

Production of human embryonic haemoglobin (Gower II) in a yeast expression system

Ruth M. MOULD, Oliver M. HOFMANN and Thomas BRITTAI*^{*}

Department of Biochemistry, University of Auckland, Auckland, New Zealand

The cDNA coding for a human embryonic globin protein has been obtained from an erythroleukaemic cell line. A plasmid expression system for human embryonic haemoglobin Gower II containing cDNA copies of the appropriate pair of globin genes coupled to synthetic galactose-regulated hybrid promoters has been engineered. Transformation of *Saccharomyces cerevisiae*

with this plasmid yields a cellular system capable of high-level production of fully functional tetrameric embryonic haemoglobin. We have developed a purification scheme which gives high yields of pure human embryonic haemoglobin suitable for structural and functional studies. Preliminary characterization studies are reported.

INTRODUCTION

The process of human development can be arbitrarily divided into three temporal phases: (1) preplacental; (2) placental; (3) postnatal. Within each of these developmental phases the individual is faced with O₂-supply problems peculiar to each phase. To satisfy the different O₂-supply demands made at each stage of development, the individual synthesizes haemoglobins optimized for function during each period. Thus the human embryo synthesizes three embryonic haemoglobins (Gower I $\zeta_2\epsilon_2$, Gower II $\alpha_2\epsilon_2$ and Portland $\zeta_2\gamma_2$) during the preplacental phase, two fetal haemoglobins ($\alpha_2\gamma^A_2$ and $\alpha_2\gamma^G_2$) during the placental phase and predominantly a single haemoglobin ($\alpha_2\beta_2$) during the postnatal phase (Wood et al., 1977; Villeponteau and Martinson, 1981).

The vast number of studies on the adult human haemoglobin molecule probably make it the most thoroughly investigated protein to date. All of its functions appear to have been identified and most explained from the physiological to the atomic level (Imai, 1982; Dickerson and Geiss, 1983). Human fetal haemoglobin has been extensively studied and its major role, of scavenging oxygen from the maternal blood supply via the placenta, has been rationalized in terms of its significantly reduced interaction with intraerythrocytic 2,3-diphosphoglycerate under physiological conditions (Tyuma and Shimizu, 1970; Fermi and Perutz, 1981). In the case of the human embryonic haemoglobins, the situation is quite different. Only two extremely limited reports have appeared concerned with their function (Tuchinda et al., 1975; Huehns and Farooqui, 1975). In the face of the extreme ethical, moral and practical problems associated with the study of early human embryonic material, many workers in the past have confined themselves to studies of animal model systems (Jelkmann and Bauer, 1978; Popp et al., 1981; Purdie et al., 1983; Brittain and Wells, 1983; Brittain, 1987; Weber and Braunitzer, 1987). The appropriateness of these models has always been the source of much debate, and hence the interpretation of human embryonic haemoglobin function in terms of the animal data remains in doubt.

In the present studies we have taken the view that it is necessary to study the human proteins directly. However, we have obtained the proteins by a route that obviates the need for access to embryonic tissues. Our approach has been based on the

work of Wagenbach et al. (1991) who first reported the synthesis of fully functional adult haemoglobin from a single plasmid introduced into a yeast expression system. Their work clearly identified the utility of this approach and we have applied an analogous strategy. To these ends we have obtained cDNA corresponding to the embryonic ϵ -globin gene from an erythroleukaemic cell line (Lozzio and Lozzio, 1975; Andersson et al., 1979; Gale et al., 1979; Rutherford et al., 1981) and introduced it into an expression vector. We have introduced this plasmid into yeast for the controlled production of large quantities of functionally active embryonic haemoglobin protein. This material provides us with a starting point for the wide range of functional and structural investigations now in progress.

MATERIALS AND METHODS

Cloning of cDNA encoding globin proteins

A sample of the erythroleukaemic cell line K562 was grown to a density of 5×10^6 cells/ml using the conditions described by Rutherford et al. (1979), Testa et al. (1982) and Mattia et al. (1986). Total mRNA was obtained from the K562 cells and used to produce single-stranded cDNA using the RiboClone cDNA synthesis system and an oligo(dT) primer (Promega). The first-strand cDNA product was used as a template in PCR reactions at a dilution of 10^{-4} . Using the appropriate pair of oligonucleotide primers in a PCR reaction, the embryonic ϵ -globin gene was then amplified to yield high levels of the corresponding cDNA. Final reaction conditions were 100 mM Tris/HCl, pH 8.8, 500 mM KCl, 1.5 mM MgCl₂, 0.25 mM dNTPs, 0.01% gelatin, primers at 0.4 pM/ μ l and 0.025 units of AmpliTaq (Perkin-Elmer-Cetus). The sample was heated for 5 min at 94 °C then subjected to 25 cycles of 72 °C for 2 min and 94 °C for 1 min before a final incubation at 72 °C for 10 min.

In order to facilitate the subsequent DNA manipulations, it is necessary to produce cDNA capable of providing *Nco*I and *Sal*I cohesive ends. The *Sal*I site was created by extending the 3' terminal primer to include a *Sal*I restriction site. Unfortunately the ϵ -globin gene contains a naturally occurring internal *Nco*I restriction site. Thus the 5' terminal primer was extended to include a *Bsa*I restriction site. By appropriate positioning of the recognition site in the primer, it was possible to produce cDNA which, when treated with the *Bsa*I enzyme, yielded the necessary

* To whom correspondence should be addressed.

NcoI cohesive end, containing the ATG initiation codon of the globin gene, without the use of *NcoI* restriction endonuclease. cDNA produced by this procedure was subjected to electrophoresis on agarose in order to verify its purity and to establish that full-length copies had been synthesized by the above procedure. cDNA corresponding to the embryonic ϵ -globin gene was then blunt-end ligated into Bluescript, and the resultant plasmid used to transform *Escherichia coli* DH5 α . Transformed clones were selected and checked by restriction mapping. At this stage, both strands of positive clones were sequenced using a cycle-sequencing strategy employing site-specific primers and fluorescently tagged terminators (McBride et al., 1989). Extension products were analysed on an automated laser-activated fluorescence-emission DNA sequencer (Applied Biosystems model 373A) in order to confirm their integrity.

Construction of the haemoglobin expression plasmid

Using a pSK⁺ (EMBL sequence bank accession no. X52325)-derived plasmid [pGS189; kindly provided by Dr. G. Stetler, Somatogen; see De Llano et al. (1993)], containing an $\alpha\beta$ -globin expression cassette, we next engineered an $\alpha\epsilon$ -globin cassette. This process consisted of first liberating the α -globin gene and its associated galactose-regulated synthetic promoter from the pGS189 by digestion with restriction endonuclease *SalI*. The β -globin gene was then removed from the cut plasmid by digestion with *NcoI*, yielding a plasmid with *SalI* and *NcoI* ends. The ϵ -globin gene was isolated from Bluescript by digestion with *SalI* and *NcoI* and ligated into this plasmid. The plasmid was then recut with *SalI*, and the α -globin gene together with its associated promoter ligated back into the vector to produce pRMAE189 containing a cassette appropriate for the production of the $\alpha_2\epsilon_2$ protein. At each stage of the process, the progress of the construction was monitored by restriction mapping of the intermediate constructs.

The expression cassette for $\alpha_2\beta_2$ haemoglobin was removed from pGS389 by digestion with *NotI* and replaced by the *NotI* fragment from pRMAE189, containing the expression cassette for $\alpha_2\epsilon_2$, to yield the final expression vector pRMAE389. Before transformation, the introduced genes were excised and both strands resequenced to confirm their integrity.

Expression of embryonic haemoglobin in yeast

The expression plasmid was then used to transform yeast strain GSY112 employing the method of Hill et al. (1991). The resulting transformed yeast was grown in selective media (–uracil and –leucine) to stationary phase in order to maximize plasmid copy number within each cell. A sample of 200 ml of this stationary phase was then used to inoculate 1 litre of YP medium containing 2% ethanol as the sole carbon source. This culture was grown at 30 °C until a cell density of 2.5×10^8 cells/ml was obtained. At this point, 2% galactose was used to induce the synthesis of the haemoglobin proteins and a further 2% of galactose was added for every 7.5×10^7 cells/ml increase in cell density. After 15 h, the culture was centrifuged at 5000 *g* for 20 min. The cells were disrupted by agitation with glass beads in the presence of 10 mM Tris/HCl buffer at pH 8.0. The clear red supernatant obtained by centrifugation of the cell lysate at 10000 *g* for 20 min was then adsorbed on to a small column (2 cm \times 3 cm) of CM52 CM-cellulose at pH 6.0 in 10 mM Tris/HCl buffer and batch-eluted with 15 mM Tris/HCl buffer at pH 8.3. The haemoglobin-containing fraction was then chromatographed on a column (2 cm \times 6 cm) of Q-Sepharose using a 100 ml linear gradient from 20 mM Tris/HCl, pH 8.3, to 20 mM Tris/HCl chloride

+ 160 mM NaCl at pH 8.3. The haemoglobin-containing fraction was finally rechromatographed on a Mono S column (0.5 cm \times 5 cm) using a 50 ml linear gradient from 10 mM Bistris at pH 6.8 to 20 mM Bistris at pH 6.8 containing 200 mM NaCl. The resulting haemoglobin sample was concentrated and stored at –70 °C.

Characterization of recombinant human embryonic haemoglobin

The globin protein chains were separated by reversed-phase h.p.l.c. on a C8 micro-bore Applied Biosystems capillary column using a 40–70% acetonitrile gradient. Each of the resulting protein chains were subjected to 12 cycles of N-terminal amino acid analysis using an Applied Biosystem 470A gas phase sequencer. O₂-binding curves were obtained using a HEMOX analyser (Wagenbach et al., 1991; Adachi et al., 1992). All samples were maintained in the ferrous form by the use of the methaemoglobin-reducing system described by Hayashi et al. (1973). Absorption spectra were recorded on an Aminco DW2a spectrophotometer. The native molecular mass of the protein was determined by gel-exclusion chromatography on Superose 12 employing a Pharmacia f.p.l.c. system. The molecular mass of the protein subunits was determined by SDS/PAGE. The presence and concentration of haemoglobin in cells and solutions was determined from reduced–(reduced + CO) difference spectra.

RESULTS AND DISCUSSION

Embryonic haemoglobin production and purification

Cultured K562 cells have been shown to provide a good source of mRNA coding for the human embryonic ϵ -globin protein. Yeast cells which have been transformed with a 2 μ m plasmid-based expression system, containing the α - and ϵ -globin genes coupled to synthetic galactose-activated promoters (Figure 1), grow vigorously on YP medium utilizing ethanol as the sole

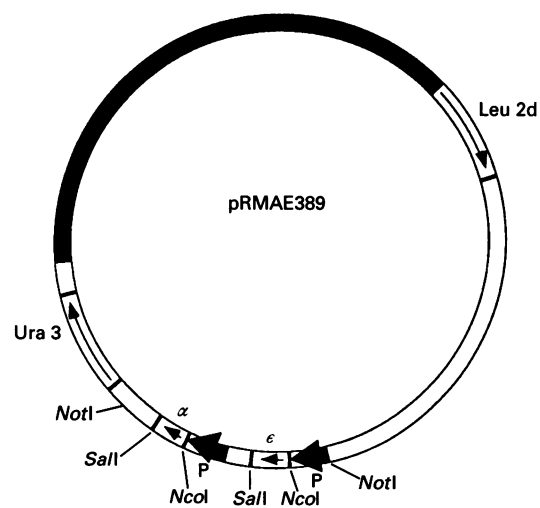


Figure 1 Structure of the human embryonic haemoglobin expression plasmid, pRMAE389

The solid line represents a full-length copy of the B-form of the 2 μ m yeast plasmid. The open line represents a copy of the full-length pBR322 plasmid. Synthetic galactose-regulated promoters (P) are coupled to the α - and ϵ -globin structural genes. Leu 2d and Ura 3 are two selectable markers for growth in yeast. The entire construct is 15.5 kb, and, except for the α - and ϵ -globin genes, represents the plasmid pGS389 (see Wagenbach et al., 1991).

carbon source. Difference spectroscopy performed on whole cell suspensions after 15 h of growth after galactose induction indicated the intracellular production of haemoglobin to a level of 20–30 mg/l of culture. Although extended growth of the cells was found to produce yields of up to 70 mg of haemoglobin/l, after 24 h of growth, traces of sulphaem were found to be present in the haemoglobin, reaching levels as high as 18% after 3 days of culture (Hofmann et al., 1993). Consequently cells used in this study were only cultured for 15 h after induction in order to avoid complications arising from the presence of sulphaem. It was found that active aeration of the yeast cultures with saturating quantities of air produced more rapid cell growth but a decrease in haemoglobin production. Addition of exogenous haem to the yeast cultures had no effect on haemoglobin synthesis. This is in direct contrast with the situation reported previously for the synthesis of human adult haemoglobin in *E. coli*, where exogenous haem is necessary for the production of fully functional proteins (Hoffman et al., 1990; Looker et al., 1993). In the absence of direct data, we can only concur with other workers (Wagenbach et al., 1990) in assuming that the yeast mitochondrial haem synthesis capacity is sufficient to provide haem for both its natural needs and to satisfy the extra demands posed by haemoglobin synthesis.

Cell disruption by agitation with glass beads produced a breakage of up to 70% after 3 min, with extended agitation leading to denaturation and precipitation of some of the haemoglobin. Early attempts to improve the efficiency of cell breakage employing buffers containing 0.1% Triton X-100 were abandoned when it was discovered that this procedure produced a crude extract which cryoprecipitated, even at 4 °C, leading to major losses in haemoglobin yields. Adsorption of the haemoglobin from the crude extract on to CM-cellulose followed by batch desorption at elevated pH yielded a haemoglobin fraction of more than 90% purity. Subsequent chromatography of this haemoglobin fraction on Q-Sepharose yielded a single symmetrical peak with a purity of more than 98%, as judged from its 280 nm/420 nm ratio. Final chromatography of the resulting haemoglobin fraction on Mono S yielded a protein sample that was essentially pure, as judged by both SDS/PAGE and its spectroscopic ratio 280 nm/420 nm.

Physicochemical characterization of recombinant human embryonic haemoglobin

The resulting protein showed bands on SDS/PAGE corresponding to molecular masses of approx. 15.5 kDa as expected for normal-length α - and ϵ -subunits. Preparative samples of the α - and ϵ -chains were obtained by reversed-phase h.p.l.c. on a C8 column. The 30–60% acetonitrile gradient containing 0.1% trifluoroacetic acid commonly used to separate α - and β -globin chains (Hoffman et al., 1990), yielded a chromatograph with a single peak coincident with that obtained for an authentic sample of α -chains. However, when the gradient was changed to a 40–70% acetonitrile gradient containing 2.5% trifluoroacetic acid, the α - and ϵ -chains were well resolved (Figure 2). The ratio of the haem/ β absorbance obtained from the h.p.l.c. traces was found to be identical with that obtained for the haem/ ϵ ratio and is thus taken to indicate the presence of four haems per tetramer of the $\alpha_2\epsilon_2$ protein. Amino acid sequencing of the ten N-terminal amino acids of each chain gave sequences identical with those predicted from the DNA sequences with no indication of N-terminal acetylation or residual N-terminal methionine residues.

Absorption spectra in the uv-visible region indicated the correct binding of haem to the globin protein as judged by the wavelength of maximum absorption and the relative absorption

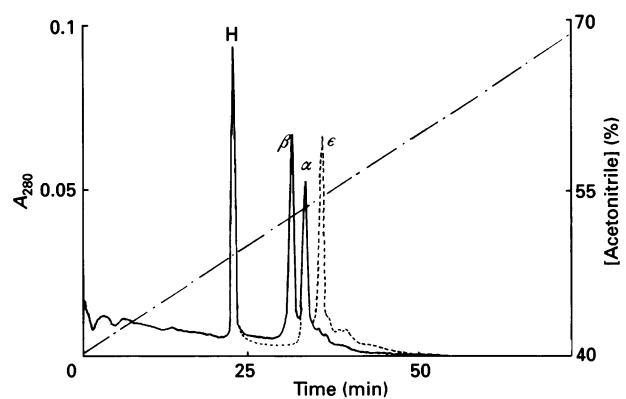


Figure 2 Globin separation by h.p.l.c.

Samples of adult human ($\alpha_2\beta_2$, —) and embryonic human Gower II ($\alpha_2\epsilon_2$, ----) were subjected to separation by h.p.l.c. as described in the text. Peaks β , α and ϵ were collected for subsequent amino acid sequencing. H represents haem. The column was monitored at 280 nm. — · —, Acetonitrile gradient.

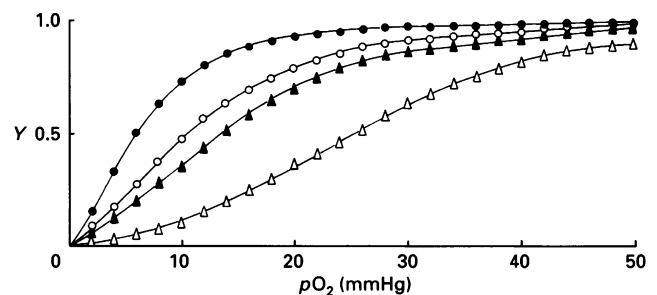


Figure 3 Equilibrium curves for the binding of O_2 to adult and embryonic human haemoglobins

Equilibrium curves are shown for the binding of O_2 to human embryonic (closed symbols) and human adult (open symbols) haemoglobins. The curves were recorded at 37 °C in 50 mM Bistris buffer containing 100 mM NaCl at pH 7.4, in the absence (circles) and presence (triangles) of 2 mM 2,3-diphosphoglycerate. Y, fractional saturation of haemoglobin with oxygen.

coefficients of the various bands. The tetrameric structure of the human embryonic haemoglobin Gower II was identified by gel-exclusion chromatography, which yielded a native molecular mass of approx. 65 kDa.

O_2 -binding curves obtained for the embryonic haemoglobin at pH 7.4 and 37 °C showed a co-operative pattern of binding with an associated $p50$ (the oxygen partial pressure necessary to half-saturate the haemoglobin) of 5.5 mmHg and a Hill coefficient of 2.2. Furthermore the embryonic haemoglobin also showed a significant response to the presence of 2,3-diphosphoglycerate, producing O_2 -binding curves with $p50$ of 13.5 mmHg and a Hill coefficient of 2.4 (Figure 3). However, whether this response to the presence of organic phosphates is of physiological significance or not is not known, as no measurements of the intracellular concentration of organic phosphates in human embryonic erythrocytes has been reported. It is interesting to note that the ϵ -globin retains the amino acids His-2, Lys-82 and His-143 which together with the N-terminal amino group provide the binding site for 2,3-diphosphoglycerate seen in the β -globin protein. This is in contrast with the later expressed fetal globin γ -chain in which His-143 is replaced by Ser with a concomitant lowering of its interaction with 2,3-diphosphoglycerate.

Table 1 Characteristics of human embryonic haemoglobin Gower II

Molecular mass (kDa): Native 65	Subunits 16		
α -Globin sequence expected:	V-L-S-P-A-D-K-T-N-V-K-A-		
α -Globin sequence found:	V-L-S-P-A-D-K-T-N-V-K-A-		
ϵ -Globin sequence expected:	V-H-F-T-A-E-E-K-A-A-V-T-		
ϵ -Globin sequence found:	V-H-F-T-A-E-E-K-A-A-V-T-		
Spectrum of carbonmonoxy haemoglobin	λ_{\max} 419 nm; 540 nm; 564 nm		
	419 nm/280 nm = 4.9		
O_2 binding			
pH 7.4:	$p50 = 5.5$ mmHg	$h = 2.2$	
pH 7.4 + 2 mM diphosphoglycerate	$p50 = 13.5$ mmHg	$h = 2.4$	

A summary of the physicochemical characteristics of human embryonic haemoglobin Gower II is presented in Table 1.

Financial support from the Health Research Council, Lottery Board, Auckland Medical Research Foundation, New Zealand is gratefully acknowledged. R.M.M. is the recipient of a N.Z.V.C.C. postdoctoral fellowship. Our thanks go to Professor J. Marbrook for providing a sample of K562 cells, to Dr. N. Birch, Dr. D. Christie and Professor R. Wells for help and advice, particularly on matters of molecular biology, and to Dr. G. Stetler, Dr. A. Mathews and Dr. J. Aitken of Somatogen for their help and in making available plasmids pGS189 and pGS389.

REFERENCES

- Adachi, K., Konitzer, P., Kim, J. and Surrey, S. (1992) *Protein Engineering* **5**, 807–810
 Andersson, L. C., Nilsson, D. and Gahrberg, C. G. (1979) *Int. J. Cancer* **23**, 143–147
 Brittain, T. (1987) *J. Exp. Biol.* **127**, 443–447
 Brittain, T. and Wells, R. M. (1983) in *Development in Mammals* (Johnson, M. H., ed.), vol. 5, pp. 135–155, Elsevier Science, Amsterdam

- De Llano, J. J. M., Schneewind, O., Stetler, G. and Manning, J. M. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 918–922
 Dickerson, R. E. and Geiss, I. (1983) *Haemoglobin*. Benjamin Cummings, Menlo Park, CA
 Fermi, G. and Perutz, M. F. (1981) in *Haemoglobin and Myoglobin Atlas of Molecular Structures in Biology*. (Phillips, D. C. and Richards, F. M., eds.), vol. 2, pp. 71–73, Clarendon Press, Oxford
 Gale, R. E., Clegg, J. B. and Huehns, E. R. (1979) *Nature (London)* **280**, 162–164
 Hayashi, A., Suzuki, T. and Shin, N. (1973) *Biochim. Biophys. Acta* **310**, 309–316
 Hoffman, S. J., Looker, D. L., Roehrich, J. M., Cozart, P. E., Durfee, S. L., Tedesco, J. L. and Stetler, G. I. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8521–8525
 Hofmann, O., Mould, R. and Brittain, T. (1993) *Protein Engineering*, in the press
 Hill, J., Donald, K. A. G. and Griffiths, D. E. (1991) *Nucleic Acids Res.* **19**, 5791
 Huehns, E. R. and Farooqui, A. M. (1975) *Nature (London)* **254**, 335–337
 Imai, K. (1982) *Allosteric Effects in Haemoglobin*, Cambridge University Press, Cambridge
 Jelkman, W. and Bauer, C. (1978) *Pflügers Arch.* **312**, 149–156
 Looker, D., Mathews, A. J., Neway, J. O. and Stetler, G. L. (1993) *Methods. Enzymol.*, in the press
 Lozzio, C. B. and Lozzio, B. B. (1975) *Blood* **45**, 321–334
 Mattia, E., Josic, D., Ashwell, G., Klausner, R. and Renswarde, J. (1986) *J. Biol. Chem.* **261**, 4587–4593
 McBride, J. K., Koepf, S. M., Gibbs, R. A., Salsler, W., Mayrand, P. E., Hunkapiller, M. W. and Kronick, M. N. (1989) *Clin. Chem.* **35**, 2196–2201
 Popp, R. A., Marsh, C. L. and Skown, C. L. (1981) *Dev. Biol.* **85**, 123–128
 Purdie, A., Wells, R. M. and Brittain, T. (1983) *Biochem. J.* **215**, 377–385
 Rutherford, T. R., Clegg, J. J. and Weatherall, D. J. (1979) *Nature (London)* **280**, 164–165
 Rutherford, T. R., Clegg, J. J., Higgs, D. R., Jones, R. W., Thomson, J. and Weatherall, D. J. (1981) *Haemoglobin in Development and Differentiation* (Stamatoyannopoulos, G. and Nienhuis, A. W., eds.), pp. 459–472, Alan R. Liss, New York
 Testa, U., Vainchenker, W., Beuzard, Y., Royer-Fessard, P., Guerrasio, A., Titeux, M., Laporte, P., Bouget, J., Breton-Gouris, J. and Rossa, J. (1982) *Eur. J. Biochem.* **121**, 649–655
 Tuchinda, S., Nagai, K. and Lehmann, H. (1975) *FEBS Lett.* **49**, 390–391
 Tyuma, I. and Shimizu, K. (1970) *Proc. Natl. Acad. Sci. U.S.A.* **29**, 1112–1114
 Villeponteau, B. and Martinson, H. (1981) *Nucleic Acids Res.* **9**, 3731–3746
 Wagenbach, M., O'Rourke, K., Vitez, L., Wiczorek, A., Hoffman, S., Durfee, S., Tedesco, J. and Stetler, G. (1991) *Bio. Technol.* **9**, 57–61
 Weber, R. E. and Braunitzer, G. (1987) *Respir. Physiol.* **69**, 347–357
 Wood, W. G., Clegg, J. J. and Weatherall, D. J. (1977) *Prog. Haematol.* **10**, 43–90