SUPPORTING INFORMATION

FOR

In silico Collision Cross Section (CCS) Calculations to Aid Metabolite Annotation

Susanta Das[a], Kiyoto Aramis Tanemura[a], Laleh Dinpazhoh[a], Mithony Keng[a], Christina Schumm[a], Lydia Leahy[a], Carter K Asef[c], Markace Rainey[c], Arthur S. Edison[b], Facundo M. Fernández[c,d], Kenneth M. Merz, Jr.[a,*]

[a]Department of Chemistry, Michigan State University, 578 S. Shaw Lane, East Lansing, Michigan 48824, United States
[b]Departments of Genetics and Biochemistry, Institute of Bioinformatics and Complex Carbohydrate Center, University of Georgia, 315 Riverbend Rd, Athens, GA 30602, United States
[c]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States
[d]Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, Georgia 30332, United States
*E-mail: merzjrke@msu.edu

Table S 1. Electrospray positive instrument settings. Parameters not included were left at their automatic values.

ES+	
Capillary (kV)	1.8
Sampling Cone	30
Source Offset	40
Source Temperature (°C)	80
Desolvation Temperature (°C)	250
Cone Gas (L/h)	50
Desolvation Gas (L/h)	650
Nebuliser (Bar)	7.0
IMS Gas flow (mL/min)	60
IMS wave height (V)	18
IMS Wave Velocity (m/s)	500-200 ramped
Stepwave 1 Height (V)	5
Stepwave 2 Height (V)	5
Stepwave 2 Offset	15
Stepwave RF	150.0
Ion Guide RF	200.0
Trap Entrance	5.0
Trap Bias	25.0
IMS Entrance	5.0
Helium Cell DC	25.0
Helium exit	-15.0
IMS Bias	15.0

Table S 2. Electrospray negative instrument settings. Parameters not included were left at their automatic values.

ES-	
Capillary (kV)	2.0
Sampling Cone	40
Source Offset	40
Source Temperature (°C)	80
Desolvation Temperature (°C)	250
Cone Gas (L/h)	50
Desolvation Gas (L/h)	650
Nebuliser (Bar)	7.0
IMS Gas flow (mL/min)	60
IMS wave height (V)	18
IMS Wave Velocity (m/s)	500-200 ramped
Stepwave 1 Height (V)	5
Stepwave 2 Height (V)	5
Stepwave 2 Offset	15
Stepwave RF	150.0
Ion Guide RF	200.0
Trap Entrance	2.0
Trap Bias	30.0
IMS Entrance	5.0
Helium Cell DC	30.0
Helium exit	-15.0
IMS Bias	3.0

Table S3. Boltzmann Weighted CCS values and % error in the *in silico* predicted CCS value as compared to the literature. N_2 gas was used as the drift gas. Errors in brackets are calculated between *in-silico* values and TWIMS experimental values (this study).

no.	Compound Name	Experimental CCS (Literature)	Experimental CCS (this work)	Molecular State	Boltzmann Weighted CCS	Error (%)
1	Carnosine	152.18	147.9	Model 1	137.15	10.96(7.84)
				Model 2	166.54	8.62(11.19)
				Model 3	163.57	6.96(9.58)
				Model 4	150.21	1.31(1.54)
				Model 5	217.50	30.03(32.00)
2	L-Anserine	155.97	151.7	Model 1	141.80	9.99(6.98)
				Model 2	159.74	2.36(5.03)
				Model 3	163.96	4.88(7.48)
3	Abscisic Acid	160.56	165.0	Model 1	154.38	4.00(6.88)
				Model 2	162.86	1.41(1.31)
				Model 3	166.43	3.53(0.86)
4	O-Succinyl-L-homoserine	147.46	142.3	Model 1	134.40	9.72(5.88)
				Model 2	145.45	1.38(2.16)
5	L-Tyrosine	145.75	139.0	Model 1	118.68	22.81(17.12)
				Model 2	148.53	1.87(6.41)
6	L-Citrulline	139.51	131.2	Model 1	122.85	13.56(6.80)
				Model 2	141.86	1.66(7.51)
				Model 3	141.96	1.72(7.58)
7	Quinolinic Acid	138.98	131.1	Model 1	105.66	31.53(24.08)
				Model 2	142.14	2.22(7.77)
8	Nicotinic Acid	128.35	126.6	Model 1	90.78	41.39(39.46)
				Model 2	132.17	2.73(4.21)
9	Guanidinoacetic Acid	126.88	127.2	Model 1	90.15	40.75(41.10)
				Model 2	131.49	3.51(3.26)
				Model 3	131.10	3.22 (2.97)
				Model 4	130.21	2.56(2.31)
10	Citramalic Acid	124.87	117.7	Model 1	98.98	26.16(18.92)
				Model 2	122.52	1.92(3.93)
				Model 3	119.62	4.39(1.61)
				Model 4	121.58	2.70(3.19)
11	N-Methyl-L-glutamate	133.71	130.0	Model 1	105.34	26.93(23.41)
				Model 2	129.30	3.41(0.54)
12	Serotonin	151.9	144.2	Model 1	131.36	15.64(9.77)
				Model 2	165.04	7.96(12.63)
				Model 3	161.60	6.00(10.77)
13	L-Mimosine	143.47	143.1	Model 1	127.93	12.15(11.86)
				Model 2	150.90	4.92(5.17)
				Model 3	145.36	1.30(1.55)
14	L-Tryptophan	154.22	145.8	Model 1	136.84	12.70(6.55)
				Model 2	161.55	4.54(9.75)
				Model 3	159.69	3.42(8.70)
15	L-Ornithine	129.79	127.4	Model 1	110.09	17.90(15.73)
				Model 2	127.39	1.89(0.01)
				Model 3	126.63	2.50(0.61)
16	N,N-Dimethylglycine	123.90	127.2	Model 1	100.10	23.77(27.07)
				Model 2	118.19	4.83(7.63)
17	Kynurenine	151.07	144.2	Model 1	137.15	10.15(5.14)
				Model 2	146.94	2.81(1.86)
18	L-Asparagine	131.54	125.8	Model 1	111.86	17.59(12.46)
				Model 2	124.49	5.66(1.05)
				Model 3	128.93	2.03(2.43)
19	L-2-Aminoadipic Acid	131.54	131.6	Model 1	112.99	16.42(16.47)
				Model 2	129.56	1.53(1.57)
20	Glutamine	133.45	127.9	Model 1	115.50	15.54(10.74)
				Model 2	129.92	2.72(1.55)

Table S 4. Conformation number, Relative energy, Mol fraction and the CCS value of Carnosine (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.36	0.22	139.18
2	4.57	0.00	141.14
3	6.54	0.00	143.40
4	5.39	0.00	139.95
5	0.00	0.41	133.40
6	5.47	0.00	138.82
7	7.97	0.00	148.68
8	13.47	0.00	140.07
9	6.29	0.00	140.21
10	2.50	0.01	137.78
11	9.90	0.00	140.89
12	8.87	0.00	141.81
13	3.70	0.00	144.85
14	0.15	0.32	140.57
15	8.58	0.00	145.71
16	1.45	0.04	136.67

Table S 5. Conformation number, Relative energy, Mol fraction and the CCS value of Carnosine (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	16.05	0.00	177.46
2	28.76	0.00	171.27
3	0.00	0.50	166.47
4	16.05	0.00	175.40
5	14.15	0.00	171.97
6	0.00	0.50	166.61
7	12.26	0.00	165.42
8	9.80	0.00	168.37
9	3.05	0.00	167.25
10	8.24	0.00	167.03
11	14.14	0.00	172.25
12	15.90	0.00	172.85
13	13.19	0.00	171.57

Table S 6. Conformation number, Relative energy, Mol fraction and the CCS value of Carnosine (Model 3).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	22.50	0.00	167.66
2	8.59	0.00	163.29
3	24.49	0.00	166.82
4	1.53	0.07	161.74
5	8.57	0.00	162.10
6	1.61	0.06	169.39
7	0.00	0.88	163.32
8	7.32	0.00	157.71
9	23.02	0.00	174.22
10	16.40	0.00	173.35
11	22.69	0.00	168.35
12	7.38	0.00	158.44
13	16.57	0.00	175.35
14	17.30	0.00	180.82
15	22.46	0.00	169.91

Table S7. Conformation number, Relative energy, Mol fraction and the CCS value of Carnosine (Model 5).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value ($Å^2$)
1	5.40	0.00	219.77
2	0.00	0.46	216.28
3	1.53	0.04	223.71
4	5.47	0.00	222.93
5	9.47	0.00	215.53
6	7.35	0.00	222.15
7	7.35	0.00	219.73
8	21.55	0.00	229.39
9	2.44	0.01	217.35
10	7.35	0.00	222.60
11	1.54	0.03	224.38
12	12.80	0.00	217.12
13	5.40	0.00	217.90
14	0.00	0.46	217.75

Table S8. Conformation number, Relative energy, Mol fraction and the CCS value of L-Anserine (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value ($Å^2$)
1	4.89	0.00	148.98
2	4.15	0.00	147.76
3	3.23	0.00	147.99
4	5.18	0.00	160.66
5	2.69	0.01	146.99
6	8.77	0.00	148.37
7	0.00	0.97	141.59
8	2.43	0.02	148.98

Table S9. Conformation number, Relative energy, Mol fraction and the CCS value of L-Anserine (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value ($Å^2$)
1	10.44	0.00	179.95
2	0.00	0.77	159.25
3	5.87	0.00	172.88
4	8.49	0.00	177.75
5	6.09	0.00	171.44
6	13.67	0.00	171.46
7	0.72	0.23	161.27
8	10.31	0.00	183.57
9	5.74	0.00	171.68
10	3.56	0.00	175.91

Table S10. Conformation number, Relative energy, Mol fraction and the CCS value of L-Anserine (Model 3).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value ($Å^2$)
1	10.44	0.00	179.95
2	0.00	0.77	163.25
3	5.87	0.00	172.88
4	8.49	0.00	177.75
5	6.09	0.00	171.44
6	13.67	0.00	171.46
7	0.72	0.23	166.27

8	10.31	0.00	183.57
9	5.74	0.00	171.68
10	3.56	0.00	175.91

Table S11. Conformation number, Relative energy, Mol fraction and the CCS value of Abscisic Acid (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value ($Å^2$)
1	6.71	0.00	154.93
2	1.97	0.01	152.26
3	0.00	0.40	156.69
4	3.69	0.00	157.21
5	0.13	0.32	151.35
6	5.65	0.00	160.34
7	8.42	0.00	149.06
8	9.10	0.00	167.53
9	2.36	0.01	149.49
10	1.97	0.01	151.40
11	0.30	0.24	155.06
12	5.40	0.00	153.94
13	6.25	0.00	152.06

Table S12. Conformation number, Relative energy, Mol fraction and the CCS value of Abscisic Acid (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value ($Å^2$)
1	9.11	0.00	164.72
2	13.42	0.00	158.15
3	13.78	0.00	171.36
4	2.22	0.02	170.96
5	7.37	0.00	170.31
6	2.21	0.02	170.09
7	7.01	0.00	171.22
8	0.00	0.95	162.50
9	7.00	0.00	171.09

Table S13. Conformation number, Relative energy, Mol fraction and the CCS value of Abscisic Acid (Model 3).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value ($Å^2$)
1	8.47	0.00	165.76
2	5.15	0.00	168.35
3	1.84	0.04	168.00
4	0.00	0.85	165.91
5	5.52	0.00	169.86
6	1.80	0.04	169.31
7	5.53	0.00	169.95
8	5.52	0.00	169.98
9	1.53	0.06	170.50
10	3.90	0.00	164.79
11	6.69	0.00	169.77

Table S14. Conformation number, Relative energy, Mol fraction and the CCS value of O-Succinyl-L-homoserine (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value ($Å^2$)
1	0.17	0.24	137.05
2	3.67	0.00	133.90
3	1.36	0.03	133.09

4	6.04	0.00	136.82
5	0.89	0.07	134.01
6	6.61	0.00	127.35
7	2.87	0.00	130.36
8	9.26	0.00	129.79
9	1.00	0.06	136.26
10	0.00	0.32	135.64
11	6.83	0.00	135.02
12	2.08	0.01	132.68
13	2.93	0.00	134.35
14	13.66	0.00	133.08
15	7.57	0.00	136.53
16	7.34	0.00	134.65
17	5.72	0.00	131.76
18	8.22	0.00	128.78
19	0.10	0.27	130.52
20	2.96	0.00	132.86
21	5.51	0.00	134.12

Table S15. Conformation number, Relative energy, Mol fraction and the CCS value of O-Succinyl-L-homoserine (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value ($Å^2$)
1	25.76	0.00	168.90
2	16.97	0.00	165.21
3	17.84	0.00	163.38
4	2.77	0.01	149.90
5	6.65	0.00	153.21
6	15.62	0.00	166.34
7	4.38	0.00	142.40
8	20.03	0.00	169.91
9	20.30	0.00	171.26
10	30.64	0.00	169.32
11	2.58	0.01	147.13
12	0.54	0.28	145.58
13	8.13	0.00	144.34
14	14.42	0.00	157.31
15	25.47	0.00	167.59
16	15.30	0.00	154.28
17	0.00	0.70	145.33

Table S16. Conformation number, Relative energy, Mol fraction and the CCS value of L-Tyrosine (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value ($Å^2$)
1	4.00	0.00	120.82
2	1.91	0.03	118.66
3	2.89	0.01	118.43
4	2.53	0.01	116.38
5	2.19	0.02	120.18
6	2.53	0.01	116.69
7	5.64	0.00	117.07
8	1.91	0.03	119.12
9	0.00	0.88	118.69

Table S17. Conformation number, Relative energy, Mol fraction and the CCS value of L-Tyrosine (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.21	0.14	148.73
2	1.38	0.02	143.58
3	0.21	0.14	151.15
4	0.00	0.19	145.11
5	1.38	0.02	144.53
6	1.38	0.02	142.86
7	0.00	0.19	145.67
8	0.21	0.14	157.40
9	1.38	0.02	142.40
10	0.21	0.14	148.64

Table S18. Conformation number, Relative energy, Mol fraction and the CCS value of L-Citrulline (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.00	0.24	122.59
2	1.20	0.03	122.28
3	0.52	0.10	123.43
4	1.49	0.02	122.04
5	0.00	0.24	123.11
6	0.84	0.06	123.01
7	0.63	0.08	124.74
8	0.85	0.06	122.40
9	0.83	0.06	120.56
10	1.70	0.01	124.57
11	2.22	0.01	121.93
12	2.52	0.00	124.34
13	2.47	0.00	117.13
14	0.54	0.10	122.52

Table S19. Conformation number, Relative energy, Mol fraction and the CCS value of L-Citrulline (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	92.17	0.00	152.81
2	87.25	0.00	182.49
3	97.39	0.00	164.10
4	1.49	0.07	145.50
5	1.28	0.10	145.60
6	2.54	0.01	145.18
7	90.52	0.00	154.03
8	92.84	0.00	161.73
9	85.10	0.00	162.61
10	87.07	0.00	153.77
11	88.85	0.00	169.86
12	0.00	0.83	141.09
13	85.10	0.00	161.94
14	94.76	0.00	140.08
15	92.11	0.00	142.70
16	10.68	0.00	157.03
17	86.54	0.00	167.19

Table S20. Conformation number, Relative energy, Mol fraction and the CCS value of L-Citrulline (Model 3).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	13.07	0.00	162.31
2	12.23	0.00	164.94
3	4.56	0.00	149.79
4	11.11	0.00	154.41
5	13.91	0.00	158.38
6	13.91	0.00	158.48
7	12.45	0.00	169.34
8	0.00	0.58	141.67
9	6.81	0.00	148.03
10	11.59	0.00	156.28
11	2.13	0.02	141.00
12	11.59	0.00	157.02
13	11.53	0.00	157.18
14	13.91	0.00	158.46
15	11.63	0.00	154.81
16	0.30	0.35	142.23
17	1.41	0.05	143.54

Table S21. Conformation number, Relative energy, Mol fraction and the CCS value of Quinolinic Acid (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	5.71	0.00	117.56
2	11.35	0.00	108.81
3	5.00	0.00	122.08
4	3.17	0.00	110.08
5	3.17	0.00	110.05
6	5.00	0.00	122.03
7	0.00	0.99	105.61

Table S22. Conformation number, Relative energy, Mol fraction and the CCS value of Quinolinic Acid (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	1.42	0.07	138.70
2	1.42	0.07	138.18
3	0.00	0.77	142.95
4	1.65	0.05	140.87
5	1.65	0.05	141.05

Table S23. Conformation number, Relative energy, Mol fraction and the CCS value of Nicotinic Acid (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.27	0.39	93.51
2	7.88	0.00	94.74
3	0.00	0.61	89.05

Table S24. Conformation number, Relative energy, Mol fraction and the CCS value of Nicotinic Acid (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.94	0.18	125.51
2	8.28	0.00	127.74
3	0.00	0.82	122.05

Table S25. Conformation number, Relative energy, Mol fraction and the CCS value of Guanidinoacetic Acid (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.00	0.49	89.87
2	0.00	0.49	90.41
3	7.71	0.00	91.23
4	4.02	0.00	90.35
5	7.71	0.00	91.29
6	4.63	0.00	90.16
7	5.95	0.00	91.25
8	4.81	0.00	88.69
9	7.96	0.00	89.98
10	2.04	0.02	90.54

Table S26. Conformation number, Relative energy, Mol fraction and the CCS value of Guanidinoacetic Acid (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	2.03	0.01	130.03
2	5.38	0.00	130.57
3	6.29	0.00	131.39
4	0.61	0.15	130.51
5	0	0.74	131.45
6	0.78	0.1	130.32

Table S27. Conformation number, Relative energy, Mol fraction and the CCS value of Guanidinoacetic Acid (Model 3).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	1.26	0.02	131.44
2	4.82	0.01	130.56
3	0.00	0.96	131.50
4	1.82	0.01	130.37
5	1.97	0.00	131.46
6	7.71	0.00	128.90

Table S28. Conformation number, Relative energy, Mol fraction and the CCS value of Guanidinoacetic Acid (Model 4).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.00	0.18	130.45
2	0.00	0.18	130.37
3	0.24	0.14	130.51
4	2.75	0.00	131.45
5	0.12	0.18	130.51
6	2.75	0.00	131.45
7	2.36	0.00	130.32
8	0.00	0.18	130.61
9	0.24	0.14	128.19
10	2.75	0.00	125.54

Table S29. Conformation number, Relative energy, Mol fraction and the CCS value of Citramalic Acid (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	1.10	0.13	98.92
2	3.01	0.01	91.73
3	3.10	0.00	91.32
4	5.11	0.00	90.68
5	0.00	0.83	99.45
6	6.63	0.00	91.47
7	2.33	0.02	91.62

8	5.04	0.00	93.39
9	2.22	0.02	89.21
10	4.54	0.00	90.11

Table S30. Conformation number, Relative energy, Mol fraction and the CCS value of Citramalic Acid (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	9.72	0.00	122.13
2	31.50	0.00	120.51
3	10.63	0.00	123.37
4	34.51	0.00	120.13
5	7.39	0.00	121.92
6	40.15	0.00	121.49
7	0.00	0.93	122.58
8	11.16	0.00	124.18
9	1.57	0.07	121.61
10	41.86	0.00	123.67
11	39.12	0.00	121.14

Table S31. Conformation number, Relative energy, Mol fraction and the CCS value of Citramalic Acid (Model 3).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	15.91	0.00	122.28
2	5.20	0.00	119.52
3	6.31	0.00	118.25
4	5.67	0.00	121.36
5	14.34	0.00	121.57
6	14.34	0.00	122.28
7	5.67	0.00	121.62
8	15.92	0.00	121.98
9	0.00	0.50	119.91
10	16.11	0.00	123.85
11	5.20	0.00	119.74
12	0.00	0.50	119.33

Table S32. Conformation number, Relative energy, Mol fraction and the CCS value of Citramalic Acid (Model 4).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	9.72	0.00	122.13
2	31.50	0.00	120.51
3	10.63	0.00	123.37
4	34.51	0.00	120.13
5	7.39	0.00	121.92
6	40.15	0.00	121.49
7	0.00	0.93	121.58
8	11.16	0.00	124.18
9	1.57	0.07	121.61
10	41.86	0.00	123.67
11	39.12	0.00	121.14

Table S33. Conformation number, Relative energy, Mol fraction and the CCS value of N-Methyl-glutamate (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	1.63	0.05	107.04

2	9.84	0.00	106.76
3	8.86	0.00	106.73
4	4.81	0.00	105.14
5	3.25	0.00	106.72
6	3.03	0.00	104.90
7	6.23	0.00	105.13
8	0.00	0.74	105.22
9	6.94	0.00	107.39
10	10.51	0.00	107.11
11	9.46	0.00	105.30
12	8.51	0.00	106.47
13	1.94	0.03	104.78
14	5.25	0.00	106.07
15	4.83	0.00	106.66
16	7.26	0.00	106.29
17	6.48	0.00	103.83
18	0.85	0.18	105.47

Table S34. Conformation number, Relative energy, Mol fraction and the CCS value of N-Methyl-glutamate (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	14.80	0.00	136.60
2	2.75	0.01	131.42
3	14.80	0.00	135.99
4	10.19	0.00	141.56
5	9.59	0.00	137.47
6	10.64	0.00	142.05
7	14.00	0.00	139.21
8	0.00	0.99	129.28
9	15.28	0.00	137.76
10	7.86	0.00	129.65
11	12.70	0.00	140.33
12	15.65	0.00	137.76
13	9.68	0.00	132.37
14	11.36	0.00	137.93
	12.18	0.00	134.45

Table S35. Conformation number, Relative energy, Mol fraction and the CCS value of Serotonin (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.39	0.24	131.62
2	0.50	0.20	130.91
3	1.69	0.03	134.65
4	1.69	0.03	135.40
5	2.69	0.00	131.42
6	1.69	0.03	134.79
7	0.00	0.47	130.79

Table S36. Conformation number, Relative energy, Mol fraction and the CCS value of Serotonin (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	6.14	0.00	167.57
2	4.10	0.00	168.00
3	4.10	0.00	166.07
4	0.00	0.38	164.69
5	0.69	0.12	168.14
6	0.00	0.38	163.40

7	0.69	0.12	168.31
8	6.14	0.00	169.31

Table S37. Conformation number, Relative energy, Mol fraction and the CCS value of Serotonin (Model 3).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.00	0.96	161.83
2	2.37	0.02	155.56
3	3.49	0.00	159.15
4	2.37	0.02	155.10

Table S38. Conformation number, Relative energy, Mol fraction and the CCS value of L-Mimosine (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	14.47	0.00	133.07
2	1.47	0.04	125.36
3	2.31	0.01	127.53
4	11.57	0.00	130.46
5	1.55	0.03	126.61
6	10.77	0.00	130.18
7	17.99	0.00	132.39
8	0.00	0.44	129.95
9	0.00	0.44	126.28
10	1.30	0.05	127.57

Table S39. Conformation number, Relative energy, Mol fraction and the CCS value of L-Mimosine (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	23.91	0.00	162.47
2	15.27	0.00	156.16
3	15.07	0.00	157.25
4	10.92	0.00	157.37
5	0.00	1.00	150.90
6	23.54	0.00	158.41
7	11.46	0.00	153.30
8	22.38	0.00	162.93
9	15.08	0.00	156.74
10	12.54	0.00	160.07

Table S40. Conformation number, Relative energy, Mol fraction and the CCS value of L-Mimosine (Model 3).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.74	0.15	145.94
2	8.17	0.00	148.68
3	7.23	0.00	141.68
4	0.00	0.54	144.64
5	0.56	0.21	146.18
6	8.02	0.00	149.43
7	2.46	0.01	141.57
8	1.33	0.06	147.01
9	1.58	0.04	147.07

Table S41. Conformation number, Relative energy, Mol fraction and the CCS value of L-Tryptophan (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	1.24	0.10	139.45

2	0.00	0.78	136.25
3	2.07	0.02	139.43
4	3.70	0.00	138.42
5	1.82	0.04	139.69
6	5.97	0.00	137.52
7	2.31	0.02	140.60
8	1.72	0.04	136.34
9	5.34	0.00	137.08

Table S42. Conformation number, Relative energy, Mol fraction and the CCS value of L-Tryptophan (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.29	0.20	163.24
2	0.63	0.12	160.11
3	0.00	0.33	160.14
4	7.77	0.00	161.67
5	1.47	0.03	159.38
6	3.73	0.00	156.80
7	0.63	0.12	160.65
8	3.61	0.00	159.02
9	5.32	0.00	156.42
10	0.29	0.20	163.78
11	8.94	0.00	163.89

Table S43. Conformation number, Relative energy, Mol fraction and the CCS value of L-Tryptophan (Model 3).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	3.75	0.00	158.92
2	7.03	0.00	169.49
3	9.78	0.00	170.47
4	0.00	0.54	162.72
5	3.78	0.00	166.44
6	3.77	0.00	164.48
7	8.71	0.00	161.02
8	7.41	0.00	165.97
9	0.11	0.45	156.03

Table S44. Conformation number, Relative energy, Mol fraction and the CCS value of L-Ornithine (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	2.00	0.01	99.93
2	0.61	0.09	109.37
3	7.16	0.00	100.32
4	0.80	0.06	99.96
5	1.15	0.03	100.19
6	1.35	0.02	101.25
7	2.48	0.00	98.08
8	2.12	0.01	98.99
9	0.51	0.10	111.70
10	1.44	0.02	100.49
11	0.61	0.09	97.53
12	0.05	0.22	114.18
13	0.00	0.24	115.94
14	5.85	0.00	99.86
15	3.39	0.00	98.31
16	2.01	0.01	102.09
17	0.56	0.09	113.64

Table S45. Conformation number, Relative energy, Mol fraction and the CCS value of L-Ornithine (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	9.81	0.00	120.86
2	18.55	0.00	139.09
3	3.12	0.01	126.82
4	18.41	0.00	135.77
5	19.83	0.00	131.55
6	3.11	0.01	126.21
7	0.00	0.99	127.40
8	18.46	0.00	137.53
9	22.57	0.00	133.84
10	18.55	0.00	138.17
11	19.56	0.00	131.90
12	18.41	0.00	137.52
13	6.50	0.00	121.88

Table S46. Conformation number, Relative energy, Mol fraction and the CCS value of L-Ornithine (Model 3).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	20.39	0.00	128.22
2	28.58	0.00	128.21
3	16.85	0.00	133.06
4	16.85	0.00	132.15
5	20.24	0.00	134.91
6	15.79	0.00	129.14
7	17.51	0.00	129.48
8	16.07	0.00	131.00
9	0.00	1.00	126.63
10	24.39	0.00	133.99
11	19.76	0.00	133.13
12	20.82	0.00	133.90
13	17.88	0.00	131.68
14	21.78	0.00	134.97
15	28.47	0.00	131.71

Table S47. Conformation number, Relative energy, Mol fraction and the CCS value of N, N-Dimethylglycine (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	0.00	0.21	79.01
2	0.81	0.05	79.37
3	0.00	0.21	78.41
4	0.00	0.21	78.81
5	0.82	0.05	79.16
6	0.00	0.21	78.83
7	0.82	0.05	79.30

Table S48. Conformation number, Relative energy, Mol fraction and the CCS value of N, N-Dimethylglycine (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	1.76	0.02	130.85
2	0.00	0.98	128.81
3	13.20	0.00	144.33
4	12.65	0.00	146.50
5	9.79	0.00	145.89

Table S49. Conformation number, Relative energy, Mol fraction and the CCS value of Kynurenine (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	4.25	0.00	137.31
2	6.28	0.00	133.94
3	3.67	0.00	133.63
4	3.68	0.00	133.96
5	1.52	0.06	132.86
6	1.98	0.03	132.91
7	1.52	0.06	133.25
8	4.35	0.00	132.07
9	4.60	0.00	132.44
10	0.00	0.81	138.13
11	4.77	0.00	137.71
12	2.05	0.03	131.49

Table S50. Conformation number, Relative energy, Mol fraction and the CCS value of Kynurenine (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	4.25	0.00	137.31
2	6.28	0.00	133.94
3	3.67	0.00	133.63
4	3.68	0.00	133.96
5	1.52	0.06	145.86
6	1.98	0.03	132.91
7	1.52	0.06	146.25
8	4.35	0.00	132.07
9	4.60	0.00	132.44
10	0.00	0.81	148.13
11	4.77	0.00	137.71
12	2.05	0.03	131.49

Table S51. Conformation number, Relative energy, Mol fraction and the CCS value of L-Asparagine (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	1.88	0.01	94.51
2	3.99	0.00	93.72
3	3.99	0.00	95.87
4	1.31	0.03	91.89
5	1.85	0.01	94.52
6	3.23	0.00	94.37
7	9.14	0.00	93.37
8	2.58	0.00	95.48
9	4.06	0.00	94.82
10	0.01	0.27	115.23
11	0.00	0.27	117.12
12	0.75	0.08	95.71
13	0.00	0.27	115.00
14	1.03	0.05	94.07

Table S52. Conformation number, Relative energy, Mol fraction and the CCS value of L-Asparagine (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	1.88	0.01	94.51
2	3.99	0.00	93.72
3	3.99	0.00	95.87
4	1.31	0.03	91.89
5	1.85	0.01	94.52
6	3.23	0.00	94.37
7	9.14	0.00	93.37
8	2.58	0.00	95.48
9	4.06	0.00	94.82
10	0.01	0.27	128.23
11	0.00	0.27	130.12
12	0.75	0.08	95.71
13	0.00	0.27	129.25
14	1.03	0.05	129.07

Table S53. Conformation number, Relative energy, Mol fraction and the CCS value of L-Asparagine (Model 3).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	1.88	0.01	94.51
2	3.99	0.00	93.72
3	3.99	0.00	95.87
4	1.31	0.03	129.89
5	1.85	0.01	94.52
6	3.23	0.00	94.37
7	9.14	0.00	93.37
8	2.58	0.00	95.48
9	4.06	0.00	94.82
10	0.01	0.27	130.23
11	0.00	0.27	130.12
12	0.75	0.08	129.71
13	0.00	0.27	130.25
14	1.03	0.05	126.07

Table S54. Conformation number, Relative energy, Mol fraction and the CCS value of 2-Aminoadipic Acid (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	2.74	0.01	108.86
2	1.12	0.08	108.84
3	5.27	0.00	107.87
4	6.52	0.00	109.95
5	5.49	0.00	109.20
6	5.21	0.00	111.33
7	6.39	0.00	108.55
8	6.80	0.00	108.96
9	12.45	0.00	110.34
10	7.27	0.00	110.50
11	3.22	0.00	107.17
12	0.00	0.53	107.18
13	0.91	0.11	108.62
14	0.80	0.14	108.26
15	2.24	0.01	107.68
16	0.85	0.13	106.96

Table S55. Conformation number, Relative energy, Mol fraction and the CCS value of 2-Aminoadipic Acid (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	24.59	0.00	146.02
2	8.18	0.00	140.30
3	10.30	0.00	143.95
4	8.81	0.00	142.60
5	2.15	0.00	136.60
6	0.00	0.98	132.40
7	13.78	0.00	144.97
8	3.85	0.00	131.64
9	13.60	0.00	143.42
10	8.70	0.00	140.67
11	15.74	0.00	139.50
12	1.05	0.20	134.52
13	6.12	0.00	138.97
14	10.47	0.00	142.44

Table S56. Conformation number, Relative energy, Mol fraction and the CCS value of Glutamine (Model 1).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	9.22	0.00	112.86
2	15.97	0.00	116.94
3	1.90	0.04	106.76
4	0.00	0.92	105.45
5	25.00	0.00	124.95
6	21.15	0.00	122.75
7	1.84	0.04	107.79
8	16.11	0.00	123.69
9	4.03	0.00	106.69
10	15.14	0.00	123.07
11	10.96	0.00	115.26
12	1.43	0.00	107.73

Table S57. Conformation number, Relative energy, Mol fraction and the CCS value of Glutamine (Model 2).

Conformation No.	Relative energy (kcal/mol)	Mol fraction	CCS value (Å^2)
1	2.02	0.00	129.64
2	7.14	0.00	134.82
3	1.08	0.00	128.34
4	9.62	0.00	135.53
5	2.01	0.00	129.13
6	5.36	0.00	129.41
7	0.70	0.23	124.94
8	4.56	0.00	137.47
9	0.00	0.76	127.00
10	4.55	0.00	137.63

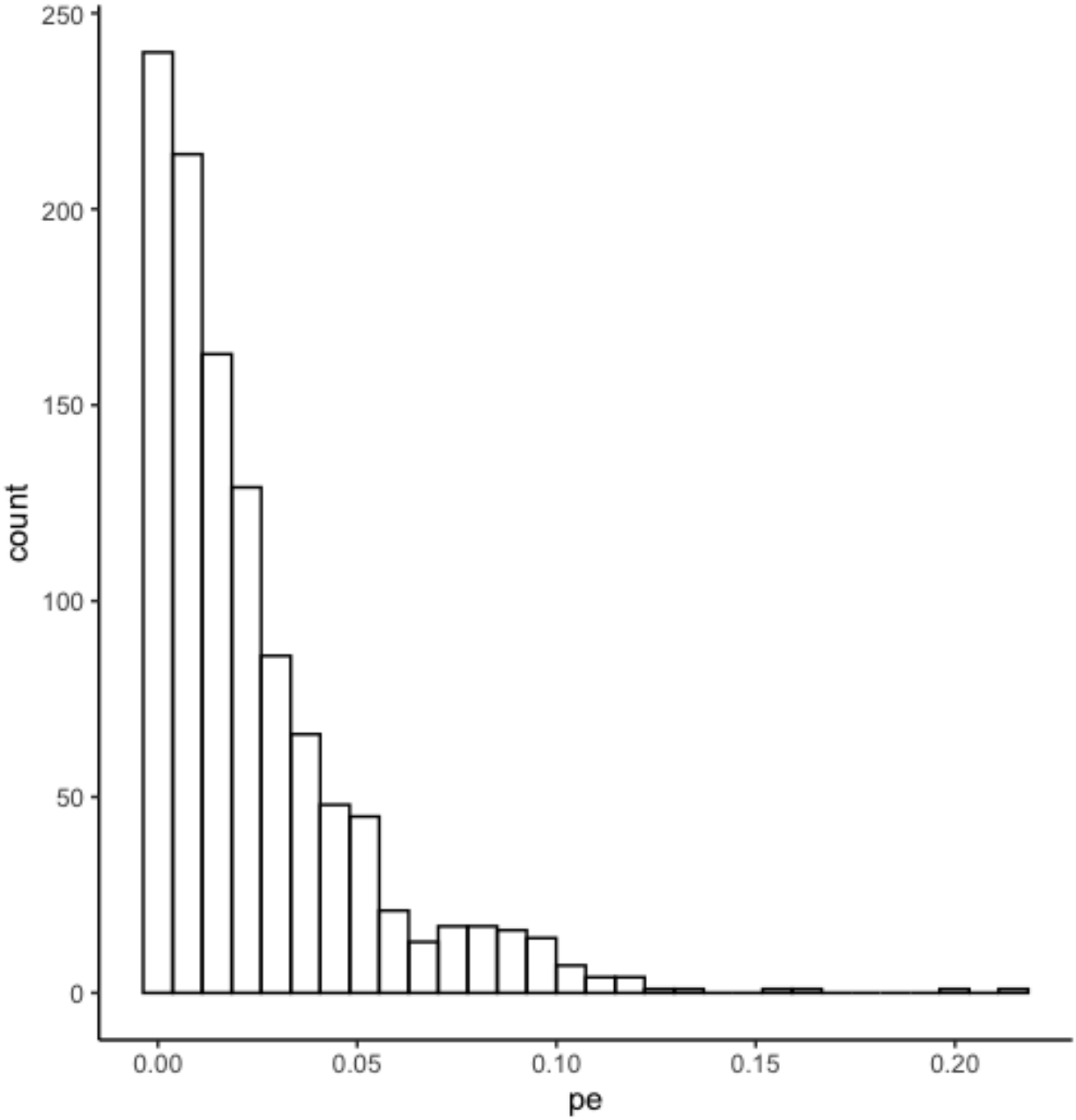

Figure S 1. Relative error distribution of the CCS values obtained from different sources.

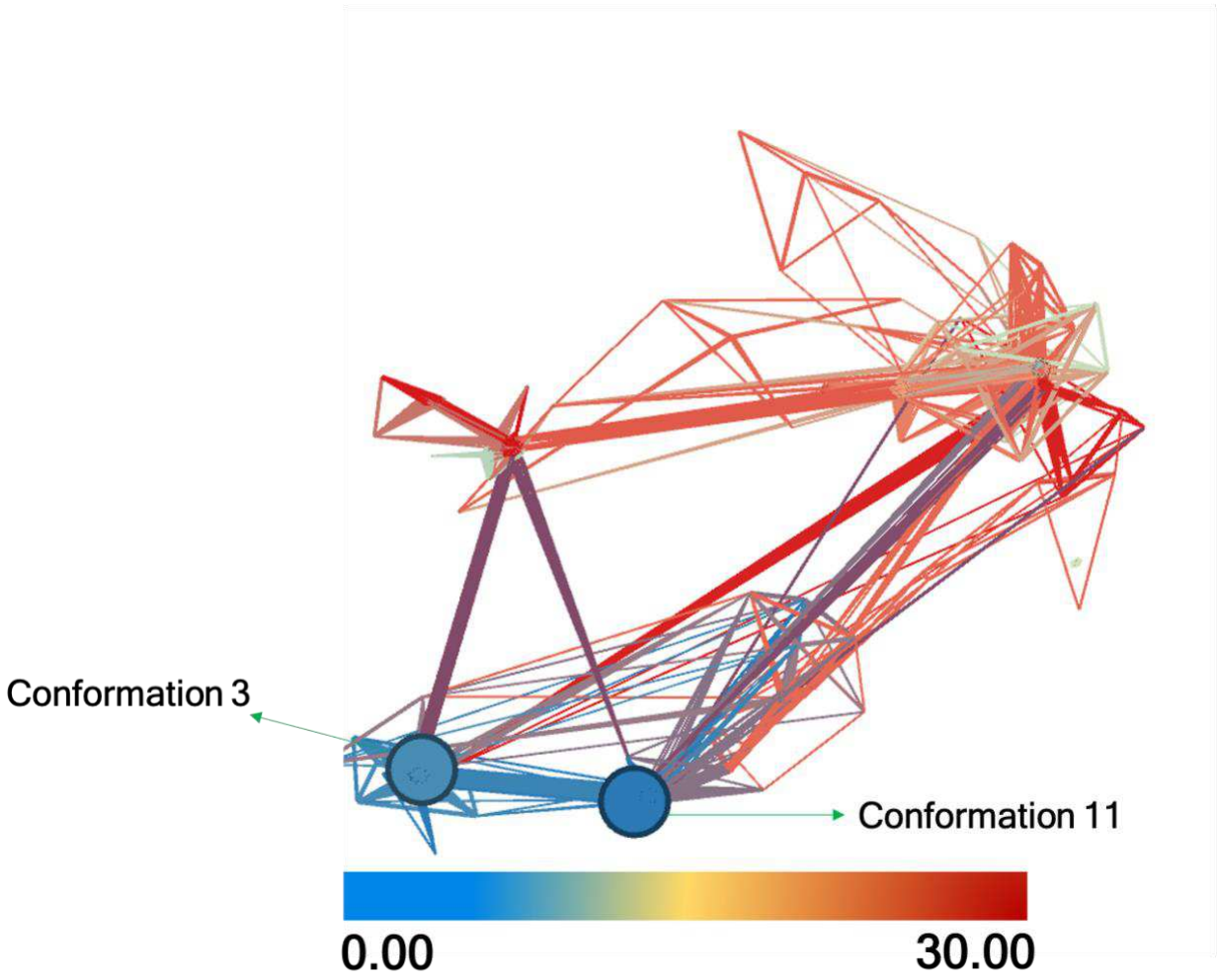

Figure S 2. Unsupervised clustering of Carnosine (Model 4) by Autograph. Low to high energies (QM energies) are represented by blue to red respectively and relative energy value is in Kcal mol^{-1}. This figure is generated by Gephi software.

Benchmark study of the computational cost of our workflow.

The computing time of the workflow was measured. We used CPU and GPU resources on the High-Performance Computing Cluster (HPCC) of the Institute of Cyber Enabled Research (ICER) at Michigan State University (MSU). We parallelized most of the computational steps as individual jobs on the cluster, therefore the throughput depends on the computational resources available. We report the summary statistics for the time assigned to calculating the CCS values of L-carnosine by particular computational steps and by the charge.

Each computational step consists of one or more jobs executed in a sequential or parallel fashion. Consider the following general computational step (Figure). It consists of jobs A, B, and C performed sequentially, however job B is a collection of *n* jobs, each having their own duration. Therefore to report an overall time cost, we must take a summary statistic of the durations of Job B, and add it to the times taken for Job A and Job C. The duration of each computational step was calculated in this manner, first taking the summary statistic of the duration of parallel jobs, and then summing those values for sequential jobs.

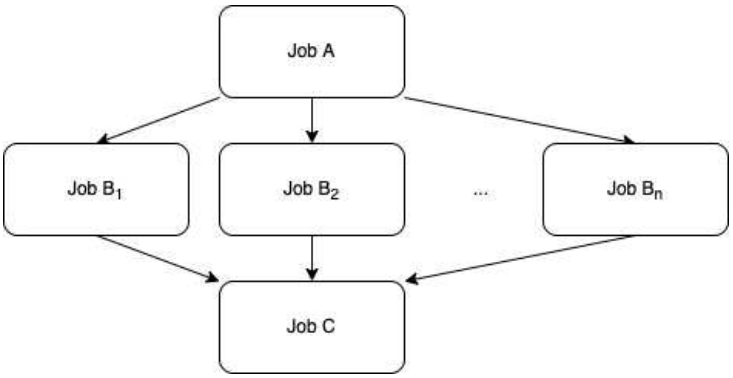

Figure S3. directed acyclic graph describing the sequence of jobs to execute in a general computational step. Here, Jobs A, B, and C are sequential, so we add the time taken at each step to obtain the overall duration of this step. However, Job B consists of *n* separate jobs running in parallel, so we take a summary statistic of the distribution of time taken for Job B.

While the mean may illustrate a representative figure for the time taken at each computational step, the job which takes the longest time dictates the overall computational time. However, the maximum value is sensitive to outliers, thus is insufficient to represent the overall computational demand. Therefore we report both the mean and the maximum time for all jobs in each step of the pipeline.

We utilize Gaussian as the quantum chemistry (QC) program in the manuscript. Because of Gaussian's license agreement that stipulates it cannot be benchmarked, we cannot provide these times. As a time benchmark, we use instead the QUICK program, which has been extensively benchmarked against other QC programs. Therefore, the reader is free to estimate the overall compute time by adjusting the steps involving QM single point energy or geometry optimization calculations by the factor of QUICK relative to other QC programs.

GPU resources are utilized at the deprotonation/protonation isomer generation, conformational refinement, and QM calculation steps. The serial GPU implementation of QUICK is used as the QM program for the deprotonation/protonation isomer generation and QM calculation step. Various GPU resources were used to run QUICK, including the NVIDIA Tesla K80 GPU and NVIDIA V100 Tensor Core GPU. Any available GPU is used to run TorchANI for the conformational refinement step.

Table S58. Time taken for each computational step of the fast version of CCS prediction workflow of L-carnosine using the NVIDIA Tesla K80 GPU for QUICK. The numbers of deprotonation/protonation isomers considered are one, two, and three respectively for charges –1, 0, and 1.

charge	step	$time_{mean}$ (min)	$time_{max}$ (min)	jobs
–1	(De)protonation Isomer Generation	7.48	12.08	10
	Conformation Generation	0.35	0.35	1
	Refinement	12.58	33.67	31
	Clustering	0.28	0.28	11
	QM Single Point Energy	3.52	5.78	33
	CCS Prediction	12.57	13.88	14

	total	36.78	66.05	100
0	(De)protonation Isomer Generation	6.53	7.40	10
	Conformation Generation	0.34	0.35	2
	Refinement	5.71	22.95	54
	Clustering	0.19	0.22	22
	QM Single Point Energy	3.67	4.35	80
	CCS Prediction	8.45	9.97	60
	total	24.89	45.23	227
1	(De)protonation Isomer Generation	6.10	6.90	5
	Conformation Generation	0.38	0.40	3
	Refinement	8.65	27.85	83
	Clustering	0.21	0.25	33
	QM Single Point Energy	4.06	5.08	97
	CCS Prediction	13.00	15.12	42
	total	32.39	55.60	261

Table S59. Time taken for each computational step of the standard version of CCS prediction workflow of L-carnosine using the NVIDIA Tesla K80 GPU for QUICK. The numbers of deprotonation/protonation isomers considered are one, two, and three respectively for charges –1, 0, and 1.

charge	step	$time_{mean}$ (min)	$time_{max}$ (min)	jobs
–1	(De)protonation Isomer Generation	7.48	12.08	22
	Conformation Generation	0.35	0.35	1
	Refinement	12.58	33.67	32
	Clustering	0.28	0.28	11
	QM Geometry Optimization	89.23	127.05	34
	CCS Prediction	12.97	14.63	18
	total	122.89	188.07	118
0	(De)protonation Isomer Generation	6.53	7.40	22
	Conformation Generation	0.34	0.35	2
	Refinement	5.71	22.95	56
	Clustering	0.19	0.22	22
	QM Geometry Optimization	82.76	150.42	82
	CCS Prediction	8.33	9.93	64

	total	103.85	191.27	248
1	(De)protonation Isomer Generation	6.10	6.90	12
	Conformation Generation	0.38	0.40	3
	Refinement	8.65	27.85	86
	Clustering	0.21	0.25	33
	QM Geometry Optimization	124.23	252.73	100
	CCS Prediction	13.04	15.30	44
	total	152.61	303.43	278

Table S60. Time taken for each computational step of the fast version of CCS prediction workflow of L-carnosine using the NVIDIA V100 Tensor Core GPU for QUICK. The numbers of deprotonation/protonation isomers considered are one, two, and three respectively for charges –1, 0, and 1.

charge	step	$time_{mean}$ (min)	$time_{max}$ (min)	jobs
–1	(De)protonation Isomer Generation	0.78	1.13	22
	Conformation Generation	0.45	0.45	1
	Refinement	3.05	4.25	32
	Clustering	0.30	0.33	11
	QM Single Point Energy	0.44	0.5	33
	CCS Prediction	12.56	14.75	11
	total	17.60	21.41	110
0	(De)protonation Isomer Generation	0.73	0.82	22
	Conformation Generation	0.53	0.62	2
	Refinement	3.33	4.13	56
	Clustering	0.23	0.27	22
	QM Single Point Energy	0.48	0.53	83
	CCS Prediction	8.11	9.27	53
	total	13.41	15.63	238
1	(De)protonation Isomer Generation	0.66	0.75	12
	Conformation Generation	0.58	0.63	3
	Refinement	3.38	4.37	86
	Clustering	0.27	0.40	33
	QM Single Point Energy	0.51	0.58	100
	CCS Prediction	12.66	14.95	43

	total	18.06	21.68	277

Table S61. Time taken for each computational step of the standard version of CCS prediction workflow of L-carnosine using the NVIDIA V100 Tensor Core GPU for QUICK. The numbers of deprotonation/protonation isomers considered are one, two, and three respectively for charges –1, 0, and 1.

charge	step	$time_{mean}$ (min)	$time_{max}$ (min)	jobs
–1	(De)protonation Isomer Generation	0.78	1.13	22
	Conformation Generation	0.45	0.45	1
	Refinement	3.05	4.25	32
	Clustering	0.30	0.33	11
	QM Geometry Optimization	63.89	187.55	33
	CCS Prediction	12.69	13.90	23
	total	81.16	207.61	122
0	(De)protonation Isomer Generation	0.73	0.82	22
	Conformation Generation	0.53	0.62	2
	Refinement	3.33	4.13	56
	Clustering	0.23	0.26	22
	QM Geometry Optimization	66.09	182.73	83
	CCS Prediction	8.31	9.53	62
	total	79.22	198.10	247
1	(De)protonation Isomer Generation	0.66	0.75	12
	Conformation Generation	0.58	0.63	3
	Refinement	3.38	4.37	86
	Clustering	0.27	0.40	33
	QM Geometry Optimization	78.12	213.17	100
	CCS Prediction	12.86	15.03	53
	total	95.87	234.35	287