

A New Format for the Efficient Storage of Molecular Structures

Peter Csizmadia and Ferenc Csizmadia
ChemAxon Ltd., Budapest

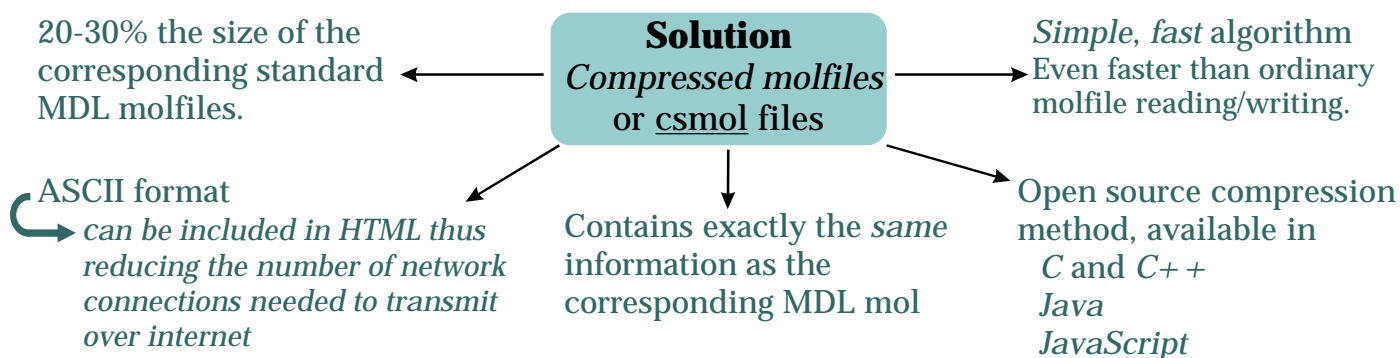


Internet is slow

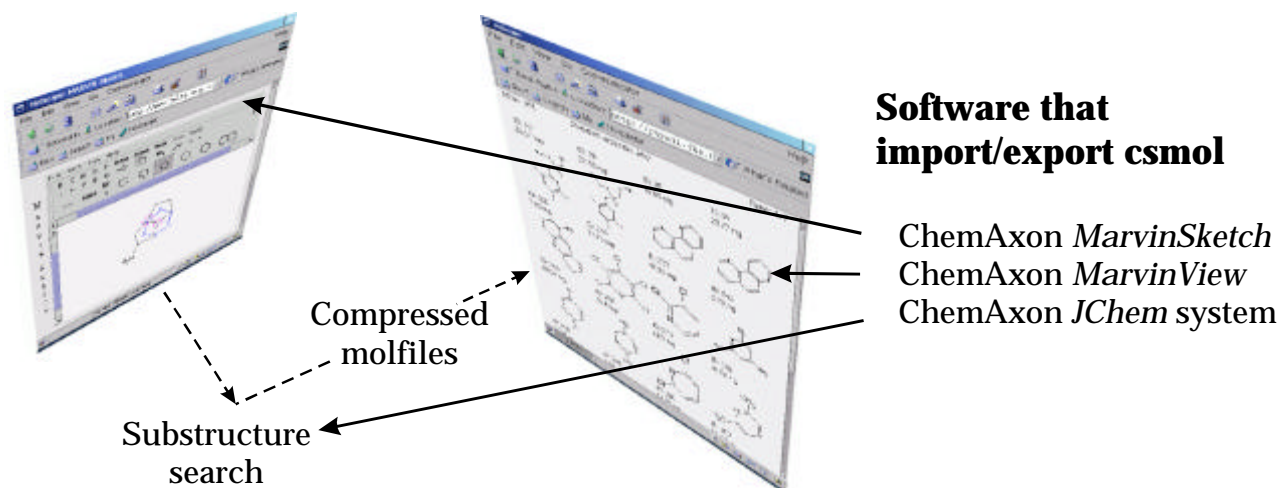
Chemistry-related internet applications need to transfer many molecules in the minimum available time.

Inefficient structure storage in databases

If structures are stored in
↳ a standard format such as MDL mol
↳ the required space is relatively large
↳ some proprietary binary format
↳ information might be lost



Searching in a networked chemical database



The molfile compression method

- ↳ atom and bond blocks are compressed
- ↳ atom coordinates, atom type, charge, stereo information
- ↳ special information in the atom and properties blocks remains uncompressed
- ↳ no information loss
- ↳ 6 bits/character encoding similar to BASE64
- ↳ HTML safe
- ↳ can be combined with usual binary compression methods
- ↳ ZIP, GZIP