A rotational investigation of the three isomeric forms of cyanoethynylbenzene ($HCC-C_6H_4-CN$): Benchmarking experiments and calculations using the "Lego Brick" approach

Electronic Supplementary Information

Jean-Thibaut Spaniol, Kin Long Kelvin Lee, Olivier Pirali, Cristina Puzzarini, and Marie-Aline Martin-Drumel

Contents

List of Tables

1 Supplementary Tables

Table S1: Rotational constants of o-, m-, and p-CEB in the ground vibrational state and comparison with predictions derived from the ωB97X-D/cc-pVQZ calculation.

	Exp. $v = 0$ /MHz	Pred.[a] /MHz	Diff.[b] /%	Pred.[c] /MHz	Diff.[b] /%
	o-CEB				
A	2027.441205 (98)	2002.00	1.27	2002.80	1.23
B	1329.456144 (39)	1329.71	−0.02	1335.70	−0.47
C	802.668552 (30)	799.01	0.46	802.17	0.06
	m-CEB				
A	2704.87469 (15)	2697.84	0.26	2698.91	0.22
B	907.033894 (26)	903.09	0.44	907.15	−0.01
C	679.066023 (24)	676.60	0.36	679.27	−0.03
	p-CEB				
A	5646.3543 (52)	5647.85	−0.03	5650.09	−0.07
B	709.161497 (30)	706.09	0.43	709.27	−0.02
C	630.029790 (29)	627.63	0.38	630.11	−0.01

[a] Results of the ωB97X-D/cc-pVQZ calculation scaled by the mean Bayesian scaling correction ($0.9866 \times A(BC)_e$) as proposed in Lee and McCarthy [2020] and reported in Buchanan et al. [2021] (see Table S6 of that reference)

[b] (Exp.-Calc.)/Calc.×100

[c] Values scaled using semi-empirical scaling factors derived as the ratio of the PhC$_3$N equilibrium values and the experimentally determined parameters reported in Buchanan et al. [2021] (see Table S6 of that reference)

Table S2: Alpha values of o-, m-, and p-CEB in singly excited vibrational states lying below $250\,\mathrm{cm}^{-1}$. Calculated values obtained at the B3LYP-D3/jun-cc-pVDZ level are compared with experimental values. All values in MHz.

		o-CEB		m-CEB		p-CEB	
		Calc.	Exp.	Calc.	Exp.	Calc.	Exp.
$v_{38} = 1$	α_A	13.33	13.22	17.45	19.93	−3.57	
	α_B	−1.27	−1.27	−0.65	−0.69	−0.71	
	α_C	−0.79	−0.88	−0.58	−0.61	−0.35	
$v_{26} = 1$	α_A	−14.84	−15.52	−19.83	−22.58	99.09	104.60
	α_B	−2.36	−2.37	−1.23	−1.28	−0.47	−0.47
	α_C	−0.42	−0.49	−0.36	−0.40	−0.72	−0.76
$v_{39} = 1$	α_A	−0.27		−0.72		−95.11	−100.38
	α_B	1.03		0.08		−0.63	−0.66
	α_C	−0.64		−0.65		−0.27	−0.30
$v_{27} = 1$	α_A	−3.87		−5.62			
	α_B	−1.07		−0.83			
	α_C	−0.08		−0.23			
$v_{25} = 1$	α_A					3.99	
	α_B					−0.62	
	α_C					−0.64	

Table S3: Tentative spectroscopic constants for m-CEB in excited vibrational states. Values in parentheses are errors (1σ) in the unit of the last digit. Parameters without an error are fixed (to the value in $v = 0$). For both vibrational states, lower K_a'' values have been assigned but are not included in the fit (too perturbed).

Parameter[a]		$v_{27} = 1$			$v_{39} = 1$		
		Exp. /MHz	Pred.[b] /MHz	Diff.[c] /%	Exp. /MHz	Pred.[b] /MHz	Diff.[c] /%
E			111[d]			119[d]	
A		2651.61 (30)	2710.49	2.17	2771.69 (83)	2705.59	2.39
B		908.417 (37)	907.86	0.06	906.218 (92)	906.95	0.08
C		679.6571 (35)	679.30	0.05	679.5889 (53)	679.72	0.02
Δ_J	$\times 10^3$	0.05737 (25)			0.02318 (37)		
Δ_{JK}	$\times 10^3$	1.079 (23)			-1.763 (38)		
Δ_K	$\times 10^3$	-27.26 (37)			35.98 (71)		
δ_J	$\times 10^3$	0.01709415			0.01709415		
δ_K	$\times 10^3$	0.0593386			0.0593386		
Φ_J	$\times 10^9$	-0.2285 (91)			0.225 (13)		
Φ_{JK}	$\times 10^9$	24.43 (82)			-28.3 (15)		
Φ_{KJ}	$\times 10^9$	46. (17)			173. (33)		
Φ_K	$\times 10^9$	$-4615.$ (133)			3694. (236)		
ϕ_J	$\times 10^9$	0.011514			0.011514		
ϕ_{JK}	$\times 10^9$	2.1395			2.1395		
N/n[e]		741/417			476/375		
$J''_{\max}, K''_{a\,\min} - K''_{a\,\max}$		$150, 10 - 19$			$151, 9 - 19$		
RMS[f]		0.132			0.175		
σ[g]		1.76			2.34		

[a] In MHz, unless otherwise noted

[b] For $v_i = 1$, Pred.=Exp.$(v = 0) - \alpha_{i\,\text{Calc.}}$; for $(v_{26}, v_{39}) = (1, 1)$, the experimental αs have been used

[c] $|$Exp.-Calc.$|$/Calc.$\times 100$

[d] in cm^{-1}

[e] Number of lines (N) and number of lines with different frequencies (n), unitless

[f] Root mean square, in MHz

[g] Weighted standard deviation, unitless

Table S4: Tentative spectroscopic constants for p-CEB in excited vibrational states. Values in parentheses are errors (1σ) in the unit of the last digit. Parameters without an error are fixed (to the value in $v = 0$).

Parameter[a]		$v_{38} = 1$			$v_{25} = 1$			$(v_{26}, v_{39}) = (1,1)$		
		Exp. /MHz	Pred.[b] /MHz	Diff.[c] /%	Exp. /MHz	Pred.[b] /MHz	Diff.[c] /%	Exp. /MHz	Pred.[b] /MHz	Diff.[c] /%
E			184[d]			187[d]			197[d]	
A		5632.392 (94)	5649.93	0.31	5627.069 (32)	5642.37	0.27	5683.4 (23)	5642.13	0.73
B		710.27945 (48)	709.87	0.06	709.49007 (36)	709.78	0.04	709.435 (38)	710.30	0.12
C		631.10601 (59)	630.38	0.12	630.437041 (69)	630.67	0.04	631.11063 (16)	631.09	< 0.01
Δ_J	$\times 10^3$	0.008360 (32)			0.007930 (35)			0.0201 (10)		
Δ_{JK}	$\times 10^3$	0.1311 (18)			0.1544 (72)			−1.45 (14)		
Δ_K	$\times 10^3$	8.38 (54)			1.134			−36.8 (43)		
δ_J	$\times 10^3$	0.0011493 (94)			0.001056 (18)			0.00706 (51)		
δ_K	$\times 10^3$	0.2137 (86)			0.1193 (18)			0.11990		
Φ_{JK}	$\times 10^9$	−1.283 (47)			0.05162			0.05162		
Φ_{KJ}	$\times 10^9$	122.3 (72)			−0.9319			−0.9319		
N/n^e		1069/870			601/356			352/204		
$J''_{\max}, K''_{a\,\min} - K''_{a\,\max}$		161, 2 − 17			173, 0 − 7			172, 0 − 7		
RMS ^{f}		0.168			0.060			0.114		
σ^g		1.68			0.60			1.14		

^{a} In MHz, unless otherwise noted
^{b} For $v_i = 1$, Pred.=Exp.$(v = 0)-\alpha_i$ Calc.; for $(v_{26}, v_{39}) = (1,1)$, the experimental αs have been used
^{c} |Exp.-Calc.|/Calc.$\times 100$
^{d} in cm^{-1}
^{e}Number of lines (N) and number of lines with different frequencies (n), unitless
^{f} Root mean square, in MHz
^{g} Weighted standard deviation, unitless

Table S5: Spectroscopic constants of p-CEB in excited vibrational states involving more than one quantum of excitation. Values in parentheses are errors (1σ) in the unit of the last digit.

Parameter[a]	$v_{26}=2$ Exp. /MHz	Pred.[b] /MHz	Diff.[c] /%	$v_{26}=3$ Exp. /MHz	Pred.[b] /MHz	Diff.[c] /%	$v_{39}=2$ Exp. /MHz	Pred.[b] /MHz	Diff.[c] /%
E		154[d]			231[d]			240[d]	
A	5442.5356 (82)	5437.16	0.10	5348.297 (15)	5343.31	0.09	5851.400 (14)	5847.11	0.07
B	710.111301 (51)	710.11	< 0.01	710.587003 (90)	710.59	< 0.01	710.484959 (84)	710.49	< 0.01
C	631.550239 (44)	631.55	< 0.01	632.309989 (75)	632.31	< 0.01	630.623169 (69)	630.62	< 0.01
$\Delta_J \times10^3$	0.00817097 (89)			0.0082832 (15)			0.0080541 (14)		
$\Delta_{JK} \times10^3$	0.142646 (43)			0.142857 (68)			0.142184 (89)		
$\Delta_K \times10^3$	−14.634 (29)			−20.199 (46)			20.409 (53)		
$\delta_J \times10^3$	0.00103304 (49)			0.00101290 (82)			0.00112863 (82)		
$\delta_K \times10^3$	−0.06944 (19)			−0.14847 (27)			0.31733 (25)		
$\Phi_{JK} \times10^9$	0.10763 (90)			0.1336 (19)			−0.0169 (19)		
$\Phi_{KJ} \times10^9$	−0.307 (25)						−1.370 (46)		
N/n[e]	4657/2926			3664/2314			3189/2050		
$J''_{\max}, K''_{a\,\max}$	171, 41			171, 40			173, 40		
RMS[f]	0.053			0.081			0.070		
σ[g]	1.06			1.08			0.94		

[a] In MHz, unless otherwise noted
[b] For $v_i=2$, Pred.=Exp.$(v=0)$ $-2\times\alpha_{v_i=1\,\text{Exp.}}$; For $v_i=3$, Pred.=Exp.$(v=0)$ $-\alpha_{v_i=1\,\text{Exp.}}$ $-2\times\alpha_{v_i=2\,\text{Exp.}}$.
[c] |Exp.-Calc.|/Calc.×100
[d] in cm^{-1}
[e] Number of lines (N) and number of lines with different frequencies (n), unitless
[f] Root mean square, in MHz
[g] Weighted standard deviation, unitless

2 Computational insights

2.1 "Lego Brick" approach: details and results

The TM+LR approach —where, for a given molecule, suitable molecular fragments are seen as "Lego bricks"— is set-up as follows:

- For the "Lego bricks", we resort to the TM approach, which –as mentioned in the main text– is based on identifying, within the molecule, "known" fragments that belong to a smaller system for which an highly accurate equilibrium geometry is available. These fragments are the TMs and are used to determine accurately the corresponding structural parameters of the larger molecular systems ($r_e(\text{lms})$):

$$r_e(\text{lms}) = r_e^{low-cost}(\text{lms}) + \Delta r_e(\text{TM}) \tag{1}$$

  where $\Delta r_e(\text{TM})$ is defined as

$$\Delta r_e(\text{TM}) = r_e^{\text{accu}}(\text{TM}) - r_e^{low-cost}(\text{TM}) \tag{2}$$

  $r_e^{low-cost}$ is the geometrical parameter of interest calculated at the same level of theory for both the molecule under consideration (lms) and the TM. The chosen level for $r_e^{low-cost}(\text{lms})$ and $r_e^{low-cost}(\text{TM})$ of Eqs. 1 and 2, respectively, is the double-hybrid rev-DSD-PBEP86 functional [Santra et al., 2019] in conjunction with the jun-cc-pVTZ basis set [Dunning, 1989, Papajak and Truhlar, 2011]. Hereafter, shortly denoted as revDSD. For $r_e^{\text{accu}}(\text{TM})$, we resort to semi-experimental equilibrium structures available in the literature or databases.

- Once we have the "Lego bricks", we need to put them together. For this task, we resort to the LR approach for improving the revDSD determinations of the inter-fragment parameters. For them, the LR approach replaces the $\Delta r_e(\text{TM})$ correction in Eq. 1 with an estimate ($\Delta r_e(\text{LR})$) based on a linear regression model:

$$\Delta r_e(\text{LR}) = A r_e^{low-cost} + B \tag{3}$$

  Therefore, the corrected parameter is given by:

$$r_e(\text{lms}) = (1 + A)r_e^{low-cost} + B \tag{4}$$

  The A and B coefficients only depend on the atomic numbers of the involved atoms and were obtained by a statistical analysis of a large number of molecules. In the present context, for the C-C bond length, we employ the corrective factors: A = -0.00184 and B = 0.0. These have been obtained from a study employing nearly 100 semi-experimental values.

Table S6: Computed and template values of the structural parameters of o-, m-, and p-CEB. Distances in Å and angles in degrees.

ZMAT		Parameters[a]	
	Name	revDSD[b]	TM (TM+LR)
o-CEB			
H	R1	1.0643	1.0614
C 1 R1	R2	1.2100	1.2074
X 2 1.5 1 90.	R3	1.4280	1.4280 (1.4253)
C 2 R2 3 A1 1 180.	R4	1.4093	1.4070
X 4 1.5 2 90. 3 0.	R5	1.3981	1.3958
C 4 R3 5 A2 3 180.	R6	1.3896	1.3873
C 6 R4 4 A3 5 180.	R7	1.3947	1.3924
C 7 R5 6 A4 4 180.	R8	1.3900	1.3877
C 8 R6 7 A5 6 0.	R9	1.0825	1.0787
C 9 R7 8 A6 7 0.	R10	1.0831	1.0793
C 10 R8 9 A7 8 0.	R11	1.0828	1.0790
H 11 R9 10 A8 9 180.	R12	1.0826	1.0788
H 10 R10 9 A9 8 180.	R13	1.4343	1.4343 (1.4316)
H 9 R11 8 A10 7 180.	R14	1.1630	1.1588
H 8 R12 7 A11 6 180.	A1	90.62	90.62
C 7 R13 6 A12 4 0.	A2	90.22	90.22
X 16 1.5 7 90. 6 180.	A3	120.68	120.68
N 16 R14 17 A13 7 180.	A4	120.42	120.42
	A5	120.03	120.03
	A6	119.91	119.91
	A7	120.35	120.35
	A8	120.63	120.63
	A9	120.04	120.04
	A10	119.81	119.81
	A11	119.17	119.17
	A12	120.18	120.18
	A13	88.73	88.73
m-CEB			
H	R1	1.0642	1.0614
C 1 R1	R2	1.2102	1.2076
X 2 1.5 1 90.	R3	1.4313	1.4313 (1.4286)
C 2 R2 3 A1 1 180.	R4	1.3986	1.3963
X 4 1.5 2 90. 1 180.	R5	1.0822	1.0785
C 4 R3 5 A2 3 180.	R6	1.3964	1.3941
C 6 R4 4 A3 5 180.	R7	1.3994	1.3971
H 7 R5 6 A4 4 0.	R8	1.3907	1.3884
C 7 R6 6 A5 4 180.	R9	1.3910	1.3887
C 9 R7 7 A6 8 180.	R10	1.0828	1.0790
C 10 R8 9 A7 7 0.	R11	1.0828	1.0790
C 11 R9 10 A8 9 0.	R12	1.0825	1.0787
H 12 R10 6 A9 4 0.	R13	1.4360	1.4360 (1.4333)
H 11 R11 12 A10 6 180.	R14	1.1628	1.1586
H 10 R12 11 A11 12 180.	A1	90.11	90.11
C 9 R13 7 A12 8 0.	A2	89.91	89.91
X 16 1.5 9 90. 7 0.	A3	120.12	120.12
N 16 R14 17 A13 9 180.	A4	120.05	120.05
	A5	119.78	119.78
	A6	120.63	120.63
	A7	119.37	119.37
	A8	120.42	120.42
	A9	119.17	119.17
	A10	119.81	119.81
	A11	120.91	120.91
	A12	119.58	119.58
	A13	90.19	90.19
			(continued)

ZMAT	Parameters[a]		
	Name	revDSD[b]	TM (TM+LR)
p-CEB			
X	R1	1.2926	1.2881
X 1 1.5	R2	1.4304	1.4304 (1.4277)
C 1 1.5 2 90.	R3	1.2104	1.2078
C 1 R1 2 90. 3 180.	R4	1.4347	1.4347 (1.4320)
X 4 1.5 1 90. 2 0.	R5	1.4023	1.4001
X 3 1.5 1 90. 2 0.	R6	1.3997	1.3974
C 3 R2 6 90. 2 180.	R7	1.0825	1.0787
X 7 1.5 3 90. 6 90.	R8	1.0826	1.0788
C 7 R3 8 90. 3 180.	R9	1.0642	1.0614
X 9 1.5 7 90. 8 0.	R10	1.1631	1.1589
C 4 R4 5 90. 2 180.	A1	59.72	59.72
C 3 R5 1 A1 2 TM90	A2	60.08	60.08
C 3 R5 1 A1 2 90.	A3	119.38	119.38
C 4 R6 1 A2 2 90.	A4	119.70	119.70
C 4 R6 1 A2 2 -90.			
H 12 R7 3 A3 1 180.			
H 13 R7 3 A3 1 180.			
H 14 R8 4 A4 1 180.			
H 15 R8 4 A4 1 180.			
H 9 R9 10 90. 8 180.			
X 11 1.5 4 90. 5 90.			
N 11 R10 21 90. 4 180.			

[a] TM and TM+LR values differ only for the inter-fragment parameters. Therefore, for the latter, the TM+LR values are reported in parentheses.

[b] revDSD: revDSD-PBEP86-D3(BJ)/jun-cc-pVTZ.

2.2 Quartic centrifugal distortion constants

In addition to the quartic centrifugal distortion constants at the fc-MP2/cc-pVTZ level purposely computed for this work, the calculations at the B3LYP-D3/jun-cc-pVDZ level carried out in Melli et al. [2021] can provide these parameters as well. In Table S7, the experimental values are compared to the computed MP2 and B3LYP results. It is interesting to note that, for m-CEB and p-CEB, B3LYP provides reliable values in qualitative agreement with experiment. Instead, for o-CEB, most of the parameters are overestimated by one order of magnitude (or more in the case of δ_K). In order to investigate whether such a disagreement is related to the level of theory, quartic centrifugal distortion constants of o-CEB have been computed at the fc-MP2 level with the cc-pVDZ basis set using both Gaussian 16 and CFour. Interestingly, the results from CFour are very similar to those obtained with the larger basis set (i.e. those reported in Table S7), while the results from Gaussian 16 show the same disagreement with experiment as B3LYP computations. This suggests the presence of a bug in Gaussian 16.

Table S7: Quartic centrifugal distortion constants of o-, m-, and p-CEB in the ground vibrational state: comparison of experiments with computed values.

Parameter		o-CEB			m-CEB			p-CEB		
		Exp.[a]	MP2[b]	B3LYP[b]	Exp.[a]	MP2[b]	B3LYP[c]	Exp.[a]	MP2[b]	B3LYP[c]
Δ_J	$\times 10^3$	0.0722	0.0715	0.7267	0.0420	0.0418	0.0387	0.0080	0.0077	0.0073
Δ_{JK}	$\times 10^3$	-0.3998	-0.4087	-2.3471	-0.4342	-0.4416	-0.4013	0.1454	0.1462	0.1342
Δ_K	$\times 10^3$	1.8159	1.8992	1.6310	2.6319	2.6584	2.4354	1.134	1.0785	1.0424
δ_J	$\times 10^3$	0.0302	0.0299	0.3017	0.0171	0.0171	0.0158	0.0011	0.0011	0.0010
δ_K	$\times 10^3$	-0.0630	-0.0682	-1.4980	0.0593	0.0562	0.0534	0.1199	0.1201	0.1120

[a] Results truncated to four decimals. Results with all significant decimals and corresponding errors are avaialble in the main text. [b] Level of theory: fc-MP2/cc-pVTZ (CFour). [c] Level of theory: B3LYP-D3/jun-cc-pVDZ (Gaussian 16)

References

Z. Buchanan, K. L. K. Lee, O. Chitarra, M. C. McCarthy, O. Pirali, and M.-A. Martin-Drumel. A rotational and vibrational investigation of phenylpropiolonitrile ($C_6H_5C_3N$). *J. Mol. Spectrosc.*, 377:111425, Mar. 2021. doi:10.1016/j.jms.2021.111425.

T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90:1007–1023, 1989. doi:10.1063/1.456153.

K. L. K. Lee and M. McCarthy. Bayesian Analysis of Theoretical Rotational Constants from Low-Cost Electronic Structure Methods. *J. Phys. Chem. A*, 124(5):898–910, Feb. 2020. doi:10.1021/acs.jpca.9b09982.

A. Melli, F. Tonolo, V. Barone, and C. Puzzarini. Extending the Applicability of the Semi-experimental Approach by Means of "Template Molecule" and "Linear Regression" Models on Top of DFT Computations. *J. Phys. Chem. A*, 125(45):9904–9916, Nov. 2021. doi:10.1021/acs.jpca.1c07828.

E. Papajak and D. G. Truhlar. Convergent partially augmented basis sets for post-Hartree-Fock calculations of molecular properties and reaction barrier heights. *J. Chem. Theory Comput.*, 7:10–18, 2011. doi:10.1021/ct1005533.

G. Santra, N. Sylvetsky, and J. M. Martin. Minimally empirical double-hybrid functionals trained against the GMTKN55 database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4. *J. Phys. Chem. A*, 123(24):5129–5143, 2019. doi:10.1021/acs.jpca.9b03157.