Storage Techniques and Analysis Tools Founding a Knowledge-Based System for Conformation Prediction†

Elke Lang*

Deutsches Krebsforschungszentrum, Zentrale Spektroskopie, Im Neuenheimer Feld 280, 69120 Heidelberg, FRG

Jürgen Brickmann

Technische Hochschule Darmstadt, Institut für Physikalische Chemie, Petersenstrasse 20, 64287 Darmstadt, FRG

Received March 8, 1993

A conformation analysis system has been created that operates on the basis of a relational database network containing spectroscopic, crystallographic, and substance data, where a similarity-oriented structure retrieval method allows three-dimensional structures of the crystallographic database to be found by graphical input of a two-dimensional structure template and specification of the required degree of similarity. The set of hits is submitted as a whole to geometric conformation analysis by indicating the structure region which is common to all of the hit structures and specifying the positions and kinds of internal coordinates within the common region that should be computed. Results can be tabulated in files for further interpretation. This conformation analysis system helps in finding convenient starting structures for molecular modeling and in revealing the conformative variability of specified structure classes for the detection of leads for the conformation prediction of similar structures in question.

INTRODUCTION

Three-dimensional structure elucidation of biologically active compounds by means of spectroscopic methods and, related with this challenge, the detection of structure activities and computer-aided drug design using molecular modeling methods[1] have become very important working fields. Rapid development in both concerned areas, NMR spectroscopy and molecular modeling, has opened new application possibilities.

One more stage in this improvement is the use of structure database techniques for conformation retrieval and analysis, which is practiced at the Department of Spectroscopy at the German Cancer Research Center (DKFZ). Traditionally, molecular modeling methods have been applied to singular structures for the investigation of their conformational behavior and related properties. Soon it turned out that a collection of molecular structures with reliable 3D information would be able to support molecular modeling studies in many regards if it could provide flexible substructure search and appropriate conformation analysis facilities. Very often, the Cambridge Crystallographic Data Files[2] as the largest collection of experimentally obtained 3D structure coordinates are chosen as 3D structure source. For the intensive and fruitful use of an additional data source, however, it is indispensable to integrate it into the existing database environment. Only a homogeneous data access which spans all involved databases can provide synergistic effects and quality enhancement of the network of databases as a whole. Therefore, 3D structure access must be consistent with existing structure retrieval techniques in order to profit from their efficiency.

Current methods for the detection of pharmacophoric patterns and the analyzation of conformational freedom in 3D databases[3] gain atom coordinates by searching in a database of three-dimensional structure information with geometric algorithms. Database searching based on geometric properties needs the derivation of "bit screens" (description of the structure features in form of a bit map) for the sake of fast search and conformative flexibility of the queries. These bit screens must be produced once by a time-consuming process before the database can be used for structure retrieval. The definition of 3D bit screens is very crucial and can be a limiting factor for completeness and exactness of the result. Therefore, a broad range of 3D bit screen sets in combination with various geometric search algorithms have been reported. The database search is performed in at least two stages. The first stage is a fast screenout of entries that do not match the required bit screen pattern. The second and further steps are time-consuming geometric analysis algorithms applied to the coordinate set of each of the structures which have passed the first step. Bit screen sets can be derived from atom distance distributions of calibration data.[4] Other approaches tabulate the chemical function (e.g. H donor/acceptor, charge) of certain substructures as bit screens[5] or relate to the topological nature of the molecular structure.[6] The first stage can be followed by a distance screenout[7] or directly by geometric comparison of the required coordinate pattern with the coordinate sets of the entries. Several types of algorithms[8-11] are used to detect the target substructure in the possible hit structures. The remaining problem is that the two similar coordinate sets of query pattern and hit entry do not necessarily correspond to the same set of atoms linked together by chemical bonds.[12] This problem must be overcome by estimating the degree of similarity between the two sets of coordinates. Several types of similarity estimation are compared by Pepperrell and Willett.[13]

The aim of the present study was to overcome the various problems which have been described above to detect the desired target substructures in the possible hit structures which may become quite large. Our main strategy is the 3D structure access using 2D structure information. This way requires mapping of the 2D and the 3D structure information during the procedure of structure access and analysis of conformational freedom. This special approach offers the advantage

† This work is part of the dissertation by Elke Lang, TH Darmstadt, D17.

0095-2338/93/1633-0763$04.00/0

of using the available 2D structure retrieval features directly. The amount of structures which have to be investigated by time-consuming conformation analysis algorithms is dramatically lower than that which is produced by screenout with known methods. In contrast to screen-based systems, the presented method will not find any false positives.

MATERIALS AND METHODS

The system which will be outlined has been built on the basis of the database network of the DKFZ Department of Spectroscopy, which contains spectroscopic, crystallographic, and substance data. This network provides substructure search facilities and, due to its structure-oriented design, offers the possibility of structure identification and structure-oriented information linkage. The requirements which should be fulfilled by the system are raised by its purposes, structure elucidation in spectroscopy, and support of molecular modeling: (i) substructure search according to user-specified degrees of similarity, (ii) 3D structure access of the retrieved structure entries, and (iii) automatic conformation analysis performed on the set of structures without the need of making the atom numbering of single entries consistent to the reference structure by the user himself.

In the following, the most important aspects will be described in detail. This includes preconditions as well as special implementations.

Database Techniques, Structure-Oriented Information Network. In very clear contrast to most of the known databases in chemistry, the presented system has not been implemented as a dedicated file-management system but by means of a relational[14] database system product, SQL/DS, which provides features for the creation and maintenance of databases, input of data items, and property-oriented retrieval. This strategy assures that further databases can be added easily, using or enhancing the existing user surface and joining the contents of different databases by specifying their common properties (e.g. structure information for databases in chemistry). Database administrators and application programmers do not have to deal directly with questions of physical storage representation, since a data description language (DDL) is available for database conception and retrieval planning such as creating indexes for fast access. The data manipulation language (DML) ISQL is used to formulate queries by describing the desired properties. Avoiding programming and maintenance overhead is a valuable help for scientists who, in fact, want to focus on the problems they want to solve with a database and not on the problems of database management.

The information network of the DKFZ Department of Spectroscopy (Figure 1) has been founded with the spectroscopic database SPEKTREN II,[15] comprising a structure data table and several tables for experimental data. With the structure database as the core of the system, a toxicologic and a crystallographic[16] (Cambrige Crystallographic Data Files, CCDF) database have been added and a modeling database for the storage of structures which have been obtained by modeling techniques is currently developed.

The structure database contains 2D structure descriptions in several degrees of similarity sharpness which serve to find structures which are identical or similar in constitution with a certain degree of similarity. Finding identical structures is very fast, as the structure-describing hash code derived using the Morgan algorithm[17] serves as primary key for the data entries. Lower degrees of similarity slow the retrieval down within still acceptable ranges of time, as the display of results

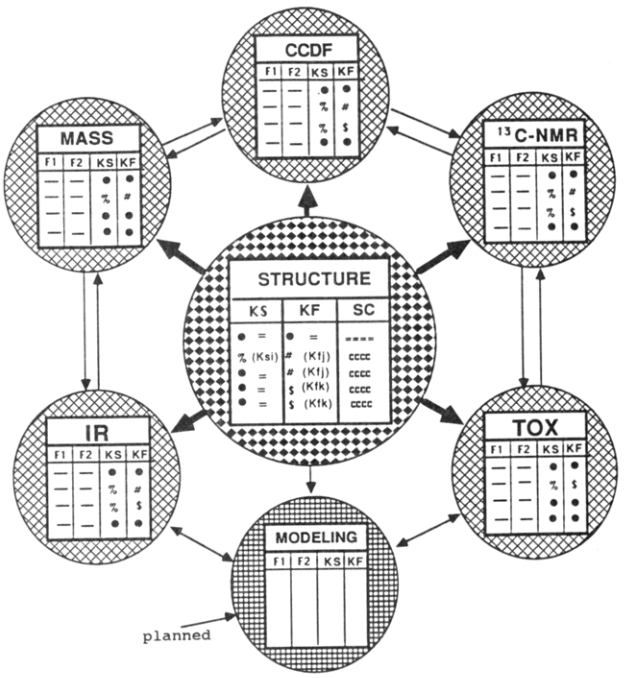

Figure 1. Information network of the DKFZ Department of Spectroscopy containing spectroscopic, crystallographic, and toxicologic information. The different databases are joined together by a canonical description of the chemical structure. Key: KS, numerical key for the constitution as generated by the Morgan algorithm; KF, numerical key for the configuration as generated by the SEMA algorithm; SC, similarity code; F1, F2, ..., stored information.

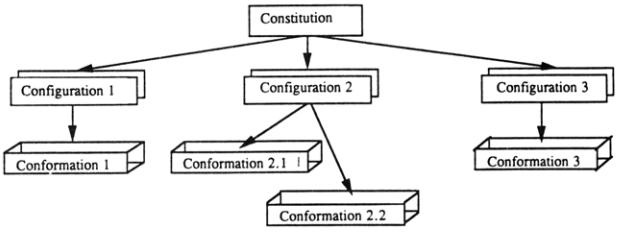

Figure 2. Three degrees of structure description accuracy: constitution, representation by the connection table (atom types, bonds and bond types); configuration, configuration of stereocenters indicated by desciptors derived from 3D data; conformation, numeric values of torsion angles considered as distinctive parameters between different conformations.

already starts during the search process. A structure editor for the input of search structure templates and a specification menu which helps to indicate the desired degree of similarity prepare the information from which the SQL structure search argument is built. The results of structure search queries are presented at the screen (2D and 3D structure representation of each hit entry) with a colored indication of the search structure template in both structure representations.

Dealing with 3D Information. Within the information network, a substance can be handled with three degrees of structure description accuracy (Figure 2) (structure description accuracy relates to different levels of description accuracy for one structure and should not be confused with the degrees of 2D structure similarity described above). The first degree is the description of the constitution (2D structure) according to the Morgan algorithm. Many of the database entries can possess the same constitution. One constitution description, if it shows at least one stereocenter, can lead to several configurations which can all be present in the database. Therefore, the configuration description is the second degree of accuracy. The third and highest degree of accuracy is the structure conformation, represented by the (3D) coordinates

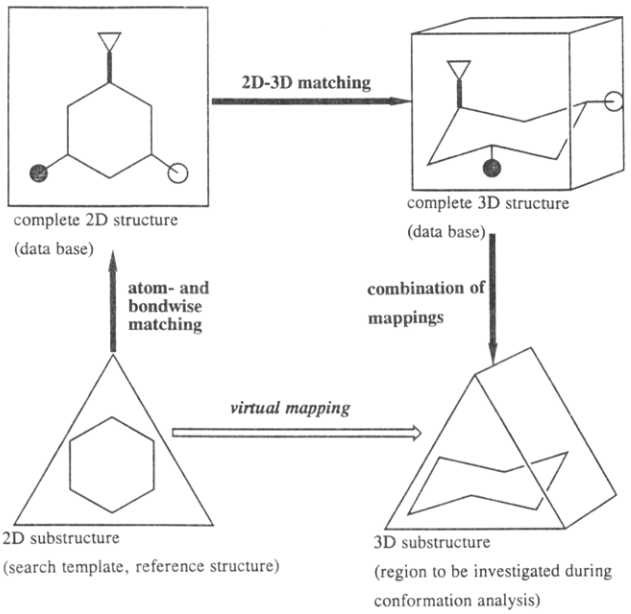

Figure 3. Concept of structure handling: 3D substructures in the database are found by linkage of two mappings, atom- and bondwise matching and 2D–3D matching.

of the structure. Conformation analysis can only lead to meaningful results if the geometric comparison is performed on structures that are similar at the lower degrees of accuracy (and hence comparable) within the structure region being under investigation. For this reason, observing the restriction that only structures with known topology can be handled, a special concept of structure handling has been developed (Figure 3): Any type of structure access, 2D, stereo-enhanced 2D, and 3D, begins at the 2D level as the access level and satisfies the desired degree of accuracy by mounting in accuracy as far as needed. This proceeding avoids geometric 3D search and the need of deriving 3D bit screens but offers access to 3D substructures via their corresponding 2D substructures. The correspondence between 2D and 3D structure representation (recognizing a certain atom of the 3D representation as corresponding to a certain atom of the 2D representation and vice versa) is found by a 2D to 3D matching procedure.[18] The intermediate level of stereosensitive 2D description has not been made available so far for CCDF users with the original QUEST software. Within the described information network, stereochemistry is represented on the base of the stereochemically extended Morgan algorithm (SEMA)[19] for all of the involved databases. From the SEMA representation and the 3D coordinates, an R/S indication of stereocenters according to Cahn–Ingold–Prelog (CIP) nomenclature[20] can be derived[21] and offers a description which is familiar to most of the users. The stereodescription of structures offers the possibility of stereoselective structure retrieval and comparison.

Conformation Analysis. Structure data which have been extracted from the database can be submitted to conformation analysis. A set of structures bound for investigation must possess a structure region in common. This common substructure is specified in the graphics-oriented user menu and forms the basis for analysis requests. Distances, bond angles, and torsion angles are computed for the atom sets and operations which have been required by the user by means of the menu. If stereocenters are detected in the common structure region, the set of structures is partitioned into the resulting classes of different stereoisomers which contribute to different sets of results. The analysis results are listed by using the numbering scheme of the common substructure and indicating its relationship to the numbering scheme of the individual entries so that the derivation of histograms and distributions, but also a closer look to certain single entries, is possible. Details about the proceeding of the conformation analysis module have been reported in ref 22.

The conformation analysis feature serves as a tool for structure comparison and evaluation of the effects of 2D structure similarity on conformative similarity. The second challenge of conformation analysis is the derivation of numeric data (internal coordinates, mainly torsion angles) for a conformation description. The current implementation of the system provides the possibility of operating on the conformations of structures which have been retrieved. The further step which is now being implemented is a conformation description which can be searched directly. Here again the benefits of using a relational database system product must be stressed: the flexibility of the database architecture and the system features for creating and modifying data tables or data rows in existing tables makes it easy to add data fields to the existing data tables and to create test versions of conformation description concepts.

RESULTS: APPLICATION OF THE SYSTEM TO THE INVESTIGATION OF PHOTOSENSITIZING PORPHYRIN DERIVATIVES

Similarity-based substructure retrieval and conformation analysis techniques as described above have been applied to several problems in the field of computer-aided drug design which arose during the work with medicinal chemists and pharmacologists. One of these cases will be outlined below in order to demonstrate how the questions posed by users of computer-aided drug design techniques can be satisfied by the system.

Porphyrins are used as photosensitizers for the photodynamic cancer therapy and in tumor diagnosis, as porphyrin-enriched tumor tissue produces characteristic fluorescence effects. Several properties of porphyrins must be optimized to make them satisfying tools in cancer therapy: Photosensibilization must concern a light spectrum which does not overlap with normal daylight spectra as patients should be able to be exposed to daylight during therapy. Enrichment in tumor tissue should be high, whereas enrichment in normal tissue, especially in liver and kidneys, should be minimal to avoid necrosis of tumor-adjacent tissue during irradiation. Tetraphenylporphyrins normally do not satisfy these conditions as their water solubility and molecular weight are too low. A possible solution is to increase the molecular weight and the volume of the structure by adding further substituents to the phenyl groups.

The above-mentioned therapeutic requirements can be translated into terms of structure chemistry:

(1) Are the phenyl rings conjugated with the porphyrin skeleton?

(2) Are the four phenyl substituents rotatable or "frozen" in their positions?

(3) Which constellations of torsion angles do the phenyl rings exprime, and which frequencies of constellation patterns can be observed?

(4) Which conformations of the porphyrin skeleton do exist? How does the skeleton adapt to central ions of different sizes?

The CCDF were searched for tetraphenylporphyrins by substructure retrieval (Figure 4). A set of hits with complete and reliable structure information (52 structures which were

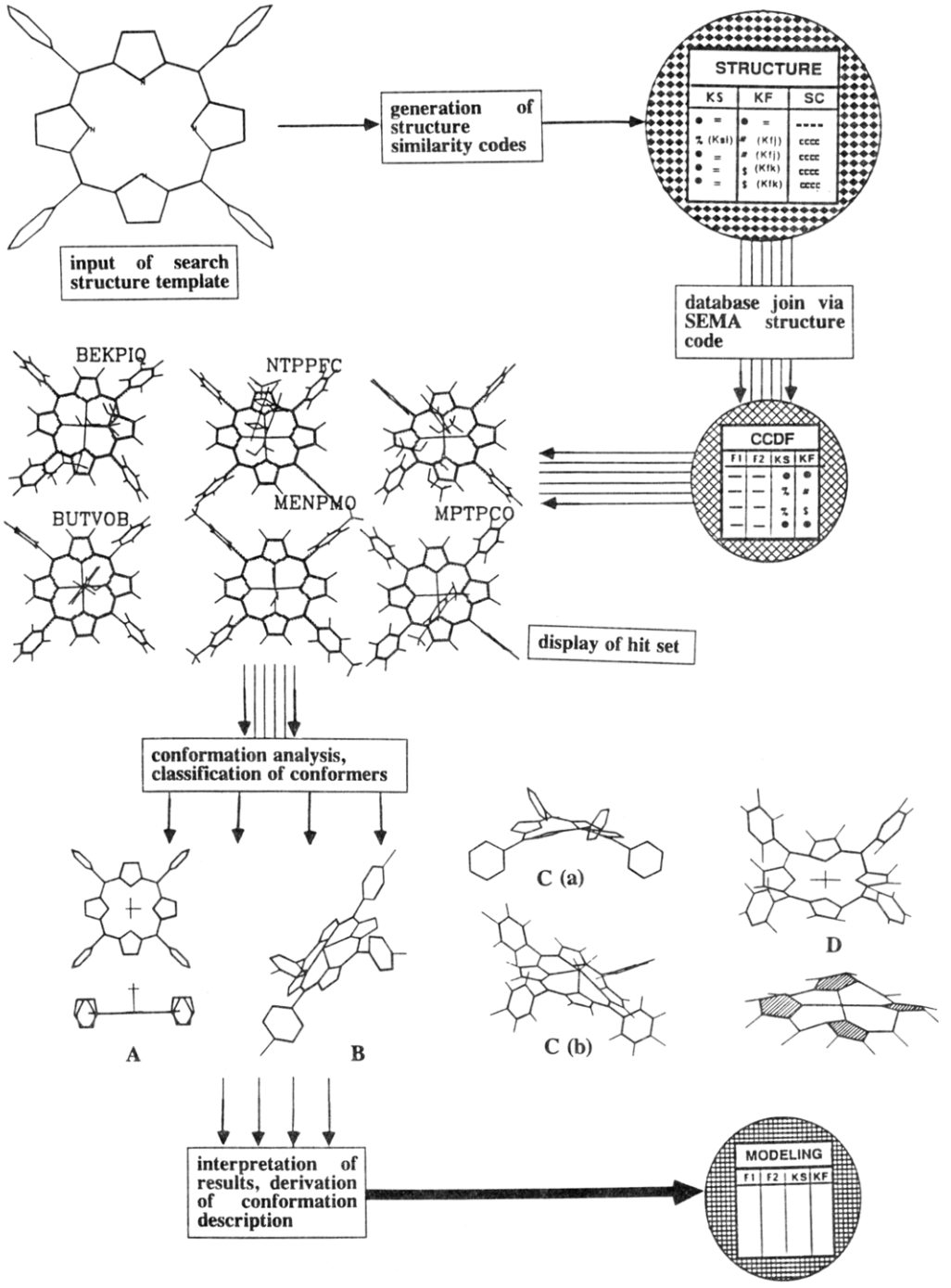

Figure 4. Information flow in the system. After input of the desired structure, structure description codes are derived and used as search arguments for the structure database. For each of the hit entries, structure-related information is found in the CCDF data tables via the SEMA code linkage key. Conformation analysis and interpretation of the results lead to the derivation of conformation descriptions, which can be stored in the modeling database.

suitable according to substitution type and size) was submitted to conformation analysis. The conformation parameters which were chosen to be computed were tabulated and allowed to estimate plausible ranges for the concerned parameters. Consideration of characteristic differences of torsion and bond angles allowed, by numeric criteria, detection of the four distinct shapes of the porphyrin skeletons which had already been reported by Hamor et al.[23] The distinctive parameters and their values are given in Figure 5 and Table I. The above questions can be answered as follows:

(1) Torsion angles between 80° and 120° and bond length values between 1.49 and 1.51 Å show that phenyl groups and porphine skeleton are not conjugated.

(2) Visual consideration of the van der Waals radii gives an evident proof that free rotation of the phenyl groups is impossible for geometric reasons.

(3) The orientations of the phenyl groups toward the porphine skeleton (defined as positive or negative torsion senses of the four torsion angles from the porphine skeleton to the phenyl groups) is nearly equally distributed on the sequence patterns "four equal signs", "alternating signs", "– – + + alternating", and "three equal signs" and implies that the orientation of the ring closure constellation is conserved in the structure shape.

(4) The conformation types could be detected applying the numeric criteria listed in Table I. The question of the adaption

Table I. Distinctive Conformation Parameters Which Were Used for the Detection of Porphine Skeleton Shapes (See Figure 5 for Localization of the Angles)[a]

	i	ii	iii	iv	II	III	N–Met–N		i	ii	iii	iv	II	III	N–Met–N
A	7.4	170.6	5.3	174.3	175	175	180	C	7.4	170.6	5.3	174.3	175	150/170	<180
B	7.4	170.6	5.3	174.3	175 180	175 150	<180	D	25.5	154.7	0.1	158.8	160	170	<180

[a] Shapes and parameters: A, flat skeleton; B, domed skeleton; C, saddle-shaped skeleton; D, tilted skeleton; i, ii, iii, iv, dihedral angles spanning the skeleton; II, III, dihedral angles around the skeleton; N–Me–N, bond angle between two opposite nitrogen atoms and the central ion. Two values, as for skeleton shape C-dihedral angle III and for D–II and D–III, indicate that alternating appearance of the values is significant for the shape detection.

Table II. Differences in Search Requirements and Retrieval Proceeding between the DKFZ Information System and Current Bit Screen-Based Methods

aspect	DKFZ information network	screen-based systems
information required for 2D and 3D structure retrieval	2D: bond information (connection table) 3D: bond information and atom coordinates	2D: bond information (connection table) 3D: atom coordinates
information which must be prepared for retrieval	2D structure codes describing structure features in several degrees of similarity stereodescription	2D structure bit screens 3D structure bit screens distance maps (optional)
proceeding of 2D structure retrieval	input of 2D structure template specification of desired similarity derivation of structure code as search argument in SQL command	indication of 2D structure properties which should be present (bit screens set) (input as command list or graphically)
configuration and conformation (3D) retrieval	=first input: 2D structure template, as for 2D search further input according to level of steric information: =indication of configurations of stereocenters 2D–3D mapping of hit entries conformations with appropriate configuration are found =restriction on certain conformations (*in preparation*) specified conformation forms are found	=input of 3D coordinate set derivation of 3D bit screen settings or of distance map first database search (prescreening), extraction of a candidate subset exhaustive examination of the candidates (distances, similarity to input coordinate set)
inherent problems in structure identification	for certain entries: 2D–3D mapping failure because of data errors of 2D and/or 3D structure representation	distance flexibility: high flexibility causes false positive hits, high rigidity looses hits similarity can be caused by artifacts, bonded or nonbonded adjacents cannot be distinguished

mechanisms of the skeleton to central ions of different size was revealed to be based on false assumptions, as the example structures did not show any interdependence between crystal ion radii and skeleton shapes. Literature studies gave several explication models for the phenomena of skeleton shape formation, from which the most recent are based on ligand field theory (strong or weak axial ligands) and consideration of the spectroscopic spin state. Hoard[24] and, in a detailed manner, Scheidt and Reed[25] reported these phenomena.

Another application of the system, conformation studies of cyclophosphamide derivatives, could provide decisive information for their spectroscopic structure elucidation.[26] Dihedral angles for the interpretation of NMR coupling constants, the steric orientation of the chloroethyl groups, and an appropriate starting conformation for molecular modeling helped in revealing the structure of the cyclophosphamide derivatives which had been synthesized.

DISCUSSION

For the implementation of the presented system, considerable advantages were gained with a restriction which has been shown to be acceptable for our applications so far: The strategy of 3D structure access by 2D structure retrieval and 2D–3D mapping allowed use of the existing structure retrieval facility which works very fast and efficient. The second advantage was that all parts of the work which has been described could be done without the existence of 3D bit screens for the 3D structure access. The conformation analysis results, however, can be used in developing description systems which will help to implement further components of the system.

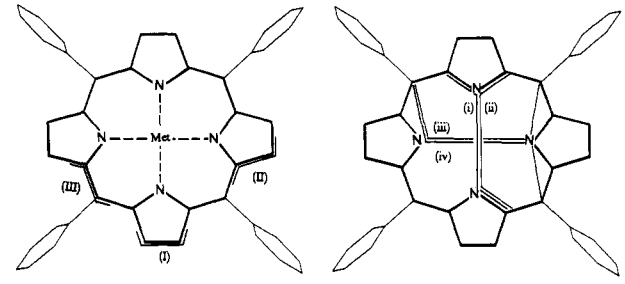

Figure 5. Localization of the torsion angles and the angle between two opposite nitrogen atoms and the central ion which had been considered for automatic distinction between the different skeleton shapes (see Table I).

The use of the unique and unambiguous Morgan structure code as primary key in the database network allows identification of structures and linkage of information concerning one structure but being stored in several substance data tables. The set of structures which are common to spectroscopic and crystallographic data tables, for example, can be retrieved with a single SQL command and be used for investigations and evaluations in the enhancement of spectroscopic data by 3D information. Table II gives an overview of the most important differences between the proceeding of our system and that of bit screen-based methods.

The restriction of the presented system is that only structure templates with bonded atoms can be searched. Long-range distances between nonbonded atoms or excluding volumina cannot be specified as retrieval arguments. This restriction did not affect the investigation of the presented examples, as

it is not valid for the conformation analysis phase: in the structure template which is used for the specification of the required conformation parameters, arbitrary pairs of (nonbonded) atoms can be indicated and their distances will be performed, as they will be found in the structures by 2D–3D mapping.

Of course, important parts of the results of the porphyrin study had already been reported years before. It must be stressed, however, that the system presented here is not intended to be used by chemists working on certain types of substances and being on the current state of knowledge in this special field but to be an easy-to-use tool for medicinal chemists and pharmacologists dealing with a wide realm of substance families and not knowing too much about recent findings on special aspects and the background of structure chemistry. Indeed, the porphyrin example and the cited list of questions report the state of problem-specific knowledge and the requirements of the involved pharmacologists, and it cannot always be assumed that molecular modeling experts possess much more special knowledge of the problem before having browsed current literature. From this point of view, it can be stated that the presented system could help a lot in solving the problem fast and correctly. The system provides retrieval methods to gain 3D coordinates of reference structures and means for problem-oriented conformation analysis and the interpretation of results. Therefore, experimental work can be supported by analytical work, and analytical work is elegant and fast compared to non-automatized processing of retrieved molecular structures.

Further components of the system which will allow the distinction of conformations during the search process are currently under development. The open architecture of the relational database system allows addition and removal of test implementations or the comparison of several prototypes, so that the results of the current system can be interpreted and verified in a multitude of test applications before the establishment of irreversible access data sets.

ACKNOWLEDGMENT

E.L. thanks Claus-Wilhelm von der Lieth for his support and contributions to the integration of the presented system into the database network of the DKFZ Department of Spectroscopy.

REFERENCES AND NOTES

(1) v. d. Lieth, C. W. Spectroscopic Information and Molecular Modeling. *Chemom. Intell. Lab. Syst.* **1990**, *8*, 53–58.
(2) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. The Development of Versions 3 and 4 of the CSD System. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187–204.
(3) Martin, Y. C. 3D Database Searching in Drug Design. *J. Med. Chem.* **1992**, *35*, 2145–2154.
(4) Jakes, S. E.; Willett, P. Pharmacophoric pattern matching in files of 3-D chemical structures: selection of interatomic distance screens. *J. Mol. Graph.* **1986**, *4* (1), 12–20.
(5) Murrall, N. W.; Davies, E. K. Conformational Freedom in 3-D Databases. 1. Techniques. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 312–316.
(6) Carhart, R. E.; Smith, D. H.; Venkataraghavan R. Atom Pairs as Molecular Features in Structure–Activity Studies: Definitions and Applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64–73.
(7) Jakes, S. E.; Watts, N.; Willett, P.; Bawden, D.; Fisher, J. D. Pharmacophoric pattern matching in files of 3D chemical structures; evaluation of search performance. *J. Mol. Graph.* **1987**, *5* (1), 41–48.
(8) Crandell, C. W.; Smith, D. H. Computer-Assisted Examination of Compounds for Common Three-Dimensional Substructures. *J. Chem. Inf. Comput. Sci.* **1990**, *23*, 186–197.
(9) Barrow, H. G.; Burstall, R. M. Subgraph Isomorphism, Matching Relational Structures and Maximal Cliques. *Inf. Proc. Lett.* **1976**, *4*, 83–84.
(10) Bron, C.; Kerbosch, J. Algorithm 457. Finding All Cliques of an Undirected Graph. *Commun. ACM* **1973**, *16*, 575–577.
(11) Ullman, J. R. An algorithm for subgraph isomorphism *J. ACM* **1976**, *23*, 31–42.
(12) Brint, A. T.; Willett, P. Algorithms for the Identification of Three-Dimensional Maximal Common Substructures. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 152–158.
(13) Pepperrell, C. A.; Willett, P. Techniques for the calculation of three-dimensional structural similarity using inter-atomic-distances. *J. Comput.-Aid. Mol. Des.* **1991**, *5*, 455–457.
(14) Codd, E. F. A relational model of data for large shared data bases. *Commun. ACM* **1970**, *13*, 377–387.
(15) Förster, T.; v. d. Lieth, C. W.; Opferkuch, H. J. Data base applications in spectroscopy: Data base design and retrieval. *GIT Fachz. Lab.* **1989**, *33*, 319–328.
(16) Lang, E.; Förster, T.; v. d. Lieth, C. W. The integration of the Cambridge Crystallographic Data Files into the relational information network of the German Cancer Research Center. In *Software Development in Chemistry*; Gasteiger, J., Ed.; Springer-Verlag: Heidelberg, 1990; Vol. 4, pp 43–51.
(17) Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113.
(18) Lang, E.; Förster, T. Matching of 2D and 3D structure descriptions in the Cambridge Crystallographic Data Files: Outline of the method and some applications. In *Software Development in Chemistry*; Gmehling, J., Ed.; Springer-Verlag: Heidelberg, 1991; Vol. 5, pp 77–84.
(19) Wipke, W. T.; Dyott, T. M. Stereochemically Unique Naming Algorithm. *J. Am. Chem. Soc.* **1974**, 96, 4834–4842.
(20) Prelog, V.; Helmchen, G. Fundamental principles of the CIP system and proposals for revision. *Angew. Chem.* **1982**, *94*, 614–631.
(21) v. d. Lieth, C. W.; Schätz, R. To be published.
(22) Lang, E.; Förster, T.; v. d. Lieth, C. W. Automatic conformation analysis applied on large databases. *Anal. Chim. Acta* **1992**, *265*, 277–281.
(23) Hamor, M. J.; Hamor, T. A.; Hoard, J. L. The Structure of Crystalline Tetraphenylporphine. The Stereochemical Nature of the Porphine Skeleton. *J. Am. Chem. Soc.* **1964**, *86*, 1938–1942.
(24) Hoard, J. L. Stereochemistry of hemes and other metalloporphyrins-the remarkably varied stereochemistry of the iron porphyrins is utilized in the hemoprotein function. *Science* **1971**, *174*, 1295–1302.
(25) Scheidt, W. R.; Reed, C. A. Spin-State/Stereochemical Relationship in Iron Porphyrins: Implications for the Hemoproteins. *Chem. Rev.* **1981**, *81*, 543–555.
(26) Schmidt, B. F.; Tang, W. C.; Eisenbrand, G.; v. d. Lieth, C. W.; Hull, W. E. The use of two-dimensional NMR and molecular modeling for the structure determination of novel cyclophosphamide derivatives: the diastereomers of 1-aza-2-bis-(2-chloroethyl)-amino-3-oxa-2-oxo-2-phospha-7-thio-bicyclo-[4.4.0] decane and -[4.3.0] nonane. *J. Magn. Res. Chem.* **1992**, *30*, 1224–1240.