

The Protein Data Bank

A Computer-Based Archival File for Macromolecular Structures

Frances C. BERNSTEIN, Thomas F. KOETZLE, Grahame J. B. WILLIAMS, and Edgar F. MEYER Jr

Chemistry Department, Brookhaven National Laboratory, Upton, New York

Michael D. BRICE, John R. RODGERS, and Olga KENNARD

University Chemical Laboratory, Cambridge

Takehiko SHIMANOCHI and Mitsuo TASUMI

Faculty of Science, University of Tokyo, Hongo, Tokyo

The Protein Data Bank is a computer-based archival file for macromolecular structures. The Bank stores in a uniform format atomic co-ordinates and partial bond connectivities, as derived from crystallographic studies. Text included in each data entry gives pertinent information for the structure at hand (*e.g.* species from which the molecule has been obtained, resolution of diffraction data, literature citations and specifications of secondary structure). In addition to atomic co-ordinates and connectivities, the Protein Data Bank stores structure factors and phases, although these latter data are not placed in any uniform format. Input of data to the Bank and general maintenance functions are carried out at Brookhaven National Laboratory. All data stored in the Bank are available on magnetic tape for public distribution, from Brookhaven (to laboratories in the Americas), Tokyo (Japan), and Cambridge (Europe and worldwide). A master file is maintained at Brookhaven and duplicate copies are stored in Cambridge and Tokyo. In the future, it is hoped to expand the scope of the Protein Data Bank to make available co-ordinates for standard structural types (*e.g.* α -helix, RNA double-stranded helix) and representative computer programs of utility in the study and interpretation of macromolecular structures.

The Protein Data Bank¹ [1,2] was established in 1971 as a computer-based archival file for macromolecular structures. The purpose of the Bank is to collect, standardize, and distribute atomic co-ordinates and other data from crystallographic studies. As the number of solved protein and nucleic-acid structures has grown to the point where some 10⁷ characters are necessary to represent the co-ordinate information currently held, the need for such a computer-readable file has become very clear, and demands for the Bank's services have increased accordingly. The Protein Data Bank is one of several data base activities in the field of crystallography, *e.g.* the Bibliographic [3] and Structural [4] Data Files for organic and organometallic compounds, the Atlas of Macromolecular Structure on Microfiche (AMSOM) [5], the Bond Index to the Determination of Inorganic

Crystal Structures (BIDICS)² and the Powder Diffraction File³.

Scope

The Protein Data Bank covers atomic co-ordinates, structure factors and phases from diffraction studies of macromolecules. Since most of this information is not generally published in the primary literature, the Bank depends for comprehensiveness on data supplied directly by the investigators. It is essentially a depository of data, held in computer-readable form, in contrast to other data banks that are based on data abstracted from scientific publications. The Bank contains 77 atomic co-ordinate entries for 47 macromolecules (Table 1)⁴, and 13 sets of struc-

¹ Protein Data Bank is a misnomer of historical origin, since the file now contains entries for a nucleic acid.

² I. D. Brown, Bond Index to the Determination of Inorganic Crystal Structures, McMaster University, Hamilton, Ontario, Canada, L8S 4M1.

³ American Society for Testing Materials, 1916 Race Street, Philadelphia, Pennsylvania, U.S.A. 19103.

⁴ In addition to current co-ordinate entries shown in Table 1, the Bank contains obsolete entries (for adenylate kinase, tosyl α -chymotrypsin, concanavalin A, lactate dehydrogenase, horse methemoglobin, papain, rubredoxin, benzamidine-inhibited trypsin and pancreatic trypsin inhibitor), which have been superseded by later, more accurate data. These obsolete data are available on special request.

Table 1. Protein-Data-Bank holdings

IDENT CODE	MOLECULE	DEPOSITOR	STATUS CODE
1ADK	ADENYLATE KINASE	G. SCHULZ	A
1ADH	ALCOHOL DEHYDROGENASE (ADP-RIB)	C.-I. BRANDEN	
2ADH	ALCOHOL DEHYDROGENASE (ORTHOPHEN)	C.-I. BRANDEN	
2CHA	ALPHA-CHYMOTRYPSIN (TOSYL)	D. BLOW	R
3CHA	ALPHA-CHYMOTRYPSIN	A. TULINSKY	
1FAB	ANTIGEN BINDING FRAGMENT (NEW)	R. POLJAK	
1REI	BENCE-JONES IMMUNOGLOBULIN REI	O. EPP, R. HUBER	
1CPV	CALCIUM-BINDING PARVALBUMIN SET 6A	R. KRETSINGER	
2CPV	CALCIUM-BINDING PARVALBUMIN SET 6H	R. KRETSINGER	
3CPV	CALCIUM-BINDING PARVALBUMIN SET 6I	R. KRETSINGER	
1CAB	CARBONIC ANHYDRASE B	K. KANNAN	
1CAC	CARBONIC ANHYDRASE C	K. KANNAN	
1CPA	CARBOXYPEPTIDASE A	W. LIPSCOMB	
1CHG	CHYMOTRYPSINOGEN	J. KRAUT	
2CNA	CONCAVALIN A	G. REEKE, G. EDELMAN	N
3CNA	CONCAVALIN A	K. HARDMAN	R
1B5C	CYTOCHROME B5	F. S. MATHEWS	
1CYT	CYTOCHROME C (ALBACORE, OXIDIZED)	R. DICKERSON	
2CYT	CYTOCHROME C (ALBACORE, REDUCED)	R. DICKERSON	
1CYC	CYTOCHROME C (BONITO, HEART)	M. KAKUDO	
1C2C	CYTOCHROME C2	J. KRAUT	
155C	CYTOCHROME C550	R. TINKOVICH	
1EST	ELASTASE	H. WATSON	
1FDX	FERRDOXIN	L. JENSEN	
1FXN	FLAVODOXIN (CLOSTRIDIUM MP)	M. LUDWIG	
1GCH	GAMMA-CHYMOTRYPSIN	COHEN, DAVIES, SILVERTON	P
1GPD	GLYCERALDEHYDE-3-P-DEHYDROGENASE (LOBSTR)	M. ROSSMANN	N
2HMB	HEMOGLOBIN (HORSE, ADUO MET)	LADNER, HEIDNER, PERUTZ	RP
1DHB	HEMOGLOBIN (HORSE, DEOXY)	M. PERUTZ, G. FERM1	
1HMB	HEMOGLOBIN (HUMAN, DEOXY)	M. PERUTZ, G. FERM1	
1FDH	HEMOGLOBIN (HUMAN, FETAL, DEOXY)	J. FRIER	
1LHB	HEMOGLOBIN (LAMPREY)	W. HENDRICKSON	
1YHX	HEXOKINASE (YEAST) BIII	T. STEITZ	B
1HIP	HIGH POTENTIAL IRON PROTEIN	J. KRAUT	
2LDH	LACTATE DEHYDROGENASE	M. ROSSMANN	PD
3LDH	LACTATE DEHYDROGENASE/NAD/PYRUVATE	M. ROSSMANN	PD
1LYZ	LYSOZYME (HEN EGG-WHITE, SET W2)	R. DIAMOND	P
2LYZ	LYSOZYME (HEN EGG-WHITE, SET RS5D)	R. DIAMOND	P
3LYZ	LYSOZYME (HEN EGG-WHITE, SET RS6A)	R. DIAMOND	P
4LYZ	LYSOZYME (HEN EGG-WHITE, SET RS9A)	R. DIAMOND	P
5LYZ	LYSOZYME (HEN EGG-WHITE, SET RS12A)	R. DIAMOND	P
6LYZ	LYSOZYME (HEN EGG-WHITE, SET RS16)	R. DIAMOND	P
1MDH	MALATE DEHYDROGENASE	L. BANASZAK	A
1MBN	MYOGLOBIN (SPERM WHALE)	H. WATSON	
2MBN	MYOGLOBIN (SPERM WHALE, MET)	T. TAKANO	
3MBN	MYOGLOBIN (SPERM WHALE, DEOXY)	T. TAKANO	
3PTI	PANCREATIC TRYPSIN INHIBITOR	R. HUBER	R
8PAP	PAPAIN, NATIVE	J. DRENTH	R
2PAP	PAPAIN (ACE-ALA-ALA-PHE-ALA, CYS-25)	J. DRENTH	
3PAP	PAPAIN (CYS DERIV OF CYS-25)	J. DRENTH	
4PAP	PAPAIN (OXIDIZED CYS-25)	J. DRENTH	
5PAP	PAPAIN (TOS-LYS, CYS-25)	J. DRENTH	
6PAP	PAPAIN (BZOX-GLY-PHE-GLY, CYS-25)	J. DRENTH	
7PAP	PAPAIN (BZOX-GLY-PHE-ALA, CYS-25)	J. DRENTH	
1PGK	PHOSPHOGLYCERATE KINASE (YEAST)	H. WATSON	A
2PGK	PHOSPHOGLYCERATE KINASE (HORSE)	P. EVANS, D. PHILLIPS	B
1PAB	PREALBUMIN (HUMAN, PLASMA)	S. OATLEY, D. PHILLIPS	
1RNS	RIBONUCLEASE S	H. WYCKOFF	
2RXN	RUBREDOXIN	L. JENSEN	ND
1SNS	STAPHYLOCOCCAL NUCLEASE	F. A. COTTON, E. HAZEN	
1SGB	STREPTOMYCES GRISEUS PROTEINASE B	M. JAMES	A
1SBT	SUBTILISIN BPN'	J. KRAUT	
2SBT	SUBTILISIN NOVO	J. DRENTH	
1SOD	SUPEROXIDE DISMUTASE	J. AND D. RICHARDSON	A
1TLN	THERMOLYSIN (UNREFINED)	B. MATTHEWS	
2TLN	THERMOLYSIN (REFINED)	B. MATTHEWS	
1SRX	THIOREDOXIN	B.-O. SODERBERG	A
1TNA	TRANSFER RNA (YEAST, PHE)	J. SUSSMAN, S.-H. KIM	N
2TNA	TRANSFER RNA (YEAST, PHE)	M. SUNDARALINGAM	P
3TNA	TRANSFER RNA (YEAST, PHE)	JACK, LADNER, KLUG	P
1TIM	TRIOSE PHOSPHATE ISOMERASE	I. WILSON, D. PHILLIPS	
1PTN	TRYPSIN (NATIVE, PH8)	FEHLHAMMER, BODE, SCHWAGER	N
2PTB	TRYPSIN (BENZAMIDINE INHIBITED, PH7)	FEHLHAMMER, BODE, SCHWAGER	RN
1PTC	TRYPSIN/TRYPSIN INHIBITOR COMPLEX	BODE ET AL.	N

STATUS CODES

BLANK	STANDARD ENTRY AVAILABLE FOR DISTRIBUTION
A	ALPHA CARBON ATOMS ONLY
B	BACKBONE ONLY
D	NEW DATA HAS BEEN PROMISED
N	NEW ENTRY WITH DEPOSITOR FOR APPROVAL
P	IN PREPARATION
R	REPLACES AN OUT OF DATE PARAMETER SET

ture factors and phases. The atomic co-ordinate entries, which include descriptive text and partial bond connectivities, conform to a uniform format (see below), but the structure factors and phases

are stored in the format received from depositors. All co-ordinate entries are referred to depositors for verification, before being made available publicly through the Bank.

Record Structure of Atomic Co-ordinate Entries

Atomic co-ordinate entries consist of records each of 80 characters⁵. Using the punched-card analogy, columns 1 to 6 contain a record type identifier, and columns 7 to 70 contain data⁶. Columns 71 to 80 are normally blank, but may contain sequence information which is added by the library-file management program UPDATE⁷ used to maintain the file on the Brookhaven CDC CYBER 70/76 computing system. In order to facilitate retrieval of data from the file, the first four characters of each record define the unique record type, and the syntax of each record is independent of the order of records within any entry for a particular macromolecule. (In the master file, this order is always fixed.) Atomic co-ordinate data contributed by depositors are processed into the standard format with program MACMOL⁸, which also subjects the data to certain nomenclature and connectivity checking procedures.

A sample partial entry for the protein ribonuclease S is shown in Table 2⁹. The unique code 1 RNS identifying this entry is given in the HEADER record, along with the date these data were entered into the Bank, and a provisional classification based on function, intended for future use in indexing and subdividing the file. Text giving the name of molecule, species from which it has been obtained, authors, literature citations, and other general description are presented in records COMPND through REMARK. SEQRES gives the amino-acid sequence, and FTNOTE records are footnotes keyed to particular residues or atoms. Records HELIX through TURN describe the secondary structure as stated or approved by the depositor. Record CRYST1 defines the unit cell, while ORIGX and SCALE respectively give transformations relating the orthogonal Ångström co-ordinates stored in the file to those originally supplied by the depositor (these frequently are referred to an oblique or non-isometric system) and to standard crystallographic fractional co-ordinates. ATOM records give the IUPAC-IUB (1969) standard atom names [8], and residue abbreviations [9], along with

sequence identifiers (*cf.* SEQRES, above), co-ordinates in Ångström units, and occupancies and thermal motion factors, if these latter data are provided. Within each residue, atoms are ordered in a standard manner, starting with the backbone (N—C_α—C—O) and proceeding in increasing remoteness from the alpha carbon atom along the side-chain. C_α has been encoded CA, C_β as CB, *etc.* Where the sequence is known, but atoms have not been located in the structure analysis, gaps have been left in the atom serial numbers, to allow for future insertion. A TER record denotes an explicit chain-terminating residue. CONECT records give bond connectivity, for all atoms where the covalent connectivity is not specified completely by the atom name and order of serial numbers within the entry (*i.e.* the primary structure of standard residues). CONECT records may also be used to denote hydrogen bond and salt-bridge interactions. Each entry is terminated with a MASTER record, which gives checksums of the number of records, broken down by record type, and an END of data record.

Services

The activities of the Protein Data Bank are to collect and standardize data from laboratories engaged in the analysis of macromolecular structures, and to distribute these data within the scientific community. As a service to depositors, data are checked for errors of a clerical nature but no exhaustive verification is attempted. The Data Bank is located at Brookhaven National Laboratory and input of data to the Bank and general maintenance functions are carried out at Brookhaven¹⁰.

Duplicate copies of the Brookhaven master file are maintained at Cambridge and Tokyo. Data are available on magnetic tape for public distribution, from Brookhaven¹¹ (to laboratories in the Americas), Tokyo (Japan), and Cambridge (Europe and worldwide). The data are also available in a limited way within the United States over the Crystallographic Computing Network [10]. Retrieval programs to access data in the file have been described ([11]; M. Tasumi, personal communication). These efforts represent modest initial attempts at interactive access: up to the present the Protein Data Bank has served almost exclusively the demands of off-line users. A statistical summary of the Bank's activities is given

⁵ Originally, the Bank used a 140-character format, similar to that employed in the protein refinement programs of Diamond [6, 7]. The 140-character format has been superseded by the 80-character format.

⁶ A detailed description of the file formats is available from Brookhaven on request.

⁷ Control Data Corporation, UPDATE Reference Manual, Publication No. 60342500, Control Data Corporation, Arden Hills, Minnesota, 1974.

⁸ G. J. B. Williams, unpublished. For the 140-character data, program PROIN by E. F. Meyer was utilized.

⁹ The file is organized in a similar way for proteins and nucleic acids, although certain differences exist, *e.g.* with regard to details of atom and residue names.

¹⁰ Details of operations are announced periodically in a Newsletter (current issue is number 3, November 1976). Copies may be obtained from Brookhaven.

¹¹ Requests should be accompanied with a new 2400-ft reel of magnetic tape, and a check or purchase order for U.S. \$34.30 made to the order of Brookhaven National Laboratory, to cover postage and handling. This charge is subject to change in the future.

Table 2. *Abbreviated sample atomic co-ordinate entry (ribonuclease S)*

```

HEADER      HYDROLASE (PHOSPHORIC DIESTER, RNA)      01-APR-73      1RNS
COMPND      RIBONUCLEASE-S (E.C. 3.1.4.22)
SOURCE      BOVINE (BOS TAURUS) PANCREAS
AUTHOR      F. M. RICHARDS AND H. W. WYCKOFF
JRNL        R.J. FLETTERICK AND H. W. WYCKOFF, PRELIMINARY REFINEMENT
JRNL        OF PROTEIN COORDINATES IN REAL SPACE, ACTA CRYST., VOL. A31,
JRNL        P698 (1975).
REMARK      1
REMARK      1 REFERENCE 1. F. M. RICHARDS AND H. W. WYCKOFF, ATLAS OF
REMARK      1 STRUCTURES FOR MOLECULAR BIOLOGY, VOL. 1. RIBONUCLEASE-S,
REMARK      1 CLARENDON PRESS (1973).
REMARK      1 REFERENCE 2. F. M. RICHARDS AND H. W. WYCKOFF, BOVINE
REMARK      1 PANCREATIC RIBONUCLEASE, THE ENZYMES, EDITED BY P. D.
REMARK      1 BOYER, VOL. IV, THIRD EDITION, P647, ACADEMIC PRESS (1971)
REMARK      1 REFERENCE 3. F. M. RICHARDS, H. W. WYCKOFF, W. D. CARLSON,
REMARK      1 N. M. ALLEWELL, B. LEE AND Y. MITSUI, PROTEIN STRUCTURE,
REMARK      1 RIBONUCLEASE-S AND NUCLEOTIDE INTERACTIONS, COLD SPRING
REMARK      1 HARBOR SYMPOSIA ON QUANTITATIVE BIOLOGY, VOL. XXXVI, P35
REMARK      1 (1971).
REMARK      1 REFERENCE 4. N. M. ALLEWELL AND H. W. WYCKOFF,
REMARK      1 CRYSTALLOGRAPHIC ANALYSIS OF THE INTERACTION OF CUPRIC
REMARK      1 ION WITH RIBONUCLEASE S, J. BIOL. CHEM., VOL. 246, P4657
REMARK      1 (1971).
REMARK      1 REFERENCE 5. H. W. WYCKOFF, D. TSEHNOGLOU, A. W. HANSON,
REMARK      1 J. R. KNOX, B. LEE AND F. M. RICHARDS, THE THREE-
REMARK      1 DIMENSIONAL STRUCTURE OF RIBONUCLEASE-S. INTERPRETATION
REMARK      1 OF AN ELECTRON DENSITY MAP AT A NOMINAL RESOLUTION OF 2
REMARK      1 ANGSTROMS, J. BIOL. CHEM., VOL. 245, P305 (1970).
REMARK      1 REFERENCE 6. H. W. WYCKOFF, K. D. HARDMAN, N. M. ALLEWELL,
REMARK      1 T. INAGAMI, D. TSEHNOGLOU, L. N. JOHNSON AND F. M.
REMARK      1 RICHARDS, THE STRUCTURE OF RIBONUCLEASE-S AT 6 ANGSTROM
REMARK      1 RESOLUTION, J. BIOL. CHEM., VOL. 242, P3749 (1967).
REMARK      2
REMARK      2 RESOLUTION. 2.0 ANGSTROMS.
REMARK      3
REMARK      3 REFINEMENT. BY A STEEPEST-DESCENTS PROCEDURE. REFER TO THE
REMARK      3 JRNL CITATION ABOVE.
REMARK      4
REMARK      4 THIS COORDINATE SET IS DESIGNATED 6D BY THE DEPOSITOR.
REMARK      5
REMARK      5 THE *S-PEPTIDE* (RESIDUES 1-20) WHICH FORMS A SEPARATE
REMARK      5 CHAIN FROM THE REMAINDER OF THE MOLECULE IS GIVEN THE
REMARK      5 CHAIN IDENTIFIER S.
SEQRES      1 S 20 LYS GLU THR ALA ALA ALA LYS PHE GLU ARG GLN HIS MET
SEQRES      2 S 20 ASP SER SER THR SER ALA ALA
SEQRES      1 104 SER SER SER ASN TYR CYS ASN GLN MET MET LYS SER ARG
SEQRES      2 104 ASN LEU THR LYS ASP ARG CYS LYS PRO VAL ASN THR PHE
SEQRES      3 104 VAL HIS GLU SER LEU ALA ASP VAL GLN ALA VAL CYS SER
SEQRES      4 104 ALV LYS ASN VAL ALA CYS LYS ASN GLY GLN THR ASN CYS
SEQRES      5 104 TYR GLN SER TYR SER THR MET SER ILE THR ASP CYS ARG
SEQRES      6 104 ALU THR GLY SER SER LYS TYR PRO ASN CYS ALA TYR LYS
SEQRES      7 104 THR THR GLN ALA ASN LYS HIS ILE ILE VAL ALA CYS GLU
SEQRES      8 104 GLY ASN PRO TYR VAL PRO VAL HIS PHE ASP ALA SER VAL
FTNOTE      1
FTNOTE      1 THE MAIN CHAIN AND MOST OF THE ASSOCIATED SIDE CHAINS ARE
FTNOTE      1 NOT WELL-DEFINED IN THE REGIONS OF RESIDUES 2, 65-72 AND
FTNOTE      1 119-123.
FTNOTE      2
FTNOTE      2 THE MAIN CHAIN IS VERY POORLY DEFINED OR NOT VISIBLE AT ALL
FTNOTE      2 IN THE ELECTRON DENSITY MAP IN THE REGIONS OF RESIDUES 1,
FTNOTE      2 18-20, 21-23 AND 124.
HELIX       1 H1 THR 5 3 MET S 13 1
HELIX       2 H2 ASN 24 ASN 34 1

```

in Table 3. These statistics show rapid growth, both in number of holdings and in requests filled.

Future Developments

In the future, it is hoped to expand the scope of the Protein Data Bank to make available co-ordinates for standard structural types (e.g. α -helix, RNA double stranded helix) and representative computer programs of utility in the study of macromolecular structures. A small number of programs will be written

to calculate useful quantities derived from the co-ordinates (namely torsion angles, wire-model bender angles, full covalent connectivities, *etc.*). In addition to this software, the Bank will undertake to distribute contributed programs, provided that documentation is deposited in machine-readable form with source code¹². In order to facilitate use of atomic co-ordinate data to assemble models from standardized parts,

¹² The Bank will assume no responsibility for checkout or correction of errors in such deposited programs.

[illegible]

Year	Co-ordinate entries held at end of year	Co-ordinate entries distributed				Laboratories receiving data			
		Brookhaven	Cambridge	Tokyo	total	Brookhaven	Cambridge	Tokyo	total
1973	15	106	30	—	136	14	2	—	16
1974	17	99	102	—	201	14	6	—	20
1975	40	513	241	—	754	31	6	—	37
1976	69	1920	600	246	2766	47	14	9	70

it is intended to offer a 'model-builder's kit', to consist of header information, a compact list of co-ordinates to 0.1-Å (10-pm) precision, and torsion angles. This kit would be distributed principally in the form of printed listings, but also would be available on magnetic tape.

As the size of the Bank grows, possibilities for wide interactive access to the atomic co-ordinate data *via* a network will become more attractive. This type of access would increase the file's utility, particularly to those outside the fields of crystallography and analysis of protein and nucleic-acid conformations, by removing the necessity for repeated development of retrieval programs. For example, graphics terminals could be used to draw pictures of macromolecules for research or instructional purposes. Network access will be a topic of continued exploration in the next few years.

Suggestions regarding possible future improvements in Protein Data Bank services will be appreciated by the authors.

The late Walter C. Hamilton was one of the founders of this project. Helen M. Berman participated in the initial organization of the Data Bank and also was active in preparation of some of the first data entries as were Betty R. Davis and Daniel D. Jones. Numerous individuals have offered helpful suggestions and criticisms: Herbert J. Bernstein, David M. Blow, John S. Coggins, Robert Diamond, Anthony C. T. North, Michael G. Rossmann,

and David G. Watson. Richard J. Feldmann generously has transferred numerous atomic co-ordinate entries originally collected for inclusion in AMSOM. The members of the Protein-Data-Bank Advisory Group, David R. Davies, Kenneth Neet and Frederic M. Richards have overseen our operations and engaged in many useful discussions.

This work was performed under the auspices of the U.S. Energy Research and Development Administration and supported by the U.S. National Science Foundation under grants AG-370, GJ 33248X, DCR-75-07702, and PCM75-18956.

REFERENCES

1. Protein Data Bank (1971) *Nature New Biol.* 233, 223.
2. Protein Data Bank (1973) *Acta Crystallogr. Sect. B*, 29, 1746.
3. Kennard, O., Watson, D. G. & Town, W. G. (1972) *J. Chem. Doc.* 12, 14–19.
4. Allen, F. H., Kennard, O., Matherwell, W. D. S., Town, W. G. & Watson, D. G. (1973) *J. Chem. Doc.* 13, 119–123.
5. Feldmann, R. J. (1977) *Atlas of Macromolecular Structure on Microfiche*, Tracor Jitco Inc., Rockville.
6. Diamond, R. (1966) *Acta Crystallogr.* 21, 253–266.
7. Diamond, R. (1971) *Acta Crystallogr. Sect. B*, 436–452.
8. IUPAC-IUB Commission on Biochemical Nomenclature (1970) *Eur. J. Biochem.* 17, 193–201.
9. IUPAC-IUB Commission on Biochemical Nomenclature (1972) *Eur. J. Biochem.* 27, 201–207.
10. Koetzle, T. F., Andrews, L. C., Bernstein, F. C. & Bernstein, H. J. (1975) in *Computer, Networking and Chemistry* (Lykos, ed.) ACS Symposium Series, vol. 19, p. 1, American Chemical Society, Washington, D. C.
11. Meyer, E. F. (1974) *Biopolymers*, 13, 419–422.

F. C. Bernstein T. F. Koetzle* and G. J. B. Williams,
Chemistry Department, Brookhaven National Laboratory, Upton, New York, U.S.A. 11973

M. D. Brice, J. R. Rodgers, and O. Kennard* (External Staff, Medical Research Council),
Chemical Laboratories, University of Cambridge, Lensfield Road, Cambridge, Great Britain, CB2 1EW

E. F. Meyer, Department of Biochemistry and Biophysics,
Texas A. & M. University, College Station, Texas, U.S.A. 77843

T. Shimanouchi and M. Tasumi, Faculty of Science, University of Tokyo, Hongo, Tokyo, Japan 113

* To whom correspondence should be addressed.