ISSN 0005-1055, Automatic Documentation and Mathematical Linguistics, 2015, Vol. 49, No. 6, pp. 213–220. © Allerton Press, Inc., 2015.
Original Russian Text © O.M. Nefedov, L.M. Koroleva, S.V. Trepalin, Yu.E. Bessonov, N.I. Churakova, 2015, published in Nauchno-Tekhnicheskaya Informatsiya, Seriya 2, 2015, No. 12, pp. 17–25.

The Development of an Integrated System for Structural Chemical Information

O. M. Nefedov[a], L. M. Koroleva[b], S. V. Trepalin[c], Yu. E. Bessonov[b], and N. I. Churakova[b]

[a] *The Presidium of the Russian Academy of Sciences, Moscow, Russia*
e-mail: onefedov@ras.ru
[b] *All-Russian Institute of Scientific and Technical Information, Russian Academy of Sciences, Moscow, Russia*
[c] *Institute of Physiologically Active Substances, Russian Academy of Sciences, Chernogolovka, Russia*

Received September 16, 2015

Abstract—Molecular and reaction structural databases on chemistry are considered. The history of how the software-technological complex of the Database of Structural Information on Chemistry of the VINITI RAS developed is described. The potential of searching for structural information based on the generalization of structural diagrams is shown.

Keywords: chemical information, Database of Structural Information on Chemistry of the VINITI RAS, search for chemical information, Internet access, algorithms for generalization of polycyclic structures, structural diagrams

DOI: 10.3103/S0005105515060060

INTRODUCTION

The presence of information on the structure and properties of substances is an integral part of chemical information that is needed by specialists in chemistry. The subject field of chemistry, along with abstract databases, is widely represented in the world information space by various information resources, including databases of structural information on chemistry that contain information on the structures and properties of chemical compounds. The largest center that generates chemical information is the Chemical Abstracts Service (CAS), which is a department of the American Chemical Society. Since 1957, the CAS has been creating the CAS REGISTRY^{SM} specialized registration system, which contains the structural information on individual organic and inorganic chemical compounds, coordination compounds, alloys, minerals, polymers, salts, as well as mixtures of chemical substances. In addition, this information system contains information on more than 66 mln bio-sequences. More than 15000 chemical compounds are registered every day at the CAS REGISTRY^{SM} [1].

The latest achievements of information technologies permit the processes of processing, storing, and searching for chemical information to be significantly improved and optimized. Thus, for example, the Elsevier Company has created its own web-resource of structural chemical information (Reaxys®) that provides access to two databases: Beilstein and Gmelin. The former contains information on the structure, properties, and chemical reactions of organic compounds, while the latter contains information on inorganic compounds [2]. The brief characteristic of some web resources on chemistry that provide information on the molecular structure, chemical reactions, and properties of chemical compounds is presented in Table 1.

In Russia, the Database of Structural Information on Chemistry (below, the SI Database), which is one of the world's largest databases on chemistry, is generated at the VINITI RAS. It contains more than 9 mln chemical structures, more than 4 mln chemical reactions, and 15 mln properties of chemical compounds. The SI Database started to be formed in 1975, and in 1998 the VINITI RAS obtained the certificate (no. 980007 as of January 26, 1998) about the official registration of the SI Database at the Russian Agency for Legal Protection of Computer Software Programs, Databases, and Integrated Circuit Layouts (RosAPO). The SI Database is annually replenished with approximately 180000 compounds and 100000 reactions from more than 6000 documents. The major property of the information contained in the SI Database is the fact that this is information on individual low-molecular-weight, hetero-organic, and coordination compounds, as well as low-molecular-weight natural compounds and their synthetic analogs.

The major principles for processing structural chemical information were developed in 1963 by the VINITI RAS at the "Chemistry" Department in the form of the so-called arbitrary-block coding system (or the AB-code) [3]. The AB-code is an instrument for presenting information on chemical compounds in

Table 1. Molecular and recreated structural databases on chemistry (the data for 2015)

Name	Number of compounds	Number of reactions	Producer	Address of a web resource
CAS REGISTRY^{SM} CASREACT$^{®}$	101 mln	67.5 mln	CAS	https://www.cas.org
Beilstein	11.7 mln	23 mln	Elsevir	http://www.elsevier.com/solutions/reaxys
Gmelin	2.4 mln	1.8 mln	Elsevir	http://www.elsevier.com/solutions/reaxys
Spresi	5.7 mln	4.4 mln	Infochem	http://www.infochem.de/products/databases/spresi.shtml
Index Chemicus$^{®}$	2.6 mln		ISI Thomson Scientific	http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/index-chemicus.html
Current Chemical Reactions®		880 ths		http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/current-chemical-reactions.html
Cambridge Structural Database	750 ths		The Cambridge Crystallographic Data Centre	http://www.ccdc.cam.ac.uk
Inorganic Crystal Structure Database	177 ths		FIZ-Karlsruhe	https://icsd.fiz-karlsruhe.de

* In 2014, the CAS REGISTRY registered a greater amount of chemical compounds than for the entire period from 1965 to 1990.

the textual form by describing the fragments of chemical structure and interconnections between them. This coding system permitted two-dimensional and three-dimensional structures of chemical compounds to be presented as linear records, which could be implemented using the equipment and software of this time.

Subsequently, the developments on the creation of the automated information system (AIS) on chemistry were carried out in close cooperation with scientists and specialists from the institutes of the Academy of Sciences, Universities and higher education institutes of the Soviet Union, Central Institute of Information and Documentation (CIID), and the Centralized Processing of Information on Chemistry People's Enterprise (CICh) of GDR. As a result, the SPRESI Automated System of Search for Structural Chemical Information (Speicherung und Recherche strukturischer Information: Storage and Search of Structural Information) was developed by 1972. The so-called fragmentary code was proposed for the operation of the SPRESI System. The Spresi format is a binary code of a chemical compound, which is readable only using special software. This being the case, chemical compounds were coded manually using the system of the arbitrary-block code, according to which a table of connections was automatically built, and the SPRESI fragmentary code was automatically developed from it [4, pp. 176–177]. Consequently, the SI Database was created in 1975–1996 at the VINITI based on the SPRESI software.

In 1996, the VINITI RAS developed and introduced a new domestic CBASE16 software shell (Chemical Base, 16-class version), which was designed to introduce and process structural chemical information based on the graphic input of structural data using the CHED structural editor that was developed by S.V. Trepalin [5–7]. The CBASE16 software shell enabled the professional and qualitative input, processing, storage, and search of information about chemical compounds and reactions, in which these compounds take part. However, the impetuous development of information technologies, software, and computer engineering led to the obsolescence of the developed CBASE16 software shell. Therefore, the VINITI RAS developed the CBASE32 software complex, which is an improved 32-class version of the CBASE software shell. A significant distinction of CBASE32 from the previous version is that not only single-stage, but also multi-stage chemical reactions are presented and automatically entered into the database. This presentation of reactions enables the most complete description of experimental details from an original document and makes it possible to reflect the actual sequence of reaction stages that yield a target product [8]. In addition, the system for automated recognition of the configuration of asymmetric centers of chemical compounds that takes account of the spatial

Table 2. The general characteristic of the structural chemical information of the SI Database of the VINITI RAS

Period	Structures chemical compounds (%)	Chemical reactions (%)	Form of presentation data
1975–1995	62.8	70.7	AB-code, Spresi, T-graph
1996–2009	27.0	20.1	CBASE16, T-graph
2010–2014	10.2	9.2	CBASE32, SDF, RDF
Total	100	100	

structure properties of almost all types of indexed compounds was improved for the adequate and unambiguous presentation of stereochemical information [9]. The developed software also enabled the implementation of unique information models of coordination compounds and hetero-organic compounds that are significantly superior to the common world standards in their descriptive potential. In order to improve the quality of forming the SI Database and the efficiency of searching for structural chemical information, the CBASE32 software complex was supplemented with a new component, viz., a directory of chemical compounds that are the most frequently occurring in chemical literature [10]. The improved CBASE32 software-technological complex for processing, storing, recognizing, and using chemical information while creating the SI Database enables the efficient use of programs for searching for both point chemical structures and structural chemical information on a fragment of structure. The developed and introduced instruments for searching for structural chemical information enabled the interactive access through the Internet to the data array, which is available in the SI Database [11, 12]. The relative volumes of structural chemical information and formats of their presentation in the SI Database on Chemistry of the VINITI RAS are presented in Table 2.

As Table 2 shows, more than 60% of the chemical structures that are contained in the SI Database were processed by the mid-1990s. Converters for presenting the structures stored in the AB-code and Spresi formats in the form of T-graphs were developed in the late 1990s. A T-graph means a linear (textual record) of the structure of a chemical compound (a molecular graph) [13]. A T-graph makes it possible to code organic, inorganic, and complex compounds. Most retrofund structures (accumulated for the period from 1975 to 1995) are stored as T-graphs as a result of the conducted works.

It is known that chemical information simultaneously features not only its rapid growth, but also slow aging. This being the case, the major objects of studies are chemical compounds and reactions, and practical tasks include the synthesis of new compounds with necessary and assigned properties and the development of new chemical reactions based on those that are already known. The determination of the modern state of scientific research and developments in the field of chemistry does not depend on the depth of searching for special and professional information.

The data in the CBASE16 and CBASE32 formats contain information on the images of molecular structures and can be unloaded in files of the SDF and RDF formats (for structures and reactions, respectively). A converter for transforming information from the CBASE16 format into the CBASE32 format has been developed and the data that accumulated in the period from 1996 to 2009 are transformed into the CBASE32 format using this converter. The absence of information on the images of structural formulas in the retrofund for the period from 1975 to 1995 requires the transformation into the common CBASE32 format.

The full presentation of a chemical compound requires the graphic imaging of a structural formula. There is an opportunity to convert the data about atoms and topology of molecules presented by the AB-code or T-graphs into the CBASE32 or SDF format. However, the automated formation of an acceptable two-dimensional image (visualization) of the structural form of a molecule only based on topology (a table of connections) is a rather complex task. In this regard, the visualization of retrofund structures must transform the greatest possible number of structures with minimal expenses and losses.

THE PROBLEMS OF VISUALIZATION: THE LIBRARIAN APPROACH

The librarian approach has been known for a fairly long time [14]. Its essence is that a set (a library) of previously imaged patterns is created. Patterns usually correspond to the cyclic fragments of a chemical structure and their images satisfy the standard requirements for images of chemical structures. The topology of a molecule is analyzed for the presence of fragments, which are isomorphic to patterns from the library. After all of the fragments are identified, the image of the entire structure is assembled [15, 16]. This principle was the basis for the experimental software complex that was developed by the VINITI RAS in 1993. The so-called chemical blocks that are maximally connected cyclic components of a molecular graph were used here as patterns. The complex permitted the structures of chemical compounds to be automatically imaged qualitatively, if all of the chemical

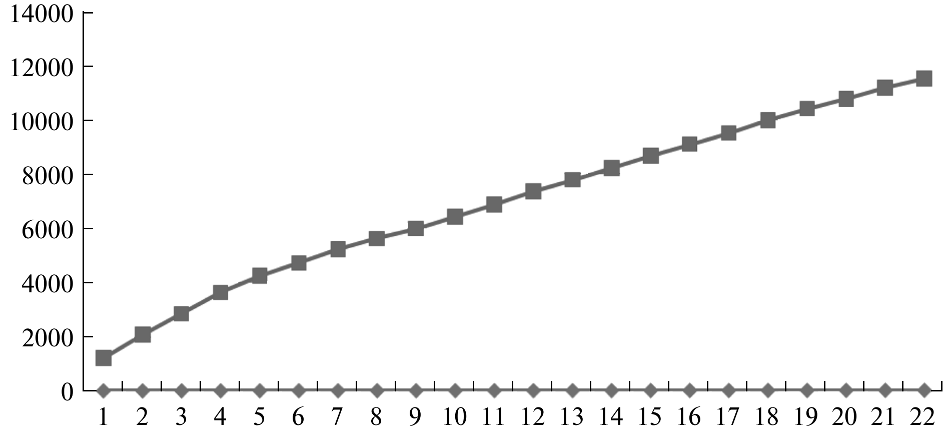

Fig. 1. The quantity of unique chemical blocks-patterns per number of the AJ of Chemistry for 1992.

blocks were included in the pattern library. If some chemical block was absent in the library, it was automatically imaged using the heuristic algorithm. However, such an image required subsequent manual correction. The research on the visualization of chemical structures using this experimental software complex, which was conducted based on the data array of 1992, showed that the number of different chemical blocks grew stably, as the number of compounds in the database increased (Fig. 1). The X-axis in Fig. 1 presents the numbers of the Abstract Journal of Chemistry for 1992, which correspond to the order of adding structures to the SI Database during a year. Each number presents approximately 20000 structures. The Y-axis presents the total number of unique chemical blocks that are needed for the visualization of structures of chemical compounds introduced into the SI Database. Extrapolating the curve for the entire database (9161021 compounds), we obtain the value, which is approximately equal to 250000 chemical blocks.[1]

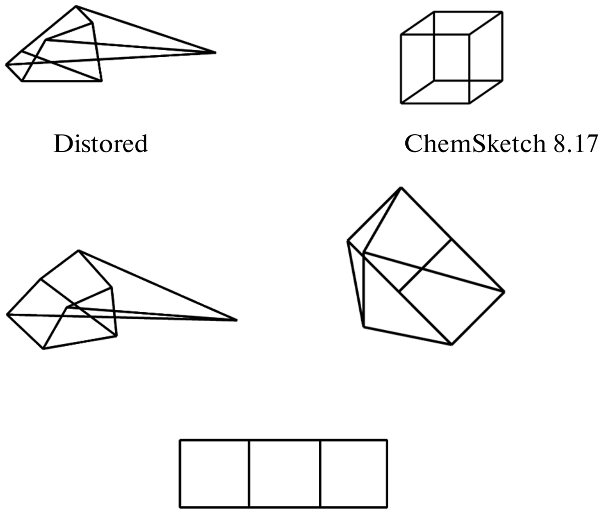

Fig. 2. The result of performing the Clean command in popular programs ChemSketch 8.17, ISIS/Draw 2.5, ChemDraw Ultra 8.0, and JChemPaint 2.0.6 Pre.

Let us consider the method for building the images of polycyclic compounds we have developed, which is also based on the librarian principle. However, the pattern library of this method has a much lower size (approximately 200 fragments).

THE GENERATION OF STRUCTURAL DIAGRAMS IN POLYCYCLIC COMPOUNDS

The conducted scientific research on the transformation of information for 1975–1995 into the CBASE32 format enabled the development and testing of the mechanism for the generation of structural diagrams into polycyclic compounds.

In order to demonstrate the structures that are stored as linear codes, it is necessary to generate 2D-Cartesian coordinates of atoms, i.e., a structural diagram. It is evident that this structure has an infinite number of solutions. Chemical structures look the best visually if all of the lengths of bonds are equivalent and the angles between bonds are $2\pi/3$. This cannot always be performed, but must be aspired to. The literature describes the algorithms for the generation of structural digrams [17, 18].

Most of the editors of chemical structures have the “Clean Structure” command [19–21]. When this command is performed, a structural diagram is generated, in which the lengths of bonds are most often equal and valence angles are optimal. However, the problems in the generation of polycyclic structures arise in the three most frequently used commercial editors of chemical

[1] This is true if there is no “saturation,” i.e., the number of unique chemical blocks does not become a constant at some time. The study of this problem is of scientific interest and is an individual subject

structures: ISIS/Draw [20], ChemDraw [21], ChemSketch [19], and free JChemPaint [22]. Figure 2 shows a cuban (C_8H_8), in which 2D-coordinates of atoms were determined randomly. The optimal structure was generated in different chemical editors for such a “distorted” structure by the “Clean Structure” command. The results are shown in the same figure. As we see, only ChemSketch (ACD Labs) generates suitable coordinates of atoms. However, ChemSketch cannot create an optimal structure in the case of a complicated task either.

In order to solve the problem of showing polycyclic structures, it is suggested that a small library of fragments with previously assigned coordinates of atoms should be used. Fragments from the library are searched for to be embedded into a compound, and, if they are found, all found atoms are assigned coordinates of atoms in a fragment with a suitable scaling factor, shift, and rotation in order to reach an optimal configuration. Since fragments are used to image any structures, all atoms and bonds are regarded as equivalent and compatible with any atoms (bonds) in the structure. Correspondingly, a compact data warehouse can be created; this is important for the generation of 2D-coordinates.

Examples of the used fragments are presented below (Fig. 3). Fragments include both polycyclic structures (adamantane and noradamantane) and large cycles (cyclooctodecane). The inclusion of large cycles makes their imaging better visually. When a cycle is imaged as a convex polyhedron with a size of over 12 atoms, the angles between the bonds on carbon atoms cease to be noticeable and it becomes difficult to make a visual estimate of the size of the cycle. If there are several ways to embed a fragment into a structure, the mapping of atoms after searching for graph embedding is random. Any symmetric fragments have several ways to embed a fragment into a structure. If structural diagrams are generated, other fragments can be joined to an assigned fragment.

This being the case, the location of their junction can be chosen in arbitrary places in equivalent atoms, including “dense” places, where there are many atoms and bonds, and it does not seem possible to add new ones. Examples of badly generated structural digrams are presented in Fig. 4. In order to solve this problem, the atoms to which other fragments cannot be joined are marked in the fragments. Such a fragment is shown in Fig. 3: adamantane with three marked atoms and cyclooctodecane. However, adamantane itself is stored in the library of fragments for the possible generation of the structural diagram of polysubstituted adamantane.

When the embedding of a fragment into a structure is searched for, the marked atoms are regarded as exactly coordinated. The atoms that are found in a chemical structure must have the same number of neighbors and joined bonds, as that indicated in a fragment. In case of such a search, it is guaranteed that new neighbors will not be added to “dense” atoms during generation of structural diagrams. To make a correct search, a library must be sorted according to the size of fragments. The search starts with the fragment that contains the maximal number of atoms; if the quantity of atoms is equal, it must start with the fragment with the maximal number of bonds. If there is a fragment both with and without atoms marked for the absence of joining of other fragments, the search starts with the marked fragment. Otherwise, there may be success in imaging small sections of a compound but failure to find the fragment that is the most proximate to the structure.

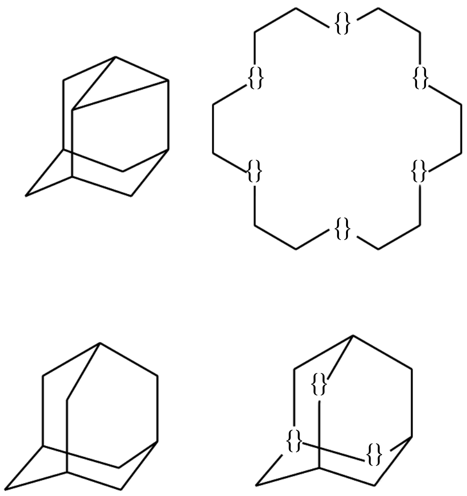

Fig. 3. Examples of fragments that were used to generate 2D-coordinates of atoms. The atoms that cannot be bound with other atoms are marked with {}.

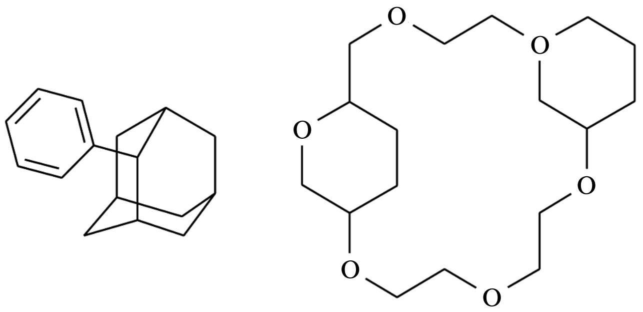

Fig. 4. An example of incorrectly generated 2D-atom coordinates.

The following problem that is to be solved is to use several fragments from the library in order to generate structural diagrams or to multiply use the same fragment. The solution requires a huge library of fragments. For example, in order to image 1,1'-diadamantane, it must be wholly kept in a library of fragments.

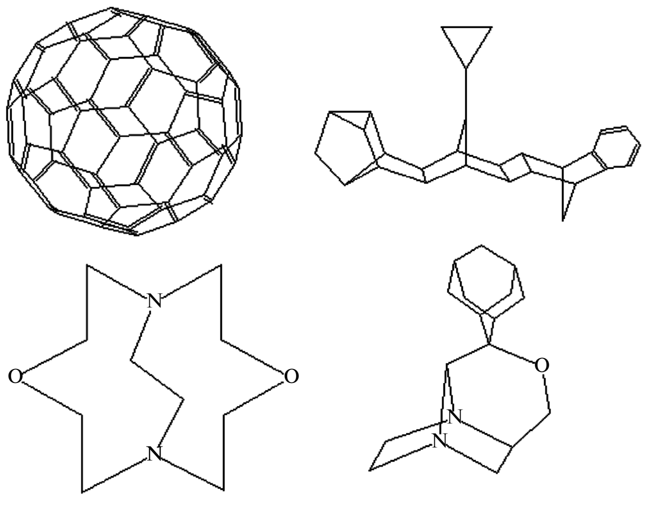

Fig. 5. An example of structural diagrams that were generated from linear strings.

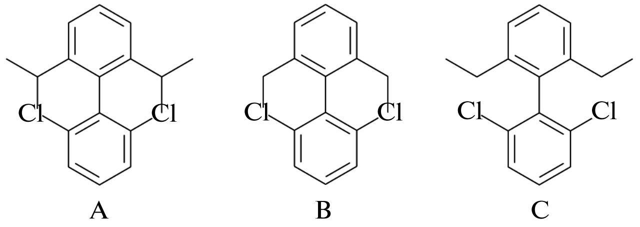

Fig. 6. Structures that contain overlapping fragments (A, B), and a structure without overlapping (C).

Multiple searches for fragments in a target molecule must be used for this purpose. Correspondingly, if the coordinates of atoms have already been reconstructed, then they are marked and become inaccessible for searching. The marked atoms give “false” when it is attempted to combine them with any of atoms of a fragment in the base.

The sequence of describing a structure according to fragments can be presented in the form of the following stages.

1. A set of minimal cycles in a compound is calculated using the algorithm from [23] and is saved in the list.

2. A search is made over the library of fragments, and, if a fragment is found, the coordinates of the found atoms are reconstructed. If a fragment is not found, then a cycle of maximal size from the list that was obtained at point 1 is imaged. If there are no cycles, then a pair of connected maximally coordinated atoms is used as an original fragment to generate structural diagrams.

3. A fragment is searched for in the library among the atoms whose coordinates were not determined. This fragment must be bound with an atom for which 2D-coordinates are already known.

4. If the fragment is found, then the coordinates of atoms of the structure that correspond to the found fragment, become known. Then point 3 is next, i.e., the next fragment is searched for. If there are no fragments bound with the atom with the known coordinates, then point 5 is the next one.

5. All cycles from the list (point 1) for which there is at least one atom with known coordinates are added to the structure. If the coordinates of one atom are known, then a spirocycle with standard lengths of bonds and angles determined from its size is added. If the coordinates of two bound atoms are known, then the condensed cycle with the bond length that is equal to the known bond length and 2pi/N angles (N is the size of the cycle) is added. If the coordinates of three and more atoms (a polycycle) are known, then the chain is closed; in this case, if possible, new are atoms are located so that bonds do not intersect with already available bonds. Further, point 3 is next. If the coordinates of none of the atoms were determined in this point, then point 4 is next.

6. Coordinates are calculated for the atoms connected by acyclic bonds, for which coordinates are known. The standard bond length and hexagonal system of coordinates (the angle between the axes is 2pi/3) is used. If there is a long chain of atoms, then coordinates are calculated for only one atom bound with the atom, whose coordinates were calculated when performing points 2–5. If the coordinates of at least one atom are determined in this point, then point 3 is next; otherwise, point 7 is next.

7. It is determined whether coordinates were obtained for all atoms. If they were, then the process finishes. If they were not, then the compound is an unbound graph, which contains two or several fragments. The process repeats starting with point 1 for the next fragment; this is done until they are completely covered.

The generated structural diagrams for some polycyclic compounds are presented in Fig. 5.

It certain that polycyclic structures for which the described algorithm generates bad diagrams can be found. However, it is possible to add them to the library of fragments and then generate qualitative structural diagrams.

OVERLAPPING FRAGMENTS

While structural diagrams are generated, individual atoms and bonds are often overlapped if there are bulky structural substitutes. In many cases overlapping can be avoided if some fragments are rotated around acyclic bonds by 180 degrees. However, for example, for 2,6-diisopropyl-2',6'-dichlorbiphenyl, as is shown in Fig. 6A, the overlapping of chlorine atoms with a methyl group cannot be avoided using any rotation. In this case, it is necessary to change the lengths of some bonds or use nonstandard angles between bonds, which leads to the generation of a low-quality structural diagram. Meanwhile, a high-quality diagram (Fig. 6D), in which overlapping atoms and bonds are absent, can be generated for the overlapping 2,6-diethyl-2',6'-dichlorbiphenyl (Fig. 6C) by means of rotations around C—C bonds.

A slightly simplified algorithm was used in order to solve this problem, namely, the rotation around acyclic bonds. Initially, when a structural diagram is generated, new atoms of a chain are added towards its growth. This yields chains of a maximal length. If in this case there are overlapping fragments, there is a chance to remove the overlapping by the rotation around acyclic bonds. The number of non-equivalent substitutes is initially determined for all acyclic bonds. In this case, a spherical fragment is taken for each atom in a molecule, as was described for the generation of the HOSE code [24]; moreover, the radius of this fragment is taken to be equal to the topological length of the molecule. After this, atom-centered indices are counted [25]. If the indices coincide for a pair of atoms that are joined to one of atoms in case of an acyclic bond, then they are equivalent and the rotation around such a bond is useless. The list of such bonds is then formed and rotation by 180 degrees is implemented for all possible rotation combinations. The total number of rotations is 2^N (N is the number of rotating bonds). In order that the time of the calculations not be very long, the maximal number of rotating bonds was limited to 16. Within the framework of this study, the structural diagram with overlapped bonds and atoms was not corrected for fragments with a large number of rotating bonds.

CONCLUSIONS

1. The introduction of new information technologies for processing structural chemical data and the development of the modern software-technological complex ensure the rational use of the entire information array of the SI Database of the VINITI RAS for a period of more than 40 years.

2. The specific property of the SI Database of the VINITI RAS is that the principles for selecting, processing, and storing the information about individual low-molecular organic, hetero-organic, and coordination compounds, as well as low-molecular-weight natural compounds and their synthetic analogs are strictly observed.

3. The task of generating structural diagrams in polycyclic compounds from linear codes into the graphic image of the structure of a chemical compound has been solved.

4. A glossary, which is of independent significance for specialists not only as a reference book, but also for organizing the formal logical control at all information processing stages, was formed within the framework of improving the linguistic support of the SI Database in order to standardize and normalize the description of structural chemical data.

5. The performed modernization of the SI Database of the VINITI RAS permits the available information array to be brought to uniformity and ensures the possibility of its interactive use for a diverse contingent of users.

ACKNOWLEDGMENTS

The work was financially supported by the Russian Foundation for Basic Research, project no. 13-07-00488.

REFERENCES

1. Chemical Substances—CAS REGISTRY. http://www.cas.org/content/chemical-substances. Cited September 14, 2015.
2. Reaxys and Reacxys Medical Chemistry. Chemistry research solutions. http://www.elsevier.com/solutions/reaxys. Cited September 14, 2015.
3. Vleduts, G.E. and Geivandov, E.A., *Avtomatizirovannye informatsionnye sistemy dlya khimii* (Automated Information Systems for Chemistry), Moscow: Nauka, 1974.
4. Chernyi, A.I., *Vserossiiskii institut nauchnoi i tekhnicheskoi informatsii: 50 let sluzheniya nauke* (Russian Institute for Scientific and Technical Information: 50 Years of Service to Science), Moscow: VINITI, 2005.
5. Alfimov, M.V., Avakyan, V.G., Trepalin, S.V., Voronezheva, N.I., and Churakova, N.I., A universal program shell for creating databases of chemical compounds and reactions, *Dokl. Akad. Nauk.*, 1999, vol. 366, no. 5, pp. 639–642.
6. Trepalin, S.V. and Yarkov, A.V., CheD, chemical database compilation tool, Internet server and client for SQL servers, *J. Chem. Inf. Comput. Sci.*, 2001, vol.41, pp. 100–107.
7. Trepalin, S.V., Gerasimenko, V.A., Kozyukov, A.V., Savchuk, N.Ph., and Ivaschenko, A.A., New diversity calculations algorithms used for compound selection, *J. Chem. Inf. Comput. Sci.*, 2002, vol. 42, pp. 249–258.
8. Voronezheva, N.I., Trepalin, S.V., Churakova, N.I., Nechaeva, K.S., and Koroleva, L.M., A system for representation and input of information about multi-stage chemical reactions by means of software complex CBASE32, *Nauchn.-Tekhn. Inform., Ser. 2. Protsessy Sist.*, 2005, no. 7, pp. 7–11.
9. Nemirovskaya, I.B., Trepalin, S.V., and Koroleva, L.M., Presentation of stereochemical information in the structural database of VINITI, *Nauchn.-Tekhn. Inform., Ser. 2. Protsessy Sist.*, 2006, no. 4, pp. 1–6.
10. Voronezheva, N.I., Trepalin, S.V., Churakova, N.I., Nechaeva, K.S., and Koroleva, L.M., Glossary as an element of data input standardization in the CBASE32 program complex, *Autom. Doc. Math. Linguist.*, 2007, vol. 41, no. 3, pp. 124–129.
11. Nefedov, O.M., Trepalin, S.V., Koroleva, L.M., and Bessonov, Yu.E., Quick search of exact chemical struc-

tures in large databases using InChI Key coding of structures, *Nauchn.-Tekhn. Inform., Ser. 2. Protsessy Sist.*, 2013, no. 12, pp. 27–33.

12. Nefedov, O.M., Trepalin, S.V., Koroleva, L.M., Bessonov, Yu.E., and Churakova, N.I., A base of structural data on chemistry of VINITI RAS: Problem of search in structural fragments, *Nauchn.-Tekhn. Inform., Ser. 2. Protsessy Sist.*, 2014, no. 12, pp. 19–29.
13. Aref'ev, V.B., Some features of the mathematical concept of molecular graph representation in the registration system of chemical compounds of VINITI, *Materialy konferentsii NTI-99* (Proc. Conf. NTI-99), Moscow, 1999, pp. 24–26.
14. Carhart, R.E., A model-based approach to the teletype printing of chemical structures, *J. Chem. Inf. Comput. Sci.*, 1976, vol. 16, pp. 82–88.
15. Dittmar, P.C., Moskus, J., and Couvreur, K.M., An algorithmic computer diagrams, *J. Chem. Inf. Comput. Sci.*, 1977, vol. 17, no. 3, pp. 186–192.
16. Leonenko, Yu.P., Software package for automatic imaging of molecular structures, in *Voprosy algoritmicheskogo analiza strukturnoi informatsii. Vychislitel'nye sistemy* (Problems of Algorithmic Analysis of Structural Information. Computer Systems), Novosibirsk, 1987, vol. 119, pp. 102–111.
17. Helson, H.E., Structure diagram generation, *Rev. Comput. Chem.*, 1999, vol. 13, pp. 313–398.
18. Fricker, P.C., Gastreich, M., and Rarey, M., Automated drawing of structural molecular formulas under constraints, *J. Chem. Inf. Comput. Sci.*, 2004, vol. 44, pp. 1065–1078.
19. MDL/CrossFire. http://www.mimas.ac.uk/crossfire/autonom.html. Cited August 24, 2015.
20. MDL^{TM} ISIS Draw 2.5. http://www.mdl.com/products/framework/isis_draw/index.jsp. Cited August 24, 2015.
21. Cambridge ChemDraw Ultra 8.0. http://www.cambridgesoft.com/products/family.cfm?FID=2. Cited August 24, 2015.
22. Steinbeck, C., Han, Y., Kuhn, S., Horlacher, O., Luttman, E., and Willighagen, E., The chemistry development kit (CDK): An open-source java library for chemo- and bioinformatics, *J. Chem. Inf. Comput. Sci.*, 2003, vol. 43, pp. 493–500.
23. Figueras, J., Ring perception using breadth-first search, *J. Chem. Inf. Comput. Sci.*, 1996, vol. 36, pp. 986–991.
24. Bremser, W., HOSE – a novel substructure code, *Anal. Chim. Acta*, 1978, vol. 103, pp. 355–365.
25. Trepalin, S.V., Yarkov, A.V., Dolmatova, L.M., Zefirov, N.S., and Finch, S.A.E., Windat: an NMR database compilation tool, user interface and spectrum libraries for personal computers, *J. Chem. Inf. Comput. Sci.*, 1995, vol. 35, pp. 405–411.

Translated by L. Solovyova