Comput. Biol. Med. Vol. 13, No. 2, pp. 89–97, 1983
Printed in Great Britain.

0010-4825/83 $3.00 + .00

TRI-AXIAL ELLIPSOIDS AS MODELS FOR MACROMOLECULES IN SOLUTION: PROCEDURES FOR NUMERICAL INVERSION OF THE SHAPE FUNCTIONS LEADING TO A STABLE UNIQUE SOLUTION

STEPHEN E. HARDING
Department of Biochemistry, University of Bristol
Medical School, Bristol BS8 1TD, U.K.

(*Received* 3 *March* 1982)

Abstract—Two FORTRAN IV algorithms are given for determining the two axial ratios of a macromolecule (as modelled by a tri-axial ellipsoid) from its hydrodynamic parameters. The first involves a simple graphical inversion procedure of the volume independent Λ and R functions but can only be applied to a restricted range of macromolecules. The second algorithm is more general and involves an R-function constrained non-linear least squares fit to birefringence decay data.

Macromolecular shape Tri-axial ellipsoid Graphical intersection method
Constrained minimization

1. INTRODUCTION

In an earlier communication [1] an algorithm was presented for evaluating the hydrodynamic parameters of a macromolecule (as modelled by a general tri-axial ellipsoid) from given values of its axial ratios. The development of an algorithm for performing the reverse operation [viz. evaluation of the two axial ratios $(a/b, b/c)$ of the ellipsoid of semiaxes $a \geqslant b \geqslant c$] is not so simple. The difficulties derive from the fact that, although explicit solutions of the hydrodynamic parameters in terms of axial ratio are now available, analytic inversion procedures are not.

The principles of two numerical inversion procedures have been outlined before [2, 3]. A given hydrodynamic parameter has a line solution of the two axial ratios $(a/b, b/c)$ characterizing a general ellipsoid of semi-axes $a \geqslant b \geqslant c$. Hence if two or more are combined graphically a unique solution may be found from the intersection. The parameters would thus be chosen according to how 'orthogonal' their intersection was, their experimental measurability and their sensitivity. These criteria however proved very restrictive and the only generally applicable method was one involving intersection of the R function (derived from viscosity and sedimentation measurements) with $\delta_{\pm}$ functions (derived from viscosity and the two electric birefringence decay constants, θ_+ θ_-) [2, 4]. Even here, hitherto available methods for extraction of the two decay constants from a two-term exponential decay were impossible with data of current experimental precision. However, it was shown that a new method, where information from the R function (which can be relatively easily determined) is placed in the analysis for the extraction of the decay constants as a constraint provided adequate resolution for data of current experimental precision. We now present the FORTRAN IV algorithm on which this method is based.

The alternative method [3] is to combine R graphically with the Λ function (determined from viscosity and harmonic mean relaxation data). This method is not generally applicable to all macromolecules but only to those asymmetric enough (one axial ratio $\gtrsim 3$) so that the Λ function is sufficiently sensitive but not *so* asymmetric that the measured harmonic mean relaxation time is not affected by internal rotations within the macromolecule. However, because this method is simpler we present its algorithm briefly first.

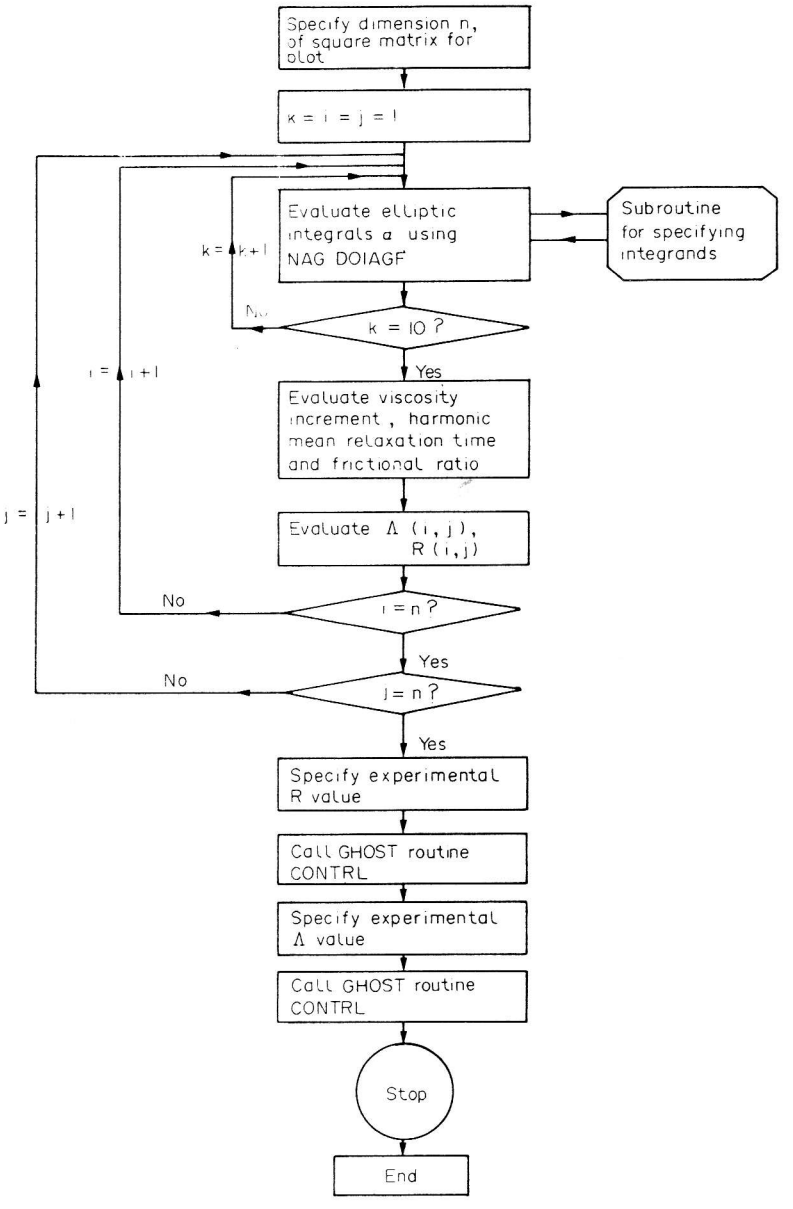

Fig. 1. Flow chart for the $\Lambda - R$ algorithm.

2. A SIMPLE GRAPHICAL INVERSION PROCEDURE

The essence of the procedure is described in the flow chart of Fig. 1. The value of Λ or R for any specified value of the pair of axial ratios $(a/b, b/c)$ can be evaluated numerically [1] using a standard numerical package for solving the elliptic integrals. Arrays of such values are evaluated in the $(a/b, b/c)$ plane (a 21 $\times$ 21 matrix takes approximately 1000 s of computer time in a CDC cyber 72). A contour plotting routine ('CONTRL' in the CDC GHOST graphics system) interpolates between these matrix points and can plot the Λ and R functions in the $(a/b, b/c)$ plane.

3. AN R-CONSTRAINED LEAST SQUARES ALGORITHM

This algorithm is more complicated and it will be instructive to present it in some detail. The problem revolves around extracting the two decay constants for an asymmetric particle

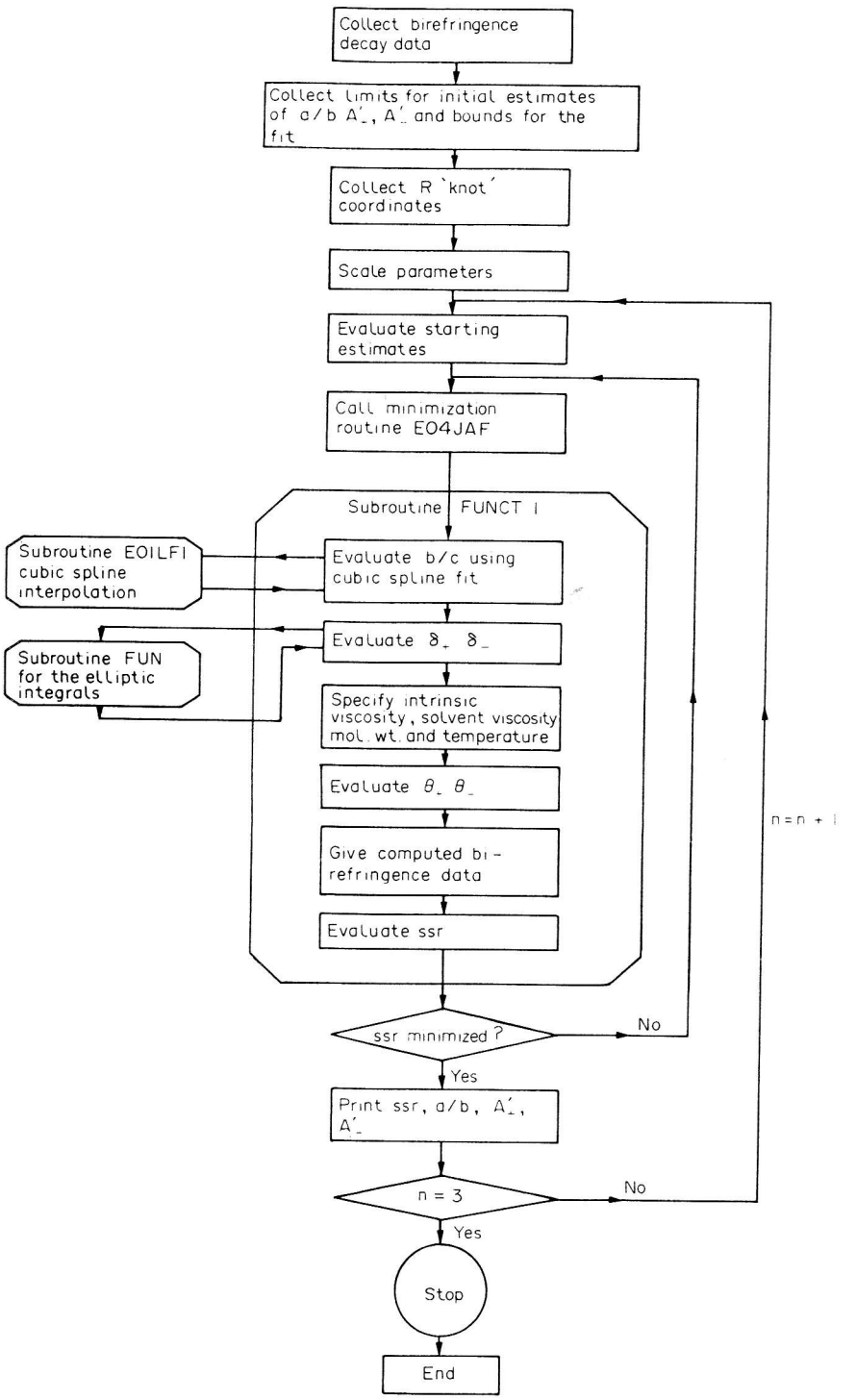

Fig. 2. Flow chart for the R-constrained least squares algorithm.

from an electric birefringence decay curve. The resolution of exponentials is notoriously difficult even for very precise data [5, 6] — this is clearly illustrated in Fig. 1 of Small and Isenberg [7]. A consideration of two recent papers by Jost and O'Konski [5], and O'Connor *et al.* [6] reveals that, for a two-term exponential decay a non-linear least squares method is the most applicable. Even so, the requirements on the precision of the experimental birefringence data are about two orders of magnitude greater than the current limits [2, 8].

The procedure introduced before [2] was to include information from the parameter R as a

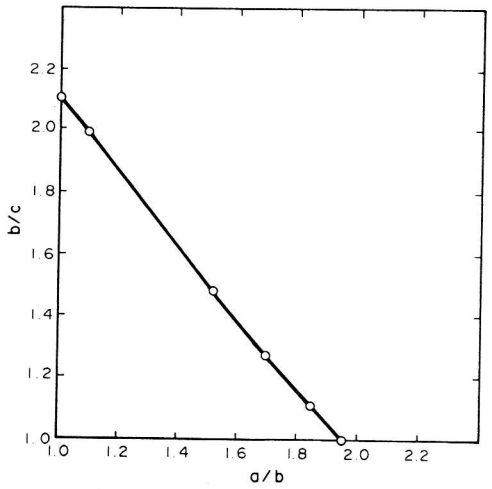

Fig. 3. Plot of the R-function for a hypothetical protein of axial ratio $(a/b, b/c) = (1.5, 1.5)$. The coordinates of the encircled points are used for cubic spline interpolation in the R-constrained least squares algorithm.

constraint in the analyses. In this way the problem is effectively reduced from four independent variables (two pre-exponential factors $A'_{\pm}$ and two decay constants $\theta_{\pm}$) to three $(A'_{\pm}, a/b)$ since a/b should now specify a unique value for b/c, and hence $\theta_{\pm}$.

4. COMPUTATIONAL METHOD

An outline of the procedure is given in the flow chart of Fig. 2. The R function is assumed to have already been measured from the ratio of the sedimentation concentration regression coefficient to the intrinsic viscosity. It is then plotted in the $(a/b, b/c)$ plane over a suitable range of axial ratio (see Section 2 above). The coordinates of 'knots' in the curve (Fig. 3) can then be specified in our routine (or alternatively if the facility is available the whole R-curve can be digitized). Other preliminary data required for converting a given estimate for $(a/b, A'_+, A'_-)$ into decay constants through the $\delta_{\pm}$ functions are the intrinsic viscosity, solvent viscosity, molecular weight and temperature of the birefringence experiment.

The program (Table 1, with key to computer symbols given in Table 2) employs the NAG quasi-Newtonian algorithm E04JAF [9] for minimising a function F, subject to bounds on the estimates for the independent variables. The subroutine FUNCTI (Table 3) specifies F as the sum of the squares of the residuals for each current estimate of $(a/b, A'_+, A'_-)$: the value of the axial ratio b/c corresponding to a given estimate for a/b is first evaluated with the aid of the cubic spline interpolation routine E01LF1† applied to the R 'knot' points introduced earlier. The value for $(a/b, b/c)$ specifies a unique value for the swelling independent hydrodynamic shape functions $\delta_{\pm}$, which can be evaluated with the aid of the NAG numerical integral routine D01AGF. Using the values for the intrinsic viscosity, molecular weight, solvent viscosity and temperature of the birefringence experiment entered by the programmer, the routine will then evaluate the birefringence decay constants $\theta_{\pm}$ from equation (4) of Harding and Rowe [2]. Hence the computer estimate for the decay curve

$$\Delta n = A'_+ \, e^{-6\theta_+ t} + A'_- \, e^{-6\theta_- t}$$

is specified, and hence the sum of the squares of the residuals (ssr).

†This package from the University of Leicester Computer Centre will not be given here since the NAG package E01ADF will perform the same interpolation.

Table 1. Main program for the *R*-constrained algorithm. The sections denoted || correspond to the generation of synthetic data (cf. [2]).

```
C ***********
C  PROTEIN 1
C ***********
C LINEAR TIME INCREASE, 100PTS.
C CUT OFF TIME: 100NS
C STREAM OF RANDOM NUMBERS: 3
C 0.1 DEG STANDARD ERROR ON EACH DATA PT.
C
      PROGRAM MAIN(INPUT,OUTPUT,TAPE2=INPUT,TAPE3=OUTPUT)
      COMMON/PARAM/GAMMA(101),T(101),A,C,D,NN,AA(6),AD(5)
      REAL R(101)
      REAL AA1,AA2,AA3,AA4,AA5,AA6,AD1,AD2,AD3,AD4,AD5,AD6
      REAL BL(3),BU(3),X(3),F,W(39)
      INTEGER IBOUND,IFAIL,J,LIW,LW,N
      INTEGER IW(5)
      REAL Q1,Q2,Q3,Q4
      WRITE(3,38)
   38 FORMAT(' '
     + 100NS, 100PTS')
      WRITE(3,39)
   39 FORMAT(' '
     +STREAM 3')
      P=-1.1
      DO21 I=1,101
      P=P+1.1
      R(I)=P
      T(I)=R(I)*1.0E-09
      R(I)=T(I)*1.0E09
   21 CONTINUE
C G05BAF(0.X) SPECIFIES THE XTH STREAM OF THE RANDOM NUMBERS
      CALL G05BAF(0.3)
      DO50 I=1,101
C CALCULATE THE UNPERTURBED DECAY CURVE
      GAMMA(I)=0.07*EXP(-T(I)*5.0*5.815383E06)+
     +0.05*EXP(-T(I)*6.0*4.1564612E06)
C NOW PERTURB EACH OF THE 100 DATA POINTS USING A NORMAL PSEUDO RANDOM
C NUMBER GENERATOR G05ADF
      GAMMA(I)=GAMMA(I)+((1.74555E-03)*G05ADF(Y))
   50 CONTINUE
C READ IN THE VALUES FOR THE LOWER LIMITS FOR THE INITIAL GUESSES OF THE
C A/B, A+, A-, THE LOWER (L) AND UPPER (U) LIMITS FOR THE COMPUTER
C ESTIMATES, THE COORDINATES OF THE A/B KNOT VALUES AND THEN THE
C COORDINATES OF THE B/C VALUES FOR THE R-CURVE TO WHICH THESE
C ESTIMATES ARE CONSTRAINED
      READ(2,*)XX1,XX2,XX3,BBL1,BBL2,BBL3,BBU1,BBU2,BBU3,
     +AA1,AA2,AA3,AA4,AA5,AA6,AD1,AD2,AD3,AD4,AD5,AD6
      AA(1)=AA1
      AA(2)=AA2
      AA(3)=AA3
      AA(4)=AA4
      AA(5)=AA5
      AA(6)=AA6
      AD(1)=AD1
      AD(2)=AD2
      AD(3)=AD3
      AD(4)=AD4
      AD(5)=AD5
      AD(6)=AD6
      N=3
      IBOUND=0
      Q1=1.0
C SCALE UP A+ AND A- (SEE NAG MANUAL)
      X(2)=X(2)*100.0
      X(3)=X(3)*100.0
      DO10 I=1,3
C GENERATE STARTING ESTIMATES FOR A/B, A+ AND A-
      Z=G05AAF(X)
      X(1)=XX1+0.2*Z
      X(2)=XX2+2.0*Z
      X(3)=XX3+2.0*Z
      BL(1)=BBL1
      BL(2)=BBL2
      BL(3)=BBL3
      BU(1)=BBU1
      BU(2)=BBU2
      BU(3)=BBU3
      LIW=5
      LW=39
      IFAIL=1
      CALL E04JAF(N,IBOUND,BL,BU,X,F,IW,LIW,W,LW,IFAIL)
      IF(IFAIL.NE.0)WRITE(3,100)IFAIL
      IF(IFAIL.EQ.1)GOTO 20
  100 FORMAT("ERROR EXIT TYPE",I3,"SEE ROUTINE MANUAL")
C SCALE BACK DOWN AGAIN A+ AND A-
      X(2)=X(2)*.01
      X(3)=X(3)*.01
      WRITE(3,110)F
```

```
      WRITE(3,120)(X(1),X(2),X(3))
  110 FORMAT( " FUNCTION VALUE ON EXIT IS",F15.12)
  120 FORMAT( " A/B= ",F8.5," A+=",F15.12,
     +"  A-=",F15.12)
      IF(Q1.LE.F)GOTO 30
      Q1=F
      Q2=X(1)
      Q3=X(2)
      Q4=X(3)
   30 CONTINUE
      WRITE(3,130)
  130 FORMAT( " ")
      WRITE(3,140)
  140 FORMAT( " ")
   10 CONTINUE
      WRITE(3,150)Q1
  150 FORMAT( " BEST LEAST SQUARES VALUE = ",F15.12)
      WRITE(3,160)Q2
  160 FORMAT( "                        A / B = ",F8.5)
      WRITE(3,170)Q3
  170 FORMAT( "                        B / C = ",F8.5)
      WRITE(3,180)Q3
  180 FORMAT( "                          A+ = ",F7.4)
      WRITE(3,190)Q4
  190 FORMAT( "                          A- = ",F7.4)
   20 STOP
      END
```

```
Program data: 1.4,.06,.04,1.3,.05,.03,1.7,.09,.07,1.0,1.0970,1.5186,
              1.6958,1.8458,1.9520,2.1073,1.9973,1.4820,1.2737,
              1.1119,1.0
```

5. RESULTS

In the main program (Table 1) we have incorporated a method for generating our own synthetic birefringence curve of mock random standard absolute error of 0.1 degrees, (within the limits of current experimental precision [8]) on each data point.

Owing to the possible danger of the routine falling into subsidiary minima [5] it is necessary to repeat the processing for several different initial starting estimates for the three independent variables. Provision for this is made in the program: the user enters the lower limit for the three parameters and the routine uses NAG pseudo-random number generator G05AAX to produce the initial estimates within reasonable limits.

Table 2. Key to computer symbols used*

Computer Symbol	Parameter
A, A/B	a/b
B	b
B/C	b/c
C	c,
A_+	A'_+
A_-	A'_-
GAMMA	Birefringence (radians)
†$X(1)$	Starting estimate for a/b
†$X(2)$	Starting estimate for A'_+
†$X(3)$	Starting estimate for A'_-
$XC(1)$	Current estimate for a/b
D	Value of b/c corresponding to this estimate
S	Value of ν corresponding to this estimate
V	Value of δ_+ corresponding to this estimate
W	Value of δ_- corresponding to this estimate
THPLUS	Value of θ_+ corresponding to this estimate
THMNUS	Value of θ_- corresponding to this estimate
$XC(2)$	Current estimate for A'_+
$XC(3)$	Current estimate for A'_-
AA1—AA6	a/b coordinates for 6 pts on the R curve
AD1—AD6	b/c coordinates for 6 pts on the R curve
Q	square of a residual
FC	sum of squares of the residuals (ssr)

* Other symbols are defined in the NAG manual [9] or ref [1].
† On exit, X (1), X (2), X (3) contains the best estimates for these parameters.

Table 3. Subroutine for evaluating the sum of the squares of the residuals for a given estimate of a/b, A'_+ and A'_-

```
C SUBROUTINE FOR CALCULATING THE SUM OF THE SQUARES OF THE RESIDUALS FOR
C CURRENT ESTIMATES OF A/B, A+ AND A-
C KB = BOLTZMANNS CONSTANT, NA = AVOGADROS NUMBER
C
      SUBROUTINE FUNCT1(N,XC,FC)
      COMMON/PARAM/GAMMA(101),T(101),A,C,D,NN,AA(6),AD(5)
      EXTERNAL FUN
      REAL Q,FC
      INTEGER N
      DIMENSION ALPHA(9)
      REAL XC(N)
      REAL X2,X3
      REAL THPLUS,THMNUS,TEMP,ETA,M,O,P,KB,NA,MW
      NN=0
      A=XC(1)
      NM=6
      IEXIT=0
C CALL LIBRARY ROUTINE FOR THE CUBIC POLYNOMIAL FIT TO THE R-CURVE POINTS
C A LISTING OF THIS IS GIVEN BELOW
      CALL E01LF1(NM,AA,AD,A,D,IEXIT)
      C=1.0/D
      B=1.0
      DO45 K=1,9
      AZ=0.0
      BZ=10000000
      NN=NN+1
      MAXDIV=50
      EPS=1.0E-08
      ACC=0.0
      IFAIL=0
      CALL D01AGF(AZ,BZ,FUN,MAXDIV,EPS,ACC,ANS,ERROR,NOFUN,IFAIL)
      ALPHA(NN)=ANS
   45 CONTINUE
      S=(1.0/(A*B*C))*((4.0/15.0)*((ALPHA(7)+ALPHA(8)
     + +ALPHA(9))/(ALPHA(8)*ALPHA(9)+ALPHA(9)*ALPHA
     + (7)+ALPHA(7)*ALPHA(8)))+(1.0/5.0)*(((ALPHA(2)
     + +ALPHA(3))/(ALPHA(4)*(B*B*ALPHA(2)+C*C*ALPHA(3)
     + )))+((ALPHA(3)+ALPHA(1))/(ALPHA(5)*(C*C*ALPHA(3)
     + +A*A*ALPHA(1))))+((ALPHA(1)+ALPHA(2))/(ALPHA(6)
     + *(A*A*ALPHA(1)+B*B*ALPHA(2))))))
      M=(B*B+C*C)/(B*B*ALPHA(2)+C*C*ALPHA(3))
      O=(C*C+A*A)/(C*C*ALPHA(3)+A*A*ALPHA(1))
      P=(A*A+B*B)/(A*A*ALPHA(1)+B*B*ALPHA(2))
      Z=((A*B*C)/(12.0))*(((1.0/M)+(1.0/O)+(1.0/P
     + ))-(((1.0/M**2.0)+(1.0/O**2.0)+(1.0/P**2.0))-
     + ((1.0/(M*O))+(1.0/(O*P))+(1.0/(P*M
     + ))))**0.5))
      U=((A*B*C)/(12.0))*(((1.0/M)+(1.0/O)+(1.0/P
     + ))+(((1.0/M**2.0)+(1.0/O**2.0)+(1.0/P**2.0))-
     + ((1.0/(M*O))+(1.0/(O*P))+(1.0/(P*M
     + ))))**0.5))
      V=6.0*U*S
      W=6.0*Z*S
C     ENTER INTRINSIC VISCOSITY
      ETA=2.746
C     ENTER TEMPERATURE
      TEMP=293.0
C     ENTER MOLECULAR WEIGHT
      MW=71744.0
      KB=1.38046E-16
      NA=6.0248E23
C     ENTER SOLVENT VISCOSITY
      SOLV=.01
      THPLUS=(NA*KB/(SOLV*6.0))*(TEMP/(ETA*MW))*V
      THMNUS=(NA*KB/(SOLV*6.0))*(TEMP/(ETA*MW))*W
      X2=XC(2)
      X3=XC(3)
      FC=0.0
      DO75 I=1,101
      Q=(GAMMA(I)-(0.01*X2*EXP(-6.0*THPLUS*T(I))+0.01*X3*EXP
     + (-6.0*THMNUS*T(I))))**2.0
      FC=FC+Q
   75 CONTINUE
      RETURN
      END
```

Despite the need for these precautions it is evident from a simulation for a typical globular protein of true axial ratios (a/b, b/c) = (1.5, 1.5) that the algorithm is extremely stable (Table 5).

The algorithm has also been tested for the effect of error in the intrinsic viscosity and molecular weight, for different true axial ratios and for different true pre-exponential factors and in every case remains stable. In order to achieve the highest accuracy with the method it is

Table 4. Subroutine for specifying the integrands in the numerical evaluation of the elliptic integrals (cf. [1])

```
C SUBROUTINE FOR CALCULATING THE ELLIPTIC INTEGRALS USED FOR DETERMINING
C THE S VALUE FROM THE CURRENT ESTIMATE FOR A/B
C
      REAL FUNCTION FUN(Y)
      COMMON/PARAM/GAMMA(101),T(101),A,C,D,NN,AA(6),AD(5)
      B=1.0
      GOTO(10,20,30,40,50,60,70,80,90),NN
   10 FUN=1.0/((A*A+Y)**1.5*(B*B+Y)**0.5*(C*C+Y)**0.5)
      RETURN
   20 FUN=1.0/((A*A+Y)**0.5*(B*B+Y)**1.5*(C*C+Y)**0.5)
      RETURN
   30 FUN=1.0/((A*A+Y)**0.5*(B*B+Y)**0.5*(C*C+Y)**1.5)
      RETURN
   40 FUN=1.0/((A*A+Y)**0.5*(B*B+Y)**1.5*(C*C+Y)**1.5)
      RETURN
   50 FUN=1.0/((A*A+Y)**1.5*(B*B+Y)**0.5*(C*C+Y)**1.5)
      RETURN
   60 FUN=1.0/((A*A+Y)**1.5*(B*B+Y)**1.5*(C*C+Y)**0.5)
      RETURN
   70 FUN=Y/((A*A+Y)**0.5*(B*B+Y)**1.5*(C*C+Y)**1.5)
      RETURN
   80 FUN=Y/((A*A+Y)**1.5*(B*B+Y)**0.5*(C*C+Y)**1.5)
      RETURN
   90 FUN=Y/((A*A+Y)**1.5*(B*B+Y)**1.5*(C*C+Y)**0.5)
      RETURN
      END
```

necessary to develop procedures for determining the optimum cut-off time for the birefringence decay and also to allow for the concentration dependence of the decay constants; these have already been outlined [2].

SUMMARY

Two methods are now available for modelling biological macromolecules in solution as tri-axial ellipsoids. Both involve numerical inversion of swelling-independent hydrodynamic shape functions followed by graphical intersection of two or more of these. The necessary computational procedures on which these methods are based are presented here. The first involves data from fluorescence depolarization, viscosity and sedimentation velocity but is applicable only to a restricted class of particle asymmetries. The second involves data from electric birefringence, viscosity and sedimentation velocity, and although more generally applicable, requires constraints in the minimization procedure for extracting the relevant birefringenece decay parameters.

Acknowledgements—The author is grateful to the Lister Institute of Preventitive Medicine for financial assistance during part of this study.

REFERENCES

1. S. E. Harding, A computer program for evaluating the hydrodynamic parameters of a macromolecule in solution for any given value of its axial dimensions, *Comput. Biol. Med* **12**, 75–80 (1982).
2. S. E. Harding and A. J. Rowe, Modelling biological macromolecules in solution II: the general tri-axial ellipsoid, *Biopolymers* (in press).

Table 5. Results for a synthetic protein (cf. [2]) of true axial ratio ($a/b, b/c$) (1.5, 1.5), true A'_+ 0.07 and true A'_- 0.05 rad

```
PROTEIN 1,  0.1 DEG. ABSOLUTE ERROR
                                                    110NS, 100PTS
                                                    STREAM3

FUNCTION VALUE ON EXIT IS  .000399962998
A/B=  1.48309  A+=  .070458728408  A-=  .048831535604

FUNCTION VALUE ON EXIT IS  .000399962998
A/B=  1.48309  A+=  .070458845275  A-=  .048831534073

FUNCTION VALUE ON EXIT IS  .00399962998
A/B=  1.48309  A+=  .070458881237  A-=  .048831534081

BEST LEAST SQUARES VALUE =   .000399962998
                   A / B =  1.48309
                      A+ =  .0707
                      A-=  .0488
```

3. S. E. Harding and A. J. Rowe, Modelling biological macromolecules in solution III: the $\Lambda - R$ intersection method for tri-axial ellipsoids, *Int. J. Biol. Macromol.* **4**, 357–361 (1982).
4. S. E. Harding Modelling biological macromolecules in solution: the general triaxial ellipsoid. Ph.D. thesis, University of Leicester (1980).
5. J. W. Jost, and C. T. O'Konski, Electro-optic data aquisition and processing. *Mol. Electro-Optics* **2**, 529–564 (1978).
6. D. V. O'Connor, W. R. Ware and J. C. Andre, Deconvolution of fluorescence decay curves: a critical comparison of techniques, *J. Phys. Chem* **83**, 1333–1343 (1979).
7. I. Isenberg, R. D. Dyson and R. Hanson, Studies on the analysis of fluorescence decay data by the method of moments, *Biophys. J.* **13**, 1090–1115 (1973).
8. B. Jennings and V. Morris (private communication).
9. NAG FORTRAN Library Manual Mk 7 Vol. 1. Numerical Algorithms Group, Oxford (1978).

About the Author—Dr Harding graduated in Physics and Molecular Biophysics from Oxford in 1976. In 1977 he worked at the Atomic Energy Research Establishment, Harwell, Oxfordshire, receiving an M.Sc. for his work on the improvement in resistance to wear of ion implanted materials. From 1977 to 1980 he researched on modelling macromolecules in solution at the Department of Biochemistry, Leicester University, from where he received his Ph.D. He is now working at the University of Bristol on physico-chemical studies of bronchial glycoproteins.