J Mol Model (2007) 13:823–837
DOI 10.1007/s00894-007-0205-8

ORIGINAL PAPER

Computational prediction of the regio- and diastereoselectivity in a rhodium-catalyzed hydroformylation/cyclization domino process

Giuliano Alagona · Caterina Ghio · Silvia Rocchiccioli

Received: 7 November 2006 / Accepted: 27 March 2007 / Published online: 22 May 2007

Abstract The regioselectivity of the hydroformylation reaction of 2-methyl-3-(3-acetylpyrrol-1-yl)prop-1-ene catalyzed by an unmodified Rh catalyst has been investigated at the B3LYP/6-31G* level with Rh described by effective core potentials in the LANL2DZ valence basis set. Considering the population of all the H-Rh$(CO)_3$-olefin transition state complexes, a regioselectivity ratio (B:L) of 12:88 has been obtained, in satisfactory agreement with the experiment producing the chiral linear aldehyde as the only product. The aldehyde, after complete diastereoselective cyclization, yields a 1:1 mixture of 1-acetyl-6*R(S)*-methyl-8*R(S)*-hydroxy-5,6,7,8-tetrahydroindolizine (having the same configuration on both stereogenic carbon atoms) and 2-acetyl-6-methyl-5,6-dihydroindolizine [Lett Org Chem (2006) 3:10–12]. The reason for such a high degree of diastereoselectivity has been elucidated examining the B3LYP/6-31G* potential energy surface for the reactions leading to the *RR* and *RS* diastereomers on a model system (without the acetyl substituent) and the actual compound. In the absence of a catalyst, a very high barrier is found along the reaction pathway, whereas spontaneous annulation occurs to a protonated pentahydroindolizine in the presence of H^+. When a counterion (F^-) is added, the proton on the newly formed tetrahedral carbon is abstracted, obtaining a structure closer to the final product (tetrahydroindolizine). Replacing H^+ with Rh^+, an initial adduct along the *RS* path much more favorable than any of those computed along the *RR* one is located because of the presence of the acetyl group. Tentative approaching paths obtained using [Rh$(CO)_3]^+$, bound to the aldehyde O, feature a higher barrier along the *RS* one, and offer a convincing explanation for the observed diastereoselectivity.

Keywords LANL2DZ · B3LYP/6-31G* · Annulation · Substrate-induced diastereoselectivity

Abbreviations
HF Hartree-Fock
MP2 Møller-Plesset second order perturbation

Introduction

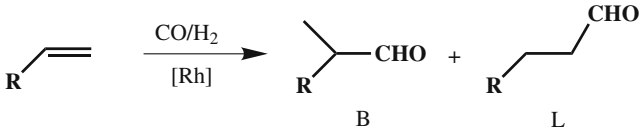
The control of stereoselectivity is one of the prominent issues in the chemistry of chiral compounds. The use of quantum mechanics to solve it, although pioneering works date back to the middle of the last century [1–3], is still a challenging task, especially when transition element complexes are involved [4]. In the case of hydroformylation, a typical homogeneous reaction catalyzed by low valence Co or Rh complexes, for instance, a considerable amount of theoretical investigations were carried out, mostly using phosphine modified Rh(I) catalysts, on ethene [5–7], the smaller substrate, that however cannot even address regioselectivity. Going to the next olefin, propene [8–10], regioselectivity can be taken into account, as easily derivable from Scheme 1.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-007-0205-8) contains supplementary material, which is available to authorized users.

G. Alagona · C. Ghio (✉)
CNR-IPCF (Institute for Physico-Chemical Processes),
Molecular Modelling Lab,
Via Moruzzi 1, 56124 Pisa, Italy
e-mail: C.Ghio@ipcf.cnr.it

G. Alagona
e-mail: G.Alagona@ipcf.cnr.it

S. Rocchiccioli
DCCI, University of Pisa,
Via Risorgimento 35, 56126 Pisa, Italy



Scheme 1 Hydroformylation of a terminal olefin. Branched and linear aldehydes are produced for R ≠ H

Although a chiral center appears when R is an alkyl group greater than CH_3 or an aromatic group, nonetheless hydroformylation with unmodified rhodium catalysts produces the racemic branched aldehydes without any stereoselectivity. In contrast, when the substrate is a chiral olefin itself, diastereoselectivity might emerge as shown for two particular cases in Scheme 2.

The regioselectivity issue for a fairly large series of substrates in the presence of an unmodified rhodium catalyst has been successfully examined by us in a previous article, where the irreversible olefin insertion into the Rh–H bond was demonstrated to be the step determining the regioselectivity by comparing theoretical and experimental results [11]. The B:L regioisomeric ratio, theoretically evaluated from the internal energy difference of branched and linear alkyl-rhodium transition states (TS):

$$\mathrm{B:L} = k_B{:}k_L = \sum \mathrm{k}_B[\mathrm{C}]{:}\sum \mathrm{k}_L[\mathrm{C}] = \sum \mathrm{e}^{-\Delta \mathrm{G}_\mathrm{B}^{\neq}/\mathrm{RT}} {:}\sum \mathrm{e}^{-\Delta \mathrm{G}_\mathrm{L}^{\neq}/\mathrm{RT}} = \sum \mathrm{e}^{-\Delta\Delta \mathrm{G}^{\neq}/\mathrm{RT}} \approx \sum \mathrm{e}^{-\Delta\Delta \mathrm{E}^{\neq}/\mathrm{RT}} \quad (1)$$

(where k = reaction rate, k = rate constant, [C] = concentration of the olefin-Rh complex, $\Delta G^{\neq}$ = TS free energy, $\Delta E^{\neq}$ = TS internal energy) was very close to the experimental result for all the substrates considered.

Analogously, the theoretical diastereomeric ratios (b:b′) for the substrates displayed in Scheme 2, which represent interesting cases of substrate asymmetric inductions, turned out in good agreement with the relevant experimental results [12]. An expression similar to that reported in Eq. (1) was used to evaluate the diastereomeric ratio:

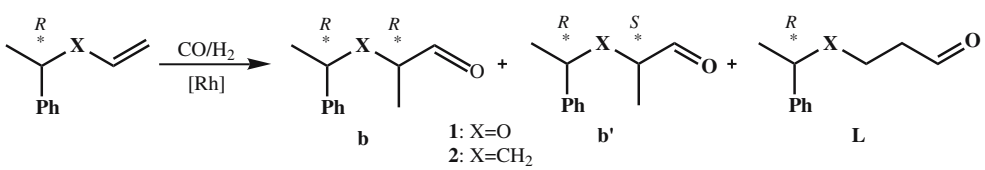
$$\mathrm{b:b'} = k_b{:}\,k_{b'} = \sum \mathrm{e}^{-\Delta \mathrm{G}_\mathrm{b}^{\neq}/\mathrm{RT}} {:}\sum \mathrm{e}^{-\Delta \mathrm{G}_\mathrm{b'}^{\neq}/\mathrm{RT}} = \sum \mathrm{e}^{-\Delta\Delta \mathrm{G}^{\neq}/\mathrm{RT}} \approx \sum \mathrm{e}^{-\Delta\Delta \mathrm{E}^{\neq}/\mathrm{RT}} \quad (2)$$

Method and basis set effects, taken into account for the substrates of Scheme 2 including in the test set even a compound with the stereocenter directly linked to the olefin double bond (i.e., without any X separator), produced only limited changes in the computed ratios [13].

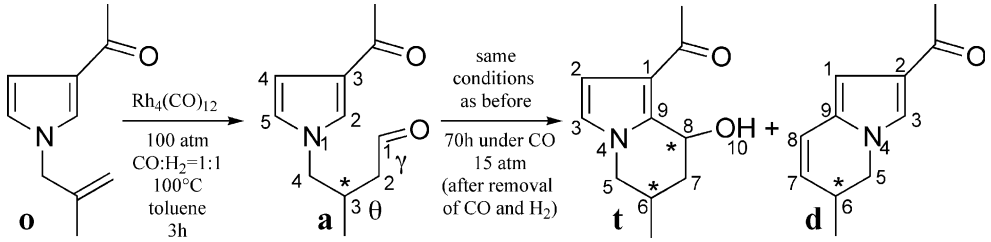
Hydroformylation represents an attractive synthetic transformation, because it allows the introduction of an aldehyde function, thus producing a starting material for carbon skeleton expanding operations, via a sequence of coupled hydroformylation-Wittig reactions, that after each step returns an olefin with an additional carbon atom with respect to the starting olefin and so on. The potentiality of this reaction is even enhanced when it is part of a domino process and/or leads to the formation of new stereocenters for the occurrence of additional reaction steps. An interesting example of further reaction has been reported earlier this year [14] (see Scheme 3).

During hydroformylation, the starting olefin (**o**, a prochiral substrate), is completely converted into 3-methyl-4-(3-acetylpyrrol-1-yl)butanal (**a**), the corresponding *linear* chiral aldehyde. This sharp regioselectivity is not surprising, since either aryl or alkyl 1,1-disubstituted olefins are known to afford the linear aldehyde as the only product [15–18]. If after aldehyde formation the CO/H_2 gas mixture is removed, to avoid reduction of the aldehyde carbonyl group to the hydroxyl function, pressurizing the reactor with CO, the linear chiral aldehyde disappears and the alcohol **t** in a 1:1 mixture with the dihydroindolizine **d** is formed. According to the literature [19, 20], the formation of **d** is explainable by an electrophilic attack of the aldehyde **a** carbonyl group onto the C_5-pyrrole carbon atom (Scheme 3), producing a bicyclic alcohol that very easily undergoes water elimination to give a double bond conjugated with the pyrrole ring. An analogous attack on the C_2-pyrrole carbon atom leads to the alcohol **t** that, conversely, is perfectly stable under reaction conditions and, even after manipulation, does not give water elimination. Moreover, an intramolecular H-bond between the hydroxy H and the acetyl carbonyl O is strongly supported by IR and ^{1}H-NMR measurements [14]. Concerning the reaction selectivity, **t** turns out to have the same absolute configuration at C_6 and C_8 (i.e. 6*R*,8*R* or 6*S*,8*S*), depending on the chirality at C_6.

This intriguing new example of complete substrate-induced diastereoselectivity spurred us to examine the whole reaction mechanism, starting from the elucidation of the hydroformylation regioselectivity via the substrate-catalyst TS complexes. The final transformation was studied in the presence of a number of species in turn, in an effort to explain the origin of the observed diastereoselectivity.



Scheme 2 Hydroformylation of a chiral (*R*) terminal olefin. Two branched diastereomers, *RR* (b) and *RS* (b′) as well as the linear aldehyde can be obtained



Scheme 3 Hydroformylation of 2-methyl-3-(3-acetylpyrrol-1-yl)prop-1-ene (**o**) followed by the reaction of **a** producing a 1:1 mixture of 1-acetyl-6-methyl-8-hydroxy-5,6,7,8-tetrahydroindolizine and 2-acetyl-6-methyl-5,6-dihydroindolizine, **t** and **d**, respectively, with atom numberings and torsions: $\gamma = OC_1C_2C_3$ and $\theta = C_1C_2C_3C_4$

Computational details

All calculations have been carried out with the Gaussian 03 system of programs [21], in the density functional theory (DFT) framework, making use of B3LYP, i.e., the Becke gradient-corrected three-parameter hybrid exchange and Lee-Yang-Parr correlation functionals [22, 23]. Some geometry optimizations in vacuo have been carried out ab initio at the Hartree-Fock (HF) and MP2/6-31G* levels [24] for comparison. Coupled to the B3LYP/6-31G* description for C, O and H, effective core potentials that implicitly include some relativistic effects for the electrons near the nucleus in the LANL2DZ valence basis set have been used for Rh [25]. However, reference is made to 6-31G* only, even when Rh is present, to simplify the notation.

Results and discussion

Hydroformylation regioselectivity

In the case of 2-methyl-3-(3-acetylpyrrol-1-yl)prop-1-ene, the substrate of the hydroformylation reaction under scrutiny, there are several torsional degrees of freedom to be considered for determining the TS populations needed to compute the selectivity ratio (Eq. (1)). It is in fact necessary to take into account not only the rotation about the $C_1C_2C_3N$ dihedral, but also those about C_2C_3NC and CCCO, since the substituent to the pyrrole ring can interact with the catalyst through space in a completely different manner depending on its geometric arrangement. Because of this reason, 12 local minima have been located for each of the linear and branched TS at the B3LYP/6-31G* level, reported in Table 1. They have been grouped in three sets (corresponding to the threefold C-CH_2Pyr axis) of four conformations each, since there are just two stable positions for the pyrrole ring and the acetyl group rotations, keeping the linear energies in ascending order as far as possible. The branched conformers differ primarily from the linear ones placed on the same row of Table 1 because of the catalyst arrangement.

The lowest energy linear and branched structures, 1L (displayed in Fig. 1) and 2B, belong to the first set with the pyrrole substituent in the half-plane opposite to the catalyst. They are quite similar apart the catalyst arrangement, as stated above. Also the pyrrole ring is rotated with its 3-position and the carbonyl group of its acetyl substituent in the same orientation as the Rh–H bond. Interestingly enough, the second most stable structures of both the linear and branched conformers (2L and 1B (the latter also displayed in Fig. 1)), just about 0.1 kcal mol^{-1} higher in energy, show the switched orientation. Of course, the energies are too close to call, but both contribute most to the regioselectivity ratio. The 3 and 4 structures have the pyrrole ring as in 1 and 2, respectively, with the acetyl group rotated by about 180°. In the second set, the pyrrole N makes about 180° with the olefin C_1 carbon atom. The 5

Table 1 B3LYP/6-31G* relative energies (kcal mol^{-1}) for the linear and branched H-Rh(CO)$_3$ / 2-methyl-3-(3-acetylpyrrol-1-yl)prop-1-ene TS complexes

	positions[a]	linear	branched
B3LYP/6-31G*	C_3/CO	ΔE	ΔE
1	ll	0[b]	1.0870
2	rr	0.0777	0.9715
3	lr	0.4562	1.4865
4	rl	0.5630	1.6595
5	dd	0.4719	2.2614
6	ud	0.8999	2.8292
7	uu	0.9447	2.7244
8	du	1.1670	3.1342
9	dd	1.2582	2.7932
10	ud	1.5302	2.9370
11	uu	1.6051	3.3429
12	du	2.1113	3.7977

[a] Rough indication of the substituent positions: l = left, r = right, d = down, u = up with respect to a standard view as in Fig. 1; [b] Reference energy=−968.962730 E_h.

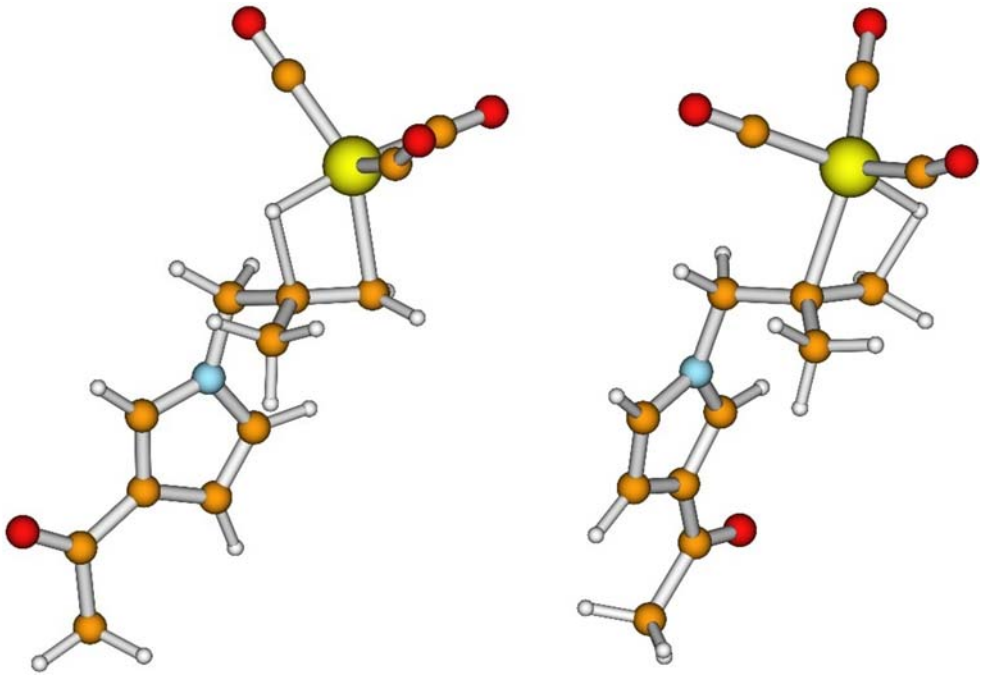


Fig. 1 B3LYP/6-31G* optimized structures of 1L (left hand side) and 1B (right hand side)

and 8 structures feature the pyrrole ring with its 3-position in the half-plane opposite to the catalyst and the acetyl carbonyl group pointing down (5) or up (8). The pyrrole substituent in 7 and 6 is rotated by 180° with respect to 5 and 8, respectively. In the third set, the pyrrole N is about perpendicular to the olefin C_1 carbon atom in the linear structures whereas it is syn in the branched ones to keep the pyrrole ring as far away as possible from the catalyst CO groups. The 9 and 12 structures have the pyrrole 3-position pointing down and the acetyl carbonyl group pointing down (9) or up (12). In 11 and 10 the substituent is rotated by 180° relative to 9 and 12, respectively.

Concerning the computed regioselectivity, seven linear conformers are lower in energy than any of the branched

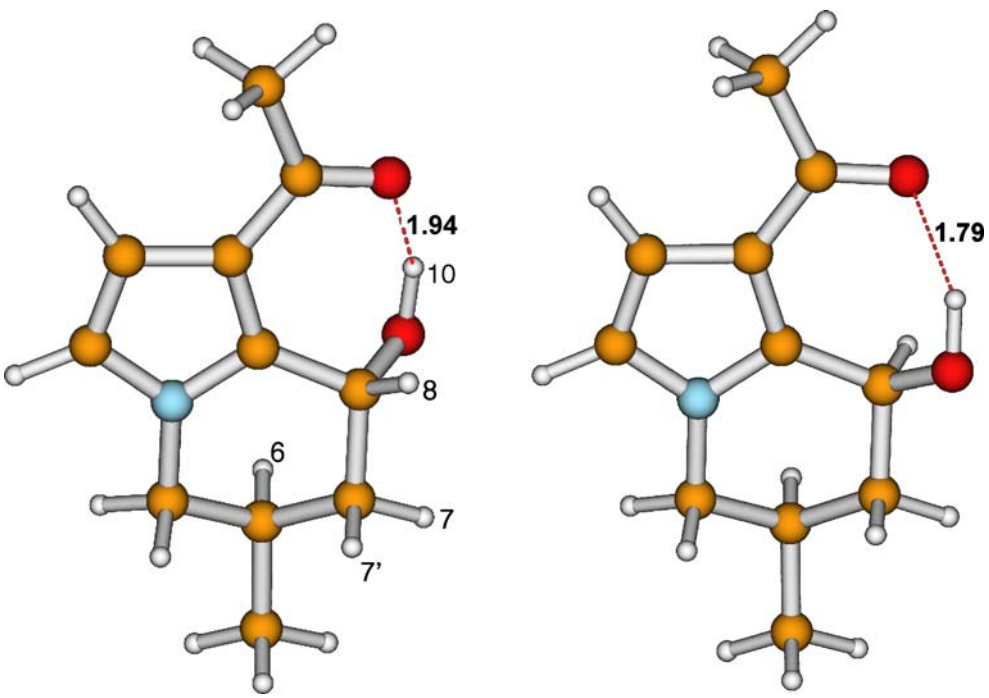


Fig. 2 B3LYP/6-31G* gas-phase optimized structures of (6*R*,8*R*) **t**, with the numbering of some hydrogens (left hand side) and (6*R*,8*S*) (right hand side)

Table 2 HF, MP2 and B3LYP/6-31G* relative energies (kcal mol^{-1}) and H-bond distances (Å) for the 1-acetyl-6-methyl-8-hydroxy-5,6,7,8-tetrahydroindolizine diastereomers

	8*R* [8*S*]		8*S* [8*R*]	
HF/6-31G*	ΔE	H···O	ΔE	H···O
6*R* [6*S*]	0[a]	2.118	0.36	1.901
$6R_a$ [$6S_a$]	1.88	1.916	1.77	2.048
MP2/6-31G*				
6*R* [6*S*]	0[b]	2.016	0.85	1.846
$6R_a$ [$6S_a$]	1.38	1.864	0.52	1.946
B3LYP/6-31G*				
6*R* [6*S*]	0.08	1.947	0[c]	1.797
$6R_a$ [$6S_a$]	1.49	1.808	1.27	1.903

[a] Reference energy= −629.464073 E_h. [b] Reference energy=−631.403453 E_h. [c] Reference energy=−633.409100 E_h.

ones. The stability of the last five linear conformers is comparable to that of the most favorable branched ones. Therefore, although the large regioselectivity of vinylidenic olefins under hydroformylation has been mainly ascribed to one of the last reaction steps (CO insertion into the Rh–C bond [17, 18]), in this case the computed ratio (B:L=12:88) decidedly favors the linear conformers, consistently with the experimental result and prior calculations on 2-methylpropene [11]. An additional hybrid method, B3P86 [23, 26], satisfactorily employed in previous studies [11, 13], has been used as well, since the metal-olefin bond strength and the transition state stability are sensitive to the electron correlation description [5]. The B3P86 results are fairly comparable to the B3LYP ones as far as the relative stabilities of TS complexes are concerned [27], although both of them are not recommended from calculations on test-sets (6 databases) made up of smaller systems [28].

For this particular substrate, however, the reaction proceeds further as shown in Scheme 3.

Product stabilities and properties

Structures and energies of the two final diastereomers (6*R*,8*R*) and (6*R*,8*S*) of 1-acetyl-6-methyl-8-hydroxy-5,6,7,8-tetrahydroindolizine (**t**) have been computed in the gas phase with full geometry optimization at the ab initio HF/MP2 and DFT/B3LYP levels, using the 6-31G* basis set. Aim of these calculations was to compare the computed structures with those derived from experiment and evaluate the relative stability of the products. Both half-chair conformers have been considered for the six-membered heteroring, namely *R* (with C_6 below the ring plane and the methyl group at C_6 in equatorial position) as in Fig. 2, and R_a (with C_6 above the ring plane and the methyl group at C_6 in axial position).

The relative stabilities at the three levels are reported in Table 2 together with the relevant H-bond separations. Of course, the (6*S*,8*S*) and (6*S*,8*R*) structures (*SS* and *SR*, respectively, for short) are the mirror images of the *RR* and *RS* ones, displayed in Fig. 2. The same holds for the other diastereomers in square parentheses in Table 2.

From a perusal of the table, energetically close-lying *RS* and *RR* diastereomers do not allow a reliable assignment of the relative stabilities (even including MP2 electron correlation), preventing any of them from being recognized as the most stable one. Their B3LYP/6-31G* structures, however, have been used to compare NMR theoretical and experimental results. Both the ^1H and ^{13}C NMR chemical shifts using the gauge-including atomic orbital method (GIAO), which achieves gauge invariance with basis functions having an explicit magnetic field dependence [29–33], have been computed at the HF/6-311+G(2d,p) level. The results for the *RR* and *RS* diastereomers are reported in Table 3.

Since the H_8 and H_9 chemical shifts, measured on a Varian Gemini 600 MHz using TMS as an internal standard, are 5.08 and 5.38 ppm, respectively [14], the

Table 3 Chemical shifts (in ppm) with respect to TMS[a] computed at the HF/6-311+G(2d,p)//B3LYP/6-31G* level for the the *RR* and *RS* diastereomers of **t**

	RR		RS	
atom[b]	^{13}C	^1H	^{13}C	^1H
C_1	120.7		120.3	
C_2/H_2	115.9	6.8	117.6	6.9
C_3/H_3	123.3	6.7	122.8	6.6
$C_5/H_5/H_{5'}$	49.6	3.6/3.1	49.5	3.5/3.2
C_6/H_6	24.9	2.1	27.9	1.7
$C_7/H_7/H_{7'}$	35.1	2.0/1.2	36.2	2.1/1.4
C_8/H_8	56.0	4.6	58.9	4.5
C_9	156.1		158.1	
H_{10}		5.6		7.2
C_{11}	207.6		207.4	
$C_{12}/H_{12}/H_{12'}/H_{12''}$	27.3	2.2/2.4/2.5	27.7	2.2/2.5/2.4
$C_{14}/H_{14}/H_{14'}/H_{14''}$	18.7	1.2/0.7/1.1	18.9	1.2/0.9/1.1

[a] Reference values for ^{13}C and ^1H are 192.5929 and 32.0725 ppm, respectively.
[b] For the atom numbers, refer to Scheme 3.

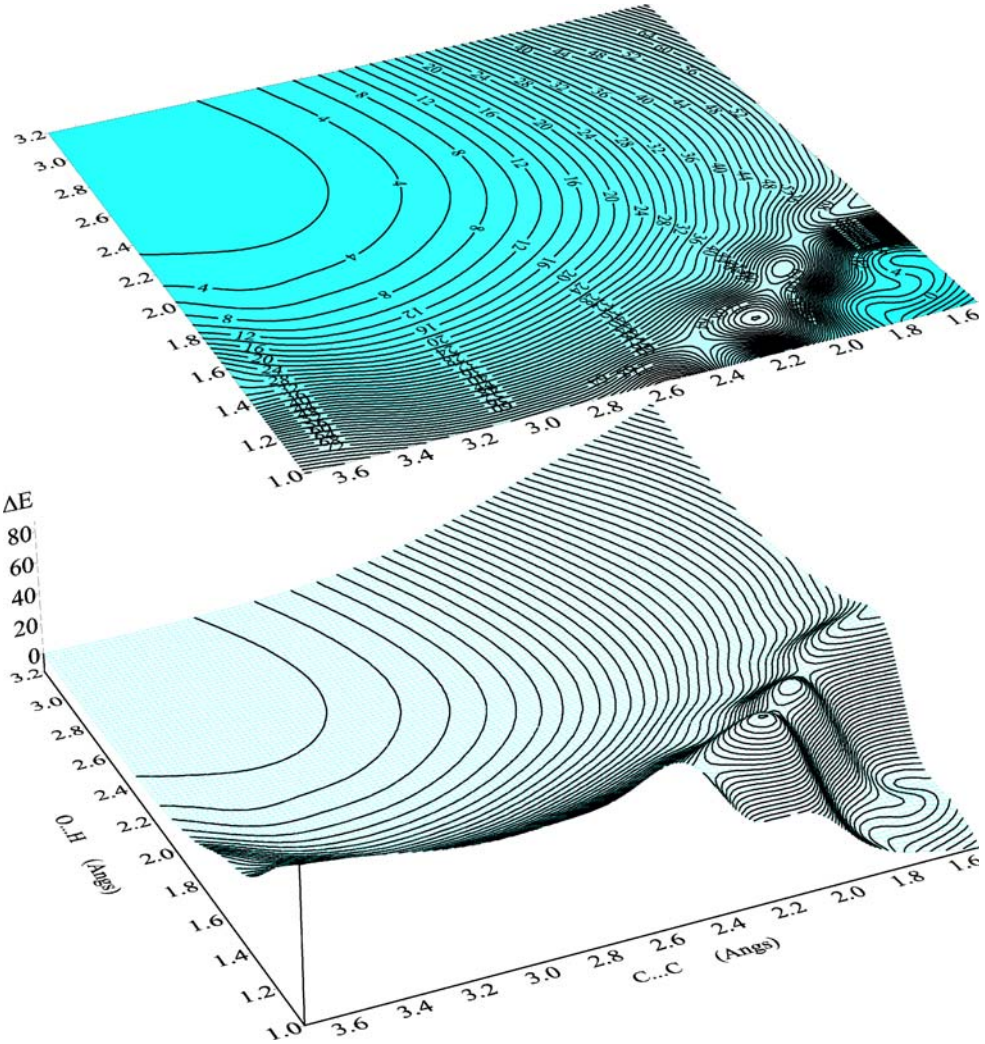


Fig. 3 B3LYP/6-31G* potential energy surface for the cyclization reaction of (3*R*) 3-methyl-4-(3-acetylpyrrol-1-yl) butanal to (6*R*,8*R*) 1-acetyl-6-methyl-8-hydroxy-5,6,7,8-tetrahydroindolizine with the $C_8\cdots C_9$ and $O_{10}\cdots H_9$ separations as leading parameters (isopotential curves spaced by 2 kcal mol^{-1})

theoretical results for *RR* (4.6 and 5.6 ppm) are in a much better agreement with them than those computed for *RS* (4.5 and 7.2 ppm). This result is consistent with the experiment because, as already mentioned, experimental data support a much larger population for *RR* and *SS* than for *RS* and *SR*, indicating once again that the reaction diastereoselectivity does not reside in the product different stabilities. This is not unexpected given our previous investigations that pointed to the differential transition state (TS) stability, indeed, as the key factor to explain, and hopefully predict, selectivity [11–13].

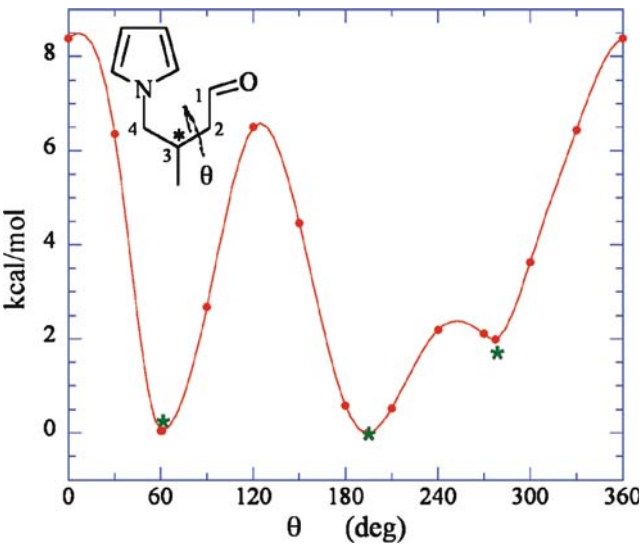
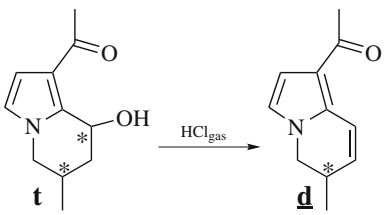


Fig. 4 B3LYP/6-31G* energy profile for the rotation of butanal about $C_1C_2C_3C_4$ (θ). Stars correspond to the minima for the acetyl-substituted aldehyde (the lowest energy is taken as zero for both systems (see Table 4))

Scheme 4 In the presence of HCl_g t dehydrates to **d**



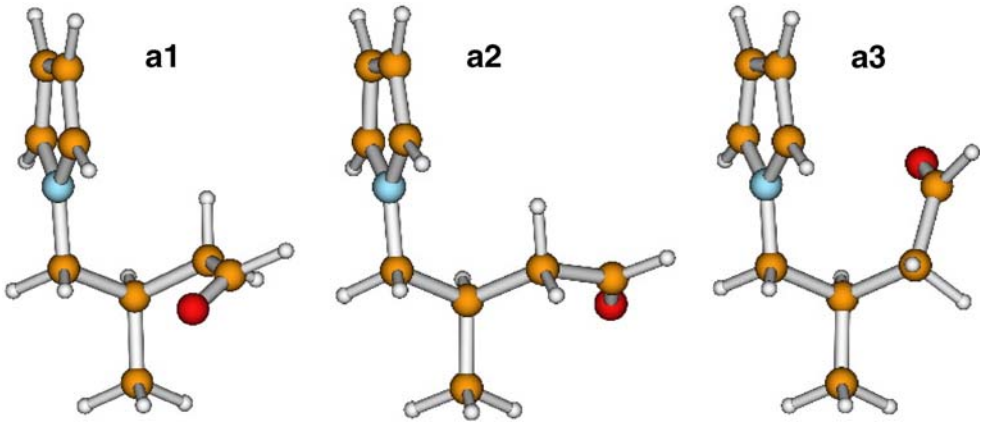


Fig. 5 B3LYP/6-31G* minimum energy structures along the profile in Fig. 3

Annulation mechanism

In order to elucidate the reaction mechanism, the potential energy surface (PES, displayed in Fig. 3) for the reaction producing the *RR* diastereomer has been computed at the B3LYP/6-31G* level to locate the TS, using as leading parameters the $C_8\cdots C_9$ and $O_{10}\cdots H_9$ separations.

From the inspection of the map, three almost equivalent TS can be found (one of them is displayed in Fig. S1 of the Electronic Supplementary Material (ESM)) with barriers of ~69 kcal mol^{-1} separating the reactant from the product, which is 11.94 kcal mol^{-1} more stable than the reactant. Those barrier height values indicate that the $C_8\cdots C_9$ cyclization cannot occur in the absence of a suitable catalyst, as we confirmed experimentally by heating the pure aldehyde **a** at high temperature for a long time in the absence of a catalyst.

Considering that the cyclization occurs as a result of an intramolecular electrophilic aromatic substitution at the electron-rich pyrrole α-positions, all those species, which make the carbonyl carbon atom more electron-poor, can promote the process. In fact the treatment of **a** with a simple Brønsted acid, such as HCl(gas) or HCl(aq), immediately gives cyclization, although followed by dehydration to the corresponding dihydroindolizines, **d** and **d** (Scheme 4) [14, 34]. Consequently, H^+ and H_3O^+ have been preliminarily employed as catalysts also in theoretical investigations. The rationale behind this is that they should satisfactorily depict the annulation process although from the experimental results they are expected to be even too effective.

The cyclization of **a** to produce the alcohol **t** under oxo conditions is likely to be rhodium-promoted. Nonetheless, the proper species is not an Rh(0) as the linear aldehyde **a** does not give annulation in the presence of $Rh_4(CO)_{12}$ under inert atmosphere (Argon), at high temperature for long time. An Rh(I) species acting as a Lewis acid is presumably involved. Therefore Rh^+ has been tentatively employed for preliminary scans, since the two catalysts considered, i.e., $[Rh(CO)_3]^+$ and $HRh(CO)_3$, are heavily computer-intensive, thus preventing their use in extensive calculations.

In the following, unless otherwise specified, 3-methyl-4-(pyrrol-1-yl)butanal was used as a model compound (i.e., no acetyl substituent to the pyrrole ring). This model system, that experimentally undergoes hydroformylation much more easily than when acetyl substituted [35–37], was tested computing the $C_8\cdots C_9$ and $O_{10}\cdots H_9$ PES, that turned out to be just slightly smoother (by about 5 kcal mol^{-1}) than the actual system one (shown in Fig. 3): barrier heights were ~64 kcal mol^{-1} and the product was more stable than the reactant by ~7 kcal mol^{-1}.

Table 4 B3LYP/6-31G* relative energies (kcal mol^{-1}), $C_8\cdots C_9$ separations (Å) and θ values (deg) for the minima along the θ profile for 3-methyl-4-(pyrrol-1-yl)butanal (**a**) and 3-methyl-4-(3-acetylpyrrol-1-yl)butanal (**A**)

	3-methyl-4-(pyrrol-1-yl)butanal			3-methyl-4-(3-acetylpyrrol-1-yl)butanal		
	ΔE	$C_8\cdots C_9$	θ	ΔE	$C_8\cdots C_9$	θ
a1/A1	0.04	3.611	60.90	0.26	3.700	61.31
a2/A2	0[a]	4.849	194.68	0[b]	4.862	195.00
a3/A3	2.03	3.256	277.34	1.69	3.256	277.28

[a] Reference energy=−480.737196 E_h. [b] Reference energy=−633.390358 E_h.

Table 5 B3LYP/6-31G* relative energies (kcal mol^{-1}) and configurations of the lowest energy closed adducts for 3-methyl-4-(pyrrol-1-yl) butanal and 3-methyl-4-(3-acetylpyrrol-1-yl)butanal

	H^+		H_3O^+	
3-methyl-4-(pyrrol-1-yl)butanal				
a1	6.65	*RS(R)*	5.16	*RS(S)*
a2	–	–	0[b]	*RR(R)*
a3	0[a]	*RR(R)*	0[b]	*RR(R)*
3-methyl-4-(3-acetylpyrrol-1-yl)butanal				
A1	6.50	*RS(R)*	–	–
A2	–	–	–	–
A3	0[c]	*RR(R)*	–	–

[a] Reference energy=−481.115772 E_h. [b] Reference energy=−577.550229 E_h. [c] Reference energy=−633.761449 E_h.

To study the cyclization mechanism, the aldehyde chain arrangement has been examined first, taking into account the θ and γ torsions (Scheme 3), respectively, in Figs. 4 and S2.

The minimum energy structures along the θ profile, namely **a1** (θ ≈ 60°), **a2** (θ ≈ 195°), and **a3** (θ ≈ 270°), shown in Fig. 5, correspond to minima even including the acetyl substituent to the pyrrole ring (stars in Fig. 4)[1], although the relative stabilities (reported in Table 4) change somewhat; it is likely that barriers are affected as well, but they have not been investigated, because they should remain quite high. The barrier heights separating **a1** from **a2**/**a3** and the relevant basins, shown in Fig. 4, possibly give a hint on the reaction stereoselectivity.

A number of additional structures, described in Table S1 of the ESM, have been taken into account such as the R_a structures (i.e., with the methyl group at C_3 in axial position).

The rotation of the aldehyde head (γ = $OC_1C_2C_3$), important for the stereochemistry of the reaction, has also been considered with the flexible scan for **a1** and **a2**, the two lowest energy minima, occurring for γ ≈ 0°. The B3LYP/6-31G* energy profiles, displayed in Fig. S2 of the ESM, are fairly similar, although the profile for **a1** is slightly higher, probably due to the fact that the aldehyde group is in somewhat closer vicinity to the pyrrole ring. The local minimum for **a2** at γ ≈ 240° occurs when the aldehyde H faces the pyrrole N, since the pyrrole ring plane is, in general, nearly perpendicular to the aldehyde chain (Fig. 5).

Reaction mechanism in the presence of H^+ or H_3O^+

The **a1**, **a2**, and **a3** structures have been used as starting conformations for the addition of H^+ or H_3O^+ to the aldehyde oxygen. When annulation occurs (C_8–C_9≤1.6 Å), a tetrahedral carbocation is formed at C_8 thus producing a third stereocenter (transient), indicated in parenthesis. From the stabilities of the optimized closed adducts, reported in Table 5, it appears evident that the *RS* configurations are much less stable than the *RR* ones. It should be stressed however that, in the *RS* structures obtained for the **a1** protonated aldehyde (**a1H$^+$**), the six membered heteroring is somewhat distorted: the methyl group at C_6 is equatorial with respect to C_6–C_5–N, and axial with respect to C_6–C_7–C_8. In the case of **a2H$^+$**, the system does not cyclize: the $C_8\cdots C_9$ separation decreases from 4.85 to 3.99 Å at most. In contrast **a3H$^+$**, reachable from **a2** for rotation about θ (Fig. 4), produces *RR(R)*.

Also the protonated 3-methyl-4-(3-acetylpyrrol-1-yl) butanal was studied to assess the validity of the model system under this respect as well. The behavior of the actual compound is very similar to that of the model system in the case of protonation. Geometry optimization starting from **A1H$^+$** produces annulation (C_8–C_9=1.595 Å) to an *RS (R)* 1-acetyl-6-methyl-8-hydroxy-5,6,7,8,9-pentahydroindolizine with a distortion of the six-membered ring analogous to what observed for **a1H$^+$**. Starting from **A2H$^+$**, albeit the $C_8\cdots C_9$ separation decreases to ~4 Å, the optimization gets stuck into a local minimum, probably because there is a small barrier to rotation about the C_6–C_7 bond (to reach the **a3** arrangement from **a2** ~2 kcal mol^{-1} had been spent). When considering **A3H$^+$** where the aldehyde head is in a favorable position, the cyclization occurs producing an *RR (R)* pentahydroindolizine where the six-atom ring adopts a chair conformation. In the presence of a counterion (F^-), an

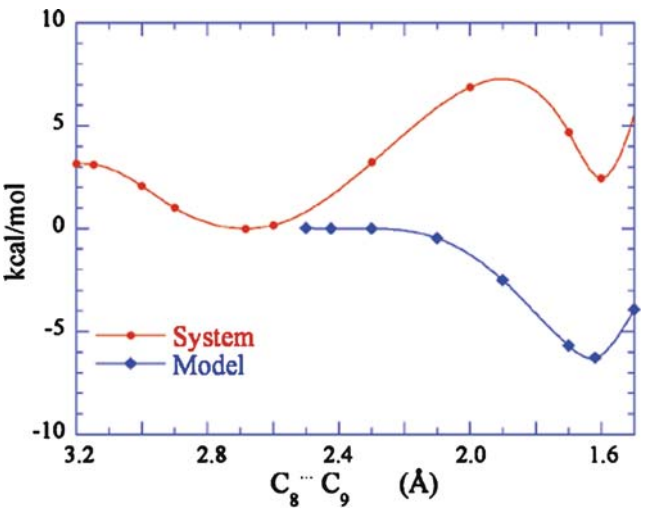


Fig. 6 B3LYP/6-31G* potential energy profiles along the $C_8\cdots C_9$ approaching path in the presence of Rh^+ for: (solid diamonds) the model aldehyde (without acetyl substituent), and (solid circles) the real system

[1] The 3-acetyl-substituted arrangements corresponding to **a1**, **a2**, **a3** are named **A1**, **A2**, **A3**. The acetyl group, unless otherwise specified, is invariably oriented with the CO group pointing as in Fig. 2.

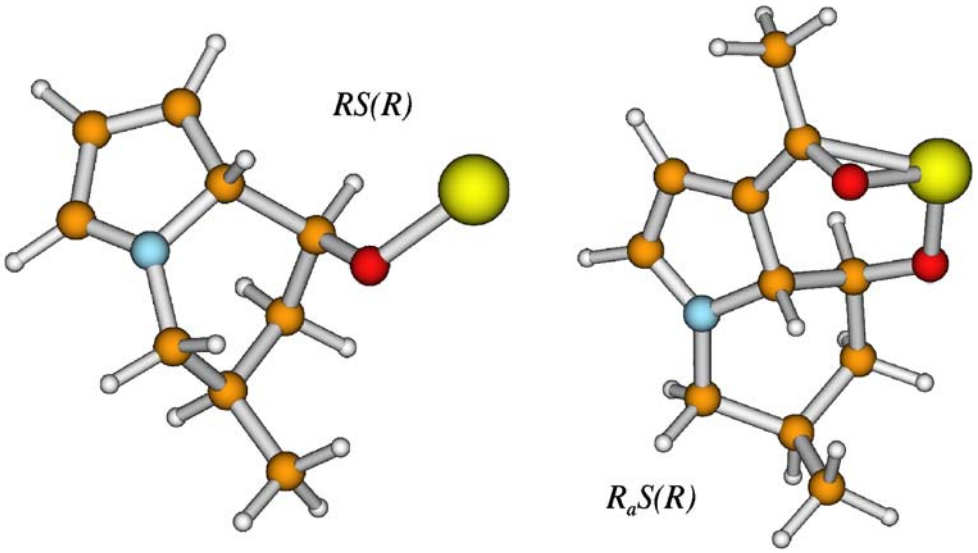


Fig. 7 B3LYP/6-31G* minimum energy structures for the model system along the profiles in Fig. 6

RS or *RR* tetrahydroindolizine, depending on the case, is formed together with hydrofluoric acid.

When the hydronium ion is used as a catalyst, neither arrangement of 3-methyl-4-(3-acetylpyrrol-1-yl)butanal produces annulation. **A1** and **A3** spontaneously cyclize to produce 1-acetyl-6-methyl-8-hydroxy-5,6,7,8,9-pentahydroindolizine only if the acetyl moiety points its carbonyl group away from the aldehyde carbonyl O. Otherwise, the hydronium ion prefers to make a hydrogen bond and eventually to protonate the acetyl carbonyl group, thus preventing the annulation reaction. Therefore, H_3O^+ cannot be considered a viable model for the catalyst. In contrast, the model system invariably cyclizes due to the absence of the acetyl group: $\mathbf{a1H_3O^+}$ gives the *RS(S)* diastereomer, while both $\mathbf{a2H_3O^+}$ and $\mathbf{a3H_3O^+}$ produce the same *RR(R)* one, as reported in Table 5. The driving force that makes also $\mathbf{a2H_3O^+}$ cyclize is the H-bonding interaction between H_3O^+ and the pyrrole ring π density.

In summary, at this computational level, annulation spontaneously occurs in the presence of H^+ only when starting from favorable conformations (**1** and **3**) of either the acetyl substituted compound or the model one. In addition, the only difference between diastereoisomers, at this point, is the lower stability of the *RS* intermediates. But the main questions remain: which is the reaction catalyst? Does Rh participate in the reaction? And, should the answer be positive, might a simple Rh(I) model help elucidate the reaction mechanism?

Reaction mechanism in the presence of Rh^+

In order to clarify the matter, the tentative species taken into account was Rh^+, because Rh(I) complexes are too computationally demanding, for number of basis functions and, mainly, number of degrees of freedom, to allow a systematic search of the various attack positions. Small variations in CO bond distances and mutual orientations, in fact, can produce significant differences in arrangement and stability among very similar (under all the other respects) substrate structures due to the Berry pseudorotation mechanism [13]. Of course, a bare Rh^+, without accompanying ligands, is chemically highly unlikely to be present in the reaction vessel, but its use is more appropriate than that of H^+, because it is expected to be less reactive.

A few different attack possibilities have been explored, starting with the insertion of the Rh cation into the C_9–H bond, since there are several examples in the literature of Rh-catalyzed C–H bond activations [38–42]. An example is displayed in Fig. S3 of ESM.

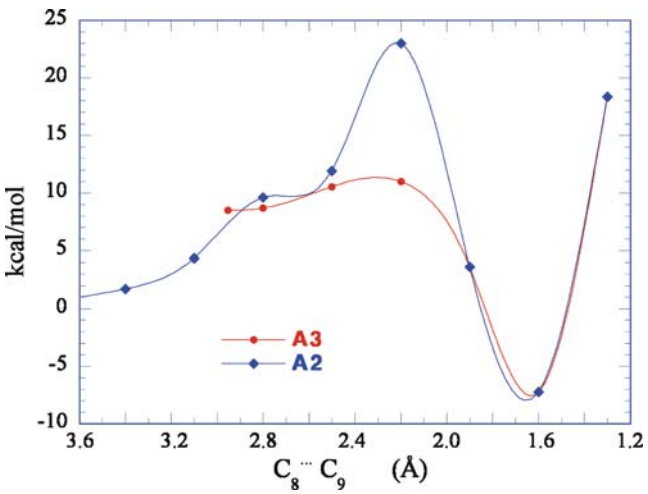


Fig. 8 B3LYP/6-31G* potential energy profiles along the $C_8 \cdots C_9$ approaching path in the presence of Rh^+ starting from **A2** (solid diamonds), and **A3** (solid circles)

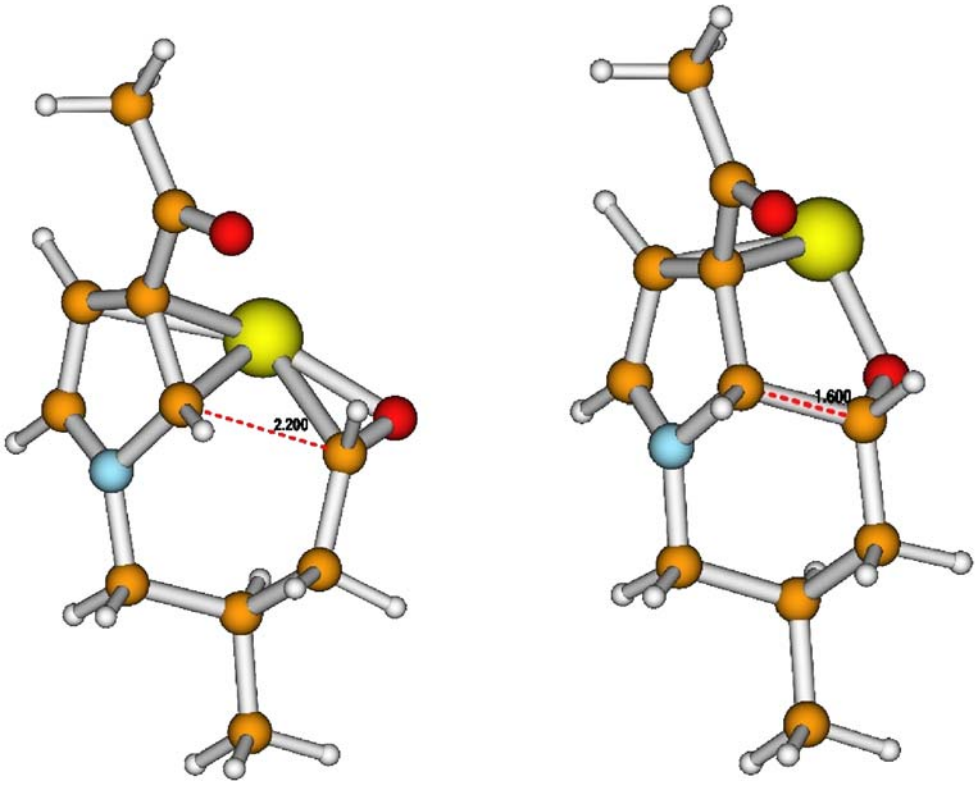


Fig. 9 B3LYP/6-31G* minimum energy structures along the **A2** profile in Fig. 8 at $C_8\cdots C_9$=2.2 Å (left) and 1.6 Å (right)

Since even starting from an ample variety of Rh^+ attack positions the model aldehyde does not spontaneously cyclize, the $C_8\cdots C_9$ approaching path has been studied, using either the model system, i.e., without the acetyl group, or the actual compound, in the presence of Rh^+, located in the vicinity of one of the aldehyde O lone pairs. The profiles are displayed in Fig. 6.

When the model system is used, **a1** produces the *RS(R)* structure at the left hand side in Fig. 7, with a significant stabilization with respect to the starting complex, despite the distorted six-atom ring (the methyl group is equatorial relative to C_5N, axial relative to C_7C_8). Conversely, when using the acetyl substituted pyrrole ring, 6*R* soon becomes a $6R_a$ structure and annulation occurs with a much less favorable profile to give $R_aS(R)$ (circles), with the six-atom ring in a boat arrangement (right hand side in Fig. 7). Interestingly, Rh^+ is bridged between the oxygens.

When **A2** and **A3** are employed as starting structures, they produce the *RR(R)* diastereomer, as expected. The relevant potential energy profiles are plotted in Fig. 8.

Along the path starting from **A2**, Rh^+ is initially located on the pyrrole ring mid-point. Despite being unconstrained, Rh^+ remains close to its original position with a sharp energy increase. When the $C_8\cdots C_9$ distance is below 2 Å, unique structures can be obtained regardless the initial arrangement, i.e., the two curves coincide to the same minimum, shown at the right hand side in Fig. 9. The addition of a counterion should then lead to the incipient 1-acetyl-6-methyl-8-hydroxy-5,6,7,8-tetrahydroindolizine, as stated above when the minimization was carried out in the presence of F^-. The release of Rh^+ and the proton migration to form the final product, however, have not been investigated.

The PES in the presence of Rh^+ for the paths producing the *RR* and *RS* configurations of the acetyl-substituted compound, using the $C_8\cdots C_9$ and $Rh\cdots N$ separations as leading parameters, have been taken into account, instead, to shed some light on the reaction energetics. The remarkably rough and hilly PES producing *RR(R)*, displayed in Fig. 10, features three adjacent local minima for $C_8\cdots C_9\leq 4.1$ Å, that do not coincide with any of the minima relevant to the isolated compound. Rather, in the region corresponding to **A3** ($C_8\cdots C_9 \approx 3.25$ Å), for the complex with Rh^+ there is a saddle point that leads to a shallow local minimum. Then, after surmounting a small barrier, the reaction coordinate steeply descends to give pentahydroindolizine (a structure very close to that shown at the right hand side of Fig. 9).

The profile along the reaction coordinate is plotted in Fig. 11 (diamonds), as compared to that obtained for the *RS(R)* PES (circles) computed using an *RS(R)* structure model built[2] starting from the $R_aS(R)$ one in Fig. 7, with Rh^+ bifurcated between the oxygens. Geometry optimization

[2] The six-atom ring was put in a half-chair conformation with the methyl group in equatorial position.

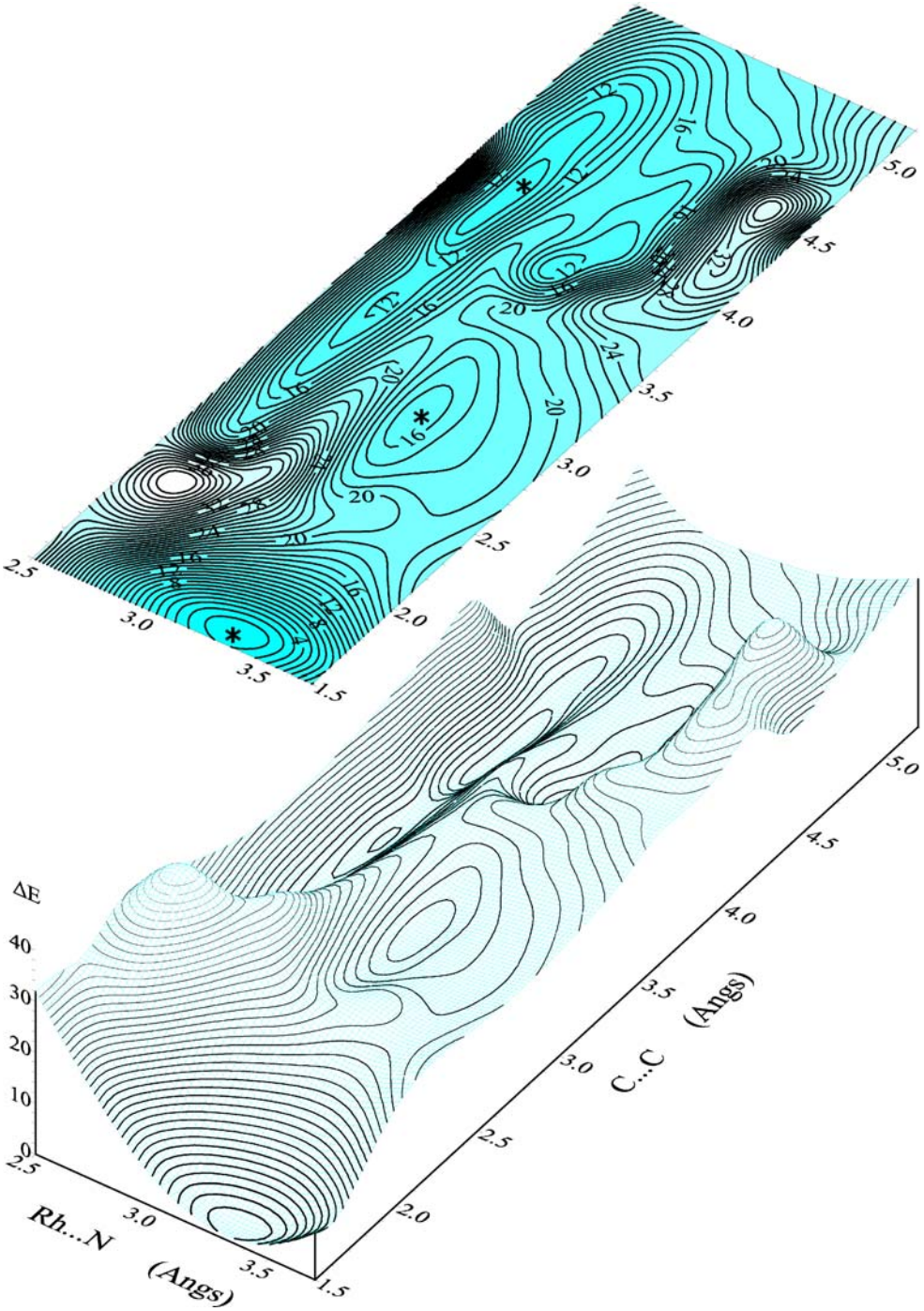


Fig. 10 B3LYP/6-31G* potential energy surface for the *RR(R)* adduct of Rh^+ to the acetyl substituted compound with the $C_8\cdots C_9$ and $Rh\cdots N$ separations as leading parameters (isopotential curves spaced by 1 kcal mol^{-1})

starting from this structure produced a much more stable adduct, although still 7.2 kcal mol^{-1} higher in energy than the *RR(R)* one. The barrier itself decreased from 7.3 to 6.6 kcal mol^{-1} for $R_aS(R)$ and *RS(R)*, respectively, but remained about twice as high as the *RR(R)* one (3.5 kcal mol^{-1}). The *RR(R)* and *RS(R)* minimum energy structures are shown in Fig. 12.

The *RS(R)* PES computed starting from the aforementioned model-built arrangement is displayed in Fig. 13. Interestingly, for large $C_8\cdots C_9$ separations the Rh^+ approach to the aldehyde is very favorable and produces the lowest energy adduct ever encountered (−742.674491 E_h), more stable by 13.8 kcal mol^{-1} than the best *RR(R)* one. For decreasing $C_8\cdots C_9$ separations (3.0 to 1.5 Å) the reaction path on the *RS(R)* map lies on the *RS(R)* profile in Fig. 11. Therefore, the *RS* diastereomer has to climb a steep uphill path (~24 kcal mol^{-1}) in order to reach the local minimum for $3.0 > C_8\cdots C_9 > 2.5$ Å before surmounting the 6.6 kcal mol^{-1} barrier, shown in Figs. 11 and 13, eventually leading to the pentahydroindolizine.

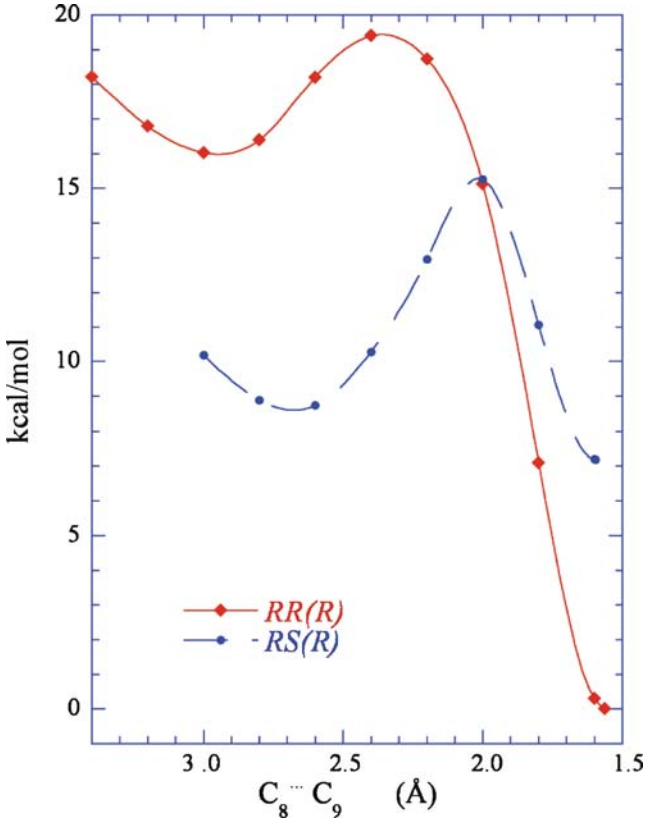


Fig. 11 B3LYP/6-31G* minimum energy profiles for the reaction paths leading to the *RR(R)* and *RS(R)* adducts of Rh^+ to the acetyl substituted compound (see legend)

It is worth noting that the *RR(R)* profile derived from the map in Fig. 10 is similar to the one starting from **A3** in Fig. 8, while the *RS(R)* one is more favorable than the corresponding profile shown in Fig. 6, especially for short $C_8\cdots C_9$ separations.

In an attempt to describe the trend of the whole *RS* profile, an incipient *RS(R)* structure, reached during the grid calculation for $C_8\cdots C_9$=3.4 Å and Rh···N=3.8 Å, was optimized relaxing the Rh···N separation. A structure nearly as stable as the lowest energy one, with Rh···N= 3.08 Å and $Rh\cdots O_{ac}$ / $Rh\cdots O_9$ equal to 2.05 / 2.09 Å, respectively, was obtained. The steep reaction profile obtained starting from that arrangement is shown in Fig. S4 of ESM.

From a perusal of the PES (Fig. 13), it appears evident that the steep reaction profile in Fig. S4 corresponds to the less favorable path at short Rh···N separations. Therefore it is compulsory to compute the potential energy surfaces not to obtain misleading pictures of the interactions. Nevertheless, even though the minimum energy reaction path is not as steep as that shown in Fig. S4, the long-range-complex stability produces a barrier that is too large for the reaction to occur. From the computed barriers both *RS(R)* arrangements are much less favorable than the *RR(R)* one, supporting a large predominance of the *RR* diastereomer. Of course, since an exhaustive surface scan along several other approaching paths cannot be performed, it is difficult

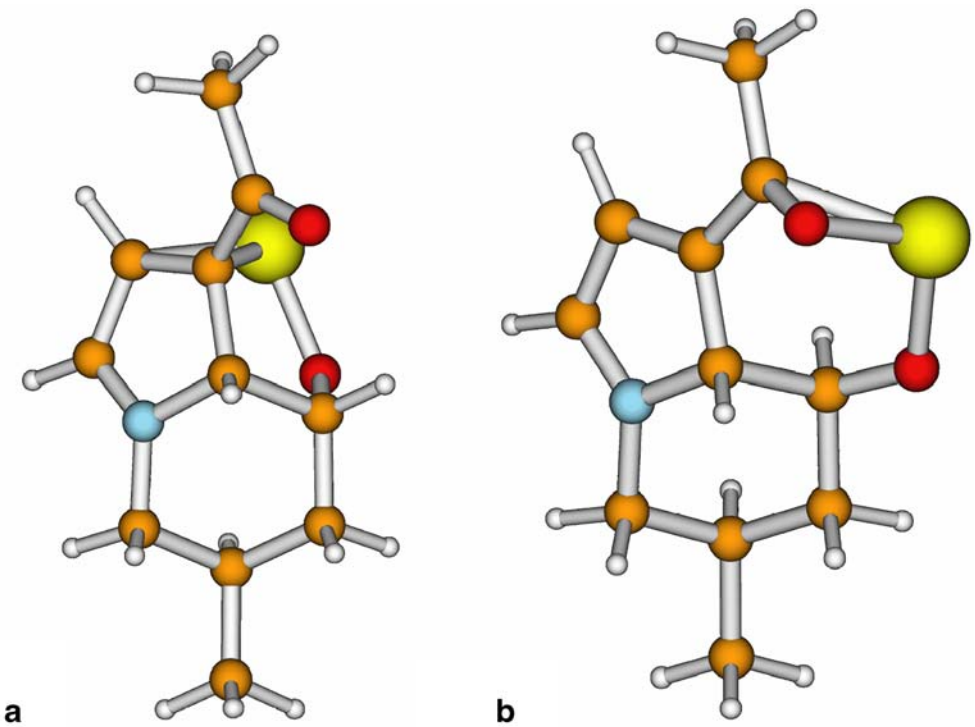


Fig. 12 B3LYP/6-31G* minimum energy structures along the reaction paths in Fig. 11 for (**a**) the *RR(R)* and (**b**) the *RS(R)* adducts of Rh^+ to the acetyl substituted compound

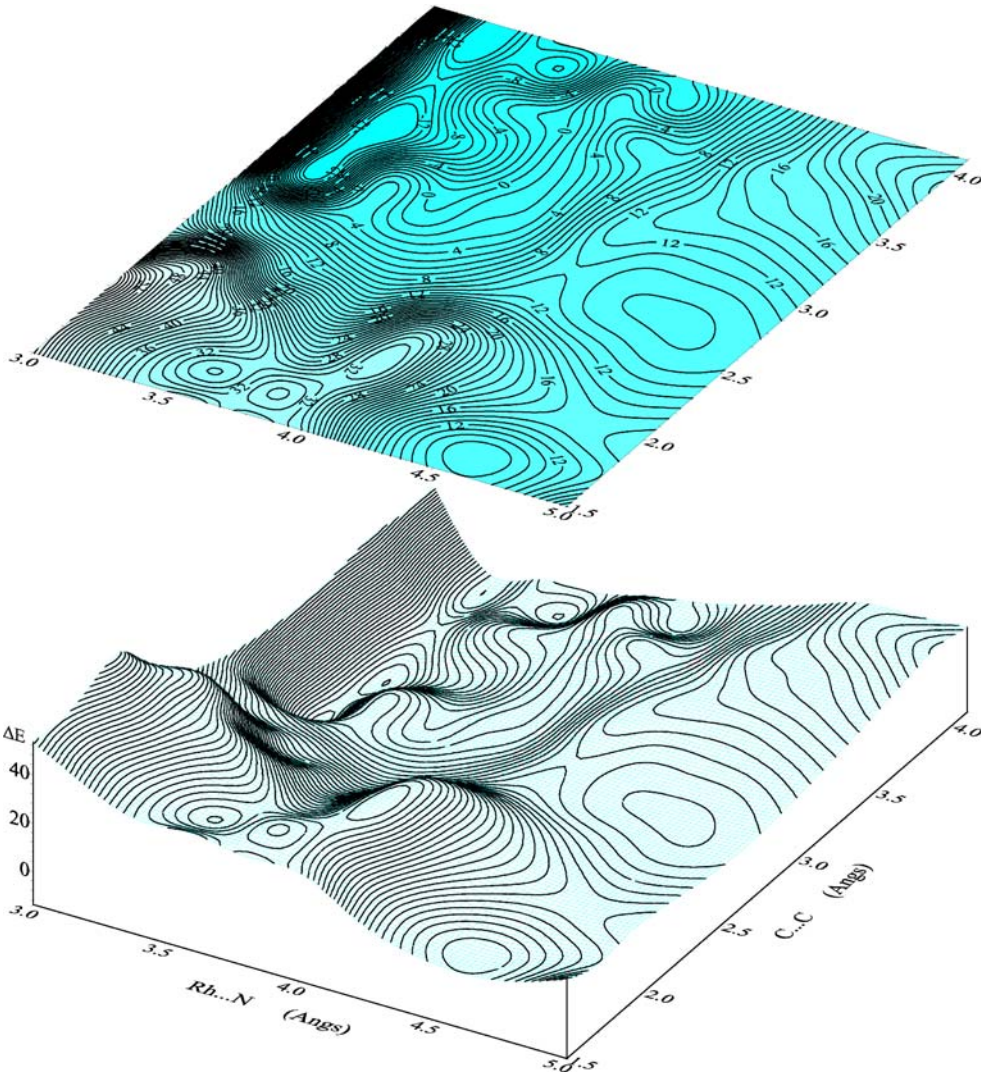


Fig. 13 B3LYP/6-31G* potential energy surface for the *RS(R)* adduct of Rh^+ to the acetyl substituted compound with the $C_8\cdots C_9$ and $Rh\cdots N$ separations as leading parameters (isopotential curves spaced by 1 kcal mol^{-1})

to state which is the real energy gap. However, this result, largely consistent with experimental data, suggests an explanation for the observed diastereoselectivity of the subsequent reaction, confirming as well that some Rh(I) species should be involved in the reaction.

Therefore, some starting arrangements have been considered in the presence of either H-$Rh(CO)_3$ or $[Rh(CO)_3]^+$ indeed, exploiting the knowledge gained with simpler models. In most cases, in fact, a complex is formed with the Rh carbonyl group further apart from the aldehyde, in a square planar arrangement on its anti lone pair. This occurs both at the B3LYP and B3P86 levels. The investigation of the $C_8\cdots C_9$ approaching paths leading to the *RR(R)* and *RS (R)* arrangements produced structures consistent with the reaction under scrutiny only when the $[Rh(CO)_3]^+$ species was substituted to Rh^+ onto the inner aldehyde O lone pair (syn). The resulting profiles, plotted in Fig. 14 with respect to the isolated partners taken as zero, show significantly different barriers for *RR* and *RS* (2.8 and 7.8 kcal mol^{-1}, respectively), in agreement with the experiment that gives the same chirality on both stereogenic centers as a result of this domino reaction. Interestingly, the rationale behind this fact is the *RS* incipient complex much larger stability than the *RR* one, in analogy to what suggested also by the calculations carried out in the presence of Rh^+.

Consideration of the PES and of the subsequent reaction steps even with this kind of Rh complexes (small as compared to the phosphine-modified ones) is however not affordable with our presently available computational resources.

Concluding remarks

The reaction outcome obtained from hydroformylation of a prochiral substrate (2-methyl-3-(3-acetylpyrrol-1-yl)prop-1-ene), followed by cyclization to 1-acetyl-6*R(S)*-methyl-8*R (S)*-hydroxy-5,6,7,8-tetrahydroindolizine, has been investi-

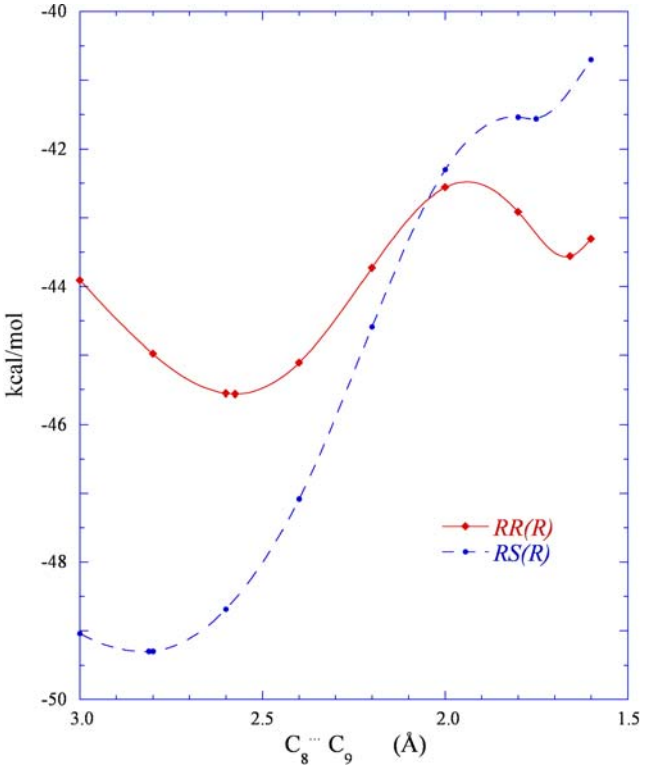


Fig. 14 B3LYP/6-31G* minimum energy profiles (with respect to the isolated partners taken as zero) for the reaction paths leading to the *RR (R)* and *RS(R)* adducts of $[Rh(CO)_3]^+$ to the acetyl substituted compound (see legend)

gated with a particular interest in assessing (a) the computationally predicted hydroformylation regioselectivity and (b) the origin of the observed diastereoselectivity of the subsequent annulation in the domino process. The obtainment of the linear aldehyde as the sole hydroformylation product had been actually attributed in the literature not to the alkyl rhodium formation step, but to the CO insertion into the Rh–C_2 bond that was considered disfavored with respect to that into the Rh–C_1 bond. The computational prediction at the alkyl-rhodium TS level, however, supports an early large preference for the linear aldehyde (B:L=12:88). As far as point (b) is concerned, the *RR* and *RS* cyclic diastereomers coming from the annulation process are almost exactly isoenergetic. Of course, in order to determine the reaction diastereoselectivity, the pathways leading to them are to be examined. This cannot be done in the absence of a suitable catalyst because of the very high barriers to be surmounted in that case. It is nonetheless unaffordable the use of Rh(I) complexes, albeit not modified with phosphines, due to the number of basis functions, degrees of freedom and, primarily, attack orientations and positions. The use of H^+ as a catalyst produced spontaneous annulation, in agreement with experimental observations even though the reaction proceeds further to give the dehydrated compounds. The tentative evaluation of the reaction barriers along the pathways leading either to *RR* or *RS* tetrahydroindolizine in the presence of Rh^+ yielded a barrier for *RS* much higher than that for *RR*, due to the great stability of the *RS* adduct featuring the cation bifurcated with respect to the aldehyde and carbonyl oxygens. The *RR* adduct cannot assume such an extremely favorable arrangement because of its aldehyde group orientation. Among the bulkier Rh(I) catalysts considered, only $[Rh(CO)_3]^+$ produced structures consistent with this domino reaction along the $C_8 \cdots C_9$ approaching paths leading to *RR* and *RS*, with the *RS* barrier almost thrice as high as the *RR* one. Thus, a convincing explanation of the reaction outcome has been put forward, exploiting the knowledge gained using a bare Rh^+ as the active species in the annulation process.

Acknowledgments We thank Professor Raffaello Lazzaroni for bringing to our attention this domino reaction.
Additional structures and plots as well as further details on this investigation, as specified in the text, can be found in the Supplementary Material.

References

1. Anh NT, Eisenstein O (1977) Nouv J Chim 1:61–70, (references therein)
2. Eliel EL, Wilen SH, Mander LN (1994) Stereochemistry of organic compounds. Wiley, New York, pp 875–886
3. Houk KN (2000) Theor Chem Acc 103:330–331
4. Torrent M, Solà M, Frenking G (2000) Chem Rev 100:439–493 (references therein)
5. Koga N, Jin SQ, Morokuma K (1988) J Am Chem Soc 110:3417–3425
6. Matsubara T, Koga N, Ding Y, Musaev DG, Morokuma K (1997) Organometallics 16:1065–1078
7. Gleich D, Hutter J (2004) Chem Eur J 10:2435–2444
8. Rocha WR, De Almeida WB (2000) Int J Quantum Chem 78:42–51
9. Gleich D, Schmid R, Herrmann WA (1998) Organometallics 17:4828–4834
10. Carbó JJ, Maseras F, Bo C, van Leeuwen PWNM (2001) J Am Chem Soc 123:7630–7637
11. Alagona G, Ghio C, Lazzaroni R, Settambolo R (2001) Organometallics 20:5394–5404
12. Alagona G, Ghio C, Lazzaroni R, Settambolo R (2004) Inorg Chim Acta 357:2980–2988
13. Alagona G, Ghio C (2005) J Organomet Chem 690:2339–2350
14. Settambolo R, Rocchiccioli S, Lazzaroni R, Alagona G (2006) Lett Org Chem 3:10–12
15. Matsui Y, Orchin M (1983) J Organomet Chem 246:57–60
16. Amer I, Alper H (1990) J Am Chem Soc 112:3674–3676
17. Botteghi C, Cazzolato L, Marchetti M, Paganelli S (1995) J Org Chem 60:6612–6615
18. Lazzaroni R, Settambolo R, Uccello-Barretta G, Caiazzo A, Scamuzzi S (1999) J Mol Cat A Chemical 143:123–130
19. Lazzaroni R, Settambolo R, Caiazzo A, Pontorno L (2000) J Organomet Chem 601:320–323
20. Settambolo R, Caiazzo A, Lazzaroni R (2001) Tetrahedron Lett 42:4045–4048
21. Gaussian 03, Revision C.02: Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery Jr JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi

J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian H. P, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian Inc. Wallingford CT
22. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785–789
23. Becke AD (1993) J Chem Phys 98:5648–5652
24. Hehre WJ, Radom L, Schleyer PvR, Pople JA (1986) Ab initio molecular orbital theory. Wiley, New York
25. Hay PJ, Wadt WR (1985) J Chem Phys 82:270–283
26. Perdew JP (1986) Phys Rev B 33:8822–8824
27. Alagona G, Ghio C, work in progress
28. Schultz NE, Zhao Y, Truhlar DG (2005) J Phys Chem A 109:11127–11143
29. London F (1937) J Phys Radium 8:397–409
30. McWeeny R (1962) Phys Rev 126:1028–1034
31. Ditchfield R (1974) Mol Phys 27:789–807
32. Dodds JL, McWeeny R, Sadlej AJ (1980) Mol Phys 41:1419–1430
33. Wolinski K, Hilton JF, Pulay P (1990) J Am Chem Soc 112:8251–8260
34. Rocchiccioli S (2006) Substrate induced Diastereoselectivity in Rhodium Catalyzed Hydroformylation, PhD Thesis, University of Pisa
35. Settambolo R, Caiazzo A, Lazzaroni R (2001) Tetrahedron Lett 42:4045–4048
36. Settambolo R, Miniati S, Lazzaroni R (2003) Synth Commun 33:2953–2961
37. Settambolo R, Guazzelli G, Mandoli A, Lazzaroni R (2004) Tetrahedron: Asymmetry 15:1821–1823
38. Tan KL, Bergman RG, Ellman JA (2002) J Am Chem Soc 124:3202–3203
39. Tan KL, Bergman RG, Ellman JA (2002) J Am Chem Soc 124:13964–13965
40. Wiedemann SH, Lewis JC, Ellman JA, Bergman RG (2006) J Am Chem Soc 128:2452–2462 (references therein)
41. Rybtchinski B, Oevers S, Montag M, Vigalok A, Rozenberg H, Martin JML, Milstein D (2001) J Am Chem Soc 123:9064–9077
42. Rybtchinski B, Cohen R, Ben-David Y, Martin JML, Milstein D (2003) J Am Chem Soc 125:11041–11050 (references therein)