A theoretical study of the Duff reaction: insights into its selectivity†

Nicolás Grimblat,* Ariel M. Sarotti, Teodor S. Kaufman and Sebastian O. Simonetti*

The Duff reaction is one of the most employed methods for the ortho-formylation of phenols; however, not much is truly known about its mechanism. Using DFT calculations, we disclose the first theoretical study regarding the selectivity-determining step of the reaction. We have found that this stage is governed by a hydrogen bond, that gives rise to a cyclohexa-2,4-diene intermediate and establishes the position where the formylation will take place. These findings were evaluated by analysis of the reaction outcome of several non-symmetrically substituted phenols.

Introduction

The synthesis of salicylaldehyde derivatives often entails the installation of a formyl moiety ortho to a phenol. This can be accomplished through different classical approaches, such as the Duff, Reimer–Tiemann and Vilsmeier–Haack reactions, among others.4

The Duff formylation involves the treatment of a phenol with HMTA in acidic media followed by hydrolysis in order to afford the desired salicylaldehyde (Scheme 1). Compared to the other approaches, the Duff formylation is advantageous in many synthetic scenarios, because it combines the use of inexpensive and easily available reagents and tolerance to traces of moisture, with its great operational simplicity and compatibility with a wide variety of functional groups.

However, in order to be still more synthetically useful, it faces the same challenges that the other approaches. These defiances include the need of increasing product yields and reaction robustness, as well as to have a handy means to predict its regioselectivity when applied to non-symmetrically substituted phenolic substrates.7

Despite its importance, in nearly one century after its first description7 no major studies nor improvements have been made in the setup of the Duff reaction, except for the employment of trifluoroacetic acid (TFA) as solvent, to enable or enhance the formylation of the less activated aromatic rings9 and the novel use of Cu2O to upgrade the yields of the reaction.10 In addition, an efficient methodology for the synthesis of lactams based on the Duff formylation was recently developed by Moreau et al.11

As a result, the most demanding applications of the Duff formylation, which involve the elaboration of advanced intermediates in multistep synthesis, are currently limited to only specific examples. Among them, in the recent syntheses of (−)-kendomycin, parvistemin A and the structure originally assigned to aspergillitine, the Duff formylation played a key role in order to install the required formyl moieties.12

Most probably, the reaction complexity and the lack of a better understanding of its mechanism have discouraged its wider application toward complex products. In consequence, its use was mainly applied to the preparation of differently substituted salicylaldehydes, which usually served as starting materials.13

The Duff formylation is a Mannich-type reaction, where a phenol acts as the carbonyl component, which is initially subjected to an ortho-aminomethylation with hexamethyleneenetetramine (HMTA), which in turn serves as the needed amino component and a latent source of the formyl moiety. The reaction, which requires acid catalysis, is usually carried out in hot acetic acid or TFA and involves a complex series of steps. Finally, a mineral acid or a water-mediated treatment are employed to hydrolyze the late stage intermediates, leading to the final product.

From a mechanistic point of view, the Duff formylation was postulated as a C–C bond formation process, between a

Scheme 1  The Duff reaction.

†Electronic supplementary information (ESI) available: Computational methods with full list of authors in the Gaussian 09 reference; optimized geometries and reaction coordinates for phenols not included in the paper; Cartesian coordinates, energies, and number of imaginary frequencies of all stationary points. See DOI: 10.1039/c6ob01887d
The reaction under study. Currently accepted structure of the initial Duff coupling product and outline of the currently accepted mechanism for the Duff reaction.

Scheme 2

The reaction under study. Currently accepted structure of the initial Duff coupling product and outline of the currently accepted mechanism for the Duff reaction.

phenol (i) and the iminium species HMTA+ (ii), resulting from protonation and ring opening of HMTA, which acts as the Mannich acceptor. This gives rise to a β-aminoketone-type Mannich base (iii) as the main intermediate, which carries a protonated oxygen atom (Scheme 2). This intermediate, which defines the reaction selectivity, undergoes further oxidation, followed by rearomatization to afford the intermediate iv, and a final hydrolysis to furnish the desired salicylaldehyde product (v).

Although the same picture of the first step of this reaction is widely accepted in order to obtain the intermediate iii, several postulates may be found through the literature, explaining how the iminium intermediate iv could be formed. Mainly, these deal with how the different structural changes take place in the more advanced intermediates. These modifications most commonly include intra-molecular system rearrangements, with C–N bond formation and breaking, under or without the assistance of the solvent. In addition, a rearomatization sequence, followed by dehydrogenation, has also been proposed.

Interestingly, if the starting phenol is non-symmetrically substituted, two potential regioisomers could be formed (R2 and R6). A beforehand knowledge of the outcome of the transformation could be beneficial during the planning stage of multistep syntheses in order to make the most accurate strategic decisions, which would avoid arriving at the wrong regioisomer and, indirectly, favor an increase in the frequency of use of this reaction.

Considering the desirability and potential usefulness of this information, and with the aim of gaining a deeper insight into the origins of the regioselectivity of the transformation, we have performed a theoretical study of the key step of the Duff formylation, in which the selectivity is defined, and herein we disclose our results.

Results and discussion

In order to carry out the projected study, we employed the M06-2X functional coupled with the 6-311+G(d,p) basis set based on recent work by Vilarrasa and coworkers, including acetic acid as solvent using SMD as implemented in Gaussian 09 at the corresponding temperature for each studied reaction (70–140 °C).

The theoretical calculations were initially performed with 3-methylphenol (1, meta-cresol), as a simple model of a meta-substituted, non-symmetrically functionalized phenol. This substrate carries two well-differentiated free ortho sites (R2 and R6), and an available para position (R3), designated according to substituent priority rules. The Duff formylation of 1 was reported to generate only one isomer (R6).

Since there are some scattered reports informing that with suitable substrates the Duff reaction is also able to yield the para isomer, the selection of compound 1 also made us possible to study all three possible substitution points. The availability of reported data on the reaction regioselectivity for this molecule, which also contains three potentially reactive positions, seemed to make of it an adequate substrate in order to study the reaction.

To this end, we examined the reaction mechanism, shown in Scheme 2, where the key initial C–C bond is formed. In this currently accepted proposal, the charge is transferred from the protonated HMTA (HMTA+) to the phenolic oxygen.

Unexpectedly, we realized that the distribution of products that arose from a Boltzmann analysis of the activation energies associated to each possible substitution position was 13 : 87 : 0 (R3/R2/R6). This selectivity is completely different from the product ratio presented by Zhang and coworkers of 0 : 0 : 100 (R3/R2/R6). Furthermore, when the same analysis was performed considering the energies of the reaction products, a 3 : 50 : 47 (R3/R2/R6) selectivity was obtained, also contradicting the experimental results.

The poor agreement between our initial theoretical calculations and the corresponding experimental results with regards to the reaction selectivity prompted us to perform a closer examination of the model behind the selectivity-determining step. After further investigations, we found that during the transition state (TS) of this step (for both ortho- and the para-regioisomers), one of the nitrogen atoms of the tetraamino moiety of the formylating agent captures the phenolic proton, causing dearomatization of the substrate and giving rise to a reactive conjugated cyclohexadienone intermediate.
However, during this process, no changes in the net charge of the HMTA⁺ portion were observed. This is because this hydrogen migration does not entail a second protonation of a nitrogen atom of the intermediate, but rather a system rearrangement, concomitant with the desired C–C bond formation, so the reaction may take place. The generation of a keto-enamine species has been previously observed in similar cases and confirmed through X-ray diffraction, being attributed to the relatively higher basicity of the nitrogen against that of the phenolic oxygen.²¹

The comparison of the traditionally admitted and the newly proposed TSs, revealed that in the accepted approximation the C–C bond being formed has smaller distances (1.81–1.87 Å) and smaller Wiberg bond indexes (WBI, 0.56–0.59) than in this newly discovered attack mode (Fig. 1). For the latter approximation, bond distances up to 2.24 Å and WBI > 0.34 were found.

Interestingly, for TS-1⁺HMTA⁻-R₆H, these parameters are roughly similar to those computed for the TS without the hydrogen migration, although they differ in approximately 7 kcal mol⁻¹. This energy difference may be an evidence of the important role played by the O⋯H–N interaction.

For the R₂ approximation, this newly found geometry of the TS (TS-1⁺HMTA⁻-R₂H) was almost 10 kcal mol⁻¹ lower in energy than that computed for the accepted one (TS-1⁺HMTA⁻-R₃A). Interestingly, for the para (R₄) attack, no energy difference was found between the newly discovered (TS-1⁺HMTA⁻-R₄H) and the accepted TS (TS-1⁺HMTA⁻-R₄A).

On the other hand, the comparatively higher energy required for this approximation (up to 8 kcal mol⁻¹) was in perfect agreement with the recognized preference of the Duff reaction to afford ortho formylation products. Unfortunately, the ratio distribution resulting from this newly found TSs (97:0:3, R₂/R₄/R₆) was also in collision with the experimental results reported in the literature for phenol 1.¹⁹

When we analyzed and compared the energies of the products, corresponding to all the approximations under study, we found that the products resulting from the newly found TSs geometry were between 17 kcal mol⁻¹ and 21 kcal mol⁻¹ lower in energy than those of the currently accepted products of this reaction step (Fig. 2). Considering the most stable products, we also took note that the calculated R₂/R₆ distribution (26:74) was also contradicting the literature.

In light of these adverse results, we decided to more carefully scrutinize the literature, and found a report where two isomeric products were described, as resulting from the Duff formylation of 3-methylphenol (1), in a 1:1 ratio.²² This finding prompted us to carry out the reaction experimentally, in order to clear out all of our doubts.

Taking into account that our calculations were performed in acetic acid, the most widely employed solvent for the Duff reaction, we decided to set up the reaction in this solvent, following a modification of the procedure disclosed by Rychnovsky (Scheme 3).¹²

To our delight, we obtained a mixture of both expected ortho-isomers, in a 21:79 (R₂/R₆) distribution ratio according to ¹H NMR integration of the mixture, in perfect agreement with that resulting from the calculated energies of the products of the newly proposed TSs (R₂/R₆ = 26:74).

Bearing in mind the analysis performed for 1 and the newly discovered mechanism for this key-step, seven additional representative phenols, entailing different non-symmetric substitution patterns and considering steric hindrance, electron donating and withdrawing groups, were selected from the recent literature for closer examination.¹³⁻²³

To our delight, we obtained a mixture of both expected ortho-isomers, in a 21:79 (R₂/R₆) distribution ratio according to ¹H NMR integration of the mixture, in perfect agreement with that resulting from the calculated energies of the products of the newly proposed TSs (R₂/R₆ = 26:74).

Bearing in mind the analysis performed for 1 and the newly discovered mechanism for this key-step, seven additional representative phenols, entailing different non-symmetric substitution patterns and considering steric hindrance, electron donating and withdrawing groups, were selected from the recent literature for closer examination.¹³⁻²³

Fig. 1  Top: The TSs for the accepted mechanism. Bottom: The newly proposed mechanism, where migration of phenolic hydrogen is involved. The distances are in angstroms and the Wiberg bond indexes in parentheses.

Scheme 3  Duff reaction of 3-methylphenol (1).
In agreement with the thermodynamics of the reaction for 1, we found that the formylation process of all the studied phenol derivatives (2–8) is endergonic, with energies between 1.50 and 15.05 kcal mol$^{-1}$ (Table 1), and that the calculated product ratios based on their energies were in excellent agreement with the reported regioselectivities.

For all cases the energy barrier of this step demonstrated to be relatively low, between 17.61 and 24.42 kcal mol$^{-1}$. Curiously, in the cases of phenols 7 and 8, the TS-based calculations predicted selectivities opposite to those observed experimentally. Nevertheless, the real outcome of this forecast was corrected to finally agree with the experimental observations, by the thermodynamic analysis, as is encountered by the initial study with phenol 1.

In accordance with the analyzed 3-methylphenol 1, when the distances and WBI values were of phenols 2–8 were examined, we observed a scenario similar to the above described. The distances varied from 1.86 to 2.25 Å and the WBI values changed between 0.27 and 0.57 for the C–C bond formation. On the other hand, the hydrogen bond distances varied from 1.45 to 1.67 Å and the WBI fluctuated between 0.08 and 0.24. Interestingly, as detected in the case of phenol 1, the smaller the distance between carbons in the new C–C bond, the shorter was the length of the hydrogen bond. A concomitant WBI increase was also noticed.

The formylation of phenol 8 seems to be a suitable example for the conclusions observed herein. According to the studied TSs of the key intermediate of this reaction, it is expected that the transformation will take place to give the product, with a $R_2/R_6$ selectivity of ≈80 : 20; that is, favoring the $R_2$ product (Fig. 3), contrary to the experimental results ($R_2/R_6 = 0 : 100$).23a

However, examination of the TS structure of 8 corresponding to an $R_6$-substitution (TS-8$^\dagger$-HMTA$^-$-$R_6$) revealed that the proton has not yet migrated from the oxygen to the nearby nitrogen; this detail may be at the roots of its higher energies. Contrarily, however, a correct prediction of the selectivity could be achieved when the thermodynamics of the reaction was also analyzed.

![Fig. 3](image-url) Comparison of the reaction first stage for phenols 5 and 8 with distances in Å and Wiberg bond indices between parentheses. Energies in kcal mol$^{-1}$.

In accordance with the thermodynamics of the reaction for 1, we found that the formylation process of all the studied phenol derivatives (2–8) is endergonic, with energies between 1.50 and 15.05 kcal mol$^{-1}$ (Table 1), and that the calculated product ratios based on their energies were in excellent agreement with the reported regioselectivities.

For all cases the energy barrier of this step demonstrated to be relatively low, between 17.61 and 24.42 kcal mol$^{-1}$. Curiously, in the cases of phenols 7 and 8, the TS-based calculations predicted selectivities opposite to those observed experimentally. Nevertheless, the real outcome of this forecast was corrected to finally agree with the experimental observations, by the thermodynamic analysis, as is encountered by the initial study with phenol 1.

In accordance with the analyzed 3-methylphenol 1, when the distances and WBI values were of phenols 2–8 were examined, we observed a scenario similar to the above described. The distances varied from 1.86 to 2.25 Å and the WBI values changed between 0.27 and 0.57 for the C–C bond formation. On the other hand, the hydrogen bond distances varied from 1.45 to 1.67 Å and the WBI fluctuated between 0.08 and 0.24. Interestingly, as detected in the case of phenol 1, the smaller the distance between carbons in the new C–C bond, the shorter was the length of the hydrogen bond. A concomitant WBI increase was also noticed.

The formylation of phenol 8 seems to be a suitable example for the conclusions observed herein. According to the studied TSs of the key intermediate of this reaction, it is expected that the transformation will take place to give the product, with a $R_2/R_6$ selectivity of ≈80 : 20; that is, favoring the $R_2$ product (Fig. 3), contrary to the experimental results ($R_2/R_6 = 0 : 100$).23a

However, examination of the TS structure of 8 corresponding to an $R_6$-substitution (TS-8$^\dagger$-HMTA$^-$-$R_6$) revealed that the proton has not yet migrated from the oxygen to the nearby nitrogen; this detail may be at the roots of its higher energies. Contrarily, however, a correct prediction of the selectivity could be achieved when the thermodynamics of the reaction was also analyzed.

![Fig. 3](image-url) Comparison of the reaction first stage for phenols 5 and 8 with distances in Å and Wiberg bond indices between parentheses. Energies in kcal mol$^{-1}$.

Table 1: Calculated activation, formation energies and theoretical and experimental Duff distribution ratios$^{24}$

<table>
<thead>
<tr>
<th>Phenol Isomer</th>
<th>$\Delta G^{\neq}$</th>
<th>Calc. ratio ($R_2/R_6$)$^{b}$</th>
<th>$\Delta G^{\dagger}$</th>
<th>Calc. ratio ($R_2/R_6$)$^{b}$</th>
<th>Exp. ratio ($R_2/R_6$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 $R_2$</td>
<td>21.17</td>
<td>97:3</td>
<td>8.97</td>
<td>26:74</td>
<td>21:79</td>
</tr>
<tr>
<td>$R_6$</td>
<td>23.91</td>
<td>8.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 $R_2$</td>
<td>17.73</td>
<td>95:5</td>
<td>6.03</td>
<td>0:100</td>
<td>100:0$^{23d}$</td>
</tr>
<tr>
<td>$R_6$</td>
<td>20.03</td>
<td></td>
<td>13.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 $R_2$</td>
<td>24.36</td>
<td>0:100</td>
<td>11.08</td>
<td>0:100</td>
<td>100:0$^{23b}$</td>
</tr>
<tr>
<td>$R_6$</td>
<td>20.22</td>
<td>1.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 $R_2$</td>
<td>17.61</td>
<td>95:5</td>
<td>4.99</td>
<td>100:0</td>
<td>100:0$^{23e}$</td>
</tr>
<tr>
<td>$R_6$</td>
<td>19.93</td>
<td>11.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 $R_2$</td>
<td>20.84</td>
<td>96:4</td>
<td>10.09</td>
<td>98:2</td>
<td>89:11$^{23c}$</td>
</tr>
<tr>
<td>$R_6$</td>
<td>23.29</td>
<td></td>
<td>12.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 $R_2$</td>
<td>22.31</td>
<td>94:6</td>
<td>12.38</td>
<td>99:1</td>
<td>100:0$^{23a}$</td>
</tr>
<tr>
<td>$R_6$</td>
<td>24.42</td>
<td></td>
<td>15.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 $R_2$</td>
<td>19.88</td>
<td>73:27</td>
<td>9.50</td>
<td>5:95</td>
<td>0:100$^{23f}$</td>
</tr>
<tr>
<td>$R_6$</td>
<td>20.63</td>
<td>7.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 $R_2$</td>
<td>22.64</td>
<td>83:17</td>
<td>11.57</td>
<td>0:100</td>
<td>0:100$^{23a}$</td>
</tr>
<tr>
<td>$R_6$</td>
<td>23.87</td>
<td>7.87</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^{a}$ Energies in kcal mol$^{-1}$. $^{b}$ According to the Boltzmann distribution.

Fig. 3: Comparison of the reaction first stage for phenols 5 and 8 with distances in Å and Wiberg bond indices between parentheses. Energies in kcal mol$^{-1}$.
On the basis of these results, we rationalized that R₃ is preferably functionalized in cases where the aromatic ring was substituted with electron-withdrawing groups (phenols 5 and 6), whereas substitution at R₄ prevailed when donor groups were involved (phenols 1, 3, 7 and 8).

The influence of similar electronic effects of the aromatic ring substituents was previously noticed by the group of Marsaïoli in their study on the selectivity of the Claisen rearrangement of allyl phenyl ethers.²⁵ However, the electronic effects do not seem to be the only variable affecting the reaction selectivity.

On the basis of the results obtained for phenols 1 and 8, it is clear that in some cases it should also be considered the steric factor as a relevant influence to the reaction outcome. On the other hand, the exact nature of the behavior of the substituents in the more complex heterocycles (2 and 4) is still unknown, being probably related to the impact of electronic effects and other factors.

Based on all these observations, we propose that the C–C bond formation step of the Duff reaction should be represented as depicted in Scheme 4, and not as commonly accepted (Scheme 2). In addition we anticipate that, on the basis of these observations, the remaining stages of the reaction should also be re-examined, since our findings suggest that the transformation takes place through a path different from that traditionally accepted.

**Experimental section**

**Computational methods**

Conformational searches for the reactants, the transition structures (TSs) and the products were run to locate the global minima employing the M06-2X functional¹⁵ coupled with the 6-311+G(d,p) level of theory. Initially, a large number of geometries were generated using the conformational search module of Hyperchem²⁶ with the MM+ method for the TSs and Spartan’s conformer distribution module, employing MMFF for both, reactants and products.²⁷

The selected structures were then optimized at the M06-2X/6-311+G(d,p) level of theory, using Gaussian 09.¹⁸ The geometries for all structures were fully optimized including acetic acid (ε = 6.2528) as solvent using SMD.¹⁷

The reported thermochemical properties include zero-point energies (ZPEs) without scaling and were calculated at 1 atm, and the corresponding temperature for each reaction.²⁴

Normal mode analysis was used to confirm the nature of the stationary points and to evaluate the thermochemical properties. All transition structures were confirmed to have only one imaginary frequency corresponding to the formation of the expected bonds. Intrinsic reaction coordinate (IRCs) calculations were run to verify the connectivity between reactants, TSs and products.

**Duff formylation of 3-methylphenol (1).** Anhydrous hexamethylenetetramine (512 mg, 3.65 mmol) was added in one portion to a stirred solution of the phenol (200 mg, 1.85 mmol) in AcOH (4 mL), under a nitrogen atmosphere. The resulting yellow solution was heated to 110° during 2.5 h, when the reaction was cooled to room temperature and treated with 6 M HCl (10 mL) for 40 min. Then, water (20 mL) was added and the aqueous solution was extracted with CH₂Cl₂ (3 × 15 mL). The combined organic extracts were washed with brine, dried over Na₂SO₄ and concentrated under reduced pressure. The crude product was chromatographed, furnishing a 79:21 mixture of 2-hydroxy-4-methyl benzaldehyde and 2-hydroxy-6-methyl benzaldehyde (30 mg, 12%). The spectral data of the so obtained compounds were in agreement with those reported.²⁸

**2-Hydroxy-4-methylbenzaldehyde.** ¹H NMR (δ, CDCl₃, 300 MHz): 2.38 (s, 3H, ArMe), 6.80 (s, 1H, H-3), 6.82 (d, 1H, J = 6.8 Hz, H-5), 7.43 (d, 1H, J = 6.8 Hz, H-6), 9.83 (s, 1H, CHO) and 11.03 (s, 1H, OH) ppm; ¹³C NMR (δ, CDCl₃, 75 MHz): 22.2 (ArMe), 117.6 (C-3), 118.7 (C-1), 121.1 (C-5), 133.5 (H-6), 148.9 (C-4), 161.7 (C-2) and 195.8 (CHO) ppm.

**2-Hydroxy-6-methylbenzaldehyde.** ¹H NMR (δ, CDCl₃, 75 MHz): 2.61 (s, 3H, ArMe), 6.72 (d, 1H, J = 7.0 Hz, H-5), 6.82 (d, 1H, J = 7.0 Hz, H-3), 7.38 (t, 1H, J = 6.8 Hz, H-4), 10.33 (s,
1H, CHO) and 11.90 (s, 1H, OH) ppm; 13C NMR (δ, CDCl₃, 75 MHz): 18.1 (ArMe), 116.1 (C-3), 118.6 (C-1), 121.8 (C-5), 137.4 (C-4), 142.1 (C-6), 163.2 (C-2) and 195.3 (CHO) ppm.

Acknowledgements

This research was supported by the Agencia Nacional para la Promoción Científica y Tecnológica (ANPCyT), PICT-2012-0970 and PICT-2014-0445. N. G. and S. O. S. acknowledge CONICET for awarding their fellowships.

Notes and references


20 It should be noted that most of the reactions that generate 4-hydroxybenzaldehydes derivatives, usually afford mixtures of products, with one being prevalent, or are carried out with phenols where both ortho positions are substituted.


24 Temperatures employed for the energy calculation of each phenol: 1 = 110 °C; 2 = 90 °C; 3 = 100 °C; 4 = 70 °C; 5 = 75 °C; 6 = 140 °C; 7 = 110 °C; 8 = 110 °C.


27 Spartan'08, Wavefunction, Irvine, CA.