|  |  |  |
| --- | --- | --- |
| QUERY NAME | DESCRIPTION | SMARTS |
| CH3 (All) | all CH3 groups | [CH3] |
| CH3 singlet (All) | any CH3 isolated spin system (CH3-Cq,O,N,S...) | [CH3][\*H0] |
| CH3-Cq (singlet) | CH3 bonded to a quaternary, sp3 carbon (Cq) | [CH3][CX4H0] |
| CH3-CH (doublet) | CH3 bonded to an alif, sp3 CH | [CH3][CX4H] |
| CH3-CH2 (triplet) | CH3 bonded to an alif, sp3 CH2 | [CH3][CX4H2] |
| CH3 (Isopropyl) | isopropyl groups -CH(CH3)CH3 | [CHX4]([CH3X4])[CH3X4] |
| CH3-Ar | CH3 bonded to an aromatic, sp2, carbon | [CH3]c |
| CH3-Csp2 | CH3 bonded to any sp2 carbon (arom, olef, carbonyl...) | [CH3][CX3] |
| CH3-C=C (Vinyl) | CH3 in a C=C double bond, not aromatic | [CH3][CX3]=[CX3] |
| CH3-CO-O (Acetyl) | CH3 in an acetyl group | [CX4H3][CX3](=[OX1])[OX2] |
| CH3-CO-N (acetamide) | CH3 in an acetamide group | [CH3]C(=O)[#7] |
| CH2=CH- (vinyl) | vinyl group | [CX3H2]=[CX3H1] |
| CH2=Cq di-subst | CH2 terminal | [CX3H2]=[CX3H0] |
| CH3-C-CO- | CH3 two bonds from a carbonyl group (from HMBC) the internal C can be completely substituted, part of a ring, sp2 or sp3 | [CH3][#6][CX3](=[OX1])[\*] |
| CH3-O (All) | all CH3 bonded to an oxygen (metoxyl) | [CH3][OX2] |
| CH3-O-Ar (Methoxy) | only aromatic metoxyl groups | [CH3][O]c |
| CH3-N (All) | CH3 bonded to any nitrogen (arom, alif,...) | [CH3][#7] |
| CH3-N-CH2- | CH3 linked to an aliphatic nitrogen which is linked to a methylene. HMBC, N can be part or rings and attached to anything | [CH3][N][CH2] |
| CH3-N-CH- | CH3 linked to an aliphatic nitrogen which is linked to a methine.HMBC, N can be part or rings and attached to anything | [CH3][N][#6H] |
| CH3-N-CO- | CH3 bonded to an aliphatic nitrogen which is conected to a carbonyl group. HMBC, N can be part or rings and attached to anything | [CH3][N][CX3](=[OX1])[\*] |
| CH2 (All) | any CH2 group (terminal, aliph) | [#6;H2] |
| CH2 (sp3) | only aliphatic CH2 | [CX4H2] |
| CH2 (sp2) | only terminal CH2 | [CX3H2] |
| CH2-O (All) | any CH2 connected to one oxygen (CHECK!!) | [CX4H2]([!#8])[OX2] |
| CH2-(O)O | any CH2 connected to two oxygen atoms (methylenedioxy groups) | [OX2][CX4H2][OX2] |
| CH2-N (All) | any CH2 connected to a nitrogen | [CH2][#7] |
| CH2-N-CH2 | HMBC, N can be part or rings and attached to anything | [CH2][#7][CH2] |
| CH2-N-CH | HMBC, N can be part or rings and attached to anything | [CH2][#7][CH] |
| CH (All) | all CHs in the molecule | [#6;H1] |
| CH sp3 | only aliphatic CHs | [CX4H] |
| CH sp2 (non arom) | only olefinic CHs | [CX3H] |
| CH sp2 (arom) | only aromatic CHs | [H]c |
| CH sp | only alkyne CHs | [CX2H] |
| CH-X (All) | all CH bonded to one heteroatom | [#6H]([#6])([#6])[#7,#8,#16] |
| CH-X sp3 | aliphatic CH bonded to one heteroatom | [CH]([#7,#8,#16])([#6])[#6] |
| CH-x (arom) | CH arom vicinal to a heteroaromatic N, O, S (1JCH<200Hz) | [H][c]([n,o,s])[c] |
| CH-N (All) | CH connected to two carbon atoms (any) and one nitrogne | [CH]([#7])([#6])[#6] |
| CH-O (All) | all CH bonded to one oxygen CHECK! | [#6H]([#6])([#6])[#8] |
| CH-O (sp3) | aliphatic CH bonded to one oxygen | [CX4H]([OX2])([#6])[#6] |
| CH-(X)X (sp3) | all CH bonded to two heteroatoms (including anomeric) | [#6H]([#6])([#7,#8,#16])[#7,#8,#16] |
| CH-(x)x (arom) | CH arom flanked by two heteroaromatic N, O, S (1JCH>200Hz) | [H][c]([n,o,s])[n,o,s] |
| CH-(O)O (All) | all CH bonded to two oxygen atoms (ketal, acetal, anomeric) | [#6H]([#8])([#8])[#6] |
| CH-(O)N | CH bonded to an oxygen, a nitrogen, and a carbon (arom, aliph, olef, ring, ...) | [CH]([#8])([#7])[#6] |
| CH-(O)(O)O | CH bonded to tree oxygen | [#6H]([#8])([#8])[#8] |
| CH (Peptide) | CH in peptide bonds | [CH]([CX3]=O)[NX3]([CX3]=O) |
| CH (Aldehyde) | CH in aldehydes | [CX3H1](=O)[#6] |
| CH=CH cis | cis 1,2-disubst alkene | \*/[CH]=[CH]\\* |
| CH=CH trans | trans 1,2-disubst alkene | \*/[CH]=[CH]/\* |
| CH=Cq tri-subst | 1,1,2-trisubst alkene | [CX3H]=[CX3!H] |
| Cq (All) | all quaternary carbon | [#6;H0] |
| Cq sp3 (All) | only aliphatic quaternary carbon | [CX4;H0] |
| Cq sp3 (spiro) | only Cq in a spyro connection between rings with 4, 5 and 6 members | [X4;R2;r4,r5,r6](@[r4,r5,r6])(@[r4,r5,r6])(@[r4,r5,r6])@[r4,r5,r6] |
| Cq sp2 (All) | any sp2 (olefinic, aromatic, C=O...) quaternary carbon | [CX3,c;H0] |
| Cq sp2 (non arom) | only non-aromatic sp2 quaternary carbon (olefinic and CO) | [CX3H0] |
| Cq sp2 (arom) | only aromatic sp2 quaternary carbon | [cH0] |
| Cq sp (alkyne) | only alkyne quaternary carbon | [CX2;H0] |
| Cq sp (nitrile) | quaternary carbon in nitrile only | [CX2]#[NX1] |
| Cq sp (isonitrile) | quaternary carbon in isonitrile only | [CX1-]#[NX2+] |
| Cq sp (allene) | quaternary carbon in allenes | [$([CX2](=C)=C)] |
| CO carbonyl (All) | all carbonyl groups, used with HSQC (aldehydes), HMBC or 13C | [CX3]=[O] |
| CO carbonyl (Ester/Lactone) | carbonyl from ester and lactone, used with HMBC or 13C | [#6][CX3](=O)[OX2H0][#6] |
| COOH | COOH (not charged) | [OX2H][CX3]=[OX1] |
| OH (All) | all OH groups (alcohol, acid...) | [OX2H] |
| OH (Alcohol) | OH in aliphatic alcohol only | [CX4][OH] |
| OH (Phenol) | only phenolic OH | [c]:[cX3]([OX2H]):[c] |
| OH (Acidic) | OH in acids only | [$([OH]-\*=[!#6])] |
| NH (All) | aliphatic and aromatic NH | [#7H] |
| NH (Arom) | only aromatic NH | [nH] |
| NH, NH2 (amides) | NH or NH2 in amides | [CX3](=O)[#7X3;H1,H2] |
| NH2 (All) | any NH2 group | [#7H2] |
| benz 1-monosubst | ? | [cX3!H]1[cX3H][cX3H][cX3H][cX3H][cX3H]1 |
| benz 1,2-disubst | ? | [cX3!H]1[cX3!H][cX3H][cX3H][cX3H][cX3H]1 |
| benz 1,3-disubst | ? | [cX3!H]1[cX3H][cX3!H][cX3H][cX3H][cX3H]1 |
| benz 1,4-disubst | ? | [cX3!H]1[cX3H][cX3H][cX3!H][cX3H][cX3H]1 |
| benz 1,2,3-trisubst | ? | [cX3!H]1[cX3!H][cX3!H][cX3H][cX3H][cX3H]1 |
| benz 1,3,5-trisubst | ? | [cX3!H]1[cX3H][cX3!H][cX3H][cX3!H][cX3H]1 |
| benz 1,2,4-trisubst | ? | [cX3!H]1[cX3H][cX3!H][cX3!H][cX3H][cX3H]1 |
| benz 1,2,3,4-tetrasubst | ? | [cX3!H]1[cX3!H][cX3!H][cX3!H][cX3H][cX3H]1 |
| benz 1,2,3,5-tetrasubst | ? | [cX3!H]1[cX3H][cX3!H][cX3!H][cX3!H][cX3H]1 |
| benz 1,2,4,5-tetrasubst | ? | [cX3!H]1[cX3!H][cX3H][cX3!H][cX3!H][cX3H]1 |
| benz pentasubst | ? | [cX3!H]1[cX3!H][cX3!H][cX3!H][cX3!H][cX3H]1 |
| benz hexasubst | ? | [cX3!H]1[cX3!H][cX3!H][cX3!H][cX3!H][cX3!H]1 |

Change Log

|  |  |  |  |
| --- | --- | --- | --- |
| Date | Issue | status | Req/Sug by |
| 17/Jul/17 | UNPD InChI and InChIKeys substituted using INDIGO2 InChi Node (KNIME) | Done | Prof. John Blunt |
| 17/Jul/17 | 8541 duplicated structures (using DW) | removed | Zani |
| 14/Jul/17 | Manually correct 40 structures with wrong charges | Done | Zani |
| 14/Jul/17 | Included Indigo2 Standardizer node to correct wrong charges | Done | Zani |
| 5/Jul/17 | Cq sp in DEREP is not giving the correct counts. It is currently including a terminal alkyne as 2 | Fixed | Prof. John Blunt |
| 5/Jul/17 | Cq sp is missing CN and C=C=C | Fixed | Prof. John Blunt |
| 5/Jul/17 | it would be more logical to have CH3 singlet as arising from CH3-C to be consistent with CH3 doublet and triplet. The count of CH3 All - CH3 singlet would reveal other CH3s on O or N or S. | Fixed | Prof. John Blunt |
| 5/Jul/17 | There is currently no count of acetyl groups on N - quite a common feature. Count of amides would also be very useful. | Included | Prof. John Blunt |
| 5/Jul/17 | it is very useful to have counts of 1,1 and 1,2 disub alkenes, trisub alkenes and vinyl groups. Another useful count is fully substituted benzenes | Included | Prof. John Blunt |
| Jun17 | CH-O(all) is counting CH anomeric (O-CH-O) twice | Fixed | Prof. A. Carroll |