SHIFTS Users' Manual
(Version 5.6)

Xiaoping Xu, Seongho Moon, Jason Swails and David A. Case

2/13/2021

The SHIFTS package includes the functionality of our earlier proton-shift-prediction code (versions 3 and earlier), augmented with new ideas based on density functional calculations on peptides. The earliest codes were written by Klara Osapay and David Case; later Doree Sitkoff added new ideas and subroutines. Annick Dejaegere and Richard Bryce extended the scope to nucleic acids, and Jason Swails refined the random coil contributions of the nucleic acid protons to improve agreement with experiment. Xiaoping Xu then developed an alternative approach that is applicable to nitrogen and carbon shifts. Procedures for fitting ring current intensities to results from quantum calculations were developed by David Case. SHIFTS 5.6 is an attempt to put all of these ideas into a coherent software package. This version has plenty of rough edges, but we hope it will be useful.

The basic literature references are:

1. K. Osapay and D.A. Case. A new analysis of proton chemical shifts in proteins. *J. Am. Chem. Soc*. **113,** 9436-9444 (1991).

2. D.A. Case. Calibration of ring current effects in proteins and nucleic acids. *J. Biomol. NMR* **6,** 341-346 (1995).

3. A.P. Dejaegere, R.A. Bryce and D.A.Case. An empirical analysis of proton chemical shifts in nucleic acids. In *Modeling NMR Chemical Shifts*, J.C. Facelli and A.C. de Dios, eds. (Washington, American Chemical Society, 1999), pp. 194-206.

4. X.P. Xu and D.A. Case. Automated prediction of ^{15}N, ^{13}C, ^{13}Cα and ^{13}C′ chemical shifts in proteins using a density functional database. *J. Biomol. NMR* **21,** 321-333 (2001).

5. X.P. Xu and D.A. Case. Probing multiple effects on ^{15}N, ^{13}C, ^{13}Cα and ^{13}C′ chemical shifts in peptides using density functional theory. *Biopolymers* **65,** 408-423 (2002).

6. S. Moon and D.A. Case. A new model for chemical shifts of amide hydrogens in proteins. *J. Biomol. NMR* **138,** 139-150 (2007)

Proton shift predictions for proteins should be compatible with earlier versions of SHIFTS, and should match results reported in references [1], above. Predictions for proton shifts in DNA and RNA have been re-calibrated from those in Ref. [3] by adjustment of the reference (or "random-coil") shifts; see Section

2.5 below. Carbon and nitrogen shifts in proteins use the algorithm described in references [4] and [5]. Amide proton shifts are computed as described in reference [6]. Shift predictions using density functional theory were a part of earlier versions of *shifts*, but have now been moved to a separate package: see github.com/dacase/afnmr.git.

1 Installation

The *shifts* package is available on the web at:

```
casegroup.rutgers.edu
```

(Click on the "SHIFTS" menu item.) SHIFTS (version 5) is written in the NAB programming language, so you must first download and install AmberTools20, ambermd.org/AmberTools.php.

The *shifts* installation uses the compiler settings specified for AmberTools to make sure the NAB-compatible compilers are used, so AmberTools must be installed and configured properly, and the `$AMBERHOME` environment variable must be set correctly.

The first step in installing shifts is to extract files using the UNIX commands:

```
tar xvf shifts-5.6.tar.gz
```

The path to this new directory should be defined as the environment variable `$SHIFTSHOME`.

```
export SHIFTSHOME="insert-your-path-here/shifts-5.6"
```

Next,

```
make install
```

will make the required executable files, and put them into `$SHIFTSHOME/bin`. Again, this assumes that AmberTools has been properly installed. If you see an error about a file config.h that does not exist or if the "nab" command cannot be found, then your AmberTools installation either does not exist or is incomplete.

After you have installed *shifts*, the command

```
make test
```

will run some test calculations and report results. This is important to make sure that your installation is working correctly.

2 SHIFTS

SHIFTS uses empirical formulas to estimate proton chemical shifts in proteins and nucleic acids; it can also use a database of density-functional shifts to predict ^{15}N and ^{13}C shifts in proteins. The basic command-line for computing proton shifts is:

```
shifts [options] atom_expr basename > output
```

The program takes an input pdb-file basename.pdb, and optionally a file of observed shifts basename.obs. Shifts are calculated for atoms matching "atom_expr". Basically, this is a string with three fields, separated by the ":" character. The first field identifies the molecule, the second the residue, and the third field the atom. For example, to compute shifts for all protons, set atom_expr to '::H*'; for all Hα protons in residues 3-27, set atom_expr to ':3-27:HA'. (Use single quotes, as in these examples, to protect wild-cards from possible shell expansion.) Full details of atom-expressions are given in the NAB documentation found at the end of the AmberTools manual (http://ambermd.org/doc12/).

For example, the following line executes the BPTI test case (assuming you are in the `$SHIFTSHOME` directory):

```
cd test; ../bin/shifts '::H*' 6pti
```

This reads in the file 6pti.pdb, a puts human-readable output in 6pti.emp.

The syntax is slightly different for nitrogen and carbon shifts. Here the command line is:

```
$SHIFTSHOME/bin/shifts -qdb [options] basename
```

Here "basename" is as described above; the atom-expression is not needed (since the qdb option automatically computes N, C, C and C′ shifts); and the output goes to "basename.qdb" (see below).

Finally, there is a special section for using a neural net model to predict shifts for amide protons. Here the command line is:

```
$SHIFTSHOME/bin/shifts -HN '::H' basename
```

Again, input comes from basename.pdb, and the output goes to basename.HN.

2.1 Options

-csa If the -csa flag is set, chemical shift anisotropy tensors will be computed, based on susceptibilities for rings and peptide groups. Note that this option is not yet fully parameterized: see ringinfo.h for information about what is being assumed.

-reslib If the flag is not set, the program will accept coordinates for an "arbitrary molecule" in basename.pdb. This allows calculations on molecules other than proteins or nucleic acids. Charges will be set to zero, so that there will be no electrostatic contributions to the computed shifts. Since the protein charge model from Ref. [1] is hard-coded, there is no need to set the -reslib flag for proteins. It is also not needed when -qdb is set, since that model ignores charges. Hence, currently, the -reslib flag is only needed for proton shift calculations in nucleic acids, or if one wants to investigate a different charge model.
Note: if you use this option, it will generally help to process the input pdb file with *pdb4amber* (from the AmberTools suite) before running *shifts*.

-nocoil If the -nocoil flag is set, the program will not attempt to compare the computed shifts to "random coil" reference values. Otherwise, the program will use fragment shifts as reference values: see sbcoil.nab for more information. The option is only relevant when -qdb is not set.

-readobs If this flag is set, the program will read in observed shifts (in the file basename.obs), and will compare these to calculated shifts.

-details When this flag is set, individual contributions to proton shift predictions will be printed out.

-sander When this flag is set for proton calculations, an input file appropriate for the sander module of AMBER will be output.

-qdb If this flag is set, a "quantum data base" (qdb) approach will be used to compute 13C and 15N shifts in proteins. Otherwise, empirical formulas will be used for proton shifts.

-refine If this flag is set, after computing shifts for the given structure, the program will attempt to find side chain 1 angles that significantly improve the qdb fit. The results for the original structure will be in the file basename.out1, and the "refined" results will be in basename.out2. Only applicable if -qdb is also set.

-HN This flag initiates a special module that predicts amide proton shifts in accordance with the model in Ref. [6], above.

2.2 Less common options

-rinfo If the -rinfo flag is set, read the ring current information from ringinfo.in

-rmd Use the ring magnetic dipole approximation

-swap Pro-chiral calculated shifts will be swapped

-noreduce Do not call the reduce program to add hydrogens.

2.3 Input files

Basic parameters for the calculations are contained in the header files constants.h and ringinfo.h. These files need to be edited if you wish to change the parameterization, add new types of aromatic rings, and so on. There is an experimental option to read a file of bond anisotropies; this is not a part of the standard calculations, but might be useful in extending the current model. For now, to use this, you will have to read the code to see how it works.

The basename.pdb file can generally be taken from Brookhaven. If there are several "models" in the file (typical for NMR structures), the user must edit the file to select the model of interest; this can be with the *pdb4amber* program from AmberTools. For proton calculations, Only ATOM records are considered, so that water molecules and co-factors are ignored. For carbon and nitrogen, ATOM records and HETATM records that refer to waters are included. Ligands and HETATM groups can be present, generally as a separate chain, separated with a "TER" card from the polypeptide backbone; see the test/1mbc.pdb file for an example. Ring currents are defined for the porphyrin ring, as in reference [1] above; for other ligands, you will need to edit the ringinfo.h file to establish the rings and their intensities.

The optional observed-shifts-file basename.obs or basename.str is only used here to prepare a list that compares calculated and observed shifts. The format of the input file is dicussed in the tutorial below. The "str" format is taken from the BMRB database (see http://www.bmrb.wisc.edu). As an example, a portion of the file for ubiquitin (bmr68.str) is reproduced here:

```
1 1 M HA H 4.22 1
2 1 M HB2 H 2.07 2
3 1 M HB3 H 2.17 2
4 2 Q H H 8.82 1
5 2 Q HA H 5.27 1
6 2 Q HB2 H 1.67 2
7 2 Q HB3 H 1.88 2
8 2 Q HG2 H 2.23 1
9 2 Q HG3 H 2.23 1
10 2 Q HE21 H 6.71 2
11 2 Q HE22 H 7.52 2
12 3 I H H 8.33 1
13 3 I HA H 4.21 1
14 3 I HB H 1.8 1
15 3 I HG12 H .88 2
16 3 I HG13 H 1.1 2
17 3 I HG2 H .66 1
18 3 I HD H .61 1
19 4 F H H 8.57 1
20 4 F HA H 5.61 1
```

The columns (in order) are: index number, residue number, residue name, atom name, atom symbol, chemical shift, ambiguity type. The last column is ignored (currently). This information can generally be extracted from a BMRB entry (if it exists), or prepared by some other program. If it is not convenient for you to prepare shifts in this format, you may wish to just run shifts with just the pdb file as input, and use a separate program to prepare any comparisons you wish to make.

2.4 Output files

For proton shifts, the basic output comes to basename.emp; the program also produces an NMR-star file (basename.bmrb, similar to that shown above. If there is an observed-shift file, a flat-file database, basename.rdb will be created; the specific format for this matches Walt Hobbs' rdb programs (see ftp://ftp.rand.org/pub/RDB-hobbs), but is ASCII-encoded, and should human-readable as well.

A somewhat different convention is followed for heavy atom shifts. Here, the basic output before any side-chain refinement is sent to basename.qdb, and some extra geometrical information goes to basename.par. If the -refine flag is set, two additional files are created (basename.qdb2 and basename.par2)

that hold the same information after refinement. To see what effect (if any) the sidechain refinement has made, the user should diff these two sets of files.

2.5 Updates to DNA and RNA reference shifts in Version 5

For version 5, we re-computed the reference shifts used for proton shift calculations for DNA and RNA. The old parameters (from Ref. [3]) and the new ones are in the *sbcoil.nab* file. In brief, we constructed a database of DNA, RNA and DNA/RNA hybrids that had corresponding chemical shift entries in the BMRB database; the PDB id's are given below.

DNA:

```
1BJD 1BWT 1C11 1C32 1COC 1CQO 1EMQ 1EVM 1EVO 1IR5 1JRW
1JVE 1K2K 1KKV 1KKW 1KR8 1NP9 1ONM 1RDE 1SKP 1U64 2F8U
```

RNA:

```
1A60 1BN0 1ESY 1F5H 1HWQ 1I4C 1IDV 1JO7 1JU7 1K8S 1KKA
1KPY 1L1W 1LC6 1LDZ 1LMV 1LPW 1LUU 1LUX 1M82 1MFJ 1MFY
1MNX 1MUV 1MV1 1MV2 1MV6 1N8X 1NA2 1NC0 1NZ1 1OW9 1PJY
1Q75 1R2P 1R4H 1S34 1S9S 1T28 1TJZ 1XHP 1YG4 1YMO 1YSV
1Z2J 1Z30 1ZC5 28SR 2ADT 2AHT 2AU4 2DD3 2F87 2FDT 2GM0
2H49 2KOC 2O33 2TPK 3PHP
```

DNA/RNA hybrids:

```
1BYX  1C2Q  1NTQ  1NTS  1NTT  1OO7
```

Using this data, the reference shifts for DNA and RNA were optimized to best fit the experimental data. There are pros and cons to this update: the original reference shifts were taken from model compounds, as described in Ref. [3]. This means that the calculations provide a model of how nucleic acid structures (say in duplexes) should differ from those in similar small molecules. By fitting the reference shifts, we get somewhat better agreement with experiment, but at the expense of additional adjustable parameters. Since we anticipate that most users will want to use *shifts* in a predictive mode, the new values are the default. But some users may wish to edit the code in *sbcoil.nab* to use the earlier values (or to try their own ideas).

2.6 Bugs

As just a program to run, there is not much flexibility in how things are calculated. Basically, if you wish to modify the program to do something other than the standard calculations, you are required (and invited!) to modify the code. We have tried to make the code easy enough to read to make this possible.

If the program cannot interpret certain atoms in a pdb file, it will emit messages of the form: "unmatched atom in file ...." This means that these atoms are not being used in the calculation. In many cases this is fine; if not, you may need to modify the input pdb-file to fix up atom names, or update the residue libraries in NAB to reflect the molecules you are working on.

Parameters for DNA/RNA and proteins were developed separately, and there is no evidence (one way or the other) of how they would work together, say on protein-DNA complexes.

There is no provision in the observed-shifts-file for distinguishing between situations where pro-chiral proton pairs are or are not stereo-assigned. For situations where stereo assignments are not available, using the -swap flag to shifts may be helpful.

3 Tutorial: ^{13}C and ^{15}N chemical shift prediction using SHIFTS

This tutorial will provide an example of running a 13C and 15N chemical shift prediction for protein profilin (PDB ID: 1ACF). It is assumed that the programs SHIFTS as well as NAB have been installed in you site and the environment for running SHIFTS has been set up, as described above. The following topics will be covered in this tutorial.

Preparation of input file(s) - Running SHIFTS - Output files

The input and output files for this example are in the test subdirectory: 1acf.pdb and 1acf.obs are input files, and the output files we obtained on our machine are in the examples subdirectory – you can compare your results to these.

3.1 Preparation of input file(s)

The program takes an input pdb-file basename.pdb, and optionally a file of observed shifts basename.obs. The basename.pdb file can generally be taken from Brookhaven. If there are several "models" in the file (typical for NMR structures), the user must hand-edit the file to select the model of interest. Only ATOM records and HETATM for water molecules are considered. Ligands can be present but will not be included for calculation in current version.

For the example of protein profilin, the original pdb-file 1acf.pdb is used without any modification. Then a 1acf.obs file was prepared with the following format:

```
#1ACF.obs
#No. res N Ca Cb C'
#---------------------------------------
3 Q 118.26 58.53 28.48 0.00
4 T 114.00 66.23 68.58 0.00
5 Y 119.55 60.68 38.29 177.50
6 V 115.86 66.72 32.26 177.10
7 D 118.29 57.94 40.88 178.90
8 T 115.72 65.55 68.82 175.00
9 N 116.09 56.61 39.01 174.20
10 L 114.26 55.50 41.90 176.70
11 V 122.33 66.35 31.61 179.60
...
```

The optional observed-shifts-file 1acf.obs here is used to compare calculated and obseved shifts as well as do side-chain orientation refinement. You may also use a 1acf.str file instead: if the program doesn't find basename.obs, it looks for basename.str. As mentioned above, the ".str" format file is like a subset of the data in an "NMR -star file from BMRB (see http://www.bmrb.wisc.edu). The data example with .str format is:

```
0 1 ALA C C 172.0 1
1 1 ALA CA C 51.0 1
2 1 ALA HA H 4.40 1
3 1 ALA CB C 17.9 1
4 1 ALA HB H 1.57 1
5 2 PRO C C 176.8 1
6 2 PRO CA C 63.1 1
7 2 PRO HA H 4.62 1
8 2 PRO CB C 31.9 1
9 2 PRO HB2 H 2.39 2
```

(Note that the above format is *not* exactly what is in the .str files from BMRB, so you may have to do some editing. We are working on a better parser.)

3.2 Running SHIFTS

With input file(s) ready, the running of SHIFTS is pretty easy. The commands are "`shifts -qdb basename`" for prediction only; "`shifts -qdb -readobs basename`" for comparison of predicted and observed data; and "`shifts -qdb -refine basename`" for prediction, observed data and side-chain orientation refinement.

Take the protein profilin as the example here and assume you have a directory containing 1acf.pdb and 1acf.obs (these files are available in `$SHIFTSHOME/test`). Then you can run SHIFTS as below:

```
$SHIFTSHOME/bin/shifts -qdb -readobs 1acf
```

You will get two output files, 1acf.out1 and 1acf.par1.

3.3 Output files

There are two types of output files being formed after running SHIFTS. One of them is a structural parameter file (.par) with backbone conformational parameters (phi and psi), side-chain orientation parameters (chi1 and chi2), and hydrogen bonding information. Another type is predicted chemical shifts file(s) giving a breakdown of the detailed contributions. Depending on if an observable shift file is included and if side-chain orientation refinement is selected, the following output files will be given.

Flags Used	Output File Suffixes	Comment
-qdb	.par1, .out1	no observed file and no refinement
-qdb -readobs	.par1, .out1	observed file included and no refinement
-qdb -refine	.par2, .out2	observed file included and with refinement

In the 1acf.par1 file, the data in columns phi, psi, chi1 and chi2 corresponds to torsion angles. H1 means the shortest hydrogen bond length between the (*i-1*)th carbonyl oxygen and the NH proton among all possible O(*i-1*)...HN pairs. H2 is the next shortest distance for the same oxygen as the case in H1. Similarly, O1 and O2 in the .par file represent the distances between the carbonyl oxygens with the shortest and next shortest HB lengths to the *i*th nitrogen. Nw and Ow mean the the hydrogen bonding on the *i*th nitrogen and the (*i-1*)th oxygen, respectively, by waters. '*i*' in the file is the residue number to which the atom belongs.

Abbreviations in the output files (for both proton and heavy-atom shifts) are listed here:

Abbreviations for heavy atom shift contributions:

```
bb-p ..... preceding backbone effect
bb-s ..... self backbone effect
bb-f ..... following backbone effect
chi-p ..... preceding chi effect
chi-s ..... self chi effect
HB-D....... direct HB effect (NH &lt;-)
HB-I....... indirect HB effect (C=O &lt;)
REF ....... reference shifts for this amino acid
pred....... predicted shift = sum of the above
obs ....... observed shift (if provided)
```

Lines marked "diff" indicate some difference in sequence between the pdb file and the obs file (*not* that the observed and predicted shifts differ by a given amount).

Abbreviations for the proton shift contributions:

```
RingCur.... ring currents from aromatic rings
El ........ electrostatic contribution (backbone only for now)
P_anis .... peptide group anisotropy
Const ..... constant contribution
RC ........ random coil shift
pred ...... predicted shift = sum of the above
```

Note that the message "unmatched atom ...OXT" can be ignored. In the output, shifts for methyl protons are reported as the average for the three protons, and the atom name of the last proton is used; for example, the (averaged) methyl shift for alanine would be labeled "HB3".