Journal of Materials Science and Engineering B 4 (2) (2014) 55-63

A Tool for Interrogation of Macromolecular Structure

Francisco Torrens[1] and Gloria Castellano[2]

1. University Institute for Molecular Science, University of Valencia, Building of Institutes of Paterna, P. O. Box 22085, Valencia, 46071, Spain

2. Faculty of Veterinary and Experimental Sciences, Valencia Catholic University Saint Vincent Martyr, Guillem de Castro-94, Valencia 46001, Spain

Received: January 12, 2014 / Accepted: January 31, 2014 / Published: February 25, 2014.

Abstract: Our program BABELPDB allows browsing and interrogating the native and derived structural features of biomacromolecules using data obtained from the Protein Data Bank (PDB). Major features of BABELPDB are: (1) convert from PDB to other formats, (2) add or remove H-atoms, (3) strip the crystallization water molecules and (4) separate the α-carbons (C^{α}). The co-ordinates obtained with BABELPDB permit characterizing the presence of H-bonds. The algorithm for detecting H-bonds is implemented in our program TOPO for the theoretical simulation of the molecular shape. An example is given to illustrate the capabilities of the software: the calculation of the fractal dimension of the lysozyme molecule with (1.908) and without (1.920) H-atoms. The figures compare well with reference calculations performed with our version of program GEPOL and results from Pfeifer et al. For proteins, C^{α}-skeleton extracted with BABELPDB allows drawing the ribbon image, which determines their secondary structure.

Key words: Information retrieval, chemical structure, secondary structure, solvation water, carbon skeleton.

1. Introduction

The three-dimensional (3D) structure of a protein is critical to its function in biosystems. The availability of an increasing number of protein structures facilitated the teaching of protein chemistry. All biochemistry textbooks display selected 3D illustrations of protein structures. The structural data of proteins and other biomacromolecules are maintained by the Protein Data Bank (PDB), which can be accessed from http://www.rcsb.org/pdb or other mirror sites, e.g., Entrez, http://www.ncbi.nlm.nih.gov/Entrez [1, 2]. Tsai [3] described classroom applications of a freeware program, Windows-based PDB (WPDB), which compresses the structure files of PDB into a set of indexed files that can be retrieved, manipulated and analyzed locally [4, 5]. Structures of 3D can be displayed within the program or invoking freeware program RasMol [6].

Structure data on biomacromolecules as maintained by PDB are growing at a near exponential rate. The PDB contains 98 359 crystalline structures of proteins, nucleic acids and viruses, and complexes of these with small molecules. While trends in the price *vs.* performance of computer hardware make handling of large amounts of data manageable, software strategies for the efficient storage and retrieval of data are necessary. A number of strategies were employed for maintenance and querying of macromolecular structure data and fall into three broad categories according to the used storage method: indexed files as in WPDB, relational databases [7, 8] and object-oriented databases [9, 10]. Associated with every storage method are one or more query methods, e.g., structured query language (SQL) [11], macromolecular query language (MMQL) [12]. It is beyond the scope of the

Corresponding author: Francisco Torrens, Ph.D., lecturer, research fields: physical chemistry, theoretical chemistry. E-mail: torrens@uv.es.

present report to describe the advantages and disadvantages of each approach in detail; for further details *cf*. References [13, 14].

Our program BABELPDB includes procedures that allow the following options to examine a particular PDB structure: (1) convert from PDB to other formats; (2) add or remove H-atoms; (3) strip the water molecules of crystallization and (4) separate C -atoms. BABELPDB would seem particularly suited to educational purposes and an example of how it might be used is given.

2. Chemical Databanks

The databanks most used in chemistry are the Brookhaven PDB and the Cambridge Structural Data Bank (CSD) [15, 16]. The PDB is a computer-based archival file for macromolecular structures. It 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 was obtained, resolution of diffraction data, literature citations and specifications of secondary structure). In addition to atomic co-ordinates and connectivities, PDB stores structure factors and phases although these latter data are not placed in any uniform format. Input of data to PDB and general maintenance functions are carried out at Brookhaven National Laboratory. All data stored in PDB are available on magnetic tape and *ftp* for public distribution from Brookhaven, Tokyo and Cambridge. A master file is maintained at Brookhaven and duplicate copies are stored in Cambridge and Tokyo. The PDB can be accessed from *http://www.rcsb.org/pdb* or other mirror sites, *e.g*., Entrez, *http://www.ncbi.nlm.nih.gov/Entrez*. The scope of PDB was expanded to make available co-ordinates for standard structural types (*e.g*., α-helix, deoxyribonucleic acid double-stranded helix) and representative computer programs of utility in the study and interpretation of macromolecular structures.

The CSD comprises files of bibliographic, chemical connectivity and numeric structural data for organics, organometallics and metal complexes studied by X-ray and neutron diffraction. The files, covering the literature from 1935 and maintained on a current basis, contain information on more than 700,000 structural studies. Certain categories of information, particularly bibliographic, are disseminated in printed form via the Molecular Structures and Dimensions series. The full potential of CSD depends, however, on its response to specific user queries. The retrieved data may be used for extensive and systematic geometric analysis, and the visual display of crystal and molecular structures.

3. The Protein Data Bank

Every structure is addressed in a file whose name is coded as *i*ABC.BRK (*i* = 1,… 9), where *i* is PDB code; e.g., file 2LYM.BRK contains the co-ordinates of hen egg-white lysozyme. The header of this file is shown in Table 1.

Table 1 Header of the file 2LYM.BRK (lysozyme).

HEADER	HYDROLASE (O-GLYCOSYL) 08-JUN-87 2LYM	2LYM	3
COMPND	LYSOZYME (E.C.3.2.1.17) (1 ATMOSPHERE, 1.4 M NA*CL)	2LYM	4
SOURCE	HEN (GALLUS $GALLUS) EGG WHITE	2LYM	5
AUTHOR	C.E.KUNDROT,F.M.RICHARDS	2LYM	6
REVDAT	2 16-JUL-88 2LYMA 1 REMARK	2LYMA	1
REVDAT	1 16-OCT-87 2LYM 0	2LYM	7
JRNL	AUTH C.E.KUNDROT,F.M.RICHARDS	2LYM	8
JRNL	TITL CRYSTAL STRUCTURE OF HEN EGG-WHITE LYSOZYME AT A	2LYM	9
JRNL	TITL 2 HYDROSTATIC PRESSURE OF 1000 ATMOSPHERES	2LYM	10
JRNL	REF J.MOL.BIOL. V. 193 157 1987	2LYM	11
JRNL	REFN ASTM JMOBAK UK ISSN 0022-2836 070	2LYM	12

A PDB file has two parts. The first part contains the authors, group, secondary spatial structure and sequence. The second part contains the co-ordinates (X, Y, Z), atoms (i), ions, connectivities among atoms and Debye–Waller temperature factors (B); *X, Y, Z, B, i*. A simple partial entry for lysozyme is shown in Table 2.

4. Computational Method

Program BABEL implements a general framework for converting between file formats used for molecular modelling [17, 18]. Code BABEL will read the file types given in Table 3.

Code BABEL will write the file types listed in Table 4.

4.1 Using Program BABEL

Code BABEL may be invoked using command line options or menus. The menu interface can be accessed typing: babel –m.

In the command line input extensive online help is available. The command line input presents the following format:

babel [-v] -i<itype> <infile> [keywords] -o<out type> <outfile> [keywords2]

All arguments surrounded by *[]* are optional. The *-v* flag is also optional and is used to produce verbose output. The *–i* flag is used to set the input type. The input type codes that are currently supported are collected in Table 5.

The *-o* flag is used to set the output file type. The output type codes that are currently supported are resumed in Table 6.

For instance, to convert an MM2 output file named mm2.grf to a MOPAC internal co-ordinate input file

Table 2 Abbreviated sample atomic co-ordinate entry 2LYM (lysozyme).

HEADER	HYDROLASE (O-GLYCOSYL) 08-JUN-87 2LYM	2LYM 3
COMPND	LYSOZYME (E.C.3.2.1.17) (1 ATMOSPHERE, 1.4 M NA*CL)	2LYM 4
SOURCE	HEN (GALLUS $GALLUS) EGG WHITE	2LYM 5
AUTHOR	C.E.KUNDROT,F.M.RICHARDS	2LYM 6
REVDAT	2 16-JUL-88 2LYMA 1 REMARK	2LYMA 1
JRNL	AUTH C.E.KUNDROT,F.M.RICHARDS	2LYM 8
REMARK	1 AUTH C.E.KUNDROT,F.M.RICHARDS	2LYM 15
SEQRES	1 129 LYS VAL PHE GLY ARG CYS GLU LEU ALA ALA ALA MET LYS	2LYM 57
FTNOTE	1 SEE REMARK 4.	2LYM 68
FORMUL	2 HOH *151(H2 O1)	2LYM 69
HELIX	1 H1 GLY 4 GLY 16 1 RESIDUE 16 IS PARTIALLY 3/10	2LYM 70
SHEET	1 S1 2 LYS 1 PHE 3 0	2LYM 77
TURN	1 T1 LEU 17 TYR 20 TYPE II	2LYM 82
SSBOND	1 CYS 6 CYS 127	2LYM 95
CRYST1	79.170 79.170 37.960 90.00 90.00 90.00 P 43 21 2 8	2LYM 99
ORIGX1	1.000000 0.000000 0.000000 0.00000	2LYM 100
ORIGX2	0.000000 1.000000 0.000000 0.00000	2LYM 101
ORIGX3	0.000000 0.000000 1.000000 0.00000	2LYM 102
SCALE1	0.012631 0.000000 0.000000 0.00000	2LYM 103
SCALE2	0.000000 0.012631 0.000000 0.00000	2LYM 104
SCALE3	0.000000 0.000000 0.026344 0.00000	2LYM 105
ATOM	1 N LYS 1 3.280 10.157 10.354 1.00 12.97	2LYM 106
TER	1002 LEU 129	2LYM 1107
HETATM	1003 O HOH 130 -1.193 11.292 19.201 1.00 20.49	2LYM 1108
CONECT	48 47 981	2LYM 1259
MASTER	47 2 0 7 5 13 0 6 1152 1 8 10	2LYMA 9
END		2LYM 1268

Table 3 Types of files read by BABEL.

Alchemy	AMBER PREP	Ball and Stick
MSI BGF	Biosym .CAR	Boogie
Cacao Cartesian	Cambridge CADPAC	CHARMm
Chem3D Cartesian 1	Chem3D Cartesian 2	CSD CSSR
CSD FDAT	CSD GSTAT	Dock Database
Dock PDB	Feature	Free Form Fractional
GAMESS Output	Gaussian Z-Matrix	Gaussian 92 Output
Gaussian 94 Output	GROMOS96 (A)	GROMOS96 (nm)
Hyperchem HIN	MDL Isis SDF	M3D
Mac Molecule	Macromodel	Micro World
MM2 Input	MM2 Output	MM3
MMADS	MDL MOLfile	MOLIN
Mopac Cartesian	Mopac Internal	Mopac Output
PC Model	PDB	PS-GVB Input
PS-GVB Output	Quanta MSF	Schakal
ShelX	SMILES	Spartan
Spartan Semi-Empirical	Spartan Mol. Mechanics	Sybyl Mol
Sybyl Mol2	Conjure	UniChem XYZ
XYZ	XED	

Table 4 Types of files written by BABEL.

DIAGNOTICS	Alchemy	Ball and Stick
Batchmin Command	Cacao Cartesian	Cacao Internal
CAChe MolStruct	Chem3D Cartesian 1	Chem3D Cartesian 2
ChemDraw Conn. Table	Conjure	Conjure Template
CSD CSSR	Feature	Fenske-Hall Z-Matrix
Gamess Input	Gaussian Cartesian	Gaussian Z-matrix
Gaussian Z-matrix tmplt	Hyperchem HIN	Icon 8
IDATM	Mac Molecule	Macromodel
Micro World	MM2 Input	MM2 Ouput
MM3	MMADS	MDL Molfile
Mopac Cartesian	Mopac Internal	PC Model
PDB	Report	Spartan
Sybyl Mol	Sybyl Mol2	MDL Maccs
XED	UniChem XYZ	XYZ

named mopac.dat, the user would enter: babel -imm2out mm2.grf -oai mopac.dat.

In order to perform the above conversion with the keywords PM3 GEO-OK T = 30000 in the file mopac.dat the user would enter:

babel-imm2out mm2.grf-oai mopac.dat “PM3 GEO-OK T = 30000”

Notice the use of the double quotes around the keywords.

4.2 Hydrogen Addition and Deletion

Program BABEL has the ability to add and delete H atoms from any file format. The H-atoms can be added supplying the -h flag; H-atoms may be deleted supplying the –d flag; e.g., to add H-atoms to a CSD fractional co-ordinate file called input.cssr and output the file as an MOPAC internal co-ordinate input file named output.add, the user would type:

babel -icssr input.cssr -h -oai output.add.

Table 5 Input type codes currently supported by BABEL.

alc:	Alchemy file	prep:	AMBER PREP file
bs:	Ball and Stick file	bgf:	MSI BGF file
car:	Biosym .CAR file	boog:	Boogie file
caccrt:	Cacao Cartesian file	cadpac:	Cambridge CADPAC file
charmm:	CHARMm file	c3d1:	Chem3D Cartesian 1 file
c3d2:	Chem3D Cartesian 2 file	cssr:	CSD CSSR file
fdat:	CSD FDAT file	gstat:	CSD GSTAT file
dock:	Dock Database file	dpdb:	Dock PDB file
feat:	Feature file	fract:	Free Form Fractional file
gamout:	GAMESS Output file	gzmat:	Gaussian Z-Matrix file
gauout:	Gaussian 92 Output file	g94:	Gaussian 94 Output file
gr96A:	GROMOS96 (A) file	gr96N:	GROMOS96 (nm) file
hin:	Hyperchem HIN file	sdf:	MDL Isis SDF file
m3d:	M3D file	macmol:	Mac Molecule file
macmod:	Macromodel file	micro:	Micro World file
mm2in:	MM2 Input file	mm2out:	MM2 Output file
mm3:	MM3 file	mmads:	MMADS file
mdl:	MDL MOLfile file	molen:	MOLIN file
mopcrt:	Mopac Cartesian file	mopint:	Mopac Internal file
mopout:	Mopac Output file	pcmod:	PC Model file
pdb:	PDB file	psin:	PS-GVB Input file
psout:	PS-GVB Output file	msf:	Quanta MSF file
schakal:	Schakal file	shelx:	ShelX file
smiles:	SMILES file	spar:	Spartan file
semi:	Spartan Semi-Empirical file	spmm:	Spartan Mol. Mechanics file
mol:	Sybyl Mol file	mol2:	Sybyl Mol2 file
wiz:	Conjure file	unixyz:	UniChem XYZ file
xyz:	XYZ file	xed:	XED file

Table 6 Output type codes currently supported by BABEL.

diag:	DIAGNOTICS file	t:	Alchemy file
bs:	Ball and Stick file	bmin:	Batchmin Command file
caccrt:	Cacao Cartesian file	cacint:	Cacao Internal file
cache:	CAChe MolStruct file	c3d1:	Chem3D Cartesian 1 file
c3d2:	Chem3D Cartesian 2 file	d:	ChemDraw Conn. Table file
con:	Conjure file	contmp:	Conjure Template file
cssr:	CSD CSSR file	feat:	Feature file
fhz:	Fenske-Hall Z-Matrix file	gamin:	Gamess Input file
gcart:	Gaussian Cartesian file	g:	Gaussian Z-matrix file
gotmp:	Gaussian Z-matrix tmplt file	hin:	Hyperchem HIN file
icon:	Icon 8 file	i:	IDATM file
macmol:	Mac Molecule file	k:	Macromodel file
micro:	Micro World file	mi:	MM2 Input file
mo:	MM2 Ouput file	mm3:	MM3 file
mmads:	MMADS file	mdl:	MDL Molfile file
ac:	Mopac Cartesian file	ai:	Mopac Internal file
pc:	PC Model file	p:	PDB file
report:	Report file	spar:	Spartan file
mol:	Sybyl Mol file	mol2:	Sybyl Mol2 file
maccs:	MDL Maccs file	xed:	XED file
unixyz:	UniChem XYZ file	x:	XYZ file

In order to delete H-atoms from a Macromodel file named benzene.dat and output the file as an XYZ file named benzene.new the user would type:

babel -imacmod benzene.dat -d -ox benzene.new

Algorithm BABELPDB was written for computer-based search, retrieval, analysis and display of information from database PDB. Several options are allowed: (1) convert from PDB to other formats; (2) add or remove H-atoms. (3) strip the water molecules of crystallization and (4) keep only C -atoms. Code BABELPDB is available from the author (torrens@uv.es).

5. Calculation Results and Discussion

With program BABELPDB, the PDB co-ordinates of several proteins were converted to Cartesian co-ordinates and H-atoms were added. With these co-ordinates the presence of *H-bonds* was tested in the macromolecules [19-22]. The geometric analysis of H-bonds (X–H…Y), observed in crystal structure data retrieved from PDB, reveals lone-pair directionality, and H-acceptor separation, angle sublaid at H-atom (H), angle at acceptor atom (Y) and displacement of H-atom from a defined plane containing the lone-pair orbitals of the acceptor atom [23]. The H-bonds are characterized by the presence of H-bond interactions X–H…Y, where atoms *X* and *Y* are N, O, F or Cl, distance X–Y < 3.25Å and bond angle X–H–Y > 90°. The algorithm for detecting H-bonds was implemented in our program TOPO for the theoretical simulation of molecular shape [24-26]. Code TOPO allows calculating geometric descriptors and topological indices of macromolecules, *e.g.*, the *fractal dimension* of the *solvent-accessible surface*. After H-atoms were added with BABELPDB, the molecular image of lysozyme is shown (*cf*. Fig. 1). Protein lysozyme consists of 129 amino-acid residues (1906 atoms) and presents a molecular weight of 14 307Da. There are 151 water molecules around the enzyme.

The molecular image of lysozyme is displayed (*cf*. Fig. 2), after the solvation water molecules were stripped and H-atoms were added with BABELPDB.

For the experimentally well-studied enzyme lysozyme, the *fractal dimension D* was calculated with and without H-atoms [27, 28]. The calculation was performed using X-ray atomic co-ordinates of lysozyme (2LYM), extracted with BABELPDB [29]. For lysozyme with H-atoms the results show a *D* value of 1.908 [30]. This figure compares well with reference calculations performed with our version of program GEPOL (1.930), being the difference 1.15%. The fractal dimension averaged for *non-buried* (solvent-accessible) atoms *D'* results 2.201, which is greater than *D* by 15% indicating that the central atoms

Fig. 1 Lysozyme after H atoms were added with BABELPDB. *Cf*. a number of water molecules around.

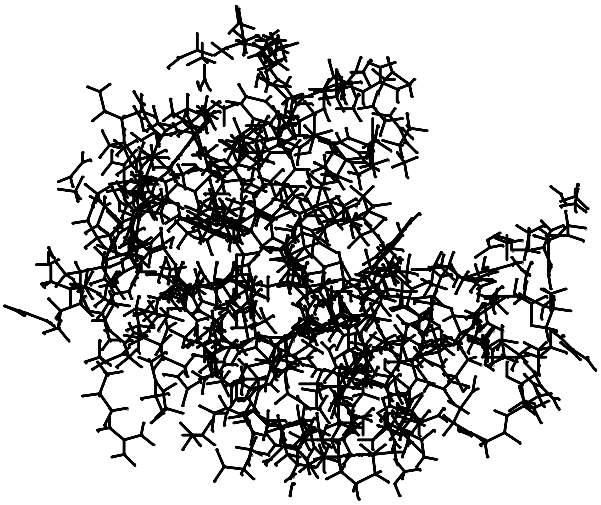

Fig. 2 Lysozyme after water molecules were stripped.

of the enzyme are buried. For lysozyme without H-atoms a similar trend is observed with all results increased by 0.6%.

Clementi et al. calculated the water solvent-accessible surface (SAS) quantum chemically as locus of the repulsive barrier in the interaction potential between the lysozyme molecule and water [31-33]. Notice that they based their calculation on X-ray data for the positions of the atoms. The SAS involved lysozyme conformation in the crystalline state. However, this is presumably no restriction because for lysozyme, the crystal structure analysis is known to provide an accurate picture of enzymatic action under native conditions. With these data presented in the form of molecular plots, Pfeifer et al. calculated for lysozyme a fractal surface dimension D = 2.17, using the silhouette and section variations of the box method [34]. Notice that our results for the lysozyme molecule compare well with Pfeifer et al.'s results, which are free of debate.

Lysozyme SAS can be compared with a self-avoiding random walk (SAW) surface. The fractal dimension results 1.908 on average corresponding to the short range of distances (1.25-3.5Å). This value can be compared with the fractal dimension corresponding to a 3D SAW. The SAW consists of identical rectilinear elements, one after the other and random oriented without attractions or repulsions among its elements, and its fractal dimension is 7/3. The SAW was proposed as a model of protein molecular surface [35]. Notice that at these short distances the fractal dimension for lysozyme is lower than SAW (1.883 < 2.333). The corresponding interpretation is that in the short range of distances, the molecule is more lengthened than an SAW because of steric repulsion between nearest atoms. Notice also how the idea of repulsive interaction in the range of short distances is translated in a difference in fractal coefficient, in comparison with the case without interactions.

Skeleton of C extracted from the lysozyme molecule with BABELPDB is shown (*cf*. Fig. 3).

The skeleton above allows drawing the ribbon image of lysozyme (*cf*. Fig. 4), where the ribbon links C -atoms. The ribbon image determines the elements of the secondary structure (α-helix, β-sheet, β-turn, etc.) [36].

The regions of helix and sheet above are summarized in Table 7. The four helical regions can be distinguished (three in Fig. 4, bottom and one in the middle). Three of them are distorted α-helices and the other is a 3.0_{10}-helix. Lysozyme contains one antiparallel β-sheet (Fig. 4, right). Finally, a disulphide linkage -S-S- between Cys-6 and Cys-127 joins both extremes.

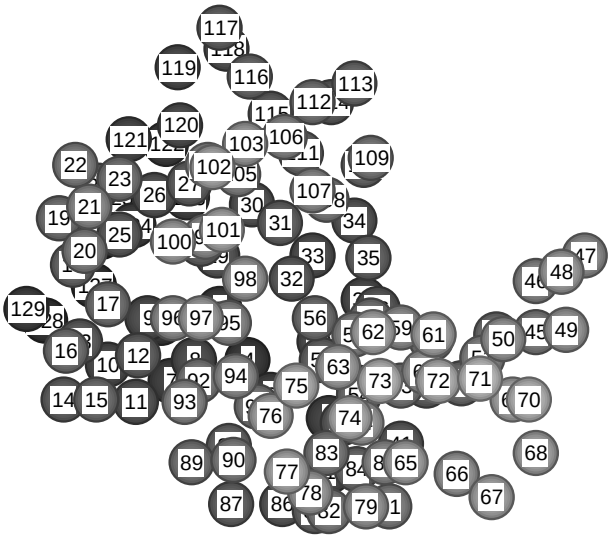

Fig. 3 C^{α} skeleton extracted from the lysozyme molecule.

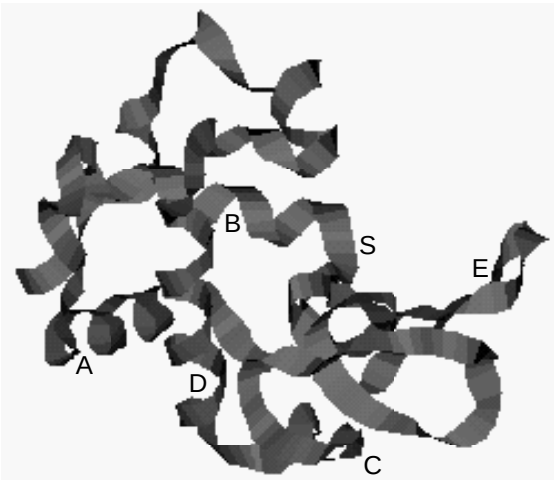

Fig. 4 Ribbon image of lysozyme linking C^{α} skeleton.

Table 7 The parameters of secondary structure regions in lysozyme.

Structure	Region	Type	Residue	Number	Percentage
Helix	A	α	5-15	11	8.5
	B	α	24-34	11	8.5
	C	3.0_{10}	80-85	6	5
	D	α	88-96	9	7
Total helix				37	29
β-sheet	E	antiparallel	41-54	14	11
Total helix+sheet				51	40
Total				129	100

6. Conclusions

From the preceding discussion the following conclusions can be drawn.

(1) Our program BABELPDB was written for the search, retrieval, analysis and display of information from database PDB. Several options are allowed: strip water molecules, separate C^{α}-atoms, etc.;

(2) The co-ordinates obtained with BABELPDB allowed characterizing the presence of H-bonds. The algorithm for detecting H-bonds was implemented in our program TOPO for the theoretical simulation of molecular shape;

(3) The fractal dimension of lysozyme was calculated with and without H-atoms. The figures compare well with reference calculations performed with our version of program GEPOL and with results from Pfeifer et al.;

(4) For proteins, C^{α}-skeleton extracted with BABELPDB allows drawing the ribbon image, which determines their secondary structure.

References

[1] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, et al., The Protein Data Bank: A computer-based archival file for macromolecular structures, J. Mol. Biol. 112 (1977) 535-542.

[2] R.A. Blevins, A. Tulinsky, Comparison of the independent solvent structures of dimeric α-chymotrypsin with themselves and with γ-chymotrypsin, J. Biol. Chem. 260 (1985) 8865-8872.

[3] C.S. Tsai, A computer-assisted tutorial on protein structure, J. Chem. Educ. 78 (2001) 837-839.

[4] I.N. Shindyalov, P.E. Bourne, WPDB-PC Windows-based interrogation of macromolecular structure, J. Appl. Crystallogr. 28 (1995) 847-852.

[5] I.N. Shindyalov, P.E. Bourne, Protein data representation and query using optimized data decomposition, CABIOS 13 (1997) 487-496.

[6] R.A. Sayle, E.J. Milner-White, RASMOL: Biomolecular graphics for all, Trends Biochem. Sci. 20 (1995) 374-376.

[7] M. Huysmans, J. Richelle, S.J. Wodak, SESAM: A relational database for structure and sequence of macromolecules, Proteins: Struct. Funct. Genet. 11 (1991) 59-76.

[8] S.A. Islam, M.J.E. Sternberg, A relational database of protein structures designed for flexible enquires about conformation, Protein Eng. 2 (1989) 431-442.

[9] P.M.D. Gray, N.W. Paton, G.J.L. Kemp, J.E. Fothergill, An object-oriented database for protein structure analysis, Protein Eng. 3 (1990) 235-243.

[10] I.N. Shindyalov, J. Cooper, W. Chang, P.E. Bourne, Design and use of a software framework to obtain information derived from macromolecular structure data, in: Proceedings of the 28th Annual Hawaii International Conference on System Sciences, IEEE Computer Society, Los Alamitos (CA), 1995, Vol. 5, pp. 207-216.

[11] C.J. Date, An Introduction to Database Systems, Addison Wesley, Reading (MA), 1990, p. 1.

[12] I.N. Shindyalov, W. Chang, C. Pu, P.E. Bourne, Macromolecular query language (MMQL): Prototype data model and implementation, Protein Eng. 7 (1994) 1311-1322.

[13] I.N. Shindyalov, P.E. Bourne, The shape of database to come, DEC Prof. 11 (1993) 38-44.

[14] W. Chang, I.N. Shindyalov, C. Pu, P.E. Bourne, Design and application of PDBlib, a C++ macromolecular class library, Comput. Appl. Biosci. 10 (1994) 575-586.

[15] F.H. Allen, S. Bellard, M.D. Brice, B.A. Cartwright, A. Doubleday, H. Higgs, et al., The Cambridge Crystallographic Data Centre: Computer-based search, retrieval, analysis and display of information, Acta Crystallogr. Sect. B 35 (1979) 2331-2339.

[16] O. Kennard, CCDC Newslett. (1990).

[17] P. Walters, M. Stahl, Program BABEL, University of

Arizona, Tucson (AZ), 1996.

[18] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: An open chemical toolbox, J. Chemoinf. 3 (2011) 33–1-14.

[19] L.C. Allen, A simple model of hydrogen bonding, J. Am. Chem. Soc. 97 (1975) 6921-6940.

[20] G. Alagona, C. Ghio, P. Kollman, Bifurcate vs. linear hydrogen bonds: Dimethyl phosphate and formate anion interactions with water, J. Am. Chem. Soc. 105 (1983) 5226-5230.

[21] G.L. Sosa, N. Peruchena, R.H. Contreras, E.A. Castro, Topological analysis of hydrogen bonding interactions involving C-H…O bonds, J. Mol. Struct. (THEOCHEM) 401 (1997) 77-85.

[22] D.E. Williams, Improved intermolecular force field for crystalline oxohydrocarbons including O-H…O hydrogen bonding, J. Comput. Chem. 22 (2001) 1-20.

[23] A. Vedani, J.D. Dunitz, Lone-pair directionality in hydrogen bond potential functions for molecular mechanics calculations: The inhibition of human carbonic anhydrase II by sulfonamides, J. Am. Chem. Soc. 107 (1985) 7653-7658.

[24] F. Torrens, E. Ortí, J. Sánchez-Marín, Vectorized TOPO program for the theoretical simulation of molecular shape, J. Chim. Phys. Phys.-Chim. Biol. 88 (1991) 2435-2441.

[25] F. Torrens, J. Sánchez-Marín, I. Nebot-Gil, Characterizing cavities in model inclusion molecules: A comparative study, J. Mol. Graphics Mod. 16 (1998) 57-71.

[26] F. Torrens, Characterizing cavities in model inclusion fullerenes: A comparative study, Int. J. Mol. Sci. 2 (2001) 72-89.

[27] K. Hamaguchi, K. Hayashi, Lysozyme, in: Proteins Structure and Function, Halsted, New York, 1972, Vol. 1, pp. 85-222.

[28] S. Blackburn, Enzyme Structure and Function, Dekker, New York, 1976, p. 1.

[29] C.E. Kundrot, F.M. Richards, Crystal structure of hen egg-white lysozyme at a hydrostatic pressure of 1000 atmospheres, J. Mol. Biol. 193 (1987) 157-170.

[30] F. Torrens, Calculation of partition coefficient and hydrophobic moment of the secondary structure of lysozyme, J. Chromatogr., A 908 (2001) 215-221.

[31] E. Clementi, G. Ranghino, R. Scordamaglia, Intermolecular pontentials: Interaction of water with lysozyme, Chem. Phys. Lett. 49 (1977) 218-224.

[32] G. Ranghino, E. Clementi, Enzyme-water interaction in solution and other properties related to computed net charges, Gazz. Chim. Ital. 108 (1978) 157-170.

[33] E. Clementi, G. Corongiu, M. Gratarola, P. Habitz, C. Lupo, P. Otto, et al., Theoretical and computational chemistry of complex systems: Solvation of DNA and proteins, Int. J. Quantum Chem. Symp. 16 (1982) 409-433.

[34] P. Pfeifer, U. Welz, H. Wippermann, Fractal surface dimension of proteins: Lysozyme, Chem. Phys. Lett. 113 (1985) 535-540.

[35] Y. Isogai, T. Itoh, Fractal analysis of tertiary structure of protein molecule, J. Phys. Soc. Jpn. 53 (1984) 2162-2171.

[36] W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577-2637.