Monte™ v2.02
User Manual

by T. Kevin Hitchens
April 2003

T. Kevin Hitchens
Jonathan A. Lukin
Yiping Zhan
Gordon S. Rule

Corresponding addresses:

rule@andrew.cmu.edu
hitchens@andrew.cmu.edu

Disclamer

Monte is provided as is. Neither the authors, nor Canegie Mellon University provide any warranty, or guarantee of program function or correctness of results. Individual users are responsible for the use, and inferences of Monte results.

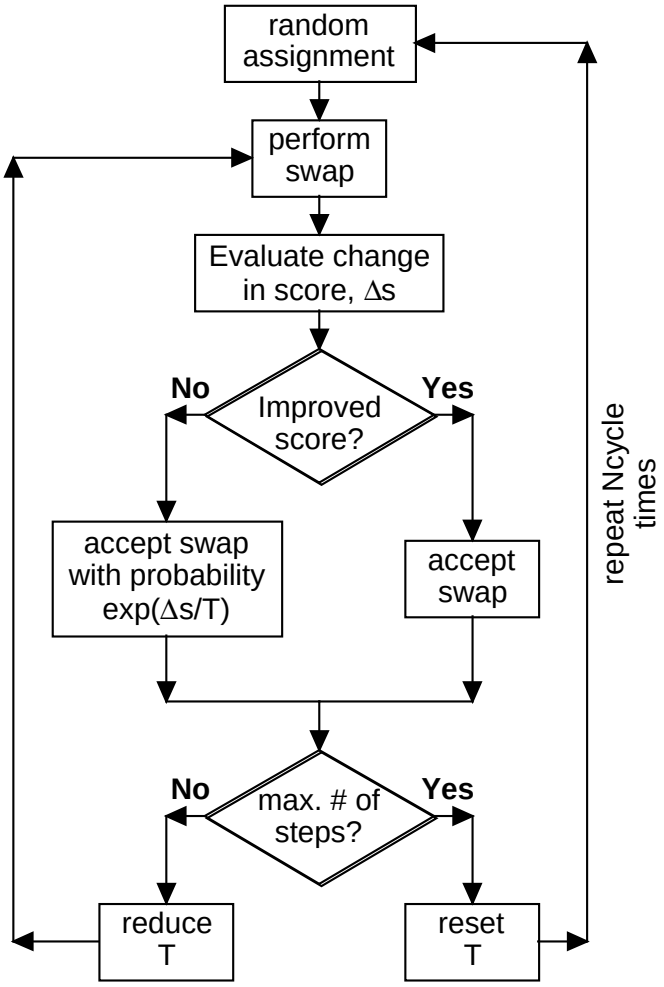
Table of Contents

1. Introduction

Monte is a program that applies a Monte Carlo/ Simulated-Annealing approach to obtain and identify a unique solution for the NMR chemical shift assignment problem of ^{15}N, ^{13}C labeled proteins. The motivation behind the Monte program is to provide a general software package for chemical shift assignments of proteins independent of particular "required" experimental data collection that may not be well suited for large, or small proteins, or protonated, or deuterated proteins.

The method starts with random assignments of the spin systems to residues in the primary sequence. A spin system is defined as all of the information associated with a particular amide peak. This includes inter-residue connectivities, NOE peaks, and intra-residue chemical shifts. All of this information is indexed by the nitrogen and proton shifts of the amide proton. The initial assignment solution is scored based upon any or all of the following criteria:

- Chemical shift matching of correlated spin pairs. How well chemical shifts identify the (*i*) and (*i*-1) residue type for a given secondary structure.
- How well amide nitrogen/proton chemical shifts match data obtained from specifically labeled protein samples.
- How well amide NOE's match the pattern predicted from a pdb file, or identified secondary structure.
- How consistent are 3D cross-peaks between the current problem, and a related (solved) problem (i.e. unliganded and ligand-bound forms of the same protein).



Following the initial scoring, two spin systems (or, two blocks of spin systems) are chosen at random, and swapped. The new solution is scored based upon the same criteria. If the swap results in an improvement in the score, the swap is accepted. If the swap results in a poorer score, the difference in score is compared to the "temperature" of the system and the swap may or may not be accepted. At infinite (high) temperature, all swaps are accepted. As the temperature of the system is lowered (simulated annealing), swaps that result in an improvement in the overall score will be preferentially accepted.

Monte is usually executed for 10 to 50 cycles, starting from different random starting positions. The best solution from each of these independent cycles are compared in order to evaluate the reliability of an assignment.

2. License Agreement

Academic and non-profit institutions are welcome to use Monte at no cost. However, users are required to sign an license agreement that is available on the program web site. Return the *two* signed copies of the license agreement to:

Dr. Gordon S. Rule
Department of Biological Sciences
4400 Fifth Avenue
Carnegie Mellon University
Pittsburgh, PA, 15213

After receipt of the license agreement you will be emailed a license key. This license key must by placed in a file called *monte.license* that must reside in the same directory as Monte. The license key itself is a 12 character string that should be the first (and only) line in the license file. To insure that you are using a relatively current version of Monte, the license key expires one year after its issue. Registered users need not sign a new license agreement. Simply drop me an email and I will send you a new key.

For-profit organizations are welcome to use Monte gratis for a limited trial period, after completing the necessary right-to-use agreement. After this period it is necessary to negotiate a license agreement. Please contact Gordon Rule for additional information.

3. Getting Started

Extract the Monte distribution tar file:

tar –xvf FMv2.02.tar

Directory Structure

The Monte home directory, FM2 contains the following four files:

fm.exe	Monte pre-compiled executable code.
aaconv	Chemical shift database for protonated shifts
aaconv_deut	Chemical shift database for deuterated proteins
fm.tcl	A graphical user interface for editing the *project.par* file.

Once you receive the license file, you should also place this file in the main directory. Within the home directory, one should create subdirectories for each chemical shift assignment project. These directories contain the parameter (.par) file, and all user input files. In addition, all output files will be placed in the subdirectory. The subdirectory path and experimental *filename* is provided to the parameter (.par) file for a given assignment problem. All input files must share the same root *filename*, and each individual file is identified by a distinct file extension (see below).

Examples: In addition to the above files, two sample project directories, called rho and rho_dna, are included with the distribution. The rho directory contains sample files of all of the data types that Monte can utilize. It also includes a parameter file *(.par)*, that contains all of the parameters that control Monte. Although this is a plain text file, it is highly recommended that one does not directly edit this file. Changes to program parameters should only be made through the tcl/tk Graphical User Interface (fm.tcl). The rho_dna directory is provided as an example of how to use known chemical shifts of an unliganded form of the protein to aid in the assignment of a liganded form, in this case a protein-DNA complex. The file that contains the chemical shifts for the unliganded form of the protein is called *rho_final.out*. When you run the rho_dna example you will have to select this file using the tcl interface in order to get the correct directory path.

Running tcl/tk: To run the tcl GUI, one should first make sure versions 8.0.4 or later of tcl and tk are installed. For SGI/IRIX these are available through http://freeware.sgi.com. Remember to change the symbolic links for /usr/bin/tclsh -> /usr/freeware/bin/tclsh8.0, and /usr/bin/wish -> /usr/freeware/bin/wish8.0. To do this, rm /usr/bin/tclsh and /usr/bin/wish and then make the links as follows:

- ln -s /usr/freeware/bin/wish8.0 /usr/bin/wish
- ln -s /usr/freeware/bin/tclsh8.0 /usr/bin/tclsh

Running and Customizing Monte: Helpful hints on how to customize Monte and define run parameters are included in section 10 of this manual.

4. Input files

4. 1 Overview

Minimum Requirements

aaconv and aaconv_deut [provided]
Parameter file (*filename*.par) [provided, user modified]
Chemical shift data base (*filename*.cs)
Sequence File (*filename*.seq)
License file (*monte.license)*

Optional Input

pdb file (*filename*.pdb)
Secondary structure information (*filename*.ss)
Specific label data
    ^{15}N Specific Labels (*filename*.nsl)
    1-^{13}C/^{15}N (*filename*.csl)
NOESY data
  Single
    Amide 4D NH-NH NOESY (*filename*.4dnoe)
or a *pair* of
    Amide 3D H-NH NOESY (*filename*.3dh_noe)
    Amide 3D N-NH NOESY (*filename*.3dn_noe)
Chemical shift assignment output from a related solution

4.2 Files aaconv and aaconv_deut

These are formatted text files that must reside in the HOME directory of the Monte program. The file contains the mean chemical shifts for CO, CA, CB, CG, and N nuclear spins, found for secondary structures of random coil, α-helix, and β-sheet for each of the 20 common amino acids (aaconv) or the same information corrected for deuterium isotope effects (aaconv_deut). The correction for the deuterium isotope effect is selected in the Tcl/tk interface.

The data found in these files is used by Monte to determine amino acid type probabilities for the (*i*) and (*i*-1) residues of a amide pair spin system. One may wish to view this file to determine if a chemical shift offset should be added to or subtracted from the user input shifts (N, CO, CA, CB, or CG shifts) for the purpose residue type calculations (See Parameter file section). A slight difference in chemical shift references may affect the assignment of residue type.

4.3 Chemical shift Data Base

The Chemical shift data base contains all J-correlated chemical shifts for a particular spin system. A spin system is defined here as all chemical shifts that may be correlated to a particular amide nitrogen–proton pair. These chemical shifts are used to both identify the residue type of the (*i*) and (*i*-1) residues, as well as linking spin systems by matching complementary pairs of chemical shifts. Chemical shifts may include data collected from any, or all of these types of experiments:

HNCA, HN(CO)CA
HNCACB, CBCACONH
HNCO, HN(CA)CO
HN(CACB)CG, HN(COCACB)CG
HNHA, HA(CACO)NH
HNHB, HBHA(CBCACONH)
(H)N(COCA)NH
(H)N(CACO)NH

The format of the chemical shifts data base data follows the following example:

```
##             N      HN   CA(i)  CA(i-1)  HA1(i)  HA2(i) HA1(i-1)  HA2(i-1) ....
100.0 GLU12  124.23  7.65  56.32   60.54   3.43     0.00     0.00    0.00  ......
101.0 GLY?   101.31  7.05   0.00   43.41   2.43     0.00     3.21    2.12  ......
```

Any line that is started by either '##' or '%%' is considered a comment text and will be ignored.

The first column in any data entry is a peak identification number. This is a four digit number with one decimal place. The last digit may be incremented, for example, to indicated a single parent amide resonance that may give rise to more than one carbon or proton cross peaks, i.e. from minor protein conformations or overlapping resonance signals.

The second entry is an optional comment of up to eight characters. If there is a number in this field, the software will assume that this is a tentative assignment. This number is not used in the Monte Carlo protocol; however, is carried over to the output files. Using this field to add/change tentative assignment is useful in comparing output from multiple Monte runs.

Subsequent entries in each row are chemical shift frequencies and may be separated by spaces or comas. The first and second columns must contain the parent amide nitrogen and amide proton shifts, in either order (defined by the .par file). The following chemical shifts may be in any order, but must be consistent throughout the table. The order is user defined in the .par file. (see section on Parameter File)

Any chemical shifts at the very end of the list may be left blank if the shift is unknown. However, a unknown chemical shift field positioned within a series of known shifts must be entered 0.00 to hold its place (see example peak 101.0 above, no shift is identified with the intra residue CA).

4.4 Sequence File

The sequence file provides a primary sequence of the protein being studied. If a pdb file is available, the sequence file is not necessary. The sequence file is identified by the extension ".seq", and must be in the assignment sub-directory. The sequence file contains, in primary sequence order, the three letter codes (capital letters) for each amino acid, separated by one or more spaces, and or, carriage returns. Each string of text may not exceed 100 characters. (25 Amino Acids, if separated by a single space)

```
MET ALA GLY PRO
HIS GLY LYS LEU
```

The sequence file, in conjunction with the secondary structure file is used to create a pseudo pdb file for NOE matching. For a coil, or sheet secondary structures, a linear peptide is created, for residues identified as helical segments, an α-helical peptide will be created. This pdb file is written to "temp.pdb".

4.5 PDB File

The pdb file provides both the primary sequence of the protein being studied, as well as amide proton coordinates for matching NOE's. Monte will add amide protons to a crystal structure. The pdb file is identified by the extension ".pdb". If a pdb file is used as input, a seq file is not used, or needed. **Note:** The pdb file should end in "TER"

4.6 Secondary Structure File

The secondary structure file is identified by the extension ".ss". The secondary structure file provides information on the secondary structure of peptide segments. This information is used for both matching residue type probabilities as well as building a pseudo pdb file for NOE matching, in the absence of an X-ray or NMR structure file. The default secondary structure is random coil unless the helical or sheet segments are identified.

If a pdb file is available, then MONTE will calculate the secondary structure from the pdb file. However, the information in the secondary structure file will override the calculation from the pdb file. The existence of a ss file prevents the calculation of secondary structure from the pdb file.

The format of the ".ss" file is the number of helices, followed by subsequent lines that indicate the residues at the start and the end the helical segments. This is followed by the number of β-sheet segments followed by the starting and ending residues for each segment. In the example below, there are 3 helices, segments 10-15, 22-40, and 42-50, there are also 2 β-sheet segments, residues 2-5 and 17-20. These residues will match best with spin systems that more closely match the α-helical or β-sheet mean chemical shifts. All other residues will match the mean chemical shifts for a random coil.

```
3
10 15
22 40
42 50
2
2 5
17 20
```

If there are no β-sheets, or no α-helices, a zero must be entered followed by no segments. In the following example, there is no beta sheet secondary structures identified.

```
2
10 15
22 40
0
```

4.7 Specific Amino Acid Labels

Monte will use amide chemical shift data acquired on either a specifically ^{15}N labeled protein sample, or a specifically 1-^{13}C (carbonyl) and uniformly ^{15}N labeled protein sample in conjunction with shifts obtained from a 2-D HNCO filtered HSQC. In the former case, the pairs of amide chemical shifts identify the residue type of the (*i*) amide spin system. In the latter case, the pairs of amide chemical shifts identify amide spin systems of residues that are adjacent (*i*+1) to the label. Our experience is that the 1-$^{13}C/^{15}N$ labeling scheme is more robust, and does not suffer from transaminase dilution of the specific label. Matching is based on a comparison of the amide proton and nitrogen shifts in the specific labeling file to the spin systems that have been assigned to that particular residue type. The score is based on the same Gaussian distribution that is used to score inter-residue connectivities.

The specific label files are identified the extensions ".nsl" for the specific ^{15}N label HSQC experiment and ".csl" for the 1-$^{13}C/^{15}N$ 2D HNCO experiment. The format of these files are the three letter code (capital letters) for the amino acid followed by the amide nitrogen and amide proton chemical shifts. The order of these two chemical shifts MUST be consistent with the chemical shift data base file, and the order defined in the .par file. The following example is a pairs of shifts for an alanine specific label.

```
ALA  123.4  7.44
ALA  120.5  8.54
```

Identification of residues containing methyl groups and other residue specific experiments.
Additional information on the potential residue type of spin systems can also be incorporated into Monte. If the residue type is uncertain, e.g. for methyl containing residues, then all possible residues should be indicated. For example, if the an amide peak at 120.2 and 7.3 ppm was observed in a (H)(CC)(CO)NH TOCSY spectrum of a methyl protonated protein then the following should be include in the "csl" file:

```
ALA 120.2 7.3
VAL 120.2 7.3
LEU 120.2 7.3
ILE 120.2 7.3
```

4.8 NOESY data

Monte may use input from either a 4D Amide-Amide NOESY, or two 3D NOESY spectra, an H-NH HSQC NOESY, and an N-NH HSQC NOESY. The NOESY cross peaks are matched with those which are predicted from the pdb file, or secondary structure, within the NOE distance cut off (see Parameter file).

The 4D NH-NH NOE file is identified by the extension ".4dnoe". Each set of NOE data must be assigned to a parent amide peak. The file format contains the peak identification number followed by the parent amide nitrogen and amide proton chemical shifts. The order of these shifts and subsequent NOE pairs must be consistent with the chemical shift data base file, and the definition in the .par file.

```
##         PARENT       NOE1              NOE2         NOE3
##
100.0  124.2  7.65  123.3  8.54   120.4  7.24  128.4  9.00
```

The 3D H-NH and N-NH HSQC NOESY files are identified by either the file extension ".3dh_noe" and ".3dn_noe", respectively. Monte will run with only the 3D H-NH NOESY, however, this information alone is often too general to provide a useful constraint for finding a unique chemical shift assignment solution. Therefore, it is highly recommended that the complementary pair of information is used for input. The format of the 3D NOESY files is similar to the 4D input file described above; however, in these cases, the parent amide shifts are followed by the single proton, or nitrogen cross peak. For example:

```
##  3d H-NH HSQC NOESY
##         PARENT       NOE1    NOE2   NOE3
100.0  124.2  7.65    8.54    7.24   9.00

##  d3 N-NH HSQC NOESY
##         PARENT     NOE1    NOE2    NOE3
100.0  124.2  7.65  123.3   120.4   128.4
```

5. Parameter File

5.1 Overview

The parameter (.par) file is unique to each chemical shift assignment project, and must reside in the assignment project sub-directory. This is a plain text file; however, it is strongly suggested that this file only be edited via the graphical user interface (fm.tcl).
The parameter file provides Monte with several pieces of information, such as the root *filename* for the parameter, data , and output files. In addition, the directory path for the assignment problem is also in the parameter file. The parameter file also identifies to Monte what data files are available for input, the column

definitions of the chemical shift data base ".cs." file, and the parameters for chemical shift matching, as well as weighting factors for different scoring criteria.

When one launches fm.tcl, three windows will appear: the main window and one chemical shift definition window. An additional chemical shift definition window for utilizing existing chemical shifts will be opened if this option is selected. If the tcl code does not run, make sure the lastest versions of tcl/tk are installed and the symbolic links are set. (see Getting Started)

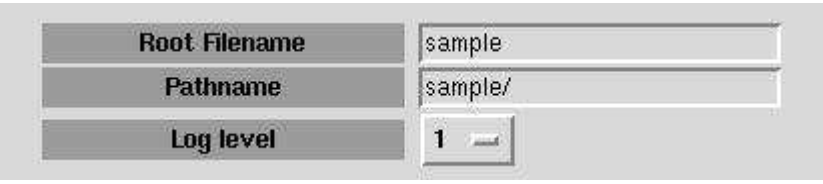
The main window is divided into two parts. The upper part contains fields for setting general parameters, and the lower part contains fields for setting the simulated annealing parameters. The very bottom of the man window contains a series of buttons to allow you to control a number of processes:

The function of these buttons is as follows:

- **HTML:** Launches a browser to view the HTML output file[1].
- **Corr Plot:** Launches a postscript viewer to view the correlation plot.
- **Solution Output:** Launches a postscript viewer to view the summary output.
- **Open CS ID table:** Opens the table that identifies the atom types in the input chemical shift file.
- **Open par file**: Opens a parameter file.
- **Save par file:** Saves the parameter file.
- **Start:** Runs Monte, prompts for a parameter file.
- **Kill:** Terminates current run of Monte
- **Quit**: Closes the tcl window.

5.2 Input Files

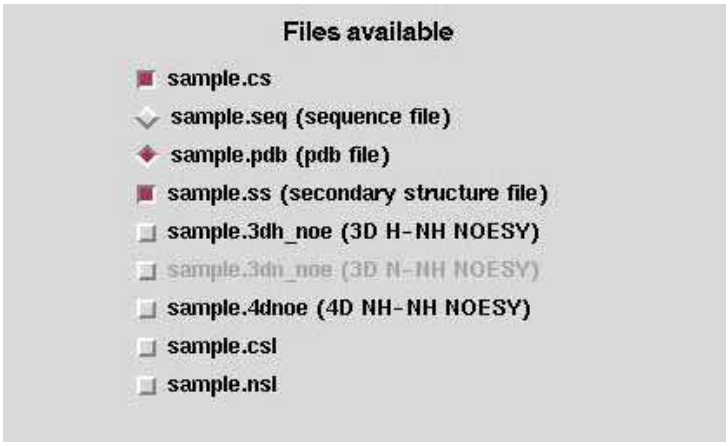
At the top of the GUI, there are two fields in which one should enter the root filename for the assignment problem and the subdirectory path (Pathname) for the assignment project.



When the *filename* in the Root Filename field above is changed, all the expected filenames for the next section are updated.
This section of the GUI also allows the user to select the level of information that is written to the *filename.log* file. The highest log level (3) should be used for diagnostic purposes. A log level of one is satisfactory in most situations.

[1] In order for the HTML, Corr Plot, and Solution Output buttons to work properly, the names of the appropriate applications have to be included in the fm.tcl file. See the last section of this guide.

Next, identify what files are available for the Monte run by clicking on the appropriate radio button (see below)



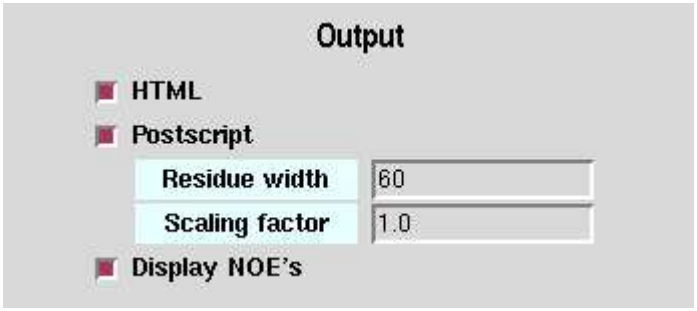
The ".cs" file MUST be available. Either the ".seq" or ".pdb" file MUST also be available. The other files are optional. One may only use the N-NH HSQC NOESY data if there is also H-HN HSQC NOESY data available, hence in the above example, the ".3dn_noe" choice may not be selected.

5.3 Output Files

Monte produces the following output files:

main.html: An HTML document summarizing chemical shift matching, NOE matching, and residue type prediction.
*filename*_out: The best assignment solution.
*filename*_out.ps: A postscript figure that summarizes the assignment data
solutions/*filename*_nn.soln: The best assignment solutions for *each* cycle of the Monte run
solutions/corr.ps: A postscript correlation plot

The GUI can be used to select whether the HTML or postscript output is desired. If the postscript output is selected then you can select how many residues to print/line and a scaling factor for the final output. For large proteins (e.g. >200 residues) it may be necessary to reduce the scaling factor to keep the output on one page. The NOE output on the postscript plot can be turned off if desired.



Additional detail regarding the output files can be found in section 8.

5.4 Run Parameters

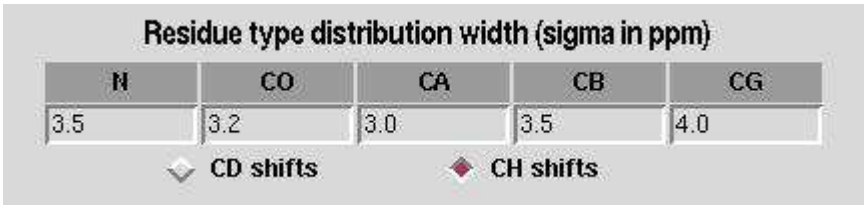
The parameters that control the scoring are defined in the next section of the top portion of the main window. You will probably have to scroll down in order to see these windows.

5.41 NOE distance cut-off.

In the next field enter the desired distance cut-off (Å) for NOE matching. Monte will match NOE chemical shifts that are consistent with the current assignment solution and are within the distance cut-off radius of the parent amide proton in the pdb structure file. If the pdb file is not available, Monte will create a pseudo pdb file (temp.pdb). In this

case, random coil or β-sheet defined residues are constructed as a linear peptide. If an NOE distance cut-off of 4.5A is selected then NOEs between adjacent residues will be generated for assignment purposes.

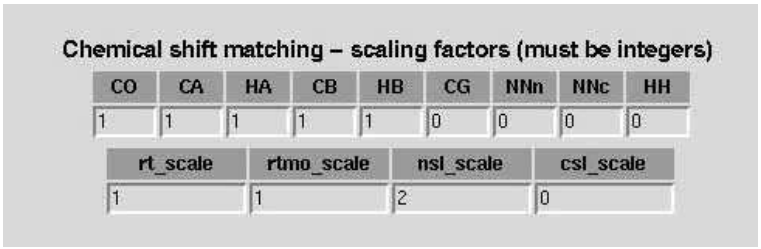
5.42 Residue Type Distribution Width & Deuterium Isotope Shifts



Monte applies a Gaussian distribution to the mean chemical shifts found in the file aaconv or aaconv_deut to determine residue type for the (i) and (*i*-1) residues of a spin system. One may wish to widen or narrow the width of the gaussian function. To do so, enter the half-height peak width (ppm) in the box for the appropriate atom type. Generally, these distributions need not be adjusted.

Directly below the distribution width table are two buttons, CD shifts and CH shifts. Select the button that is appropriate for your data, either protonated proteins (CH shifts) or perdeuterated (CD shifts).

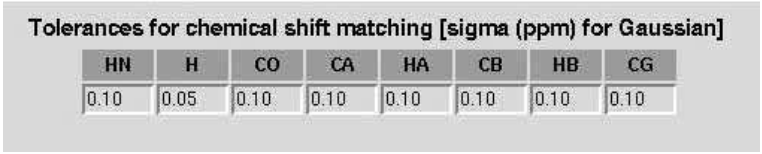
5.43 Weighting Factors



Monte internally scores a perfect score as 100 points. This maximum applies for all matching including for matching chemical shift for an atom type, matching a spin system to residue type, or specific labeling, or NOE matching. Less than ideal matching is scored appropriately (less than 100) based upon user-defined parameters. One may wish to give more mathematical "weight" to a particular data type. The next section is used to adjust the relative scaling of different data types. (The only exception is the relative NOE scale, this may be changed in the simulated annealing parameters, see Simulated Annealing parameter section.)

Entering a scaling factor of zero will turn off a scoring of that particular data type. In the example above, data from the ^{15}N specific label file will be multiplied by 2 during the Monte run. The entry NNn and NNc are for the N-terminal Nitrogen and C-terminal Nitrogen chemical shift matching (see Chemical Shift Definitions section).

5.44 Tolerence for Chemical Shift Matching



Monte scores matching between complementary chemical shift pairs [i.e. the CA^{-1} of residue (*i*), with the CA of residue (*i*-1)] by comparing the difference between the chemical shifts to a Gaussian distribution function. An exact match will be scored as 100 (multiplied by the weighting factor). The N entry is used for matching J-coupled shift obtained from the (H)N(COCA)NH type experiment, as well as amide NOE matching, and matching specific label shift in conjunction with the H atom entry.

The chemical shift matching is truncated at 1.00 ppm for all chemical shift matching. In the example in Table 1, a score of 0 is given to any pair of shifts that have a difference in chemical shift that is 3 times the matching tolerance. If the Gaussian extends past 1.00 ppm, All shifts differences that are 1.00 ppm or greater are set to a score equal to the baseline value. In some cases, the baseline may be set below zero (see Annealing Schedule section on Repulsive Terms). An example of the distribution for scoring is shown in Table 1.

5.45 Chemical Shift offsets

To determine reside type for the (*i*) and (*i*-1) residues, Monte compares the input chemical shifts to the median shifts for each residue type and secondary structure type found in the aaconv file (see aaconv section). One's chemical shift referencing may differ from the standard shifts. In this case, one may instruct Monte to apply a correction to the input data for residue type identification. Enter the values in the appropriate boxes provided. These corrections are only applied when determining residue type probabilities, and Monte will not actually change the chemical shifts between the input and output files.

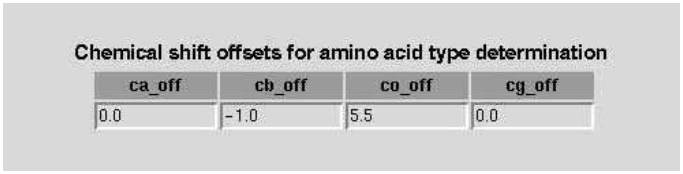


Table 1: Example of distribution of scoring for chemical shift matching for three different Guassian tolerence widths. The chemical shift difference between complementary shifts is indicated in the first column (ppm). In this example a tolerence is 0.3 ppm for CO, 0.15 ppm for CA, and 0.05 ppm for HA.

ppm	CO	Ca	Ha
0.00	100	100	100
0.01	99	99	98
0.02	99	99	92
0.03	99	98	83
0.04	99	96	72
0.05	98	94	60
0.06	98	92	48
0.07	97	89	37
0.08	96	86	27
0.09	95	83	19
0.10	94	80	13
0.11	93	76	8
0.12	92	72	5
0.13	91	68	3
0.14	89	64	1
0.15	88	60	1
0.16	86	56	0
0.17	85	52	0
0.18	83	48	0
0.19	81	44	0
0.20	80	41	0
0.21	78	37	0
0.22	76	34	0
0.23	74	30	0
0.24	72	27	0
0.25	70	24	0
0.26	68	22	0
0.27	66	19	0
0.28	64	17	0
0.29	62	15	0
0.30	60	13	0
0.31	58	11	0
0.32	56	10	0
0.33	54	8	0
0.34	52	7	0
0.35	50	6	0
0.36	48	5	0
0.37	46	4	0
0.38	44	4	0
0.39	42	3	0
0.40	41	2	0
0.41	39	2	0
0.42	37	1	0
0.43	35	1	0
0.44	34	1	0
0.45	32	1	0
0.46	30	0	0
0.47	29	0	0
0.48	27	0	0
0.49	26	0	0
0.50	24	0	0
0.51	23	0	0
0.52	22	0	0
0.53	21	0	0
0.54	19	0	0
0.56	17	0	0
0.57	16	0	0
0.58	15	0	0
0.59	14	0	0
0.60	13	0	0
0.61	12	0	0
0.62	11	0	0
0.63	11	0	0
0.64	10	0	0
0.65	9	0	0
0.66	8	0	0
0.67	8	0	0
0.68	7	0	0
0.69	7	0	0
0.70	6	0	0
0.71	6	0	0
0.72	5	0	0
0.73	5	0	0
0.74	4	0	0
0.75	4	0	0
0.76	4	0	0
0.77	3	0	0
0.78	3	0	0
0.79	3	0	0
0.80	2	0	0
0.81	2	0	0
0.82	2	0	0
0.83	2	0	0
0.84	1	0	0
0.85	1	0	0
0.86	1	0	0
0.87	1	0	0
0.88	1	0	0
0.89	1	0	0
0.90	1	0	0
0.91	1	0	0
0.92	0	0	0
0.93	0	0	0
0.94	0	0	0
0.95	0	0	0
0.96	0	0	0
0.97	0	0	0
0.98	0	0	0
0.99	0	0	0
1.00	0	0	0

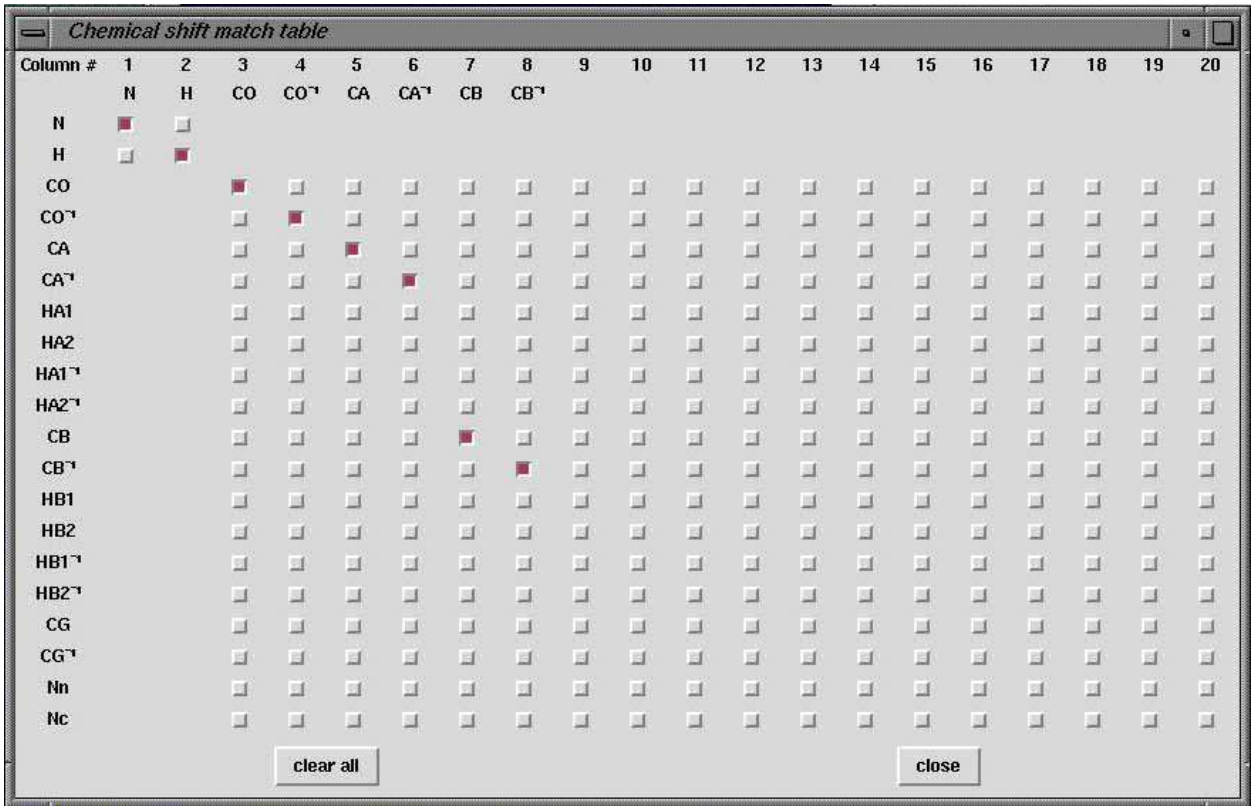
6. Chemical Shift Identification Table

The chemical shift identification table allows the user to define the chemical shift types that are listed in each column of the ".cs" file. The "N" and the "H" entries are the parent amide nitrogen and amide proton chemical shifts. These two shifts must be in either of the first two columns of the chemical shift data base file.

Only one column may be selected for one atom type listed in each row. To redefine a row or colum, the row or column must first be deselected. The intra residue (*i*) chemical shifts for the parent amide nitrogen pair are identified by CO, CA, HA, etc. Inter-residue shifts (*i*-1) (i.e. from the preceeding residue) are identified by the same atom names followed by the superscript "-1". The atom types Nn and Nc refer to the N-terminal nitrogen shift and the C-terminal nitrogen shift. These correlations are derived from the following experiments

Nn : (H)N(CACO)NH
Nc : (H)N(COCA)NH

In the following example, the Chemical shift matching table is consistent with the ".cs" input file that contains the parent amide nitrogen shift in column 1, the parent amide proton shift in column 2, the inter-residue CO shift in column 3 (i.e. from the HN(CA)CO experiment), the inter-residue CO shift in column 4 (i.e., HNCO experiment), the intra-residue CB shift in column 7, and the inter-residue CB shift in column 8. If there happen to be shifts in columns 9 and above, the chemical shifts will be carried over to the Monte output, however, they will NOT be used by the program to obtain the solution.



Note about aliphatic Proton matching: Monte will match HA1 and $HA1^{-1}$ chemical shifts without the HA2 shifts being defined. If both shifts are identified (as in the case of Glycine residues), however, Monte will find the best combination of the two shifts. If one desires to match HB chemical shifts, however, both

HB1, HB2 and the complement $HB1^{-1}$ and $HB2^{-1}$ must be defined. Monte will find the best combination between these four chemical shifts.

If the window shown above is closed, it may be opened again by selecting the "Open CS ID table" button at the bottom of the Main window.

7. Matching old Assignments

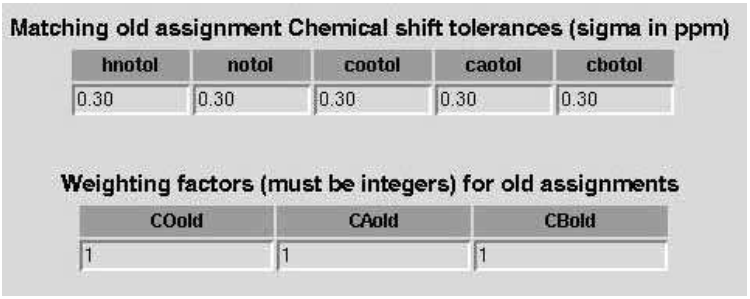
One of the unique features of Monte is that the program can match a chemical shift assignment solution from an previous Monte output file. An example of this approach is provided in the rho_dna directory that is included in the distribution. This feature is useful when working on chemical shift assignments of a particular protein, and subsequent ligand bound forms of the same protein. The assumption here is that the presence of the ligand only significantly perturbs chemical shifts of residues that are in the vicinity of the binding site, or are involved in structural changed due to the ligand binding. In this case many of the chemical shifts obtained for the unliganded form of the protein provide a useful constraint for assignment of the ligand bound form of the protein (or, vice versa). To match an old assignment solution, Monte compares how well cross peaks in the same type of 3D spectra match. Monte will match all or some of HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCB, and HN(CO)CB types data.

To match an old output, select "match old output". One must also tell Monte what columns to find the desired data. These column definitions identified in a manner similar to that described above. (See Chemical Shift Definitions). Open the old CS match table to set the column definitions.

The user must also provide the path and filename of the old Monte output file. One may bowse the directory structure for the desired file.

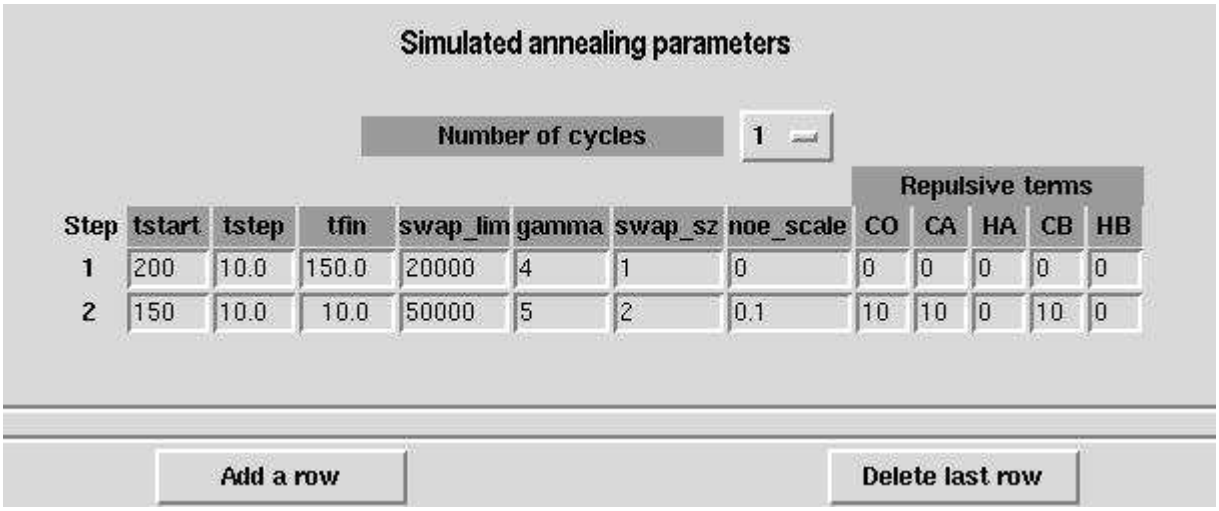
Monte will only match chemical shift assignments in this old output file that are indicated as a confident assignment. A confident assignment is indicated in the file with a "+" symbol (see Output File section). One may wish to first edit this file to remove some "+" symbols (see Advice section)

The tolerance for chemical shift matching follows the rules discussed above. Enter the desired matching tolerance and weight factor in the appropriate field.



8. Simulated Annealing Parameters

The lower section of the GUI is used to set the simulated annealing schedule. The annealing schedule may have a single, or multiple steps. To add, or delete a step, click "Add a row" or "Delete last row", respectively.



8.1 Annealing Schedule

Each Annealing schedule step has several user adjustable features.

8.11 Number of cycles:

This specifies the number of times, or cycles, the annealing schedule will be run. Since the initial assignments are random, these will be independent solutions.

8.12 tstart/tstep/tfin:

The parameter tstart defines the starting "temperature" for that annealing step. After the number of swapping cycles is reached, the temperature of the system is lowered by tstep until the temperature is less than or equal to tfin. Following which, the next annealing step is started. At the end of each annealing step (in addition to the initial random assignment) a text output file is written to the experiment sub-directory. The output file extension is incremented as the annealing step in incremented (see Output Files section).

Note on annealing temperature: Monte is designed to provide a general method for chemical shift assignments. Different chemical shift assignment problems will result in higher or lower overall scores, as well as variable magnitude of score differences while comparing swapped blocks of assignments. These score difference are a function of the amount of data and the relative weighting factors used for the Monte run. With this in mind, the "melting temperature", or transition between the temperature at which all swaps are accepted and better scores a preferentially accepted will vary. The annealing schedule should be adjusted accordingly. Monte provides some scoring and swapping data during simulated annealing process to help the user adjust the annealing schedule.

8.13 swap_lim/gamma

The parameter ncyclim defines the number of cycles that should be completed before the temperature is lowered. Since lower temperatures result in a fewer number of accepted swaps, it is useful to increase the number of cycles per interation as the temperature is lowered. The parameter gamma defines the rate of increase of ncyclim as temperature is lowered for a particular annealing step. The desired result is to normalize the number of successful swaps over the course of the annealing step.

8.14 swap_sz

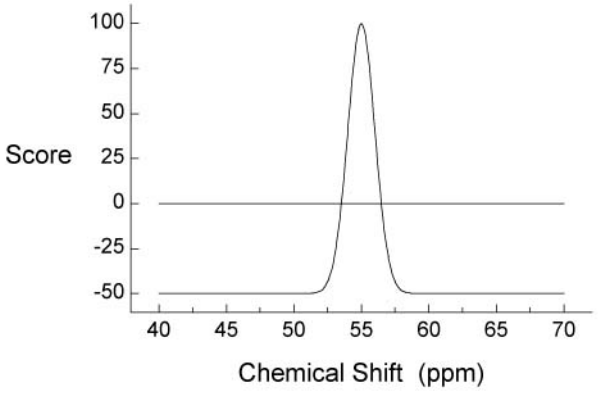
The parameter n_swap defines the maximum number of adjacent spin systems that will be swapped during that annealing step. The size of the blocks that are swapped is chosen at random from 1 to n_swap. Initially, n_swap should be low, i.e. 1, since it is unlikely that the random assignment resulted in correctly assembling linked spin systems. As the temperature is lowered, more spin systems will be better linked, and the problem becomes finding the correct position in the primary sequence of the protein. Therefore, at later stages of the annealing schedule, it may be advantageous to swap larger blocks of spin systems.

8.15 noe_scale

Matching predicted NOE's (from a pdb file) to experimental NOE's is a very powerful assignment tool. This information becomes most useful to Monte after some spin systems become linked and better assignment. In addition, since, Monte internally scores all perfect matches to 100, and a single parent amide may have multiple NOE's, NOE matching may over power J-coupled matching. For these reasons, the NOE scaling factor may be adjusted at each annealing step. Also, in contrast to other scaling factors, the NOE scaling factor may be less than 1. It may be advantageous to start out a Monte run with a scaling factor of zero. As the "melting temperature" of the system is reached slowly increase the NOE scale to desired value.

8.16 Repulsive Terms

As Monte assembles linked spin systems, it may be advantageous to penalize linked assignments that have chemical shift differences much larger than the desired tolerance set in the parameter file. In this case, one may add a repulsive term to the chemical shift matching calculation. Enter the desired repulsive term in the field for the appropriate atom type. The baseline of the Gaussian matching function will be lowered by this value. For example, if a repulsive term is set to 50, The matching score for an ideal match will be given a score of 100, a chemical shift difference greater than ca. 3 times the tolerance will be given a score of –50, as shown in the figure to the left. In all cases; however, a missing chemical shift (i.e. and entry of 0.0) will not penalize a solution.



8.2 Saving the Parameter File

After all the desired parameter values are set, one may save the parameter file by clicking "Save Parameter File". The software will save the parameter file and check to see if the necessary files are available, the gui will warn the user if there are any inconsistencies between the files marked as available and files present in the experimental sub-directory.

9. Data Output files

Monte writes several forms of output during the simulated annealing run, as well as at the end of the run. These output files are useful in determining the quality of an assignment solution, and help the user make decisions concerning the annealing schedule and other parameters. The log file is also useful to trouble shoot problems that may arise with files that may have a format, which is incompatible with Monte.

9.1 Log file

Monte creates a ".log" file in the experimental sub-directory. The amount of information written to this file is controlled by the log level parameter.

log level=0	Nothing is logged.
log level=1	Only the filenames of input data, and standard output are logged.
log level=2	In addition to log level=1, most data read in from input files is written to the log file This includes chemical shift database, pdb file, specific label files, etc
log level=3	In addition to log level=2, all internally calculated data is logged.

9.2 Standard Output

During the simulated annealing, Monte with write text to the unix shell indicating the progress of the annealing schedule. An example is shown below

```
anneal                                                    max    max   NOE
sched  temp      ncyc    n-suc    n-rej     score      score  swap  scale
2     140.0     10001     2712     7289     28971      30875    2   0.1
-------------------------------------------------------------------------
|     CO      Ca       Ha      Cb      Hb      Cg     Nn/Nc            |
|    2745    2436    4304    2132    6387       0       0              |
-------------------------------------------------------------------------
|   rt(i)   rt(i-1)  N(15)  1-C(13)  | NH      NH      HA       HB     |
|    6111    4308       0       0    | NOE     548                     |
-------------------------------------------------------------------------
```

The top row of this output shows the current step number of the annealing and the current temperature. The value ncyc indicated the number of cycles for that temperature. The values n-suc and n-rej indicate the number of successful swaps and number of rejected swaps, respectively. Score and max Score indicate the current overall score for the solution and the highest score that has been reached over the current annealing step. Max swap indicated the maximum number of residues that are swapped, and NOE scale indicated the current scaling factor for NOE matching.

The next section of text indicated the total score calculated for each type of chemical shift matching. (These values are a result of the internal score multiplied by the scaling factor). The values reported here might be negative if repulsive terms are greater than zero in the annealing step parameters. (see Repulsive Terms)

The following section reports the score given for residue type matching. The first two indicate the total score for matching residue type probabilities (based upon chemical shifts) for the (*i*) and (*i*-1) residues, respectively. The value for N(15) reports the total score given for matching ^{15}N specific label data, and likewise the 1-C(13) value reports the total score for any 1-^{13}C/^{15}N type specific label data.

The next section in this row reports the total score for NOE matching. Currently Monte only takes advantage of amide/amide NOE data. Hence the HA and HB fields are blank.

9.3 Text Output

At the end of each annealing step (as well as the initial random assignment), Monte writes a text file to the experimental subdirectory of the current best assignment solution. The text files have the name *filename*_out.00, *filename*_out.01, etc, where *filename* is the root filename of the experimental data. The best solution from all of the different runs is placed in the fine; *filename*_final. The format of the text output file is indicated by the following example:

```
202  PRO   0.0   0   0.00   0.00   0.00   0.00   0.00   0.00     0.00   0.00
203+ ILE  91.0 203 119.56   8.49  64.46  58.76  35.59  28.70   174.56 173.09
204+ ASN  19.0 204 110.61   8.51  48.27  64.42  39.30  35.61   173.11 173.82
205  GLY   0.0   0   0.00   0.00   0.00   0.00   0.00   0.00     0.00   0.00
206+ ASN  52.0 206 116.07   7.42  48.60  43.08  35.83  -1.00   170.42 174.08
207+ GLY  10.0 207 106.79   8.10  42.68  48.57  -1.00  35.81   174.04 171.06
208+ LYS  60.0 208 117.39   6.97  51.77  42.75  29.85  -1.00   171.08 173.04
209+ GLN 110.0 209 122.11   7.59  54.03  51.84  25.72  29.85   173.02 174.24
%%
%%
%%  Peaks not used
%%
161.0   0 115.71   6.63  52.90  61.01  37.81   0.00   0.00   0.00   0.00
```

The first field in each row is the residue number. This number may be followed by a "+". The plus symbol indicated that the reported spin system was assigned that residue number in all of the cycles. For example, if 10 cycles were selected, and a spin system was assigned to the same residue in all of the cycles, then a '+' would appear in the 1st column. Entries that do not have a "+" indicate that two or more different spin systems have been assigned to that residue.

> **Warning:** The "+" residues is an indication of a ""possible" correct assignment. One must further verify the goodness of this assignment by evaluating the goodness of chemical shift matching, NOE matching, residue type prediction, and specific labeled sample matching, As well as the consistency of this assignment over the course of several Monte runs. (Note: the inclusion of an assignment in the 50 highest scored solutions is dependent upon the amount and quality of the data input, relative scaling factory and annealing schedule used. Care should be taken when analyzing these results). (also see HTML section)
>
> **Additional note:** if one chooses to match an old Monte output file, this "+" symbol is used to indicate a confident assignment that will be matched to the new solutions. Before doing so, one may wish to remove some "+" symbols from the old output file.

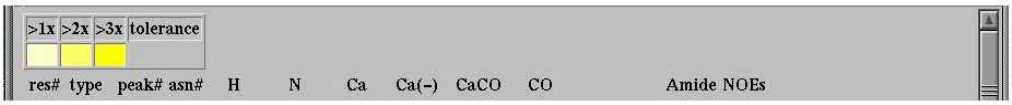
The next column indicated the three-letter code for the residue. This is followed by the spin system number, any "assignment" that is found in the comment line of the ".cs" file, followed by a list of the chemical shifts in the order of the ".cs" file.

Monte uses an additional swap area that does not add any score to the solution. If a best fit solution is not found for a spin system, it will listed at the end of this text file under "peaks not used". This line is preceded by "%%", so that, if one would like to use this output for matching chemical shifts, these lines will be ignored when read by Monte.

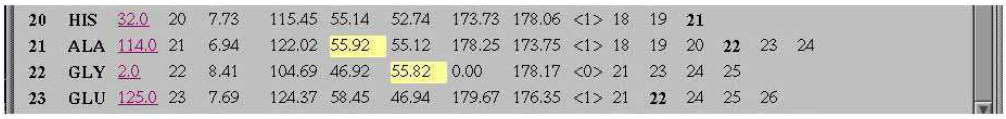
9.4 HTML

To help in visualizing the assignment solution and determining the quality of the assignment solution, at the end of a run, Monte writes four html files. Use an Internet browser to open the main.html file in the experimental subdirectory. This html document will set up several frames on you browser.

The top frame contains the assignment information. At the top of this frame there is a color code key which indicates the colors that highlight chemical shift mismatches – shades of yellow indicating mismatches that are greater than 1 times the set tolerance, greater than 2 times the tolerance, and greater than 3 times the set tolerance. The next row lists the column headers. The residue number and residue type are taken from the pdb file or sequence file. The next columns are the peak number and assignment number (if found in the ".cs." file). These columns are followed by the chemical shifts atom types that are identified in the .par file.



Shown below is an example of the body of this html frame. In this example the CA matching tolerance was set to 0.04, therefore the CA shifts between residue 21 and 22 are highlighted in the medium yellow shade indicating that the difference in the shifts between 2-3 times the chemical shift tolerance.



20	**HIS**	32.0	20	7.73	115.45	55.14	52.74	173.73	178.06	<1>	18	19	**21**			
21	**ALA**	114.0	21	6.94	122.02	55.92	55.12	178.25	173.75	<1>	18	19	20	**22**	23	24
22	**GLY**	2.0	22	8.41	104.69	46.92	55.82	0.00	178.17	<0>	21	23	24	25		
23	**GLU**	125.0	23	7.69	124.37	58.45	46.94	179.67	176.35	<1>	21	**22**	24	25	26	

Residue numbers and residue type names that are shown in bold face type indicate that this assignment is consistent with the highest 50 scores solutions for the last annealing step. Entries that are not bold face type indicated that the peak assignment for this residue changed over the 50 highest scores. The bold type residue is an indication of a ""possible" correct assignment. One must further verify the goodness of this assignment by evaluating the goodness of chemical shift matching, NOE matching, residue type prediction, and specific labeled sample matching, As well as the consistency of this assignment over the course of several Monte runs. (Note: the inclusion of an assignment in the 50 highest scored solutions is dependent upon the amount and quality of the data input, relative scaling factory and annealing schedule used. Care should be taken when analyzing these results)

Following the chemical shift is a number in brackets "<>" this number indicated the number of NOE shifts that were input for that spin system. The following numbers residue numbers that are predicted to have NOE's based on the pdb file and the distance cut-off value. A number in bold face type indicates that the predicted NOE is satisfied by an NOE chemical shift that was provided.

The peak number in the top frame is a hyper link that will change the data displayed in the lower frame.

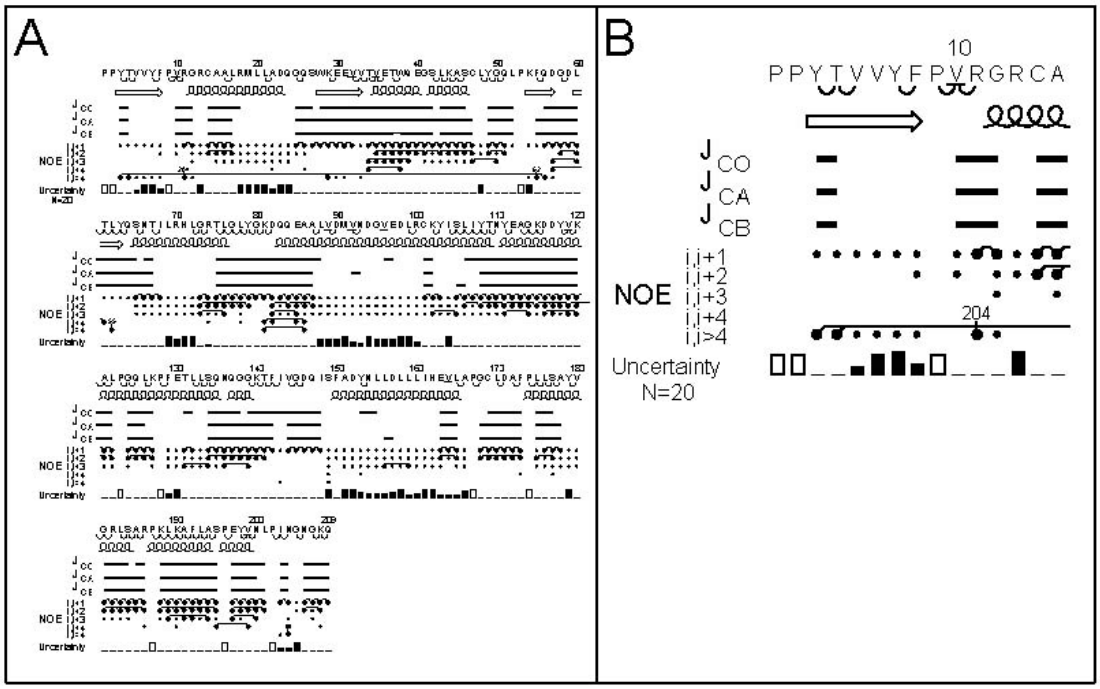
The lower fame indicated the residue type probabilities calculated by Monte for the (*i*) and (*i*-1) residues for that spin system. The greatest value is indicated in bold face type. In the example here, one can see that peak# 2.0 has an 86% probability of being a glycine based upon the N, and CA shifts (note than no CO shift is input for this residue type). This window also reveals that the inter residue CA and CO shifts alone do not define the residue type of the preceding residue of this spin system very well.

res	ALA	VAL	LEU	THR	SER	ILE	GLY	ARG	ASN	GLU	GLN	ASP	TRP	TYR	PHE	LYS	CYS	MET	HIS	PRO
Peak # 2.0																				
(i-1)	3	2	7	1	4	3	0	7	3	**8**	6	6	7	6	7	**8**	7	**8**	5	2
(i)	2	0	1	1	1	0	**86**	1	2	1	1	1	1	0	0	1	1	1	1	0

This information is helpful to determine if a chemical shift offset should be added/or subtracted to the input shifts in calculating residue type information (see Parameter File section).

9.5 Assignment Summary (Postscript)

A graphical summary, in postscript form, of the final assignment solution is provided in the main directory as the file *filename*_out.ps. An example output file is shown below (left side) and an enlarged view of the image is shown on the right.



The amino acid sequence is given at the top of each row. The half-circle symbol under two adjacent residues indicates that this assignment agrees with information from specific ^{13}C-1 labeling. For example, the amide of Thr4 was identified as being coupled to a ^{13}C carbonyl in a sample that was uniformly labeled with ^{15}N and specifically labeled with ^{13}C-1 Tyr. An underlined residue indicates that the assignment is consistent with specific ^{15}N labeling. For example, the amide resonance for the spin-system that was assigned to Val10 was present in a sample labeled with ^{15}N Val. The following line indicates the secondary structure, either provided by the user or calculated from the three-dimensional structure of the protein. The next three lines, labeled with J_x, indicate the fidelity of matching of inter- and intra-residue chemical shifts. The thickness of the bar is related to how well the chemical shifts match. Thinner lines indicate poorer matches, such as the J_{CB} coupling between residues 37 and 38 (panel A). In panel B, all of the indicated matches are within one σ of the Gaussian function that was used in scoring, hence all have the same thickness. The rows that display the information from NOE data show consecutive NOEs (i, i+1), local NOEs, as well as long-range NOEs (i, i>4). An open circle indicates that the NOE was predicted from the input structure (either primary or tertiary). A filled circle indicates that the NOE was found in the experimental data. Where possible, filled circles are connected by a solid line between the two coupled amide protons. For example, panel A shows long-range NOEs between residues 3 and 4 and 55 and 56, respectively. In the three dimensional structure these residues are across from each other on a β-sheet. If it is not possible to connect the coupled residues by a line, because they reside on separate lines in the output figure, then the NOE partner is indicated by the residue number. For example, Arg11 shows an NOE to Asn204. The last line of this figure provides information on the uncertainty of the assignments. A bar is printed under each residue, with a height that is proportional to the number of different spin systems that were assigned to this particular residue in the independent trials. A zero height bar (thin line) indicates that the same solution was found in all of the independent trials. For example Tyr3, Thr4, Val10, Arg11, Gly12, Cys14, and Ala15 appear to be uniquely assigned with this data. In contrast, residues Val5-Phe8 are not assigned for reasons discussed in the text. Open rectangles mark the location of Pro residues.

9.6 Correlation Plot

Additional detail on the existence of alternative assignments is presented in a correlation plot. This is a postscript file that is found in the 'solutions directory.

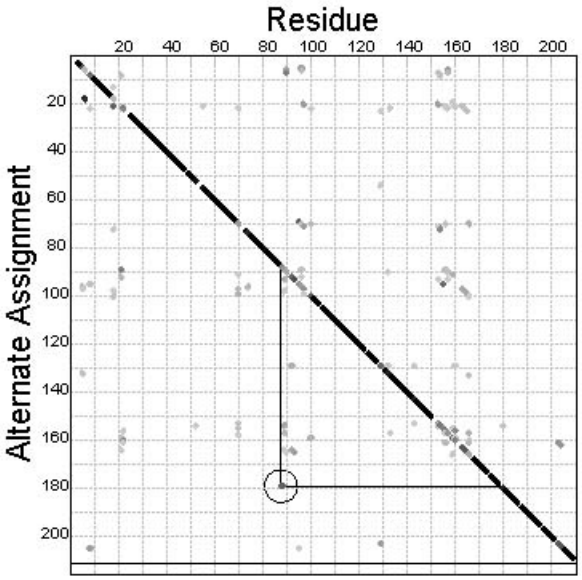


Figure Legend: The x-axis indicates the residue number and the y-axis indicates to what other residues, besides that found in the best solution, the spin system was assigned to. For example, the circled point in the plot indicates that the spin system that was most frequently assigned to residue 88 is also assigned to 179 in one or more of the 20 solutions. The intensity of the plotted point is proportional to the frequency that a spin-system is assigned to a particular residue. The more darkly colored the point, the higher the frequency of the assignment of a single spin-system to a particular residue. For example, the dark diagonal in this plot shows that most spin-systems are uniquely assigned to a single-residue. The rectangular area below the lower horizontal line represents the cache area. Points found in this region indicate that spin-system was placed in the cache (i.e., unassigned) in some or all of the assignment solutions.

10. Customizing and Running Monte.

10.1 Customizing.

If you desire to use the buttons to launch the browser and postscript viewers for viewing the output files then it is necessary to tell Monte what the names of the applications are. This is done by editing two lines at the beginning of the tcl/tk script, fm.tcl:

- set browser netscape
- set psviewer xpsview

You should replace "netscape" and "xpsview" with the names of the browser and postcript viewer on your system.

10.2 Running Monte

1. The most important thing to remember is to save the parameter file before starting a run, Monte always reads the parameter file, it does not directly obtain the information from the tcl/tk window.
2. The initial annealing temperature should be high enough so that most proposed swaps are accepted.
3. The final annealing temperature should be low enough so that the score has stabilized.
4. The annealing schedule should be slow enough to maintain equilibrium. Try doubling the number of attempted swaps (swap_lim) at each stage of annealing, and check whether the final score improves.
5. Inspect the html output. If all the chemical shift matches are within tolerance (i.e. no yellow, orange or red blocks), you can afford to tighten the tolerances for chemical shift matching.
6. View the correlation plot (corr.ps). If long stretches of connected chemical shifts are swapped between different residues, you may want to increase the scaling factors for residue type RT(i), RT(i-1), and specific labeling (nsl and csl).