

# **CSEARCH-Technologie und Daten am Internet – ein Überblick und Vergleich mit anderen Systemen**

**Wolfgang Robien**

**Institut für Organische Chemie  
Fakultät für Chemie der Universität Wien  
Währingerstr. 38, A-1090 Wien  
Email: [Wolfgang.Robien@univie.ac.at](mailto:Wolfgang.Robien@univie.ac.at)**

Das NMR-Datenbanksystem CSEARCH (1,2) wurde in den letzten Jahren mit einer Rate von etwa 20,000 Spektren/Jahr ausgebaut. Gleichzeitig ermöglichte diese ungeheure Fülle an Daten umfangreiche statistische Auswertungen über die doch recht heterogene Qualität von Literaturdaten. Es wurde daher in der Algorithmenentwicklung konsequent ein Schwerpunkt im Bereich Qualitätsmanagement verfolgt – die Konsequenz war eine intensive Überarbeitung des vorhandenen Datenmaterials mit einer beträchtlichen Qualitätssteigerung. Parallel dazu wurden verschiedene Zugangsmöglichkeiten zu diesem Datenpool von nahezu einer halben Million NMR-Spektren (ca. 750,000 intern) durch neu hinzugekommene Internet-Technologien eröffnet.

Die Möglichkeit C-NMR Spektren mit vernünftiger Genauigkeit innerhalb weniger Millisekunden vorherzusagen, eröffnete die Möglichkeit mit den PUBCHEM-Strukturen (etwa 16 Millionen zu diesem Zeitpunkt) einen Server zur spektralen Ähnlichkeitssuche zu etablieren, wobei die Suchzeiten bei 16,000,000 Referenzspektren mit durchschnittlich 4 Sekunden akzeptabel sind.

Die Zusammenarbeit mit emolecules und Wiley-VCh wurde Anfang 2007 begonnen und sämtliche Strukturen der Systeme ‚SPECINFO‘ und ‚CSEARCH‘ sind dadurch mittels (Teil)Struktursuche zugänglich (3). Die Vorhersagealgorithmen von CSEARCH mit dem gesamten Datenvolumen von CSEARCH und SPECINFO ist als ‚NMRPredict Online Full‘ via MESTRELAB RESEARCH (4) verfügbar.

Seitens der IUPAC wurde der InChI durch den InChIKey ergänzt, die entsprechende Software stand ab September 2007 zur Verfügung. Sämtliche Strukturen, die in CSEARCH verfügbar waren, wurden in InChIKeys konvertiert und damit ein Portal geschaffen, bei dem (auch automatisch) die Existenz von NMR-Daten in computerlesbarer Form zu einer konkreten Verbindung abgefragt werden kann. Innerhalb dieses Portals wurden sowohl kommerzielle Systeme als auch Open-Access Systeme wie CSEARCH, SPECINFO, CHEMGATE, NMRPredict, NMRPredict Online, KnowItall, NMRShiftDB und die in-house Datenbank der Universität Mainz indexiert.

Die Daten innerhalb des Open Source/Open Access Systems ‚NMRShiftDB‘ (5) wurden nach CSEARCH-Kriterien auf Fehler geprüft. Unabhängig wie man die daraus resultierende Diskussion bewertet – die nachfolgende Anstrengung insgesamt 72 Fehler in etwa 3 Monaten zu korrigieren, haben dieser Datenbank sicherlich gut getan.

- (1) <http://nmrpredict.orc.univie.ac.at>
- (2) <http://csearch-nmr-data.blogspot.com>
- (3) <http://chemgate.emolecules.com>
- (4) <http://www.mestrec.com>
- (5) <http://nmrshiftdb.org>