1. RDKit descriptors

1.1 Basic phys-chem / size

Descriptor(s)	Family	What it means
`MolWt`, `HeavyAtomMolWt`, `ExactMolWt`	Mass / composition	Total molecular weight, only heavy atoms, exact isotopic mass.
`NumValenceElectrons`, `NumRadicalElectrons`	Electronic composition	Total count of valence and radical electrons.
`MolLogP`	Lipophilicity	cLogP (octanol/water). Hydrophobicity and indirectly electron distribution.
`MolMR`	Molar refractivity	Approximate **polarizability** of the electron cloud.
`qed`	Drug-likeness	Composite drug-likeness score (Lipinski-like).
`SPS`, `Phi`	Size/shape scores	Simple global shape/flexibility measures.

1.2 Topological / structural indices

Descriptor(s)	Family	What it means
`BalabanJ`	Topological index	Graph-based index with low degeneracy; encodes branching/connectivity.
`BertzCT`	Complexity	Increases with size, heteroatoms, branching; a "complexity" score.
`Ipc`, `AvgIpc`	Information content	Info-theoretic measure of diversity of degrees in the molecular graph.
`Chi0–Chi4`	Connectivity (χ)	Path-based indices using atom degree (topology only).
`Chi0n–Chi4n`	Normalized χ	Size-normalized variants of χ indices.
`Chi0v–Chi4v`	**Valence connectivity**	χ indices weighted by **valence**: this is your *valentnost'/valencia* coverage.
`HallKierAlpha`	HK alpha	Branching + electronic environment (used in logP/QSAR).
`Kappa1–Kappa3`	Kier kappa shape	Shape/branching indices, sensitive to rings and linearity.

1.3 Surface & polar / electronic distribution

Descriptor(s)	Family	What it means
`LabuteASA`	Solvent-accessible surface	Approximate SASA (Å²); how much surface a molecule presents to solvent.
`TPSA`	Topological polar surface area	Sum of polar fragments; correlates with permeability, H-bonding.
`PEOE_VSA1–14`	**Charge × VSA**	Van der Waals surface area binned by **Gasteiger partial charge**; where positive/negative charge is on the surface.
`SMR_VSA1–10`	MR × VSA	VSA binned by **molar refractivity** (polarizability-related).

Descriptor(s)	Family	What it means
`SlogP_VSA1–12`	logP × VSA	VSA binned by fragment logP (hydrophobic/hydrophilic surface regions).
`EState_VSA1–11`	E-state × VSA	VSA bins weighted by **electrotopological state**.
`VSA_EState1–10`	VSA × E-state	Complementary combination of VSA and E-state (slightly different aggregation).

1.4 E-state (electrotopological) indices

Descriptor(s)	Family	What it means
`MaxEStateIndex`, `MinEStateIndex`	E-state extremes	Max/min atomic electrotopological state in the molecule.
`MaxAbsEStateIndex`, `MinAbsEStateIndex`	Absolute extremes	Same, but based on absolute values (very strong donors/acceptors).

E-state = intrinsic electronic character of an atom + correction for its topological environment.

1.5 Atomic / ring / H-bond counts

Descriptor(s)	Family	What it means
`HeavyAtomCount`, `RingCount`	Size / rings	Number of heavy atoms and total ring systems.
`NumHAcceptors`, `NumHDonors`	H-bonding	Hydrogen bond acceptor and donor counts (your “atomic descriptors” from article).
`NHOHCount`, `NOCount`	Polar atoms	Count of N/O or NH/OH moieties.
`NumHeteroatoms`	Heteroatom count	Total N, O, S, halogens, etc.
`NumRotatableBonds`	Flexibility	Number of rotatable single bonds.
`NumAliphaticCarbocycles`, `NumAromaticCarbocycles`, `NumSaturatedCarbocycles`	Carbocycles	Counts of carbon-only rings (aliphatic/aromatic/saturated).
`NumAliphaticHeterocycles`, `NumAromaticHeterocycles`, `NumSaturatedHeterocycles`	Heterocycles	Heteroatom-containing rings by type.
`NumAromaticRings`, `NumSaturatedRings`	Ring type	Total aromatic vs saturated ring systems.
`NumHeterocycles`	Heterocycle count	Number of heterocyclic ring systems.
`NumSpiroAtoms`, `NumBridgeheadAtoms`	Special ring atoms	Ring junction complexity.
`NumAtomStereoCenters`, `NumUnspecifiedAtomStereoCenters`	Chirality	Count of defined/undefined chiral centers.

1.6 Fragment (functional group) descriptors `fr_*`

Pattern: `fr_XXX`

Family: Fragment / functional group counts

Examples & meaning:

- `fr_COO`, `fr_COO2`, `fr_ester`, `fr_lactone` → carboxylic/ester functionalities.
- `fr_NH0`, `fr_NH1`, `fr_NH2`, `fr_N_O`, `fr_Nhpyrrole`, `fr_guanido`, `fr_amide`, `fr_amidine`, `fr_priamide` → N-containing groups.
- `fr_halogen`, `fr_sulfide`, `fr_sulfone`, `fr_sulfonamd`, `fr_phos_acid`, `fr_phos_ester` → heteroatom/acid/sulfur/phosphorus groups.
- `fr_nitro`, `fr_nitrile`, `fr_azide`, `fr_azo`, `fr_diazo`, `fr_nitro_arom`, `fr_nitro_arom_nonortho`, `fr_nitroso` → strong EWG / special electronic motifs.
- Aromatic fragments: `fr_benzene`, `fr_benzodiazepine`, `fr_furan`, `fr_oxazole`, `fr_thiophene`, `fr_pyridine`, `fr_tetrazole`, etc.

These encode **chemically meaningful patterns** important for reactivity and drug-likeness.

---

2. ChemDes – extra *electronic* descriptors

2.1 Quantum frontier orbital descriptors

Descriptor(s)	Family	What it means
`HOMO_eV`	Frontier orbital	Energy (eV) of the highest occupied MO → electron donating ability.
`LUMO_eV`	Frontier orbital	Energy (eV) of the lowest unoccupied MO → electron accepting ability.
`HL_Gap_eV`	Reactivity / hardness	HOMO–LUMO energy gap: small gap → softer/more reactive.

These are true **quantum** descriptors (ChemDes uses MOPAC or similar under the hood).

2.2 Global charge / electronegativity descriptors (`Q*`)

Example names	Family	What they mean
`Qmax`, `Qmin`	Global charge	Maximum and minimum atomic partial charges in the molecule.
`Qindex`, `Qass`	Charge indices	Global indices summarizing charge asymmetry / spread.
`QCmax`, `QCmin`, `QCss`	Carbon charge	Max/min/sum/spread of charges over carbon atoms.
`QOmax`, `QOmin`, `QOss`	Oxygen charge	Same, but only for O atoms.
`QHmax`, `QHmin`,	Hydrogen	Charge statistics over hydrogens.

Example names	Family	What they mean
`QHss`	charge	
`QNmin`, `QNss`	Nitrogen charge	Charge statistics over nitrogens.
`QHmax`, `QHmin` (again)	H extremes	Most positive/negative H charges.

In words: **how charge is distributed globally and per element** – more detailed than RDKit's Max/Min partial charge.

2.3 Charge/valence-weighted connectivity (`Chi*` / `Chiv*` variants)

We have χ0–χ4 and χv0–χv4 in RDKit. ChemDes adds **longer path and charge/polar variants** like:

Example names	Family	What they mean
`Chi3ch`, `Chi4ch`, `Chi5ch`, `Chi6ch`	Charge-weighted χ	Connectivity indices including **charge correction** (`ch` = charged). Longer paths (3–6) than RDKit.
`Chi3c`, `Chi4c`, `Chi8`, `Chi9`, `Chi10`	Higher-order χ	High-order path/chains; more global connectivity.
`Chi4pc`	Polar-corrected χ	χ index with **polar correction** (`pc`), emphasising electronegative atoms.
`Chiv3c`, `Chiv4c`	Valence/ normalized χ	`Chiv` = valence/normalized χ variant; encoded with additional normalization or valence weighting.
`Chiv3ch`, `Chiv4pc`, `Chiv5ch`, `Chiv6ch`	Charge/valence χ	Combined valence + charge/polar corrections over longer path lengths.
`mChi1`	Mean/modified χ	Average/modified connectivity index (variant of χ1).

These are **advanced "valentnosť + electronic" topological descriptors**, extending RDKit's shorter χ set.

2.4 Topological autocorrelation descriptors (ATS / MATS / GATS)

These are the big "global electron distribution" families.

2.4.1 ATS (Broto–Moreau autocorrelation)

- Names like: `ATSe1–ATSe8`, `ATSv1–ATSv8`, `ATSp1–ATSp8`

Pattern	Property	Meaning
`ATSe*`	`e` = electronegativity/intrinsic state	Autocorrelation of **electronegativity / intrinsic state** over topological distances.
`ATSv*`	`v` = van der Waals volume	Autocorrelation of atomic volumes (related to electronic size).
`ATSp*`	`p` = polarizability	Autocorrelation of **polarizability** – how deformable electron cloud is along the graph.

Number (1–8) = **topological lag** (approx. number of bonds between correlated atoms).

2.4.2 MATS (Moran autocorrelation)

- Names like: `MATSe*`, `MATSv*`, `MATSp*`

Same properties as ATS, different statistical formula (Moran's I):

Pattern	Property	Meaning
`MATSe*`	electronegativity	Moran autocorrelation of EN along the graph.
`MATSv*`	volume	Moran autocorrelation of atomic volume.
`MATSp*`	polarizability	Moran autocorrelation of polarizability.

2.4.3 GATS (Geary autocorrelation)

- Names like: `GATSe*`, `GATSv*`, `GATSp*`

Pattern	Property	Meaning
`GATSe*`	electronegativity	Geary autocorrelation of EN.
`GATSv*`	volume	Geary autocorrelation of atomic volume.
`GATSp*`	polarizability	Geary autocorrelation of polarizability.

Intuition:
These ATS/MATS/GATS families encode **how similar or different electronic properties are between atoms at various distances across the graph** → "global electron distribution" view.

2.5 VSA / surface electronic descriptors (if you kept them)

Pattern	Family	What it means
`PEOEVSA*`	PEOE charge × VSA	Like RDKit `PEOE_VSA*`, but ChemDes' implementation/parameters.
`slogPVSA*`	logP × VSA	Like RDKit `SlogP_VSA*`, but alternative binning.
`VSAEstate*`, `EstateVSA*`	E-state × VSA	Similar to RDKit `EState_VSA*` / `VSA_EState*`.

2.6 Other single electronic scalar(s)

Descriptor	Family	What it means
`Pol`	Polarizability	Global polarizability estimate (similar idea to `MolMR`, but different method).

---

3. Legend

3.1 Property letters in ATS/MATS/GATS and similar

- `v` → **van der Waals volume** (atomic volume / size)
- `e` → **electronegativity / intrinsic state** (often Sanderson EN or E-state–derived)
- `p` → **polarizability** (how easily the electron cloud is distorted)

So for example:

- `MATSe3` = Moran autocorrelation of **electronegativity** at topological distance 3.
- `GATSv6` = Geary autocorrelation of **atomic volume** at distance 6.
- `ATSp2` = Broto–Moreau autocorrelation of **polarizability** at distance 2.

3.2 Other common tokens

- `VSA` → Van der Waals **Surface Area** (Å²).
- `PEOE` → **Partial Equalization of Orbital Electronegativities** (Gasteiger charges).
- `EState` → **Electrotopological state**.
- `BCUT` → Burden eigenvalue descriptors (eigenvalues of weighted adjacency matrix).
- `ATS` → Broto–Moreau **A**uto**T**opological **S**tructure autocorrelation.
- `MATS` → **M**oran **A**uto**T**opological **S**tructure autocorrelation.
- `GATS` → **G**eary **A**uto**T**opological **S**tructure autocorrelation.
- `Chi` / `Chiv` → Connectivity indices (valence versions = `v`, charge-corrected = `ch`, polar-corrected = `pc`).