

# НТИ-2002

Москва, 16-18 октября 2002 г.

МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ ПОСВЯЩАЕТСЯ  
50-ЛЕТИЮ ВИНТИ

ИНФОРМАЦИОННОЕ  
ОБЩЕСТВО  
ИНТЕЛЛЕКТУАЛЬНАЯ  
ОБРАБОТКА ИНФОРМАЦИИ  
ИНФОРМАЦИОННЫЕ  
ТЕХНОЛОГИИ

*INFORMATION SOCIETY*  
*INTELLIGENT INFORMATION*  
*PROCESSING*  
*INFORMATION*  
*TECHNOLOGIES*

МАТЕРИАЛЫ КОНФЕРЕНЦИИ  
PROCEEDINGS OF THE CONFERENCE

# **БАНК ДАННЫХ «СТРУКТУРНЫЕ ДАННЫЕ ПО ХИМИИ» ВИНТИ И НЕКОТОРОЕ ЕГО СОПОСТАВЛЕНИЕ С КЕМБРИДЖСКОЙ СТРУКТУРНОЙ БАЗОЙ ДАННЫХ**

М. С. Стуклова, Н. Н. Кочанова, В. П. Соловьев, Н. В. Качурина, А. Г. Сухова  
ВИНИТИ, ИФАВ РАН

## **DATABASE «CHEMICAL COMPOUNDS STRUCTURE DATABASE» AND ITS COMPARISON WITH CAMBRIDGE STRUCTURAL DATABASE**

M. S. Stuklova, N. N. Kochanova, V. P. Solov'ev, N. V. Kachurina, A. G. Sukhova

Database «Chemical Compounds Structure Database» (SD VINITI) is being formed since 1975 and contains more than 6 million chemical structures. This database has been developing as an information platform for scientific researches, for chemical and biological technologies. The database provides the users of an opportunity of the

information searching on various aspects (on PC under WINDOWS): structure and substructure searching, the methods of the researches of the compounds (X-ray or neutron diffraction analysis; NMR, EPR, IR, Raman, UV-VIS spectroscopy; magnetochemistry; electrochemistry; thermodynamics; kinetics; etc.), physical, chemical and biological properties, chemical reactions and the applications of the substances. This database contains the following fields: the bibliography (the full source reference), key words, abstract, the 2D structure diagram of the compound, formula, compound name, special subject terms (key words) about all factual information (properties and applications) in the source. The software of the SD database allows to store, to look through, to search or to print the data, also to create local databases on important fields of the science or concrete technological tasks.

Database fields and their representation for the users in SD VINITI and in the Cambridge Structural Database (CSD CCDC) have been compared for the intensively studied coordination compounds. The Cambridge Structural Database contains crystal structure information for over 257 000 organic and metal organic compounds. All of these crystal structures have been analysed using X-ray or neutron diffraction techniques. For each crystallographic entry in the CSD there are three distinct types of information (bibliographic, 2D chemical connectivity, 3D molecular structure and 3D crystal structure). Available crystallographic data are accessible in SD VINITI in a textual format only. In CSD, as well as in SD, the 2D structure of molecules is submitted, which comparison in the image of the coordination compounds represents doubtless interest. The basic difference in the 2D structure representation in the CD and CSD databases is identical drawing for  $\pi$ - and coordination bonds in the CSD database. This creates considerable difficulties for the correct imaging of the structure and for visual recognition. As a result of this, double bond is represented in some cases by means of single bond to keep the valency of coordinated atom. For example, the complexes of metals (M) with phosphin oxide derivatives are given in CSD as  $R_3P=O-M$  and in SD as  $R_3P=O...M$ . At the analysis of CSD we have detected, that cyclopentadienyl ring has 5 equivalent bonds with metal for the 3D structures, and this ring is connected to metal by one bond in the 2D structure representation; while ethen has two bonds for 2D as well as for 3D structures. It is necessary to note, that the information contents of SD and CSD databases well supplements one another.

The results of the databases comparison are used for the concept development of the SD VINITI, for methods of the information, diagnostic and technical maintenance of the advanced model for the representation of co-ordination compounds in uniform information environment of the researchers, the information and library workers.

Банк данных ВИНТИ «Структурные данные по химии» (БД СД) пополняется на протяжении 27 лет и содержит более 7 млн. химических структур. База данных развивается, как информационная платформа для научных исследований, химических и биологических технологий, предоставляя пользователю в среде WINDOWS возможности поиска информации по различным аспектам: структурный и подструктурный поиск (химическая 2D структура), методы исследований соединений (РСА, ЯМР, ЭПР, ПМР, ИК, КР, УФ, магнетохимия, электрохимия, термодинамика, кинетика и др.), физические, химические и биологические свойства, данные по химическим реакциям и применению веществ. Эта база данных ВИНТИ содержит следующие поля: библиография, ключевые слова, реферат РЖ «Химия», химическую 2D структуру соединения, брутто-формулу, название соединения, предметные термины (ключевые слова) о всей фактографической информации реферируемой работы. Программные средства базы позволяют не только хранить, просматривать, проводить поиск или распечатывать упомянутые данные, но и создавать локальные БД по приоритетным областям науки или конкретным технологическим задачам.

Для того чтобы отчасти проанализировать качественный уровень базы данных ВИНТИ БД СД, был выбран интенсивно изучаемый класс химических веществ - координационные соединения, и проведена сравнительная характеристика представления координационных соединений в БД СД и в Кэмбриджской структурной базе данных (CSD CCDC). CSD является специализированной базой в области кристаллографических структурных данных и содержит информацию о кристаллической структуре более 257 000 органических и металлоорганических соединений. Для каждого соединения в CSD содержатся библиографические данные, молекулярные 2D и 3D структуры и кристаллическая 3D структура. CSD включает все характеристики о пространственной 3D структуре молекулы, манипуляции с ней, поиск и др. В БД СД имеющиеся кристаллографические данные представлены только в текстовом формате. В CSD, как и в БД СД, представлена 2D структура молекул, сравнительная характеристика которой в изображении координационных соединений представляет несомненный интерес. Основным отличием 2D структуры БД СД от CSD является эквивалентное представление в CSD валентных и координационных связей, что создает немалые трудности для правильного изображения структуры и ее визуального восприятия. Так, двойная связь в некоторых случаях изображается посредством одинарной, с тем чтобы не вступать в конфликт с валентностью донорного атома. Например, комплексы металлов (M) с производными фосфиноксида изображаются в CSD, как  $R_3P=O-M$ , а в БД СД, как  $R_3P=O...M$ . При анализе CSD мы столкнулись с неоднозначностью в изображении металлоорганических соединений: в формате 3D циклопентадиенил имеет 5 равнозначных связей с металлом, а в 2D формате - кольцо связано с металлом одной связью; в то время как этен и в 2D, и в 3D-формате имеет две связи. Следует отметить, что информационное содержимое баз данных СД и CSD хорошо дополняет одно другое.

Результаты проведенного анализа использованы в процессе разработки концепции развития БД СД, методов и средств информационно-диагностического и технического обеспечения усовершенствованной модели представления координационных соединений в единой информационной среде химиков-исследователей, информационных и библиотечных работников.