

# The calculation of molecular structural similarity: Principles and practice

Peter Willett

*Information School, University of Sheffield, 211 Portobello Street, Sheffield S1 4DP, United Kingdom*

Similarity measure = {Structure representation, Weighting scheme, Similarity coefficient}
----------------------------------------------------------------------------------------------

Measures of structural similarity play an important role in chemoinformatics for applications such as similarity searching, database clustering and molecular diversity analysis. A similarity measure comprises three components: a structure representation; a weighting scheme; and a similarity coefficient. The paper introduces these and describes methods for comparing different measures. The use of similarity measures in chemoinformatics research is illustrated by recent projects in the author's laboratory on: the interactions between a weighting scheme and a similarity coefficient; the design of comparative studies of similarity measures; the use of 2D fingerprints for scaffold-hopping searches; and the registration of orphan drugs for rare diseases