

## Supporting Information

### Improved Quantum Chemical NMR Chemical Shift Prediction of Metabolites in Aqueous Solution Toward the Validation of Unknowns

Felix Hoffmann,<sup>1</sup> Da-Wei Li,<sup>2</sup> Daniel Sebastiani,<sup>1</sup> and Rafael Brüschweiler<sup>2,3,4\*</sup>

<sup>1</sup>Institute of Chemistry, Martin-Luther-University Halle-Wittenberg, von-Danckelmann-Platz 4, 06120 Halle, Germany

<sup>2</sup>Campus Chemical Instrument Center, The Ohio State University, Columbus, Ohio 43210, United States

<sup>3</sup>Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States

<sup>4</sup>Department of Biological Chemistry and Pharmacology, The Ohio State University, Columbus, Ohio 43210, United States

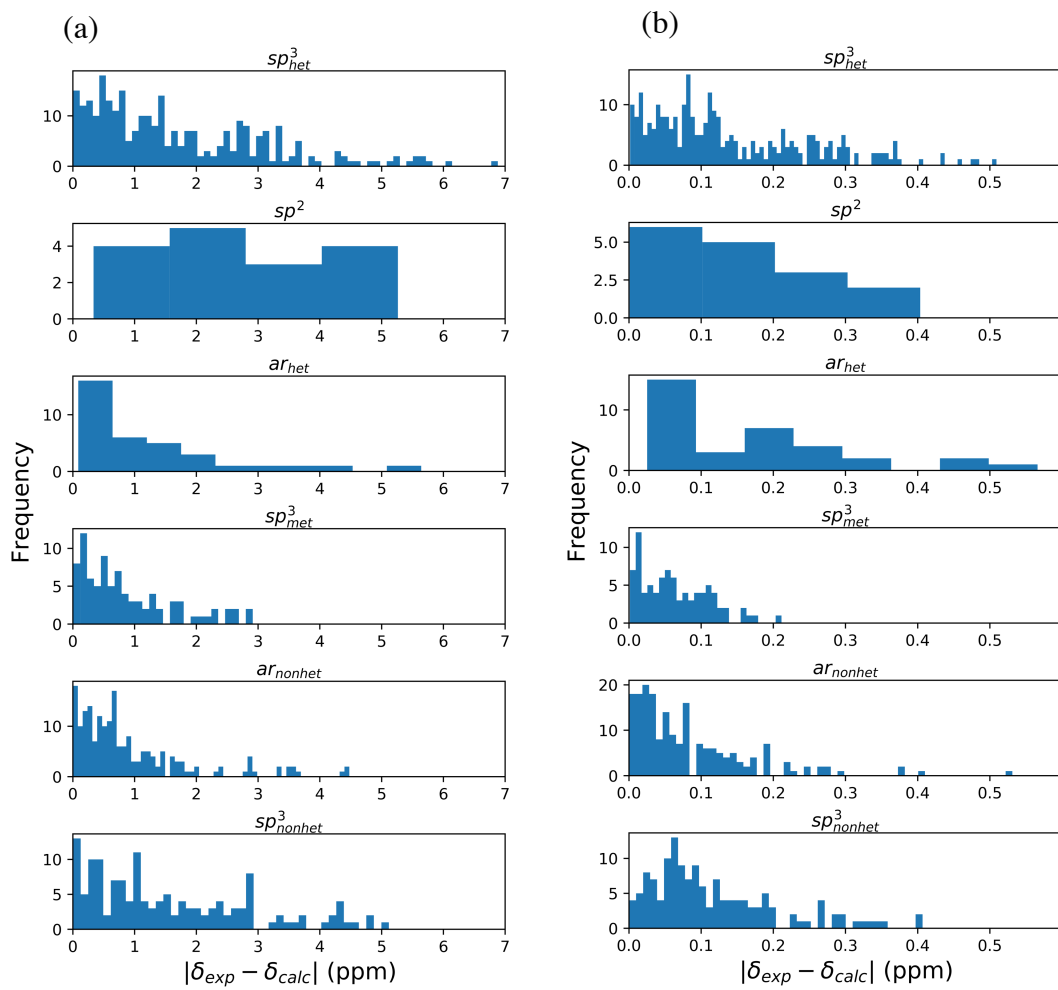
\*To whom correspondence should be addressed:

Rafael Brüschweiler, Ph.D.

Department of Chemistry and Biochemistry

The Ohio State University, Columbus, Ohio 43210

E-mail: [bruschweiler.1@osu.edu](mailto:bruschweiler.1@osu.edu)



**Figure S1.** Absolute deviations of MOSS-DFT predicted and experimental  $^{13}\text{C}$  (a) and  $^1\text{H}$  (b) chemical shifts for the various atom types.

**Table S1.**  $^{13}\text{C}$  Linear Regression Parameters (a,b) Obtained with Different DFT Functionals and Basis Sets.

| motif                             | B97-2/pcS-1 |          | B97-2/pcS-2 |          | B97-2/pcS-3 |          | B3LYP/pcS-2 |          | BLYP/pcS-2 |          |
|-----------------------------------|-------------|----------|-------------|----------|-------------|----------|-------------|----------|------------|----------|
|                                   | a           | b        | a           | b        | a           | b        | a           | b        | a          | b        |
| $\text{sp}^3_{\text{het}}$        | -0.9699     | 182.3627 | -0.9252     | 173.2275 | -0.9261     | 172.1137 | -0.8789     | 161.6165 | -0.9184    | 162.7311 |
| $\text{sp}^2$                     | -1.0200     | 186.1140 | -0.9940     | 179.5118 | -0.9936     | 177.9458 | -0.9704     | 170.1978 | -1.0120    | 170.4744 |
| $\text{ar}_{\text{het}}$          | -0.9051     | 181.0846 | -0.8699     | 174.9883 | -0.8634     | 173.6563 | -0.8478     | 166.9615 | -0.8429    | 166.4417 |
| $\text{sp}^3_{\text{nonhet,met}}$ | -0.9819     | 183.9691 | -0.9569     | 178.3115 | -0.9532     | 176.5288 | -0.9021     | 164.6078 | -0.8783    | 156.9817 |
| $\text{ar}_{\text{nonhet}}$       | -0.8504     | 178.3376 | -0.8427     | 173.7705 | -0.8433     | 172.4185 | -0.8157     | 165.5075 | -0.8126    | 164.2485 |
| $\text{sp}^3_{\text{nonhet}}$     | -0.9423     | 177.4040 | -0.9277     | 172.3199 | -0.9277     | 171.0732 | -0.8829     | 159.9842 | -0.8766    | 154.2725 |

**Table S2.**  $^1\text{H}$  Linear Regression Parameters (a,b) Obtained with Different DFT Functionals and Basis Sets.

| motif                             | B97-2/pcS-1 |         | B97-2/pcS-2 |         | B97-2/pcS-3 |         | B3LYP/pcS-2 |         | BLYP/pcS-2 |         |
|-----------------------------------|-------------|---------|-------------|---------|-------------|---------|-------------|---------|------------|---------|
|                                   | a           | b       | a           | b       | a           | b       | a           | b       | a          | b       |
| $\text{sp}^3_{\text{het}}$        | -0.8792     | 28.3608 | -0.9267     | 29.5819 | -0.9181     | 29.2769 | -0.9169     | 29.3082 | -0.9050    | 28.9257 |
| $\text{sp}^2$                     | -1.0230     | 32.3395 | -0.9840     | 30.9787 | -0.9677     | 30.4447 | -0.9871     | 31.0831 | -1.0236    | 32.0215 |
| $\text{ar}_{\text{het}}$          | -0.7870     | 26.6120 | -0.7761     | 26.2103 | -0.7725     | 26.0812 | -0.7777     | 26.2842 | -0.7501    | 25.7034 |
| $\text{sp}^3_{\text{nonhet,met}}$ | -0.9235     | 29.3315 | -0.9298     | 29.5025 | -0.9244     | 29.2860 | -0.9354     | 29.6693 | -0.9421    | 29.8497 |
| $\text{ar}_{\text{nonhet}}$       | -0.8644     | 28.2049 | -0.8173     | 26.8939 | -0.7854     | 26.0863 | -0.8201     | 27.0073 | -0.7972    | 26.5386 |
| $\text{sp}^3_{\text{nonhet}}$     | -0.9719     | 31.0221 | -0.9592     | 30.4554 | -0.9399     | 29.8153 | -0.9670     | 30.6914 | -0.9548    | 30.2499 |

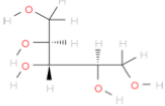
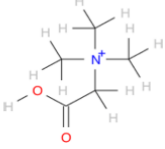
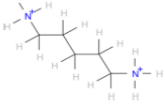
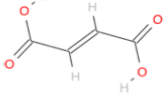
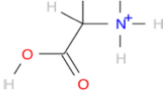
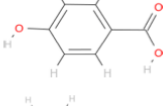
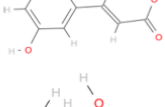
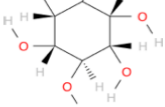
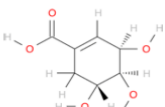
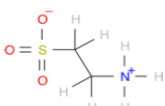
**Table S3** Linear regression chemical shift parameters for the B97-2/pcS-3 method that were calculated from a training set obtained by randomly excluding 20% of the molecules in the original MOSS-DFT set (see Table S5 for a complete list of the test set). Chemical shifts were calculated in a second step by applying the linear regression parameters obtained from the training data to the molecules in the test set. The RMSD values reported below are those for the test set (i.e. molecules that were not used for fitting). Since the test set consists of only 20% of the molecules of the original training set, the RMSD values show somewhat larger variations, but their average is very close to the original average RMSDs for both  $^{13}\text{C}$  and  $^1\text{H}$  chemical shifts.

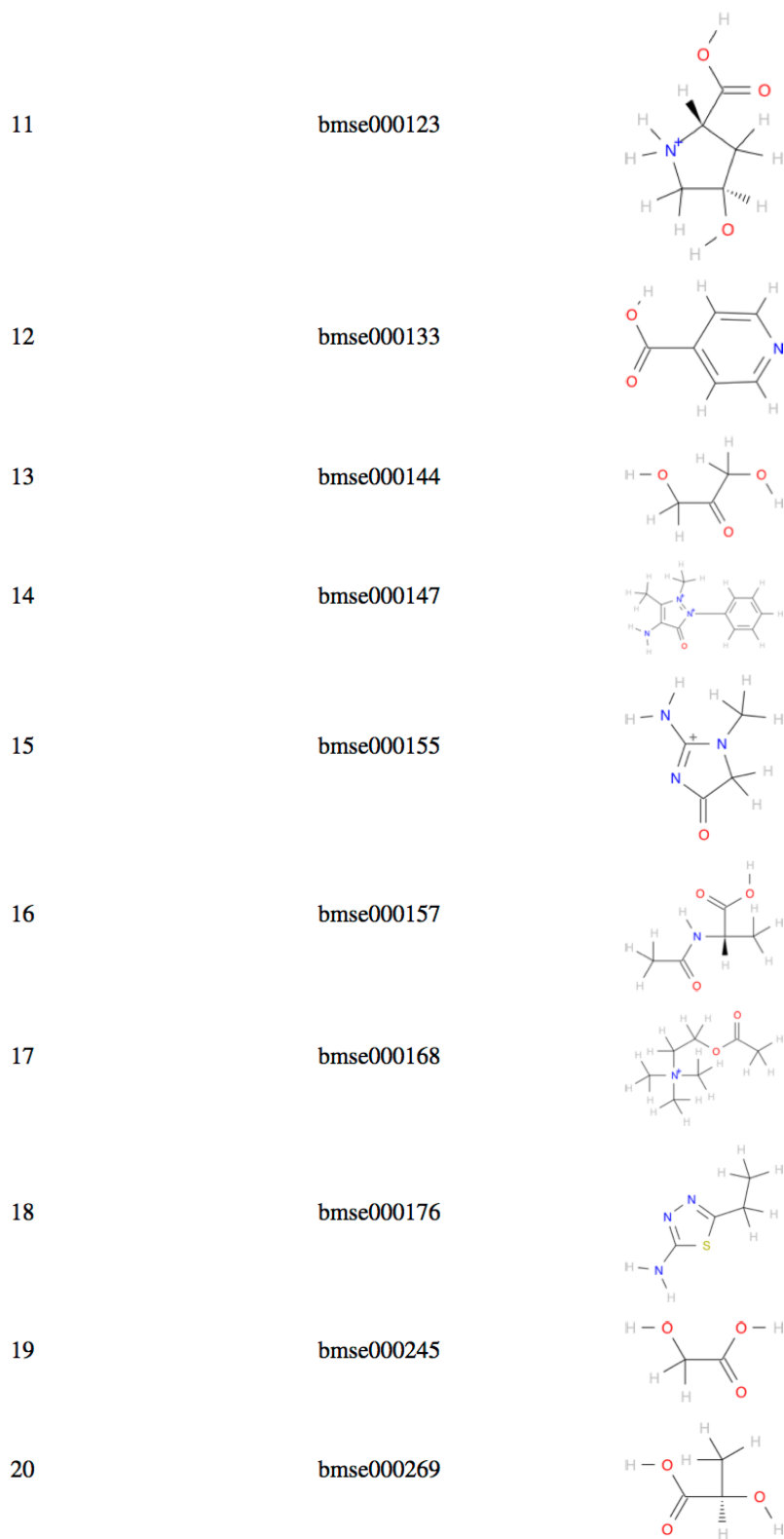
| Nucleus         |   | $\text{sp}^3_{\text{nonhet,met}}$ | $\text{ar}_{\text{nonhet}}$ | $\text{ar}_{\text{het}}$ | $\text{sp}^3_{\text{nonhet}}$ | $\text{sp}^3_{\text{het}}$ | $\text{sp}^2$ | total |
|-----------------|---|-----------------------------------|-----------------------------|--------------------------|-------------------------------|----------------------------|---------------|-------|
| $^{13}\text{C}$ | a   | -0.9456                           | -0.8382                     | -0.8466                  | -0.9300                       | -0.9255                    | -0.9916       |       |
|                 | b   | 175.2692                          | 172.0385                    | 173.4065                 | 171.4041                      | 171.9112                   | 177.7612      |       |
|                 | RMSD  | 1.31                              | 1.18                        | 2.32                     | 2.42                          | 2.17                       | 4.87          | 1.94  |
| $^1\text{H}$    | a   | -0.9168                           | -0.7818                     | -0.7742                  | -0.9412                       | -0.9146                    | -0.9706       |       |
|                 | b   | 29.0564                           | 26.0009                     | 26.1394                  | 29.8493                       | 29.1770                    | 30.5297       |       |
|                 | RMSD  | 0.061                             | 0.086                       | 0.197                    | 0.149                         | 0.163                      | 0.141         | 0.133 |
|                 | No. of<br>$^{13}\text{C}$ - $^1\text{H}$<br>pairs | 14                                | 51                          | 8                        | 22                            | 52                         | 2             | 149   |

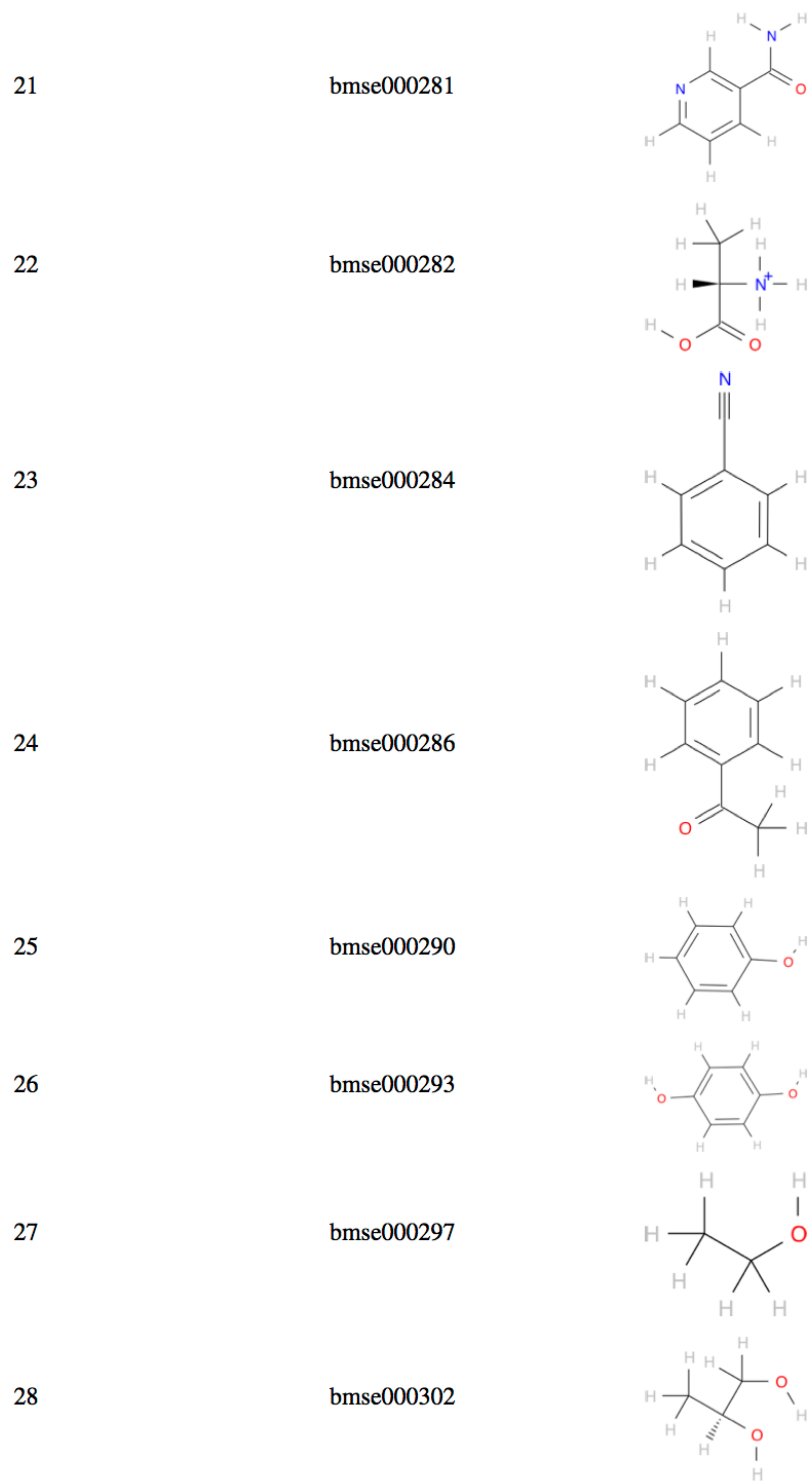
**Table S4** BMRB and HMDB IDs of Metabolite Molecules that are Part of the Test Set for Cross-Validation (see Table S3 for results)

| #  | ID         |
|----|------------|
| 1  | bmse000062 |
| 2  | bmse000089 |
| 3  | bmse000092 |
| 4  | bmse000093 |
| 5  | bmse000147 |
| 6  | bmse000157 |
| 7  | bmse000281 |
| 8  | bmse000345 |
| 9  | bmse000350 |
| 10 | bmse000389 |
| 11 | bmse000433 |
| 12 | bmse000437 |
| 13 | bmse000573 |
| 14 | bmse000619 |
| 15 | bmse000703 |
| 16 | bmse000738 |
| 17 | bmse000825 |
| 18 | HMDB00034  |
| 19 | HMDB00055  |
| 20 | HMDB00073  |
| 21 | HMDB00310  |
| 22 | HMDB00479  |
| 23 | HMDB00678  |
| 24 | HMDB00700  |
| 25 | HMDB00736  |
| 26 | HMDB00748  |
| 27 | HMDB00764  |
| 28 | HMDB00783  |
| 29 | HMDB01266  |
| 30 | HMDB01522  |
| 31 | HMDB01857  |
| 32 | HMDB01866  |
| 33 | HMDB02039  |
| 34 | HMDB02339  |
| 35 | HMDB03646  |
| 36 | HMDB04816  |

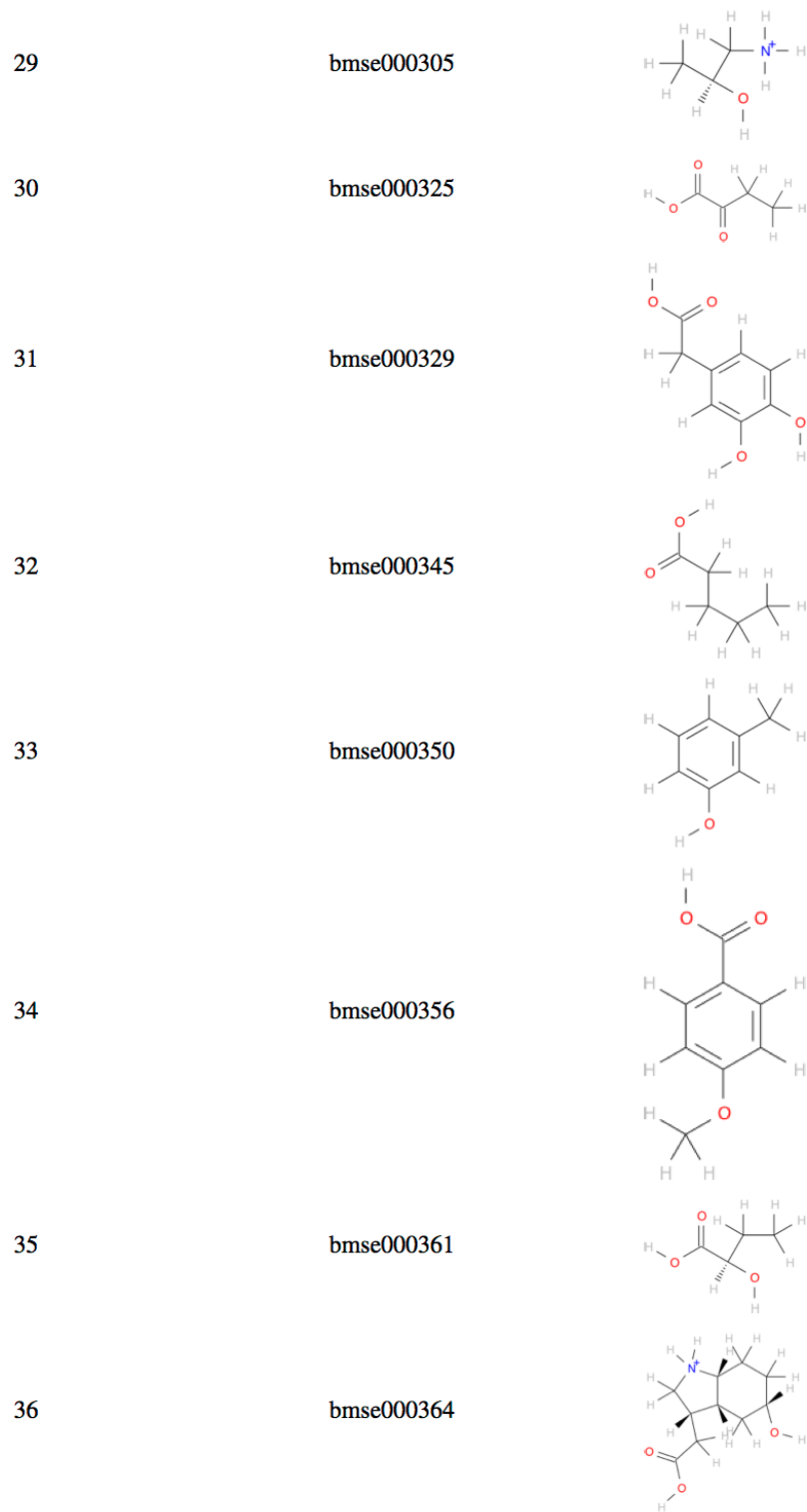
**Table S5.** BMRB and HMDB IDs and Structures of Predicted Metabolite Molecules

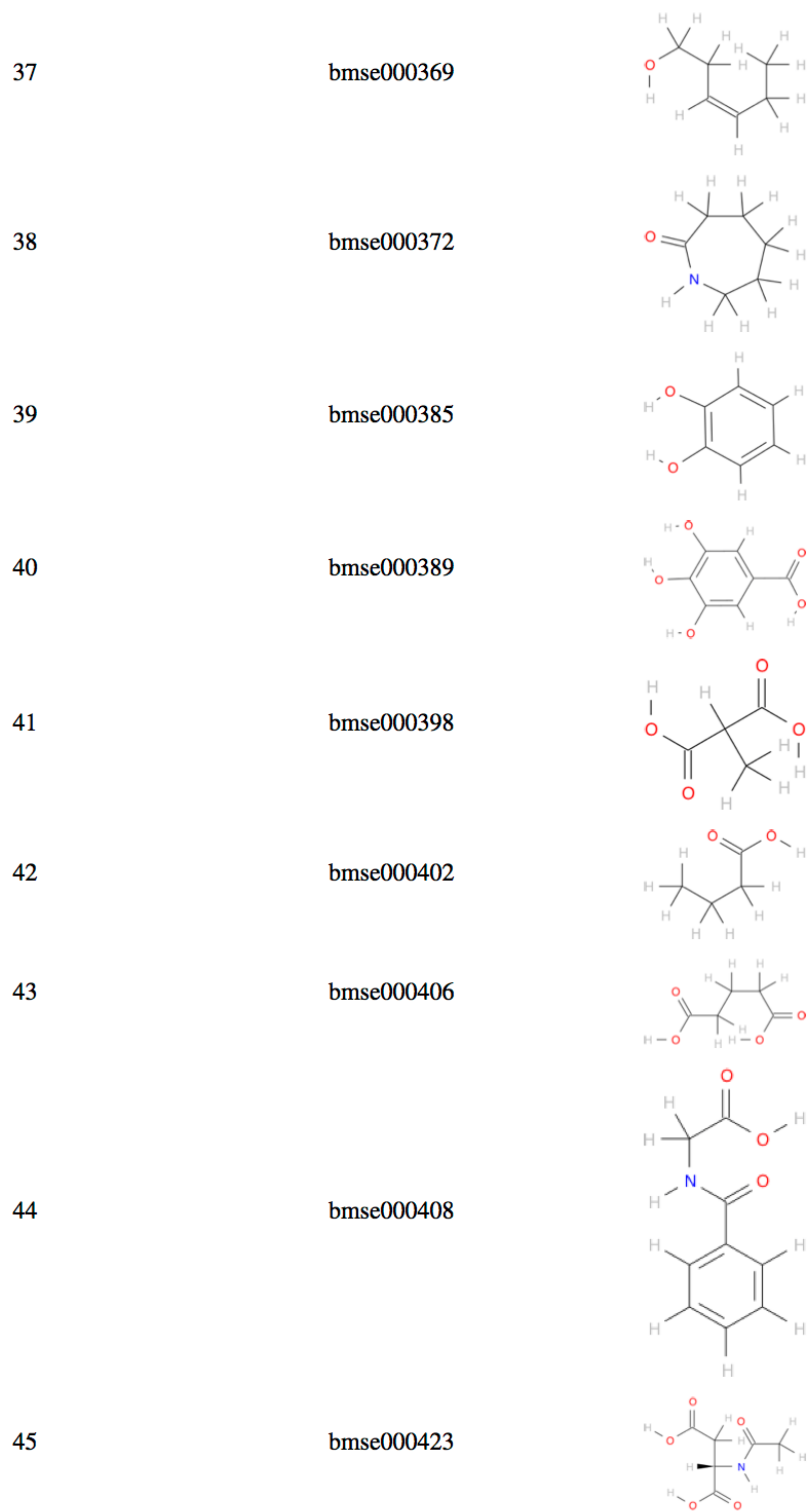
| #  | ID         | Structure  |
|----|------------|--|
| 1  | bmse000062 |    |
| 2  | bmse000069 |    |
| 3  | bmse000072 |    |
| 4  | bmse000083 |    |
| 5  | bmse000089 |   |
| 6  | bmse000092 |  |
| 7  | bmse000093 |  |
| 8  | bmse000102 |  |
| 9  | bmse000114 |  |
| 10 | bmse000120 |  |

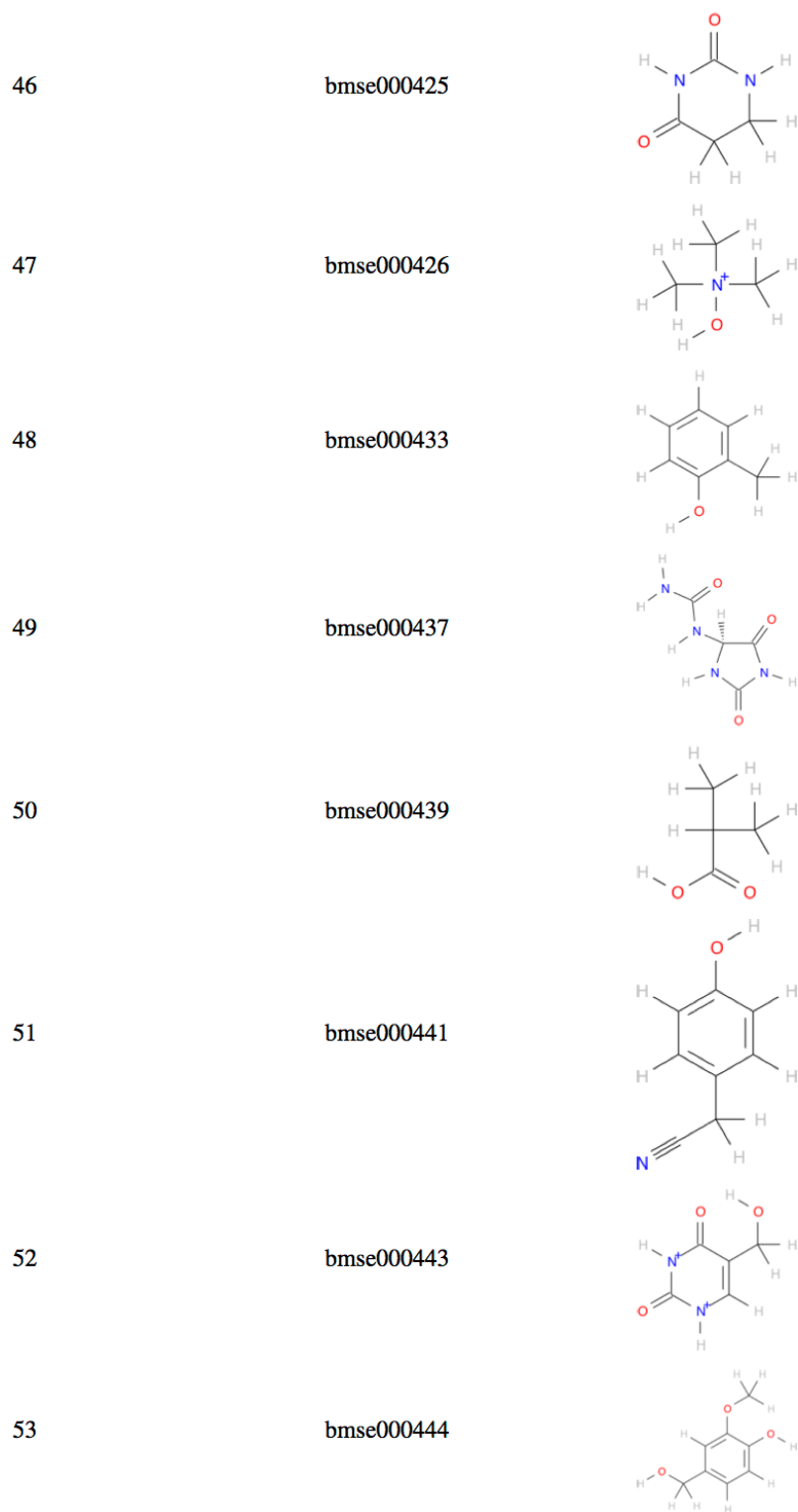


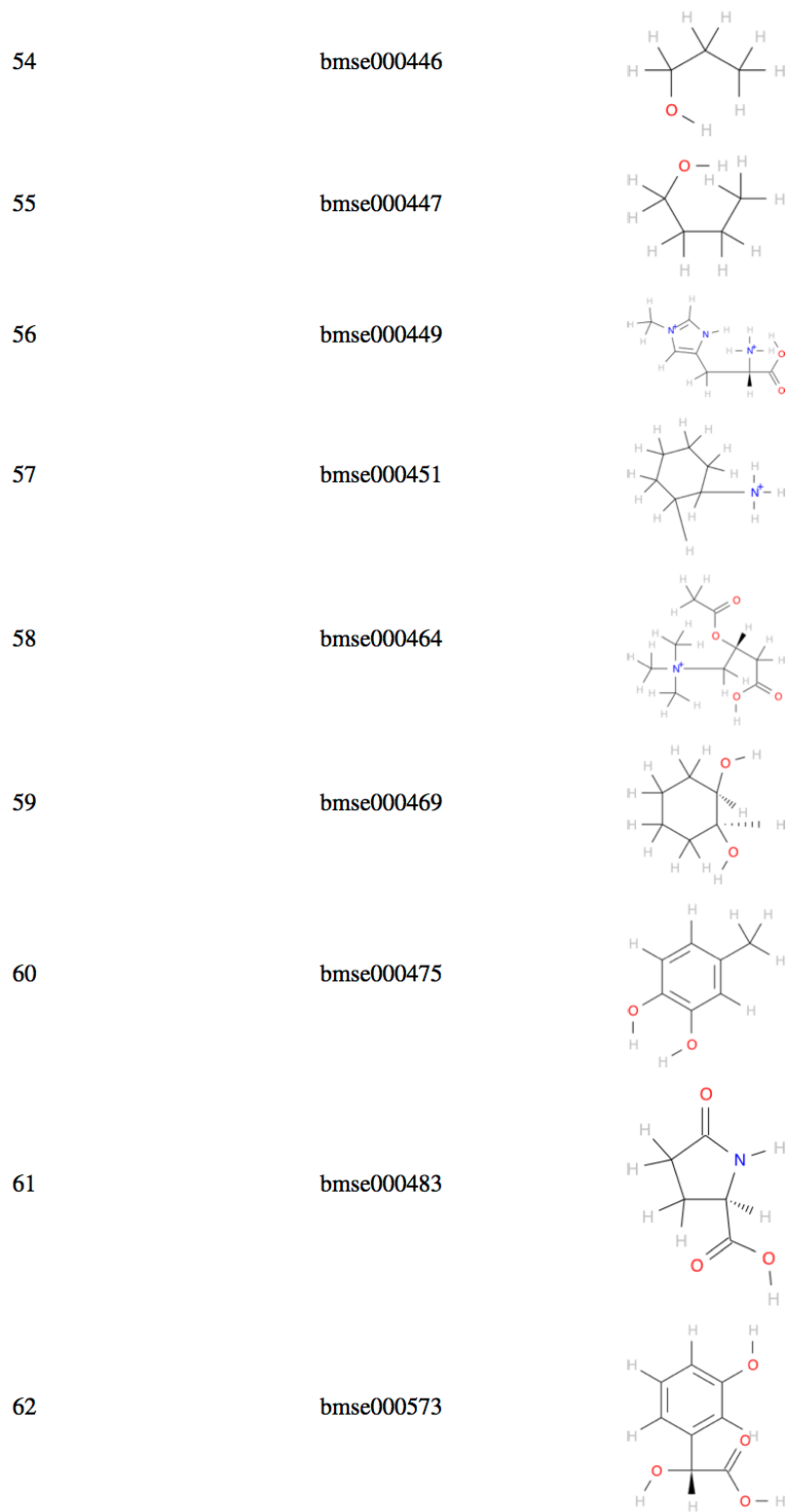


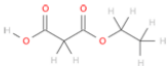
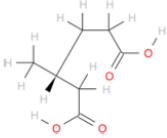
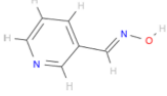
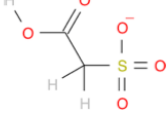
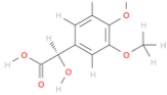
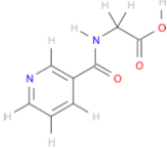
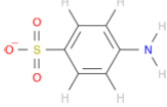
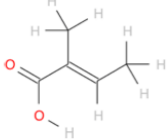
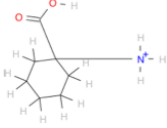
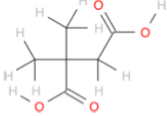
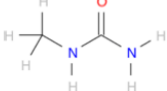


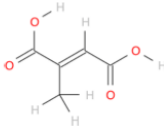
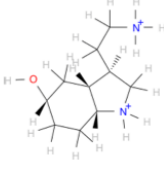
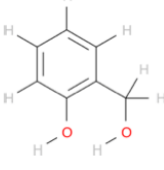
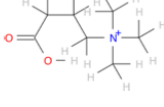
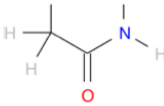
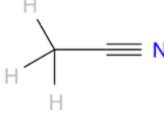
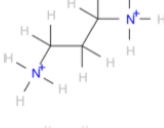
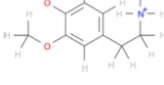
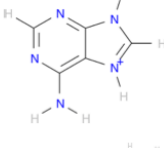
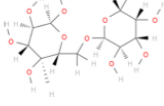


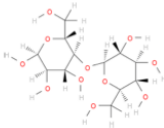
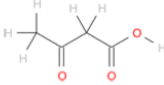
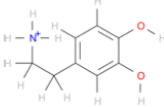
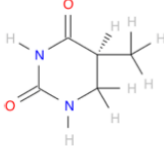
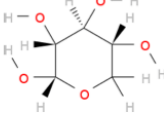
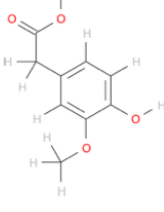
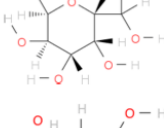
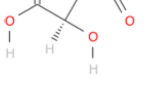
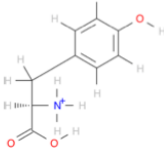
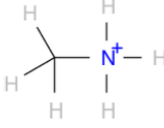


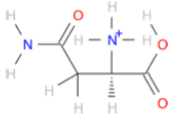
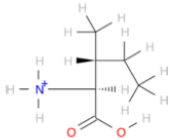
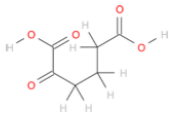
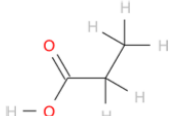
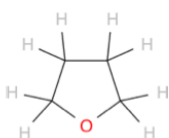
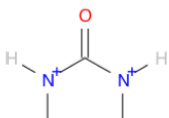
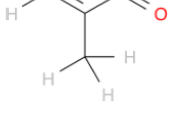
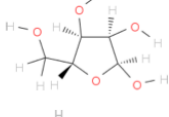
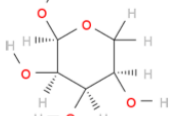
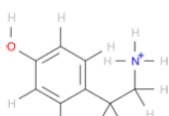




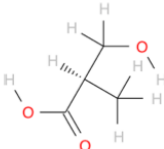
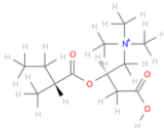
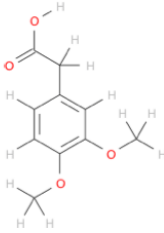
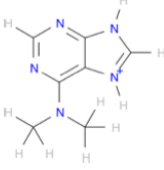
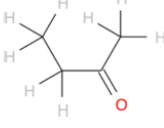
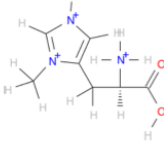
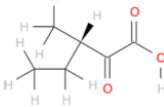
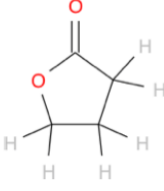
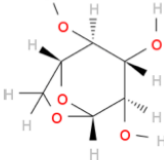
|    |            |  |
|----|------------|--|
| 63 | bmse000579 |    |
| 64 | bmse000619 |    |
| 65 | bmse000621 |    |
| 66 | bmse000632 |    |
| 67 | bmse000703 |    |
| 68 | bmse000708 |   |
| 69 | bmse000726 |  |
| 70 | bmse000727 |  |
| 71 | bmse000728 |  |
| 72 | bmse000730 |  |
| 73 | bmse000738 |  |

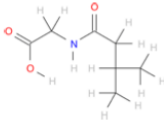
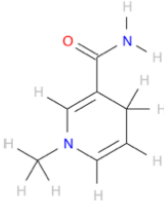
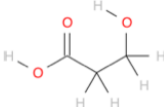
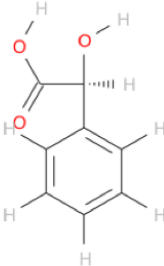
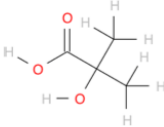
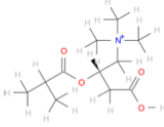
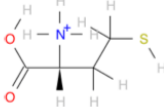
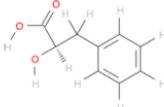
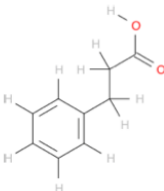
|    |            |  |
|----|------------|--|
| 74 | bmse000746 |    |
| 75 | bmse000757 |    |
| 76 | bmse000768 |    |
| 77 | bmse000794 |    |
| 78 | bmse000825 |   |
| 79 | bmse000826 |  |
| 80 | HMDB00002  |  |
| 81 | HMDB00022  |  |
| 82 | HMDB00034  |  |
| 83 | HMDB00048  |  |

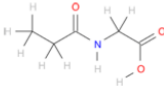
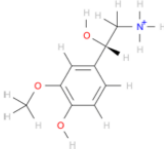
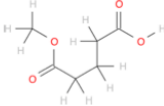
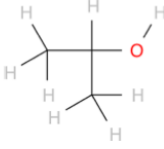
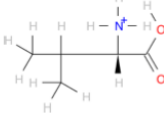
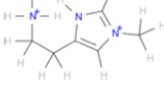
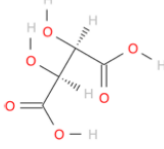
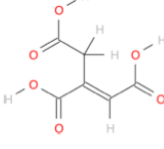
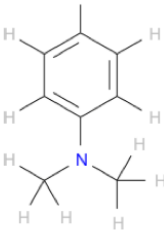
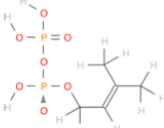
|    |           |  |
|----|-----------|--|
| 84 | HMDB00055 |    |
| 85 | HMDB00060 |    |
| 86 | HMDB00073 |    |
| 87 | HMDB00079 |    |
| 88 | HMDB00098 |    |
| 89 | HMDB00118 |   |
| 90 | HMDB00143 |  |
| 91 | HMDB00156 |  |
| 92 | HMDB00158 |  |
| 93 | HMDB00164 |  |

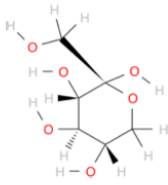
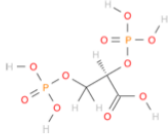
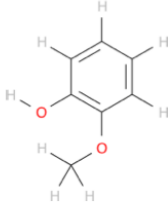
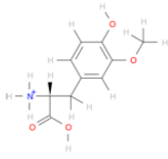
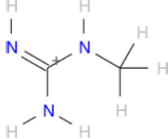
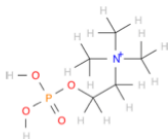
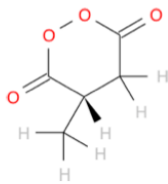
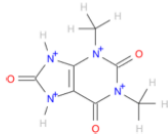
|     |                      |  |
|-----|----------------------|--|
| 94  | HMDB00168            |    |
| 95  | HMDB00172            |    |
| 96  | HMDB00225            |    |
| 97  | HMDB00237            |    |
| 98  | HMDB00246            |    |
| 99  | HMDB00262            |   |
| 100 | HMDB00283 (furanose) |  |
| 101 | HMDB00283 (pyranose) |  |
| 102 | HMDB00306            |  |
| 103 | HMDB00310            |  |

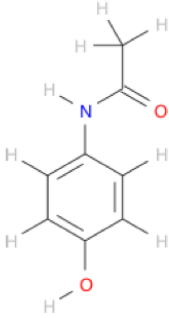
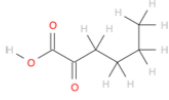
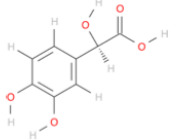
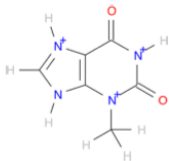
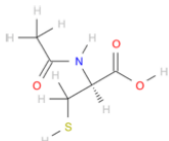
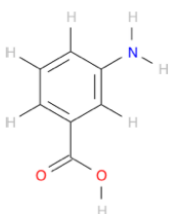
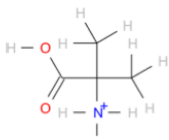
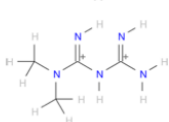


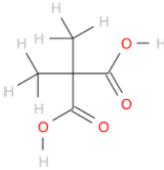
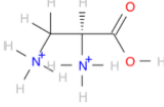
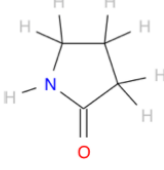
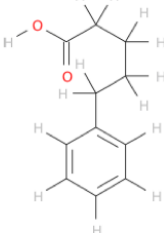
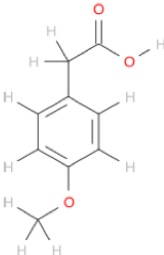
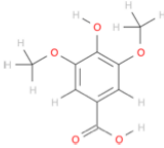
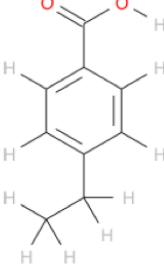
|     |           |  |
|-----|-----------|--|
| 104 | HMDB00336 |    |
| 105 | HMDB00378 |    |
| 106 | HMDB00434 |    |
| 107 | HMDB00473 |   |
| 108 | HMDB00474 |  |
| 109 | HMDB00479 |  |
| 110 | HMDB00491 |  |
| 111 | HMDB00549 |  |
| 112 | HMDB00640 |  |

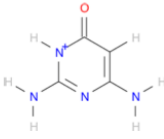
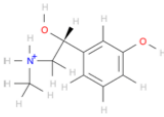
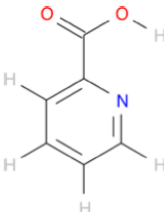
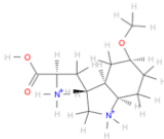
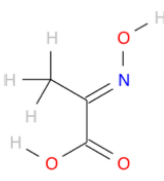
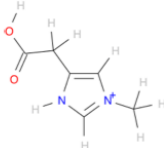
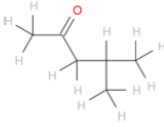
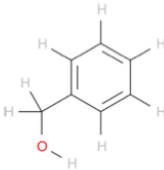
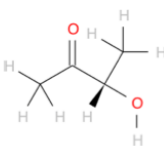
|     |           |  |
|-----|-----------|--|
| 113 | HMDB00678 |    |
| 114 | HMDB00699 |    |
| 115 | HMDB00700 |    |
| 116 | HMDB00703 |   |
| 117 | HMDB00729 |  |
| 118 | HMDB00736 |  |
| 119 | HMDB00742 |  |
| 120 | HMDB00748 |  |
| 121 | HMDB00764 |  |

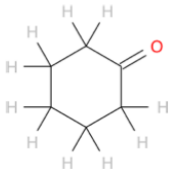
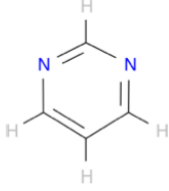
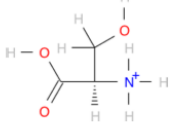
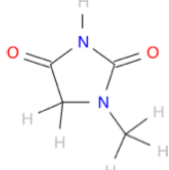
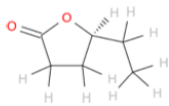
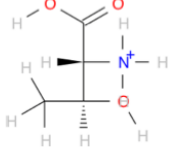
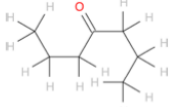
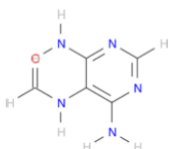
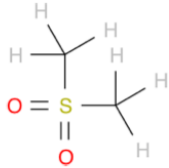
|     |           |  |
|-----|-----------|--|
| 122 | HMDB00783 |    |
| 123 | HMDB00819 |    |
| 124 | HMDB00858 |    |
| 125 | HMDB00863 |    |
| 126 | HMDB00883 |    |
| 127 | HMDB00898 |   |
| 128 | HMDB00956 |  |
| 129 | HMDB00958 |  |
| 130 | HMDB01020 |  |
| 131 | HMDB01120 |  |

|     |           |  |
|-----|-----------|--|
| 132 | HMDB01266 |    |
| 133 | HMDB01294 |    |
| 134 | HMDB01398 |    |
| 135 | HMDB01434 |   |
| 136 | HMDB01522 |  |
| 137 | HMDB01565 |  |
| 138 | HMDB01844 |  |
| 139 | HMDB01857 |  |

|     |           |  |
|-----|-----------|--|
| 140 | HMDB01859 |    |
| 141 | HMDB01864 |    |
| 142 | HMDB01866 |    |
| 143 | HMDB01886 |   |
| 144 | HMDB01890 |  |
| 145 | HMDB01891 |  |
| 146 | HMDB01906 |  |
| 147 | HMDB01921 |  |

|     |           |  |
|-----|-----------|--|
| 148 | HMDB02001 |    |
| 149 | HMDB02006 |    |
| 150 | HMDB02039 |    |
| 151 | HMDB02043 |   |
| 152 | HMDB02072 |  |
| 153 | HMDB02085 |  |
| 154 | HMDB02097 |  |

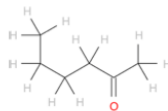
|     |           |  |
|-----|-----------|--|
| 155 | HMDB02128 |    |
| 156 | HMDB02182 |    |
| 157 | HMDB02243 |    |
| 158 | HMDB02339 |    |
| 159 | HMDB02455 |   |
| 160 | HMDB02820 |  |
| 161 | HMDB02939 |  |
| 162 | HMDB03119 |  |
| 163 | HMDB03243 |  |

|     |           |  |
|-----|-----------|--|
| 164 | HMDB03315 |    |
| 165 | HMDB03361 |    |
| 166 | HMDB03406 |    |
| 167 | HMDB03646 |    |
| 168 | HMDB03843 |  |
| 169 | HMDB04041 |  |
| 170 | HMDB04814 |  |
| 171 | HMDB04816 |  |
| 172 | HMDB04983 |  |



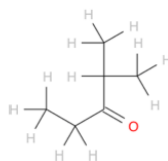
173

HMDB05842



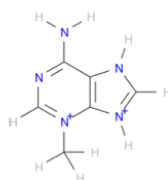
174

HMDB05846



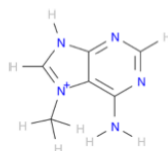
175

HMDB11600



176

HMDB11614



177

HMDB11745

