Principal Component Analysis on the effect of nucleophiles on the reactivity of α-acylenaminoketones †

Concetta Kascheres,*[a] Giuseppina Negri,[b] Márcia M. C. Ferreira[c] and Luciana C. Sabino[c]

[a] *Universidade Paulista (UNIP), Instituto de Ciências da Saúde, Avenida Independência, 412. CEP–18087-050. Sorocaba, São Paulo, Brazil. E-mail: conniek@iqm.unicamp.br*
[b] *Universidade Bandeirante de São Paulo (UNIBAN), Rua Maria Cândida, 1813. CEP-02071-013. São Paulo, São Paulo, Brazil. E-mail: negrijun@zaz.com.br*
[c] *Instituto de Química, Universidade Estadual de Campinas (UNICAMP), Caixa Postal 6154. CEP–13083-970. Campinas, São Paulo, Brazil. E-mail: marcia@iqm.unicamp.br*

Received (in Cambridge, UK) 24th April 2001, Accepted 21st September 2001
First published as an Advance Article on the web 8th November 2001

This report shows the importance of Principal Component Analysis for grouping types of products observed when α-acylenaminoketones **K1**–**K3** react with four substituted hydrazine nucleophiles in five organic solvents. The reactions were carried out with the goal of obtaining substituted pyrazoles and determining which of the carbonyls would preferentially be attacked by the nucleophile. The reaction products were submitted to GC-MS analysis and the results were subjected to Principal Component Analysis (PCA). The data set was separated in four groups (scores). The deacetylated pyrazoles **P5** were separated from the other pyrazoles by the first principal component PC1. The second principal component PC2 separated the pyrazoles **P4**, derived from nucleophilic attack on the acetyl carbonyl group, from the pyrazoles **P6**, derived from nucleophilic attack on the carbonyl bonded to the more bulky group R (loadings analysis). The simultaneous analysis of the scores-loadings shows the relationship between the mechanisms (types of reaction products-loadings) and the reaction conditions (solvent, nucleophile). Frontier orbital considerations were also included to complete the analysis.

Introduction

Data interpretation and even experimental optimization have been the domains of statisticians, engineers and managers. Such a division of labor was no doubt necessary before the advent of highly computerized laboratories when obtaining a single datum was time-consuming and costly. However, the microprocessor revolution has enabled scientists to acquire and store great quantities of data easily and cheaply; the bottleneck has become data processing and interpretation.[1] Instead of generating data and analysing them manually, chemists are now able to use multivariate data analysis methods[2–5] to help uncover the meaning of the chemical information they produce.

Most chemical applications of data analysis are by nature multivariate and one of the most suitable methods for analysing these cases is PCA.[2,4] This method is based on the correlation of variables, and is particularly effective when these variables show any degree of correlation. Its aim is to group these correlated variables, generating new sets called "principal components" (PCs) onto which the data are projected. These PCs have the property of being completely uncorrelated and are built as simple linear combinations of original variables. The important point here is that the PCs contain the maximum variability in the data set, in a much lower dimensional space. The first principal component, PC1, is defined in the direction of maximum variance in the data set, and the subsequent components are orthogonal to one another and describe the maximum of the remaining variance. Once the redundancy is removed, only the first few principal components are required to describe most of the information contained in the original data.

The raw data matrix, represented by $\boldsymbol{X}$ ($N \times M$), has in each row the experimental results on a single sample, while each column contains the experimental measurements for a particular variable. Each sample corresponds to a point in the M-dimensional space.

The original data matrix is decomposed into two matrices, represented by $\boldsymbol{T}$ and $\boldsymbol{V}$.

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{V}^{\mathrm{T}} \quad (1)$$

The matrix $\boldsymbol{T}$, known as "scores" matrix, represents the position of the samples in the new coordinate system where the PCs are the axes. The second matrix, $\boldsymbol{V}$, is the "loadings" matrix whose columns (variables) describe how the new axis, *i.e.* the PCs, are built from the old axes.

Hierarchical Cluster Analysis, HCA[2,4] is another important multivariate method of data analysis. Its primary purpose is to display the data in such a way as to emphasize its natural clusters and patterns in the two dimensional space. The results, qualitative in nature, usually are presented in a form of dendograms, making it possible to visualize the similarities among samples or variables. In HCA, the distances between samples or variables are calculated, transformed into a similarity matrix $\boldsymbol{S}$ and then compared. For any two samples k and l, the similarity index is defined as eqn. (2) where S_{kl} is an element of

$$S_{kl} = 1.000 - d_{kl}/d_{\mathbf{max}} \quad (2)$$

$\boldsymbol{S}$, $d_{\mathbf{max}}$ is the largest distance among each pair of samples in the data and d_{kl} is the Euclidean distance between samples k and l. The similarity scale ranges from zero to one. It is clear that the larger the index S_{kl}, the smaller the distance between k and l. Therefore, S_{kl} directly reflects their similarity.

Chemometrics in its present form was started in the 1960s to cope with the ever increasing size of chemical data sets. In analytical chemistry, spectroscopy and gas chromatography started

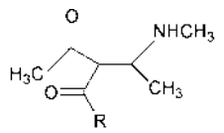
† Electronic supplementary information (ESI) available: calculations for the hydrazines and the α-acylenaminoketones studied in this work. See http://www.rsc.org/suppdata/p2/b1/b103660m/

DOI: 10.1039/b103660m

to provide many variables per analytical sample, often several hundreds.[5] Similarly, other branches of chemistry were becoming increasingly flooded by large data sets from spectroscopy, kinetics, electrophoresis, process sensors, *etc.* In organic chemometrics, methods were transplanted from psychology, *i.e.* factor and principal component analysis (PCA) and similar approaches, for the analysis of both reactivity and other data.[6] Although analytical chemists have been using chemometrics for a number of years, the same is not true for organic chemists.[7–10] The screening of discrete variations in organic synthesis is based upon principal properties, *i.e.* principal component characterization of the constituents defining the reaction system.[5] The development of high-throughput screening (HTS) assays has resulted in the possibility of testing a large number of compounds for biological activity in a fast and automated fashion. HTS often tends to give simple results such as active or not and usually requires identified hits to be tested again in order to verify the result.[6] Fourier transform infrared (FTIR) microspectroscopy, in combination with chemometrics, was investigated as a novel method to discriminate between cyanobacterial strains.[11] Applications of the methods such as response surface methods, simplex optimization with exponential weighing of multiple responses and PLS modelling were utilized by optimization of the $TiCl_4$-mediated synthesis of the morpholine enamine from pinacolone.[12] Optimization of the synthesis of *p*-substituted phenylacetic acid thiomorpholides was achieved by using a fractional factorial experimental design combined with response surface methods.[13] Linear free energy relationships (LFERs) and extra-thermodynamic relationships (ETRs), *i.e.* similarity and analogy models of physical organic chemistry, are mathematically and statistically equivalent to the models much used in chemometrics and data analysis, PCA, PLS and SIMCA (soft independent modelling by class analogy).[14] Chemometric methods such as PCA and SIMCA have been extensively applied to the analysis of infrared spectra in a variety of different areas within the fields of medicine, biology and forensic science.[11]

A successful interpretation of the complex manner by which the GC retention indices of methylalkanes produced by insects were related to chemical structure was achieved using the quantitative structure–property relationship (QSPR) method.[15] A high-speed quantitative analysis of aromatic isomers in a jet fuel sample was performed using comprehensive two-dimensional gas chromatography (GC-GC) and chemometrics.[16] The description of substituent effects by means of mathematical equations containing more than one explanatory variable has been a method widely used over the last decade.[17,18] The dimensionality of the basicity-dependent behavior in the condensed phase of nonprotogenic organic molecules commonly used as solvents, was approached by PCA of a set of basicity-dependent properties related to hydrogen bonding, proton transfer, and interactions with hard and soft Lewis acids.[7] This was reduced to the 5 most informative scales. During the last 10 years combinatorial chemistry, along with HTS, has developed as a new route to drug discovery.[19,20] Traditional descriptors for QSAR models were compared with two alternative blocks obtained by computer chemistry tools: electronic densities and principal properties.[21] Eight common descriptors of solvent properties for 82 different solvents were analysed by PCA.[22,23] The liquid viscosity of 361 organic compounds containing C, H, N, O, S and/or halogens was investigated using a QSPR approach.[24,25] A QSRR[26,27] study of the decarboxylation rates of 6-nitrobenzisoxazole-3-carboxylic acid employing the CODESSA program correlated the effect of 24 solvents with theoretical descriptors to provide a straightforward interpretation of these solvent effects in terms of molecular parameters.

In this work we wish to apply the PCA method to study enaminone chemistry. Enaminones[28–31] are important synthetic intermediates, especially with regard to the formation of nitrogen heterocyclic systems. They contain three sites that



K1, R=CH(Ph)$_2$ **K2**, R=CH(CH$_3$)Ph **K3**, R=CH(CH$_3$)$_2$

Fig. 1 α-Acylenaminoketones **K1**–**K3**.

are vulnerable to electrophilic attack (O, C_α, N) and two to nucleophilic attack (C=O, C_β). Utilizing an enaminoketone as starting material, 3-substituted chromones were synthesized by the reactions of acid anhydride derivatives under mild conditions.[32] Our interest in these systems dates back to the late seventies when we discovered that enaminones react with diphenylcyclopropenone as nucleophilic species through their nitrogen to form 5-functionalized 1,5-dihydropyrrol-2-ones.[33] On the other hand, they react with Cu(II) stabilized ketocarbenes at the C_α position to form pyrroles.[34] Diazoketones react with enaminones *via* ketenes under noncatalytic thermal conditions[35,36] to form nucleophilic addition products, such as the α-acylenaminoketones **K1** [3-acetyl-1,1-diphenyl-4-(methylamino)pent-3-en-2-one], **K2** [3-acetyl-1-phenyl-1-methyl-4-(methylamino)pent-3-en-2-one] and **K3** [3-acetyl-1,1-dimethyl-4-(methylamino)pent-3-en-2-one]. The study of the reactivity of α-acylenaminoketones **K1**–**K3** (Fig. 1) was of interest to us because of the differences in the two ketonic carbonyls. With hydrazines, these systems form different pyrazoles depending on the reaction conditions.[36]

The most important derivatives of pyrazole are pyrazolones, which have important pharmacological properties and of which a few naturally-occurring examples exist. For this reason, there is increasing interest in the development of new procedures for the synthesis of pyrazoles and their derivatives. *N*-Substituted pyrazoles are of interest as chiral auxiliaries for stereoselective synthesis and for the resolution of certain racemic compounds.[37]

Treatment of β-diketones and the corresponding β-enaminoketones having modified carane and *p*-menthane skeletons, with aryl and alkylhydrazines resulted in regioselective formation of *N*-substituted pyrazoles or stable pyrazolinols depending on the nature of the substituent at the hydrazine nitrogen.[38] The enaminoimine hydrochlorides were transformed *in situ* to the corresponding pyrazoles in moderate to high yields by the addition of hydrazine.[39]

Palladium-catalyzed cross-coupling of 1-(benzyloxy)pyrazol-5-ylzinc halides prepared by transmetalation of 1-(benzyloxy)-5-lithiopyrazole with acyl chlorides produced 5-acyl-1-(benzyloxy)pyrazoles in high yields.[40] 3,5-Dimethyl-4-(benzotriazol-1-ylmethyl)-1-phenylpyrazole and 1,3,5-trimethyl-4-(benzotriazol-1-ylmethyl)pyrazoles were prepared and alkylated with various alkyl bromides. Treatment of these compounds with alkyl Grignard reagents led to a variety of 4-substituted-3,5-dimethylpyrazoles.[41] The reaction of aryl- or alkylhydrazine hydrochlorides with 4-cyano-3-oxotetrahydrothiophene in refluxing ethanol afforded, in a regioselective manner, the corresponding 2-alkyl- or 2-aryl-3-aminothieno[3,4-*c*]pyrazoles in good yields.[42,43]

Diazomethane adds to enyne sulfones regio- and stereoselectively to give the 4-alkynyl-5-phenylsulfonyl-4,5-dihydro-3*H*-pyrazoles, which are converted by MeLi into the 4-alkynyl-1*H*-pyrazoles in good yields.[44] A pyrazole aldehyde was used as a starting material for an intramolecular hetero Diels–Alder reaction, affording tetracyclic pyrazoles with high yield and diastereoselectivity.[45,46] β-Aminoenones react with monoalkyl hydrazines to give regioselectively 1,3,5-trisubstituted pyrazoles.[47–52]

Although there are a considerable number of studies involving the chemistry of simple enaminones, the same is not true

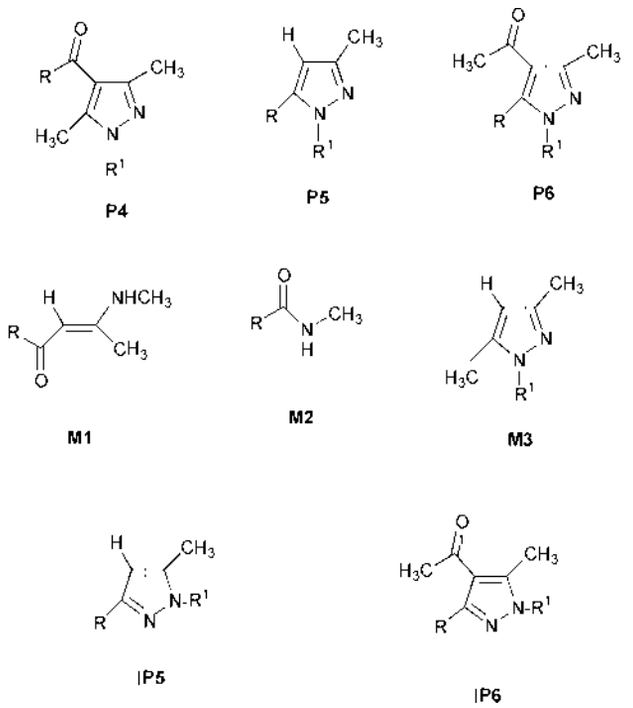


Fig. 2 Reaction products of **K1**–**K3** with hydrazines. R = $CH(Ph)_2$, $CH(CH_3)Ph$, $CH(CH_3)_2$; R^1 = *p*-NO_2Ph, CH_3, H, Ph.

for α-acylenaminoketones. Our continuing interest in the structure–reactivity relationships of enaminones has led us to examine the chemistry of 4-methylaminopent-3-en-2-ones with acetyl derivatives in the 3-position[36,53] introducing another position vulnerable to nucleophilic attack. Benzene, methylene chloride, tetrahydrofuran, methanol and *N,N*-dimethylformamide were utilized as solvents[54] in the reactions of α-acylenaminoketones **K1**–**K3** with hydrazine reagents (methylhydrazine, phenylhydrazine, *p*-nitrophenylhydrazine and hydrazine hydrate). Besides verifying which of the two carbonyls would preferentially be attacked during a nucleophilic attack, we wished to obtain information on the solvent dependence of the regiochemistry of the pyrazole formed. The reaction mixtures were submitted to GC-MS analyses, in an attempt to identify all the products and possible intermediates formed during the reactions.

The formation of the principal pyrazoles **P4** (the product formed by nucleophilic attack on the acetyl group), **P5** (deacetylated pyrazoles, with the substituent R in the 5 position) and **P6** (4-acetylpyrazoles formed by attack on RCO) (Fig. 2) can be explained by an initial Michael-type reaction. The pyrazoles **P5** and the deacetylated enaminones (2-methylaminopent-2-en-4-ones) (**M1**) were formed by a deacetylation process. Small amounts of isomeric pyrazoles (**IP5**, **IP6**) were also formed, perhaps by initial reaction on the carbonyl,[36] together with acetamides (**M2**) and 3,5-dimethylpyrazoles (**M3**) (Fig. 2).

Simple eye inspection of the distribution of the eight products, especially those obtained in low yields did not enable detection of consistent differences between solvents or nucleophiles. Therefore, in an attempt to understand how these factors could be correlated, we utilized PCA. *Ab initio* molecular orbital calculations were also undertaken to understand intrinsic properties of the nucleophiles and substrate.

Results and discussions

The experimental results are the integrated mass chromatographic peaks obtained with the reactions of compounds **K1**–**K3** with methylhydrazine (MH), phenylhydrazine (FH), *p*-nitrophenylhydrazine (NFH) and hydrazine hydrate (HH) utilizing benzene (b), methylene chloride (m), tetrahydrofuran (t), methanol (me) and *N,N*-dimethylformamide (d) as solvents. There were eight products obtained, the principal ones being the pyrazoles **P4**–**P6**. The amounts of each pyrazole were determined by integration of the areas of the corresponding peaks, which was performed using HP-chemstation software and comparison with the areas of isolated pyrazoles **P4**–**P6** with known concentrations.

The data set to be analysed using Principal Component Analysis, was arranged in a matrix (20 × 24). The variables (24) correspond to the % yield of pyrazole **P4**–**P6** and the other five compounds obtained in low yield (**IP5**, **IP6**, **M1**, **M2** and **M3**). The samples in the rows correspond to the reaction conditions, *i.e.* variation of the nucleophile and solvent.

In the study of the reactivity of α-acylenaminoketones the organization of the results show that loadings have information about the yields of products formed while the scores show that the products are affected by reaction conditions (solvent and nucleophile). Three principal components describe 87% of the total variance in the original data set. They can identify the separation of pyrazoles according to the mechanism type as can be seen in the loadings plot (Fig. 3A). Thus, the pyrazole **P5**, which was obtained by a mechanism involving a deacetylation process, was separated from pyrazoles **P4** and **P6** in the first principal component. The second component, on the other hand, has information about the structural characteristics of the α-acylenaminoketones. In this component **P4** is separated from **P6** which corresponds to pyrazoles formed by attack at different carbonyls. The pyrazoles **P4** were formed by nucleophilic attack on the acetyl group and the pyrazoles **P6** were formed by nucleophilic attack on the carbonyl neighboring the larger group (R). The results obtained using HCA confirm that each one of the pyrazoles was grouped together in the same cluster according to the compounds **K1**–**K3** used as reagents. It also shows that the pyrazoles **P5** are separated from pyrazoles **P6** and **P4** (Fig. 4A) and that **P5** is in the same cluster as products **M1**–**M3** suggesting correlation among them. This would be consistent with a mechanism in which **M1** is formed initially and then serves as the starting material for **P5**.

The separation observed in the scores graph (Fig. 3B) represents the information which allows one to evaluate the influence of solvent and nucleophile on the formation of each one of the pyrazoles. Although the separations are not as straightforward as in Fig. 3A, tendencies can be seen.

The first group **I** is exclusively composed of reactions involving formation of deacetylated pyrazoles **P5a**–**c** (Table 1). The reactions of compounds **K1**–**K3** using *p*-nitrophenylhydrazine were more sluggish, leading to the formation of **P5a**–**c** (Fig. 3B). The nucleophilic effect of *p*-nitrophenylhydrazine (NFH) is evident, showing that the decrease in the nucleophilic power of the substituted nitrogen of hydrazine favored the deacetylation process. These results show that the solvent did not affect this process, because in all solvents, the utilization of this nucleophile favored the deacetylation product. The same clustering was observed in the dendogram (HCA, Fig. 4B) which is divided into three main groups at 0.6 similarity index related to the pyrazoles.

In the second group **II** are the reactions of compounds **K1**–**K3** with methylhydrazine (MH) when methylene chloride (m), benzene (b), tetrahydrofuran (t) and *N,N*-dimethylformamide (d) were used as solvents (Table 2). Under such conditions, mixtures of **P4d**–**f** and **P6d**–**f** were obtained (Fig. 3). This group could be divided into two clusters. Thus, in the reactions of compounds **K1**–**K3** with methylhydrazine, the formation of pyrazoles **P4d**–**f** were favored in benzene, methylene chloride and tetrahydrofuran. When *N,N*-dimethylformamide, a polar aprotic solvent, was used, the pyrazoles **P6d**–**f** were preferentially formed. Very interestingly, the pyrazole **P4d** was obtained as the principal product in all solvents used in the reactions of compound **K1** with methylhydrazine (Table 2). The

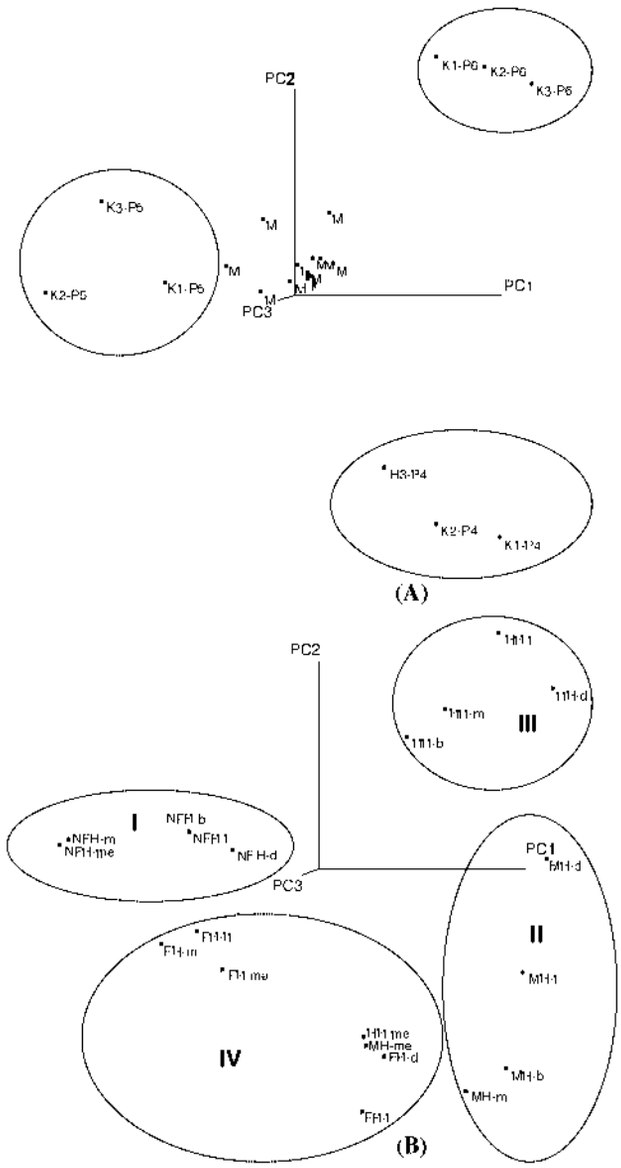


Fig. 3 Loadings plot showing how product reactions formed clusters in the first three PCs (**A**); **K1**–**K3**-α-Acylenaminoketones index (reagents), **P4**–**6**-Pyrazoles index. Scores plot showing the reaction conditions for the first three PCs (**B**). Nucleophile index: **NFH** = *p*-nitrophenylhydrazine; **FH** = phenylhydrazine; **MH** = methylhydrazine; **HH** = hydrazine monohydrate. Solvent index: **b** = benzene; **m** = methylene chloride; **me** = methanol; **t** = tetrahydrofuran; **d** = dimethylformamide. Classification of groups in relation to the type of nucleophile: Group I: reaction with *p*-nitrophenylhydrazine; Group II: reaction with methylhydrazine; Group III: reaction with hydrazine monohydrate; Group IV: reaction with phenylhydrazine in all solvents and methylhydrazine and hydrazine hydrate in methanol. **A** and **B** are related. The positions of the reaction conditions in **B** are correlated with the products in similar positions in **A**.

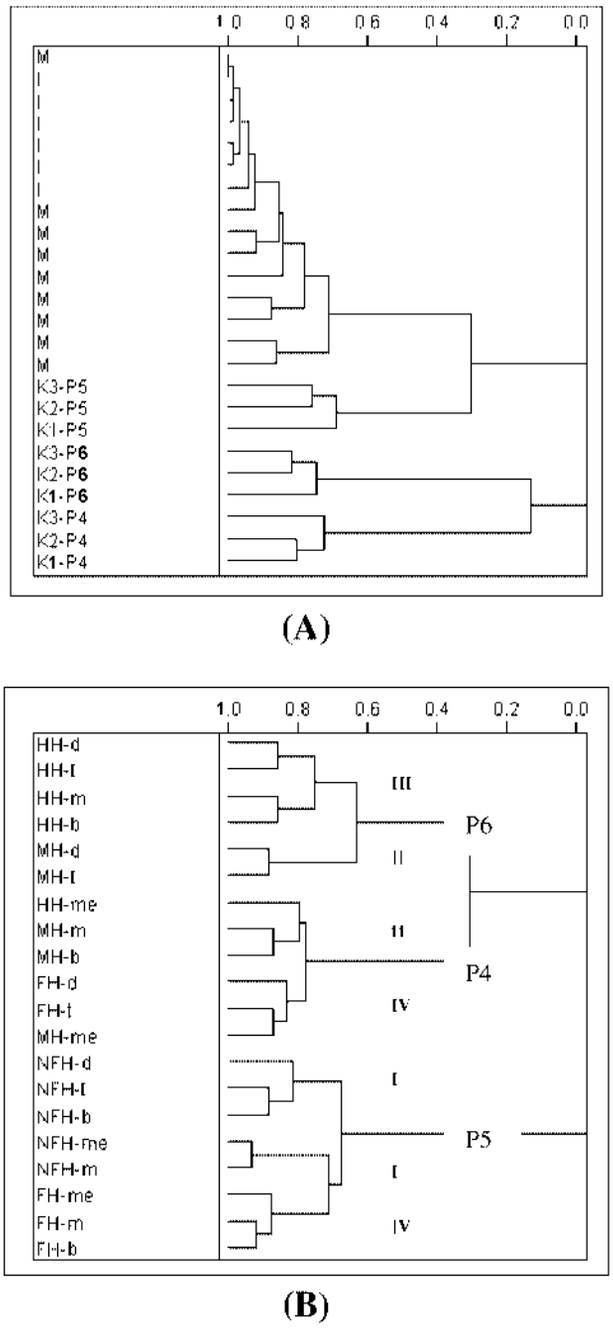


Fig. 4 Dendogram in the direction of columns connecting the variables (**A**) and in the direction of rows the standard of samples (**B**).

reactions of **K1**–**K2** with methylhydrazine using methanol, a polar protic solvent, are in group IV. In the dendogram (Fig. 4B) one observes that reactions involving methylhydrazine as nucleophile in benzene and methylene chloride form mainly **P6** while in tetrahydrofuran and *N,N*-dimethylformamide **P4** is formed. It can be seen that the reactions of compounds **K1**–**K3** with hydrazine hydrate in methanol favor the formation of pyrazole **P4g**–**i**. In PCA, the reactions of compounds **K1**–**K3** with hydrazine hydrate in methanol are in group IV.

Group **III** includes the reactions of α-acylenaminoketones **K1**–**K3** with hydrazine hydrate, utilizing benzene, methylene chloride, tetrahydrofuran and *N,N*-dimethylformamide as solvents, with predominance of the pyrazoles **P6g**–**i** (Table 3, Fig. 3). In the dendogram (Fig. 4B, group II) these reactions are also grouped together. The nucleophilic attack on the carbonyl bonded to the R group occurred preferentially due to the high nucleophilic power of the two nitrogens of the hydrazine hydrate, thus suggesting that the formation of pyrazoles **P6g**–**i** was favored by a fast nucleophilic attack. When methanol was used as solvent, the reactions of α-acylenaminoketones **K1**–**K3** with hydrazine hydrate formed **P4g**–**i**. This reaction condition appeared in cluster **IV** (Fig. 3B). Using hydrazine hydrate as nucleophile, the formation of deacetylated products occurred in small quantities when tetrahydrofuran was used as solvent.

The reactions in which a mixture of pyrazoles **P4j**–**m** and **P5j**–**m** were formed are in group **IV** (Fig. 3B). Therefore, in group **IV** we verified that there was a mixture in relation to the types of nucleophiles. It is described by reactions involving the nucleophiles phenylhydrazine (Table 4), methylhydrazine and hydrazine hydrate using methanol as solvent. This group can be divided into two clusters (Fig. 3B). In the first cluster are the

Table 1 Yields (%) of compounds **P4**–**P6**, **IP5**, **M2** and **M3** using *p*-nitrophenylhydrazine (**a**: R = $CH(Ph)_2$, R^1 = *p*-NO_2Ph; **b**: R = $CH(CH_3)Ph$, R^1 = *p*-NO_2Ph; **c**: R = $CH(CH_3)_2$, R^1 = *p*-NO_2Ph)

Entry	Substrate	Solvent	Main Products[a]	Side Products[a]
1	**K1**	Benzene	**M2a** (46)	**M3a** (31)
2	**K2**	Benzene	**P5b** (42)	**M3b** (11), **M2b** (12)
3	**K3**	Benzene	**P5c** (67)	**IP5c** (11), **M3c** (6)
4	**K1**	CH_2Cl_2	**P5a** (49)	**M3a** (18), **M2a** (23)
5	**K2**	CH_2Cl_2	**P5b** (71)	**M3b** (13), **M2b** (7)
6	**K3**	CH_2Cl_2	**P5c** (64)	—
7	**K1**	THF	**M2a** (34)	**M3a** (16)
8	**K2**	THF	**P5b** (47)	**M3b** (4), **M2b** (7)
9	**K3**	THF	**P5c** (41)	—
10	**K1**	MeOH	**P5a** (43)	**M3a** (26), **M2a** (11)
11	**K2**	MeOH	**P5b** (79)	**M3b** (8)
12	**K3**	MeOH	**P5c** (65)	—
13	**K1**	DMF	**M2a** (22), **M3a** (22)	—
14	**K2**	DMF	**P5b** (53)	**M3b** (14), **M2b** (18)
15	**K3**	DMF	**P4c** (21), **P5c** (20), **P6c** (20)	—

[a] The amounts of each pyrazole were determined by integration of the areas of the corresponding peaks, which was performed using HP-Chemstation Software and comparison with the areas of isolated pyrazoles with known concentrations.

Table 2 Yields (%) of compounds **P4**–**P6**, **IP5**, **IP6**, **M1** and **M2** using methylhydrazine (**d**: R = $CH(Ph)_2$, R^1 = CH_3; **e**: R = $CH(CH_3)Ph$, R^1 = CH_3; **f**: R = $CH(CH_3)_2$, R^1 = CH_3)

Entry	Substrate	Solvent	Main Products	Side Products
16	**K1**	Benzene	**P4d** (78)	**P6d** (14), **M2d** (5)
17	**K2**	Benzene	**P4e** (60)	**P6e** (7), **M2e** (25)
18	**K3**	Benzene	**P4f** (37), **P6f** (37)	—
19	**K1**	CH_2Cl_2	**P4d** (70)	**P6d** (8), **M2d** (11)
20	**K2**	CH_2Cl_2	**P4e** (58)	**P5e** (16), **P6e** (17)
21	**K3**	CH_2Cl_2	**P4f** (58)	**P6f** (25)
22	**K1**	THF	**P4d** (73)	**P6d** (14)
23	**K2**	THF	**P4e** (43), **P6e** (43)	**M2e** (7)
24	**K3**	THF	**P4f** (49), **P6f** (39)	—
25	**K1**	MeOH	**P4d** (42)	**P5d** (9), **M1d** (9), **IP5d** (4)
26	**K2**	MeOH	**P4e** (52)	**P5e** (23), **IP5e** (7), **P6e** (5)
27	**K3**	MeOH	**P4f** (35)	**M1f** (20), **P5f** (7), **P6f** (9), **IP6f** (2)
28	**K1**	DMF	**P4d** (62)	**P6d** (19), **M2d** (8)
29	**K2**	DMF	**P6e** (55)	**P4e** (32), **M2e** (6)
30	**K3**	DMF	**P6f** (73)	**P4f** (24)

Table 3 Yields (%) of compounds **P4**–**P6**, **M1** and **M2** using hydrazine hydrate (**g**: R = $CH(Ph)_2$, R^1 = H; **h**: R = $CH(CH_3)Ph$, R^1 = H; **i**: R = $CH(CH_3)_2$, R^1 = H)

Entry	Substrate	Solvent	Main Products	Side Products
31	**K1**	Benzene	**P6g** (35), **P4g** (25)	**M1g** (11), **P5g** (8), **M2g** (13)
32	**K2**	Benzene	**P6h** (46), **M1h** (27)	**P5h** (10), **P4h** (8)
33	**K3**	Benzene	**P6i** (38), **M1i** (30)	**P5i** (24), **P4i** (7)
34	**K1**	CH_2Cl_2	**P6g** (45), **P4g** (21)	**M1g** (9), **P5g** (13), **M2g** (8)
35	**K2**	CH_2Cl_2	**P6h** (33), **M1h** (27)	**P5h** (3), **P4h** (3), **M2h** (13)
36	**K3**	CH_2Cl_2	**P6i** (64)	**P4i** (13), **P5i** (13)
37	**K1**	THF	**P6g** (73)	**P4g** (14), **M2g** (7)
38	**K2**	THF	**P6h** (59)	**M1h** (13), **P4h** (8), **M2h** (14)
39	**K3**	THF	**P6i** (67)	**P4i** (16), **P5i** (17)
40	**K1**	MeOH	**P4g** (68)	**M1g** (6), **P5g** (10), **M2g** (8)
41	**K2**	MeOH	**P4h** (49)	**P5h** (37), **P6h** (8)
42	**K3**	MeOH	**P6i** (36), **P5i** (34), **P4i** (30)	—
43	**K1**	DMF	**P6g** (50)	**P4g** (33), **M2g** (17)
44	**K2**	DMF	**P6h** (75)	**P4h** (16), **M1h** (9)
45	**K3**	DMF	**P6i** (81)	**P4i** (19)

reactions of compounds **K1**–**K3** with phenylhydrazine using benzene, methylene chloride and methanol as solvents. In all cases the yields of pyrazoles **P5j**–**m** were greater. In the second cluster are the reactions using phenylhydrazine with tetrahydrofuran and *N,N*-dimethylformamide as solvents and the reactions of compounds **K1**–**K3** with methylhydrazine and hydrazine monohydrate using methanol as solvent. The use of phenylhydrazine as a nucleophile in the reactions of **K1**–**K3**, and the use of benzene, methylene chloride and methanol as solvents led to the formation of deacetylated pyrazoles **P5j**–**m**, while tetrahydrofuran and *N,N*-dimethylformamide proceeded to form the products corresponding to the attack of the secondary amino group of the hydrazine reagent on the carbonyl carbon of the acetyl group favoring the formation of **P4j**–**m**. Only a very small amount of pyrazole **P6m** was obtained.

One observes in the dendogram (Fig. 4B) that group **IV** was divided into two clusters. In one cluster are the reactions of phenylhydrazine using tetrahydrofuran and *N,N*-dimethylformamide as solvent and methylhydrazine using methanol as solvent. In the other cluster are the reactions of phenylhydrazine

Table 4 Yields (%) of compounds **P4**–**P6**, **M1** and **M2** using phenylhydrazine (**j**: R = $CH(Ph)_2$, R^1 = Ph; **l**: R = $CH(CH_3)Ph$, R^1 = Ph; **m**: R = $CH(CH_3)_2$, R^1 = Ph)

Entry	Substrate	Solvent	Main Products	Side Products
46	**K1**	Benzene	**P5j** (39), **P4j** (21)	**M3j** (10), **M2j** (11)
47	**K2**	Benzene	**P5l** (51)	**M3l** (8)
48	**K3**	Benzene	**P4m** (30), **P5m** (24)	**P6m** (4), **M3m** (4), **IP5m** (4)
49	**K1**	CH_2Cl_2	**P5j** (39), **P4j** (20)	**P6j** (18), **M2j** (14)
50	**K2**	CH_2Cl_2	**P5l** (61)	**P6l** (19), **P4l** (8), **M2l** (6)
51	**K3**	CH_2Cl_2	**P5m** (38), **P4m** (33)	**M3m** (5)
52	**K1**	THF	**P4j** (56)	**P5j** (9), **M3j** (14), **M2j** (7)
53	**K2**	THF	**P4l** (57)	**P5l** (25), **M3l** (14)
54	**K3**	THF	**P4m** (54)	**P5m** (15)
55	**K1**	MeOH	**P4j** (27), **P5j** (26)	**M3j** (9), **M2j** (8)
56	**K2**	MeOH	**P5l** (39), **P4l** (21)	**M3l** (13), **M2l** (10)
57	**K3**	MeOH	**P5m** (44), **P4m** (30)	**M3m** (14)
58	**K1**	DMF	**P4j** (61)	**M3j** (13), **M2j** (8), **P5j** (6)
59	**K2**	DMF	**P4l** (28), **M2l** (21)	**M3l** (14), **M1l** (6), **P5l** (4)
60	**K3**	DMF	**P4m** (38)	**M3m** (10), **M1m** (6), **P5m** (6)

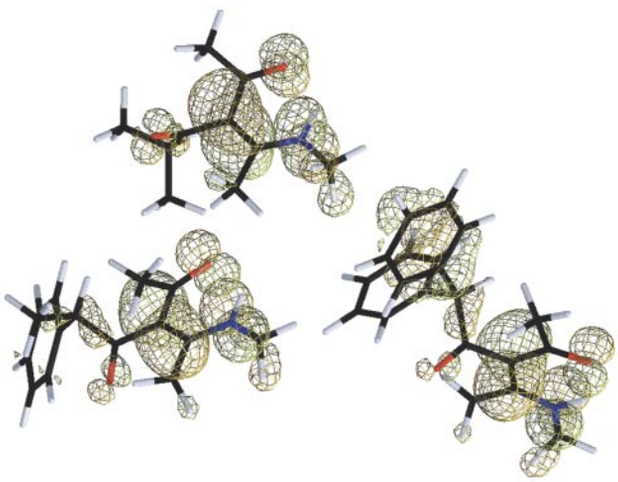


Fig. 5 HOMOs of **K1**–**K3**.

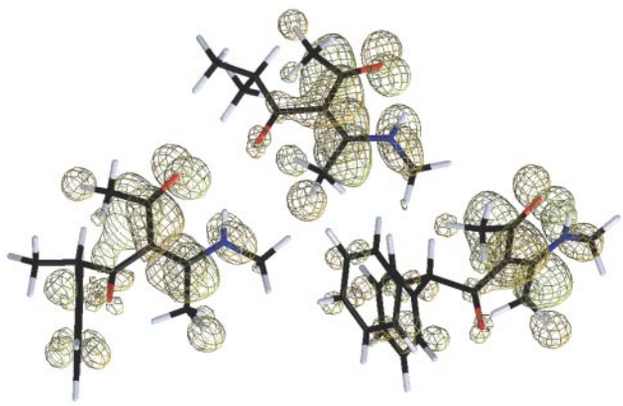


Fig. 6 LUMOs of **K1**–**K3**.

using methanol, benzene and methylene chloride as solvents. The reactions of **K1**–**K3** with methylhydrazine and hydrazine hydrate in methanol favor the formation of pyrazoles **P4**.

The separation of these reactions into four groups occurred depending on the nucleophile used in the reaction. This suggests that the utilization of methanol, the only protic solvent used, favored nucleophilic attack on the acetyl carbonyl group to form **P4**. Products obtained *via* a deacetylation process were favored by a decrease in the nucleophilicity of the secondary amino group of hydrazine. The same four groups were formed using HCA, as can be observed in the dendogram (Fig. 4).

Ab initio 6-31G** molecular orbital calculations which do not take solvent effects into account show that compounds **K1**–**K3** have HOMOs with large coefficients on both the α-carbon and nitrogen atoms (Fig. 5). The absolute magnitude of the two coefficients is somewhat greater at the α-carbon atom. **K1**–**K3** have LUMOs with large coefficients on both the β-carbon atom and the acetyl group (Fig. 6). The absolute magnitude of the two coefficients is somewhat greater at the β-carbon atom. This is consistent with initial attack of the nucleophile on the β-carbon for formation of the principal products.

Both *ab initio* 6-31G** and AM1 [36,53] geometry optimizations indicate a tendency for one intramolecular hydrogen bond with the carbonyl carbon of the acetyl group favoring *E* configuration and the conjugation of the acetyl group with the double bond. The heat of formation of the compound **K1** is −12.434 kcal mol^{-1}, for compound **K2** is −47.956 kcal mol^{-1} and for compound **K3** is −81.892 kcal mol^{-1}.

Molecular orbital calculations of the hydrazine nucleophiles show that the two nitrogens of hydrazine monohydrate have HOMOs with large coefficients (−0.783, −0.836). In the hydrazine dihydrate, the nucleophilic power of the two nitrogen atoms is higher (−0.883, −0.880). These results can explain the simultaneous attack of the two nucleophilic sites to form, preferentially, the pyrazoles **P6**.

The unsubstituted nitrogen of methylhydrazine is more nucleophilic than that of phenylhydrazine and *p*-nitrophenylhydrazine, as can be seen by comparing the coefficients of the HOMOs (−0.784 *versus* −0.708 and −0.709, respectively). The coefficient of the substituted nitrogen of methylhydrazine is similar to that of phenylhydrazine (−0.541 *versus* −0.550) while that of *p*-nitrophenylhydrazine is lower (−0.520) as would be expected because of the greater electron withdrawing power of the nitro group. These differences may explain the differences in behavior of these nucleophiles with respect to the deacetylation process. The results suggest that the coefficient or charge on the unsubstituted nitrogen of hydrazine monohydrate and methylhydrazine have the same value, but the *N*-substituted nitrogen of methylhydrazine has approximately the same charge as that of phenylhydrazine and *p*-nitrophenylhydrazine. Thus charge effect does not explain the differences in behavior.

A factor that may be influencing the behavior of nucleophiles and compounds **K1**–**K3** is the competition between the two carbonyls susceptible to nucleophilic attack, one being conjugated. For all α-acylenaminoketones studied in this work, the large coefficients of the LUMO are on the β-carbon and carbonyl carbon of the acetyl group. However, reaction sometimes occurred on the non-conjugated carbonyl, perhaps by a Michael reaction followed by a fast nucleophilic attack on COR before isomerization of the intermediate formed could occur.

Experimental

The general procedure for reactions of compounds **K1**–**K3** with hydrazine reagents is cited in ref. 36. The analyses, using GC-MS, were recorded with a Hewlett Packard model 5988A (installed at UNICAMP). The conditions for gas chromatography were: temperature range 35–250 °C at 8 °C min^{-1} with a 10 min hold at 250 °C, injector temperature 250 °C, detector temperature 250 °C.

Acknowledgements

We thank CNPq and FAPESP for financial support, R. E. Bruns for his helpful discussions, and A. T. Bruni for the *ab initio* calculations.

References

1 I. S. Scarminio and R. E. Bruns, *Anal. Chem.*, 1989, **8**, 326.
2 M. A. Sharaf, D. L. Illman and B. R. Kowalski, in *Chemometrics*; Ed. P. J. Elving, J. D. Winefordner and I. M. Lolthoff, Wiley-Interscience, New York, 1986.
3 S. Wold, K. Esbensen and P. Geladi, *Chemom. Intell. Lab. Syst.*, 1987, **2**, 37.
4 K. R. Beebe, R. J. Pell and M. B. Seasholtz, in *Chemometrics: A Practical Guide*, John Wiley & Sons, New York, 1998.
5 P.-C. Maria, J.-F. Gal, J. de Franceschi and E. Fargin, *J. Am. Chem. Soc.*, 1987, **109(2)**, 483.
6 W. P. Ashman, H. J. Lewis and E. J. Poziomek, *Anal. Chem.*, 1985, **57(9)**, 1951.
7 K. Varmuza, P. Penchev, F. Stand and W. Werther, *J. Mol. Struct.*, 1997, **408(1)**, 91.
8 S. Wold and M. Sjöström, *Acta Chem. Scand.*, 1998, **52(5)**, 517.
9 R. Carlson, J. Carlson and A. Grennberg, *J. Chemometrics*, 2001, **15**, 455.
10 P. M. Andersson, M. Sjöström, S. Wold and T. Lundstedt, *J. Chemometrics*, 2001, **15**, 353.
11 M. Kansiz, P. Heraud, B. Wood, F. Burden, J. Beardall and D. McNaughton, *Phytochemistry*, 1999, **52**, 407.
12 R. Carlson, L. Hansson and T. Lundstedt, *Acta Chem. Scand. Ser. B*, 1986, **40**, 444.
13 R. Carlson, T. Lundstedt and R. Shabana, *Acta Chem. Scand. Ser. B*, 1986, **40**, 534.
14 S. Wold and M. Sjöström, *Acta Chem. Scand.*, 1998, **52**, 517.
15 A. R. Katritzky, K. Chen, U. Maran and D. A. Carlson, *Anal. Chem.*, 2000, **72(1)**, 101.
16 C. G. Fraga, B. J. Prazen and R. E. Synovec, *Anal. Chem.*, 2000, **72(17)**, 4154.
17 C. Ebert, P. Linda, S. Alunni, S. Clementi, G. Cruciani and S. Santini, *Gazz. Chim. Ital.*, 1990, **120**, 29.
18 O. Pytela, *Collect. Czech. Chem. Commun.*, 1996, **61**, 705.
19 A. Linusson, J. Gottfries, F. Lindgren and S. Wold, *J. Med. Chem.*, 2000, **43**, 1320.
20 S. Hellberg, M. Sjöström, B. Skagerberg and S. Wold, *J. Med. Chem.*, 1987, **30(7)**, 1126.
21 D. Bonelli, G. Coata, G. Cruciani and S. Clementi, *Farmaco*, 1990, **45(3)**, 293.
22 R. Carlson, T. Lundstedt and C. Albano, *Acta Chem. Scand. Ser. B*, 1985, **39(2)**, 79.
23 R. Carlson and T. Lundstedt, *Acta Chem. Scand. Ser. B*, 1987, **41(3)**, 164.
24 A. R. Katritzky, T. Tamm, Y. L. Wang, S. Sild and M. Karelson, *J. Chem. Inf. Comput. Sci.*, 1999, **39(4)**, 684.
25 A. R. Katritzky, K. Chen, Y. Wang, M. Karelson, B. Lucic, N. Trinajstic, T. Suzuki and G. Schueuuermann, *J. Phys. Org. Chem.*, 2000, **13(1)**, 80–86.
26 A. R. Katritzky, S. Perumal and R. Petrukhin, *J. Org. Chem.*, 2001, **66(11)**, 4036.
27 F. Ignatz-Hoover, R. Petrukhin, M. Karelson and A. R. Katritzky, *J. Chem. Inf. Comput. Sci.*, 2001, **41(2)**, 295.
28 M. G. Gorbunova, *J. Chem. Soc., Perkin Trans. 2*, 1993, **3**, 559.
29 S. Seko and N. Tani, *Tetrahedron Lett.*, 1998, **39(4)**, 8117.
30 R. Bartnik, A. Bensadat, D. Cal, R. Faure and N. Khatimi, *Bull. Soc. Chim. Fr.*, 1997, **134(7)**, 725.
31 A. R. Katritzky, Y. F. Fang, A. Donkor and J. Y. Xu, *Synthesis*, 2000, **14**, 2029.
32 I. Yokoe, K. Maruyama, Y. Sugito, T. Harashida and Y. Shirataki, *Chem. Pharm. Bull*, 1994, **42(8)**, 1697.
33 C. Kascheres, A. J. Kascheres and P. S. H. Pilli, *J. Org. Chem.*, 1980, **45**, 5340.
34 M. N. Eberlin and C. Kascheres, *J. Org. Chem.*, 1988, **53(9)**, 2084.
35 M. N. Eberlin, Y. Takahata and C. Kascheres, *J. Org. Chem.*, 1990, **55**, 5150.
36 G. Negri and C. Kascheres, *J. Heterocyclic Chem.*, 2001, **38**, 109.
37 K. Makino, H. S. Kim and Y. Kurasawa, *J. Heterocycl. Chem.*, 1998, **35**, 489.
38 S. A. Yamashkin and I. A. Batanov, *Chem. Heterocycl. Compd. (Engl. Transl.)*, 1995, **31(1)**, 50.
39 A. R. Katritzky, A. Denisenko, S. N. Denisenko and M. Arend, *J. Heterocycl. Chem.*, 2000, **37**, 1309.
40 M. Franck-Neumann, P. Geoffroy and A. Winling, *Tetrahedron Lett.*, 1998, **39**, 5643.
41 J. Kristensen, M. Begtrup and P. Vedsoe, *Synthesis*, 1998, **11**, 1604.
42 J. Soloducho and R. P. Musgrave, *Pol. J. Chem.*, 1997, **71(8)**, 1075.
43 P. G. Baraldi, H. El-Kashef, S. Manfredini, M. J. Pineda de Las Infantas, R. Romagnoli and S. Giampiero, *Synthesis*, 1998, **9**, 1331.
44 O. A. Attanasi, P. Filippone, C. Fiorucci, E. Foresti and F. Mantellini, *J. Org. Chem.*, 1998, **63(26)**, 9880.
45 M. Yoshimatsu, M. Kawahigashi, E. Honda and T. Kataoka, *J. Chem. Soc., Perkin Trans. 1*, 1997, **5**, 695.
46 E. Ceulemans, M. Voets, S. Emmers and W. Dehaen, *Syn. Lett.*, 1997, **10**, 1155.
47 K. K. Kieckononowicz, X. Ligneau, J.-C. Schwartz and W. Schunack, *Arch. Pharm. (Weinheim Ger.)*, 1995, **328(5)**, 469.
48 F. Haunert, M. H. Bolli, B. Hinzen and S. V. Ley, *J. Chem. Soc., Perkin Trans. 1*, 1998, **15**, 2235.
49 A. Alberola, A. Gonzalez-Ortega, M. L. Sadaba and M. C. Sanudo, *J. Chem. Soc., Perkin Trans. 1*, 1998, **24**, 4061.
50 N. Oikawa, C. Mueller, M. W. Kunz and F. W. Lichtenthaler, *Carbohydr. Res.*, 1998, **309(3)**, 269.
51 K. Makino, H. S. Kim and Y. Kurasawa, *J. Heterocycl. Chem.*, 1998, **35(3)**, 489.
52 T. M. Krygowski, R. Anulewicz, M. K. Cyranski, A. Puchala and D. Rasala, *Tetrahedron*, 1998, **54(40)**, 12295.
53 C. Kascheres, G. Negri, M. T. P. Gambardella and R. H. A. Santos, *J. Braz. Chem. Soc.*, 1994, **5(1)**, 31.
54 C. Reichardt, in *Solvents and Solvent Effects in Organic Chemistry*, Second Edition, Ed. H. F. Ebel, VCH, Weinheim, 1988.