Iterated Reaction Graphs: Simulating Complex Maillard Reaction Pathways

Shail Patel,[†] Jeremy Rabone,[†] Stephen Russell,[‡] Jos Tissen,[‡] and Werner Klaffke[‡,§]

Unilever Research Port Sunlight Laboratory, Bebington, Wirral, L63 3JW, UK, Unilever Research Vlaardingen Laboratory, Olivier van Noortlaan 120, NL-3133AT Vlaardingen, The Netherlands, and Westfälische Wilhelms-Universität Münster, Organisch-chemisches Institut, Corrensstrasse 40, D-48149 Münster, Germany

Received October 7, 2000

This study investigates a new method of simulating a complex chemical system including feedback loops and parallel reactions. The practical purpose of this approach is to model the actual reactions that take place in the Maillard process, a set of food browning reactions, in sufficient detail to be able to predict the volatile composition of the Maillard products. The developed framework, called *iterated reaction graphs*, consists of two main elements: a soup of molecules and a reaction base of Maillard reactions. An iterative process loops through the reaction base, taking reactants from and feeding products back to the soup. This produces a reaction graph, with molecules as nodes and reactions as arcs. The iterated reaction graph is updated and validated by comparing output with the main products found by classical gas-chromatographic/mass spectrometric analysis. To ensure a realistic output and convergence to desired volatiles only, the approach contains a number of novel elements: rate kinetics are treated as reaction probabilities; only a subset of the true chemistry is modeled; and the reactions are blocked into groups.

1. INTRODUCTION

This study introduces a new method of simulating the Maillard process, a set of complex food browning reactions with parallel reactions and feedback loops. The practical purpose of this approach is to model the actual reactions that take place in the Maillard process, in sufficient detail to be able to (a) predict the volatile composition [less than 5% of mass balance] based on the reactants and pH and (b) identify the reaction pathways to both the desired flavor profile as well as off-flavors.

There are a number of approaches in the literature that simulate reaction pathways either synthetically, e.g. PSYCHO,[1] DARC-SYNOPSIS,[2] and REACTION[3] or retrosynthetically, e.g. LHASA,[4] RETROSYN,[5] OCSS,[6] and SYNCHEM.[7] Bador[8] et al. give an excellent review of these approaches. These methods are sufficient to identify reaction pathways to single product or intermediate molecules. To predict the outcome of the Maillard process, a system that can cope with inherent parallelism and feedback loops, and operate without user interaction to construct the complete reaction graph, is preferred. Prickett and Mavrovouniotis[9] have developed a theoretical system that models generic complex reaction systems. This iteratively applies known elemental reaction steps, according to theoretical chemistry, to the reactants and all intermediates. Their focus is on a theoretical description of a generic representation language and how it deals with complex molecular forms and does not model the reaction rate kinetics. This is difficult for two reasons. First, it is impossible to state in advance what the reaction pathways and intermediates are for any particular combination of reactants. Even if this were possible, there would be too many kinetic parameters to fit using regression with experimental data.

Given that our goal is to predict the final volatile composition of the Maillard reaction, we are concerned with ensuring that the simulation output matches as closely as possible with experimental values. Specifically this can be measured in terms of the presence and absence of molecules and matching the peak heights. To achieve this we have developed a system with the following novel characteristics: (1) modeling a subset of all possible reactions, i.e., those that are specific to Maillard chemistry; (2) modeling of the rate kinetics with reaction rate probabilities; and (3) blocking of the reactions into logical groups.

We describe how we have fine-tuned the reaction and rate probability databases by comparison with experimental results and demonstrate the benefits of our approach. In this work we have focused on the Maillard reactions; however, the iterated reaction graphs (IRG) framework is generic and may be applied to any chemistry.

2. METHOD

2.1. Experimental Method. While the focus of this article is on the computational simulation, an important component of the work is the validation with experimental results. A brief description is given here. Samples were produced under well-defined standard conditions. A mixture of amino acid(s) and sugar(s) was heated to 120 °C in glycerol solvent buffered to a pH of 3.0, 6.5, or 7.9, cooled, and then extracted. The composition of volatile products was determined by gas chromatography. The identity of each peak in the gas chromatogram was determined using mass spectrometry and by comparing the generated fragmentation pattern with a library. From this molecular mass distribution (MMD)

* Corresponding author email: Shail.patel@unilever.com.
[†] Unilever Research Port Sunlight Laboratory.
[‡] Unilever Research Vlaardingen Laboratory.
[§] Westfälische Wilhelms-Universität Münster.

10.1021/ci000399a CCC: $20.00
Published on Web 05/02/2001

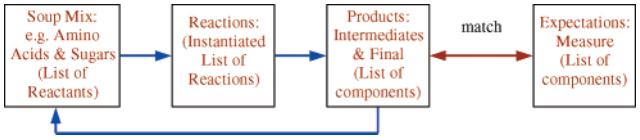


Figure 1. The IRG consists of a soup and a set of reactions. The soup mix is initialized to the desired starting reactants (amino acids and sugars). The set of reactions is a set of computational functions that operate on the soup molecules. They output intermediates and eventually final products. This process takes place iteratively, adding the intermediate outputs to the soup. At the end of the iteration, the final products are compared with experimental measures.

patterns were reconstructed, representing the frequency of masses of the product composition for each individual experiment. These MMD were then directly compared with the virtual mass distribution (VMD) output of the IRG, as discussed below.

2.2. Overview of Iterated Reaction Graphs (IRG). Iterated reaction graphs model complex reaction pathways by simulating the reaction steps in parallel. An iterated reaction graph has two main elements: (1) a "soup" of molecules representing the current state of the system and (2) a set of reactions that may take place in the system of interest.

Molecules are represented as two-dimensional connection tables within the program but for ease of understanding may be expressed as SMILES[10] a simple line notation. Each reaction is coded as a computer program that takes connection table input (reactants), carries out necessary rearrangements (reactions), and produces a connection table output (products).

The initial reactants, or "start soup", typically 100−500 molecules, are determined at the start, and the size is limited only by computer memory considerations. At the start of a run this will be composed of amino acids and sugars only, e.g. for glucose and threonine:

"100 × C(O)C(O)C(O)C(O)C(O)C=O

100 × NC(C(C)O)C(=O)O

..."

There are duplicates of molecules, as the relative number of times a molecule appears simulates the concentration of that molecule in the soup. During, and at the end of a run, the soup will contain a list of end products that is the result of iteratively applying many thousands of reaction steps. It also may contain duplicates, to simulate the relative concentration of end products, e.g.:

"2 × C(=O)(C(=O)C)O

6 × C(=O)(C(O)C(=O)C)O

1 × C(=O)(C(O)C)O"

At a simplistic level the reaction base operates on the molecular soup to form products:

reaction base: molecular soup → products

The full complexity of the possible reactions is modeled by iterating through this "equation", feeding the products back in to the molecular soup, and running through the reaction base again (Figure 1). To refine the system, the output is compared with experimental results and the reaction base is modified.

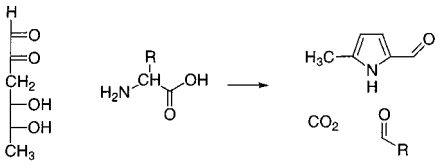


Figure 2. A reaction showing the formation of 5-formyl-2-methyl pyrrole from a hexose sugar derivative and an amino acid. The IRG coding for this reaction is given in Chart 2.

The full reaction graph,[10−14] where molecules are nodes and reactions are arcs may be defined as the set of triplets:

{<reactant> <reaction> <product>}

For example the text below is a small fragment of a reaction graph:

C(=O)C(O)C(O)C(O)C(O)C(O)	R1_1_5_asugar	CC(C(O)C(C(O)CO)=O)=O
CC(C(C(O)C(O)CO)=O)=O	R1_3_2_retroaldol	C(C(O)CO)=O
CC(C(=O)C)=O	R2_5_10a_strecker	CC(C(C)N)=O

The full graph is reconstructed by linking products to reactants and chaining through the triplets (cf. Figure 5). The IRG approach may also be described as chaining through a parallel rule-base of reactions,[4−7] or an evolutionary system[15,16] where the reaction base represents the evolutionary operators (see Chart 1 for an overview of program operation).

2.3. Designing the Reaction Base. Central to the working of the program is the reaction base: a computer simulation of the chemical reactions that actually take place during the Maillard reaction. Each virtual reaction is coded as a program function that conducts the following steps: (1) 2-D pattern match on reactive fragment of reactant (input) molecule(s); (2) check for required adjacent atoms; (3) break bonds; (4) add bonds (change atom hybridization); and (5) output product molecules(s).

The pattern-matching step allows for fragment matching on the connection table of the reactive fragment necessary for the reaction to take place. Thus the Maillard process is coded as a set of generic reactions which can act on a range of different starting molecules.

Figure 2 shows a reaction for the formation of 5-formyl-2-methylpyrrole. This reaction takes a sugar derivative and an amino acid as reactants and providing that the necessary constraints are met produces the pyrrole along with carbon dioxide and an aldehyde (see Chart 2 for the coding for this reaction in IRG).

This particular reaction and others in the reaction database could have been written as a combination of several smaller reactions. In early versions we modeled reactions at a low level of granularity; however, cycling through reversible reactions lead to a large computational overhead. For example, the rearrangement of α-hydroxycarbonyls and Schiff's base reactions accounted for the vast majority of reactions, but because of their reversibility they did not generate sufficient "concentrations" of forward reaction products. Slowing the rates of the respective reverse reactions proved detrimental to all other reactions for which these steps were vital. Therefore these, and other elementary reactions, were combined into single reactions without forming the intermediate compounds. Thus, the reaction of an amine with an aldehyde is now immediately followed by Strecker

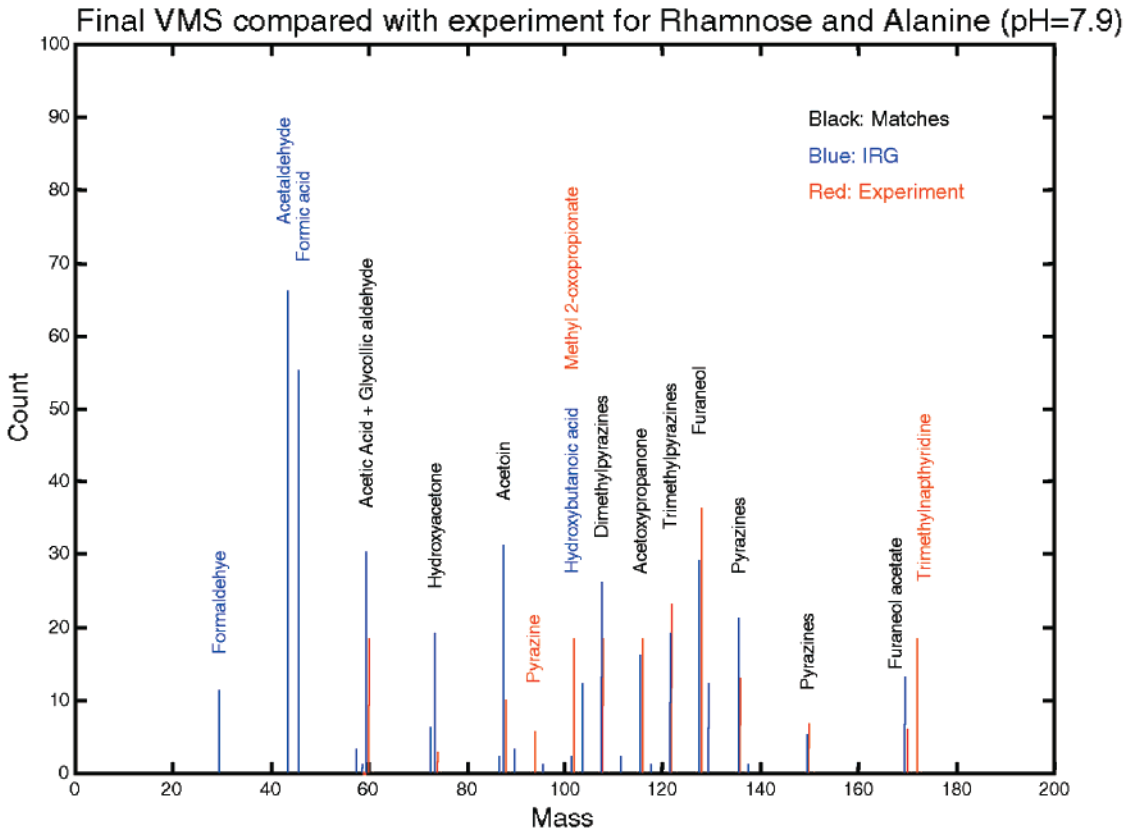


Figure 3. The graph compares the virtual mass distribution results of an IRG run (blue spikes offset 0.2 to the left) on rhamnose and alanine with experimental mass distribution results (red spikes offset 0.2 to the right). Many of the peaks have been identified and are labeled. Those in black show the matches, where the output of the IRG correctly matches the experimental GC results. Mismatches may be due to a number of causes: missing reactions from the database; chemically unfavorable reactions being allowed by the model; and compounds are not detected by the GC. The lower mass products such as formaldehyde and acetaldehyde are usually ignored along with any left over amino acid—it is quite probable that these form nonvolatile products under experimental conditions. False positives are in blue, false negatives in red (a peak due to left over alanine (mass 89, height 47) was removed for clarity).

degradation without formulation of the intermediate azomethine. This speeds up the execution of IRG considerably, although several partial reactions, e.g. the formation of a Schiff's-base, have to be coded more than once into other reactions that also use azomethines.

Furthermore, we are interested only in the volatile components, which represent less than 5% of the mass balance. So for efficiency we have designed the simulation so as to model not the true chemistry but a subset of the chemistry that leads directly to the volatiles, for example by omitting keto−enol tautomerizations and writing enol outputs in the keto form.

Based on approximately 1500 references on the mechanistic aspects of the Maillard reaction (a selection[17−36] of these is referenced here), we have written a set of about 100 Maillard reaction steps.

2.4. Blocking the Reaction Database. For further efficiency we have split the reaction database into blocks of reactions. Each block consists of a subset of reactions chosen roughly according to the order in which reactions occur in the Maillard process. The blocks are run sequentially with the output from one block being used as the input to one or more further blocks. There are two reasons for using the following approach: (1) time is saved because reactions are not called before their reactants are available; for example one reaction which produces pyrazines requires aminocarbonyl reactants which are only made after several reactions (cf. Figure 5) and (2) reactions which would otherwise produce large, unwanted products, i.e., nonvolatiles, are not run until their reactants have been reduced in size by e.g. the sugar degradation reactions.

Using blocks in this way is not as strongly sequential as it may appear; parallel reactions may take place within each block, the same reaction may occur in more than one block, and there is a high level of traffic between the blocks (cf. Figure 6). There is the risk that certain side reactions are missed, but the benefit is that the program finishes within tractable time scales and that the simulation is pushed toward the volatile end products (cf. Table 1).

Finally, a mass limit on the molecules in the soup was introduced to prevent polymerization and focus on volatile production.

2.5. Simulating Rate Kinetics/Reaction Rate Probabilities. For a simple chemical reaction A + B → P, where A and B are reactants and P is the product molecule, we have

$$\frac{\mathrm{d[P]}}{dt} = \frac{-d[\mathrm{A}]}{\mathrm{d}t} = k_{\mathrm{ABP}}[\mathrm{A}][\mathrm{B}] \qquad (1)$$

where k_{ABP} is the rate constant for that reaction.

Discretising eq 1 we obtain

$$\Delta[\mathrm{A}] = -k_{\mathrm{ABP}}[\mathrm{A}][\mathrm{B}]\Delta t \qquad (2)$$

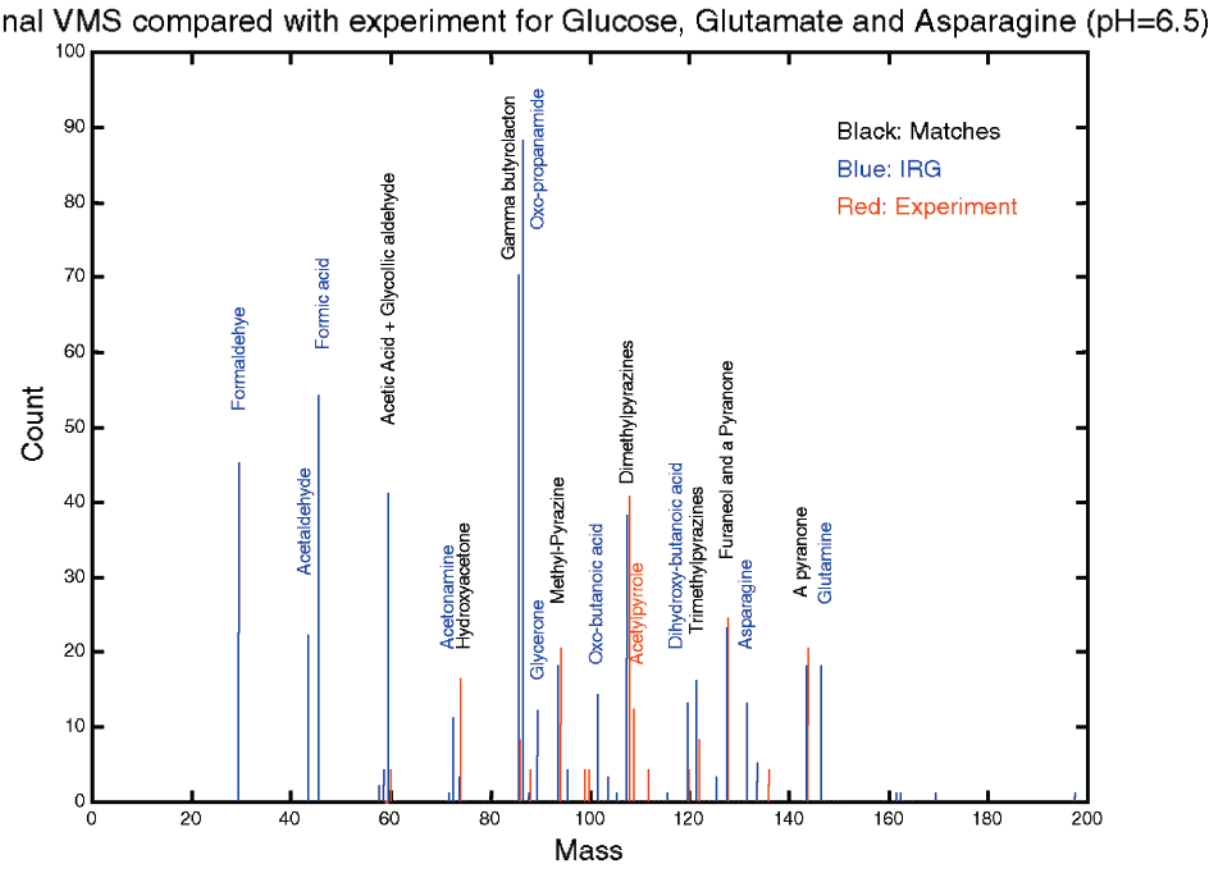


Figure 4. A similar graph to Figure 3 for glucose, glutamate, and asparagine. There are significantly more unmatched smaller peaks and IRG produces more compounds which are not found in experiment. Several of these are derived from the amino acid side chains for which we are still identifying and testing appropriate reactions.

Losing the time step Δt in the constant of proportionality and describing values as relative probabilities, this may be written as

$$\Delta(n(\mathrm{A})) \propto -p(\mathrm{R_{ABP}})p(\mathrm{A})p(\mathrm{B}) \qquad (3)$$

where $n(\mathrm{A})$ = number of molecules of A in the soup, $p(\mathrm{R_{ABP}})$ = relative "probability" of reaction A + B → P, and $p(\mathrm{X})$ = probability of selecting molecule X from the soup.

In IRG the joint probability $p(\mathrm{A})p(\mathrm{B})$ is simulated by picking a random pair of molecules {<molecule1>, <molecule2>}. This selection is biased by the "concentrations" of molecule1 and molecule2 in the soup and, therefore, over successive selections, is a reasonable approximation to the probability. $p(\mathrm{R_{ABP}})$ is simulated by assigning a "probability of reacting" to each reaction R and probabilistically selecting random reactions on a roulette wheel basis.[15,16] Initially the probabilities were all set to unity and then updated with reference to experimental data (2.6 below). If the selected molecules match the requirements of the reaction R, then it reacts and the products are added to the soup. The implication is that reaction−molecule matches are tried many thousands of times, sometimes with repetition. For efficiency considerations a look-up table for common reaction−molecule matches may be implemented; however, this should take into account that there may be multiple (potentially isomeric) reaction sites on any particular reactant, resulting in quite different products.

2.6. Updating the Reaction Database and Rate Probabilities. After each run the simulation output was compared with the experimental data to identify the false negatives, i.e., molecules produced by the experimental method and not the IRG, and false positives (vice versa). The false negatives were dealt with by two methods. First, the reaction database was modified to include any missing essential reactions on the path to the end product or by altering the pattern-matching step. In most cases, these changes would be referenced from the literature. Second, if there were potential reaction pathways to the molecules produced by experimentation, the reaction rate probabilities for the reactions were altered so as to favor those potential pathways. The plots of reaction pathways (Figure 5) were useful in determining which pathways to alter.

An example is the formation of pyrazines, which are significant products formed in the Maillard reaction. Pyrazine precursors result from the condensation reaction of two α-amino ketone units that are only formed after several reaction steps. Initially it was thought that the α-amino ketones only resulted from a Strecker reaction between a diketone and an amino acid. The yield of pyrazines was found to be very small, but the reaction paths showed that those diketones, which were formed, almost exclusively went into forming pyrazines. Therefore, reactions forming α-amino ketones from hydroxy ketones were sourced from the literature and added to the reaction database. Furthermore, the rate-limiting step was found to be the formation of amino carbonyls which initially were formed slowly resulting in a low yield of pyrazines. Therefore the reaction rate probabilities were increased for all of the reactions responsible for the production of amino carbonyls.

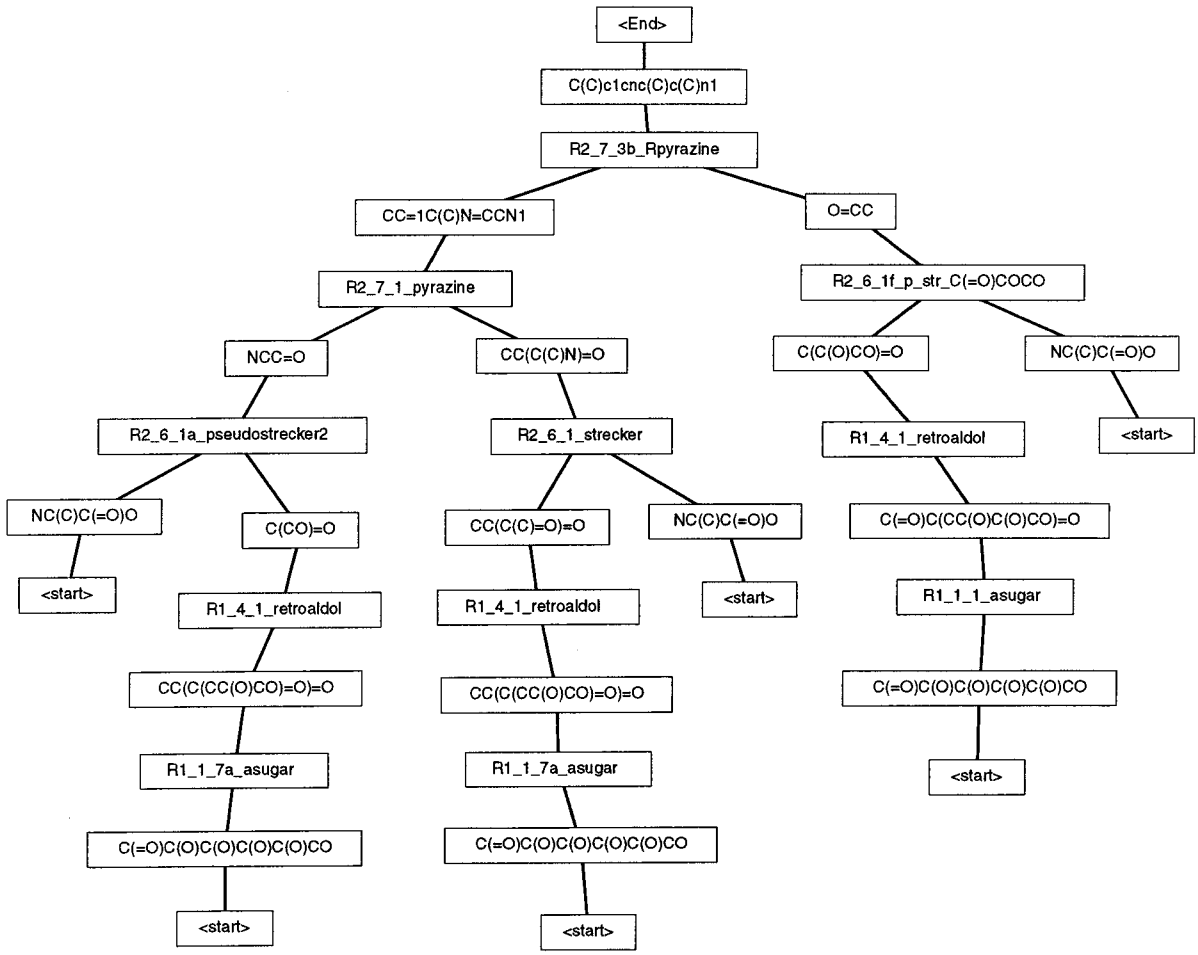


Figure 5. This is a single pathway from the reaction graph for glucose and alanine at pH 7. The beginnings of the path are shown with ⟨*start*⟩, the end of the path by ⟨*end*⟩. The graph only shows the reaction products that go into the production of the resulting pyrazine (i.e. there are many more branches from the reactions which lead to other products).

Chart 1. Program Operation[a]

```
Initialise Soup, Reactions, Database etc
 Loop
   Loop through Reaction Blocks
        Select Random reaction
        If (reaction probability > random number)
              Select random reactant(s)
              If reactant(s) are correct for reaction
                    Remove bonds
                    Add bonds & change atom hybridisation
                    If (mass of product < mass limit)
                          Remove reactants from soupmix
                          Add product(s) to soupmix
                    Endif
              Endif
        Endif
   Endloop
 Endloop
```

[a] The top level process iterates through the reactions base, selecting molecules from the "soup" and reactions from the list of reactions [italics indicate extra code for block-approach]. IRG was initially implemented in the Sybyl Programming Language (SPL),[37] but has now been ported to C++ with 2 orders of magnitude speed improvement.

The false positives were initially less of a concern as these may be side products of important reactions and may react further in reactions that lead to nonvolatiles. The results show these were dealt with by the blocking procedure.

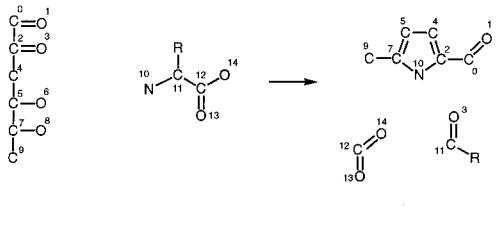
3. RESULTS

3.1. Experimental Validation with Virtual Mass Distribution. The final output of the computational IRG contains the "soup" of molecules at the end of the run, the full set of reaction paths to all molecules, and various run statistics. The output soup may be represented as a "virtual mass distribution" (VMD) by taking relative frequencies bucketed by molecular weight. We may then compare the experimental MMD with the VMD. In Figure 3, the MMD is in red, the VMD is in blue, and the matches have been labeled in black. For this mixture IRG clearly produces most of the compounds identified in the MMD with the exception of the two pyrazines and methyl 2-oxopropionate. IRG also produces additional compounds such as the lower mass aldehydes and some left over amino acid (removed from Figure 3) which are not found in the MMD. It must be borne in mind however that the MMD only analyzes the 5% volatiles produced by the experiment and that the material left over by IRG could well be the reactants which would produce the nonvolatile components of the reaction.

Similar results for glucose, glutamine, and asparagine are shown in Figure 4.

3.2. Performance Results of Blocking and Reaction Rate Probabilities. For objective comparison of the performance of our system we have devised four indicative measures. These are as follows: (1) matches, the number of molecules the IRG has correctly identified as being present in the final soup; (2) false_positives, the number of molecules the IRG has incorrectly identified as being present in the final soup; (3) false_negatives, the number of molecules the

Chart 2. Reaction Coding[a]



1. BINARY
2. R2_3_10b_rhamMeCHOpyrrol
3. C(=O)C(=O)CC(O)C(O)C
4. NCC(=O)O
5. 0 4 5 6 7 8 9 10 11 14
6. H H H H H H H HC3 H
7. N N N N N N N N Y N
8. NA
9. 2 3 2 4 5 6 5 7 7 8 10 11 11 12 12 14
10. 2 4 2 2 10 1 5 7 2 7 10 1 11 3 2 12 14 2

[a] The diagram shows a schematic of the reaction to form 5-formyl-2-methylpyrrole from a hexose sugar derivative and an amino acid. Below it is the corresponding entry in the reaction database. The top lines describe the reaction type (binary) and its ID. Lines 3 and 4 are the pattern matching templates for the reactive fragments of each reactant. Each atom is identified by its position in the strings on lines 3 and 4, shown also in the schematic. Lines 5, 6, and 7 restrict the types of atoms which may be bonded to any nonterminal atoms in the fragment. For example, atom 11 may be bonded to a hydrogen or sp^3 carbon (i.e. allowing an R-group), the Y meaning that it has to be bonded to at least one hydrogen. Line 8 allows for ubiquitous molecules, e.g. water, none in this case. Line 9 identifies the pairs of atoms between which bonds are broken, and line 10 identifies the pairs of atoms between which bonds are created followed by the bond type.

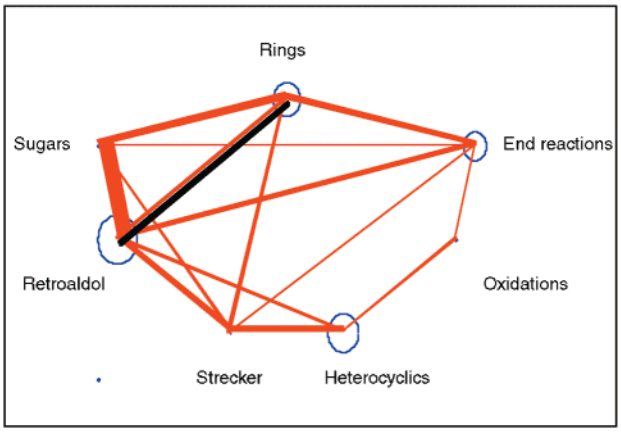


Figure 6. This graph shows the reaction traffic, with the thickness of the line proportional to the number of intermediate molecules output by one block, that are used as input to the next block. Red lines are in the forward direction; black lines are reverse reactions. The size of the circle shows the traffic within a reaction block.

IRG has failed to identify as being present in the final soup; and (4) difference, the difference in peak heights = $\Sigma_i|MMD\_peak\_height_i - n*VMD\_peak\_height_i|^p$—the absolute difference of normalized peak heights, summed over j, the correct matches and raised to some power p (typically = 1). The normalization factor n corrects for the difference in scaling between the experimental data and the IRG output and was equal to $0.5*\Sigma_i MMD\_peak\_height_i/\Sigma_i VMD\_peak\_height_i$.

Summary results for the IRG on different combinations of amino acids and sugars are given in Table 1 below.

To compare the performance of the blocking procedure and the use of reaction rate probabilities, we have run the IRG over the 12 combinations for which we have full experimental data and varying two dimensions: (a) (i) no blocking of reactions in the reaction database; (a) (ii) full blocking; (b) (i) all reaction rate probabilities set = 1; and (b) (ii) reaction rate probabilities set by heuristic updating. The results for the four indicative measures are shown in Table 2. We can see a number of qualitative characteristics: (1) The false negatives are already fairly low, largely due to the careful restriction of reactions in the reaction database to those that are relevant to the Maillard process. (2) Blocking removes about half of the false positive, i.e., the blocking procedure helps to drive the IRG toward the key volatiles and prevents many of the sideways reaction paths. (3) Without blocks the reaction rate probabilities show little improvement on performance. (4) With blocking the rates reduce the peak height error i.e., bringing the IRG output to match the concentrations of the experimental results more closely.

Table 1. Comparison of IRG Output with the Experimental Data on 12 Combinations of Amino Acids and Sugars[a]

amino acid	sugar	matches	false +ve	false −ve	diff
alanine	rhamnose	9	14	4	10.18
tyrosine + asparagine	fructose	5	8	12	20.39
valine + glutamate	fructose	8	12	9	19.28
glutamate + asparagine	glucose	10	9	5	14.24
histidine + alanine	rhamnose	9	14	5	25.68
lysine + alanine	rhamnose	8	18	10	18.99
lysine + threonine	rhamnose	11	13	9	16.92
glutamine	xylose + fructose	10	12	5	11.07
glutamine + threonine	xylose	11	14	4	11.79
proline + glutamine	xylose	6	16	9	18.07
lysine	xylose + rhamnose	13	12	4	8.54
serine + proline	xylose	5	17	6	33.02

[a] Analysis in the mass range 50−200. Many of the false positives can be discounted as these represent intermediates which would react further to form nonvolatile final products, e.g. aldehydes and unreacted amino acids.

Table 2. Average Results Using 12 Different Sets of Starting Materials[a]

	matches			difference	
	unblocked	blocked		unblocked	blocked
probs = 1	7.08	7.92	probs = 1	25.99	23.67
heuristic	7.00	8.75	heuristic	25.13	14.69
	false −ve			false +ve	
	unblocked	blocked		unblocked	blocked
probs = 1	8.50	7.67	probs = 1	23.42	14.75
heuristic	8.58	6.83	heuristic	23.91	13.25

[a] Analysis in the mass range 50−200. This table shows the results of comparing the performance of the IRG with and without blocking and with and without heuristically set reaction rate probabilities. The blocking procedure removes about half the false positives. Setting the reaction rates has little effect without blocking, but with the blocking procedure gives a better match in peak heights (difference measure).

3.3. Visualizing the Reaction Graph. The full reaction pathways to each end product are given in the program output. This is very large and difficult to interpret. A single pathway from the graph for glucose and alanine is shown in Figure 5; this gives the sequence of reactions and intermediates that lead to one of the end products (a pyrazine). The reaction paths give important insights into how each product is formed from the starting materials and how each reaction may depend on previous reactions.

A simplified visualization of the reaction graph for glucose and threonine is shown in Figure 6; this shows the traffic of

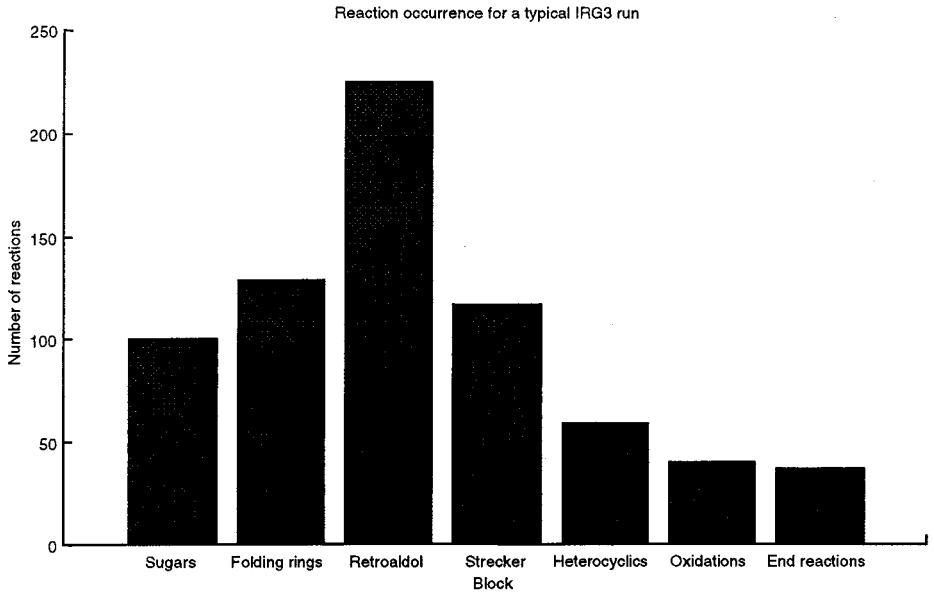


Figure 7. This histogram summarizes the relative frequency of reactions, grouped by the reaction block, of which occurred during a typical IRG run with glucose and threonine. The dominant reactions reflect the relative concentrations of reactants, e.g. retroaldol reactions may occur many times starting with a single sugar molecule whereas a Strecker reaction can only occur once with a single amino acid molecule.

molecules between the reaction blocks (red = forward direction, black = reverse direction). It is interesting to note that there is considerable traffic between the various blocks. For example retroaldol reactions (block 3) are often followed by end reactions (block 7) as unreacted formaldehyde and glycolic aldehyde undergo oxidation and condensation reactions, respectively. The relative size of the circles indicates when a reaction in a block is followed by another reaction in the same block. Sugar reactions (block 1), oxidation reactions (block 6), and Strecker (block 4) have no size because these reaction "blocks" contain reactions that do not feed into each other but simply take an input and pass it through to the next block.

The frequency of each reaction successfully reacting is given in the program output. The total number of reactions in each block may be summarized in a histogram (e.g. Figure 7). This shows the predominance of the retro-aldol reactions in the Maillard process for glucose and threonine.

3.4. Learning about Maillard Chemistry. During the process of optimizing rates and the reaction database we have gained a number of insights: (1) Certain reactions can be "assumed" giving a much higher performance when simulating a complex system without losing much generality. Reactions such as keto−enol tautomerism and Schiff-base reactions occur so readily in both the forward and backward directions that their products may be taken directly into a subsequent reaction. (2) Some compounds tend to react mainly via one pathway e.g. rhamnose always forms a lot of furaneol. (3) The result of the initial sugar degradation reactions largely determines the final outcome of an IRG run. These reactions also tend to be highly pH dependent. Therefore obtaining a balance of the reaction rate probabilities for these reactions is important. (4) Several reactions in the database were not found by directly researching the Maillard reaction but were identified as reactions necessary for the production of some of the products found in the experiments. E.g. gamma butyrolactone is an important Maillard product when xylose or glutamate is present, but there are no published mechanisms for either case. (5) Redox mechanisms in Maillard chemistry play at least a small role. They are currently a gray area with little published information but are suspected to involve smaller fragments such as formic acid and formaldehyde.

4. FURTHER WORK

We are currently investigating ways of automatically optimizing the reaction rate probabilities to match more closely with experimental results. An error function related to the experimental measures for a particular reactant mix may be defined as

$$\text{Error}(R, S) = w1*\text{False\_Positives}(R,S) + w2*\text{False\_Negatives}(R,S) + w3*\text{Difference}(R,S)$$

where R = the set of reaction rate probabilities at a specified pH; S = the start soup; and w1,w2,w3 are a set of weights $>= 0$ (cf 3.2 for other definitions).

An objective function summed over the reactant mixes, or start soups for which there is experimental data, may be defined: $O(R) = \Sigma_S \text{Error}(R, S)$. Clearly as O(R) approaches 0, the IRG is producing results closer to the experimental values. We may therefore define the optimization problem to be to optimize R, i.e., the reaction rate probabilities for a given pH, such that O(R) is minimized. This is computationally expensive, but we are investigating how to achieve this using standard optimization algorithms such as sequential quadratic programming or more recent evolutionary methods, e.g. genetic algorithms.[15,16]

5. CONCLUSION

We describe a new approach, iterated reaction graphs (IRG), that simulates complex chemical reaction systems. The IRG is composed of two parts, a soup of molecules and a reaction base. An iterative procedure loops through these elements, simulating the reactions that take place. To ensure

a realistic output and a convergence on desired volatiles (ca. 5% of experimental output), the approach contains a number of novel elements: (1) rate kinetics are treated as reaction probabilities; (2) only a subset of the true chemistry is modeled; and (3) the reactions are blocked into groups.

This approach is generic to any system of complex chemical reactions, with specific knowledge contained within the reaction base. We have applied this to the Maillard process, a set of food browning reactions, with parallel reactions and feedback loops. We have refined our understanding of the reactions, and the rates of reactions, by comparing the IRG output with experimental data (gas-chromatography/mass-spectrometry). In the process of developing the reaction base, we have increased our understanding of the Maillard reactions, with the potential for recognizing new pathways and intermediates.

The main benefits of this approach is that it can predict the volatile composition of the products and identify the principle reaction pathways to both desired flavor profiles as well as the production of off-flavors even without an a priori understanding of the reaction kinetics. Once set up correctly, it can operate on arbitrary N-ary mixtures of amino acids and sugars.

REFERENCES AND NOTES

(1) Jauffret, P.; Laurenco C.; Kaufmann G. Psycho − A Computer Program for Organic Synthesis. *Technique Sci. Informatiques* **1986**, *5*, 375−390

(2) Picchiottino, R.; Georgoulis G.; Sicouri G. Darc-Synopsis − Designing Specific Reaction Data Banks − Application to Keto-React. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 241−249.

(3) Blurock, E. S. REACTION, System for modelling chemical reactions. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 607−616.

(4) Corey E. J.; Long A. K.; Rubenstein S. D. Computer-Assisted Analysis in Organic Synthesis. *Science* **1985**, *228*, 408−418.

(5) Blurock, E. S. Computer Aided Synthesis Design at RISC−Linz − Automatic Extraction and Use of Reaction Classes. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 505−510.

(6) Corey, E. J.; Long A. K.; Rubenstein S. D. Computer-Assisted Analysis in Organic Synthesis. *Science* **1985**, *228*, 408−418.

(7) Gelernter, H.; Rose J. R.; Chen C. Building and Refining a Knowledge Base for synthetic Organic Chemistry via the Methodology of Inductive and Deductive machine Learning. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 492−504.

(8) Bador, P.; et al. Les Systemes Informatiques de Recherche d'Information sur les Reactions Chimiques et les Systemes de Synethese Assistee par Ordinateur. *New J. Chem.* **1992**, *16*, 3, 413−423.

(9) Prickett, S. E.; Mavrovouniotis, M. L. Construction of Complex Reaction Systems − II Molecule Manipulation and reaction application algorithms. *Comput. Chem. Eng.* **1997**, *21*, 11, 1237−1254.

(10) Weininger, D. SMILES, a chemical language and information system. *J. Chem. Inf. Comput. Sci.* **1998**, *28*, 1, 31−36.

(11) Lohn, J. D. *Evolving Catalytic Reaction Sets using Genetic Algorithms, IEEE World Congress Computational Intelligence Anchorage, Alaska*; 1998; pp 487−492.

(12) Schuster, P. *Dynamical Systems and Cellular Automata;* Demongeot, J., et al., Eds.; Academic Press: 1985; pp 255−267.

(13) Banzhaf, W.; et al. Emergent Computation by Catalytic Reactions. *Nanotechnology* **1996**, *7*, 307−314.

(14) Kauffman, S. A. *The Origins of Order*; Oxford University Press*:* 1993; pp 303−305.

(15) Goldberg, D. E. *Genetic Algorithms in Search, Optimisation, and Machine Learning*; Addison-Wesley: Reading, MA, 1989; pp 148−200.

(16) Holland, J. H. *Adaptation in natural and artificial systems*; The University of Michigan Press: Ann Arbor: MI, 1975.

(17) Skog, K.; Johansson, M.; Jaegerstad, M. A review of mutagenic heterocyclic amines. *Scand. J. Nutr.* **1994**, *38*, 30−33.

(18) Baltes, W. *Maillardreaktionen in Lebensmitteln*; Lebensmittelchemie, 1993; Vol. 47, pp 9−14.

(19) Pikul, J. Lipid oxidation and formation of aroma and flavour in heated and stored meat. II. *Gospodarka Miesna* **1992**, *44*, 22−26.

(20) Friedman, M. Dietary impact of food processing. *Annu. Rev. Nutr.* **1992**, *12*, 119−137.

(21) Whitfield, F. B. Volatiles from interactions of Maillard reactions and lipids. *CRC Crit. Rev. Food Sci. Nutr.* **1992**, *31*, 1−58.

(22) Anon. Maillard reaction products: Safety and physiologic effects. *Comments Agric. Food Chem.* **1994**, *3*, 111−128.

(23) Horiuchi, S. Advanced glycation end products (AGE)-modified proteins and their potential relevance to atherosclerosis. *Trends Cardiovascular Med.* **1996**, *6*, 163−168.

(24) Feather, M. S. The mechanism of nonenzymic browning reactions occurring during cookie and cracker making. In *Sci. Cookie Cracker Prod;* Faridi, H., Ed.; Chapman & Hall: New York, NY, Department Biochemistry, University Missouri, Columbia, MO, 1994; pp 439−453.

(25) Rewicki, D.; Kersten, E.; Helak, B.; Nittka, C.; Tressl, R. Mechanistic studies on the formation of Maillard products from [1−13C]-D-fructose. *Spec. Pub. − The Royal Soc. Chem.* **1994**, *151*, 61−68.

(26) McCarthy, A. D. Nonenzymic glycosylation of proteins: its role in the chronic complications of diabetes and aging. *Acta Bioquim. Clin. Latinoamericana* **1995**, *29*, 173−190.

(27) Eichner, K.; Reutter, M.; Wittmann, R. Detection of Amadori compounds in heated foods. In *Thermally Generated Flavors*; ACS Symposium Series; 1994; Vol. 543, pp 42−54.

(28) Feather, M. S. Dicarbonyl sugar derivatives and their role in the Maillard reaction. *Comments Agric. Food Chem.* **1994**, *3*, 69−85.

(29) Cohen, M. P.; Ziyadeh, F. N. Role of Amadori-modified non-enzymically glycated serum proteins in the pathogenesis of diabetic nephropathy. *J. Am. Soc. Nephrol.* **1996**, *7*, 183−190.

(30) Yaylayan, V. A.; Huyghues-Despointes, A. Chemistry of Amadori rearrangement products: analysis, synthesis, kinetics, reactions, and spectroscopic properties. *Crit. Rev. Food Sci. Nutr.* **1994**, *34*, 321−369.

(31) Sakurai, T.; Nakano, M. Glycation and free radicals. Generation of free radicals. *Kassei Sanso, Furi Rajikaru* **1993**, *4*, 126−133.

(32) Moll, N. Nonenzymic browning: reaction scheme. *Ind. Alimentary Agric.* **1993**, *110*, 301−307.

(33) Eichner, K. Chemical changes in food preservation and their inhibition. *Wiss. Z. − Technol. Hochsch. Koethen* **1991**, 1−14.

(34) Hunt, J. V.; Wolff, S. P. Oxidative glycation and free radical production: a causal mechanism of diabetic complications. *Free Radical Res. Commun.* **1991**, *12/13*, 115−123.

(35) Weisburger, J. H. Specific Maillard reactions yield powerful mutagens and carcinogens. *Spec. Pub. − The Royal Soc. Chem.* **1994**, *151*, 335−340.

(36) Ledl, F. Der Abbau von reduzierenden Zuckern und Aminen bei der Maillard-Reaktion. *Z. Ernaehrungswissens.* **1991**, *30*, 4−17.

(37) Tripos Inc.; S. Hanley Road, St. Louis, MO, 1996.

CI000399A