The Amino Acid Sequence of Cytochrome *c* from the Locust, *Schistocerca gregaria* Forskal

By ANDREW LYDDIATT* and DONALD BOULTER

Department of Botany, University of Durham, Durham DH1 3LE, U.K.

(Received 11 October 1976)

The amino acid sequence of locust cytochrome *c* was determined, although the overlap between chymotryptic and tryptic peptides at residues tyrosine-97 and leucine-98 was not observed, owing to an anomalous tryptic break duplicating the chymotryptic digestion. The molecule consists of a single polypeptide chain of 107 residues, homologous with other mitochondrial cytochromes *c*. In common with other known insect cytochromes *c*, it possesses a non-acetylated, four-residue tail at the *N*-terminus relative to glycine-1 of the standard alignment. A molecular phylogeny for 17 species was constructed relating the cytochrome *c* molecules of *Schistocerca gregaria* and other invertebrates with those of representative taxonomic groups. Experimental details are given in a supplementary paper deposited as Supplementary Publication SUP 50077 (24 pages) at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1977) 161, 1.

The use of amino acid-sequence data for cytochrome *c* in the construction of molecular phylogenies is well documented (Dayhoff, 1972; Boulter, 1973). However, within the animal kingdom, there has been a strong bias towards the vertebrates, which provide 29 sequences compared with the six known from invertebrate sources (see Dayhoff, 1972, 1973). The cytochrome *c* of aerobic eukaryotes functions as an electron carrier within the terminal oxidation chain (see Lemberg & Barrett, 1973), and an organism may be judged as a potential source of the protein from the extent of its muscular activity. The disregard for invertebrate cytochromes *c*, despite their phylogenetic interest, has been based on the expectation of low yields from the majority of members of this animal group. An exception is the phylum Arthropoda, whose members are frequently muscularly active and from which sequences of five insect cytochromes *c* have been reported [Chan & Margoliash (1966); S. K. Chan, J. Tulloss & E. Margoliash, unpublished work cited by Dayhoff (1972); C. Nolan, L. J. Weiss, J. J. Adams & E. Margoliash, unpublished work cited by Wojciech & Margoliash (1968); Chan, 1970; Fernandez-Sousa *et al.*, 1975). The determination of the sequence of *Schistocerca* cytochrome *c* provides the first sequence data for the insect order Orthoptera and allows comparisons with the orders Diptera and Lepidoptera.

Experimental

Materials

Locusts were bred by Dr. N. Jago at the Centre for Overseas Pest Research, Wrights Lane, London W.8, U.K., and supplied by Dr. R. P. S. Jefferies of the British Museum (Natural History), London S.W.7, U.K. They were stored at -20°C until required. Trypsin (Tos-Phe-CH₂Cl*-treated, twice recrystallized) and α -chymotrypsin (twice recrystallized) were from Worthington Biochemical Corp., Freehold, NJ, U.S.A. Carboxypeptidase A (di-isopropyl phosphorofluoridate-treated) was from Sigma (London) Chemical Co., London S.W.6, U.K. 9,10-Phenanthraquinone was from Koch-Light Laboratories, Colnbrook, Bucks., U.K. Pyridine (no-ninhydrin grade) was from Rathburn Chemicals, Walkerburn, Peeblesshire, Scotland, U.K. Triethylamine and chemical kits for the Beckman 890C Sequencer (Beckman, 1972) were from Pierce Chemicals, Rockford, IL, U.S.A.

Other materials were as described by Thompson *et al.* (1970, 1971).

Methods

The isolation and purification of cytochrome *c* from *Schistocerca* was essentially as described for plant material (Richardson *et al.*, 1970, 1971), with

Abbreviations: Tos-Phe-CH₂Cl, 1-chloro-4-phenyl-3-tosylamidobutan-2-one (TPCK); dansyl-, 5-dimethylaminonaphthalene-1-sulphonyl-; SE-Sephadex, sulphoethyl-Sephadex.

* Present address: Department of Genetics, Trinity College, Dublin 2, Ireland.

Fig. 1. Amino acid sequence of *Schistocerca cytochrome c*

Residues were identified by dansylphenylisothiocyanate analysis (→), amino acid-composition data (→), carboxypeptidase-A digestion (→) and as the free C-terminal amino acid after the final Edman degradation (→). Arrows (↓, ↑, ↓ and ↓) indicate points of complete, partial and secondary enzyme cleavage respectively; the arrows point up for trypsin and down for chymotrypsin. Prefixes T and C refer to tryptic and chymotryptic peptides respectively, and peptides derived from partial cleavage have a letter added to the major peptide. Peptide C2 was digested with chymotrypsin after removal of the haem moiety (Nolan & Margoliash, 1966). Residues -4 to 37 were determined by using an automatic sequencer.

the following modifications. (1) Frozen locusts (3 kg) were homogenized with an equal volume of ice in 2 vol. of 10 mM- $\text{Al}_2(\text{SO}_4)_3$ maintained at pH 4.5 with 2 M- H_2SO_4 . The homogenate was stirred at 2°C for 2 h. (2) The filtered homogenate was adjusted to pH 8.0 with 2 M-NaOH and pumped through a column (18 cm $\times$ 10 cm) of Amberlite CG-50 at 15 litres/h. Cytochrome *c* was eluted from Amberlite CG-50 in 2 M-NaCl maintained at pH 8.0 with 2 M-NaOH. (3) *Schistocerca* cytochrome *c* showed no precipitation in satd. $(\text{NH}_4)_2\text{SO}_4$ solutions at pH 8.0; 80% satd. solutions were centrifuged at 30000g (r_{av} , 4.2 cm) for 20 min at 2°C and the supernatant was exhaustively dialysed against 10 mM- NaH_2PO_4 /Na₂HPO₄ buffer, pH 7.2, before concentration on a column (1 cm $\times$ 5 cm) of CM-52 cellulose equilibrated in the same buffer.

The amino acid sequence was determined as described by Thompson *et al.* (1970, 1971), with the following modifications. (1) Peptides were purified by electrophoresis at pH 6.5 and pH 1.9, together with descending chromatography on Whatman 3 MM paper with butan-1-ol/acetic acid/water/pyridine (15:3:12:10, by vol.) as solvent (Waley & Watson, 1953). (2) Peptides were located on guide strips by using a combination of the phenanthraquinone, ninhydrin and *p*-dimethylaminobenzaldehyde reagents (Yamada & Itano, 1966). (3) Purified peptides were sequenced by using a variation of the dansyl-phenylisothiocyanate method (Gray & Hartley, 1963) described by Ambler & Wynn (1973). Freeze-dried peptide samples were dansylated with 10 μ l of a 1:1 (v/v) mixture of 0.1 M-triethylamine and a solution containing 5 mg of dansyl chloride/ml of acetone. The reaction was for 30 min at 45°C and terminated by drying *in vacuo*. Dansyl derivatives were identified after chromatography on 5 cm $\times$ 5 cm polyamidesheets in the solvent system described by Ramshaw *et al.* (1970). C-Terminal residues were determined by carboxypeptidase-A analysis as described by Ramshaw *et al.* (1974a). (4) Automatic sequence analysis was performed by using a Beckman 890C sequencer as described by Ramshaw *et al.* (1974b).

Results

Pure cytochrome *c* (oxidized), which had a spectral ratio A_{410}/A_{280} of 4.0, was obtained in a yield of 20 mg/kg of starting material. N-Terminal analysis of the total protein by the methods of Gros & Labouesse (1969) and Gray (1972) gave a single residue, glycine, as the N-terminus.

The amino acid composition was determined from three quadruple 50 nmol samples of protein hydrolysed for 22, 48 and 72 h, and the results are given in Table 1. The values are in agreement with the sequence results, although in preliminary analyses a discrepancy of one residue was found between the

Table 1. *Amino acid composition of Schistocerca cytochrome c*

The composition was determined from three quadruple 50 nmol samples hydrolysed for 22, 48 and 72 h. The results are expressed as residues/molecule. The mean values are expressed as a mean of the twelve hydrolyses; the values extrapolated to zero time (*) were found by assuming first-order kinetics for destruction (Moore & Stein, 1963).

Time of hydrolysis (h) ...	Composition (residues/molecule)				
	Experimental values			Mean or extrapolated value	Sequence
	22	48	72		
Asp	8.4	8.2	8.6	8.4	8
Thr	7.4	7.0	6.7	7.7*	8
Ser	3.0	2.6	2.2	3.4*	3
Glu	10.3	10.6	10.8	10.6	11
Pro	6.5	6.1	6.6	6.4	6
Gly	12.7	13.0	13.2	13.0	13
Ala	7.3	7.6	7.4	7.4	7
Val	5.0	5.0	5.2	5.1	5
Met	1.0	1.0	1.0	1.0	1
Ile	4.9	4.9	4.9	4.9	5
Leu	7.4	7.4	7.4	7.4	7
Tyr	3.1	4.0	3.8	3.6	4
Phe	4.6	4.7	4.7	4.7	5
His	2.8	2.5	2.7	2.7	3
Lys	14.7	14.7	14.3	14.6	15
Arg	3.2	3.1	3.2	3.2	3
Cys	2.1	2.6	1.9	2.2	2
Trp	—	—	—	—	—

amino acid composition and sequence composition for serine, glutamic acid, phenylalanine and lysine. Tryptophan was not found by amino acid analysis, but spectrophotometric (Goodwin & Morton, 1946) and sequence analysis indicated that a single residue was present.

The amino acid sequence was determined from the analysis of two chymotryptic digestions, one tryptic digestion and from results obtained with a Beckman 890C sequencer. (The details of the experimental evidence for the sequence of locust cytochrome *c* is given in Supplementary Publication SUP 50077.) The sequence is shown in Fig. 1, together with the points of enzyme cleavage and the overlapping peptides from which the sequence was deduced. The observed enzyme specificities were consistent with those expected (Smyth, 1967; Kasper, 1970), although partial tryptic cleavage was observed at phenylalanine-65 and phenylalanine-82, and full tryptic cleavage at tyrosine-97. All residues were positively identified in both chymotryptic and tryptic digestions except for residues 8, 39 and 59, which were placed from the evidence of the chymotryptic digestions or the automatic-sequencer data.

Fig. 2. *Phylogeny constructed by the ancestral-sequence method relating 17 amino acid sequences of cytochrome c*. Sequences were aligned relative to the cysteine residues; absent residues at the *N*- and *C*-terminus were computed as amino acid differences. No equal alternatives were found for this minimum topology. A computation of the 103 residues common to the sequences gave an identical, minimum topology with shortened branch-lengths. All sequences were taken from Dayhoff (1972, 1973) except: (a) Brown *et al.* (1972); (b) Fernandez-Sousa *et al.* (1975); (c) Meatyard & Boulter (1974).

All possible overlaps between chymotryptic and tryptic peptides were observed except in the region of peptides C2CA–C2CB/T4CA–T4CB (see the Supplement for details of terminology), where the order was made plain from the sequencer data, and in the region of peptides C14–C15/T13A–T13B, where the order was clear from the overlaps apparent at peptides C13–C14/T13A and C15(C15A)/T13B–T14. All acidic and amide residues indicated in the sequence were placed from a combination of electrophoretic mobilities of intact or partially degraded peptides and automatic-sequencer analysis.

Discussion

Schistocerca cytochrome *c* was readily denatured in 95% (v/v) ethanol (Margoliash *et al.*, 1962) and was

adequately digested with chymotrypsin and trypsin, which showed the specificities generally expected (Smyth, 1967; Kasper, 1970). However, Tos-Phe-CH₂Cl-treated trypsin totally cleaved a tyrosine-leucine peptide bond, and partially cleaved a phenylalanine-isoleucine bond and a phenylalanine-alanine bond. The activity of trypsin at tyrosyl bonds in peptides has been reported (Matsubara *et al.*, 1967) and was attributed to chymotryptic contamination removable by treatment with Tos-Phe-CH₂Cl and SE-Sephadex chromatography (Keil-Dlouha *et al.*, 1971). There are no reports of trypsin acting at phenylalanyl bonds in peptides, although other anomalies have been reported (Bachmayer *et al.*, 1968; Plapp *et al.*, 1967).

A comparison of the six insect cytochromes *c*

shows that they are all 107 residues in length and have a non-acetylated *N*-terminus and four additional residues in the *N*-terminal region relative to glycine-1 of the standard alignment of cytochromes *c* (see Dayhoff, 1972). This 'tail' is identical in all cytochromes except the locust, which has glutamine instead of alanine in the '-1'-position (see Fig. 1); it remains to be seen if this is a general characteristic of arthropod cytochrome *c*.

The amino acid sequence of *Helix* (garden snail) cytochrome *c* (Brown *et al.*, 1972) and the insect sequences were compared with representative sequences of other taxonomic groups. A molecular phylogeny relating 17 sequences was constructed by using a computer program based on the methods of Dayhoff & Eck (1966) and Dayhoff (1972) (see Fig. 2).

The results indicate that the primary structure of insect cytochrome *c* is sufficiently variable to resolve the orders of Insecta within a molecular phylogeny. Although this one additional sequence of a member of the insect order Orthoptera simply confirms classical views in the group, an extension of the study to the Arthropoda as a whole should add to the discussion concerning the mono- or poly-phyletic origins for this group of organisms (see Siewing, 1960; Sharov, 1966; Dahl, 1969; Manton, 1973; Cisne, 1974).

We are grateful to Dr. D. Peacock for the computer analysis and to Dr. R. P. S. Jefferies for the supply of locusts.

References

- Ambler, R. P. & Wynn, M. (1973) *Biochem. J.* **131**, 485–498
- Bachmayer, H., Yasunobu, K. T., Peel, J. L. & Mayhew, S. (1968) *J. Biol. Chem.* **243**, 1022–1030
- Beckman (1972) *Instruction Manual for Model 890C Sequencer*, Beckman Instruments, Palo Alto
- Boulter, D. (1973) *Pure Appl. Chem.* **34**, 539–552
- Brown, R. H., Richardson, M., Boulter, D., Ramshaw, J. A. M. & Jefferies, R. P. S. (1972) *Biochem. J.* **128**, 971–974
- Chan, S. K. (1970) *Biochim. Biophys. Acta* **221**, 497–501
- Chan, S. K. & Margoliash, E. (1966) *J. Biol. Chem.* **241**, 335–348
- Cisne, J. L. (1974) *Science* **184**, 13–18
- Dahl, E. (1969) *Opusc. Entomol.* **34**, 1–9
- Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, vol. 5, National Biomedical Research Foundation, Silver Spring
- Dayhoff, M. O. (1973) *Atlas of Protein Sequence and Structure (Suppl.)*, vol. 5, National Biomedical Research Foundation, Silver Spring
- Dayhoff, M. O. & Eck, R. V. (1966) *Atlas of Protein Sequence and Structure*, vol. 2, National Biomedical Research Foundation, Silver Spring
- Fernandez-Sousa, J. M., Gavilanes, J. G., Municio, A. M., Paredes, J. A., Perez-Aranda, A. & Rodriguez, R. (1975) *Biochim. Biophys. Acta* **393**, 358–367
- Goodwin, T. W. & Morton, R. A. (1946) *Biochem. J.* **40**, 628–632
- Gray, W. R. (1972) *Methods Enzymol.* **25**, 121–138
- Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89**, 379–380
- Gros, C. & Labouesse, B. (1969) *Eur. J. Biochem.* **7**, 463–470
- Kasper, C. B. (1970) *Mol. Biol. Biochem. Biophys.* **8**, 137–184
- Keil-Dlouha, V., Zylber, N., Imhoff, J. M., Tong, W. T. & Keil, B. (1971) *FEBS Lett.* **16**, 287–290
- Lemberg, R. & Barrett, J. (1973) *The Cytochromes*, pp. 327–384, Academic Press, London and New York
- Manton, S. M. (1973) *J. Zool.* **171**, 111–130
- Margoliash, E., Kimmel, J. R., Hill, R. L. & Schmidt, W. T. (1962) *J. Biol. Chem.* **237**, 2148–2150
- Matsubara, H., Sasaki, R. M. & Chain, R. K. (1967) *Proc. Natl. Acad. Sci. U.S.A.* **57**, 439–445
- Meatyrd, B. T. & Boulter, D. (1974) *Phytochemistry* **13**, 2777–2782
- Moore, S. & Stein, W. H. (1963) *Methods Enzymol.* **6**, 819–831
- Nolan, C. & Margoliash, C. (1966) *J. Biol. Chem.* **241**, 1049–1059
- Plapp, B. V., Raftery, M. A. & Cole, R. D. (1967) *J. Biol. Chem.* **242**, 265–270
- Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1970) *Biochem. J.* **119**, 535–539
- Ramshaw, J. A. M., Scawen, M. D., Bailey, C. J. & Boulter, D. (1974a) *Biochem. J.* **139**, 583–592
- Ramshaw, J. A. M., Scawen, M. D. & Boulter, D. (1974b) *Biochem. J.* **141**, 835–843
- Richardson, M., Laycock, M. V., Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1970) *Phytochemistry* **9**, 2271–2280
- Richardson, M., Richardson, D., Ramshaw, J. A. M., Thompson, E. W. & Boulter, D. (1971) *J. Biochem. (Tokyo)* **69**, 811–813
- Sharov, A. G. (1966) *Basic Arthropodan Stock, with Special Reference to Insects*, Pergamon Press, Oxford
- Siewing, R. (1960) *Z. Wiss. Zool. Abt. A* **164**, 238–270
- Smyth, D. G. (1967) *Methods Enzymol.* **11**, 214–231
- Thompson, E. W., Laycock, M. V., Ramshaw, J. A. M. & Boulter, D. (1970) *Biochem. J.* **117**, 183–192
- Thompson, E. W., Richardson, M. & Boulter, D. (1971) *Biochem. J.* **121**, 439–446
- Waley, S. G. & Watson, J. (1953) *Biochem. J.* **55**, 328–337
- Wojciech, R. & Margoliash, E. (1968) in *Handbook of Biochemistry* (Sober, H. A., ed.), pp. C160–C161, Chemical Rubber Co., Cleveland
- Yamada, S. & Itano, H. A. (1966) *Biochim. Biophys. Acta* **130**, 538–540