Problem Session (4) 2018.4.28. Daiki Kuwana

Please answer below problems.

1) Here shows how to calculate NMR chemical shifts of *cis*-3-methylcyclohexanol (**1-1**). First, conformational search was conducted in Macromodel. As a result, 6 conformers were outputted. The structures of these conformers were optimized in Gaussian, and NMR shifts were predicted. Resultant ^{13}C NMR shift of each conformer and its energy is shown below.

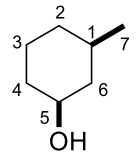
Configuration	cis-conf-1.out	cis-conf-2.out	cis-conf-3.out	cis-conf-4.out	cis-conf-5.out	cis-conf-6.out
C1	34.15624406	33.37701958	33.51862764	29.46188937	30.09199772	30.16688842
C2	34.96436039	34.85839194	34.82883482	31.28188557	32.45856301	32.69730089
C3	25.77950960	25.32693404	25.17686752	15.00940886	16.31571945	16.24158905
C4	37.59228284	33.79186466	37.52480517	33.15842996	32.59693975	35.82788443
C5	71.09095229	70.74824178	70.76563391	68.24852690	68.04333777	68.22514731
C6	46.24187417	46.13932712	42.23731230	38.73246531	39.78331116	36.32778939
C7	20.91665083	20.84318571	20.87226763	19.86200342	19.69521004	19.51881772
energy (a.u.)	-350.098736	-350.099026	-350.099049	-350.091363	-350.093640	-350.093436

i) The probability that conformer *i* (*i* = 1-6) exists can be written as:

$P(i) = \frac{1}{Z}\exp(-\frac{E_i}{RT})$ (Z: partition function, E_i: Energy of *i*, R: Gas constant, T: temperature)

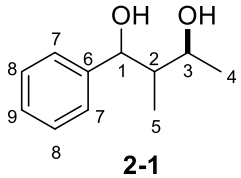
For each conformer *i*, calculate the probability that it exists at 298.15 K.

ii) Calculate the ^{13}C chemical shifts of **1-1** at 298.15 K.

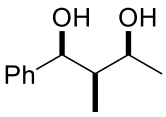


1-1

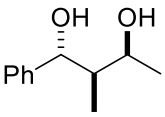
2) **2-1** is a product of an experiment, and one of **a**-**d**. But the stereochemistry is not determined.



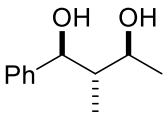
2-1



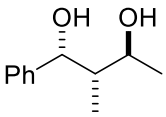
a



b



c



d

To determine the stereochemistry of **2-1**, NMR shifts were calculated. Resultant ^{13}C shifts are shown below.

Structure	2-1	a	b	c	d
C1	74.7	79.45	80.33	82.58	75.11
C2	45.7	47.23	46.50	49.75	46.02
C3	70.9	73.36	68.02	73.36	72.61
C4	21.9	23.23	21.00	22.74	22.37
C5	11.3	5.91	15.37	14.50	14.37
C6	142.7	139.85	140.82	139.93	139.28
C7	127.0	121.40	121.54	122.85	121.91
C8	128.0	122.38	122.94	123.23	122.43
C9	126.1	120.84	121.61	122.45	121.07

Now, DP4 analysis is going to be conducted. When there are *m* possible conformers and we have observed data δ_1-δ_N, the probability that conformation *i* (*i* = 1-*m*) is the correct one is described below.

$$P(i|\delta_1, \delta_2, ..., \delta_N) = \frac{\prod_{k=1}^{N}(1 - T^{\nu}(|(\delta^{i}_{\mathrm{scaled},k} - \delta_{\mathrm{exp},k}) - \mu|/\sigma))}{\sum_{j=1}^{m}[\prod_{k=1}^{N}(1 - T^{\nu}(|(\delta^{j}_{\mathrm{scaled},k} - \delta_{\mathrm{exp},k}) - \mu|/\sigma))]}$$

δ_{scaled}: Scaled chemical shifts, $T^{\nu}(x)$: Cumulative *t* distribution function, ν: Degrees of freedom, μ: Average error, σ: Standard deviation

i) For each diastereomer, derive δ_{scaled} from calculated chemical shifts (δ_{calc}) and experimental chemical shifts (δ_{exp}) by following method.

I. Plot δ_{calc} (y axis) vs δ_{exp} (x axis).

II. Get slope and interception of the regression line.

III. Calculate δ_{scaled} by using the formula δ_{scaled} = (δ_{calc} – intercept) / slope.

ii) In DP4 analysis, parameters are given (ν = 11.38, μ = 0 and σ = 2.306). Calculate P(*i*|**2-1**) for **a-d** respectively.
Hint: If you calculate in excel, function 'TDIST' is useful.

iii) Choose the most appropriate structure of **2-1** from **a**–**d**.

Problem Session (4) - Answer 2018.4.28. Daiki Kuwana

Theme: Prediction of NMR chemical shift

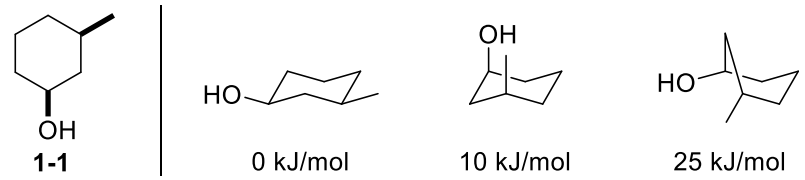
Useful method for correction of known compound or determination of unknown compound

(Sometimes used as evidence)

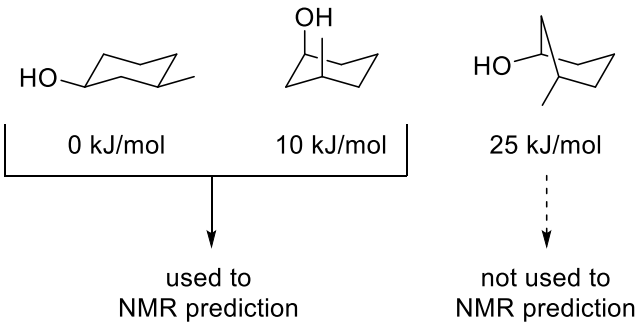
1. Derivation of calculated chemical shifts

1-1. NMR shift calculation

I. Conformational search in Macromodel → Possible conformations and their energies are calculated.



(II. Selection of low energy conformations → Possible conformations and their energies are calculated.)



III. Opt+Freq calculation using functional and basis

Opt: Optimization of structure

Freq: Vibrational frequency (If there is no imaginary frequency, the structure is regarded as “stable”)

Functional PM6, M06-2X, B3LYP, APFD, mPW1PW91, ωB97X,…

Basis: 6-31G(d), 6-31+G(d,p), …

IV. Calculation of NMR shielding constant by GIAO method for each conformer.

GIAO: Gauge Independent Atomic Orbitals.

V. Boltzmann averaging of shielding constant

VI. Calculation of chemical shift using shielding constant of TMS and that of step V.

(In Problem 1, step VI is conducted before step V)

1-2. Answer

i) A comparison of energies indicates that conformation 3 is the most stable (= lowest energy).

Substitution of *I* for the formula gives

$P(1) = \frac{1}{Z}\exp(-\frac{E_1}{RT})$, $P(3) = \frac{1}{Z}\exp(-\frac{E_3}{RT})$ $\frac{P(1)}{P(3)} = \exp\left(-\frac{E_1-E_3}{RT}\right)$

R = 8.31 J/(mol*K), T = 298.15 K, E_1-E_3 = 0.000313 a.u. = 824.9115 J/mol (1 a.u. = 2625 kJ/mol)

$\therefore \frac{P(1)}{P(3)} = \exp\left(-\frac{824.9115}{8.31*298.15}\right) = 0.717$

Using the same method, abundance ratio of each conformer to conformer 3 is obtained.

$\frac{P(2)}{P(3)} = 0.976, \frac{P(4)}{P(3)} = 2.81 \times 10^{-4}, \frac{P(5)}{P(3)} = 3.17 \times 10^{-3}, \frac{P(6)}{P(3)} = 2.55 \times 10^{-3}$

Conformers of **1-1** are only conformer 1-6, so P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 1

$\therefore$ P(1) = 0.266, P(2) = 0.362, P(3) = 0.371, P(4) = $1.04*10^{-4}$, P(5) = $1.18*10^{-3}$, P(6) = $9.46*10^{-4}$,

filename	energy (a.u.)	relative energy (a.u.)	relative energy (kJ/mol)	Boltzmann factor	the possibility
cis-conf-1.out	-350.098736	0.000313	0.8249115	0.716818103	0.265620357
cis-conf-2.out	-350.099026	2.3E-05	0.0606165	0.975832099	0.361599224
cis-conf-3.out	-350.099049	0	0	1	0.370554755
cis-conf-4.out	-350.091363	0.007686	20.256453	0.000281472	0.000104301
cis-conf-5.out	-350.09364	0.005409	14.2554195	0.003171762	0.001175312
cis-conf-6.out	-350.093436	0.005613	14.7930615	0.002553068	0.000946051

ii) Calculated shift of **1-1** is the sum of (calculated shift of *i*)*(probability of *i*).

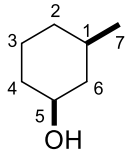
For example: C1

(calculated shift of C1) = 34.2*0.266 + 33.4*0.362 + 33.5*0.371 + 29.5*($1.04*10^{-4}$) + 30.1*($1.18*10^{-3}$) + 30.2*($9.46*10^{-4}$) = 33.6

Using the same formula, chemical shift of **1-1** can be calculated.

C1: 33.6 ppm, C2: 34.9 ppm, C3: 25.4 ppm, C4: 36.2 ppm, C5: 70.8 ppm, C6: 44.7 ppm. C7: 20.9 ppm.

Configuration	cis-conf-1.out	cis-conf-2.out	cis-conf-3.out	cis-conf-4.out	cis-conf-5.out	cis-conf-6.out	Average shift
the possibility	0.265620357	0.361599224	0.370554755	0.000104301	0.001175312	0.000946051	
C1	34.15624406	33.37701958	33.51862764	29.46188937	30.09199772	30.16688842	33.62916478
C2	34.96436039	34.85839194	34.82883482	31.28188557	32.45856301	32.69730089	34.8703487
C3	25.7795096	25.32693404	25.17686752	15.00940886	16.31571945	16.24158905	25.37127714
C4	37.59228284	33.79186466	37.52480517	33.15842996	32.59693975	35.82788443	36.18504765
C5	71.09095229	70.74824178	70.76563391	68.2485269	68.04333777	68.22514731	70.8398906
C6	46.24187417	46.13932712	42.2373123	38.73246531	39.78331116	36.32778939	44.7031305
C7	20.91665083	20.84318571	20.87226763	19.86200342	19.69521004	19.51881772	20.8707715



1-1

2. DP4 analysis: Comparison based on one set of experimental data and multiple possible structures

$$P(i|\delta_1, \delta_2, ..., \delta_N) = \frac{\prod_{k=1}^{N}(1 - T^{\nu}(|(\delta^{i}_{\mathrm{scaled},k} - \delta_{\mathrm{exp},k}) - \mu|/\sigma))}{\sum_{j=1}^{m}[\prod_{k=1}^{N}(1 - T^{\nu}(|(\delta^{j}_{\mathrm{scaled},k} - \delta_{\mathrm{exp},k}) - \mu|/\sigma))]}$$

1- $T^{\nu}(|(\delta_{\mathrm{scaled},k}-\delta_{\mathrm{exp},k})-\mu/\sigma|)$: the probability of getting an error larger than $e_k = \delta_{\mathrm{scaled},k}-\delta_{\mathrm{exp},k}$ when observed error of C_k is e_k. When the error is small, this value is large.

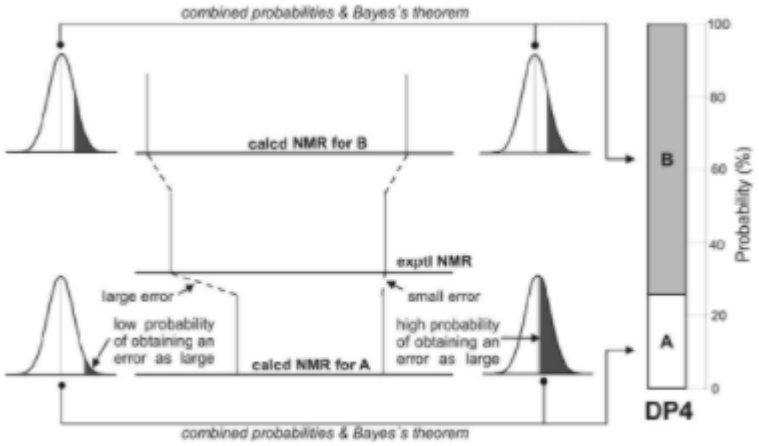


Figure 8. Schematic illustration of the DP4 probability.

(Sarotti *et al. Chem. Eur. J.* **2016**, 12246.)

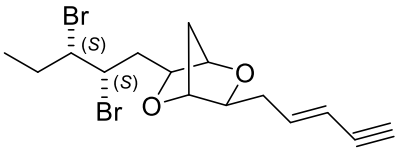
For each nucleus, this probability can be calculated. They are independent from each other, so when we consider at all shifts of conformation *i*, the product is described.

Parameters ν, μ, σ are determined by analyses of already isolated/synthesized compounds.

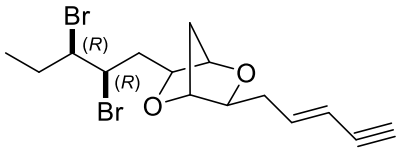
(1717 ^{13}C and 1794 ^{1}H shifts)

Parameter μ is 0 because an individual error is like equally to be positive and negative.

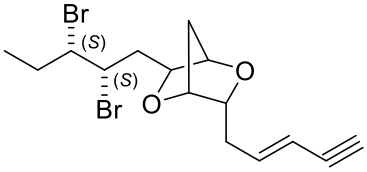
2-1. Recently reported DP4 assessment: (*E*)-Ocellenyne



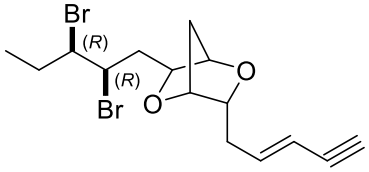
SS-syn (originally proposed)
<1%



RR-syn (Originally proposed)
<1%



SS-anti (biogenetically proposed)
15%



RR-anti
85%

→ ***RR-anti*** is the most likely structure of (*E*)-Ocellenyne computationally.

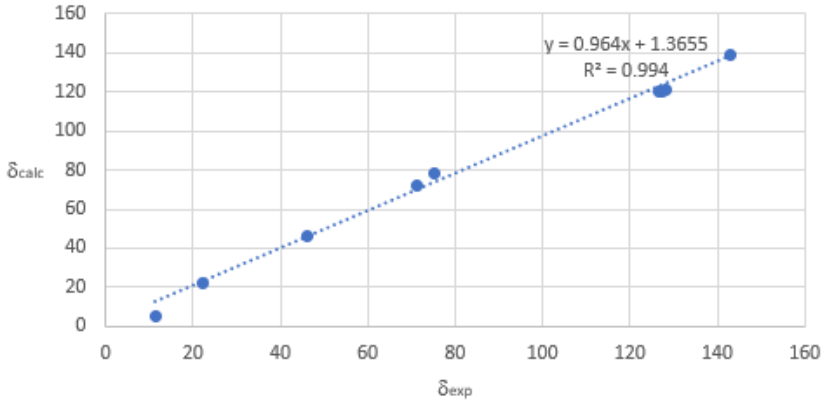
2-2. Answer

i)~ iii) Scaled chemical shifts and P(*i*|**2-1**) of **a**-**d**

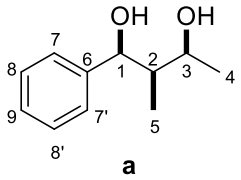
a

	δ_{exp} (2-1)	δ_{calc}	δ_{scaled}	$(\delta_{scaled}-\delta_{exp})/\sigma$	$1-T((\delta_{scaled}-\delta_{exp})/\sigma)$
C1	74.7	79.45	81.00	2.732228392	0.009750253
C2	45.7	47.23	47.58	0.814085931	0.216432897
C3	70.9	73.36	74.68	1.640542622	0.064573624
C4	21.9	23.23	22.68	0.338688897	0.370610469
C5	11.3	5.91	4.71	2.855935985	0.007814933
C6	142.7	139.85	143.66	0.414622867	0.343194434
C7	127.0	121.40	124.52	1.076705905	0.152329457
C8	128.0	122.38	125.53	1.069508373	0.153874005
C9	126.1	120.84	123.94	0.938333339	0.184109819

a



slope
0.964
intercept
1.3655
Product
5.84526E-10
P(a|2-1)
5.45941E-05



a

For example: C1

i) δ_{calc} = 79.45 ppm is substituted for regression line y = 0.964x+1.3655, 79.45 = 0.964*δ_{scaled} +1.3655

δ_{scaled} = (79.45-1.3655)/0.964 = <u>81.00</u>

ii) In the formula of P(*i*|**2-1**), $\mu = 0$, $\sigma = 2.306$, $\nu = 11.38$ so

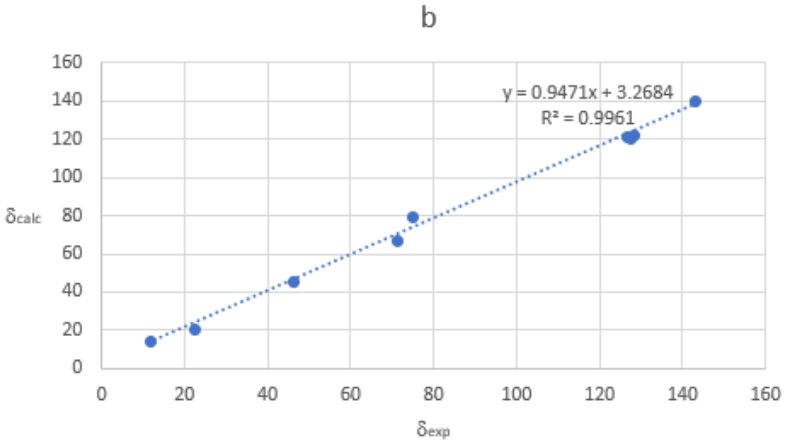
$$\frac{|(\delta_{scaled}-\delta_{exp})-\mu|}{\sigma} = \frac{|(81.00-74.7)-0|}{2.306} = 2.732,\ 1-T^{\nu}\left(\frac{|(\delta_{scaled}-\delta_{exp})-\mu|}{\sigma}\right) = TDIST(2.732, 11.38, 1) = 9.75\times10^{-3}$$

$$\prod_{k=1}^{9}\left[1-T^{\nu}\left(\frac{|(\delta_{scaled,k}-\delta_{exp,k})-\mu|}{\sigma}\right)\right] = (9.75\times10^{-3})\times0.216\times0.0646\times\ldots = 5.85\times10^{-10}$$

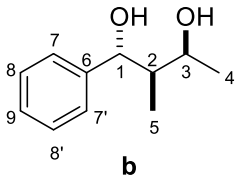
$$P(\boldsymbol{a}|\mathbf{2}\text{-}\mathbf{1}) = \frac{5.85\times10^{-10}}{5.85\times10^{-10}+2.99\times10^{-8}+1.85\times10^{-8}+1.07\times10^{-5}} = 5.46\times10^{-5}$$

b

	δexp (2-1)	δcalc	δscaled	(δscaled-δexp)/σ	1-T((δscaled-δexp)/σ)
C1	74.7	80.33	81.37	2.8906564	0.007344293
C2	45.7	46.50	45.65	0.02329199	0.490917258
C3	70.9	68.02	68.37	1.09788286	0.147853155
C4	21.9	21.00	18.72	1.378146811	0.097772015
C5	11.3	15.37	12.78	0.640733483	0.267408063
C6	142.7	140.82	145.23	1.099091644	0.147600706
C7	127.0	121.54	124.88	0.920370148	0.188560978
C8	128.0	122.94	126.36	0.71299955	0.245347656
C9	126.1	121.61	124.95	0.498032841	0.31413386



slope
0.9471
intercept
3.2684
product
2.98961E-08
P(b|2-1)
0.002784493



For example: C1

i) δ_{calc} = 80.33 ppm is substituted for regression line y = 0.9471x+3.2684, 80.33 = 0.9471*δ_{scaled} +3.2684

δ_{scaled} = (80.33-3.2684)/0.9471 = <u>81.37</u>

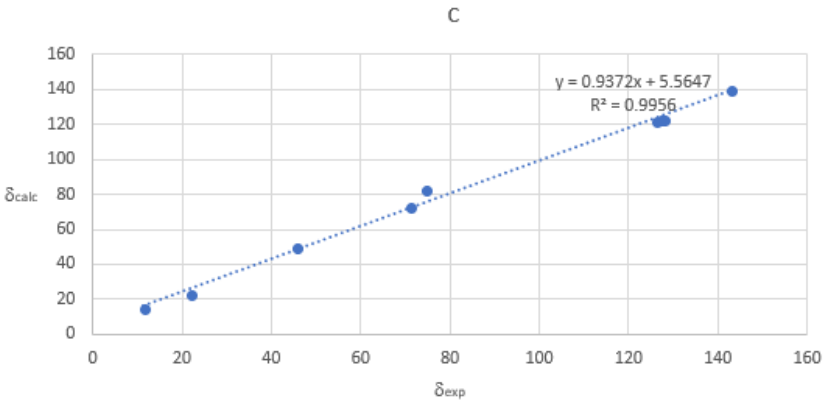
ii) $\frac{|(\delta_{scaled}-\delta_{exp})-\mu|}{\sigma} = \frac{|(81{,}37-74.7)-0|}{2.306} = 2.891,\ \ 1-T^{\nu}\left(\frac{|(\delta_{scaled}-\delta_{exp})-\mu|}{\sigma}\right) = TDIST(2.891, 11.38, 1) = 7.34\times10^{-3}$

$$\prod_{k=1}^{9}\left[1-T^{\nu}\left(\frac{|(\delta_{scaled,k}-\delta_{exp,k})-\mu|}{\sigma}\right)\right] = (7.34\times10^{-3})\times0.491\times0.148\times\ldots = 3.32\times10^{-9}$$

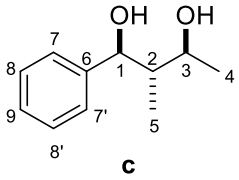
$$P(\boldsymbol{b}|\mathbf{2\text{-}1}) = \frac{2.99\times10^{-8}}{5.85\times10^{-10}+2.99\times10^{-8}+1.85\times10^{-8}+1.07\times10^{-5}} = 2.78\times10^{-3}$$

c

	δexp (2-1)	δcalc	δscaled	(δscaled-δexp)/σ	1-T((δscaled-δexp)/σ)
C1	74.7	82.58	82.18	3.24195561	0.003923317
C2	45.7	49.75	47.15	0.627091678	0.271697779
C3	70.9	73.36	72.34	0.62364912	0.272786456
C4	21.9	22.74	18.33	1.549789948	0.074733662
C5	11.3	14.50	9.53	0.765811987	0.229952793
C6	142.7	139.93	143.37	0.290054078	0.388584813
C7	127.0	122.85	125.14	0.804698093	0.219020366
C8	128.0	123.23	125.55	1.062519827	0.155384989
C9	126.1	122.45	124.72	0.599495684	0.280493682



slope
0.9372
intercept
5.5647
product
1.85362E-08
P(c|2-1)
0.001731258



c

For example: C1

i) δ_{calc} = 82.58 ppm is substituted for regression line y = 0.9372x+5.5647, 82.58 = 0.9372*δ_{scaled} +5.5647

δ_{scaled} = (82.58-5.5647)/0.9372 = <u>82.18</u>

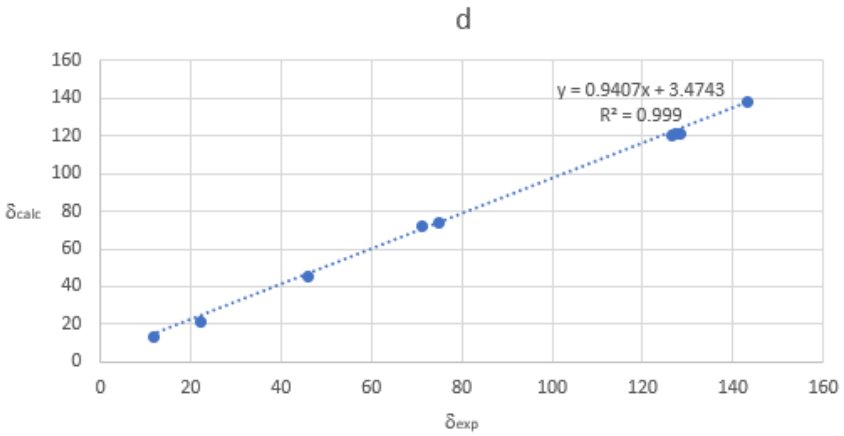
ii) $\frac{|(\delta_{scaled}-\delta_{exp})-\mu|}{\sigma} = \frac{|(82{,}18-74.7)-0|}{2.306} = 3.242,\ \ 1-T^{\nu}\left(\frac{|(\delta_{scaled}-\delta_{exp})-\mu|}{\sigma}\right) = TDIST(3.242, 11.38, 1) = 3.92\times10^{-3}$

$$\prod_{k=1}^{9}\left[1-T^{\nu}\left(\frac{|(\delta_{scaled,k}-\delta_{exp,k})-\mu|}{\sigma}\right)\right] = (3.92\times10^{-3})\times0.272\times0.273\times\ldots = 1.85\times10^{-8}$$

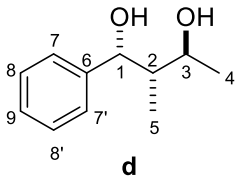
$$P(\boldsymbol{c}|\mathbf{2}\text{-}\mathbf{1}) = \frac{1.85\times10^{-8}}{5.85\times10^{-10}+2.99\times10^{-8}+1.85\times10^{-8}+1.07\times10^{-5}} = 1.73\times10^{-3}$$

d

	δ_{exp} (2-1)	δ_{calc}	δ_{scaled}	$(\delta_{scaled}-\delta_{exp})/\sigma$	$1-T((\delta_{scaled}-\delta_{exp})/\sigma)$
C1	74.7	75.11	76.15	0.629437527	0.270957341
C2	45.7	46.02	45.23	0.204812327	0.420729645
C3	70.9	72.61	73.49	1.124842815	0.142300728
C4	21.9	22.37	20.09	0.78627484	0.224156884
C5	11.3	14.37	11.58	0.122525982	0.452346179
C6	142.7	139.28	144.37	0.722741484	0.242461009
C7	127.0	121.91	125.90	0.476292728	0.32159296
C8	128.0	122.43	126.45	0.670230349	0.258267388
C9	126.1	121.07	125.01	0.473236378	0.322648321



slope
0.9407
intercept
3.4743
product
1.06876E-05
P(d|2-1)
0.998214148



d

For example: C1

i) δ_{calc} = 75.11 ppm is substituted for regression line y = 0.9407x+3.4743, 75.11 = 0.9407*δ_{scaled} +3.4743

δ_{scaled} = (75.11-3.4743)/0.9407 = <u>76.15</u>

ii) $\frac{|(\delta_{scaled}-\delta_{exp})-\mu|}{\sigma} = \frac{|(76.15-74.7)-0|}{2.306} = 0.629,\ 1-T^{\nu}\left(\frac{|(\delta_{scaled}-\delta_{exp})-\mu|}{\sigma}\right) = TDIST(0.629, 11.38, 1) = 0.271$

$$\prod_{k=1}^{9}\left[1-T^{\nu}\left(\frac{|(\delta_{scaled,k}-\delta_{exp,k})-\mu|}{\sigma}\right)\right] = 0.271\times 0.421\times 0.142\times \ldots = 1.07\times 10^{-5}$$

$$P(\boldsymbol{d}|\mathbf{2\text{-}1}) = \frac{1.07\times 10^{-5}}{5.85\times 10^{-10}+2.99\times 10^{-8}+1.85\times 10^{-8}+1.07\times 10^{-5}} = 0.998$$

iii) The probability that **d** is the correct structure is 99.8%. Therefore, **<u>d</u>** is the most appropriate structure.

2-3. Methods of analysis

Correlation: Plotting computed shifts vs experimental shifts

MAE (Mean Absolute Errors): Average of the difference between experimental shifts and calculated ones

CMAE (Corrected Mean Absoluted Errors): MAE using experimental shifts and scaled shifts

CP(Comparison Parameters)3: Goodman *et al. J. Org. Chem.* **2009**, 4597.

DP4: Goodman *et al. J. Am. Chem. Soc.* **2010**, 12946. http://www-jmg.ch.cam.ac.uk/tools/nmr/DP4/

DP4+: Sarotti *et al. J. Org. Chem.*, **2015**, 12526. https://sarotti-nmr.weebly.com/

ANN-PRA (Artificial Neural Networks – Pattern Recognition Analysis): Sarotti *et al. Org. Biomol. Chem.*, **2013**, 4847.

2-3-1. MAE

	δ_{exp} (2-1)	$\delta_{d,calc}$	$\lvert\delta_{d,calc}-\delta_{exp}\rvert$
C1	74.7	75.11	0.41
C2	45.7	46.02	0.32
C3	70.9	72.61	1.71
C4	21.9	22.37	0.47
C5	11.3	14.37	3.07
C6	142.7	139.28	3.42
C7	127.0	121.91	5.09
C8	128.0	122.43	5.57
C9	126.1	121.07	5.03

MAE
2.787777778

Experimental and calculated data are regarded as the same when MAE is under 2.0 ppm (^{13}C), 0.1 ppm (^{1}H).

So, if we use MAE, we cannot recognize **d** as correct structure of **2-1**.

2-3-2. CP3

Comparison based on two sets of experimental data and two possible structures.

In this case, there are two combinations ([exp1-calc1, exp2-calc2] or [exp1-calc2, exp2-calc1]).

For each combination, CP3 value is calculated.

Using CP3 value, the probability that assignment combination is right is calculated.

$$\mathrm{CP3} = \frac{\sum_i f_3(\Delta_{exp}, \Delta_{calc})}{\sum_i \Delta_{exp}^2} \quad \text{where} \quad f_3(\Delta_{exp}, \Delta_{calc}) = \begin{cases} \Delta_{exp}^3/\Delta_{calc} \text{ if } \Delta_{calc}/\Delta_{exp} > 1 \\ \Delta_{exp}\Delta_{calc} \text{ otherwise} \end{cases}$$

2-3-3. DP4+

In calculation of DP4, only scaled shifts are used.

→ Errors are independent from chemical environment.
There is a risk of false positive when computed chemical shifts are similar.

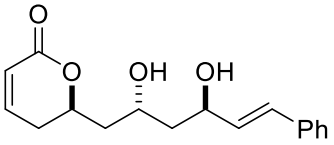
→ Original data is considered.

$$P(i) = \frac{\prod_{k=1}^{N}[1-T^{v}_{s}(e^{i}_{s,k}/\sigma_{s})]\prod[1-T^{v}_{u\text{-}spx}((e^{i}_{u,k}-\mu_{u\text{-}spx})/\sigma_{u\text{-}spx})]}{\sum_{j=1}^{m}\prod_{k=1}^{N}[1-T^{v}_{s}(e^{j}_{s,k}/\sigma_{s})]\prod[1-T^{v}_{u\text{-}spx}((e^{j}_{u,k}-\mu_{u\text{-}spx})/\sigma_{u\text{-}spx})]}$$

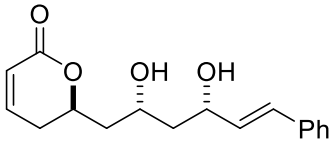
T^{v}, σ, μ, and e^{i} computed from B3LYP/6-31G* geometries

T_s, e_s etc.: scaled data, T_u, e_u etc.: unscaled(=originally calculated) data

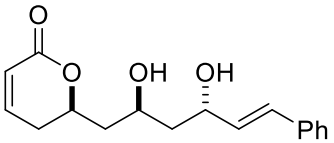
2-3-3-1. Example of successful DP4+ assessment: Cryptomoscatone Sarotti *et al. J. Org. Chem.* **2015**, 12526.



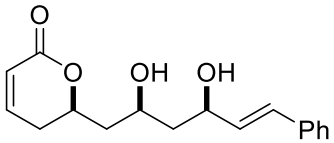
2-2a (Cryptomoscatone D2)
DP4: 1% (misassigned to **2-2d** in >95% probability)
DP4+: 99%



2-2b (Cryptomoscatone D1)
DP4: 3% (misassigned to **2-2d** in >95% probability)
DP4+: 98%



2-2c



2-2d

2-3-4. ANN-PRA

CP3, DP4, DP4+: Comparison of candidate structures

ANN-PRA: Absolute evaluation

First, 18 descriptors (MAE, CMAE etc.) are obtained using experimental and calculated data.

Then, analysis is conducted in hidden layer by weighing descriptors.

Finally whether the structure is right or wrong is judged from output layer.

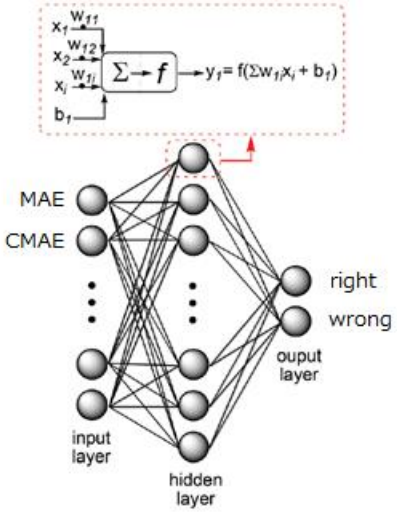


Fig. 4 Schematic illustration of a two-layer feed-forward ANN used in this study.

2-3-5. Summary

CP3: two sets of experimental data and two sets of calculated data

DP4, DP4+, MAE, CMAE: one set of experimental data and multiple sets of calculated data

ANN-PRA: one set of experimental data and one set of calculated data