The application of the CSB to the prediction of single step reactions and reaction pathways

Grażyna Nowak1*, Grzegorz Fic2†

1 Department of Physical Chemistry, Rzeszów University of Technology, 6 Powstańców Warszawy Ave., 35-041 Rzeszów, Poland
2 Department of Computer Chemistry, Rzeszów University of Technology, 6 Powstańców Warszawy Ave., 35-041 Rzeszów, Poland

Received 5 May 2004; accepted 29 June 2004

Abstract: Applications of the CSB (Common-Sense Builder) system for the logic-oriented and knowledge-assisted simulation of chemical reaction courses are described. We present the possibility of using the CSB for two ways of reaction simulation, i.e., as a multi-step process or as single step procedure. Results of the first simulation type are given to predict the course, and to model reaction mechanism. The second one is capable of complex chemical transformations such as multi-component and cascade reactions to generate structurally diverse products for combinatorial chemistry. In several experiments performed, we analyze the capabilities and limitations of the CSB modules and controlling tools for the examination and selective generation of solutions.

© Central European Science Journals. All rights reserved.

Keywords: computer chemistry, CAOS, prediction of chemical reactions, modeling the reaction course, generation of reaction tree, multi-component reactions

1 Introduction

Two concepts of using computer programs in organic synthesis - retrosynthetic analysis of Corey and Wipke [1] (continued by Bersohn [2], Gelernter [3], Barone and Chanon [4], and others) and mathematical-logical models of chemistry (such as Ugi-Dugundji matrix model [5], Hendrickson model [6], the synthon model [7]) - opened new research areas and

* E-mail: gnowak@prz.edu.pl
† E-mail: gfic@prz.edu.pl
initiated the development of several approaches for the representation and simulation of chemical conversions. In computer applications they allowed to solve many problems, for example the design and generation of possible synthetic routes for a given structure, searches for possible mechanisms for a given reaction and the prediction and discovery of new possible modes of organic reactions. Over the years, a number of computer systems of this category have been and are still being developed offering new opportunities to generate and analyze the chemical space in search of new methods for organic synthesis. The fundamentals and applications have been reviewed by Ihlenfeldt and Gasteiger [8]; a current list of CAOS (Computer Assisted Organic Synthesis) software is in [9]. The programs were successfully applied in modeling a wide variety of organic conversions such as metabolic pathways, the degradation in environment, reactions in mass spectrometer, and the generation of combinatorial libraries [10]. However, their practical utility is still not fully satisfied, but limited by low reliability or low creativity of generated reactions (which is a consequence of a compromise between two conflicting requirements: the generation of reliable reactions and the generation of new solutions). We attempt to solve this problem by the integration in the CSB system [11,12,13] of the logic-based generation of chemical reactions (capable to discover new knowledge) with the learning and knowledge processing tools utilizing information about known organic reactions. Recently we presented the newest version of the CSB extended by additional sources of knowledge, i.e., the similar reaction knowledge base and the real organic reaction database, to assist and guide the reaction generator. Due to the combination of logic models for reaction simulation with learning tools to acquire and generalize knowledge from databases, the CSB allows to search for solutions (reaction courses and products) with different levels of reliability/completeness. The user has the possibility to select the most suitable model of simulation, matched to the synthetic problem being solved (prediction of reaction products, searching the reaction mechanism, reaction discovery). Additional sources of knowledge implemented recently in the CSB may be used for verifications of results generated by the CSB.

The objective of this paper is to show the use of the improved CSB in the prediction of the course of an organic reaction and its products. We describe several simulations illustrating the potential of implemented programming tools and options that the user can choose and combine to change the line of the CSB reasoning, and to obtain more adequate results of predictions. As an example of applications of the CSB to study the organic reaction courses, the Biginelli three-component condensation was chosen. We present the results of the CSB predictions, obtained in the single step and multi-step simulation experiments, and we analyze the capabilities and limitations of the CSB for the examination and selective generation of solutions.

2 Applications of the CSB

The CSB system is an interactive computer program for simulation of single step reactions and multi-step pathways for a given set of the starting materials (from one to four).
Details of the CSB have been described elsewhere [11,12,13]; here, only its main features are given in relation to the applications. Actually the CSB system enables:

- combinatorial generation of solutions (libraries of products)
- selective generation of solutions
- database search.

In the first approach, the reactions are generated purely on the basis of the logical scheme (according to the matrix model of constitutional chemistry), were all creativity potential of the CSB is accessed. All allowed ways of a reaction may be generated without additional chemical knowledge, or with optional user knowledge involved in generating process. Examples of such applications may be degradations (photochemical, temperature, in GC), fragmentations and rearrangements of organic compounds in the mass spectrometer. In a separate paper we will present the computer simulation of these reaction types with the CSB. In the second case, the reaction generator is assisted, during the reaction simulation process, by different knowledge sources implemented in the CSB, which are hierarchically organized into three levels (section 3). In this case more rational (and less creative) predictions may be obtained depending on scope of knowledge (level) included in the generating procedure. Examples of integrated logical-oriented and knowledge-assisted generation of reactions are presented in section 4. A third possibility of using the CSB is as a database search for validation of predicted reactions.

3 Knowledge acquisition in the CSB

As already mentioned, computer simulation and prediction of reactions with the CSB can be accomplished:

- in a purely deductive way, without chemical knowledge,
- with mathematical model assisted by user knowledge and intuition,
- with mathematical model assisted by formalized, organized knowledge from optionally selected sources existing in the CSB, or
- using mathematical models assisted by the learning module of CSB, aimed at the creation of new rules for knowledge bases through the exploration, analysis and generalization of data in the reaction database.

Thus, different lines of reasoning may be realized depending on the user knowledge and intuition imposed on the reaction generator, or depending on knowledge sources, implemented in the CSB, involved in the generating process. These knowledge sources, arranged into multilevel structure, are as follows:

- reaction generator library – describing different transformation schemes of reacting centers in substrate(s) – product(s) molecules [11],
- common-sense base REACT – expert-prepared rules, describing reacting centers with structural environments and reaction conditions affecting their reactivity [12],
- similar reaction library – computer-generated rules (generalized reaction descriptions) in the form of cause-consequence correlations [13],
- database of organic reactions.
4 Examples of the CSB predictions

With the recent advances in combinatorial chemistry, one can observe the growing interest in the developing of chemical transformations that efficiently generate molecular diversity and complexity. Multi-component, domino and tandem reactions become attractive categories to explore the application of computer-assisting tools. One of these reactions, based on the Ugi-concept of multi-component (one-pot) reactions was chosen to illustrate the CSB possibilities. Computer modeling experiments were performed on the known three-component Biginelli cycloaddition. The experimental work and computational studies on this reaction were described by Kappe, Falsone et al. [14,15,16]. We selected this reaction because it has a high degree of mechanistic complexity, suitable for computer modeling experiments, and because mechanistic alternatives that were proposed by CSB may be verified relating them to the experimentally confirmed reaction mechanism. Additionally, the transformation designed on the basis of Biginelli reactions allows generation by the CSB the collection of lead-like structures with dihydropyrimidine scaffold.

This reaction can be performed with the CSB in two ways, i.e., as a multi-step process or as a classical one-pot procedure. In next sections we describe how the CSB can be applied to the simulation of these two ways of running the reaction. With several simulations performed, we demonstrate and analyze the possibilities of implementing in the CSB programming tools, knowledge- and databases, and we illustrate in detail how the user can exploit them for modeling the course of chemical reactions.

The three starting materials of this reaction (an N-substituted urea derivative, acetylacetone, and an aldehyde) are presented in Fig. 1.

4.1 Multi-step simulation

In this experiment, the reaction tree is generated by the CSB, beginning from the starting materials, via sequence of intermediates, to a set of alternatives of products. Two conceivable courses of the Biginelli reaction were modeled with the CSB to study the reaction mechanism. In the first run, the possibility that the initial step is the reaction of the aldehyde 1 and urea 2 precursors has been examined. In second step of this run, each intermediate from the generated list of candidates reacts with acetylacetone 3. In the second run, another reaction mechanism was developed by the CSB. Here, the addition of aldehyde 1 to acetylacetone 3 has been considered as a possible primary step, and after this, each of the resulting candidates reacts with substrate 2 to form a second level of intermediates.

For modeling these reaction sequences, the common-sense model was chosen as the reaction generation procedure, accessing two different sources of knowledge, i.e., the reaction generator library and knowledge base REACT. Among different transformation schemes stored in the library of the reaction generator (describing conversions of reacting centers into substrate(s) and product(s) molecules) we selected only three for this run (I,
II and III in Fig. 2), restricting in this way the size of the resulting product space. The number of products generated is further limited through the application of the second source for acquiring chemical knowledge. It allows to recognize the reaction sites and charges in the reacting molecules, and to select the proper combinations of the reactive centers that will comply with transformation schemes chosen from the reaction generator library. Besides this knowledge-based control, other evaluation tools were selected and combined in this run in order to guide the reaction generation process, and to obtain more reliable product predictions. These are: the thermodynamic product control (we specified limits of values for the changes of reaction enthalpies: \(<-20, 20>\) kcal/mole), tautomeric conversion of an unstable form into a more stable one, and some heuristics.

The resulting reaction tree predicted in the first run, according to these restrictions imposed on reaction generator is shown in Fig. 3. In the first level, the set of 17 molecules (Fig. 4) was generated, as a result of modeling the reaction between two starting materials (molecules 1 and 2) by three transformations (I, II, III in Fig. 2). From this set only the most probable, three candidates (molecules: 4, 5, and 6 in the reaction tree, Fig. 3) were selected for modeling the second step. Compounds 4 and 6 were formed by the application of the reaction transformation I, (with two bonds formed/broken) to modeling, in this case, the addition of the primary amine group, or of the secondary one (in substrate 2) to the carbonyl group of formaldehyde 1. Compound 5 was obtained with the second reaction transformation, resulting in three bonds breaking/forming. Here, the reaction transformation was used for the simulation of the addition/elimination reaction. Molecule 4 (generated by the transformation: A-B + C-D \(\rightarrow\) A-C + B-D, \(\Delta H = 0\) kcal/mole) being precursor for 5 (generated with the reaction transformation: A-B + D-C-E \(\rightarrow\) B=C + D-A-E, \(\Delta H = -9\) kcal/mole) was excluded from the second step generation. Thus, in this reaction step, each of two alternative candidates (5 or 6) reacts with a third reagent (acetylacetone) to form intermediates of the second level.

For modeling these reactions, previously activated control parameters and restrictions (kept in memory) can be reproduced, changed or deleted. Retaining the same menu controlling the simulation process (common-sense model for reaction generation, three reaction transformations chosen from library of reaction generator, thermodynamic control with the range of enthalpy \(<-20, 20>\) kcal/mole, and tautomeric control) we obtained two sets of the second generation products.

The first from these sets (molecules: 7-10), derived from the most promising precursor 5 and acetylacetone 3 is presented in the resulting reaction tree (in Fig. 3). Intermediate 7 was formed by the application of the transformation I for modeling the addition of the N-H group of 5 to the carbonyl group of 3 (\(\Delta H = 8\) kcal/mole). Intermediate 8 was the result of the same addition with the loss of water, (reaction transformation III, \(\Delta H = 18\) kcal/mole). Candidate 9 was derived from another reaction type, effected by the loss of a proton from the methylene group of 3 and then addition of this active anion to the C=N bond of precursor 5 (reaction transformation I, \(\Delta H=1\) kcal/mole). The last compound 10 from this set is the result of the application of the same transformation for modeling the similar reaction type, i.e., addition of other active anion (formed by the
loss of a proton at the methyl group of 3) to the C=N bond of 5.

The second set of candidates (11-14), resulting from the reaction of precursor 6 with acetylacetone 3, is shown in Fig. 5. Compound 11 was formed with the reaction transformation I, simulating the addition of the primary amine group of 6 to the carbonyl group of 3 (\(\Delta H = 8\) kcal/mole). Compound 12A was generated with another combination of bonds formed/broken (reaction transformation II). In consequence, a reaction as above was generated one step further, with elimination of water. The less stable tautomeric form of 12 (12B, \(\Delta H = 18\) kcal/mole) obtained in this reaction was converted into a more stable one, 12A (\(\Delta H = -19\) kcal/mole). This compound was also generated immediately from 6 and 3, with the reaction transformation III (\(\Delta H = -1\) kcal/mole). Another reaction course for these reacting molecules (6 and 3), leading to 13 and 14, was modeled by the reaction transformation I. In this case, the substitution on a carbon atom at the C-OH group in precursor 6 by anion \(-\text{CH}_2-(\text{CO-CH}_3)-\text{CO-CH}_3\) or \(-\text{CH}_2-\text{CO-CH}_2-(\text{CO})\text{CH}_3\) from acetylacetone was reproduced (\(\Delta H = -8\) kcal/mole). From structures generated in the second step, we selected only three more probable intermediates (9, 10 from the first set, and 13, 14 from the second one) to continue generating the reaction tree.

From the first competing intermediate 9 employed in reaction modeling, the sequence of ring closure reactions was generated (some of them are in the reaction tree of Fig. 3). Product 15 was a consequence of ring closure effected by the attack of \(-\text{CH}_2-\text{CO-R}\) anion on a carbon atom of the amide group, with the removal of the -CH$_3$-NH$_2$ substituent (transformation I, \(\Delta H = -7\) kcal/mole). Product 18 was predicted according to the same reaction type, but here the OH was the leaving group, followed by elimination of water (transformation III, \(\Delta H = -11\) kcal/mole). In compound 16 modeled by the reaction transformation scheme I, the C-N bond between a carbon atom at the carbonyl group, and a nitrogen atom of the amide group was formed to effect the cyclization (\(\Delta H = 7\) kcal/mole). The same substitution type, with the additional elimination of water, has given the cyclic product 17 (transformation III, \(\Delta H = 18\) kcal/mole). Other types of cyclic compounds (19-22) were obtained from the second intermediate 10 employed in modeling the third reaction step. The reactions producing candidates (19,20,21) takes a similar course as above (16,17,18), i.e., product 19 (reaction transformation I, \(\Delta H = 7\) kcal/mole) is analog of 16, candidate 20 (reaction transformation III, \(\Delta H = 18\) kcal/mole) corresponds to 17, and product 21 (reaction transformation III, \(\Delta H = -11\) kcal/mole) is the same type as 18. An additional compound 22 was obtained by the same reaction transformation scheme and reaction centers as 20 (transformation III breaking the second combination of N-H, C=O, and H-C-H bonds, \(\Delta H = 18\) kcal/mole).

Further candidate products of this level (for example 23 and 24) were generated when the third competing intermediate 13 was involved in the reaction modeling. Here similar cyclizations were modeled. The reaction leading to 23 is the same type as simulations of 16 and 19, and the second reaction providing 24 is as the above reaction, generated one step further (loss of water and formation of C=N double bond). From the last intermediate 14 involved in the third step generating procedure, four further alternatives were obtained (25-28).
In the second approach, another reaction mechanism was tested with the common-sense module of the CSB, using the same as above controlling menu. The reaction tree obtained in this run is presented in Fig. 6. Here, the reaction between aldehyde 1 and acetylacetone 3 was considered as a first reaction step. In modeling this reaction we obtained twelve intermediate candidates, seven of which are shown in the reaction tree. Molecules (29, 31, 33) are produced by the application of the same reaction transformation of breaking two bonds and making two bonds, but with different combinations of reactive centers involved in the reaction generation. In modeling of the product 29 (ΔH = -10 kcal/mole), the two bonds in substrates (the C-O single bond of aldehyde 1 and the C-H bond at methyl group of 3) are broken, and two bonds (C-C and O-H) in product are formed. It corresponds to the addition of the \(-\text{CH}_2\text{C(CO)}\text{-CH}_2\text{(CO)}\text{-CH}_3\) anion (its enol form) to the carbonyl group of aldehyde. Molecule 31 is formed by breaking the same C-O bond of aldehyde, and other, more reactive C-H bond at the methylene group of 3 (ΔH = -12 kcal/mole). With another combination of breakable bonds involved in the reaction modeling, compound 33 was formed (ΔH = -8 kcal/mole). Here, the addition of the enol anion of aldehyde 1 to the carbonyl group of 3 was simulated (this reaction was classified as less probable by the common-sense module because the C=O group of ketones is less electrophilic than that of aldehydes).

Other intermediates of the first level (30, 32, 34) were generated, by using another reaction transformation scheme (II) in which five reaction centers change their valency (breaking/making three bonds). With this scheme involved in the reaction generating procedure, the addition/elimination reactions were modeled, of similar type as in the first run with the CSB. From this sequence of intermediates, only the most probable molecule 32 was involved in the second-step modeling (reaction with substrate 2, i.e., N-substituted urea). Some of the intermediates are shown in the reaction tree (Fig. 6). Product 36 was generated with the reaction transformation II, modeling the addition of the primary amine in 2 to the C=O group of 32 followed by elimination of water (ΔH = 4 kcal/mole). The second conceivable course of the second-step reaction, effected by addition of the nucleophilic center HN(CH3)CO-NH2 to the C=C double bond of 32 (reaction scheme I, ΔH = -4 kcal/mole) has produced intermediate 37. Its unstable enol form (molecule 37A), generated with other reaction transformation scheme (transformation II, ΔH = 17 kcal/mole), was excluded from third step generation. Another reaction generated from these two reagents (32 and 2) was the addition of the second nucleophilic center (NH2-(CO)-NH-CH3) of substrate 2 to the double C=C bond of 32 with formation of 38 (reaction transformation I, ΔH = -4 kcal/mole). The CSB is not capable of distinguishing the relative reactivity of these two N-H bonds in 2. The reaction 38 has the same priority as 37 resulting from identical reactivity of both N-H bonds at the amide group and equal reaction enthalpies. These two intermediates (37, 38) were involved in modeling of third reaction step generating two sets of final products. The first set (39, 40, 41) - formed by different cyclization reactions of intermediate 37, and the second one (42, 43) - produced by analogous cyclizations of intermediate 38.
4.2 Single step simulation

In single step simulation the library of possible final products is generated from the ensemble of three starting materials according to a given reaction transformation scheme (or transformation schemes), the model of simulation, and control parameters chosen by the user. For modeling the one-pot Biginelli protocol the following reaction scheme was designed and then stored in the library of reaction generator:

\[ \text{A-B} + \text{C-D} + \text{E=F} + \text{G=H} + \text{K-I-L} \rightarrow \text{A-E-I-G-C} + \text{B-F-K} + \text{D-H-L}, \]

where A, B, ... L, denote the single atoms or larger molecular fragments.

In order to illustrate the advantages of the multi-model approach in this single step experiment, it was performed with three optionally selected models of the CSB. Using the D-U model with the above reaction transformation (128 permutations) and full control menu, the CSB has generated 515 solutions, with a considerable part of some unusual ones. When the common-sense model was applied in computer simulation with the same set of controlling parameters and reaction centers, a set of 58 products has been generated. Here, the reduction of the solution space was a consequence of the recognition of allowed combinations of reactive centers, corresponding with the reaction transformation and the charge extinction rule. However, the most adequate model for simulating this reaction type (eight reaction centers changed during the course of the reaction) was the similarity model. The application of this model caused a distinct decrease in the size of product space, eliminating some less satisfactory results. As a result, two products were generated, 17, \( \Delta H = 10 \text{ kcal/mole} \) and 24, \( \Delta H = -9 \text{ kcal/mole} \), shown in Fig. 7. In this case, the computer-produced rules in the Similar Reaction Library have constituted the basis for the simulation experiment. In order to focus the generating process onto a specific reaction category, the Real Reaction Database (source of learning and creating rules) was endowed with 144 examples of the real Biginelli reactions. During the learning procedure, every reaction from this set was reproduced by the common-sense model, and in a result the respective rules in the form of the cause-consequence correlations were generated and stored in the Similar Reaction Library. Any correlation in this knowledge-base is a general description of a collection of similar reactions, and represents a given combination of reaction centers in substrates/products molecules with the meaningful structural surroundings. The CSB system using these learned rules enables reproduction of every reaction being learned as well as generation of all the reactions that are similar to representatives in the Real Reaction Library.

5 Conclusion

In multi-step simulation experiments, two conceivable courses of the Biginelli reaction were modeled with the CSB to study the reaction mechanism. In the first run, the possibility that the initial step is the reaction of the precursors aldehyde 1 and urea 2 has been examined. From the reaction tree generated in this run (Fig. 3 and 5), we selected
the most probable reaction course shown in scheme I (Fig. 8). The same thermodynamic passage ($\Delta H = 10$ kcal/mole) has the pathway in scheme II (Fig. 8). In the second reaction course, the second less reactive C-H bond of acetylace tone 3 participates in the same addition (to the C=N bond of 5), leading to the product 20.

The presence in substrate 2 of second nucleophilic center (the secondary amine group) may be a source of third competing reaction, and the reaction sequence (with higher thermodynamic priority, $\Delta H = -9$ kcal/mole) then leads to the alternative product 24 (scheme III, Fig. 8).

In the second run with the CSB, for the same ensemble of starting materials, another reaction mechanism was developed (reaction tree in Fig. 6). Here, the addition of aldehyde 1 to acetylace tone 3 has been considered as a possible primary step. The most likely pathways in this run (Scheme IV and V, Fig. 8) have led to the following competitive products (43, 41), being analogues to the candidates 17 and 24 produced in the first simulation experiment. The above pathways have the same thermodynamic passage, but differ in the second step. Two competing reaction sites in substrate 2, i.e., N-H bonds in primary and secondary amine groups, (recognized by the CSB at the same level of reactivity) are sources for two directions of addition to the C=C bond of 32. In this case, the CSB is not able to distinguish a relative reactivity of these two reaction centers. More information is needed in knowledge bases of the CSB to determine and predict which of the scenarios would dominate: formation of candidate 43; formation of 41; or proceeding two (or more) reactions simultaneously and producing nearly equal amounts of products.

Two other possible routes (from the intermediate 5) were obtained in this run leading to the formation of the equivalent of 20 (from the first simulation), and other candidates (43, 44 in Fig. 7).

The set of preferable product candidates with some of less probable alternatives obtained in two discussed simulation experiments is presented in Fig. 7. Among these structures, compound 17 was obtained as the main product in the experimental synthesis. Closer inspection of the two different pathways involved in the formation of the product 17 (Schemes I and IV) shows that these courses differ in the first two reaction steps. Taking into account the enthalpy of these reaction steps, the preferred pathway is produced by the first simulation ($\Delta H = -9$ kcal/mole, Scheme I). Such a mechanism for this condensation was supported by the work of Falsone and Kappe in acidic conditions [17]. They have established that the key step in this sequence involves the acid-catalyzed formation of an N-acylinium ion.

In a separate test, involving acidic conditions for the reaction simulation, the CSB generated the intermediate of this type (reaction transformation V in Fig. 2). The formation of this iminium ion from 1 and 2 starting materials is energetically favored, ($\Delta H = -59$ kcal/mole), this ion is more stable than 5 (Fig. 3) intermediate by -50 kcal/mole. However, more information is needed in the CSB knowledge bases to determine the relative reactivity of competing reaction centers ($-\text{NH-CO-NH-CH}_3$ and $-\text{N(CH}_3)-\text{CO-NH}_2$) and to predict with a high degree of certainty which of the courses would dominate.
In the single step simulation experiment, the application of the common-sense model and the combination of control parameters has produced too many solutions. However, implementation of four simulation models into the CSB allows us to select the most suitable tool for the given synthetic problem. In order to obtain more satisfactory product predictions for the discussed reaction type (eight reaction centers changed by the reaction scheme with 128 permutations), the similarity model seems to be more valuable. In this case, reactions that are similar to instances in the database of real reactions are generated, avoiding the generation of the complete solution space.

The similarity model conducted with the learning procedure is aimed at the creation of complex multi-centered transformations that can be then used for simulation of different one-pot reactions. For this category of reactions, the similarity model and the single step simulation process may be used in searching for the adequate combination of reaction centers changed in substrates/products to study reaction mechanisms. Another application is the designing and generating (according to a given similarity/diversity criterion) of libraries of products for combinatorial synthesis (in the future).

References

[1] E.J. Corey and W.T. Wipke: “Computer-Assisted Design of Complex Organic Syntheses”, *Science*, Vol. 166, (1969), pp. 178–192.

[2] M. Bersohn: “Automatic Problem Solving Applied to Synthetic Chemistry”, *Bull. Chem. Soc. Jap.*, Vol. 45, (1972), pp. 1897–1903.

[3] K.K. Agarwal, D.L. Larsen and H.L. Gelernter: “Application of Chemical Transforms in SYNCHEM2, a Computer Program for Organic Synthesis Route Discovery”, *Comput. Chem.*, Vol. 2, (1978), pp. 75–84.

[4] R. Barone, M. Chanon and M.L. Contreras: “Microcomputers and Organic Synthesis. Application of an Interactive Program to the Photochemical Synthesis of Pheromones”, *Nouv. J. Chim.*, Vol. 8, (1984), pp. 311–315.

[5] J. Dugundji and I. Ugi: “An Algebraic Model of Constitutional Chemistry as a Basis for Chemical Computer Programs”, *Top. Curr. Chem.*, Vol. 39, (1973), pp. 19–64.

[6] J.B. Hendrickson: “Systematic Synthesis Design. IV. Numerical Codification of Construction Reactions”, *J. Am. Chem. Soc.*, Vol. 97, (1975), pp. 5784–5800.

[7] J. Koca: “A Mathematical Model of Realistic Constitutional Chemistry. A Synthon Approach. II. The Model and Organic Synthesis”, *J. Math. Chem.*, Vol. 3, (1989), pp. 91–115.

[8] W.D. Ihlenfeldt and J. Gasteiger: “Computer-Assisted Planning of Organic Syntheses: the Second Generation of Programs”, *Angew. Chem. Int. Ed. Engl.* Vol. 34, (1995), pp. 2613–2633.

[9] G. Fic and G. Nowak: “The CSB Approach to Prediction of the Course of Chemical Reactions”, *Chemom. Intell. Lab. Syst.* (2003, send).

[10] R. Höllering, J. Gasteiger, L. Steinhauer, K.P. Schulz and A. Herwig: “Simulation of Organic Reactions: From the Degradation of Chemicals to Combinatorial Synthesis”, *J. Chem. Comput. Sci.*, Vol. 40, (1999), pp. 482–494.
[11] G. Fic and G. Nowak: “Extended Concept of Reaction Generators in an Organic Reaction Prediction System CSB”, *Comput. Chem.*, Vol. 22, (1998), pp. 141–146.

[12] G. Nowak and G. Fic: “Implementation of Commonsense Reasoning in Organic Reaction Prediction System CSB”, *Comput. Chem.*, Vol. 22, (1998), pp. 147–152.

[13] G. Fic and G. Nowak: “Implementation of similarity model in the CSB system for chemical reaction prediction”, *Comput. Chem.*, Vol. 25, (2001), pp. 177–186.

[14] C.O. Kappe: “100 Years of the Biginelli Dihydropyrimidine Synthesis”, *Tetrahedron*, Vol. 49, (1993), pp. 6937–6963.

[15] C.O. Kappe: “Recent Advances in the Biginelli Dihydropyrimidine Synthesis. New Tricks from an Old Dog”, *Acc. Chem. Res.*, Vol. 33, (2000), pp. 879–888.

[16] C.O. Kappe, D. Kumar and R.S. Varma: “Microwave-Assisted High-Speed Parallel Synthesis of 4-Aryl-3,4-dihydropyrimidin-2(1H)-ones using a Solventless Biginelli Condensation Protocol”, *Synthesis*, Vol. 10, (1999), pp. 1799–1803.

[17] F.S. Falsone and C.O. Kappe: “The Biginelli dihydropyrimidone synthesis using polyphosphate ester as a mild and efficient cyclocondensation/dehydration reagent”, *ARKIVOC*, Vol. 2(ii), (2001), pp. 122–134; http://www.arkat-usa.org/ark/journal/2001/Sauter/FS-155F/155.asp or http://www.arkat-usa.org/ark/journal/2001/Sauter/FS-155F/FS-155F.pdf.
Fig. 1 Starting materials and reactive sites recognized by the CSB in modeling the Biginelli three-component condensation (symbols $s$, $d$ denote single, double bonds respectively, $p$ describes electron pair at the given atom.)
\[ A - B + C - D \rightarrow A - C + B - D \]  
\[ A = B + D - C - E \rightarrow D - A - E + B = C \]  
\[ A - B + C - D + E = F \rightarrow A - E - C + B - F - D \]  
\[ A - B + C - D + E = F + G = H + K - I - L \rightarrow A - E - I - G - C + B - F - K + D - H - L \]  
\[ A - B + C = D + E \rightarrow A = C + E - D - B \]

**Fig. 2** Reaction transformations used for the simulation experiments (I, II, and III – for multi-step simulation, IV- for single step modeling, and V- in acidic conditions). Subsequent symbols A, B, C, . . . , L stand for nodes of reaction centers; they may represent individual atoms, or groups of atoms in the molecule.
Fig. 3 Reaction tree generated by the CSB in first multi-step simulation experiment.
Fig. 4 Set of conceivable intermediate candidates in first level of the reaction tree.
Fig. 5 Reaction tree (further part) generated in the first simulation experiment.
Fig. 6 Reaction tree generated in the second simulation.
Fig. 7 Some of final products generated by the CSB.
Fig. 8 The most probable reaction paths generated by the CSB.