

XYZ file format for molecular structure and properties.

The XYZ format (originally developed for the XMol program by the Minnesota Supercomputer Center) is a widespread plain text format for encoding Cartesian coordinates of molecules, with no formal specification. It contains a header line specifying the number of atoms n_a , a comment line, and n_a lines containing element type and atomic coordinates, one atom per line. We have extended this format as indicated in [Table 2](#). Now, the comment line is used to store all scalar properties, Mulliken charges are added as a fifth column. Harmonic vibrational frequencies, SMILES and InChI are appended as respective additional lines

<https://www.nature.com/articles/sdata201422/tables/2>

From: [Quantum chemistry structures and properties of 134 kilo molecules](#)

Line	Content
1	Number of atoms n_a
2	Scalar properties (see Table 3)
3,..., n_a+2	Element type, coordinate (x, y, z , in Å), Mulliken partial charges (in e) on atoms
n_a+3	Harmonic vibrational frequencies ($3n_a-5$ or $3n_a-6$, in cm^{-1})
n_a+4	SMILES strings from GDB-17 and from B3LYP relaxation
n_a+5	InChI strings for Corina and B3LYP geometries

n_a = number of atoms.

Table 3: Calculated properties.

<https://www.nature.com/articles/sdata201422/tables/3>

From: [Quantum chemistry structures and properties of 134 kilo molecules](#)

No.	Property	Unit	Description
1	tag	—	‘gdb9’ string to facilitate extraction
2	i	—	Consecutive, 1-based integer identifier
3	A	GHz	Rotational constant
4	B	GHz	Rotational constant
5	C	GHz	Rotational constant
6	μ	D	Dipole moment
7	α	alpha_0^3	Isotropic polarizability

8	ϵ HOMO	Ha	Energy of HOMO
9	ϵ LUMO	Ha	Energy of LUMO
10	ϵ gap	Ha	Gap (ϵ LUMO– ϵ HOMO)
11	$\langle R^2 \rangle$	α_0^2	Electronic spatial extent
12	zpve	Ha	Zero point vibrational energy
13	U 0	Ha	Internal energy at 0 K
14	U	Ha	Internal energy at 298.15 K
15	H	Ha	Enthalpy at 298.15 K
16	G	Ha	Free energy at 298.15 K
17	C v	Cal/molK	Heat capacity at 298.15 K

File example: dsgdb9nsd_000001.xyz

5

```
gdb 1 157.7118    157.70997    157.70699    0.    13.21 -0.3877    0.1171 0.5048
      35.3641    0.044749    -40.47893    -40.476062    -40.475117    -40.498597
      6.469
```

```
C    -0.0126981359    1.0858041578    0.0080009958    -0.535689
```

```
H    0.002150416 -0.0060313176    0.0019761204    0.133921
```

```
H    1.0117308433    1.4637511618    0.0002765748    0.133922
```

```
H    -0.540815069 1.4475266138    -0.8766437152    0.133923
```

```
H    -0.5238136345    1.4379326443    0.9063972942    0.133923
```

```
1341.307    1341.3284    1341.365    1562.6731    1562.7453    3038.3205
      3151.6034    3151.6788    3151.7078
```

```
C    C
```

```
InChI=1S/CH4/h1H4 InChI=1S/CH4/h1H4
```

The **number of vibrational modes** (different types of vibrations) in a molecule is $3N-5$ for linear molecules and $3N-6$ for nonlinear molecules, where N is the **number** of atoms.

GDB-9 is a database of molecular quantum calculations which includes computed geometric, energetic, electronic, and thermodynamic properties for 134k stable small organic molecules made up of carbon, hydrogen, nitrogen, oxygen, and fluorine.