Computer Physics Communications 60 (1990) 257–263
North-Holland

A three-dimensional protein graphic program *

Donghao R. Lu and Kinam Park
School of Pharmacy, Purdue University, West Lafayette, IN 47907, USA

Received 24 October 1989; in revised form 12 January 1990

A FORTRAN program was developed to generate the three-dimensional vector graphics of protein molecules. The program used DISSPLA subroutines (Computer Associates International Inc.) to interface the computer system for graphic display. Two data formats supported by the program were the X-ray crystallographic protein data and the structure data from the program PROADS. A protein molecule was interactively rotated or enlarged for display. The program provides a simple solution to the need of protein graphics.

PROGRAM SUMMARY

Title of program: PROPLT

Catalogue number: ABTB

Program obtainable from: CPC Program Library, Queen's University of Belfast, N. Ireland (see application form in this issue)

Computer: IBM 3090-180E; *Installation*: Purdue University Computing Center, West Lafayette, In 47907, USA

Operating system: VM/SP CMS

Programming language used: FORTRAN 77

High speed storage required: 32000 words

No. of bits in a word: 60

Peripherals used: disc storage for input and output files, line printer for output and display device for protein structure plotting

No. of lines in combined program and test deck: 5348

Keywords: protein structure, protein graphics

Nature of physical problem
The three-dimensional structure of a protein is displayed by a vector graphic algorithm. A protein molecule can be interactively rotated to show a three-dimensional image in any orientation. The protein molecule can also be enlarged allowing the user to focus on a specific part of the molecule. Two data formats are supported by the program. They are the X-ray crystallographic protein structure coordinates from Brookhaven Protein Data Bank in New York, USA [1], and the protein structure data format of the protein adsorption simulation program PROADS.

Method of solution
Rotation of a protein molecule in any given orientation corresponding to two rotational angles is carried out by rotating the Cartesian coordinates of the X-ray crystallographic structure of the protein or the protein structure from the program PROADS. Determination of the drawing algorithm for each amino acid residue and hydrogen atoms is based on the specific data arrangement of the two data formats. A three-dimensional image-drawing strategy is used in which the thick bars represent the chemical bonds which are in the front and the thin bars represent the bonds which are away from the front. The program uses the DISSPLA subroutines to interface a mainframe computer and a color display device, such as a color graphic terminal to display the image.

Restrictions on the complexity of the problem
The protein structure data is limited to the two formats mentioned above. The DISSPLA subroutines are needed for the image display. It should be noted that the DISSPLA sub-

* This study was supported by the National Heart, Lung and Blood Institute of the National Institutes of Health through Grant HL 39081.

0010-4655/90/$03.50 © 1990 – Elsevier Science Publishers B.V. (North-Holland)

routines are available in the mainframe computer system installed in most universities for public use in the United States.

Typical running time
About 0.5 min, depending on the job loading on a mainframe computer system.

References
[1] Protein Data Bank, Brookhaven National Laboratory, Brookhaven, New York.
[2] D.R. Lu and K. Park, in preparation.

LONG WRITE-UP

1. Theory

The protein modelling techniques have become an essential tool in many research areas. It is known that the relationship between a protein structure and a biological function is important. The atomic coordinates of many proteins have been determined by X-ray crystallography and the coordinates can be readily obtained from the Brookhaven Protein Data Bank [1]. The computer facilities to visualize the three-dimensional structures of proteins, however, are not available in many places due to the expensive software and equipments. We have developed a FORTRAN program which generates images of three-dimensional structures of a protein on an IBM mainframe computer system.

There are usually two types of computer graphic displays for proteins. Raster graphics depict the atoms as overlapping spheres based on the size of their van der Waals radii, that is, the orbits of their outer electrons. These displays are called space-filling models. They describe steric, or spatial, constraints but leave it to the viewer to infer other interactions. Vector graphics, on the other hand, describe each atom in vector coordinates, highlighting specified properties of the molecule. The raster display focuses attention on the surface of the molecule (the inner atoms are hidden from view) and the packing of atoms. Vector displays permit views on the interior of the molecule. Since vector graphics is the simpler visual model and can easily view the interior arrangement of proteins, we have chosen vector graphics for the display of protein structures. The FORTRAN program, along with the DISSPLA graphic subroutines, was designed to support the following two data formats: (a) the X-ray crystallographic protein structure coordinates from Brookhaven Protein Data Bank; and (b) the data format of the protein adsorption simulation program PROADS [2]. The DISSPLA is a set of FORTRAN callable subroutines which allows one to write a FORTRAN program calling the subroutines to interface a computer system for specific graphics. The DISSPLA subroutines are available in mainframe computer system installed in most universities in the United States.

The program can display a protein molecule in a three-dimensional image: the thick bars represent the chemical bonds which are in the front and the thin bars represent the bonds which are away from the front. A protein molecule can be interactively rotated with two rotational angles round the Y and Z axes to be viewed in any orientation. The protein molecule can also be scaled (i.e. enlarged) to observe specific parts of the molecule in detail. A color image can be generated on many color display devices, in which the blue color represents protein backbones, the red color represents the chemical bonds for carbonyl groups in peptide bonds, the green color represents the bonds for each amino acid residue on protein side chains, and optionally, the black color represents the bonds for hydrogen atoms. A number of display devices can be used for protein structure plotting, such as a color graphic terminal or a four pen color plotter.

2. Data

2.1. The X-ray crystallographic data format of the Brookhaven Protein Data Bank

The coordinates of X-ray crystallographic structures of proteins are obtained from the

Brookhaven Protein Data Bank in a computer-readable format. All records in the original files are deleted except the records ATOM and the TER lines before the files are used for graphic plotting. The records ATOM contain atomic coordinate records for all amino acid residues and the TER lines are chain terminators. Since the coordinates of hydrogen atoms are not available in the data from the protein data bank, the protein molecule is displayed with protein backbones and amino acid residue side chains without hydrogen atoms.

2.2. The data format of the protein adsorption simulation program PROADS

The coordinate data generated from the protein adsorption simulation program PROADS can directly be used for graphic plotting. The chemical bonds for all hydrogen atoms can be optionally displayed.

3. Program description

3.1. Input

The program reads two files. They are the coordinate file, DISPRO, for protein structure, and the command file, COMDIS, for specific plotting instruction. The DISPRO file (in two data formats) is read from unit 10 and organized as follows:

(a) The X-ray crystallographic data format:

Card 1

Data: ATM, ATMNAM, RESNAM, CHANID, SEQNUM, XCO, YCO, ZCO.

Format: (A4, 9X, A4, A3, 1X, A1, I4, 4X, 3F8.3).

Details: record indicator, name of the atom, name of the residue, chain identifier, residue sequence number, *X*, *Y*, and *Z* coordinates for the atom, respectively.

Read repeatedly until the TER line is read or the end of file is reached.

Card 2

Data: TER.

Format: (A4, 49X, I1).

Details: chain terminator.

Card 1 and card 2 are repeated until the end of file is reached.

(b) The protein structure data format of program PROADS:

Card 1

Data: NUMRES.

Format: (I3).

Details: number of the total amino acid residues in the protein.

Card 2

Data: NAME.

Format: (A3).

Details: name of the amino acid residue.

Card 3

Data: NATOM.

Format: (I3).

Details: number of atoms in the amino acid residue.

Card 4 (loop reading (I = 1, N atom)

Data: NAATM(I), ZATOM(I), CLASS(I), XCH(I), YCH(I), ZCH(I).

Format: (A4, 4X, I3, 1X, I3, 11X, 3(1X, F10.5)).

Details: the name, nuclear charge, class parameter, x, y and z coordinates of each atom in the residue.

Card 2 to card 4 are repeated until all the amino acid residues are read.

The COMDIS file contains the instruction command for different data formats, display rotation angles (in radian), hydrogen atom display

option and display device type. The file is read from unit 11 and organized as follows.

Card 1
Data: DTFORM.
Format: (I1).
Details: flag for the specific protein structure data format (1, the X-ray crystallographic data format of the Brookhaven Protein Data Bank; 2, the protein structure data format of the protein adsorption simulation program PROADS).

Card 2
Data: ROTAG1, ROTAG2.
Format: (F6.3, 4X, F6.3).
Details: pre-set rotational angles (in radian) for rotating a protein molecule to any orientation.

Card 3
Data: HYDDIS.
Format: (I1).
Details: flag for the optional display of all hydrogen atoms in the protein structure data format of the protein adsorption simulation program PROADS (0, no hydrogen atom display; 1, yes, display of the hydrogen atoms).

Card 4
Data: DISDEV.
Format: (I1).
Details: flag for the type of display device (1, the IBM 3179 color graphic terminal; 2, the Calcomp 1073 roll fed four pen color plotter; 3, the VISUAL 550 graphic terminal; 4, the VT100 graphic terminal; and 5, TEKTRONIX 4010 graphic terminal). The listed display devices are installed at the Purdue University Computing Center.

Card 5
Data: INTROT.
Format: (I1).
Details: flag for the interactive rotation of a protein molecule on screen (0, no, only show static picture; 1, yes, allow user to interactively rotate the protein molecule).

Card 6
Data: INTELG.
Format: (I1).

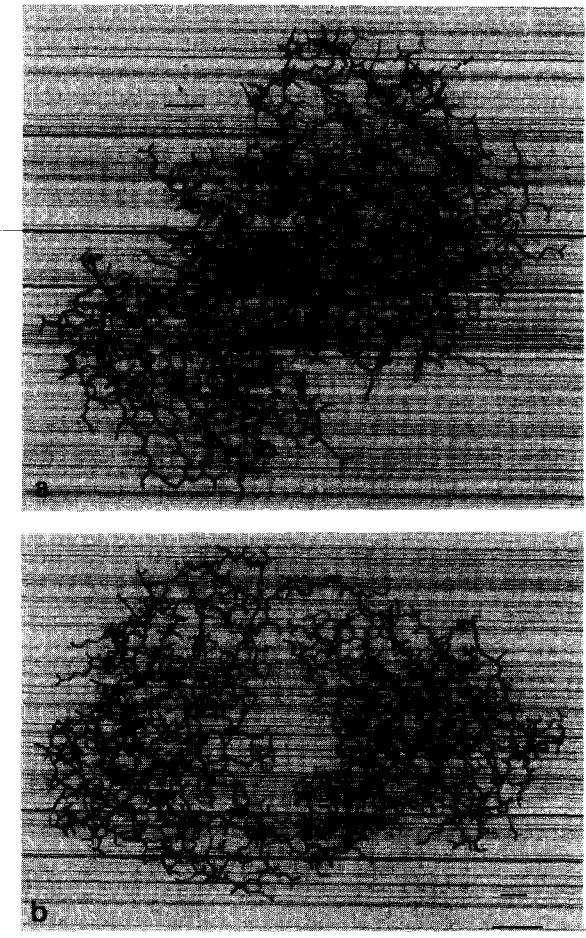

Fig. 1. Protein graphics for Test Run 1 using immunoglobulin F_{ab}: (a) the reference orientation obtained from the Brookhaven Protein Data Bank; (b) The protein was rotated by angles 1.400 (Y axis) and 2.200 (Z axis), and then interactively rotated by angles 0.500 (Y axis) and 1.700 (Z axis).

Details: flag for the interactive enlargement of any part of a protein molecule (0, no, no enlargement; 1, yes, allow user to interactively enlarge the protein molecule).

3.2. Procedure of the program

The run steps are as follows during the execution of the program:

(1) The file COMDIS is read from unit 11 to input the instruction command.
(2) The file DISPRO containing the coordinate data of protein structure is read from unit 10.
(3) Rotation of a protein molecule according to the input rotational angles to a given orientation.
(4) The ranges of the axes for plotting the protein structure are calculated based on the size of the protein and the coordinate position of the protein.
(5) Determination of the drawing algorithm for the protein backbone, calculation of the drawing bar thickness based on the relative position of each peptide bond on the Y axis and plotting of the backbone with each corresponding bar thickness.
(6) Determination of the drawing algorithm for each carbonyl group of the peptide bond, calculation of the corresponding bar thickness and plotting.
(7) Determination of the drawing algorithm (except the chemical bonds for hydrogen atoms) for each amino acid residue on the protein

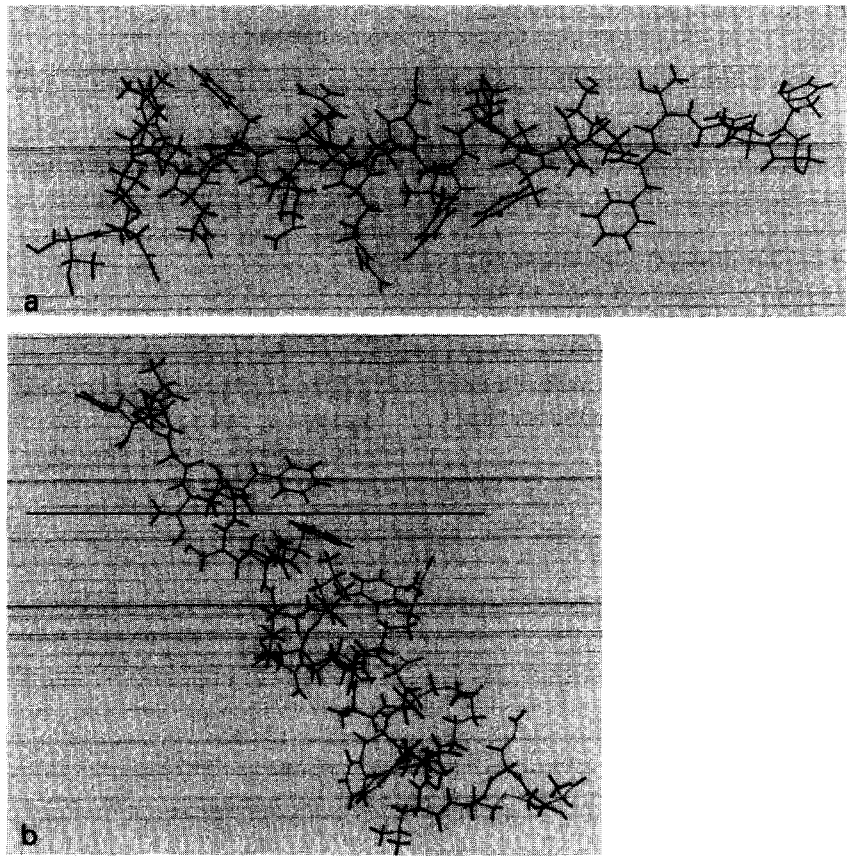

Fig. 2. Protein graphics for Test Run 2 using glucagon: (a) the orientation obtained from the protein adsorption simulation program PROADS (with hydrogen atoms); (b) the protein was rotated by angles 1.000 (Y axis) and 1.000 (Z axis).

side chains, calculation of the corresponding bar thickness and plotting.
(8) Determination of the drawing algorithm for all hydrogen atoms, calculation of the corresponding bar thickness and plotting.
(9) Interactive rotation or enlargement of a protein molecule on screen.

3.3. Output

(1) The structure display of a protein molecule.
(2) The data error messages are printed out from unit 12.

4. Main program and subroutines

(1) MAIN
Control of the execution which includes input/output, rotation of the protein molecule, calculation of the plot range, determination of the drawing algorithm for all atoms, calculation of the corresponding drawing bar thickness and plotting of the protein structure. The execution is carried out along with the following subroutines.

(2) BLOCK DATA
This section contains the drawing algorithms for following data sets corresponding to each of the 20 amino acid residues: (a) protein side chain for the X-ray crystallographic data format; (b) protein side chain for the data format of the protein adsorption simulation program PROADS; (c) the hydrogen atoms of a protein molecule.

(3) ROTPRO
Rotation of a protein molecule according to the two input rotational angles.

(4) DRWSID
Plotting of the amino acid structure on the side chain without hydrogen atoms.

(5) DRWLIN
Drawing the bonds for all hydrogen atoms.

5. Test run

(1) The X-ray crystallographic structure of immunoglobulin F_{ab} (Ig F_{ab}) from the Brookhaven Protein Data Bank. The protein has 445 amino acid residues and is plotted without hydrogen atoms (see fig. 1).

(2) The glucagon structure from the protein adsorption simulation program PROADS. The protein has 29 amino acid residues and is plotted with all hydrogen atoms (see fig. 2).

The test run input is given at the end of this paper.

References

[1] Protein Data Bank, Brookhaven National Laboratory, Brookhaven, New York.
[2] D.R. Lu and K. Park, in preparation.

TEST RUN INPUT

```
   ** FILE COMDIS (for protein data bank format) **
1
 0.000     0.000
0
2
0
0

   ** FILE DISPRO (from Brookhaven Protein Data Bank) **
ATOM      1  N   GLU L   1      38.496  53.804  50.025  0.00  0.00
ATOM      2  CA  GLU L   1      37.580  54.759  49.388  0.00  0.00
ATOM      3  C   GLU L   1      38.251  55.748  48.440  0.00  0.00
ATOM      4  O   GLU L   1      39.432  55.547  48.061  0.00  0.00
ATOM      5  CB  GLU L   1      36.699  55.486  50.421  0.00  0.00
ATOM      6  CG  GLU L   1      35.893  54.479  51.264  0.00  0.00
ATOM      7  CD  GLU L   1      34.893  55.230  52.137  0.00  0.00
ATOM      8  OE1 GLU L   1      33.723  55.412  51.715  0.00  0.00
ATOM      9  OE2 GLU L   1      35.122  55.368  53.365  0.00  0.00
ATOM     10  N   SER L   2      37.405  56.511  47.791  0.00  0.00
ATOM     11  CA  SER L   2      37.742  57.180  46.535  0.00  0.00
ATOM     12  C   SER L   2      38.303  58.590  46.661  0.00  0.00
ATOM     13  O   SER L   2      37.496  59.536  46.470  0.00  0.00
ATOM     14  CB  SER L   2      36.486  57.238  45.648  0.00  0.00
ATOM     15  OG  SER L   2      35.908  55.930  45.544  0.00  0.00
ATOM     16  N   VAL L   3      39.643  58.634  46.622  1.00 31.51
ATOM     17  CA  VAL L   3      40.586  59.749  46.332  1.00 32.14
ATOM     18  C   VAL L   3      41.974  59.808  47.020  1.00 30.41
ATOM     19  O   VAL L   3      42.099  59.803  48.274  1.00 33.61
ATOM     20  CB  VAL L   3      39.889  61.116  46.283  1.00 32.40
ATOM     21  CG1 VAL L   3      39.816  61.804  47.670  1.00 26.99
ATOM     22  CG2 VAL L   3      40.538  62.005  45.214  1.00 26.99
ATOM     23  N   LEU L   4      43.018  60.098  46.259  1.00 26.14
ATOM     24  CA  LEU L   4      44.265  60.491  46.942  1.00 20.15
ATOM     25  C   LEU L   4      44.277  62.015  47.126  1.00 16.02
ATOM     26  O   LEU L   4      43.687  62.754  46.297  1.00 15.96
ATOM     27  CB  LEU L   4      45.490  60.053  46.095  1.00 13.31
ATOM     28  CG  LEU L   4      45.503  58.545  45.836  1.00  9.37
ATOM     29  CD1 LEU L   4      46.655  58.112  44.900  1.00  8.60
ATOM     30  CD2 LEU L   4      45.658  57.837  47.181  1.00  8.60
ATOM     31  N   THR L   5      44.852  62.511  48.182  1.00 14.47
ATOM     32  CA  THR L   5      44.828  63.965  48.378  1.00 14.66
ATOM     33  C   THR L   5      46.135  64.654  48.002  1.00 15.15
ATOM     34  O   THR L   5      47.199  64.333  48.585  1.00 16.24
ATOM     35  CB  THR L   5      44.454  64.319  49.841  1.00 16.10
ATOM     36  OG1 THR L   5      43.104  63.906  50.126  1.00 30.93
ATOM     37  CG2 THR L   5      44.569  65.819  50.123  1.00 30.93

... (Residues 6-216 are omitted due to space limitation)

ATOM   3321  N   GLU H 217      28.772  96.847  10.404  1.00 30.09
ATOM   3322  CA  GLU H 217      28.779  98.193   9.840  1.00 33.26
ATOM   3323  C   GLU H 217      28.762  99.273  10.918  1.00 34.44
ATOM   3324  O   GLU H 217      27.973  99.168  11.885  1.00 33.97
ATOM   3325  CB  GLU H 217      27.509  98.314   8.965  1.00 39.11
ATOM   3326  CG  GLU H 217      27.674  99.263   7.767  1.00 44.04
ATOM   3327  CD  GLU H 217      26.329  99.392   7.060  0.00  0.00
ATOM   3328  OE1 GLU H 217      25.910  98.439   6.355  0.00  0.00
ATOM   3329  OE2 GLU H 217      25.505 100.251   7.465  0.00  0.00
ATOM   3330  N   PRO H 218      29.505 100.344  10.706  1.00 33.43
ATOM   3331  CA  PRO H 218      29.437 101.556  11.548  1.00 33.83
ATOM   3332  C   PRO H 218      28.007 102.077  11.623  1.00 36.20
ATOM   3333  O   PRO H 218      27.289 102.079  10.586  1.00 37.08
ATOM   3334  CB  PRO H 218      30.304 102.634  10.881  1.00 32.72
ATOM   3335  CG  PRO H 218      30.921 101.995   9.635  1.00 31.27
ATOM   3336  CD  PRO H 218      30.391 100.555   9.553  1.00 32.72
ATOM   3337  N   LYS H 219      27.599 102.422  12.829  1.00 39.28
ATOM   3338  CA  LYS H 219      26.287 103.022  13.075  1.00 43.68
ATOM   3339  C   LYS H 219      26.225 104.403  12.429  0.00  0.00
ATOM   3340  O   LYS H 219      27.297 104.965  12.087  0.00  0.00
ATOM   3341  CB  LYS H 219      26.040 103.137  14.589  0.00  0.00
ATOM   3342  CG  LYS H 219      25.943 101.748  15.249  0.00  0.00
ATOM   3343  CD  LYS H 219      25.799 101.843  16.779  0.00  0.00
ATOM   3344  CE  LYS H 219      25.564 100.468  17.434  0.00  0.00
ATOM   3345  NZ  LYS H 219      24.275  99.908  17.001  0.00  0.00
ATOM   3346  N   SER H 220      25.025 104.797  12.051  0.00  0.00
ATOM   3347  CA  SER H 220      24.717 105.969  11.209  0.00  0.00
ATOM   3348  C   SER H 220      25.132 107.320  11.786  0.00  0.00
ATOM   3349  O   SER H 220      24.432 108.335  11.545  0.00  0.00
ATOM   3350  CB  SER H 220      23.197 106.017  10.972  0.00  0.00
ATOM   3351  OG  SER H 220      22.739 104.752  10.476  0.00  0.00
ATOM   3352  N   CYS H 221      26.112 107.298  12.664  0.00  0.00
ATOM   3353  CA  CYS H 221      26.501 108.437  13.497  0.00  0.00
ATOM   3354  C   CYS H 221      27.901 108.961  13.199  0.00  0.00
ATOM   3355  O   CYS H 221      28.079 110.199  13.078  0.00  0.00
ATOM   3356  CB  CYS H 221      26.453 108.007  14.972  0.00  0.00
ATOM   3357  SG  CYS H 221      25.099 106.899  15.411  0.00  0.00
ATOM   3358  OXT CYS H 221      28.892 108.231  13.455  0.00  0.00

   ** FILE COMDIS (for program PROADS data format) **
2
 0.000     1.571
1
2
0
0

   ** FILE DISPRO (from program PROADS data) **
 29
HIS
 18
H2         1   1     0.28252    3.49084   23.89783   -4.03735
N          7  41    -0.52057    3.21846   23.57836   -3.23672
CA         6  38    -0.12686    3.06837   24.49247   -2.09140
C          6  35     0.54273    2.60328   23.86181   -0.77932
HA         1   2     0.21666    3.97730   24.86713   -1.97857
CB         6   7    -0.35988    2.17224   25.69813   -2.39624
ND1        7  12    -0.29787    0.11757   24.13383   -2.68216
NE2        7  15    -0.44445   -1.27350   25.45813   -3.86825
CG         6  14    -0.07329    0.75579   25.36764   -2.91506
CE1        6  14    -0.07329   -1.15274   24.18861   -3.28487
CD2        6  14    -0.07329   -0.07984   26.15785   -3.61702
HE2        1  16     0.26870   -1.99455   25.78971   -4.38173
HB1        1   2     0.21666    2.66555   26.20801   -3.05362
HB2        1   2     0.21666    2.13362   26.31268   -1.62375
HE1        1  16     0.26870   -1.81697   23.51586   -3.29746
HD         1  16     0.26870    0.06312   27.07506   -3.92611
O          8  10    -0.38432    3.30728   22.92453   -0.32037
H1         1   1     0.28252    3.28123   24.07548   -3.97086
SER
 11
H2         1   1     0.25476    1.52720   25.55903   -0.43836
N          7  41    -0.57582    2.10176   24.80557   -0.01851
CA         6  38    -0.14032    1.34693   24.58818    1.21217
C          6  35     0.48941    0.75529   23.18935    1.27420
HA         1   2     0.19538    2.07499   24.69057    2.03125
CB         6  18     0.00766    0.17594   25.53107    1.23650
OG         8   9    -0.56023   -0.72856   25.18506    2.24628
HB1        1   2     0.19538    0.54903   26.53727    1.45159
HB2        1   2     0.19538   -0.31400   25.55797    0.25766
HG         1   4     0.38692   -1.37231   24.53229    1.89578
O          8  10    -0.42510    1.21310   22.33968    2.07653

... (Residues 3-27 are omitted due to space limitation)

ASN
 14
H2         1   1     0.27013    1.38477  -18.69629    5.41081
N          7  41    -0.54523    0.57009  -19.32119    5.24431
CA         6  38    -0.13286    0.05943  -19.63982    6.59498
C          6  35     0.51893   -0.12999  -21.10182    6.97279
HA         1   2     0.20717    0.81879  -19.21649    7.28619
CB         6   7    -0.37693   -1.13762  -18.83671    7.06797
OD1        8  10    -0.40252    0.43649  -17.56914    8.33949
ND2        7  13    -0.62002   -1.66481  -17.74501    9.11188
CG         6   5     0.51893   -0.68444  -18.13342    8.33929
HD1        1   1     0.27013   -1.46507  -17.32039    9.98260
HD2        1   1     0.27013   -2.60817  -17.86894    8.80081
HB1        1   2     0.20717   -1.41019  -18.05055    6.32710
HB2        1   2     0.20717   -2.05798  -19.35614    7.24603
O          8  10    -0.40252   -1.28727  -21.53433    7.19960
THR
 16
H2         1   1     0.26732    1.93915  -21.25261    6.77487
N          7  41    -0.55081    1.01446  -21.76492    6.99356
CA         6  38    -0.13422    1.37876  -23.12984    7.44510
C          6  35     0.51354    0.28856  -24.20346    7.53279
HA         1   2     0.20501    1.94306  -23.43665    6.62608
CB         6  18     0.00804    2.10005  -23.01060    8.78980
OG1        8   9    -0.53589    1.20585  -23.23186    9.87451
CG2        6   6    -0.60450    2.76970  -21.64793    8.96274
HB         1   2     0.20501    2.93085  -23.66241    9.00402
HG21       1   3     0.20602    3.26472  -21.60897    9.92245
HG22       1   3     0.20602    2.02162  -20.87061    8.91150
HG23       1   3     0.20602    3.49563  -21.50119    8.17669
HG1        1   4     0.40601    1.64964  -22.93706   11.02640
O          8  10    -0.40664   -0.64639  -24.20864    6.70073
HO2        1   4     0.40601    0.11398  -25.97200    7.75451
O2         8   9    -0.53589    0.51847  -25.23922    8.20391
```