

# Tripod Development

Cheminformatics proving ground

[Home](#) [GInAS](#) [NPC browser](#) [Siphonify](#) [Overview](#) [Screenshots](#) [Jobs](#) [About](#)

## Do structurally similar molecules have similar hash codes?

Molecular hash codes are fixed-length alphanumeric encoding of molecular graphs. They play a key role within chemical data management systems in facilitating (among other things) structural identity and uniqueness validation. While an important component of any hash code generation approach is the structure standardization procedure (something which we briefly [discussed previously](#)), the focus of this post is on the encoding step, i.e., that of generating the hash code given a standardized structure. In the remainder of this post, we first give a brief overview of **spectral hash code**, our proposed multi-resolution hash code based on InChI. We then highlight the utility of the hash code through a number of examples. We conclude with the description of a web resource for converting between PubChem CID, InChIKey, and **spectral hash code**. The complete source code is readily available from our [bitbucket](#) repository. As always, we welcome comments and feedback!

A **spectral hash code** is a 30-character (150-bit) alphanumeric hash string that uniquely encodes an [InChI](#). The hash code has three logical blocks—denoted as  $h_1$ ,  $h_2$ , and  $h_3$ —that progressively encode the InChI graph at finer resolution where  $h_1 \subseteq h_2 \subseteq h_3$  and  $|h_1| = 9$  (45-bit),  $|h_2| = 19$  (95-bit), and  $|h_3| = 30$  (150-bit). Consider, for