(2)

SECURITY CLASSIFICATION OF THIS PAGE

DTIC ELECTE MAY 31 1990 S D

AD-A222 345

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION: (U)	1b N/A
2a SECURITY CLASSIFICATION AUTHORITY: N/A	3 DISTRIBUTION/AVAILABILITY OF REPORT: Distribution unlimited
2b DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A	
4 PERFORMING ORGANIZATION REPORT NUMBER(S): University of Maryland, College Park	5 MONITORING ORGANIZATION REPORT NUMBER(S): N/A
6a NAME OF PERFORMING ORGANIZATION: University of Maryland; 6b OFFICE SYMBOL (If applicable): N/A	7a NAME OF MONITORING ORGANIZATION: Office of Naval Research
6c ADDRESS (City, State, and ZIP Code): Department of Chemistry & Biochemistry, College Park, Maryland 20742	7b ADDRESS (City, State, and ZIP Code): 800 N. Quincy Street, Arlington, Virginia 22217-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION: Office of Naval Research; 8b OFFICE SYMBOL (If applicable): ONR	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-88-K-0323
8c ADDRESS (City, State, and ZIP Code): 800 North Quincy Street, Arlington, Virginia 22217-5000	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 61153N; PROJECT NO. 1141MB; TASK NO 441N009-01; WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
Macromolecular Calculations for the XTAL-System of Crystallographic Programs

12 PERSONAL AUTHOR(S)
Stewart, James M.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Annual	FROM 6/1/89 TO 5/31/90	1990, May, 24	4

16 SUPPLEMENTARY NOTATION
Prepared in collaboration with K.B. Ward and D.M. Collins at the Naval Research Laboratory (Code 6030). See attached summary and preprints.

17 COSATI CODES		
FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS (~~Continue on reverse if necessary and identify by block number~~)
Macromolecular crystallography, Protein crystallography, Crystallographic computing, Computing for molecular biology

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

During the last year effort has been concentrated on the completion of two programs to carry out least-squares refinement of macromolecules which preserve canonical stereochemistry. In addition, two other programs were completed, documented and placed in the XTAL system. In the coming year testing, debugging, and documentation will continue and a new program for crystallographic phase determination by maximum entropy methods will be written and tested. In addition three new links which should speed up the least-squares refinement process by approximately two orders of magnitude will be developed.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	(U)	
22a NAME OF RESPONSIBLE INDIVIDUAL: Dr. Michael Marron	22b TELEPHONE (Include Area Code): 202-696-4760	22c OFFICE SYMBOL: ONR

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

DATE: 22 May 1990

ANNUAL REPORT ON CONTRACT ONR N00014-88-K-0323

PRINCIPAL INVESTIGATOR: James M. Stewart

CONTRACTOR: University of Maryland - Code 07419

CONTRACT TITLE: Macromolecular Calculations for the XTAL System of Crystallographic Computer Programs

CONTRACT PERIOD: 1988 June 1 through 1991 May 31

RESEARCH OBJECTIVE: To produce, within the XTAL system of crystallographic computer programs, codes, documentation and tests for macromolecular crystallographic calculations.

This project is being carried out in collaboration with K. B. Ward and D. M. Collins of the Naval Research Laboratory (Code 6030).

During the last year, the implementation and documentation of least-squares refinement programs for macromolecular crystals has been completed. The documentation which has been generated for programs completed during the 1989/1990 year is included in the XTAL version 3.0 manual which will be distributed in September 1990. Preprints are attached to this report. The following brief description indicates the nature of the six programs which have been completed and documented and for which a simulated data test case has been prepared:

PROTIN: A program to read "constraint information" to describe the elements of a macromolecular crystal structure. This information, for example, can be the idealized coordinates for all the amino acids. This program is a translation of the Hendrickson-Konnert program by the same name. The purpose of this program when used in conjunction with PROLSQ is to preserve canonical stereochemistry.

PROLSQ: A program to do constrained least-squares refinement of the positional and thermal parameters of a macromolecular structure. This program is a translation into XTAL of the Konnert-Hendrickson program by the same name.

Two small service programs to support ongoing structure work were also written and documented, PHACMP and REFOUT. PHACMP allows a user to compare one reflection phase set with another on a statistical basis. REFOUT allows the user to transfer reflection information from the XTAL format to another in order to communicate with other programs available.

In addition to the programs PROTIN and PROLSQ prepared under this

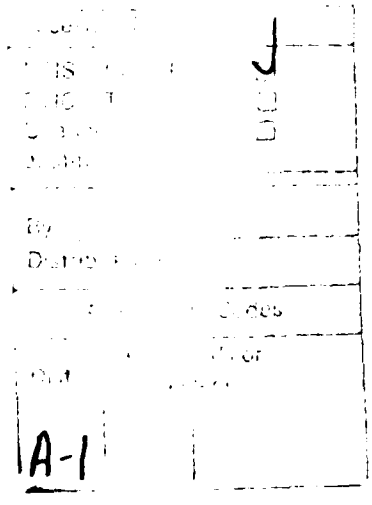
ONR grant, two other programs for XTAL have been received for testing and distribution from Dr. Keith Watenpaugh of The Upjohn Co. in Kalamazoo MI. As a part of a cooperative effort Dr. Watenpaugh has produced the programs PRECED and CEDAR which will refine macromolecules by a different algorithm than that used in PROTIN/PROLSQ. Whereas PROTIN/PROLSQ constrains atomic parameters based on geometric considerations, PRECED constrains based on intermolecular repulsion and attraction. Thus the two methods supplement one another.

Dr Watenpaugh also supplied several other service programs for inclusion which will not be described here.

Work is continuing on test cases using ideal simulation data for the purpose of verifying the correctness of the programs written for macromolecular structure determination.

Work has begun on a program MESIGN which uses a maximum entropy algorithm for the ab-initio determination of crystallographic phases. This algorithm is based on the work of D. Collins at the Naval Research Laboratory and E. Prince at the National Institute for Science and Technology. This program shows great promise and should, if the method proves applicable, have a profound impact on protein crystal structure analysis.

The following preprints for the programs PROTIN, PROLSQ, PHACMP, and REFOUT will be included in the XTAL 3.0 system User's.Manual by S.R. Hall and J.M. Stewart. This edition of the XTAL system is being distributed through Dr. S. R. Hall, Crystallography Centre, University of Western Australia, Nedlands 6009.



PROTIN: Set Constraints for Refinement of Atomic Parameters

Wayne A. Hendrickson: Department of Biochemistry/Biophysics
Columbia University, New York, NY 10032

Translated to the XTAL system, with permission, by
James M. Stewart: Department of Chemistry/Biochemistry
University of Maryland, College Park, MD 20742

SYNOPSIS

PROTIN combines the atomic parameters of the trial structure with canonical stereochemical information and sets up the constraint observational terms for PROLSQ the restrained-parameter structure-factor least-squares atomic parameter refinement program for large molecules.
This XTAL version is derived from the program of John H. Konnert (1976) which was elaborated for protein refinement by Wyane A. Hendrickson (1985). Since that time a number of others have made changes, in details, to the original code and documentation. Among these contributors are Barry Finzel, Steven Sheriff, Anita Sielecki, Janet Smith, & Alex Wlodawer.

PURPOSE

The pair of programs PROTIN and PROLSQ are designed to allow the refinement of the atomic parameters obtained from crystallographic diffraction data while restraining the parameters to conform to stereochemical information. These programs use stereochemical information as additional observations which are incorporated with the observations of reflection intensities so that the refinement of atomic parameters includes information in the observational equations about both stereochemistry and diffraction structure factors.

In the XTAL system PROATM is used to load the atomic parameters, identifying each atom in terms of its type and residue or group. PROTIN is used to load the canonical stereochemical information and connect the attributes of the ideal atoms to atoms of correspond[illegible] type and group in the given structure. Once the attachments are made the quantities for the structure being refined which are needed to cont[illegible] to the matrix of normal equations are written to a file for use by the PROLSQ program. PROLSQ takes in this information and the diffraction information and forms and solves the combined least-squares normal equations for shifts in the atomic parameters.

Before using PROTIN/PROLSQ the Henderson paper (1985) should be read. Especially important are pages 264-268 where the "Refinement Strategy" is laid out.

METHOD

The method used in PROTIN/PROLSQ is documented in publications by Konnert and Hendrickson (1976,1979,1980,1985).
From the standpoint of use of these programs Hendrickson (1985) presents

an overview of the PROTIN-PROLSQ programs which are the ones which have been translated into RATMAC and placed in the XTAL system. A summary of this paper follows in which the input lines set forth below are related to the terms used in that paper.
In general the geometric data must be placed in the input stream in a very restricted order. Therefore, the input lines are grouped according to the kinds of stereochemical features which may be restrained.

The first lines PROTIN, SPCPEP, and ALTCHN set global conditions for the structure. *SPCPEP, and ALTCHN are optional if the protein has well* defined chains which have been specified in the ATOM lines loaded by PROATM. The PROATM program marks, through the lrsequ: record of the binary data file the N terminal and C terminal group of each chain. The chain identity and terminal residue information is then generated in PROTIN from the signals set by PROATM.

DISULB is optional if no disulfide linkages in the structure are to be restrained.

The optional lines from *INTRAD* to *SYMWGT* or the *END* line supplied at this point causes all the from RESTRM through ELLRES, to be read from the "default" file, PRORES, which is supplied with the XTAL system. This file contains the images shown below. The groups contained in this file will serve for many substances. If there are other groups or restraints which must be defined it may be done either by editing the ascii PRORES file or by supplying all the lines shown below in the specified order.

NOTE WELL: That additional groups which are added must be added after those groups already present. As noted elsewhere in this document the line order is very restricted in PROTIN due to the structure of the program itself.

The lines, *RESTRM, RESLNK,* and *RESIDU contain* the canonical coordinates for the amino acids and some other well known moieties. They also serve to set up the character strings which identify groups and atoms within the groups. The substrings, TRM, LNK, and RES, are then used in forming names for lines which follow where the distinction between groups which are terminal on chains, linking in chains, or a residue in a chain is important in forming the restraint derivatives. The lines which follow use the strings established in the RESxxx lines to connect restraints to the defined groups. Thus any new groups added will require the addition of all the other corresponding restraint input lines described below.

DISTRM, DISLNK, and DISRES supply the connectivity for the groups. This permits the calculation of the important distances. Four types of distances have been distinguished: actual bond distances, the next-nearest-neighbor distances from the triple of atoms that define bond angles, first to fourth atom distances that relate to prescribed dihedral angles as within planar groups, and hydrogen bond lengths. In restraining these geometric factors all the information is encoded in terms of atom pair distances which are used to produce the factors in the normal equations formed in PROLSQ.

PLARES, PLAPRO, PLATRM, PLALNK and PLALNL serve to establish constraints on the coplanarity that exists and should be conserved in some groups. In PROTIN/PROLSQ a method is used which restrains the deviations of the atoms of the refined structure from the least-squares plane of the group.

CHIRAL serves to define chiral centers within groups. The set of interatomic distances is insensitive to handedness so additional restraints must be imposed to assure the preservation of chirality. This is done by introducing a chiral volume equal to the triple scaler product of the vectors from a central atom to three attached atoms to quantify chirality. The sign of the chiral volume depends upon the handedness of the group and its magnitude equals the volume of the parallelepiped formed by the three vectors. The connectivity given in the CHIRAL line serves to allow the calculation of the ideal and model volumes to be used in forming the observational function.

VDWDIS serves to set up a table of atomic van der WAALs diameters which may then be used in the calculations of non-bonded contacts.

VDWTRM, VDWLNK, VDWRES, serve to supply the number of contact distances characteristic of each group. VDWCON lines are used to supply the distance information in the form of pairs of atoms which might come in non-bonding contact within the group. In the case of these distances the constraints are designed to follow a potential energy function that feature a steep repulsive barrier against close contacts and a very weak attractive potential. The observational function used is taken only over possible repulsive contacts. That is when d(model) < d(minimum). The value of d(minimum) depends upon the atomic elements in contact and on the type of contact: single-torsion separated atom pairs or multiple-torsion separated atoms.

TORTRM, TORRES, and their data supplying adjuncts NEIGHB, CHIDIS and CHIWGT serve to restrain torsion angles. Torsion angle restraints are among the least restricted of stereochemical features, but certain restrictions, related to non-bonded contacts, do apply. The nature of these restrictions, both for main-chain and for side-chains conformations in proteins are known. In PROTIN a simple quadratic form for the torsional potential is used. The differences in the angle of the torsional potential minimum, chi, is calculated for both for the ideal group and for the model and the difference in the square used in the observational function. The configurations treated can include quasi-planar torsions, as in the peptide bond, staggered potentials as in aliphatic side chains, transverse preferred conformations as in aromatic side chains, and targeted main-chain conformation angles such as in alpha helices. The definition of the torsion angles omega, phi, psi, chi(1),chi(2), etc. are given in Jane S. Richardson's paper (1981) which can serve as a useful general reference for the nomenclature used in setting up the restraints used in PROTIN/PROLSQ.

SECSTR allows for the naming and characterizing of known secondary structural features in terms of characteristic chi and phi values.

ELLTRM and ELLRES serve to control the thermal parameters shifts by restraining the motion of bonded atoms relative to one another.

INTERD, INTRAD, and SPCDIS restrain the distances between chains in structures with multiple chains.

SPCPLA serves to set up constraints for special planar groups.

EXCON serves to set up restraints which will exclude contacts between chains in the structure.

SECSEL serves to set up restraints on elements of secondary structure for back-bone torsion-angles.

SPCSYM, SYMOP, and SYMWGT serve to set up constraints based on chains related by non-crystallographic symmetry.

The PROTIN program should be run once before a series of refinements using PROLSQ. The augmenting normal-equation elements pertinent to the stereochemical restraints are stored in file F for communication to PROLSQ. The values in this file will change slowly as refinement takes place so that, in general, it is not necessary to rerun PROTIN for every execution of PROLSQ. Of course after several cycles of least-squares or when additional atoms are added to the model, PROTIN must be run again before PROLSQ.

NAMES OF CHAINS AND GROUPS/RESIDUES

In the description which follows a "chain identifier" will mean the character string supplied in field 5 of an ATOM line in PDB format or field 7 of a PROAT line. All reference to chains will be by these "names". Note that the PDB format limits chain names to 1 character while PROAT lines allow up to three characters. These names are placed sequentially in a table and transformed to relative pointers for use in PROTIN and PROLSQ.

Groups are defined in the RESTRM, RESLNK, RESIDU and RESIDE lines. The symbols used for these groups must be used in the PROATM lines to identify the group to which an atom belongs. Residues are a special subset of these groups consisting of the amino acid groups which make up the chains. Each residue in a chain is assigned a "residue sequence number" during atom loading by PROATM. This number is the one in field 6 of a PDB ATOM line or in field 8 of a PROAT line. During processing these residue sequence numbers are compressed into a sequential set of pointers developed at atom loading. The user refers to his "residue sequence number" in the input lines which follow, but it is well to keep in mind, that internally these pointer will dominate. In this case error messages may give the pointer which will require thought to find which residue is being pointed to.

The term "order number" refers to the number which points to the relative position of a given atom in a group. For example, arginine, ARG contains atoms N, CA, C, O, CB, CG, CD, NE, CZ, NH1, and NH2 with order numbers 1 through 11. The order numbers of linking groups are sometimes negative as a signal that the atoms are in the previous residue. This term is used in the data lines which follow and may appear in error messages from PROTIN.

PRINTER OUTPUT

Only the essential counts of restraints and atoms are printed at priority 3. Printer output is copious at priority 4. At 4, all the input data is echoed, and the compound specific distances, angles, etc. displayed.

FILE ASSIGNMENTS

Reads atom and unit cell data from a BDF on	file A
Writes geometric constraint data for PROLSQ on	file F
Optionally reads standard constraint data from file PRORES on	profil:

CONTROL LINES

NOTE: control line order is very restricted in PROTIN. In the following list certain control lines must be used many times in groups. Within the groups the use of one line requires the use of a following line or lines. The whole group is then repeated. The groups are set off in the sequence listing which follows. All groups must follow one another in the sequence shown below.

PROTIN	Program initiation line
	Group 1 input lines set up standard groups dictionary
*DISULB	Identify disulfide bridges
*FIRSTP	Identify atoms with population parameters to be refined
*SPCPEP	Identify the C and N terminus groups for each chain
*ALTCHN	Identify alternate chains

Note that, if at this point all the lines from RESTRM through ELLRES are left out the values will be taken from the PRORES file. This is the default file of idealized groups.
A PRORES file is supplied with the XTAL system which contains all the usual groups encountered in protein structures.

	Group 2 input lines set up ideal coordinates of residues
*RESTRM	Identify and idealize terminal and main residues
*RESLNK	Identify and idealize linkage residues
*RESIDU	Identify and idealize standard chain residues and other groups
*RESIDE	Identify and idealize the last standard residue at ends of chains
	Group 3 input lines set up distance restraints
*DISTRM	Characterize terminal chain residue interatomic distances
*DISLNK	Characterize linking chain residue interatomic distances
*DISRES	Characterize normal chain residue interatomic distances
	Group 4 input lines set up plane restraints

*PLARES	Identify planar groups in standard residues
*PLAPRO	Identify multi-planar prosthetic planar groups
*PLATRM	Identify terminal planar groups
*PLALNK	Set up the four linking planar groups
*PLALNL	Specify restraining pairs for linking planar groups
	Group 5 input lines set up chiral center restraints
*CHIRAL	Characterize ideal chiral centers
	Group 6 input lines set up contact restraints
*VDWDIS	Set van der Waals diameters by atomic species
*VDWTRM	Number of non-bonded contacts for specified terminal residue
*VDWLNK	Number of non-bonded contacts for specified linking residue
*VDWRES	Number of non-bonded contacts for specified residue
*VDWCON	van der Waals contact codes for VDWTRM, VDWLNK, or VDWRES
	Group 7 input lines set up torsion angle restraints
*TORTRM	Identify atoms defining ideal torsion angles in terminal residues
*NEIGHB	Identify neighbors in torsion groups
*CHIDIS	Set neighbor distance identification
*TORRES	Identify atoms defining ideal torsion angles in chain residues
*CHIWGT	Set weighting codes for torsion angles
*SECSTR	Characterize secondary structure features by psi and phi values
	Group 8 input lines set up anisotropic thermal restraints
*ELLTRM	Set ellipsoid specifications for terminal groups
*ELLRES	Set ellipsoid specifications for main chain groups

Note that the following lines are optional. All references to chain identities and user residue numbers are in terms of the names and numbers supplied through the PROATM program. These names and numbers are translated to pointers for use in the calculations. Some error messages may display these pointers. In this case the chain pointers will be 1, 2, ... In order of chain identification encountered in the ATOM or PROAT lines loaded in PROATM. The residue pointers will be the sequence number for the residue in order of loading by PROATM.

	Group 9 input lines set up special distances restraints
*INTERD	Set up blocks of inter chain special distances
*INTRAD	Set up blocks of intra chain special distances
*SPCDIS	Set up special inter-group distances
	Group 10 input lines set up special plane restraints
*SPCPLA	Specify special planar groups

Group 11 input lines set up contact exclusion restraints

*EXCON Specify distance identifiers for contact exclusion

Group 12 input lines set up backbone torsion angle restraints

*SECSEL Specify elements of secondary structure for backbone

Group 13 input lines set up noncrystallographic symmetry restraints

*SPCSYM Identify chains related by noncrystallographic symmetry
*SYMOP Rotation and translation of a noncrystallographic symmetry
*SYMWGT Set noncrystallographic symmetry restraint weighting specs

END One and only one END line required in input stream

PROTIN Program initiation line

PROTIN

Field 1	compound ID			
2-	Function	Option code		
	Print atom coordinates	ATL		<no>
	Signal neutron data	NEU		<no>
		CEL	n	< >
		NAP	n	
	Limit for van der Waals distances	VDW	x.x	<5.0>
	Expect hydrogen atoms	HAT		<no>
	Expect individual aniso t.f.	ANI		<no>
		ATL		

DISULB Identify disulfide bridges
Field 1 Chain identifier of CYS residue on one side of the S-S bridge.
2 Residue sequence number on that side.
3 Chain identifier of CYS residue on other side of the S-S bridge.
4 Residue sequence number on the other side.

FIRSTP Identify atoms with population parameters to be refined
Field 1 Name of the first atom, in the notation of PROATM on loading, to have the population parameter refined. This atom and all atoms to the end of the atoms loaded will have their population parameters refined in the subsequent PROLSQ run.

SPCPEP Identify "cis" residues
Field 1 Chain identifier symbol e.g. A2
2 Number of cis peptide bonds in chain
3 Residue sequence number of cis peptide

ALTCHN Identify alternate chains

Field 1 Primary chain identifier for mothers of alternate chains
2 Secondary chain identifier for mothers of alternate chains

	p	s
if there is 1 chain and 1 alternate chain:	1	0
if there is 1 chain and 2 alternate chains:	1	1
if there are 2 chains and each has 1 alternate chain:	1	2
if there are 2 chains and only chain 1 has 1 alternate chain:	1	0
if there are 2 chains and only chain 2 has 1 alternate chain:	2	0

Alternate conformations of discretely disordered residues are assigned during atom loading a different chain number from the mother chain for a given residue. Only the atoms which are modeled in different positions should be included among the input atoms.

Set up groups (residues) in ideal form as a dictionary
RESTRM Terminal
RESLNK Linking
RESIDU "Standard"
RESIDE Note last of "standard" groups
Field 1 Group name for residue e.g. NAMINO ; ASP ; HIS
2 Name of atom in the group e.g. CA for alpha carbon
3 X Cartesian coordinate in angstroms relative to C alpha as the origin in each group
4 Y Cartesian coordinate
5 Z cartesian coordinate
6 Order number of atom within the group
e.g. 1 for N, 2 for C alpha, etc. For all standard groups, save glycine, this number starts at 5 because the coordinates of RESTRM MAIN are automatically taken for the first four atoms of every group. This number may also be negative in RESLNK lines in order to designate atoms belonging to a "previous" residue.
7 Single letter code which is to be used as a synonym for the group name e.g. G for GLY. Used only for standard groups.

Set up distance restraints for idealized groups
DISTRM
DISLNK
DISRES
Field 1 Group name as defined in RESxxx input lines e.g. NAMINO
2 Order number of origin atom for corresponding distance
3 Order number of terminus atom for corresponding distance
4 Code to define weight that is to be used to restrain distance
1 Bonded pair 2 Angle pair
3 Planar 1-4 distance 4 XH bond distance
5 Y-X-H angle distance 6 H-X-H angle distance
5 Code to define weight that is to be used to restrain B values
0 Any distance for which B restraint is not required

1 Bond distance between two main chain atoms
2 Angle distance between two main chain atoms
3 Bond distance involving at least one side chain atom
4 Angle distance involving at least one side chain atom
5 Bond distance involving a hydrogen
6 Angle distance involving a hydrogen

Fields 6-9, 10-13, etc. as for Fields 2-5. Note that lines may be continued by placing a "C" as the last field on a line. The next line must then have the same ID for both line type and group name. Fields will continue in groups of four to complete the required restraints.

Set up restraints on planarity for idealized groups

PLARES
PLAPRO
PLATRM
PLALNK

Field 1 Group name
2- The order number of the atoms of the group which are to be restrained to planarity. If a group has no planes fields 2- are left blank.

PLALNL

Field 1- Codes specifying all the possible "restraining pairs" among the atoms specified in the RESLNK input lines. Note that only the number of atoms specified to be in planes and given in the PROTIN line under directive NAP will be considered to restrain the plane.

Set up chiral centers for idealized groups

CHIRAL

Field 1 Group name
2 Descriptor code for how the chirality of the group is defined
1 for groups "intrinsically" chiral
0 for chirality related to nomenclature e.g. LEU VAL

Set up contact distances for idealized groups

VDWDIS

Field 1 Name of atom species e.g. HYDROGEN (limit 12 characters)
2 Atomic number of species or atomic scattering factor pointer Carbonyl is arbitrarily set to 3
3 van der Waals diameter of the species or distance between centers of two non-bonded atoms of the species

VDWTRM
VDWRES

Field 1 group name
2 number of non-bonded contacts for group

VDWLNK

Field 1 group name

2 number of non-bonded contacts for group
3 number of non-bonded contacts specified for prolyl link

VDWCON

These lines follow their corresponding VDWRES, VDWTRM, or VDWLNK lines. There must be enough of these lines to satisfy the counts given in the header lines.

Field 1 Group name
2 Order number of origin atom corresponding to a possible non-bonded contact
3 order number of terminus atom for corresponding possible non-bonded contact
4 Non-bonded distance restraint class as:
1 to indicate that the relative position of the given atoms is determined by only one torsion angle
2 to indicate that the relative position of the given atoms is determined by two or more torsion angles
5-7, 8-10, etc. as 3-4 continuing onto additional lines as required to satisfy the counts given in the preceding VDWRES, VDWTRM, or VDWLNK lines.

Set up torsion angle restraints for idealized groups

TORTRM
TORRES

Field 1 Group
2 Atom order number for C(i-1)
3 Atom order number for N(i)
4 Atom order number for C alpha(i)
5 Atom order number for C(i)
6 Atom order number for N(i+1)
7 Atom order number for C alpha(i+1)
fields after 7 are ignored in TORTRM lines
8 Atom order number for C beta(i)
9 Atom order number for C gamma(i)
10 Atom order number for C delta(i)
11 Atom order number for C epsilon(i)
12 Atom order number for C zeta(i)
13 Atom order number for N ?(i)

C(i-1)-N(i)-C alpha(i)-C(i)	specifies phi
N(i)-C alpha(i)-C(i)-N(i+1)	specifies psi
C alpha(i)-C(i)-N(i+1)-C alpha(i+1)	specifies omega
N(i)-C alpha(i)-C beta(i)-C gamma(i)	specifies chi(1)
C alpha(i)-C beta(i)-C gamma(i)-C delta(i)	specifies chi(2)

NEIGHB

Field 1 Group
2- Neighbor identification codes
-1 for atoms from residue i-1
0 for atoms from residue i
1 for atoms from residue i+1
5 for atoms from terminal groups e.g. OT of carboxyl terminus

CHIDIS

Field 1 Group
2 Distance identification codes as specified in field 4 of the DISRES input line. Negative numbers imply ??????

CHIWGT

Field 1 Group
2- Weighting code for side chain chi angles
0 for no specifications
2 for planar e.g. chi(5) of ARG
3 for staggered e.g. aliphatics
4 for orthonormal e.g. chi(2) of aromatics

SECSTR

Field 1 Name for element of structure up to 20 characters maximum
2 code for secondary structure element
1 alpha helix
2 3(10) helix
3 pi helix
4 collagen type helix
5 parallel beta sheet
6 anti-parallel beta sheet
7 classic beta bulge position 1
8 classic beta bulge position 2
9 beta bend [or turn] type I position 2
10 beta bend [or turn] type I position 3
11 beta bend [or turn] type II position 2
12 beta bend [or turn] type II position 3
13 gamma turn position 1
14 gamma turn position 2
15 gamma turn position 3
3 Characteristic phi value
4 Characteristic psi value
5 Literature reference of up to 40 characters

Set ellipsoid specifications

ELLTRM
ELLRES

Field 1 Group name
2 order number for atom for which ellipsoid is being specified
3 order number for atom p(0)
4 order number for atom p(1)
5 order number for atom s(0)
6 order number for atom s(1)
Note that negative numbers are used to point to atoms in preceding residues.
7-11, 12-16, etc. for additional ellipsoid atoms; new lines to continue.

Set intra and inter chain distance restraints

INTRAD

Field 1 Origin chain identifier

INTERD

Field 1 Origin chain identifier
2 Target chain identifier

There must be SPCDIS lines following each INTRAD or INTERD line supplied. The number required will depend upon the number of special inter-group distances for this chain block to be restrained. e.g. S gamma - S gamma from two different CYS groups that form a bridge; distance from solvent to protein; etc. One line per pair of atoms for which distance is to be restrained.

SPCDIS

Field 1 Residue sequence number of group to which origin atom belongs
2 Order number of origin atom within group
3 Residue sequence number of group to which target atom belongs
4 Order number of target atom
5 Distance in Angstroms to which the distance between the atoms should be restrained
6 Code to determine weight that should be used to restrain distances. Same as those shown for DISRES lines.
7 Code to determine weight that should be used to restrain temperature factors. Same as those shown for DISRES lines.

Specify special planar groups

SPCPLA

Field 1 Chain identifier of first atom in special plane
2 Residue sequence number of first atom in special plane
3 Order number of first atom in special plane
4-6 The same quantities as in 1-3 for the second,
7-9 third, ... atoms which define the special plane.
etc.

Specify contact exclusion

EXCON

Field 1 Chain identifier of first atom for which contacts are to be excluded
2 Residue sequence number of first atom
3 Order number of first atom
4-6 The same quantities as in 1-3 for the second, atom which defines the contact exclusion.

Specify elements of secondary structure for the backbone

SECSEL

Field 1 Chain identity; all chains of this type will be set
2 Residue sequence number of the initial group for this region of structure
3 Residue sequence number of the final group for this region of structure
4 Code for structure type

-1 restrain phi and psi values to those of initial structure
0 no restraints on phi and psi
n restrain psi and phi values specified in SECSTR line for this n

Specify special non-crystallographic symmetry

SPCSYM
Field 1-n Chain identities for chains in this special non-crystallographic symmetry group
Each SPCSYM line must be followed by one or more SYMOP lines.

SYMOP
Field 1-12 R11, R12, R13, T1, R21, R22, R23, T2, R31, R32, R33, T3

SYMWGT
Field 1 Initial residue sequence number to which the SYMOP applies
2 Final residue sequence number to which the SYMOP applies
3 weighting code

Code	Implication for Main chain	Side chain
1	1	1
2	1	2
3	1	3
4	2	2
5	2	3
6	3	3

Where 1=tight; 2=medium; 3=loose
4-6, 7-9,... as in 1-3 for all group spans specified

EXAMPLE

Note that the program PROATM must have been run before PROTIN may be run in order that the atom coordinates to be constrained will be available on file A so that the appropriate constraint information can be generated on file F for use in the refinement process.

```
TITLE  CRAMBIN FROM ABYSSINIAN CABBAGE SEED-HENDRICKSON & TEETER 4 APR 89  JMS.
PROTIN CRAMB VDW 4.51 NAP 5 ATL  GNL  CEL 3 ANISO
END
PROLSQ CRAMB   NCY 25  PCH ISO
PRINT   ALIS 1
STATS  5.0 4.0 3.0 2.0 1.0
RTEST    5  23   1   1   4    3    .25 .45 .65 1.25 .35 .65 .95
END
FINISH
```

REFERENCES

Finzel,B. C. Incorporation of Fast Fourier Transforms to speed Restrained Least-squares Refinement of Protein Structures. J. Appl. Cryst. (1987) 20, 53-55

Hendrickson, W. A. Stereochemically Restrained Refinement of Macromolecular Structures in Diffraction Methods for Biological Macromolecules Part B from Methods in Enzymology Edited by H. W. Wyckoff, C. H. W. Hirs, & S. N. Timasheff Academic Press, Inc. (1985)

Hendrickson, W. A., & Konnert, J. H. (1979) in Biomolecular Structure, Conformation, Function, and Evolution, edited by R. Srinivasan, Vol 1, pp. 43-57. Pergammon Press, New York

Konnert, J.H. A Restrained-Parameter Structure-factor Least-squares refinement Procedure for Large Asymmetric Units (1976). Acta Cryst. A32, 614-617.

Konnert, J. H. & Hendrickson, W. A. A Restrained-parameter Thermal-factor Refinement Procedure (1980). Acta Cryst. A36, 344-350.

Richardson, J. S. Protein Anatomy, page 168-
Chapter in Advances in Protein Chemistry (1981) Academic Press, New York
Edited by C. B. Anfinsen, J. T. Edsall, and F. M. Richards

Sheriff, S. Addition of Symmetry-related Contact Restraints to PROTIN & PROLSQ J. Appl. Cryst. (1987) 20, 55-57.

PROLSQ: Constrained Refinement of Atomic Parameters

Wayne A. Hendrickson: Department of Biochemistry/Biophysics
Columbia University, New York, NY 10032

John H. Konnert: Laboratory for the Structure of Matter 6030
Naval Research Laboratory, Washington D.C. 20375

Translated to the XTAL system, with permission, by
James M. Stewart: Department of Chemistry/Biochemistry
University of Maryland, College Park, MD 20742

SYNOPSIS

PROLSQ is a restrained-parameter structure-factor least-squares atomic parameter refinement program for large molecules. This XTAL version is derived from the program of John H. Konnert (1976) which was elaborated for protein refinement by Wyane A. Hendrickson (1980). Hendrickson (1985) has described the technique and philosophy of the use of PROLSQ in a readily acquirable publication.
Since that time a number of others have made changes, in details, to the original code and documentation. Among these contributors are Barry Finzel, Steven Sheriff, Anita Sielecki, Janet Smith, & Alex Wlodawer.

PURPOSE

The program will refine atomic models using both the conditional structure-factor least-squares procedure described by Waser and geometric restraints so that the canonical stereochemistry of the sub-elements of the structure will be taken into account in the least-squares process.

The information about the restraining geometry must be supplied by use of the PROTIN program of Hendrickson which supplies the geometry in terms of connectivity, covalent radii, bond angles, planarity, chirality, and other structural features of macromolecular substances. This information and the diffraction data are combined in the derivative matrix of the least-squares procedure such that elastic bounds are imposed on the geometry of the model. The restrained least-squares procedure uses a greatly reduced, "sparse", matrix. The parameter shifts are derived from this matrix not by inversion but by the iterative conjugant gradient method of Hestenes & Stiefel applied, described, and programmed by Konnert (1976).
This approach has to refinement several advantages over conventional structure factor least-squares procedures: first, the size of the matrix increases linearly rather than quadratically. Second, the restraints effectively reduce the number of parameters to be determined. Third, the method usually shows a larger radius of convergence. Fourth, all parameters may be refined simultaneously, but only correlations dictated by the geometry are included in the derivative matrix.

Before the PROTIN/PROLSQ programs are used, Hendrickson's paper (1985) should be read. In the section "Refinement Strategy", pages 264-268, is given an outline of the appropriate use of these codes for the refinement of macromolecular structures.

METHOD

The method is described by Konnert (1976). The function minimized is the weighted sums of the difference in the squared F calculated and F relative, of squared ideal and observed interatomic distances, and of other types of restraints. The list of the currently allowed restraints is given in the documentation of PROTIN and will not be reiterated here. The weighted, squared differences are multiplied by appropriate numerical differentials of the structure factors, the distances, or other factors, with respect to the parameters of the structure to form the vector of the normal equations. The matrix of the normal equations is formed from the diravative terms alone. Only correlations between atoms related by constraining features of the structure are saved in memory. This results in the formation of a "sparse" matrix since the off-diagonal elements are dispersed throughout the matrix. Since all parameters are refined simultaneously and standard matrix inversion methods are not space efficient and the problem is not separable into a number of smaller sets of equations, as in conventional block-diagonal least squares Konnert uses the conjugate gradient iterative technique for the solution of the normal equations. The conjugant gradient technique retains the sparse derivative matrix throughout the solution for the shifts in atomic parameters. The non-zero elements of the hypothetical n by n derivative

matrix are retained during the iterative solution procedure. For the refinement of a structure of n atoms and m distance restraints, for example, the number of related elements in one half of the symmetric matrix that need be stored is 6n + 9m. Since m is roughly linear with n, m is about 3n, storage space required varies linearly with n. To ensure rapid convergence, it is necessary to know or determine the optimum damping factor for the computed shifts. In the XTAL version of PROLSQ this value is set to 0.5 by default. However, use of the RTEST line will permit calculating R for a selected sub-set of reflections. Once a suitable damping factor is found it must be applied to all shifts if the desired geometry is to be retained.

REFINEMENT WEIGHTS

The optional WGTxxx input lines allow for the adjustment of refinement weights. These weights are given in terms of "standard deviations" of the class of observation. The default values set in this XTAL version are similar to those shown by Hendrickson (1985). In this paper there is a section devoted to the proper use of weights.

The setting of these weights is very important to the successful use of PROLSQ on two accounts. First, if the restraint weights are imbalanced with respect to geometric restraints and structure factor refinement, one will dominate the other giving rise to large divergence from ideal geometry or divergence in terms of the structure factors. If the restraints are too heavily weighted the R value may actually increase as the refinement goes forward since the conformity to the idealized model will dominate the structure factor information. Second, if any of the weights cause the weighted deltas squared to be very large there is the possibility that "floating point overflow" will occur during the conjugant gradient process, especially on 32 bit machines.

The numbers in the output which should be monitored are the "average weighted delta squared" values for each of the restraints and for the structure factors. In a reasonable refinement these values all approach a common value, usually near 1.0 as the refinement proceeds to completion. Furthermore, if any of the categories shows a value much larger than the others, then that restraint will apply with the greatest effect in the least-squares process.

The values of the weights may be set by the use of the WGTXXX input lines and by the inclusion of experimental sigma F values. See Hendrickson's paper for elaboration of this point. A value for the weights that will force the average to 1.0 may be found by looking at the corresponding value of the average delta given in the output and supplying this value as the sigma in the appropriate WGTXXX line.

For example consider the following "snapshot" of input and output:

Including the input lines:

```
WGTB  *2  0.5  0.5  0.5  0.5  0.5  0.5  0.5  0.5
```

```
WGTREF  1    2.5   -10.0
```

Produces the following output:

```
 >>>>>>>>>>>>>>>>>>>>>>>> THERMAL PARAMETER RESTRAINTS <<<<<<<<<<<<<<<<<<<<<<<<

 SUM(WGT*DELTA**2) =    .4163904+04
 AVG(WGT*DELTA**2) = 23.525 FOR ISOTROPIC THERMAL FACTORS

        ROOT          RMS          MEAN       RMS    SIGMA
 TYPE  1/WGT NUMBER   DELTA   SIGMA \DELB\     DELU      U THERMAL FACTOR TYPE
   1    .500     21   1.912    .500   1.296   .1281    .07 MAIN-CHAIN BOND
   2    .500     21   2.278    .500   1.676   .1457    .07 MAIN-CHAIN ANGLE
   3    .500     42   2.487    .500   1.575   .1412    .07 SIDE-CHAIN BOND
   4    .500     63   3.060    .500   1.947   .1570    .07 SIDE-CHAIN ANGLE
   5    .500      0    .000    .500    .000   .0000    .07 X-H BOND
   6    .500      0    .000    .500    .000   .0000    .07 X-H ANGLE
   7    .500      0    .000    .500    .000   .0000    .07 METAL COORD. ETC.
   8    .500      3   1.342    .500   1.341   .1303    .07 HYDROGEN BOND

 AGREEMENT FACTORS BASED ON PARAMETERS BEFORE THIS CYCLE

 SUM(WGT*DELTA**2) =    .5221144+07     AVG(WGT*DELTA**2) = 1508.57

                     NUMERATOR   DENOMINATOR     R
 STANDARD R-FACTOR      3203.9        68627.7  .047
 WEIGHTED R-FACTOR      2285.0        46805.6  .049
 ERROR R-FACTOR          686.3        68627.7  .010

    USED  REJECTED  R SURVEY      R SURVEY
    3461         0         0             0

     .93 IS THE AVERAGE \FO-FC\ DISCREPANCY

 ---------- RESOLUTION BREAKDOWN ----------
 DMIN    REFS     <SIGF>/DATA     <SIGF>/USED   <FO-FC>    R SHELL   R SPHERE
 5.00      36             .72            3.43     2.998      .041       .041
 3.00     116             .73            2.79     3.673      .050       .048
 2.50     101             .46            2.33     2.008      .044       .047
 2.00     226             .39            1.90     1.977      .051       .048
 1.70     277             .28            1.43     1.217      .043       .047
 1.50     324             .22            1.02     1.020      .047       .047
 1.30    2381             .12             .16      .568      .046       .047

  ALL    3461                                                           .047

 MODELED DELTA  = (   2.5) + (  -10.0)*(S-1/6)

 FITTED <FO-FC> = (   2.5) + (  -10.3)*(S-1/6)
```


Then, from the same starting point, the run is made again having altered only the weight line. Note the values of delta in the first run and the values of sigma in the second run:

Input:

```
WGTB  *2  1.9  2.3  2.5  3.1  0.5  0.5  0.5  1.3
WGTREF  1   2.5  -10.0
```

Output:

```
 >>>>>>>>>>>>>>>>>>>>>>>> THERMAL PARAMETER RESTRAINTS <<<<<<<<<<<<<<<<<<<<<<<<

 SUM(WGT*DELTA**2) =    .1480259+03
 AVG(WGT*DELTA**2) =    .836 FOR ISOTROPIC THERMAL FACTORS

        ROOT          RMS          MEAN       RMS    SIGMA
 TYPE  1/WGT NUMBER   DELTA   SIGMA \DELB\     DELU      U THERMAL FACTOR TYPE
   1   1.900     21   1.912   1.900   1.296   .1281    .15 MAIN-CHAIN BOND
   2   2.300     21   2.278   2.300   1.676   .1457    .17 MAIN-CHAIN ANGLE
   3   2.500     42   2.487   2.500   1.575   .1412    .17 SIDE-CHAIN BOND
   4   3.100     63   3.060   3.100   1.947   .1570    .19 SIDE-CHAIN ANGLE
   5    .500      0    .000    .500    .000   .0000    .07 X-H BOND
   6    .500      0    .000    .500    .000   .0000    .07 X-H ANGLE
   7    .500      0    .000    .500    .000   .0000    .07 METAL COORD. ETC.
   8   1.300      3   1.342   1.300   1.341   .1303    .12 HYDROGEN BOND

 AGREEMENT FACTORS BASED ON PARAMETERS BEFORE THIS CYCLE

 SUM(WGT*DELTA**2) =    .5221144+07     AVG(WGT*DELTA**2) = 1508.57

                     NUMERATOR   DENOMINATOR     R
 STANDARD R-FACTOR      3203.9        68627.7  .047
 WEIGHTED R-FACTOR      2285.0        46805.6  .049
 ERROR R-FACTOR          686.3        68627.7  .010

    USED  REJECTED  R SURVEY      R SURVEY
    3461         0         0             0

     .93 IS THE AVERAGE \FO-FC\ DISCREPANCY

 ---------- RESOLUTION BREAKDOWN ----------
 DMIN    REFS     <SIGF>/DATA     <SIGF>/USED   <FO-FC>    R SHELL   R SPHERE
 5.00      36             .72            3.43     2.998      .041       .041
 3.00     116             .73            2.79     3.673      .050       .048
 2.50     101             .46            2.33     2.008      .044       .047
 2.00     226             .39            1.90     1.977      .051       .048
 1.70     277             .28            1.43     1.217      .043       .047
 1.50     324             .22            1.02     1.020      .047       .047
 1.30    2381             .12             .16      .568      .046       .047
```

```
ALL    3461

MODELED DELTA  =  (   2.5) + (  -10.0)*(S-1/6)

FITTED <FO-FC> =  (   2.5) + (  -10.3)*(S-1/6)
```

ATOM NAMES IN LISTS

Each atom in the structure is loaded from the XTAL program PROATM. PROATM reads atom parameters and "names" in the Protein Data Bank format. The atom names consist of an atom serial number, which changes as the numbers of known atoms in the model changes; an atom name, which consists of a scattering factor symbol, a remoteness indicator code, and a branch designator; an alternate location indicator; a residue name; a chain identifier; a residue sequence number; and a code for insertion of residues. In PROLSQ listings these name parts are compressed to a 9 character string for compact listing purposes. This string is structured in the following way:

Position	Contents
1	An integer pointing to the chain to which the atom belongs
2	A single letter code for the residue to which the atom belongs
3-5	The residue sequence number from the PDB input to PROATM
6-9	The atom name; symbol, remoteness, and branch; left justified

The single letter codes used by PROLSQ are shown in the following table which shows the three character and one character code side by side.

Residue	3	1	Residue	3	1	Residue	3	1
Alanine	ALA	A	Arginine	ARG	R	Asparagine	ASN	N
Aspartic acid	ASP	D	ASP/ASN ambiguous	ASX	B	Cysteine	CYS	C
Cystine	CYS	C	Glutamine	GLN	Q	Glutamic acid	GLU	E
GLU/GLN ambiguous	GLX	Z	Glycine	GLY	G	Histidine	HIS	H
Isoleucine	ILE	I	Leucine	LEU	L	Lysine	LYS	K
Methionine	MET	M	Phenylalanine	PHE	F	Proline	PRO	P
Serine	SER	S	Threonine	THR	T	Tryptophan	TRP	W
Tyrosine	TYR	Y	Valine	VAL	V			

PRINTER OUTPUT

The total mass of printing may be controlled with the use of a PRINT line. The print line allows for selected and limited printing of shift, reflection, and selected large deviations from distance restraints at priority 3. Most echoing of constraint information is done at priority 4. The results of each cycle are printed at priority 3 as are R values which result from use of the RTEST input line. If a sampling of the matrix and vector elements is desired, print priority 5 may be specified. This feature will result in copious printed output in all but the smallest test cases.

FILE ASSIGNMENTS

Reads atom and reflection data from a BDF on	file A
Writes refined atomic parameters and Fc on	file B
Scratches an intermediate BDF on	file E
Reads geometric constraint data from PROTIN on	file F
Optionally writes reflections excluded from refinement on	file G
Optionally writes reflections for an R value survey on	file H
Optionally writes images of Protein Data Bank ATOM lines on	unitpch:

CONTROL LINES

PROLSQ	Program initiation line
*ARCHIV	Specify items to be placed in reflection records
*BOV	Overall isotropic temperature factor
*DAMPSH	Shift damping factors to control refinement
*PRINT	Print control line
*REFLIM	Limits on reflections to be included in refinement
*RTEST	Set up survey of R values to assess refinement
*SCALE	F relative scale factors
*SCANIS	Anisotropic scale factor
*SOLMOD	Solvent model parameters
*STATS	Number and limits of d shells for R value survey
*WGTB	Set thermal restraint weighting
*WGTCHI	Set chiral center restraint weighting
*WGTCON	Set contact distance restraint weighting
*WGTDIS	Set distance restraint weighting
*WGTPLA	Set plane restraint weighting
*WGTREF	Set structure factor restraint weighting
*WGTSHF	Set excessive shift restraint weighting
*WGTSYM	Set noncrystallographic symmetry restraint weighting
*WGTTOR	Set torsion angle restraint weighting
*NOREF	Control atoms included in refinement

PROLSQ

Field			Option code		
1		compound ID			
2-		Function			
		Congugent gradient cycles	NCY	n	<n=50>
		Lagrange multiplier scale	LAG	n	<n=10>
		Normal refinement	<REF>		
		Ideal structure refinement no structure factor calc.	IDE		
		Skip refinement and only calculate structure factors	SKI		
		Evaluate idealization but do not refine	IZE		
		Apply anomalous dispersion	ANO		

do not apply dispersion	<NOA>	
Temperature factors type		
overall isotropic	OVA	
individual isotropic	ISO	
Data set number	DSN n	<n=1>
Constrain origin		
along x	ORX	<default is no>
along y	ORY	<default is no>
along z	ORZ	<default is no>
"Punch" atom coordinates in PDB format on unitpch:	PCH	<default is no>

Control items in lrrefl: of output BDF

ARCHIV

Field 1 - ID numbers of data items to be added to or deleted from the BDF. Positive ID numbers signal addition, negative deletion. See long comment in program FC.

Set overall isotropic temperature factor

BOV

Field 1 Either B overall or U overall depending on <B from input BDF> magnitude. If value is 0 < B < 0.2 the quantity is multiplied by 78.9568 and stored as B. If the quantity is 0.2 < B < 40.0 it is accepted as B and stored. The line is ignored unless in overall t.f. mode.

Set parameter shift damping factors

DAMPSH

Damping factors to be applied to parameter shifts determined by the least-squares process. These factors will be used to establish the new parameters written to the output BDF on file B and, optionally on unitpch:. Factors between 0 & 4 allowed. N.B. Unequal damping factors will distort the geometry.

Field 1	Damping factor for x parameter shifts	<0.5>
Field 2	Damping factor for y parameter shifts	<0.5>
Field 3	Damping factor for z parameter shifts	<0.5>
Field 4	Damping factor for B parameter shifts	<0.5>
Field 5	Damping factor for population parameter shifts	<0.5>
Field 6	Damping factor for B11 parameter shifts	<0.5>
Field 7	Damping factor for B22 parameter shifts	<0.5>
Field 8	Damping factor for B33 parameter shifts	<0.5>
Field 9	Damping factor for B12 parameter shifts	<0.5>

Field 10	Damping factor for B13 parameter shifts	<0.5>
Field 11	Damping factor for B23 parameter shifts	<0.5>

Control the volume of printed information

PRINT

Field 1- Option control codes

Function	Option code		
List first m reflections	LFIR	m	<m=10>
List reflection data	FLIS	m	<m=0> list every mth reflection
List first n input atoms ILIS not present means no input atoms listed	ILIS	n	<n=99999>
List updated atoms m = 0/1/2 for none/all/beyond shift limits set below	ALIS	m	<m=2>
Atom parameter shifts on x, y, and z where x.x is distance in A	PARC	x.x	<x.x=0.1> only those shifts > x.x
temperature factor shifts	BCUT	x.x	<x.x=0.25>
population parameter shifts	POPC	x.x	<x.x=0.25>
Torsion angle printing			
all standards and statistics	STOR		
statistics only	<NTOR>		
single angle contacts	CTOR	x.x	<x.x=0.1>
multiple angle contacts	MTOR	x.x	<x.x=0.1>
Chiral centers printing			
worst case chiral centers	WCHI		
all chiral centers	ACHI		
statistics only	<NCHI>		
Hydrogen bond printing			
all possible H bonds	HBON		
statistics only	<NHBO>		
Restrained distances printing			
distances of type code "J" if |DIDEAL-DMODEL| >= x*SIGD(J)	DISC	x.x	<x.x=5.0> x.x=0.0 for print all
Restrained planes printing			
all planes with RMS deviation > x*SIGP	PLAN	x.x	<x.x=2.0>

Limit reflection included in refinement process

EFLIM

Field 1- Option control codes

Function	Option code	
Discard reflections based on the following criteria:		
F relative < x.x	FMIN x.x	<x.x=0.0>
sin(theta)/lambda < x.x	SMIN x.x	<x.x=0.0>
sin(theta)/lambda > x.x	SMAX x.x	<x.x=1.5>
d maximum	DMAX x.x	<x.x=1000.>
d minimum	DMIN x.x	<x.x=0.35>
F(rel) < sigma(F(rel))	SIGM x.x	<x.x=0.0>

note: Use either d or stol specs but not both.

Set up a survey of R values as a function of shift damping

TEST

Set up R testing for reflection subsets. The subsets are defined by the reflection count plus a constant modulo the frequency being zero. After the least-squares refinement structure factors are recalculated for the subsets in order to evaluate the efficacy of the shift damping and progress of the refinement.

Field 1	Sampling frequency for reflections to be included in the refinement and subsequent Fc for R. Caution: If this number is set too low and fields 5 & 6 are large a great deal of time will be needed.	<10000>
2	Sampling frequency for reflections to be excluded from the refinement but for which a subsequent structure factor R value will be calculated. This gives a stringent test of convergence, but caution is required or the number of reflections eliminated may be too great.	<no>
3	Constant to offset from the beginning of sampling of included reflection subset.	<0>
4	Constant to offset from the beginning of sampling of excluded reflection subset.	<0>
5	Number of coordinate, x,y,z, shift steps to be tested.	<1>
6	Number of B steps to be tested.	<1>
7-n	Factors to multiply shifts for testing. These are the same kind of factor as those in the DAMPSH line. e.g. x(test) = x(start) +factor*x(shift) The sum of fields 5 & 6 determines the value of n.	<0.5>

Set F relative scale factors

SCALE

Field 1	F relative scale factor <value from input BDF>	
2	Scale group number <1>	

Set an overall anisotropic scale factor

SCANIS

Overall anisotropic scaling of F calculated

Field 1	B11 component	<0.0>
Field 2	B12 component	<0.0>
Field 3	B13 component	<0.0>
Field 4	B22 component	<0.0>
Field 5	B23 component	<0.0>
Field 6	B33 component	<0.0>

Set solvent modeling parameters

SOLMOD

Field 1	Linear scaling constant for partial F calculated term extracted as A part and B part from input BDF. The partial F calc will be multiplied by this factor before being added to the atomic F calculated.	<1.0>
Field 2	Exponential scaling constant, B. The partial F calculated is multiplied by EXP(-B*stol**2) before being added to F calc.	<0.0>
Field 3	Linear scaling constant for the analytic solvent attenuation function, UK.	<1.0>
Field 4	Experimental, exponential, scaling constant for the analytic solvent attenuation function, UB. Atomic scattering factors are multiplied by (1.0 - UK*EXP(-UB*stol**2)) prior to being used in the structure factor calculation.	<0.0>
Field 5	Electron density plateau level to be used to model space not occupied by atoms. (?!)	<0.0>

Set resolution limits for R value calculations

STATS

Field 1-7 Resolution limits for determining R values for shells of reciprocal space. Maximum of seven shells; default one. Values supplied must be descending order d spacings in Angstroms.

WGTB

Field 1	Weight for temperature factor restraints, WB applied weight in isotropic case = ((WB/sigma(B))**2 * (delta(B))**2)	<1.0>
2	sigma of type 1 main-chain bond distances	<1.0>
3	sigma of type 2 main-chain angle distance	<1.5>
4	sigma of type 3 side-chain bond distance	<1.0>

5	sigma of type 4	side-chain angle distance	<1.5>
6	sigma of type 5	X-H bond	<0.5>
7	sigma of type 6	X-H angle	<1.0>
8	sigma of type 7	hydrogen-bond, metal, etc.	<0.0>

WGTCHI

Field 1 Weight for chiral center restraints, WC <1.0>
applied weight = ((WC)/sigma(chi))**2

2 Sigma(chi) <0.15>

WGTCON

Field 1 Weight for van der Waals contacts, WV <1.0>
applied weight = ((WV)/sigma(vdv))**2

2 Sigma(vdv) <0.5>

The following fields may be used to supply parameters to modify the minimum "theoretical" van der Waals contact distance of four possible types which are defined and calculated in PROTIN and passed in file F. There are four parameters corresponding to:

1) Atoms in which the relative position is determined by only one torsion angle.
2) Two atoms in which two or more torsion angles are involved, multiple torsion contact.
3) A possible (X....Y) hydrogen bond pair involving contacts between nitrogen oxygen pairs, but not N main - N main or O main - O main.
4) A possible (X-H....Y) hydrogen bond.

3	Type 1 distance factor in Angstroms	<-0.3>
4	Type 2 distance factor in Angstroms	<0.0>
5	Type 3 distance factor in Angstroms	<-0.2>
6	Type 4 distance factor in Angstroms	<0.0>

Note: The distance factors must be negative amounts. They are used to adjust the ideal contact distances supplied from PROTIN in order to establish when a contact should be restrained.

WGTDIS

Field 1 Weight for distance restraints, WD <1.0>

The weight applied to a derivative is:
((WD/sigma(i))*(ideal dist - actual dist))**2
The following fields set sigmas for various types of distances. These types for possible i are:
1) bonded distances (e.g. C alpha - C)

2) angle distances (e.g. N - C)
3) planar 1-4 distances (e.g. O i-1 - C alpha1)
4) bond distances involving H (e.g. C alpha - H)
5) angle distance involving H (e.g. H alpha - C)
6) hydrogen bond, metal
7) coordination, etc.

2	sigma of type 1	<0.03>
3	sigma of type 2	<0.04>
4	sigma of type 3	<0.05>
5	sigma of type 4	<0.1>
6	sigma of type 5	<0.08>
7	sigma of type 6	<0.0>
8	sigma of type 7	<0.0>

WGTPLA

Field 1 Weight for plane restraints, WP <1.0>

2 sigma of a plane <0.02>

The weight applied to the derivative is:
(WP/(sigma(p))**2

WGTREF

Field 1 Weighting scheme as 1 through 5 <1>

2 A term for weights <4.0>

3 B term for weights <0.0>

Where:
Sd = A + B*(sin(theta)/lambda - 0.1667)
Sa is the "sigma" used to calculate the structure factor weight = 1/Sa**2
scheme 1: Sa = Sd
scheme 2: Sa = MAX(sigma(Frel),Sd)
scheme 3: Sa = sigma(Frel) from input BDF
scheme 4: Sa = A*sigma(Frel)
scheme 5: Sa = SQRT(Sd**2 + sigma(Frel)**2)

WGTSHF

Weights to control excessive shifts

Field 1 Positional shifts magnitude restraint, PDEL <0.2>
weight = (cell dimension/PDEL)**2 for a, b, & c

2 Individual temperature factor shifts magnitude restraint (not used if t.f. overall), BDEL. <3.0>
weight = (1/BDEL)**2

3 Population parameter shifts magnitude restraint, QDEL. weight = (1/QDEL)**2 <0.1>

The following six fields are for setting sigmas associated with positional and thermal parameter restraints. They set sigmas for "tight", "medium", and "loose" restraints respectively.

4 Sigma for tight positional restraints <0.05>

5 Sigma for medium positional restraints <0.5>

6	Sigma for loose positional restraints	<2.0>
7	Sigma for tight thermal restraints	<0.5>
8	Sigma for medium thermal restraints	<1.0>
9	Sigma for loose thermal restraints	<0.0>

WGTSYM

Field 1 Weight scale for non-crystallographic symmetry restraints, WS. weight = (WS/SIGS)**2 WHAT IS SIGS??????????????????????????????? <1.0>

WGTTOR

Field 1	Weight scale for torsion angle restraints applied weight = ((Weight scale)/sigma(T))**2	<1.0>
2	Sigma(T) associated with a prespecified angle usually phi or psi of a regular secondary structure.	<15.0>
3	Sigma(T) associated with a planar angle such as omega.	<3.0>
4	Sigma(T) associated with a staggered angle such as chi 1.	<15.0>
5	Sigma(T) associated with an orthonormal angle such as chi 2 of aromatics.	<20.0>

NOREF

Field 1 - Atom identifiers for atoms which should not be refined in any way. Continue on more NOREF lines if need be.

EXAMPLE

Note that the program PROTIN must have been run before PROLSQ may be run in order that the appropriate constraint information be available on file F for use in the refinement process.

```
PROLSQ CRAMB   NCY 25  PCH ISO
PRINT   ALIS 1
STATS  5.0 4.0 3.0 2.0 1.0
RTEST    5  23   1   1   4    3    .25 .45 .65 1.25 .35 .65 .95
END
FINISH
```

In this example the atomic parameters of crambin are to be refined with individual isotropic temperature factors. Twenty five conjugant gradient cycles will be used to establish the parameter shifts, the refined coordinates will be written to unitpch: as well as to BDF B. The PRINT line specifies that a listing of the shift information for all atoms shall be written to the print file. The STATS line directs that R values be calculated for

shells of resolution as shown. The RTEST line will cause every fifth reflection to be written to a scratch BDF on file G and every twenty third reflection to be written on file H. For both of these files the sampling frequency will begin with the first reflection. Once refinement is complete these scratch files will be used to calculate R values for the subset of reflections contained in them. Four structure factor calculations will be made damping the positional parameter shifts of the atoms by .25, .45, .65, and 1.25. Then three runs will be made damping the isotropic B shifts by .35, .65, and .95. Because of the 23 in the second field of the R test line the structure factors, and therefore, the R values for the excluded reflections will also be calculated. This test using excluded reflections is a sensitive measure of the success of a refinement. If the R value of the excluded sub-set does not track downward with that of the included reflections it may well be a signal of a problem with the model or that the restraints are too slack. That is that the distribution of the restrained parameters about their "ideal value" is too great; i.e. the energy is too high.

REFERENCES

Finzel,B. C. Incorporation of Fast Fourier Transforms to speed Restrained Least-squares Refinement of Protein Structures. J. Appl. Cryst. (1987) 20, 53-55

Hendrickson, W. A. Stereochemically Restrained Refinement of Macromolecular Structures in Diffraction Methods for Biological Macromolecules Part B from Methods in Enzymology Edited by H. W. Wyckoff, C. H. W. Hirs, & S. N. Timasheff Academic Press, Inc. (1985)

Hendrickson, W. A., & Konnert, J. H. (1979) in Biomolecular Structure, Conformation, Function, and Evolution, edited by R. Srinivasan, Vol 1, pp. 43-57. Pergammon Press, New York

Konnert, J.H. A Restrained-Parameter Structure-factor Least-squares refinement Procedure for Large Asymmetric Units (1976). Acta Cryst. A32, 614-617.

Konnert, J. H. & Hendrickson, W. A. A Restrained-parameter Thermal-factor Refinement Procedure (1980). Acta Cryst. A36, 344-350.

Sheriff, S. Addition of Symmetry-related Contact Restraints to PROTIN & PROLSQ J. Appl. Cryst. (1987) 20, 55-57.

PHACMP: Compare The Phases of Two Sets of Reflections

James M. Stewart: Department of Chemistry/Biochemistry
University of Maryland, College Park, MD 20742

SYNOPSIS

The purpose of PHACMP is to allow the comparison of the values of the reflection phases determined in two different ways for the same set of reflections. For example, the phases determined by an MIR calculation with those determined by a structure factor calculation. Input may either be from two phase sources on the same BDF or from two different BDFs. If two files are used, the reflection set must have been sorted into the same order by the use of the XTAL program SORTRF.

METHOD

This program calculates the mean difference and the weighted mean difference in the phase values, in degrees, for reflections. The estimated standard deviations in the differences are also calculated. The calculations are done as a function of "d" spacing, resolution, in equal volumes of reciprocal space. In addition, counts are made of the values of the phases in both sets as a function of angle. The number of resolution intervals and the range of angle for counts are under user control. Statistics are done only for those reflections which have established phases in both sets. Reflections for which the value of the phase is undetermined in either set are rejected.

All input phases from whatever source are reduced to degrees modulo 360. The differences in phases are defined as arcos(cos(p1-p2)) in order to place them all in the range 0 to 180 degrees. Note that the sign of the difference is lost by this method.

The use of this program requires the specification of the two different phase types to be compared. This is accomplished by the use of the PHASO1 and PHASO2 lines (PHA SOurce) which specify the source of phase information in terms of the types specified to exist in an XTAL BDF. The recognized mnemonics are:

FC — The calculated structure factor
Items n700 or n801 and n802 in lrrefl:, n is the data set number

BEST m — The "best Fourier coefficient" for the "native" or "parent" substance
Items 701 and 702 + (m-1)*4, in lrrefl:
where m is the alternate solution no.

MOST m — The "most probable Fourier coefficient" for the "native" or "parent" substance
Items 701 and 702 + (m-1)*4, in lrrefl:
where m is the alternate solution no.

DENS — The calculated phase from density modification methods
Item n511 in lrrefl:, n is the data set number

DIR m — Direct methods structure factor estimate
Items n631 + (m-1)*2, in lrrefl:

where n is the data set number and m is the alternate solution no.

In the case where the two sets are to be taken from separate binary data files with different compound identities provision has been made to supply the alternate compound identity in the PHACMP input line.

PRINTER OUTPUT

All printed output is at priority 3.

FILE ASSIGNMENTS

Reads BDF from file .AA1 for one or both phase sets
Optionally reads BDF from file .AA2 for second phase set. If the compound identity on the second file is different than that on the first file the correct one must be given in the PHACMP input line.

CONTROL LINES

```
  PHACMP   Program initiation line
* MAXHKL   Maximum and minimum values of h, k, l, and sin(theta)/lambda
* PHASO1   Source of first phase set
* PHASO2   Source of second phase set for comparison
  END
```

* Indicates optional lines

PHACMP Program initiation line

Field	Function	Option code	Default
1	compound ID		
2-	List reflections modulo n	LIST n	<n=100>
	Take second set of phases from file .AA2 If the compound ID on file .AA2 is different than that on file .AA1 it must be supplied following BDF2	BDF2 s	<one file> <s=CMPID of .AA1>
	Set range of phase values in each tally "bin" in degrees	BIN x	<x=22.5>
	Set number of equal volume shells in reciprocal space for statistics	SHEL n	<n=4>
	Do not copy the BDF for the second phase set from s.000 to s.aa2 before starting the calculation.	NOCO	<copy it>

Does not apply if BDF2 not given.

MAXHKL Scope of data to be scanned

Field	Quantity	Default
1	maximum h	<from file AA1 lrdset:>
2	maximum k	<from file AA1 lrdset:>
3	maximum l	<from file AA1 lrdset:>
4	minimum sin(theta)/lambda	<from file AA1 lrdset:>
5	maximum sin(theta)/lambda	<from file AA1 lrdset:>
6	minimum h	<-1000>
7	minimum k	<-1000>
8	minimum l	<-1000>

PHASO1 and PHASO2 Sources of phase information

Field	Function	Option code	Default
1-	Function		
	Data set number	DS n	<n=1>
	Phase source		<FC>
	Calculated structure factors	FC	
	MIR "best" phases	BEST m	<m=1>
	MIR "most probable" phases	MOST m	<m=1>
	Density modification methods	DENS	
	Direct methods	DIR m	<m=1>
	m is the solution number		

EXAMPLE

```
TITLE Test of PHACMP  19 Oct 1989
PHACMP TRIDI LIST 50 BDF2 TRIDI1 BIN 11.25 SHEL 10
PHASO1 FC
PHASO2 FC
END
FINISH
```

In this example the two binary data files TRIDI.000 and TRIDI1.000 will be used for the comparison of their calculated structure factor phases. Every fiftieth reflection will be listed. The statistics will be made in ten equal volumes of reciprocal space. Counts will be made of the number of reflections having phase values in "bins" from -5.625 to 5.625, 5.625 to 16.875, ...,and 343.125 to 354.375 degrees. Counts of differences will be made up through 180 degrees.

James M. Stewart: Department of Chemistry/Biochemistry
University of Maryland, College Park, MD 20742

SYNOPSIS

The purpose of REFOUT is to allow the "punching" of selected reflection data from the standard BDF. For each reflection in a BDF an output line will be formed with h, k, l, and any other selected quantities present. Reflections for which any of the selected quantities are void will not be sent to the output file.

METHOD

This program runs in two modes. The "automatic" mode creates lines with h, k, l, |F(observed)|, |F(calculated)|, the phase in degrees, and sigma(F(relative)) as FORMAT(1X,3I5,3X,2F12.3,F12.1,F12.3). All input phases from whatever source are reduced to degrees modulo 360.

In the "user controlled" mode h, k, and l are written out in the same form, but all other quantities are chosen by use of an IDNUM input line. This line accepts ID numbers as described in the appendix on the binary data file for packets of record lrrefl:. For example F relative for the "parent" derivative is ID number 1304. The format of each quantity specified in the IDNUM input line may be specified, in XTAL notation, in a FORMAT input line. This line is optional if the format shown above is satisfactory. Finally, an optional SCALEQ input line allows the selected quantities to be multiplied by a supplied factor. The default is 1.0 for the SCALEQ quantities.

PRINTER OUTPUT

All printed output is at priority 3.

FILE ASSIGNMENTS

Reads BDF from file .AA1
Writes line images on .PCH

CONTROL LINES

```
   REFOUT   Program initiation line
*# IDNUM    lrrefl: ID numbers for selected quantities
*  FORMAT   Format, in XTAL notation for each selected quantity
*  SCALEQ   Multiplying scale factors for each selected quantity
   END
```

* Indicates optional lines
# This line switches program from "automatic" to "user controlled"

REFOUT Program initiation line

Field 1	c mpound ID		
2-	Function in "automatic" mode	Option code	Default
	List first n reflections	LIS n	<n=20>
	Scale F relative	SCA x.x	<x.x=1.0>
	Data set to be "punched"	DSE n	<n=1>

IDNUM Reflection record ID numbers for "user controlled" output
h, k, and l are always written by default. There are no other defaults.

Field	Quantity
1-n	Chosen ID numbers. First void field terminates the scan.

FORMAT Format of data for output
h, k, and l are always 1X,3I5 by default. The first void field, or a field less than 10001. terminates the scan.

Field	Quantity	Default
1	First beyond hkl	<311232.>
2	Second	<431232.>
3	Third	<551212.>
4	Fourth	<671232.>
5 -	Fifth and beyond	<uninitialized>

Note: An XTAL format consists of a real number of the form cccwwdt. Where ccc is the ending column for the item, ww is the width of the item, d is the number of places after the decimal point for reals or the minimum number of digits for integers; t is the type of output as 1 for E, 2 for F, 3 for I. For example the formats for h, k, and l are 60513., 110513., and 160513..

SCALEQ Scale to be applied to corresponding fields in FORMAT line.
Void fields are skipped as are fields less than |0.0001|.

Field		
1-n	Function	Default
	Multiplicative scale factor	<1.0>

EXAMPLE

```
TITLE TEST OF REFOUT  1 Dec 1989
REFOUT SALY
IDNUM 1600 501 1309
FORMAT 250742.  351031. 400513.
```



```
SCALEQ 1.5 1000. 10.
END
FINISH
```

In this example the normalized structure factor, E1 - item 1600, scaled by 1.5 is to be written F7.4 ending in column 25; the first interpolated scattering factor, f1 - item 501, scaled by 1000. is to be written E10.3 ending in column 35; and the scale group number, item 1309, scaled by 10 is to be written I5 ending in col 40. If ending columns before 16 had been specified h, k, and/or l would have been overwritten. Take care as very little screening of these format numbers is made by this program.

##