memphis/102/72 strain (ref. 5 and C. W. W. and T. A. D., unpublished). Sequences for the other strains were deduced from the amino acid composition of their tryptic peptides. Boxed areas indicate regions of identical sequence and the residues which changed are indicated. It is not known which of the two Gly residues of Vic/75 peptide 17 changed to Ser in Texas/77, nor how one of the two Ser residues in Aichi/68 peptide 25 changed to Gln in Eng/187/70. Peptide numbers refer to the spots shown in Fig. 1. The four zig-zag lines denote regions of sequence associated either with the large peptides 1 and 3, which have not been fully sequenced and which therefore were not reliably analysed by single hydrolysis, or with two large insoluble tryptic peptides which we were unable to separate and analyse. The sequence of one of these (between peptides 25 and 20b) for the Mem/102/72 strain has been published[5] but no sequence data for this peptide from the other strains are available. We have previously found that variants of A/Memphis/1/71 virus selected with monoclonal antibodies showed single amino acid sequence changes[15]. These were in peptide 11, where the Asn at position 4 changed to Lys; peptide 16, where the Ser changed to Tyr; peptide 17 (26) where the Pro changed to Ser, Leu, Thr or His, and peptide 25, which was missing in one monoclonal variant, although the change responsible was not determined.

Figure 1 shows a map of the soluble tryptic peptides from *S*-carboxymethylated HA1 of A/Memphis/1/71. The assumed sequences and locations of these A/Memphis/1/71 peptides, based on their amino acid composition[15] and the available sequence data for the closely related A/Memphis/102/72 strain[5], are shown in Fig. 2 with the corresponding data for the soluble HA1 tryptic peptides from the other eight Hong Kong influenza A variants examined.

Figure 2 shows that most of the primary structure of Hong Kong virus HA remained unchanged during the period 1968–77. In the soluble tryptic peptides, which contained 262 amino acids (82% of HA1), 18 changes in sequence could be detected. Of these, nine were located precisely, but in the four large peptides (1, 2, 3 and 4), where changes in composition were seen, we cannot know the exact number of sequence changes without more detailed analysis of sub-fragments. There may be more changes in sequence than detected by composition analysis. The regions of sequence not covered in this report are associated with two large insoluble tryptic peptides that will require additional enzymatic digestion before sequence analysis.

In most cases, once an amino acid residue had changed, it did not change again in any of the other variants subsequently isolated. Exceptions to this were seen in four cases where apparent reversion to the original amino acid was seen. Thus, in peptide 12, the arginine in position 1 in A/Aichi/2/68, A/England/878/69, A/England/187/70 and A/Memphis/1/71 was replaced by glycine in A/England/42/72 but is arginine in the subsequent strains. A likely explanation is that this isolate of A/England/42/72 was not the virus that gave rise to the 1972–77 variants. As reported previously[1], peptide 25 from HA1 of A/Queensland/7/70 virus had the leucine residue replaced by glutamine. Although this change was not found in

Table 1 Hong Kong (H3N2) variants examined

A/Aichi/2/68	A/Memphis/102/72
A/England/878/69	A/Port Chalmers/1/73
A/England/187/70	A/Victoria/1/75
A/Memphis/1/71	A/Texas/1/77
A/England/42/72	

All these viruses were obtained from WHO reference laboratories, with the exception of the Memphis strains, which were local isolates. The viruses were grown as recombinants with the neuraminidase from A/Bel/42 (HON1). The recombinants were cloned twice at limit dilution before analysis.

any of the Hong Kong strains examined in this report, the Eng/187/70 strain had a Ser→Gln change in the same peptide (see Fig. 2). The most likely explanation is that the A/Queensland/7/70 and Eng/187/70 strains did not persist in nature and did not give rise to subsequent strains. The third example is in peptide 4, where in Mem/1/71 there is a Gly→Gln change not seen subsequently. The fourth example of apparent reversion back to the original residue can be seen in peptides 17/26 where the glycine at position 4 is replaced by aspartic acid in the 1969 and 1970 England variants, but not in the 1971 American strain, A/Memphis/1/71. All subsequent variants (1972–77) have aspartic acid in this position. Although the nine variants studied here represent Hong Kong variants isolated in chronological order, they do not necessarily represent an exact geneological series. Variants showing this Gly→Asp change in peptides 17/26 apparently arose first in England, then spread to, or arose independently in, the US.

Almost all the changes we have observed in compositions of small peptides can be accounted for by a single nucleotide change. One exception is in peptide 25 of Eng/187/70 (and no other strain we have examined), where there is a Ser→Gln change. It would be interesting to sequence this peptide, because the sequence could show two changes, for example,

A A
CUG UCG

Gly - Leu - Ser - Ser - Arg

↓ ↓

Gly - Gln - Leu - Ser - Arg

A A
CAG UUG

The Leu→Gln change has been found in Qld/7/70, although peptide 25 in subsequent strains is the same as in 1968 and 1969 strains until Texas/1/77, where the peptide disappears (indicating some other change).

The other 2-base change appears in peptide 19 of Eng/42/72 and Mem/102/72, and persists through to Texas/1/77, being Thr→Tyr

U
ACX→UAC

In the large peptides 4 and 2, several of the observed changes in composition would require at least two nucleotide changes in the RNA.

Sequence changes in HA2

Figure 3 shows the map of the soluble tryptic peptides obtained from *S*-carboxymethylated HA2 of A/Memphis/1/71 virus. Using the sequence data available for A/Memphis/102/72 (ref. 5) it was possible to deduce the sequences of most of these A/Memphis/1/71 peptides from their amino acid compositions[15]. The location of each of these peptides in the partial sequence of HA2 is shown in Fig. 4. These peptides account for 129 residues (58% of the total HA2 sequence), the rest being insoluble and not examined.

Although we have not been able to characterise the exact changes, the data presented in Fig. 4 show that in the 20 soluble HA2 peptides examined, only two differences could be detected between the first and last Hong Kong strains. This suggested that extensive variation would not be found in the intermediate strains, and therefore these were not examined.

Relationship between sequence changes and antigenicity

The analysis of amino acid sequence changes in natural influenza variants may not necessarily indicate which portions of the sequences make up the antigenic determinants. Some of the changes observed may be unrelated to the antigenic differences between the various HAs. Considering the large antigenic difference between Hong Kong/68 and Texas/77 viruses[3], and the large number of antigenic sites on the HA molecule[16], the number of sequence changes found associated with antigenic drift was surprisingly small. In most cases, only one or two amino acid changes were seen between neighbouring drift strains. The heavy chain (HA1) from the last member of the series, A/Texas/1/77, differed from that of the first, A/Aichi/2/68, by about 18 out of 262 residues, whereas the light chain (HA2) showed only two amino acid substitutions in the soluble tryptic peptides during antigenic drift in this period.

We do not know which of the changes in sequence found to occur between 1968 and 1977 were responsible for the changes in antigenicity which occurred in the Hong Kong virus during this period. In a previous study[15], influenza variants of A/Memphis/1/71 were selected with monoclonal hybridoma antibodies and examined for changes in amino acid sequence and antigenicity. These variants showed dramatic changes in antigenicity, having completely lost the capacity to bind the monoclonal antibody used for their selection. These changes were found to be associated with a single amino acid substitution in HA1, which in every case required only a single base change in the RNA. No changes were seen in HA2. The HA1 tryptic peptides in which alterations occurred in the variants were the same as those in which changes occurred in the field strains, but in every instance the amino acid changes in the monoclonal variants were different from those found in the field strains (Fig. 2).

The recombinant viruses we used as the source of HA were cloned twice at limit dilution before analysis; this may have led to the separation of variant viruses with sequence changes in the HA. Furthermore, HA molecules of viruses isolated late in the Hong Kong era, antigenically related to the 'Victoria' and 'Texas' strains, may well have a number of sequence changes from the Victoria and Texas strains we examined. Asian (H2N2) viruses isolated in different parts of the world in 1968 at the end of the H2N2 period showed considerable differences in their HA1 and HA2 tryptic peptide maps[17].

Electron microscopy of antibody molecules attached to the HA has shown that some (and possibly all) of the antigenic determinants are situated in a region just below the tip of the HA spike[18]. In the present work we have found changes along the whole length of the HA polypeptides, which may reflect the way in which the polypeptide chains are folded within the HA spike.

On the other hand, we do not know which of the sequence changes are responsible for the antigenic differences between the various viruses and which are not. Neither do we know if the sequence changes responsible for the antigenic changes occur in

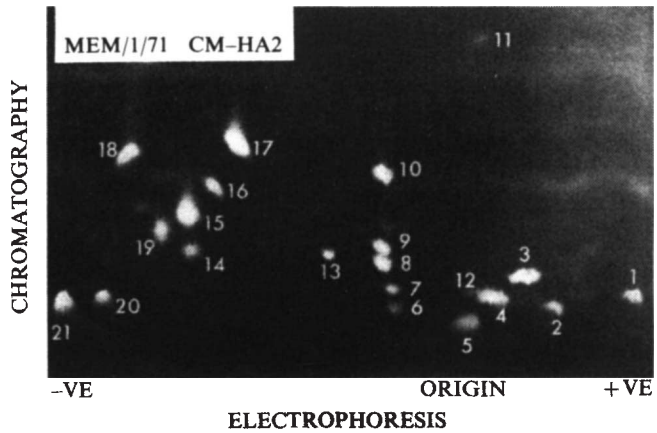

Fig. 3 Map of the soluble tryptic peptides from *S*-carboxymethylated HA2 of A/Memphis/1/71 influenza virus. The methods for virus growth, HA2 purification, enzyme digestion and peptide mapping are described in Fig. 1 legend.

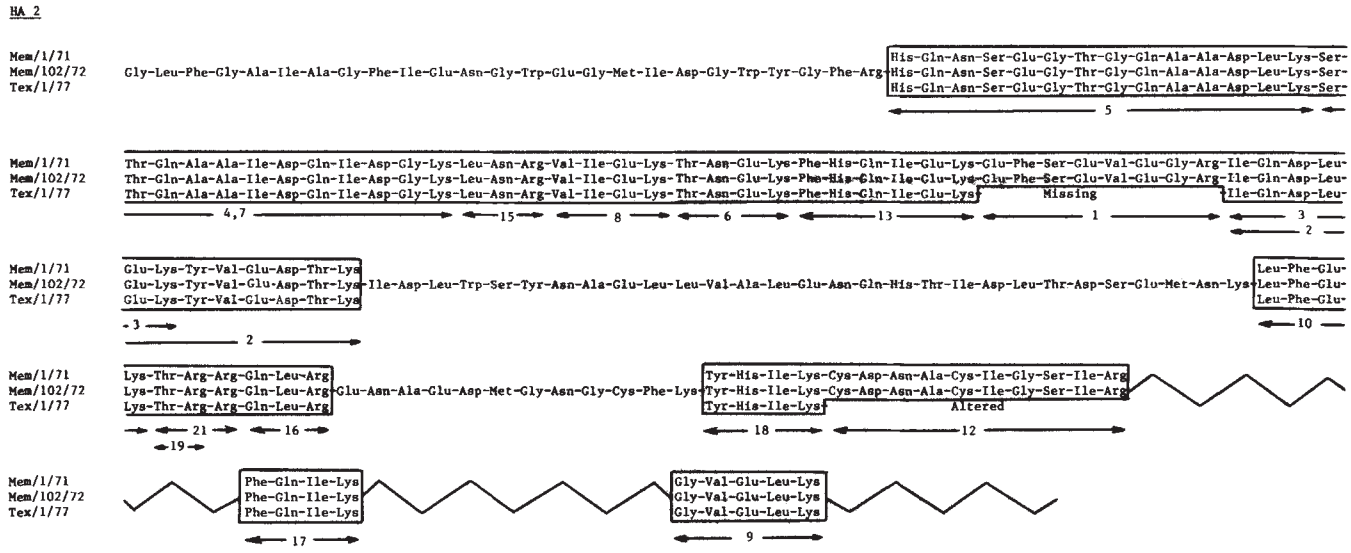

Fig. 4 Comparative amino acid sequences of A/Memphis/1/71, A/Memphis/102/72 and A/Texas/1/77 HA light chains (HA2). The amino acid sequences have been determined for the A/Memphis/102/72 strain[5]. Sequences for the other two strains were deduced from their amino acid compositions. The peptide numbers refer to the corresponding spots shown in Fig. 3. The regions of residues 1–25 and 89–117 and much of the C-terminal sequence involve large insoluble tryptic peptides not analysed in this study. In peptide 18 the sequence has been revised and corrected from that given previously[5].

those amino acids which constitute the antigenic determinants, or whether the sequence changes occur elsewhere in the molecule and cause conformational changes which alter the antigenic determinants.

Received 24 July; accepted 15 November 1979.

1. Laver, W. G. & Webster, R. G. *Virology* **34,** 193–202 (1968); *Virology* **48,** 445–455 (1972).
2. Pereira, M. S. in *Influenza, Virus, Vaccines and Strategy* (ed. Selby, P.) 25–31 (Academic, New York, 1976).
3. Pereira, M. S. *Br. med. Bull.* **35,** 9–14 (1979).
4. Periera, M. S. & Schild, G. C. *J. Hyg., Camb.* **69,** 99–103 (1971).
5. Dopheide, T. A. & Ward, C. W. *Eur. J. Biochem.* **85,** 393–398 (1978); *Virology* **92,** 230–235 (1979); *Virology* **95,** 107–118 (1979); *Br. med. Bull.* **35,** 51–56 (1979).
6. Palese, P. & Schulman, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73,** 2142–2146 (1976).
7. Scholtissek, C., Harms, E., Rohde, W., Orlich, M. & Rott, R. *Virology* **74,** 332–344 (1976).
8. Laver, W. G. *Virology* **45,** 275–288 (1971).
9. Skehel, J. J. & Waterfield, M. D. *Proc natn. Acad. Sci. U.S.A.* **72,** 93–97 (1975).
10. Ward, C. W. & Dopheide, T. A. *FEBS Lett.* **65,** 365–368 (1976).

We thank Jean Clark and Donna Boyle for technical assistance. The viruses were supplied by Dr R. Webster. This work was supported by grants AI-15343 and AI-42510 from the National Institutes of Allergy and Infectious Diseases.

11. Skehel, J. J. & Schild, G. C. *Virology* **44,** 396–405 (1971).
12. Laver, W. G. & Valentine, R. C. *Virology* **38,** 105–119 (1969).
13. Wrigley, N. G. *Br. med. Bull.* **35,** 35–38 (1979).
14. Laver, W. G., Downie, J. C. & Webster, R. G. *Virology* **59,** 230–244 (1974).
15. Laver, W. G. *et al. Virology* **98,** 226–237 (1979).
16. Yewdell, J. W., Webster, R. G. & Gerhard, W. *Nature* **279,** 246–248 (1979).
17. Laver, W. G. & Webster, R. G. *Virology* **48,** 445–455 (1972).
18. Wrigley, N. G., Laver, W. G. & Downie, J. C. *J. molec. Biol.* **109,** 405–421 (1977).
19. Laver, W. G. *J. molec. Biol.* **9,** 109–124 (1964).
20. Air, G. M. & Thompson, E. O. P. *Aust. J. biol. Sci.* **22,** 1437–1454 (1969).
21. Udenfriend, S. *et al. Science* **178,** 871–872 (1972).
22. Laver, W. G. in *Fundamental Techniques in Virology* (eds Habel, D. & Salzman, N. P.) 82–86 (Academic, New York, 1969).

LETTERS

IR sources in the Southern Coalsack

M. J. Smyth
University of Edinburgh, Department of Astronomy, Edinburgh

M. E. Sim
Royal Observatory, Edinburgh, UK

A preliminary survey of four dark globules in the Southern Coalsack, an area of conspicuous obscuration in the southern sky, has revealed seven IR sources, which are most probably heavily obscured background stars. We report here that the wavelength dependence of extinction is not significantly different from that in the general interstellar medium. Spectroscopy of these stars, and of still more heavily obscured stars that should be revealed by a deeper survey, would permit detailed analysis of 'proto-protostellar' material. Although these are probably background stars, the numbers found in our small sample at least suggest association with the globules themselves; there is growing evidence for star formation within or associated with dense dark globules.

Bok[1] has reviewed the important role of dark globules in our understanding of star birth in the Galaxy, and has suggested the use of IR photometry to probe these heavily absorbing 'proto-protostars'. The Southern Coalsack, the finest dark nebula in the southern Milky Way, contains many globules, which appear as roundish, sharply bounded dark nebulae a few arc-min in diameter. Bok, Sim and Hawarden[2] have superimposed blue-green and far-red photographs of Globule 2 of the survey of Tapia[3], the far-red images revealing peripheral reddened background stars invisible on the blue-green photographs.

We have used the 0.75 m telescope of the South African Astronomical Observatory to make a quick survey of four Coalsack globules, numbers 1, 2, 3 and 9 in Fig. 1 of ref. 2, at a wavelength of 2.2 μm (K magnitudes). The areas scanned were squares of 7 × 7 arc min in each case. For illustrations we again used two photographs taken with the SRC 1.2 m Schmidt telescope at Siding Spring, Australia. Their star images were almost superimposed in a modified Zeiss blink comparator, and Fig. 1 shows photographs of the four globules taken from the TV monitor. Black images correspond to stars in the far-red photograph ($\bar{\lambda} \sim 0.8$ μm), and displaced white images to those in the blue-green photograph ($\bar{\lambda} \sim 0.5$ μm); again, far-red images unaccompanied by blue-green images can be seen within the globule boundaries, representing heavily reddened background

0028-0836/80/050457—03$01.00