Automated strategies to identify compounds on the basis of GC/EI-MS and calculated properties.

Abstract:
The identification of unknown compounds based on GC/EI-MS spectrum and structure generation techniques has been improved by combining a number of strategies into a programmed sequence. The program MOLGEN-MS is used to determine the molecular formula and incorporate substructural information to generate all structures matching the mass spectral information. Mass spectral fragments are then predicted for each structure and compared with the experimental spectrum using a match value. Additional data are then calculated automatically for each candidate to allow exclusion of candidates that did not match other analytical information. The effectiveness of these "exclusion criteria", as well as the programming sequence, was tested using a case study of 29 isomers of formula C(12)H(10)O(2). The default classifier precision resulted in the generation of too many structures in some cases, which was improved by up to several orders of magnitude by including additional classifiers or restrictions. Combining this with the exclusion of candidates based on a Lee retention index/boiling point correlation, octanol-water partitioning coefficients, steric energies, and finally spectral match values limited the number of candidate structures further from over 1 billion without any restrictions down to less than 6 structures in 10 cases and below 35 in all but 3 cases. This method can be used in the absence of matching database spectra and brings
unknown identification based on MS interpretation and structure generation techniques a step closer to practical reality.