A NEW TOOL FOR THE INTERROGATION OF MACROMOLECULAR STRUCTURE

Francisco Torrens[1] and Gloria Castellano[2]

[1] Institut Universitari de Ciència Molecular, Universitat de València, Edifici d'Instituts de Paterna, P. O. Box 22085, E-46071 València, Spain

[2] Facultad de Veterinaria y Ciencias Experimentales, Universidad Católica de Valencia *San Vicente Mártir*, Guillem de Castro-94, E-46001 València, Spain

Our program BABELPDB allows browsing and interrogating the native and derived structural features of biological macromolecules 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 coordinates obtained with BABELPDB permit characterizing the presence of hydrogen bonds (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 numbers compare well with reference calculations performed with our version of the GEPOL program and with the results from Pfeifer *et al.* For proteins, the C^{α} skeleton extracted with BABELPDB allows drawing the ribbon image, which determines the secondary structure of proteins.

Keywords: Information retrieval, Chemical structure, Secondary structure, Solvation water, α-Carbon skeleton, Hydrogen bond, Molecular shape, Fractal dimension, Protein database, Protein Data Bank, Property visualization

INTRODUCTION

The three-dimensional (3D) structure of a protein is critical to its function in biological systems. The availability of an increasing number of protein structures facilitated greatly 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 such as Entrez, *http://www.ncbi.nlm.nih.gov/Entrez*.[1,2] Tsai[3] described classroom applications of a freeware program, WPDB, which compresses the structure files of the PDB into a set of indexed files that can be retrieved, manipulated and analyzed locally.[4,5] The 3D structures can be displayed within the program or invoking freeware program RasMol.[6]

Structure data on biological macromolecules as maintained by PDB are growing at a near exponential rate. The PDB contains *ca.* 70 000 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 such large amounts of data manageable, at least for the next few years, software strategies for the efficient storage and retrieval of these data are necessary. A number of such 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 each storage method are one or more query methods, *e.g.*, SQL[11] and MMQL.[12] It is beyond the scope of the present paper to describe the advantages and disadvantages of each approach in detail; for further details see References 13 and 14.

Our program BABELPDB includes subprograms that allows 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^{α} atoms. BABELPDB would seem particularly suited to educational purposes and an example of how it might be used is given.

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 coordinates 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 coordinates 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 such as Entrez, *http://www.ncbi.nlm.nih.gov/Entrez*. Now, the scope of PDB has been expanded to make available coordinates 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.

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, presently contain information on more than 28 000 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 then be used for extensive and systematic geometric analysis and the visual display of crystal and molecular structures.

THE PROTEIN DATA BANK

Each structure is addressed in a file whose name is coded as *i*ABC.BRK (*i* = 1,… 9), where *i* is PDB code. For instance, the file 2LYM.BRK contains the coordinates of hen egg-white lysozyme. The header of this file is shown in Table 1. A PDB file has two parts. The first part contains the authors, group, secondary spatial structure and sequence. The second part contains the coordinates (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.

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
```

Table 2. Abbreviated sample atomic coordinate 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                                               2LYM1107
HETATM 1003  O   HOH   130      -1.193  11.292  19.201  1.00 20.49       2LYM1108
CONECT   48   47  981                                                     2LYM1259
MASTER       47    2    0    7    5   13    0    6 1152    1    8   10   2LYMA  9
END                                                                       2LYM1268
```

COMPUTATIONAL METHOD

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

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	

BABEL will write the file types listed in Table 4.

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

Using Program BABEL

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 has the following format:

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

All arguments surrounded by *[]* are optional. The *-v* flag is 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.

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

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

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

For instance, to convert an MM2 output file named *mm2.grf* to a MOPAC internal coordinate input file named *mopac.dat*, the user would enter:

babel -imm2out mm2.grf -oai mopac.dat

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.

Hydrogen Addition/Deletion

BABEL has the ability to add and delete H atoms from any file format. 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 coordinate file called *input.cssr* and output the file as a MOPAC internal coordinate input file named *output.add*, the user would type:

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

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

The program BABELPDB has been 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^{α} atoms. BABELPDB is available from the author (Francisco.Torrens@uv.es).

CALCULATION RESULTS AND DISCUSSION

With program BABELPDB, the PDB coordinates of several proteins have been converted to Cartesian coordinates and H atoms have been added. With these coordinates the presence of *H-bonds* has been tested in the macromolecules.[18–21] The geometric analysis of H-bonds (X–H…Y), observed in crystal structure data retrieved from PDB, reveals lone-pair directionality as well as the H–acceptor separation, the angle sublaid at H atom (H), the angle at the acceptor atom (Y) and the displacement of H atom from a defined plane containing the lone-pair orbitals of the acceptor atom.[22] 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 has been implemented in our program TOPO

for the theoretical simulation of molecular shape.[23–25] TOPO allows calculating geometric descriptors and topological indices of macromolecules, including the *fractal dimension* of the *solvent-accessible surface*. After H atoms have been added with BABELPDB, the molecular image of lysozyme is shown in Figure 1. Protein lysozyme consists of 129 amino-acid residues (1906 atoms) and has a molecular weight of 14 307Da. There are 151 water molecules around the enzyme protein.

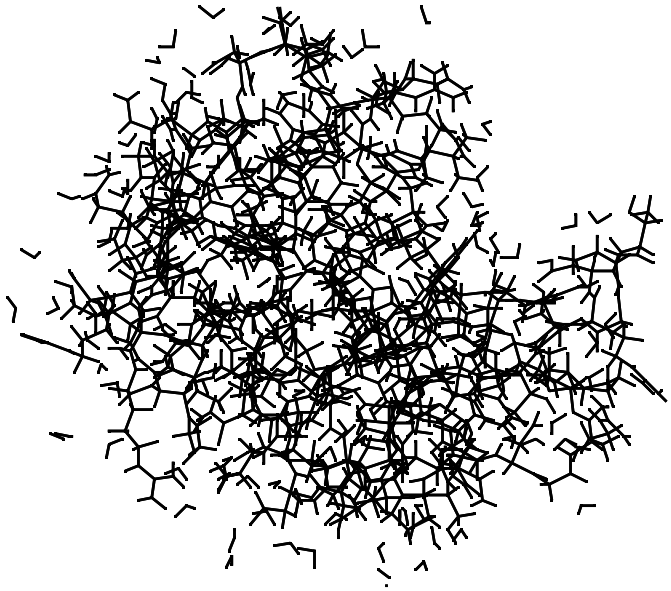

Figure 1. Lysozyme molecule after H atoms have been added with program BABELPDB. See a number of water molecules around the enzyme.

The molecular image of lysozyme is displayed in Figure 2, after the solvation water molecules have been stripped, and H atoms have been added with program BABELPDB.

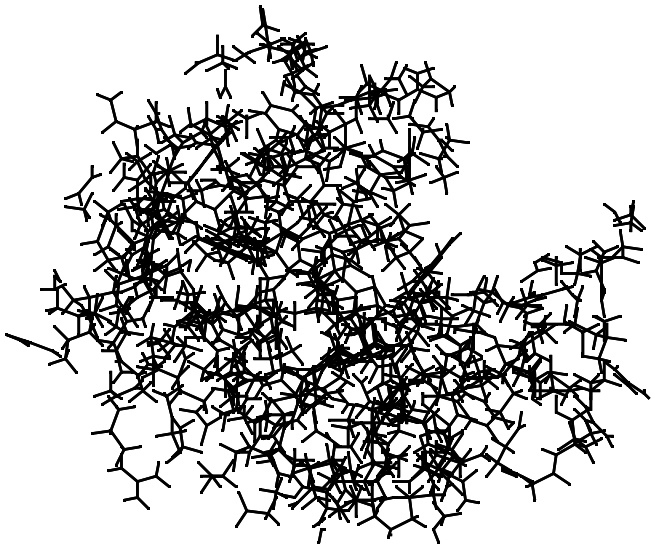

Figure 2. Lysozyme molecule after the water molecules have been stripped.

For the experimentally well-studied enzyme lysozyme, the *fractal dimension* D has been calculated with and without H atoms.[26,27] The calculation has been performed using X-ray atomic coordinates of lysozyme (2LYM), extracted with BABELPDB.[28] For lysozyme with H atoms, the results show a D value of 1.908.[29–31] This number compares well with reference calculations carried out 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, so, D' is greater than D by 15% indicating that the central atoms 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-accessible surface* quantum chemically as locus of the repulsive barrier in the interaction potential between the lysozyme molecule and water.[32–34] Notice that they based their calculation on X-ray data for the positions of the atoms. Therefore, the surface under consideration 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*.[35] From the preceding discussion it is clear that our results for the lysozyme molecule compare well with Pfeifer *et al.*'s results, which are free of debate.

The solvent-accessible surface of lysozyme can be compared with a self-avoiding random 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 self-avoiding random surface. This surface 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. This surface was proposed as a model of protein molecular surface.[36] Notice that, at these short distances, the fractal dimension for lysozyme is lower than that for the random surface (1.883 < 2.333). The corresponding interpretation is that, in the short range of distances, the molecule is more lengthened than a random surface, due to *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.

C^{α} skeleton extracted from the lysozyme molecule with BABELPDB is shown in Figure 3. This skeleton allows drawing the ribbon image of lysozyme (Figure 4), where the ribbon links C^{α} atoms. The ribbon image determines the elements of the secondary structure (α-helix, β-sheet, β-turn, *etc*.).[37] The regions of helix and sheet are summarized in Table 7. The four helical regions can be distinguished (three in Figure 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 (Figure 4, *right*). Finally, a disulphide linkage between Cys-6 and Cys-127 joins both extremes.

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

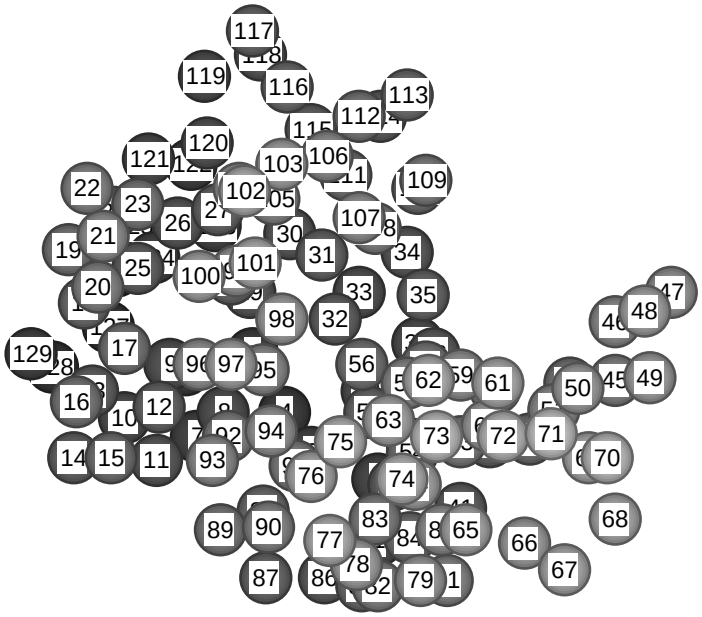

Figure 3. C^{α} skeleton extracted from the lysozyme molecule.

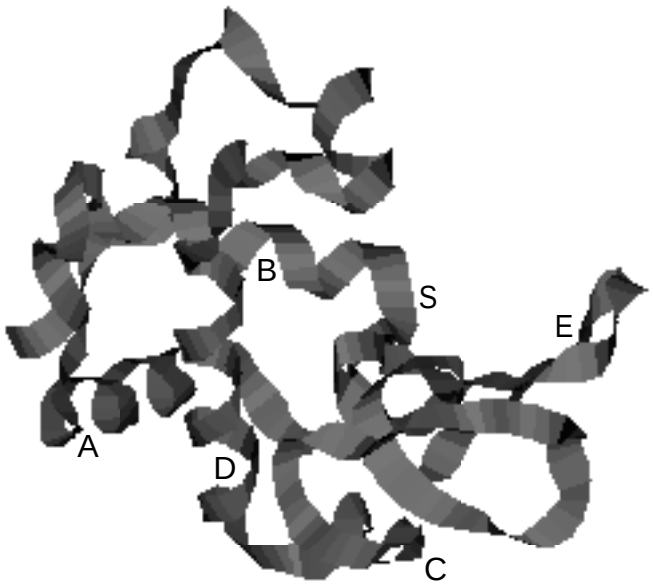

Figure 4. Ribbon image of lysozyme. The ribbon links C^{α} skeleton.

CONCLUSIONS

From the preceding discussion the following conclusions can be drawn.

1. Our program BABELPDB has been 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 coordinates obtained with BABELPDB have allowed characterizing the presence of H-bonds. The algorithm for detecting H-bonds has been implemented in our program TOPO for the theoretical simulation of molecular shape.

3. The fractal dimension of lysozyme has been calculated with and without H atoms. The figures compare well with reference calculations

carried out 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 the secondary structure of proteins.

ACKNOWLEDGEMENTS

The authors acknowledge financial support from the Spanish MEC DGI (Project No. CTQ2004-07768-C02-01/BQU), Generalitat Valenciana (DGEUI INF01-051 and INFRA03-047, and OCYT GRUPOS03-173) and Universitat de València–Mediscovery.

REFERENCES

[1] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol.* **112** (1977) 535.

[2] R. A. Blevins and A. Tulinsky, *J. Biol. Chem.* **260** (1985) 8865.

[3] C. S. Tsai, *J. Chem. Educ.* **78** (2001) 837.

[4] I. N. Shindyalov and P. E. Bourne, *J. Appl. Crystallogr.* **28** (1995) 847.

[5] I. N. Shindyalov and P. E. Bourne, *CABIOS* **13** (1997) 487.

[6] R. A. Sayle and E. J. Milner-White, *Trends Biochem. Sci.* **20** (1995) 374.

[7] M. Huysmans, J. Richelle and S. J. Wodak, *Proteins: Struct., Funct., Genet.* **11** (1991) 59.

[8] S. A. Islam and M. J. E. Sternberg, *Protein Eng.* **2** (1989) 431.

[9] P. M. D. Gray, N. W. Paton, G. J. L. Kemp and J. E. Fothergill, *Protein Eng.* **3** (1990) 235.

[10] I. N. Shindyalov, J. Cooper, W. Chang and 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, Maui, 1995, p. 207.

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

[12] I. N. Shindyalov, W. Chang, C. Pu and P. E. Bourne, *Protein Eng.* **7** (1994) 1311.

[13] I. N. Shindyalov and P. E. Bourne, *DEC Prof.* **11** (1993) 38.

[14] W. Chang, I. N. Shindyalov, C. Pu and P. E. Bourne, *Comput. Appl. Biosci.* **10** (1994) 575.

[15] F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, W. D. S. Motherwell, J. R. Rodgers and D. G. Watson, *Acta Crystallogr., Sect. B* **35** (1979) 2331.

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

[17] P. Walters and M. Stahl, Program BABEL, University of Arizona, Tucson (AZ), 1996.

[18] L. C. Allen, *J. Am. Chem. Soc.* **97** (1975) 6921.

[19] G. Alagona, C. Ghio and P. Kollman, *J. Am. Chem. Soc.* **105** (1983) 5226.

[20] G. L. Sosa, N. Peruchena, R. H. Contreras and E. A. Castro, *J. Mol. Struct. (Theochem)* **401** (1997) 77.

[21] D. E. Williams, *J. Comput. Chem.* **22** (2001) 1.

[22] A. Vedani and J. D. Dunitz, *J. Am. Chem. Soc.* **107** (1985) 7653.

[23] F. Torrens, E. Ortí and J. Sánchez-Marín, *J. Chim. Phys. Phys.-Chim. Biol.* **88** (1991) 2435.

[24] F. Torrens, J. Sánchez-Marín and I. Nebot-Gil, *J. Mol. Graphics Mod.* **16** (1998) 57.

[25] F. Torrens, *Int. J. Mol. Sci.* **2** (2001) 72.

[26] K. Hamaguchi and K. Hayashi, Lysozyme, in: *Proteins Structure and Function*, M. Funatsu, K. Hiromi, K. Imahori, T. Murachi and K. Narita, Eds., Halsted, New York, 1972, vol. 1, ch. 2, p. 85.

[27] S. Blackburn, *Enzyme Structure and Function*, Dekker, New York, 1976.

[28] C. E. Kundrot and F. M. Richards, *J. Mol. Biol.* **193** (1987) 157.

[29] F. Torrens, *J. Chromatogr., A* **908** (2001) 215.

[30] F. Torrens, *J. Chromatogr., A* (submitted for publication).

[31] F. Torrens, *Anal. Sci. (Tokyo)* (submitted for publication).

[32] E. Clementi, G. Ranghino and R. Scordamaglia, *Chem. Phys. Lett.* **49** (1977) 218.

[33] G. Ranghino and E. Clementi, *Gazz. Chim. Ital.* **108** (1978) 157.

[34] E. Clementi, G. Corongiu, M. Gratarola, P. Habitz, C. Lupo, P. Otto and D. Vercauteren, *Int. J. Quantum Chem., Symp.* **16** (1982) 409.

[35] P. Pfeifer, U. Welz and H. Wippermann, *Chem. Phys. Lett.* **113** (1985) 535.

[36] Y. Isogai and T. Itoh, *J. Phys. Soc. Jpn.* **53** (1984) 2162.

[37] W. Kabsch and C. Sander, *Biopolymers* **22** (1983) 2577.