VIBFREQ1295: A New Database for Vibrational Frequency Calculations

Supplementary Material

Juan C. Zapata Trujillo[1], Laura K. McKemmish[1,*]

[1] *School of Chemistry, University of New South Wales, Sydney, NSW, 2052, Australia*

* l.mckemmish@unsw.edu.au

1 Functional Groups

Based on the SMARTS identifiers for common functional groups in organic chemistry, we used the RDKit library in python to obatin the counts of all functional groups in the molecules present in VIBFREQ1295. In Table S1 we present the description, count, classification, and SMARTS identifier for each functional group. All functional groups are sorted into four main classes depending on their atomic composition: hydrocarbons, O-containing groups, N-containing groups, S-containing groups, and halogen-containing groups.

Table S1: Functional group descriptions, counts and SMARTS identifiers in VIBFREQ1295

Functional Group	Detailed Description	Count	Class	SMARTS Identifier
Alkane	Alkyl Carbon	49	Hydrocarbon	[CX4]
Alkene	Allenic Carbon	2	Hydrocarbon	[$([CX2](=C)=C)]
Alkene	Vinylic Carbon	18	Hydrocarbon	[$([CX3]=[CX3])]
Alkyne	Acetylenic Carbon	20	Hydrocarbon	[$([CX2]#C)]
Aromatic Hydrocarbon	Aromatic Hydrocarbon	24	Hydrocarbon	c
Carbonyl Group	Aldehyde	4	O-containing Groups	[CX3H1](=O)[#6]
Carbonyl Group	Carboxylic Acid	2	O-containing Groups	[CX3](=O)[OX2H1]
Carbonyl Group	Ether	4	O-containing Groups	[OD2]([#6])[#6]
Hydroxyl	Hydroxyl in Alcohol	3	O-containing Groups	[#6][OX2H]
Hydroxyl	Hydroxyl in Carboxylic Acid	2	O-containing Groups	[OX2H][CX3]=[OX1]
Hydroxyl	Hydroxyl_acidic	3	O-containing Groups	[$([OH]-*=[!#6])]
Peroxyde	Peroxyde	1	O-containing Groups	[OX2,OX1-][OX2,OX1-]
Amine	Amine	4	N-containing Groups	[NX3;H2,H1;!$(NC=O)]
Azo Compounds	Diazene	1	N-containing Groups	[NX2]=[NX2]
Azo Compounds	Diazo	1	N-containing Groups	[([6] = [N+] = [N−]),([#6-]-[N+]#[N])]
Azo Compounds	Azole	6	N-containing Groups	[([nr5] : [nr5, or5, sr5]),([nr5]:[cr5]:[nr5,or5,sr5])]
Hydrazine	Hydrazine	1	N-containing Groups	[NX3][NX3]
Nitrate Group	Nitrate Group	1	N-containing Groups	[([NX3](= [OX1])(= [OX1])O),([NX3+]([OX1-])(=[OX1])O)]
Nitrile	Nitrile	9	N-containing Groups	[NX1]#[CX2]
Nitrile	Isonitrile	1	N-containing Groups	[CX1-]#[NX2+]
Nitroso Group	Nitroso Group	6	N-containing Groups	[NX2]=[OX1]
Thio group	Thio analog to Carbonyl	2	S-containing Groups	[#6X3](=[SX1])([!N])[!N]
Thio group	Sulfide or Disulfide Sulfur	14	S-containing Groups	[SX2]
Thio group	Thiol	1	S-containing Groups	[#16X2H]
Sulfide	Mono-sulfide	14	S-containing Groups	[#16X2H0][!#16]
Sulfide	Di-sulfide	2	S-containing Groups	[#16X2H0][#16X2H0]
Sulfoxide	Sulfoxide	3	S-containing Groups	[([16X3] = [OX1]),([#16X3+][OX1-])]
Carbon attached to Halogen	Carbon attached to Halogen	45	Halogen-containing Groups	[#6][F,Cl,Br,I]
Halogen	Halogen	104	Halogen-containing Groups	[F,Cl,Br,I]
Three halides groups	Three halides groups	22	Halogen-containing Groups	[F,Cl,Br,I].[F,Cl,Br,I].[F,Cl,Br,I]
Acyl Halide	Acyl Halide	7	Halogen-containing Groups	[CX3](=[OX1])[F,Cl,Br,I]

2 Thresholds for Defining Frequency Regions

As noted in (1), the standard approach for producing scaling factors is more sensitive towards high- than low-frequency vibrations, and thus defining scaling factors for different frequency regions can result in better scaling of the computed harmonic frequencies.

Though previous studies have split the frequency range into low- and high-frequency vibrations using $1000\,cm^{-1}$ as the turning point between both regimes (2), a closer look into our database (see Figure 3) suggests three regions of importance: low, mid and high-frequency vibrations. To examine how sensitive the scaling factor and scaling factor's performance are to the partition of these frequency regions, we studied the effect of three different thresholds in the calculated metrics:

1. Low: $0 - 500\,cm^{-1}$, mid: $500 - 1500\,cm^{-1}$, and high: $\geq 1500\,cm^{-1}$
2. Low: $0 - 1000\,cm^{-1}$, mid: $1000 - 2000\,cm^{-1}$, and high: $\geq 2000\,cm^{-1}$
3. Low: $0 - 1500\,cm^{-1}$, mid: $1500 - 2500\,cm^{-1}$, and high: $\geq 2500\,cm^{-1}$

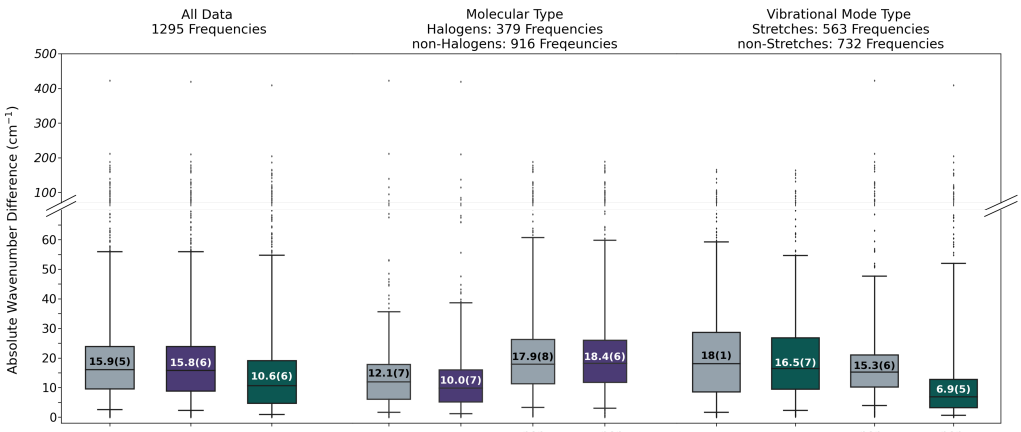
In Table S2 we present the calculated scaling factor, RMSE, mean absolute percentage error (MAPE), as well as the first, second (median), and third quartiles for each frequency range, individually, using the aforementioned frequency thresholds. The table shows that, apart from the scaling factor values, the frequency threshold used does not significantly influence the reported statistical metrics (in most cases, variations are within the reported standard deviations). Nonetheless, the second group of thresholds (low: $0 - 1000\,cm^{-1}$, mid: $1000 - 2000\,cm^{-1}$, and high: $\geq 2000\,cm^{-1}$) highlights as a good compromise in performance and stability, as evidenced by its rather small deviations when implementing frequency-range-specific scaling factors. Therefore, further discussions in this manuscript will focus on low-, mid- and high-frequency regions as defined by this frequency thresholds.

Table S2: Scaling factors (SF) and statistical metrics for the low-, mid-, and high-frequency regions defining three different thresholds. The RMSE, and quartiles (Q1, Q2, Q3) are given in cm^{-1}, whereas the Mean Absolute Percentage Error (MAPE) is given in percentage. Number sin parenthesis indicate one standard deviation in the last reported digit.

		Frequency-range-specific Scaling Factor						Global Scaling Factor					
Threshold	Count	SF	RMSE	MAPE	Q1	Q2	Q3	SF	RMSE	MAPE	Q1	Q2	Q3
Low-frequency Range													
$< 500\,cm^{-1}$	192	1.006(5)	31(8)	5.1(9)	4(1)	7(1)	13(2)	0.9617	34(8)	5.9(9)	5.8(9)	11(1)	17(2)
$< 1000\,cm^{-1}$	570	0.987(1)	29(7)	2.8(4)	2.4(4)	6.2(8)	13(1)	0.9617	34(7)	3.9(4)	8.4(8)	14(1)	21.0(9)
$< 1500\,cm^{-1}$	937	0.9775(7)	29(5)	2.2(3)	3.4(3)	6.7(4)	12.7(7)	0.9617	33(5)	3.0(3)	9.9(5)	15.4(5)	21.4(7)
Mid-frequency Range													
$500 - 1500\,cm^{-1}$	745	0.9767(6)	28(6)	1.4(1)	3.5(4)	6.7(5)	12.4(9)	0.9617	33(6)	2.2(1)	10.9(5)	16.1(5)	22.1(7)
$1000 - 2000\,cm^{-1}$	425	0.9727(6)	24(4)	1.0(1)	2.7(5)	6.0(6)	11.1(1)	0.9617	28(4)	1.6(1)	11.0(6)	15.8(7)	22(1)
$1500 - 2500\,cm^{-1}$	121	0.9706(6)	19(2)	0.74(8)	5(1)	11(2)	20(2)	0.9617	25(2)	1.07(9)	9(1)	16(3)	29(4)
High-frequency Range													
$\geq 1500\,cm^{-1}$	355	0.9576(3)	29(3)	0.94(7)	10(2)	16(1)	27(3)	0.9617	31(3)	0.94(7)	9(1)	20(2)	30(2)
$\geq 2000\,cm^{-1}$	297	0.9564(4)	30(3)	0.78(8)	7(2)	14(1)	26(5)	0.9617	34(4)	0.92(7)	10(2)	22(2)	34(3)
$\geq 2500\,cm^{-1}$	234	0.9545(3)	25(4)	0.55(7)	4.5(7)	11(2)	23(3)	0.9617	34(4)	0.86(7)	12(3)	23(1)	31(2)

3 Molecular-Type and Vibrational Mode-Specific Scaling Factors

As important as it is to examine the dependence of the scaling factor on the different frequency regions, a detailed analysis on the scaling factor dependence on the different molecular types (halogens and non-halogens) and vibrational modes (stretches and non-stretches) in the database is needed. In Figure S1 we present the absolute wavenumber difference (in cm^{-1}) between the scaled harmonic and experimental fundamental frequencies for the different molecular and vibrational mode classifications using the following scaling factors: in grey, using the global scaling



Scaling Factor	0.9617(3)	Multiple	Multiple	0.9617(3)	0.9651(9)	0.9617(3)	0.9610(4)	0.9617(3)	0.9584(4)	0.9617(3)	0.9766(6)
RMSE (cm^{-1})	32(4)	32(4)	30(4)	30(11)	30(11)	33(3)	33(3)	31(3)	30(2)	33(6)	29(6)
MAPE (%)	2.4(2)	2.4(2)	1,9(2)	2.1(2)	1.9(3)	2.5(2)	2.5(2)	1.4(1)	1.5(1)	3.1(3)	2.3(3)
Q1 25% (cm^{-1})	9.6(5)	8.8(5)	4.8(3)	6(1)	5.3(8)	11.1(7)	11.8(7)	8(1)	9.4(9)	10.1(5)	3.3(4)
Q2 50% (cm^{-1})	15.9(5)	15.8(6)	10.6(6)	12.1(7)	10.0(7)	17.9(8)	18.4(6)	18(1)	16.5(7)	15.3(6)	6.9(5)
Q3 75% (cm^{-1})	24(1)	24.1(9)	19(1)	18(1)	16.1(1)	27(1)	26(1)	29(1)	27(2)	21(1)	12.9(8)

Figure S1: Absolute wavenumber difference (cm^{-1}) between scaled harmonic and experimental fundamental frequencies using the global (grey), molecular-type-specific (purple), and vibrational mode-specific (dark-green) scaling factors. The numbers inside the boxes represent the median values for each classification. A scale discontinuity in the y-axis has been imposed to highlight both the box and whiskers, and the outliers in the data. The table at the bottom of the figure presents the scaling factor and RMSE for each classification individually.

factor (0.9617(3)) reported in Table 4 in the main manuscript; in purple, using two scaling factors optimised for halogen (0.9651(9)) and non-halogen (0.9610(4)) molecules, respectively; and in dark-green, two scaling factors optimised for stretching (0.9584(4)) and non-stretching (0.9766(6)) vibrational modes, respectively. The boxes in the middle and right-most side of the figure showcase the difference between using the global and specific scaling factors for each group. The corresponding scaling factors, RMSEs, mean absolute percentage error (MAPE), as well as the first, second (also displayed inside each box), and third quartiles are presented in the table at the bottom of the figure.

Focusing on the molecular types considered (halogens and non-halogens), the figure shows identical statistical metrics when using the global or molecule-specific scaling factors (medians of 15.9(5) vs 15.8(6) cm^{-1}, respectively). It is therefore evident that there is not explicit need to develop scaling factors for different molecular groups in the database and, thus, the global scaling factor (0.9617(3)) reported in Table 4 can be safely used without significantly compromising accuracy, at least when considering organic-like molecules. The effect of having molecules with, for instance, heavy metals in the database is unclear to us and our results may not generalise to this case. We can further validate this result by looking at the middle boxes in the figure, where using molecular type-specific scaling factors, i.e., halogen and non-halogen scaling factors (in purple), only results in minimal improvement compared to the global scaling (in grey).

One interesting aspect to highlight from the comparison between halogen and non-halogen molecules in the figure (purple middle boxes) is that similar scaling factors (0.9651(9) vs 0.9610(4)) and RMSEs (30(11) and 33(3) cm^{-1}) are found in both cases. This suggests that, despite the large counts of

halogen-containing molecules in the database (see Figures 4 and 5), applications of VIBFREQ1295 are not likely to be biased towards a particular molecular group.

Table S3: Scaling factors (SF) and statistical metrics for the low-, mid-, and high-frequency regions defining three different thresholds. The RMSE, and quartiles (Q1, Q2, Q3) are given in cm^{-1}, whereas the Mean Absolute Percentage Error (MAPE) is given in percentage. Number sin parenthesis indicate one standard deviation in the last reported digit.

Data	No. Freqs	Scaling Factor Type	SF	RMSD	MAPE	Q1	Q2	Q3
Stretch	562	Stretch	0.9584(4)	30(2)	1.5(1)	9.4(9)	16.5(7)	27(2)
Stretch	562	High	0.9564(4)	31(3)	1.5(1)	9.2(9)	17(1)	28(2)
Stretch and High	237	High	0.9564(4)	30(3)	0.79(7)	8(1)	14(2)	26(4)
Stretch and High	237	Stretch and High	0.9564(4)	30(3)	0.80(8)	8(1)	14(2)	27(4)
Non-stretch	730	Non-stretch	0.9766(6)	29(6)	2.3(3)	3.3(4)	6.9(5)	12.9(8)
Non-stretch	730	Mid	0.9727(6)	31(6)	2.4(3)	3.6(3)	7.3(5)	13.4(8)
Non-stretch and Mid	52	Mid	0.9727(6)	23(5)	0.9(1)	2.7(5)	5.4(6)	10(1)
Non-stretch and Mid	52	Non-stretch and Mid	0.9722(2)	21(5)	0.9(1)	2.8(4)	5.5(5)	10.1(9)
Non-stretch	730	Low	0.987(1)	31(6)	2.5(3)	3.4(5)	10.5(8)	21(2)
Non-stretrch and Low	678	Low	0.987(1)	31(10)	2.9(5)	2.0(3)	4.9(8)	11(1)
Non-stretrch and Low	678	Non-stretch and Low	0.992(2)	32(9)	3.1(5)	2.8(6)	5.8(8)	12(2)

Regarding the vibrational mode types, Figure S1 shows a rather small overall improvement when utilising mode-specific scaling factors instead of the global scaling factor (left-most boxes: RMSE of 33(4) and 30(4) cm^{-1} and medians of 15.9(5) and 10.6(6) cm^{-1} for the global and mode-specific scaling, respectively). The source of this improvement predominantly comes from the better scaling of the non-stretching frequencies when using a non-stretch-specific scaling factor; differently to the stretching modes where the improvement is not that significant. We can attribute this behaviour to the different degrees of anharmonicty between the two types of modes, as evidenced by their specific scaling factors (0.9584(4) and 0.9766(6)) for the stretching and non-stretching modes, respectively). Stretching modes can naturally lead to bond dissociation, a process that is not contemplated under the harmonic approximation and is thus better represented by anharmonic treatments, while most non-stretching modes are harmonically bound. This finding may potentially suggests the need of different scaling factors whenever a particular vibrational mode is considered, similar to alternative approaches that seek to scale the vibrational force constants in the calculations instead of the calculated harmonic frequencies (3; 4; 5; 6; 7; 8; 9; 10; 11).

Nonetheless, as shown in Figure 7 in the main manuscript, stretching and non-stretching modes in the database can be generally assigned to the high and low/mid-frequency ranges of the vibrational spectrum, respectively. Indeed, 100 % of the non-stretching frequencies in VIBFREQ1295 lies below 2000 cm^{-1} (the low- and mid-frequency ranges), whereas more than 50 % of the stretching frequencies lie in the high-frequency range above 2000 cm^{-1}. It is, therefore, likely that both scaling factors (mode and frequency-range-specific) are accounting for the same corrections and could be potentially interchanged.

To analyse this further, in Table S3 we present the statistical metrics for the stretching and non-stretching frequencies in VIBFREQ1295 using different combinations of scaling factors including the frequency ranges considered. Effectively, using frequency-range-specific scaling factors on the corresponding vibrational modes leads to virtually the same statistical metrics as calculated with

the mode-specific scaling factors (first two rows for each group in the table). A similar trend is observed when combining frequencies belonging to a particular vibrational mode type and frequency range, e.g., stretching frequencies in the high-frequency range ($2000\,cm^{-1}$), as shown in the last two rows for each group in Table S3. In either case, the statistical metrics remain fairly similar thus indicating that frequency-range-specific scaling factors can be used over mode-specific or more specialised scaling factors, as similar performance is achieved without significantly compromising accuracy.

4 Data Outliers

As discussed in the main manuscript, some of the *ab initio* harmonic frequencies calculated in VIBFREQ1295 differ by more than $\pm\,50\,\mathrm{cm}^{-1}$ from their experimental counterparts. The reason behind these large discrepancies is unclear to us, but we suspect that these frequencies correspond to highly coupled modes where the harmonic approximation struggles to provide reliable predictions. In Table S4 we present the frequencies with deviations larger than $\pm\,50\,\mathrm{cm}^{-1}$ from the experimental values, together with their corresponding molecule, frequency range location, and approximate mode description.

Table S4: Molecules with a frequency difference larger than $\pm\,50\,\mathrm{cm}^{-1}$ between the scaled harmonic and experimental fundamental frequencies. All frequencies are given in cm^{-1}.

Molecule	Exp Freq	Scaled Freq	FreqDiff	Freq Range	Molecule Class	Vibrational Mode Desc.
BH_3CO	2166.0	2111.4	-54.6	High	non-Halogen	C–O stretch + B–H stretch
^2BO	1772.9	1826.8	53.9	Mid	non-Halogen	B–O stretch
	578.0	520.9	-57.1	Low	non-Halogen	S–S symmetric stretch
	801.0	693.1	-107.9	Low	non-Halogen	CS_2 rock
	809.0	725.9	-83.1	Low	non-Halogen	C–S asymmetric stretch
	862.0	756.6	-105.4	Low	non-Halogen	CS_2 scissor
	948.0	839.6	-108.4	Low	non-Halogen	CH_2 rock
c-$C_2H_4S_3$	1024.0	868.1	-155.9	Low	non-Halogen	CH_2 rock
	1168.0	1066.4	-101.6	Mid	non-Halogen	CH_2 twist
	1173.0	1096.0	-77.0	Mid	non-Halogen	CH_2 twist
	1327.0	1167.9	-159.1	Mid	non-Halogen	CH_2 wag
	1355.0	1201.3	-153.7	Mid	non-Halogen	CH_2 wag
	1360.0	1414.9	54.9	Mid	non-Halogen	CH_2 scissor
c-C_3H_6S	113.0	163.9	50.9	Low	non-Halogen	Ring torsion
	663.0	732.5	69.5	Low	non-Halogen	Ring torsion
	925.0	978.8	53.8	Mid	non-Halogen	N–C–N scissor
	1002.0	1091.9	89.9	Mid	non-Halogen	C–H rock
c-CH_2N_4	1015.0	1129.4	114.4	Mid	non-Halogen	N–N stretch
	1084.0	1149.9	65.9	Mid	non-Halogen	N–N stretch
	3102.0	3157.6	55.6	High	non-Halogen	C–H stretch
	3447.0	3506.6	59.6	High	non-Halogen	N–H stretch
C_2H_3N	2266.5	2214.0	-53.0	High	non-Halogen	C≡N symmetric stretch
$C_2H_3N_1$	2161.0	2110.2	-50.8	High	non-Halogen	N≡C stretch
C_2H_4O	2715.8	2798.1	82.3	High	non-Halogen	C–H stretch
$C_2H_4O_2_1$	534.0	416.7	-117.3	Low	non-Halogen	$C_2H_4O_2$ bend

Continued on next page

Table S4 – *Continued from previous page*

	642.0	580.7	-61.3	Low	non-Halogen	CO_2 scissor
	1264.0	1318.1	54.1	Mid	non-Halogen	$C_2H_5O_2$ deformation
C_2HF	2239.2	2187.1	-52.1	High	Halogen	C≡C stretch
C_2N_2	2157.8	2103.0	-54.8	High	non-Halogen	C≡N asymmetric stretch
	2330.5	2280.2	-50.3	High	non-Halogen	C≡N symmetric stretch
C_3H_3Cl	2147.0	2094.2	-52.8	High	Halogen	C≡C stretch
C_3H_3F	2150.0	2093.9	-56.1	High	Halogen	C≡C stretch
C_3H_3N	2240.0	2187.9	-52.1	High	non-Halogen	C≡N stretch
C_3H_8	2961.7	2904.9	-56.8	High	non-Halogen	C–H asymmetric stretch
C_3O_2	787.7	841.1	53.4	Low	non-Halogen	C–C stretch
	2196.9	2146.0	-50.9	High	non-Halogen	C–O symmetric stretch
	2289.8	2214.8	-75.0	High	non-Halogen	C–O asymmetric stretch
C_4H_2	482.7	330.9	-151.8	Low	non-Halogen	C_4H_2 bend
C_4N_2	620.0	471.8	-148.2	Low	non-Halogen	Deformation
	2245.4	2193.8	-51.5	High	non-Halogen	C≡N asymmetric stretch
C_6H_8	170.0	92.6	-77.4	Low	non-Halogen	Torsion
	217.0	274.6	57.6	Low	non-Halogen	Torsion
	355.0	294.0	-61.0	Low	non-Halogen	C_6H_8 deformation
	444.0	496.2	52.2	Low	non-Halogen	C_6H_8 deformation
	590.0	663.3	73.3	Low	non-Halogen	C_6H_8 deformation
	683.0	759.2	76.2	Low	non-Halogen	CH_2 twist
	872.0	800.1	-71.9	Low	non-Halogen	C–H wag
	901.0	849.4	-51.6	Low	non-Halogen	CH_2 rock
	1186.0	1037.7	-148.3	Mid	non-Halogen	C–C symmetric stretch
	1192.0	1076.0	-116.0	Mid	non-Halogen	C–C asymmetric stretch
	1295.0	1387.4	92.4	Mid	non-Halogen	CH rock
	1579.0	1631.1	52.1	Mid	non-Halogen	C–H rock
CH_2N_2	408.9	354.8	-54.0	Low	non-Halogen	CH_2N_2 deformation
CH_3NO	289.0	103.2	-185.7	Low	non-Halogen	NH_2 wag
CH_4O	199.8	286.2	86.4	Low	non-Halogen	Torsion
CH_5N	779.6	844.1	64.4	Low	non-Halogen	NH_2 wag
	1044.8	966.0	-78.8	Low	non-Halogen	CH_5N deformation
	1130.0	1043.1	-86.9	Mid	non-Halogen	C–N stretch
	1430.0	1323.6	-106.4	Mid	non-Halogen	CH_5N deformation
	2820.0	2875.8	55.8	High	non-Halogen	C–H symmetric stretch
ClCN	2249.3	2161.5	-87.8	High	Halogen	C≡N stretch
$ClNO_2$	370.2	261.6	-108.6	Low	Halogen	Deformation
	410.0	350.2	-59.8	Low	Halogen	O–Cl stretch
	651.7	451.6	-200.1	Low	Halogen	NO_2 bend
	792.8	670.3	-122.4	Low	Halogen	N–O stretch + C–Cl stretch
	1267.3	867.0	-400.3	Low	Halogen	NO_2 scissor
^2CN	2042.4	1977.3	-65.1	High	non-Halogen	C–N stretch
CO	2169.8	2075.6	-94.2	High	non-Halogen	C≡O stretch
CO_2	2349.1	2295.6	-53.5	High	non-Halogen	CO_2 asymmetric stretch
FCN	2318.8	2259.6	-59.2	High	Halogen	FCN asymmetric stretch

Continued on next page

Table S4 – *Continued from previous page*

HCl	2990.9	2880.0	-111.0	High	Halogen	H–Cl stretch
HCN	2096.8	2039.6	-57.3	High	non-Halogen	C≡N stretch
^{2}HCO	2434.5	2579.2	144.7	High	non-Halogen	HCO stretch
HNCO	2268.9	2209.5	-59.4	High	non-Halogen	NCO asymmetric stretch
HNO	2684.0	2833.2	149.2	High	non-Halogen	N–H stretch
$^2HO_2$	3436.2	3500.0	63.8	High	non-Halogen	O–H stretch
N_2H_4	780.0	840.8	60.8	Low	non-Halogen	NH_2 wag
	3350.0	3432.6	82.6	High	non-Halogen	N–H asymmetric stretch
^{2}NH	3125.6	3196.0	70.4	High	non-Halogen	N–H stretch
NH_3	968.1	1031.7	63.6	Mid	non-Halogen	NH_3 bend
O_3	1042.1	1161.5	119.4	Mid	non-Halogen	O–O asymmetric stretch
	1103.1	1165.8	62.6	Mid	non-Halogen	O–O symmetric stretch
OCS	2062.2	2008.7	-53.5	High	non-Halogen	OCS asymmetric stretch
$^3CH_2$	963.1	1072.2	109.1	Mid	non-Halogen	CH_2 stretch

5 Supplementary Files

To ensure the accessibility and straightforward use of the data, the VIBFREQ1295 database can be found as a csv file tabulated with the following columns:

1. Molecule.
2. Molecular Formula.
3. Molecule classification, i.e., CHNOPS, Halogens, or Others.
4. Total number of atoms.
5. Total number of non-hydrogen atoms.
6. IUPAC Name.
7. SMILES identifier.
8. InChi code.
9. InChi key.
10. Citation key to the parent vibrational frequency databases (see text below for references).
11. Parent database description, i.e., No. of frequencies/No. of molecules.
12. Vibrational mode.
13. Updated experimental fundamental frequencies.
14. CCSD(T)-F12c/cc-pVDZ-F12 raw harmonic frequencies.
15. CCSD(T)-F12c/cc-pVDZ-F12 scaled harmonic frequencies using frequency-range-specific scaling factors, i.e., Low: less than 1000 cm^{-1}, Mid: between 1000 and 2000 cm^{-1}, and High: above 2000 cm^{-1}.
16. Frequency range.
17. Frequency region covered.
18. Symmetry of the vibrational mode.
19. Vibrational mode description (general).
20. Vibrational mode description (detailed).
21. Vibrational mode class, i.e., stretch or non-stretch.
22. Uncertainty in the reported experimental frequencies.

23. Uncertainty description.
24. Resolution of the experimental measurement, i.e., low-to-mid or rovibronic resolution.
25. Journal of publication.
26. Year of publication.
27. BibTex code for the original experimental publication.
28. DOI to the original experimental publication.
29. Comments.

All molecules included into VIBFREQ1295 are linked back to their parent databases through their corresponding references.

Table S5: Parent databases used in the development of VIBFREQ1295. References to the original publications are provided for all databases.

Database Alias	No. of Frequencies	No. of Molecules	Reference
04WiMo	1306	65	(12)
93PoScWo	1064	122	(13)
21UnNaOz	913	157	(14)
93HeHo	510	42	(15)
15KeBrMa	119	30	(16)
10AlZhZh	60	15	(17)

6 Sample Input File

We ran all our quantum-chemistry harmonic frequency calculations using the CCSD(T)-F12c/cc-pVDZ-F12 model chemistry, using the MOLPRO quantum chemistry package. Below is the input file template used in the calculations.

```
***, {molecule} harmonic frequencies
basis=cc-pVDZ-F12
orient,mass
symm=auto
geometry={
molecular geometry
}
mass,iso

hf
ccsd(t)-f12c
optg,gradient=0.000015,grms=0.000010,step=0.000060,srms=0.000040

frequencies,symm=no
---
```

References

[1] A. P. Scott and L. Radom, "Harmonic vibrational frequencies: An evaluation of Hartree-Fock, Møller-Plesset, quadratic configuration interaction, density functional theory, and semiempirical scale factors," *Journal of Physical Chemistry*, vol. 100, no. 41, pp. 16502–16513, 1996.

[2] P. Sinha, S. E. Boesch, C. Gu, R. A. Wheeler, and A. K. Wilson, "Harmonic vibrational frequencies: Scaling factors for HF, B3LYP, and MP2 methods in combination with correlation consistent basis sets," *Journal of Physical Chemistry A*, vol. 108, no. 42, pp. 9213–9217, 2004.

[3] P. Pulay, G. Fogarasi, G. Pongor, J. E. Boggs, and A. Vargha, "Combination of Theoretical ab Initio and Experimental Information to Obtain Reliable Harmonic Force Constants. Scaled Quantum Mechanical (SQM) Force Fields for Glyoxal, Acrolein, Butadiene, Formaldehyde, and Ethylene," *Journal of the American Chemical Society*, vol. 105, no. 24, pp. 7037–7047, 1983.

[4] G. Rauhut and P. Pulay, "Transferable scaling factors for density functional derived vibrational force fields," *Journal of Physical Chemistry*, vol. 99, no. 10, pp. 3093–3100, 1995.

[5] J. Baker, A. A. Jarzecki, and P. Pulay, "Direct scaling of primitive valence force constants: An alternative approach to scaled quantum mechanical force fields," *Journal of Physical Chemistry A*, vol. 102, no. 8, pp. 1412–1424, 1998.

[6] Z. A. Fekete, E. A. Hoffmann, T. KörtvÌlyesi, and B. Penke, "Harmonic vibrational frequency scaling factors for the new NDDO Hamiltonians: RM1 and PM6," *Molecular Physics*, vol. 105, no. 19-22, pp. 2597–2605, 2007.

[7] P. Borowski, M. Fernández-Gómez, M. P. Fernández-Liencres, and T. P. Ruiz, "An effective scaling frequency factor method for scaling of harmonic vibrational frequencies: theory and preliminary application to toluene," *Chemical Physics Letters*, vol. 446, no. 1-3, pp. 191–198, 2007.

[8] P. Borowski, A. Drzewiecka, M. Fernández-Gómez, M. P. Fernández-Liencres, and T. P. Ruiz, "An effective scaling frequency factor method for harmonic vibrational frequencies: The factors' transferability problem," *Chemical Physics Letters*, vol. 465, pp. 290–294, 11 2008.

[9] P. Borowski, "An effective scaling frequency factor method for scaling of harmonic vibrational frequencies: The use of redundant primitive coordinates," *Journal of Molecular Spectroscopy*, vol. 264, pp. 66–74, 11 2010.

[10] P. Borowski, A. Drzewiecka, M. Fernández-Gómez, M. P. Fernández-Liencres, and T. P. Ruiz, "A new, reduced set of scaling factors for both SQM and ESFF calculations," *Vibrational Spectroscopy*, vol. 52, pp. 16–21, 1 2010.

[11] J. C. Zapata Trujillo and L. K. McKemmish, "Metaâanalysis of uniform scaling factors for harmonic frequency calculations," *WIREs Computational Molecular Science*, p. e1584, 10 2021.

[12] H. A. Witek and K. Morokuma, "Systematic study of vibrational frequencies calculated with the self-consistent charge density functional tight-binding method," *Journal of Computational Chemistry*, vol. 25, no. 15, pp. 1858–1864, 2004.

[13] J. A. Pople, A. P. Scott, M. W. Wong, and L. Radom, "Scaling Factors for Obtaining Fundamental Vibrational Frequencies and ZeroâPoint Energies from HF/6â31G* and MP2/6â31G* Harmonic Frequencies," *Israel Journal of Chemistry*, vol. 33, no. 3, pp. 345–350, 1993.

[14] Y. Ünal, W. Nassif, B. C. Özaydin, and K. Sayin, "Scale factor database for the vibration frequencies calculated in M06-2X, one of the DFT methods," *Vibrational Spectroscopy*, vol. 112, p. 103189, 1 2021.

[15] E. F. Healy and A. Holder, "An evaluation of AM1 calculated vibrational frequencies," *Journal of Molecular Structure: THEOCHEM*, vol. 281, pp. 141–156, 4 1993.

[16] M. K. Kesharwani, B. Brauer, and J. M. Martin, "Frequency and zero-point vibrational energy scale factors for double-hybrid density functionals (and other selected methods): Can anharmonic force fields be avoided?," *Journal of Physical Chemistry A*, vol. 119, no. 9, pp. 1701–1714, 2015.

[17] I. M. Alecu, J. Zheng, Y. Zhao, and D. G. Truhlar, "Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries," *Journal of Chemical Theory and Computation*, vol. 6, no. 9, pp. 2872–2887, 2010.