Acta Cryst. (1978). B**34**, 3658–3662

The Crystal Structure of Met-Myoglobin from *Aplysia limacina* at 5 Å Resolution

BY L. UNGARETTI, M. BOLOGNESI, E. CANNILLO, R. OBERTI AND G. ROSSI

CNR Centro di Studio per la Cristallografia Strutturale and Istituto di Cristallografia dell'Università, via Bassi 4, 27100 *Pavia, Italy*

(*Received* 27 *April* 1978; *accepted* 14 *July* 1978)

Myoglobin from the gasteropod *Aplysia limacina* has been isolated and crystallized. The crystals are orthorhombic, space group $P2_12_12_1$, with cell dimensions $a = 52{\cdot}9$, $b = 70{\cdot}4$, $c = 32{\cdot}5$ Å; the asymmetric unit contains one molecule of the protein. The low-resolution structure has been determined by the method of isomorphous replacement with the inclusion of anomalous scattering, using five heavy-atom derivatives (mean figure of merit 0·88). The resulting electron density map at 5 Å resolution clearly shows the molecular boundary and eight segments of helix in the peculiar myoglobin fold, enclosing an ellipsoid of electron density which has been interpreted as the heme.

Introduction

In spite of the widespread investigation of myoglobins extracted from mammals, no crystallographic studies on mollusc myoglobins have yet been reported.

Chemical and kinetic studies (Antonini & Brunori, 1974) have shown that *Aplysia* myoglobin is characterized by distinct spectral and binding properties if compared with those of other hemoproteins. Recently, a spectroscopical approach (Makinen, Churg & Glick, 1978) has given evidence that in the crystals the oxyheme structure of this protein is considerably different from that of sperm-whale myoglobin. On the other hand, the sequence of *Aplysia* myoglobin, which consists of 145 amino acid residues, shows some peculiar features: among others the distal histidine is lacking and the number of phenylalanines (15) is unusually high (Tentori, Vivaldi, Carta, Antonini & Brunori, 1968). Furthermore, *Aplysia* myoglobin is reported (Brunori, Giacometti, Antonini & Wyman, 1972) to be very stable to thermal and solvent induced denaturation, and the transition proved to be fully reversible; *Aplysia* myoglobin can therefore be considered a suitable system for investigating folding phenomena in conjugated proteins.

On the basis of these considerations, the resolution of the structure of this protein was undertaken.

Materials and methods

(*a*) *Extraction and crystallization*

Purified *Aplysia* myoglobin, extracted according to the procedure of Rossi-Fanelli & Antonini (1957) from frozen buccal muscles of the mollusc, was a gift from M. Brunori. Crystals of the Met-form suitable for X-ray diffraction studies were obtained by vapour diffusion techniques as previously described (Blundell *et al.*, 1975); their shape is prismatic and those used for data collection had dimensions of approximately 0·6 × 0·3 × 0·2 mm; the space group is $P2_12_12_1$ with cell dimensions $a = 52{\cdot}9$, $b = 70{\cdot}4$, $c = 32{\cdot}5$ Å. On the basis of the primary structure (Tentori *et al.*, 1968) we calculated a molecular weight of 16 500; the unit cell can accommodate four molecules and 34% of the crystal volume is occupied by the solvent ($V_M = 1{\cdot}84$ Å^3 dalton^{-1}).

(*b*) *The preparation of heavy-atom derivatives*

In a search for derivatives, crystals were soaked in a number of solutions containing heavy-atom reagents dissolved in 3·8 *M* ammonium sulphate and 0·05 *M* phosphate or acetate buffers at pH values between 5·5 and 7·5. Precession photographs of the *hk*0 and *h*0*l* zones were used to monitor the effect of the soaking experiments. The diffraction pattern of native crystals was identical in different buffers, while some changes in intensities were noticed at pH lower than 5.

$NaUO_2(CH_3COO)_3$, K_2HgI_4 and K_2PtCl_4 markedly affected the diffraction pattern with very fast reactions. In particular, for the K_2HgI_4 derivative the changes in intensity were accompanied, in a few days, by disorder of the protein even at low concentration (0·4 m*M*), while for $NaUO_2(CH_3COO)_3$ (10 m*M*) and K_2PtCl_4 (2 m*M*) the limit of reaction could be reached in about two days without significant changes of the lattice parameters. For all these three derivatives it was not possible to interpret unambiguously all the vectors in the resulting difference Patterson maps.

Dilute solutions (2 m*M*) of uranyl acetate gave very small changes in the diffraction pattern and uranyl

complexes such as $UO_2F_5^{3-}$ and cations such as rare earths did not bind.

A good uranyl derivative was prepared by lowering the $NaUO_2(CH_3COO)_3$ concentration and reaching the final value of 5 m*M* in a step-by-step procedure. Under these conditions the intensity changes were reproducible and the crystals were suitable for measurements of heavy-atom differences. The same procedure was also tried with K_2PtCl_4 solutions without significant improvements; however, replacement of the Cl ligands with poorer leaving groups, as in $Pt(NH_3)_2Cl_2$, led to a slower reaction and to minor and different changes in the intensities.

Reasonably isomorphous Hg derivatives were obtained by using K_2HgBr_4 and $K_2Hg(SCN)_4$ which did not lead to any crystal disorder; the time of reaction and concentration of K_2HgBr_4 were varied and two derivatives (K_2HgBr_4, I and II) with different sites and occupancies were obtained.

Table 1 gives a summary of the conditions to prepare the heavy-atom derivatives used in the phase determination.

(*c*) *Data collection and determination of heavy-atom positions*

X-ray diffraction data for the native protein and the five derivatives listed in Table 1 were collected on a Philips four-circle diffractometer using a fine-focus copper tube and a graphite monochromator to select Cu $K\alpha$ radiation. The ω-scan technique was used for all the specimens and, according to the quality and dimensions of the crystals, scan speeds and angular widths of scanning were chosen in the ranges 0·025–0·040° s^{-1} and 0·8–1·6° respectively. Counting times thus varied between 20 and 60 s, whilst the background radiation was measured 5 to 10 s on each side of the reflection. The diffraction data were processed and standard deviations assigned according to Davies & Gatehouse (1973).

Anomalous-dispersion data were collected only for $NaUO_2(CH_3COO)_3$ and K_2HgBr_4 (II) derivatives; eight equivalent reflections were measured and averaged in two Bijvoet sets, while the mean of four reflections was used for each intensity value of the native and of the remaining three derivatives. Data were collected in shells of decreasing θ value and the decay of the crystals was monitored, recording three selected reflections every four hours; after a 10% decrease in the intensity of one reference reflection, data collection was interrupted.

Table 1. *Conditions for preparation of heavy-atom derivatives* (*buffer* $NaCH_3COO$ 0·05 *M*)

Derivative	Heavy-atom concentration (m*M*)	pH	Soaking duration time
$K_2Hg(SCN)_4$	0·3	6·5	40 h
$Pt(NH_3)_2Cl_2$	saturated	6·0	37 d
K_2HgBr_4 (I)	10	6·5	160 h
K_2HgBr_4 (II)	4	6·6	24 h
$NaUO_2(CH_3COO)_3$	5	6·0	72 h

The intensities were corrected for absorption using the empirical procedure of North, Phillips & Mathews (1968). The merging R factors ($R = \sum_h \sum_i |I_{hi} - I_h| / \sum_h \sum_i I_{hi}$, where I_{hi} is the ith measurement of reflection h and I_h is the mean value of the N equivalent reflections) for native and derivative data were $\leq 0{\cdot}05$; averaged intensities were processed to yield the magnitudes of the observed structure factors for the native protein F_P and heavy-atom derivatives F_{PH}.

Scale and overall temperature factors were evaluated from a Wilson plot based on 5430 native reflections at a resolution of 2·37 Å; these results were used to bring the derivatives and native data sets to a common scale.

Heavy-atom positions were determined from the analysis of three-dimensional isomorphous difference ($F_{iso} = |F_{PH} - F_P|$) Patterson maps. Difference Fourier syntheses in which the native phases (α_P) were calculated by single isomorphous replacement with the inclusion of anomalous scattering [using $NaUO_2(CH_3COO)_3$ or K_2HgBr_4 (II) data] allowed the derivatives to be brought to the same origin and confirmed the heavy-atom coordinates (Kartha & Parthasarathy, 1965). The choice of the correct enantiomorph was effected by comparing peak heights between isomorphous difference Fourier maps whose phases had been calculated on the basis of the two possible configurations. When the uranyl $|F_{PH} - F_P| \exp(i\alpha_P)$ map was calculated, α_P's were evaluated considering only the K_2HgBr_4 (II) derivative contribution and *vice versa* in the other case.

The refinement of positional and occupancy parameters of the heavy atoms was carried out by means of conventional least-squares procedures using unit weights and keeping the thermal parameters fixed to the value of 18 $Å^2$ obtained from the Wilson plot of the native. Two derivatives were refined using the function F_{HLE} (the lower estimate of the magnitude of the heavy-atom contribution to the structure factor) as observed structure amplitude and empirical values for $k = f'/f''$ (Matthews, 1966), while for the remaining three only isomorphous differences in the centric sections were considered. The reflections for which either native or derivative intensities were less than three estimated standard deviations were discarded in both refinements; no attention was given to cases in which crossing over could have occurred and no scatterers other than U, Hg and Pt were considered.

Native phases were then calculated with the refined parameters excluding the contribution of each derivative in turn; these sets of phases were used in difference Fourier syntheses to search for lower occupancy sites of that derivative which had not

contributed to the phase determination. The native data set at 5 Å resolution was phased using the results of further cycles of refinement with the inclusion of the minor sites. Once a preliminary map was available, coordinates of the heme iron were estimated and allowance was made for its contribution to the anomalous scattering of the derivatives.

The values of E (r.m.s. values of 'lack of closure' error for isomorphous differences) were calculated from the centric refinement of the heavy-atom parameters; the values of E' (r.m.s. error for anomalous differences) were set to $\frac{1}{3}$ of the E values of the corresponding angular shell.

For the 621 reflections phased in the 5 Å sphere the mean figure of merit was 0·88; the heavy-atom parameters, R values and the phasing power are shown in Table 2. The electron density map was calculated combining best phases with figures of merit and displayed on Plexiglass sheets.

Results

The Fourier map revealed unambiguously four molecules per unit cell surrounded by very low electron density regions. Several rod-shaped portions were observed within every molecule and were interpreted as helical segments; their path could be easily followed and helped in locating N and C termini. No branching or alternative course of the electron density was found; the overall shape and fold of the molecule (Fig. 1) are those reported for other respiratory hemoproteins extracted from mammals (Kendrew, Watson, Strandberg, Dickerson, Phillips & Shore, 1961) as well as from insects (Huber, Epp & Formanek, 1969) or leguminosae (Vainshtein *et al.*, 1975).

The helices, named *A*, *B*,...*H*, according to Kendrew's nomenclature, appeared to be joined by segments of somewhat weaker density. Measurements of the lengths of the segments identified as helices *A*, *B*, *E*, *F*, *G* and *H* yielded the approximate values shown in Table 3 which were calculated on the basis of a 1·5 Å shift per residue along the helix axis. No estimate was made of the span of *C* and *D* helices owing to the difficulty in evaluating short tracts of helices at this resolution. In the computation of the total helical content, this part of the molecule was considered

Table 3. *Number of residues per helix segment; figures in parentheses show the number of residues building up the polypeptide chains*

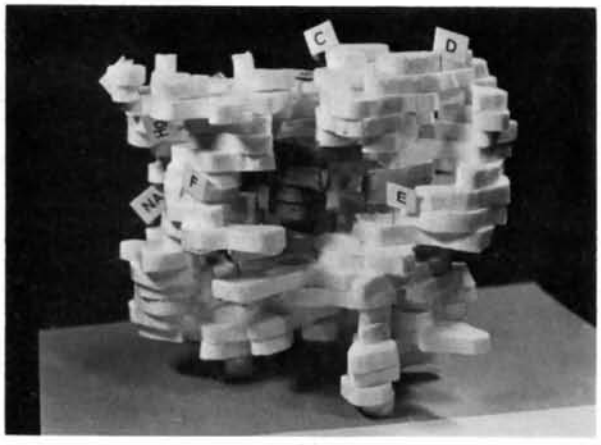
Helix	*Aplysia limacina* myoglobin (145)	Sperm-whale[a] myoglobin (153)	*Glycera dibranchiata*[b] hemoglobin (147)	*Chironomus thummi*[c] erythrocruorin (135)
A	13	16	16	16
B	14	16	16	12
E	22	20	19	21
F	11	9	12	13
G	16	19	18	19
H	17	26	22	20

References: (*a*) Perutz, Muirhead, Cox & Goaman (1968). (*b*) Padlan & Love (1968). (*c*) Huber (private communication).

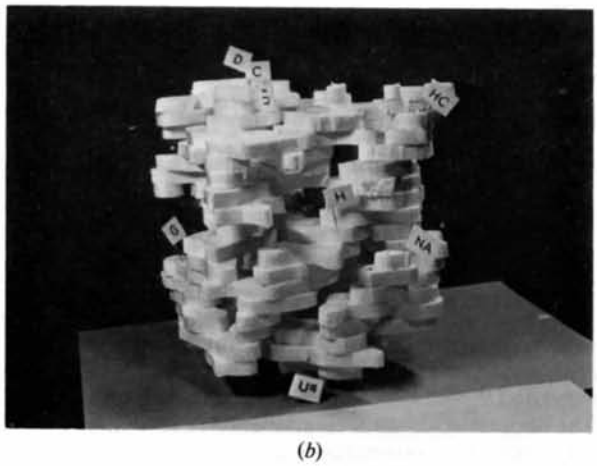
Table 2. *Heavy-atom parameters and refinement statistics*

The agreement factors are defined as: $R_C = \Sigma||F_{PH} - F_P| - |F_{Hcalc}||/\Sigma|F_{PH} - F_P|$ for the centric zones and $R_K = \Sigma|F_{PHobs} - F_{PHcalc}|/\Sigma F_{PHobs}$; $R_{FHLE} = \Sigma|F_{HLE} - F_{Hcalc}|/\Sigma F_{HLE}$ over all data, where F_{PH}, F_P and F_H are the magnitudes of the derivative, the protein and the heavy-atom structure factors. $\bar{F}_H$ and $\bar{F}_{H'}$ are the mean values of the real and imaginary parts of the calculated heavy-atom structure factor amplitudes. E and E' are the r.m.s. 'lack of closure' for isomorphous and anomalous differences for the 'best' phase angle.

Heavy-atom derivative	Site number	Occupancy	x	y	z	R_C	R_{FHLE}	R_K	Mean $\sin^2\theta/\lambda^2$ = 0·003 Å^{-2}, $\bar{F}_P$ = 405				Mean $\sin^2\theta/\lambda^2$ = 0·009 Å^{-2}, $\bar{F}_P$ = 303			
									$\bar{F}_H$	E	$\bar{F}_{H'}$	E'	$\bar{F}_H$	E	$\bar{F}_{H'}$	E'
$K_2Hg(SCN)_4$	Hg(1)	0·82	0·369	0·559	0·229	0·42		0·12	124	50			105	60		
	Hg(2)	0·24	0·360	0·690	0·200											
	Hg(3)	0·16	0·323	0·549	0·470											
$Pt(NH_3)_2Cl_2$	Pt(1)	0·55	0·235	0·430	0·142	0·49		0·11	82	54			69	50		
	Pt(2)	0·20	0·440	0·616	0·302											
	Pt(3)	0·11	0·224	0·385	0·186											
K_2HgBr_4 (I)	Hg(1)	0·97	0·357	0·544	0·216	0·44		0·14	160	68			143	60		
	Hg(2)	0·61	0·165	0·884	0·355											
K_2HgBr_4 (II)	Hg(1)	0·88	0·361	0·547	0·214	0·47	0·41	0·11	131	53	11	24	113	51	10	25
	Hg(2)	0·24	0·165	0·884	0·345											
	Hg(3)	0·15	0·352	0·410	0·234											
	Hg(4)	0·12	0·062	0·416	0·128											
$NaUO_2(CH_3COO)_3$	U(1)	0·74	0·027	0·641	0·304	0·34	0·33	0·13	164	61	19	15	150	61	21	18
	U(2)	0·70	0·184	0·545	0·476											
	U(3)	0·28	0·067	0·408	0·004											



(*a*)



(*b*)

Fig. 1. Views of the model of *Aplysia* myoglobin at 5 Å resolution. (*a*) *E* and *F* helices and the heme group are in the foreground; N terminus (NA) and one Hg binding site can be seen in the background. (*b*) N and C termini (NA and HC), *G* and *H* helices and two U binding sites are shown.

structurally homologous with *Chironomus thummi* erythrocruorin; in the *C*, *C–D*, and *D* regions in fact the two sequences share 38% of the residues (8 out of 21) and one deletion if compared to sperm-whale myoglobin (Kovàcs, Antonini, Brunori, Giacometti & Tentori, 1976).

The heme group was identified as the highest density feature of the map; it appears like a rather flattened ellipsoid located between helices *E* and *F*. At this level of resolution two weak linkages of electron density run from the *E* helix and the *C–D* corner to some of the side substituents of the heme while no significant feature pointing towards the Fe atom could be noticed.

Heavy-atom sites were found to be on the surface of the molecule and some of their positions can be seen in Fig. 1.

Even though the overall structural similarity with other myoglobins is evident, the results on *Aplysia* myoglobin reported here account for some differences which await confirmation by higher-resolution studies. The helical content of this globin is about 75%, somewhat higher than the estimate of 60% calculated on the basis of the dichroic properties in the far UV (Kovàcs *et al.*, 1976). It must be pointed out however that among the values reported in Table 3, those referring to helices *F* and *G* may be rough estimates: corners *E–F* and *F–G* are regions of low electron density and a particular secondary structure in the *E–F* region could simulate a portion of the *F* helix at this resolution. Considering that the aligned sequences of sperm-whale and *Aplysia* myoglobin show that the latter is eight residues shorter in the *C* terminal region (Kovàcs *et al.*, 1976), it is not surprising to find only 17 residues in helix *H*.

Using Protein Data Bank coordinates, we calculated an electron density map at 5 Å resolution and built a balsa-wood model of sperm-whale myoglobin. Comparison of the two models suggests that in *Aplysia* myoglobin the heme is in some way more accessible to the solvent; this seems to be particularly true along the *E* helix side, where *Aplysia* myoglobin lacks the distal histidine. A peculiar distribution of polar and apolar residues (Kovàcs *et al.*, 1976), probably combined with the different geometry of the distal position, could account for the supposed enhanced accessibility.

Quantitative evaluation of the structural analogies of *Aplysia* myoglobin with other respiratory hemoproteins is awaited in order to gain more insight into the processes of evolution; higher resolution maps will also reveal the details of the unusual heme environment. At the present level further comparisons are not possible. It is intended to continue this study with the collection of higher-resolution data.

The authors are particularly grateful to I. Tickle for helpful advice in the computational aspects of this study. Many thanks are also due to T. L. Blundell and M. Brunori who read the manuscript and offered constructive criticism.

References

ANTONINI, E. & BRUNORI, M. (1974). *Hemoglobin and Myoglobin and their Reaction with Ligands*, pp. 47–54. Amsterdam: North-Holland.

BLUNDELL, T. L., BRUNORI, M., CURTI, B., BOLOGNESI, M., CODA, A., FUMAGALLI, M. & UNGARETTI, L. (1975). *J. Mol. Biol.* **97**, 665–666.

BRUNORI, M., GIACOMETTI, G. M., ANTONINI, E. & WYMAN, J. (1972). *J. Mol. Biol.* **63**, 139–152.

DAVIES, J. E. & GATEHOUSE, B. M. (1973). *Acta Cryst.* B**29**, 1934–1942.

HUBER, R., EPP, O. & FORMANEK, H. (1969). *Naturwissenschaften*, **56**, 362–376.

KARTHA, G. & PARTHASARATHY, R. (1965). *Acta Cryst.* **18**, 749–753.

KENDREW, J. C., WATSON, H. C., STRANDBERG, B. E., DICKERSON, R. E., PHILLIPS, D. C. & SHORE, V. C. (1961). *Nature* (*London*), **190**, 666–670.

KOVÀCS, A. L., ANTONINI, E., BRUNORI, M., GIACOMETTI, G. M. & TENTORI, L. (1976). *Colloquium on Myoglobin*, edited by A. G. SCHENECK & C. VANDECASSERIE, pp. 52–74. Bruxelles: Editions de la Faculf de Science de l'Université.

MAKINEN, M. W., CHURG, A. K. & GLICK, H. A. (1978). *Proc. Natl Acad. Sci. USA*. In the press.

MATTHEWS, B. W. (1966). *Acta Cryst.* **20**, 230–239.

NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. S. (1968). *Acta Cryst.* A**24**, 351–359.

PADLAN, E. A. & LOVE, W. E. (1968). *Nature (London)*, **220**, 376–378.

PERUTZ, M. F., MUIRHEAD, H., COX, J. M. & GOAMAN, L. C. G. (1968). *Nature (London)*, **219**, 131–139.

ROSSI-FANELLI, A. & ANTONINI, E. (1957). *Biokhimya*, **22**, 335–342.

TENTORI, L., VIVALDI, G., CARTA, S., ANTONINI, E. & BRUNORI, M. (1968). *Nature (London)*, **219**, 487.

VAINSHTEIN, B. K., HARUTYUNYAN, E. H., KURANOVA, I. P., BORISOV, V. V., SOSFENOV, N. I., PAVLOVSKY, A. G., GREBENKO, A. I. & KONAREVA, N. V. (1975). *Nature (London)*, **254**, 163–164.

Acta Cryst. (1978). B**34**, 3662–3666

Nucleic-Acid Constituents. X.* The Crystal and Molecular Structure of Adenosine-5′-*O*-methylphosphate

BY J. D. HOOGENDORP, G. C. VERSCHOOR AND C. ROMERS

Gorlaeus Laboratoria, X-ray and Electron Diffraction Section, University of Leiden, Leiden, The Netherlands

(*Received* 19 *May* 1978; *accepted* 18 *July* 1978)

Adenosine-5′-*O*-methylphosphate, $C_{11}H_{16}N_5O_7P$, crystallizes in the orthorhombic system with very unusual space group *I*222. The lattice constants at 20°C are $a = 17{\cdot}244$ (8), $b = 13{\cdot}309$ (7) and $c = 15{\cdot}037$ (7) Å. There are eight mononucleotides per unit cell as well as an undefined number of methanol molecules in disordered positions. The space group and structure were determined by Patterson analysis and direct methods. The structure was refined by a least-squares procedure ($R_w = 3{\cdot}8\%$). The adenine moiety has the *anti* configuration ($\chi = 68{\cdot}6°$) with respect to the sugar. The ribose ring has an *S*-conformation [C(2′)-*endo*] with pseudorotation angle $P = 171{\cdot}7$ and $\varphi_{max} = 33{\cdot}9°$. The side-chain orientation is g^+ (*gauche–gauche*) and the backbone is folded. The unusual space group is commented on.

Introduction

This paper is a contribution to a project on synthesis, NMR and ORD studies and X-ray investigations of nucleic-acid constituents. Since the geometry and conformation of the title compound may serve as a reference for the interpretation of NMR measurements of oligonucleotides (Altona and co-workers, unpublished) the crystal structure of adenosine-5′-*O*-methylphosphate (mpA) was decided upon. The chemical synthesis of mpA will be described elsewhere (Oltshoorn, Altona & van Boom, 1979). The chemical formula and the numbering of atoms are indicated in Fig. 1(*a*).

Experimental

mpA was crystallized from a mixture of water, methanol and hydrochloric acid (pH = 3·0). Only one single crystal was found, therefore the density was not determined experimentally. The diffraction symmetry is *mmm* and reflexions *hkl* are absent for $h + k + l$ odd. Since mpA is optically active the space groups to be considered are *I*222 and $I2_12_12_1$. Taking into account the cell dimensions (Table 1) and the lack of symmetry of mpA, it is reasonable to assume that the unit cell contains eight molecules. In a later stage we located an undefined amount of methanol (less than eight

Table 1. *Crystal data at* 20°C

$a = 17{\cdot}244$ (8) Å	*I*222
$b = 13{\cdot}309$ (7)	λ(Mo $K\alpha$) = 0·71069 Å
$c = 15{\cdot}037$ (7)	μ(Mo $K\alpha$) = 2·2 cm^{-1}
1·39 < density < 1·51 g cm^{-3}	
Instrument: three-circle diffractometer, graphite-monochromatized Mo $K\alpha$ radiation, crystal mounted parallel to [010]	
Scan method: ω scan between $\theta = 4$ and $\theta = 25{\cdot}5°$	
Scanned reflexions	3074
Symmetry-independent	1645
Non-significant ($I < 2\sigma$)	167
Not observed	106
Used in analysis	1372

* Part IX: de Graaff, Martens & Romers (1978).