MULCh: ModULes for the analysis of Contrast variation data
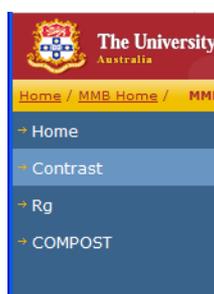
The following set of modules allows the extraction of structural parameters from neutron contrast variation data from complexes composed of two regions with differing neutron scattering densities.

1. *Contrast*: Determines the contrast ("scattering power" relative to the solvent) for each subunit in the complex for various proportions of D_2O in the solvent. Knowledge of the contrast is important for experimental planning and data analysis using the other programs. Also calculated are various parameters that are useful for the analysis of small-angle X-ray scattering data.

2. *Rg*: Analyzes the dependence of the radius of gyration upon contrast. This analysis allows the radii of gyration of each subunit and their separation to be determined.

3. *COMPOST*: Decomposes a contrast variation series into three scattering profiles representing the scattering from each subunit in isolation and a profile related to the disposition of each.

Testing and basic description of the work can be found in A. E. Whitten, S. Cai and J. Trewhella, *J. Appl. Crystallogr.* **41**, 222-226 (2008). If any of these programs are used in experimental planning or data analysis we ask that you cite this article in any resulting publications.

Contrast

The *Contrast* module can be used for both experimental planning and optimisation of contrast parameters after data collection. To start the *Contrast* module click on "*Contrast*" in the menu on the left of the web-page (see below). The other modules can also be accessed in this manner.

Experimental planning: If you have previously run the *Contrast* module and saved the input file, you can re-load all data by clicking on the [Upload Contrast File] button, otherwise you need to create the input file by filling in the required data.

Firstly the project title should be entered, which should summarise the calculation.

Project Title: | KinA:SDA 2:2 complex in 200mM NaCl, 50mM Tris, 150mM imidazole |

For planning purposes leave **Number of measured contrast points:** [0 ▾] as is; this option is relevant at later stages for optimisation of the contrast parameters. The next step is to fill in the solvent conditions. If the solvent is only composed of water, then you may leave **Number dissolved species in the solvent:** [0 ▾] and move to the next section. Otherwise set this quantity to the number of different compounds dissolved in the solvent. In the example below, three common compounds are entered as molecular formulae (note that the radio-box "*Molecule*" is selected): sodium chloride (NaCl), tris ($C_4H_{11}NO_3$), and imidazole ($C_3H_4N_2$). The concentration of each must be entered in mol/L, and the volume of <u>each formula unit</u> must be entered in $Å^3$. If the volume is left as zero, then the program will attempt to calculate the volume; however, you should check that the atomic volumes are specified and appropriate for the conditions (Table 1). Also when entering molecular formulas, brackets are not handled, so they must be expanded manually (i.e. $(CH_3)_3 = C_3H_9$). Capitalisation must also be correct (i.e. Cl is valid, while CL is not). Most additives make little difference to the scattering length density of the solvent unless they are at high concentration.

Substance Type	Formula	Conc. (mol/L)	Volume ($Å^3$)
○Protein ○DNA ⊙Molecule ○RNA	NaCl	0.200	0.0
○Protein ○DNA ⊙Molecule ○RNA	C4H11NO3	0.050	0.0
○Protein ○DNA ⊙Molecule ○RNA	C3H4N2	0.150	0.0

The next step is to enter information regarding each component of the complex studied. For testing purposes it is easiest to leave the fraction of exchangeable protons accessible to the solvent set to the default values, and let the program calculate the volumes where possible. In the example below, contrasts would be calculated for a 1:1 complex between two proteins, where the first protein is protiated (no deuterium incorporated into the protein), and the other is deuterated to a level of 85%. The sequence information can be in upper and/or lower case and may contain spaces. The bound volumes of various amino acids and nucleotides used by the program are given in Tables 2 & 3.

Number of components in subunit 1: 1

Deuteration level(0 - 1):	Fraction of exchangeable protons accessible by the solvent:		
0.0	0.95		

Substance Type	Formula	N_molecules	Volume (Å³)
⊙ Protein ○ DNA ○ Molecule ○ RNA	GSHMTEELMLKSEKLSIAGQLAAGIAHEIRNPLTAIKGFLQLMKPTMEGNEH	1	0.0

Number of components in subunit 2: 1

Deuteration level (0 - 1):	Fraction of exchangeable protons accessible by the solvent:		
0.85	0.95		

Substance Type	Formula	N_molecules	Volume (Å³)
⊙ Protein ○ DNA ○ Molecule ○ RNA	GSMRKLSDELLIESYFKATEMNLNRDFIELIENEIKRRSLGHIISVSS	1	0.0

If a given subunit is coordinated to significant numbers of metal ions, ligands or contains glycosolated residues, the number of components in that subunit can be increased to account for these differences. In the example below, the protein has six N-acetyl glucosamine molecules, twelve mannose/glucose molecules and four calcium ions bound. MULCh v1.1 can now account for exchangeable hydrogens in molecules associated with the protein. For the example below, the formula for NAG would be specified as C8H10**X3**NO5, and mannose as C6H7**X3**O5 (i.e. an 'X' is the symbol for an exchangeable hydrogen atom).

Deuteration level (0 - 1):	Fraction of acidic protons accessible by the solvent:		
0.0	1.0		

Substance Type;	Formula	N_molecules	Volume (Å³)
⊙ Protein ○ DNA ○ Molecule ○ RNA	GSMRKLSDELLIESYFKATEMNLNRDFIELIENEIKRRSLGHIISVSS	1	0.0
○ Protein ○ DNA ⊙ Molecule ○ RNA	C8H13NO5	6	0.0
○ Protein ○ DNA ⊙ Molecule ○ RNA	C6H10O5	12	0.0
○ Protein ○ DNA ⊙ Molecule ○ RNA	Ca	4	0.0

Once the input has been completed, the calculation can be submitted and the program will return relevant information regarding the complex. The top of the file shows two buttons, both of which are irrelevant for experimental planning. It also contains a link to the input file created for your calculation - if you wish to redo this calculation in future, or use it as a starting point, right-click on this and save the input file; if the output is required, save the html file. After this, the output summarises the information that has been input into the program - this information should be checked carefully to ensure a mistake has not been made, and that the program is functioning correctly. Directly below the input data summary (see below) is a table providing a summary of relevant molecular properties of the solvent and each subunit in ^1H_2O.

Solvent and particle information in H_2O

	Volume (Å^3)	Σb_X (10^{-13}cm)	Σb_N (10^{-13}cm)	ρ_X (10^{10}cm^{-2})	ρ_N (10^{10}cm^{-2})	Mass (kDa)
			Solvent			
54.56	30.0	28.20	-1.68	9.409	-0.560	0.018
0.20	29.3	78.96	13.21	26.958	4.509	0.058
0.05	167.7	194.58	0.98	11.600	0.058	0.124
0.15	74.9	101.52	23.69	13.554	3.163	0.068
Total	-	1579.23	-85.36	9.510	-0.514	-
Subunit (1) - Number of hydrogen atoms = 1819.0, Number of exchangeable hydrogen atoms = 377.0						
1.00	31347.5	38250.48	5528.61	12.202	1.764	25.338
Total	31347.5	38250.48	5528.61	12.202	1.764	25.338
Subunit (2) - Number of hydrogen atoms = 400.5, Number of exchangeable hydrogen atoms = 94.5						
1.00	6855.8	8427.57	3920.52	12.293	5.719	5.845
Total	6855.8	8427.57	3920.52	12.293	5.719	5.845

The information in this table includes scattering length densities (ρ), volumes and masses. In calculating the various properties of the solvent, the volume of water displaced by the dissolved substrates is taken into account. This means that the contrast is correctly calculated, even in the presence of significant quantities of glycerol or other compounds, and explains why in the case above the molar concentration of water (54.56) is less than 55.41 mol/L (at 20 ºC).

Listed next in the output are the various mathematical dependencies of the scattering length density and contrast upon the proportion of D_2O in the solvent. These values are also tabulated, along with the match point of both subunits in solution. The mathematical forms of the expressions allow straightforward calculation of (for example), the deuteration level required for a molecule to be solvent matched in pure D_2O.

Tabulated scattering length densities and contrasts

	ρ (10^{10}cm^{-2})			$\Delta\rho$ (10^{10}cm^{-2})		
	1	2	Solvent	1	2	Total
X-RAY	12.202	12.293	9.510	2.692	2.782	2.708
NEUTRON						
0.0	1.764	5.719	-0.514	2.278	6.233	2.987
0.1	1.883	5.862	0.170	1.712	5.692	2.427
0.2	2.002	6.006	0.854	1.147	5.151	1.866
0.3	2.121	6.149	1.538	0.582	4.611	1.305
0.4	2.239	6.293	2.223	0.017	4.070	0.744
0.5	2.358	6.436	2.907	-0.548	3.529	0.183
0.6	2.477	6.580	3.591	-1.114	2.989	-0.377
0.7	2.596	6.723	4.275	-1.679	2.448	-0.938
0.8	2.715	6.867	4.959	-2.244	1.907	-1.499
0.9	2.834	7.010	5.643	-2.809	1.367	-2.060
1.0	2.953	7.154	6.328	-3.374	0.826	-2.621
Calculated match-point (f_{D_2O})				0.403	1.153	0.533

The final section of the output files lists some additional expressions for the dependency of the molecular mass and solvent density upon D_2O, which may be of interest.

Optimisation of contrast parameters: Once a contrast variation experiment has been performed the *Contrast* module can be used to optimise (manually) the contrast parameters. The *Contrast* module is able to estimate the experimental match-point of the entire complex in solution from the zero-angle intensities.

Number of contrast points: 7

$f_{D_2O}(0 - 1)$	$I(0)$	$\sigma(I(0))$	Protein conc.
0.00	0.541	0.002	11.9
0.10	0.352	0.003	11.9
0.20	0.223	0.002	11.9
0.40	0.0745	0.002	26.9
0.80	0.19	0.001	11.9
0.90	0.332	0.002	11.9
1.00	0.537	0.001	11.9

To enable this functionality, change the number of contrast points to the appropriate value and then enter the various parameters. The program will then perform a linear fit to $\sqrt{I(0)}$ vs. f_{D_2O}, and the X-intercept of this plot is the match point for the entire complex. This quantity can then be used as a guide for adjusting the molecular volumes of the individual components. In the output, the estimated standard deviation of the 3rd decimal place of the match-point is shown (see below). This error estimate does not account for any systematic effects, so some discretion should be used when deciding how close the calculated and experimental match points should be.

1.0		2.933		7.134		0.328		-3.374	0.826		-2.621
Calculated match-point (f_{D_2O})								0.403	1.153		0.533
Experimental match-point (f_{D_2O})								-	-		0.501(4)

Using the contrast values with the other modules: At the top of the output for the *Contrast* program there are two buttons: [Continue for Rg Analysis] [Continue for Compost Analysis]. Clicking these buttons moves you to either of the other modules, copying the contrast values across, where you will see something like that below.

$\Delta\rho_1 =$	-5.65946	$f_{D_2O} +$ 2.27794	$Vol_1 =$	31353.110000 $Å^3$	[Pre-fill table]
$\Delta\rho_2 =$	-5.41482	$f_{D_2O} +$ 6.22848	$Vol_2 =$	6861.400000 $Å^3$	

If the D_2O fractions are then entered:

f_{D_2O}
0.0
0.1
0.2
0.4
0.8
0.9
1.0

Clicking on [Pre-fill table], will yield:

$\Delta\rho_1$	$\Delta\rho_2$	$Vol_1/(Vol_1 + Vol_2)$
0.219001	0.588525	0.81885856
0.16180440	0.5335068	0.81885856
0.1046078	0.4784886	0.81885856
-0.0097854	0.3684522	0.81885856
-0.2385718	0.1483794	0.81885856
-0.29576840	0.09336119	0.81885856
-0.352965	0.03834300	0.81885856

This button calculates the contrast from the entered D_2O fraction and the functional form of the contrast. The volume fraction is required to convert between the individual contrasts and the contrast of the entire complex.

If you wish to combine data from two different experiments, then you can run the appropriate module, save the output and manually edit the file. Equally you could enter all the required quantities by hand.

Rg

Once the contrast values have been calculated as described above, using the *Rg* module is straightforward. Again, if output has been saved from a previous run, then this can be reloaded.

f_{D_2O}	R_g	$\sigma(R_g)$	$\Delta\rho_1$	$\Delta\rho_2$	$Vol_1/(Vol_1 + Vol_2)$
	28.85	0.12	2.22	6.10	0.82
	28.33	0.19	1.64	5.53	0.82
	28.22	0.29	1.06	4.96	0.82
	23.40	0.70	-0.10	3.83	0.82
	23.03	0.20	-2.42	1.56	0.82
	24.16	0.14	-3.00	0.99	0.82
	25.11	0.09	-3.58	0.42	0.82

If there is no saved input, then the appropriate R_g values and their corresponding $\sigma(R_g)$ values must be entered. Along with the contrast and volume information, this is all the information required for the calculation. The program performs two different analyses, where the first is

attributed to Stuhrmann[1]. If this analysis is used, in addition to citing MULCh that work should also be cited. The first part section of the output is output that can be taken, and used to create a Stuhrmann plot:

Stuhrmann plot data

$\Delta\rho^{-1}$	R_g^2 (exp)	$\sigma(R_g^2)$	R_g^2 (calc)
0.343	832.3	6.9	815.3
0.427	802.6	10.8	822.1
0.568	796.4	16.4	825.9
1.646	547.6	32.8	543.3
-0.587	530.4	9.2	516.6
-0.438	583.7	6.8	591.9
-0.350	630.5	4.5	631.8

The next section of the output (see below) summarises the quality of the fit, the Stuhrmann coefficients and their uncertainties, as well as the radius of gyration of each of the subunits in the complex (R_1 and R_2), and their separation (D). The equations used to derive R_1, R_2 and D from the Stuhrmann coefficients are the same as those outlined by Olah[2].

Stuhrmann Analysis

Parameter	Value	ESD
χ^2	4.11	-
Stuhrmann plot coefficients		
R_m^2	752.89	8.65
$\alpha^\dagger$	263.30	15.61
β	237.26	26.39
Extracted parameters		
R_1^2	646.70	6.53
R_2^2	643.47	47.66
D^2	723.42	80.47
R_1	25.43	0.13
R_2	25.37	0.94
D	26.90	1.50
R_m	27.44	0.16

†Because the sign of alpha is positive, the component with a higher scattering density lies toward the periphery of the complex.

The parallel axis analysis provides similar information (R_1, R_2 and D still have the same meaning), obtained in a different manner, and an example of the corresponding output is shown below.

[1] K. Ibel, H. B. Stuhrmann, *J. Mol. Biol.* **93**, 255-265 (1975)
[2] G. A. Olah, S. E. Rokop, C. L. A. Wang, S. L. Blechner, J. Trewhella, *Biochemistry* **33**, 8233-8239 (1994).

Parallel Axis Analysis

Parameter	Value	ESD
χ^2	4.12	-
$R_1{}^2$	650.88	6.55
$R_2{}^2$	650.84	46.58
D^2	700.29	78.06
R_1	25.51	0.13
R_2	25.51	0.91
D	26.46	1.47

COMPOST

The use of *COMPOST* is slightly more complicated than that for *Rg*. In addition to contrast and volume fraction data, *COMPOST* requires contrast variation data in three column format. There are also some additional parameters and options:

▪ The scattering data at each contrast point must use the same *q*-scale. As the quality of the low-*q* data in each file may not be the same, a different *Start point* may be defined for each file. This allows poor quality data points to be disregarded from the fitting process, while keeping the same *q*-scale.

▪ It is also possible to specify a different *Scale factor* for each contrast point which is used to account for differences in sample concentration or collection time. It is also possible to refine this scale factor.

A typical input will resemble the following:

Number contrast points: 7

Upper qR_g **limit:** 1.3

f_{D2O}	Refine Scale	Scale factor	Start Point	Contrast variation data		$\Delta\rho_1$	$\Delta\rho_2$	$Vol_1/(Vol_1+Vol_2)$
	☑ Refine	1.00	1	D:\Projects\MULCh\NC	Browse…	2.22	6.10	0.82
	☑ Refine	1.00	1	D:\Projects\MULCh\NC	Browse…	1.64	5.53	0.82
	☑ Refine	1.00	1	D:\Projects\MULCh\NC	Browse…	1.06	4.96	0.82
	☑ Refine	1.00	1	D:\Projects\MULCh\NC	Browse…	-0.10	3.83	0.82
	☑ Refine	1.00	1	D:\Projects\MULCh\NC	Browse…	-2.42	1.56	0.82
	☑ Refine	1.00	1	D:\Projects\MULCh\NC	Browse…	-3.00	0.99	0.82
	☑ Refine	1.00	1	D:\Projects\MULCh\NC	Browse…	-3.58	0.42	0.82

Above, 7 files have been input, contrast values have been calculated, the start point for each file is the first point and the scale factor is to be refined. In reality, the scale factor for the first file is not refined, and all corrections are made relative to this file. The scale factor is determined using a Guinier fit to the low angle data, where the number of data points used in this fitting is governed by the *Upper* qR_g *limit*. This number is set to 1.3 by default, however if the particle is quite asymmetric, then this number should be reduced.

On completion of the calculation, the top of the output shows links to download not only input and output files, but also the composite scattering functions.

Following this is a summary of the Guinier analysis and scaling corrections.

Guinier analysis and scale factor calculation

	R_g		$I(0)$		qR_g		χ^2	$\Delta\rho^2$ $(10^{20}\,cm^{-4})$	Scale Factor
	Value	ESD	Value	ESD	Min	Max			
1	27.51	0.26	0.531	0.003	0.28	1.29	1.22	8.517	1.00
2	27.51	0.25	0.357	0.002	0.28	1.29	0.66	5.477	0.96
3	28.31	0.42	0.216	0.002	0.29	1.27	0.62	3.105	0.90
4	21.90	0.70	0.068	0.001	0.22	1.26	0.80	0.369	0.34
5	22.09	0.20	0.184	0.001	0.23	1.30	0.66	2.902	0.98
6	23.88	0.17	0.319	0.001	0.25	1.28	1.08	5.207	1.02
7	24.64	0.07	0.529	0.001	0.25	1.29	0.51	8.180	0.96

And finally, the statistics of the *COMPOST* fitting procedure.

Deviations between the composite scattering functions and the contrast variation series

	$[\Delta I(q)/\sigma(I(q))]^2$							χ^2
	1	2	3	4	5	6	7	
1	0.38	0.05	0.52	0.03	0.19	0.31	0.01	0.37
2	0.00	0.03	0.00	0.00	0.04	0.00	0.01	0.02
3	0.13	0.06	0.52	0.00	0.04	0.13	0.13	0.25
4	0.07	0.42	0.27	0.00	0.00	0.00	0.00	0.19
5	0.01	0.72	1.37	0.03	0.26	0.96	0.21	0.89
6	0.33	0.55	0.03	0.05	0.88	0.67	0.03	0.64

Each of the values in the matrix represents the error-weighted deviation squared of that point from the model, and the last column represents the error-weighted mean-square deviation of the model from the contrast variation data at that q-value. All points where the deviation of the model from a specific point is greater than 9.0 is marked with an asterisk. This matrix allows identification of data that that are not modelled well by the *COMPOST* fitting procedure.

Running the programs from the command line

The source code for *Contrast, Rg and COMPOST* (as well as sample input and scattering data) can be downloaded from the *MULCh* homepage (`MULCh.zip`). To compile the programs you will need the `make` utility and an appropriate C compiler (you may have to

modify `Makefile` if you are not using a common C compiler). Once you have these, at the command line, enter the following:

```
make Contrast
make Rg
make Compost
```

This will compile the executable for each program. You will need to create your own input files using the sample files provided as a guide. Input files saved from the web-interface are also valid, but for the *COMPOST* input file will need to be edited so that it includes the scattering data filenames. The programs can then be run by typing:

```
./Contrast.exe < contrast_input.txt > contrast_output.html
./Rg.exe < rg_input.txt > rg_output.html
./Compost.exe < compost_input.txt > compost_output.html
```

To prevent overwriting of the composite scattering functions created by *COMPOST*, it is possible to run `Compost.exe` in a different way, by typing:

```
./Compost.exe dir tag < compost_input.txt > compost_output.html
```

tag is a tag included in the filename, i.e. `i11_tag.dat`, `i22_tag.dat` and `i12_tag.dat` - *tag* can be composed of alphanumeric characters.

dir is the directory into which `i11_tag.dat`, `i22_tag.dat` and `i12_tag.dat` are to be written - `./` is used to write to the current directory.

You may take any of this code and modify it, but any publications arising from these changes must clearly outline what the changes are. Similarly, if a modified version of any of these programs is shared, it must be made clear to the user what changes have been made. If a bug is identified please let us know, so the version downloadable from the website can be fixed.

Table 1 *Stored element parameters*

	Volume (Å^3)		Volume (Å^3)
H (or ^{2}H)[a]	5.15	K[b]	11.01 (1.38 Å)
C[a]	16.44	Ca[b]	4.19 (1.00 Å)
N[a]	2.49	Fe[c]	7.99 (1.24 Å)
O[a]	9.13	Co[c]	7.99 (1.24 Å)
Na[b]	4.45 (1.02 Å)	Ni[c]	8.18 (1.25 Å)
Mg[b]	1.56 (0.72 Å)	Cu[c]	8.78 (1.28 Å)
P[c]	3.37 (0.93 Å)	Zn[c]	9.85 (1.33 Å)
S[b]	26.09 (1.84 Å)	Br[b]	31.54 (1.96 Å)
Cl[b]	24.84 (1.81 Å)	I[b]	44.60 (2.20 Å)

[a]Fraser, R. D. B.,MacRae, T. P. & Suzuki, E. (1978). *J. Appl. Crystallogr.* **11**, 693-694.

[b]Volumes calculated from Shannon radii (in parenthesises). Source: http://abulafia.mt.ic.ac.uk/shannon/ptable.php

[c]Volumes calculated from atomic radii (in parentheses). Source: http://periodic.lanl.gov/default.htm

Table 2 *Stored amino acid parameters*

	Formula	Volume[a] ($Å^3$)	N_{exH}
Alanine	C_3H_5ON	89.27	1
Cysteine	C_3H_5ONS	112.84	2
Aspartic acid	$C_4H_4O_3N$	114.43	1
Glutamic acid	$C_5H_6O_3N$	138.81	1
Phenylalanine	C_9H_9ON	190.84	1
Glycine	C_2H_3ON	63.76	1
Histidine[b]	$C_6H_7ON_3$	157.46	2
Isoleucine	$C_6H_{11}ON$	163.01	1
Lysine	$C_6H_{13}ON_2$	165.08	4
Leucine	$C_6H_{11}ON$	163.09	1
Methionine	C_5H_9ONS	165.82	1
Asparagine	$C_4H_6O_2N_2$	122.35	3
Proline	C_5H_7ON	121.29	0
Glutamine	$C_5H_8O_2N_2$	146.91	3
Arginine	$C_6H_{13}ON_4$	190.33	6
Serine	$C_3H_5O_2N$	93.50	2
Threonine	$C_4H_7O_2N$	119.61	2
Valine	C_5H_9ON	138.16	1
Tryptophan	$C_{11}H_{10}ON_2$	226.38	2
Tyrosine	$C_9H_9O_2N$	194.63	2

[a]Tsai, J., Taylor, R., Chothia, C. & Gerstein, M. (1999). J. Mol. Biol. **290**, 253-266

[b]In v1.0 of Contrast, the imidazole of histidine was considered to be bound to 1.5 exchangeable hydrogen atoms. In v1.1 it was erroneously changed so that 0.0 exchangeable hydrogen atoms are bound to the imidazole ring, but this was fixed in v1.1.1 to a value of 1.0.

Table 3 *Stored nucleic acid parameters*

	Formula	Volume[a] (Å^3)	N_{exH}
DNA			
Adenine	$C_5H_4N_5$	133.9	2
Cytosine	$C_4H_4ON_3$	115.6	2
Guanine	$C_5H_4ON_5$	146.6	3
Thymine	$C_4H_3O_2N_2$	133.6	1
Backbone	$C_5H_7O_5P$	168.1	0
RNA			
Adenine	$C_5H_4N_5$	139.2	2
Cytosine	$C_4H_4ON_3$	115.0	2
Guanine	$C_5H_4ON_5$	145.9	3
Uracil	$C_4H_3O_2N_2$	110.8	1
Backbone	$C_5H_7O_6P$	176.1	1

[a]Nadassy, K., Tomas-Oliveira, I., Alberts, I., Janin, J. & Wodak, S. J. (2001). *Nucleic Acids Res.* **29**, 3362-3376.;
Voss, N. R. & Gerstein, M. (2005). *J. Mol. Biol.* **346**, 477-492.

A few notes about the contrast calculations

1) For protein or nucleic acid sequences, the calculation adds the mass, volume and scattering length of H_2O (i.e. H to one end of the molecule and OH to the other end to "cap" the molecule). This is the reason why you may notice that the volume of the protein is ~ 20 $Å^3$ larger than might be expected.

2) The program does not recognise the ambiguous codes B, J, X or Z or the unassigned O and U for protein sequences. These are skipped, and noted in the output. Be aware that as of V1.1 of MULCh, X is recognised for molecular formulae and is interpreted as an exchangeable hydrogen atom.

3) The program reads both upper and lower case characters for the any of the sequences <u>but not for molecular formulae</u>. There can also be spaces etc in the sequence which the web-interface should remove, but if you are experiencing problems you should remove the extraneous characters.

4) It is assumed that neither the 3' nor 5' ends of an RNA or DNA sequence are phosphorylated.

5) The average scattering lengths and atomic masses for all elements $Z < 55$ are tabulated, but the volumes for only a fraction of these are tabulated (see Table 1). If other elements (or specific isotopes) are required you can use (for example) an Yttrium atom ($Z_Y = 39$) and a Zirconium ($Z_{Zr} = 40$) to represent a Gold atom ($Z_{Au} = 79$) for X-ray contrast calculations, where the volume of the YZr unit should be specified to that of a single gold atom. For neutron contrast calculations you may specify hydrogen or deuterium (opposite sign), and use $N_{molecules}$ to scale the scattering length. For example, the scattering length of Uranium is $b_U = 8.417 \times 10^{-13}$ cm, which can be mimicked by a deuterium atom ($b_D = 6.671 \times 10^{-13}$ cm) with $N_{molecules} = 1.26$ (the volume would then be set to $V_U / 1.26$).

6) Note that specifying a calcium atom (for example) will result in a minute discrepancy in the X-ray scattering length density calculation (as calcium will be present in solution as Ca^{2+}). For practical purposes, this discrepancy is unlikely to cause any issues as the number of electrons in your biological molecule will be orders of magnitude larger than 2 electrons missing from calcium.

7) Deuteration level is only considered for protein, RNA or DNA sequences. It is not considered for molecules. If required, deuterium can be specified in molecular formulas (this is new as of V1.1 of MULCh – in the previous version the contrast was not calculated correctly when deuterium was specified).

8) Be careful when comparing masses calculated using MULCh and ProtParam. MULCh assumes that aspartic acid and glutamic acid are deprotonated, while lysine and arginine are protonated. Thus, the difference in the mass calculated from MULCh and ProtParam will be $(N_{ASP} + N_{GLU}) - (N_{LYS} + N_{ARG})$. Further, the masses printed in the "*Solvent and particle information in H_2O*" table include non-exchangeable deuterium atoms, which will make the protein heavier than expected.