Reaction energy profile and fragment attributed molecular system energy change (FAMSEC)-based protocol designed to uncover reaction mechanism: A case study of the proline catalysed aldol reaction

Ignacy Cukrowski*[a], George Dhimba[a] and Darren L. Riley[a]

[a] Prof. I. Cukrowski, G. Dhimba, Dr D. L. Riley
Department of Chemistry, Faculty of Natural and Agricultural Sciences, University of Pretoria, Lynnwood Road, Pretoria 0002 (South Africa)

[*] E-mail: ignacy.cukrowski@up.ac.za

Abstract

A REP-FAMSEC (reaction energy profile-fragment attributed molecular system energy change) protocol designed to explain each consecutive energy change along the reaction pathway is reported. It mainly explores interactions between meaningful polyatomic fragments of a molecular system and, by quantifying energetic contributions, pin-points fragments (atoms) leading to or opposing a chemical change. Its usefulness is tested, as a case study, on the proline catalysed aldol reaction for which a number of mechanisms is being debated for over four decades. Relative stability of S-proline conformers, their catalytic (in)activity and superior affinity of the higher energy conformer to acetone is fully explained on an atomic and molecular fragment levels, but still appealing to general chemist knowledge. We found that (i) contrary to generally accepted view, the CN-bond formation cannot be explained by the $N^{\delta-}$,$C^{\delta+}$ atom pair but rather by O-atom of acetone and its strongest inter-molecular attractive interactions with N-atom as well as C-atom of the COO group of proline (at this initial stage the lower energy conformer of proline is eliminated) and (ii) the following 'first' H-transfer from N to O atoms of proline moiety is nearly energy-free even though initially H-atom interacts three times stronger with N- than O-atom; a full explanation of this phenomenon is provided.

Keywords

reaction mechanism, ab initio calculations; proline catalysed aldol reaction;
computational chemistry; FAMSEC.

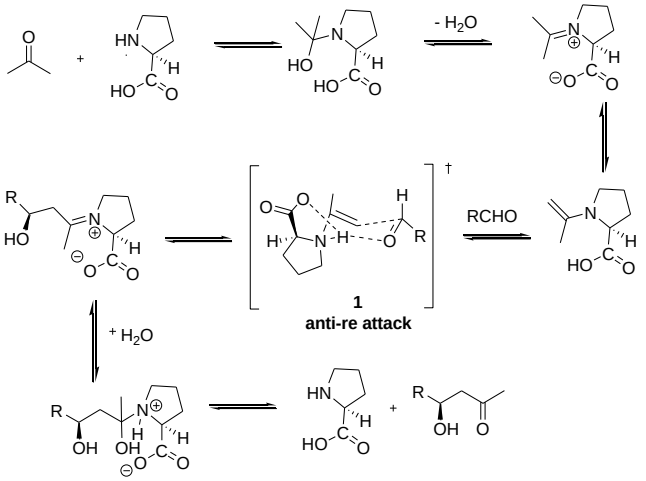
1. Introduction

The classic understanding of how bond formation and breaking occurs in organic reactions is intrinsically linked to two principles, the 3D-structure of the molecule as approximated by the linear combination of atomic orbitals and the electronic structure of the molecule.[1] Typically, in considering a multi-step chemical process a textbook approach will involve the identification of atoms of reactants with most negative ($A^{\delta-}$) and positive ($B^{\delta+}$) partial charge constituting a 2-atom fragment $\mathcal{G}$ = {A,B} of a molecular system with the expectation being that the large difference in electronegativity $\Delta\delta$(A,B), will lead to the formation of a new bond.[2–4] Although the movement of electrons from nucleophilic to electrophilic sites is successfully used to predict the formation (breaking) of covalent bonds in many instances,[5] this approach does not provide a deep understanding of processes taking place and does not guarantee that $\Delta\delta$(A,B) is the primary driving force. Moreover, in many instances the failure of reactions with the required structural and electronic features is challenging to predict (explain) by use of this approximate model.

Typically, by combining general knowledge, available experimental data and chemical intuition, a working hypothesis is put forward and used in drawing a reaction mechanism.[6,7] Clearly, it would be highly beneficial if such hypotheses could be supported (or otherwise) by computational modelling of at least most critical steps. This is then not surprising that computational/theoretical modelling of reaction mechanism and chemical reactivity has gained a lot of attention for decades and indeed it is still a very active area of research.[8–29] (and references therein). There are two general approaches used in gaining chemical insight from quantum chemical calculations, namely, making use of (i) orbitals (e.g., MO,[8–11] VB,[12–15] and NBO[16-18] methods) and (ii) topology of electron density using e.g., properties at BCPs and RCPs,[19-21] topology of the Laplacian,[22] the electron localization function (ELF),[23] the bonding evolution theory (BET),[24,25] the molecular electron density theory (MEDT)[26,27] and concerted DFT-conceptual DFT-QTAIM approach.[28,29] Although classical orbital-based approach was successful

in explaining many reaction mechanisms, the contemporary density-based approaches provide deeper insights that sometimes either do not support previous orbital-based models or are in direct conflict with them[26] (and reference therein).

The use of proline as an organic catalyst was first reported in the 1970's,[30,31] and the reagent can be used in either the L- or S-forms allowing enantioselective transformations most notably aldol condensations[32] and mechanistically related Michael, Robinson and Mannich reactions.[33–35] The mechanism of the proline catalysed aldol reaction (Scheme 1) is still the target of both theoretical[33,36–38] and experimental investigations[39–41] and to date several contrasting mechanisms have been proposed.[30,36-38,42–46] Most reports concentrated on the latter stages of the reaction mechanism[47–51] (they are at variance with each other) with less attention being given to initial proline–acetone adduct formation.[46] To date several conformers of proline have been reported but little to no attention has been paid to their role when the entire catalytic process is considered. That being said, there is a single recent report claiming that the active catalyst is the higher energy conformer,[37] with the lowest energy conformer being eliminated at the first H-transfer step.



Scheme 1. Proposed[32] mechanism of proline catalysed aldol reaction.

Clearly, prior to considering the entire process, it is of paramount importance to (dis)prove the catalytic form of *S*-proline using computational methods. Furthermore, in such modelling, when performed on an atomic and molecular fragment levels, many atoms of a molecular system should

be considered as they might play subtle yet critical mechanistic roles. This, in turn, should provide a more in-depth mechanistic insight into how reactions proceed and why, *e.g.*, 3D, substituent and electronic changes sometimes lead to dramatic variation in reactivity.

Hence, with a focus on the initial steps of the proline catalysed aldol reaction, we decided to go beyond the classical 2-atom approach and a standard analysis of energy profiles generated from computational studies. To explain every incremental step along the reaction pathway, we have developed a protocol that makes use of the energy terms computed within the interacting quantum atoms (IQA)[52,53] framework and a general concept of the fragment attributed molecular system energy change, FAMSEC,[54–56] method (computational details and coordinates for all structures are included in PART S1 of the ESI). The main (but not exclusive) focus of the protocol is on interaction energies and their changes, ΔE_{int}, computed for each incremental step (with a specific increase/decrease in the electronic energy, ΔE) along the reaction pathway. The protocol is highly flexible as a chemist can (i) select any size of a fragment, from a single atom up to entire molecule, (ii) investigate inter- and intra-fragment interactions, (iii) analyse variations in long- and short-distance interactions, or (iv) monitor a process of covalent bond's breaking/formation through $E_{\text{int}}^{\text{A,B}}$ computed for atoms A and B of interest. The wealth of data collected can then be used to rationalize computed ΔE values and identify fragments that either lead to or oppose the chemical change most.

2. Basic and relevant to this work concepts

2.1. Interacting quantum atoms method (IQA)

The IQA method is an energy partitioning scheme of a molecular system (e.g., a single molecule, adduct, or interacting molecules at a transition state) that recovers properties of atoms, such as (i) their energies confined within each atom's specific volume it occupies in a molecule and (ii) numerous di- and poly-atomic interactions as well as interactions between nuclei and electrons. Importantly, a molecular system is being considered as made of atoms that fill in the entire space

occupied by the system. From this follows that there are no voids in 3D molecular space or regions of overlapping atoms. Hence, each atom has well-defined interatomic boundaries and, as consequence, its own energy that depends mainly on the kind of an atom and somewhat (to a much lesser degree) on its placement in a molecule. From this follows that the computed electronic (or *ab initio*) molecular energy E can be recovered in the IQA scheme by summing up energies of each atom A, called total (or additive, $E_{\mathrm{add}}^{\mathrm{A}}$) atomic energies – Eq. 1,

$$E = E_{\mathrm{IQA}} = \sum_{\mathrm{A}} E_{\mathrm{add}}^{\mathrm{A}} \qquad (1)$$

In accord with a chemical intuition, all IQA atoms of a molecule are involved in interactions (either attractive or repulsive) with associated interaction energies. Importantly, regardless whether a classical chemist see atoms as covalently or otherwise (non)bonded, they all are treated on equal footing. From this follows that the total atomic energy must consist of two major components, namely the energy of an atom itself (often referred to as a self-atomic energy, $E_{\mathrm{self}}^{\mathrm{A}}$) and the sum of diatomic interaction energies, $E_{\mathrm{int}}^{\mathrm{A,B}}$, atom A is experiencing with each other atom B of a molecule. Note that to make an energy of an atom additive (in order to recover energy E of a system), the total interaction energy involving all possible unique atom pairs {A,B} is halved – Eq. 2,

$$E_{\mathrm{add}}^{\mathrm{A}} = E_{\mathrm{self}}^{\mathrm{A}} + 0.5\sum_{\mathrm{X}\neq\mathrm{A}} E_{\mathrm{int}}^{\mathrm{A,B}} \qquad (2)$$

From the above it follows that by summing up all self-atomic energies one can compute the total self-molecular energy, Eq. 3

$$E_{\mathrm{self}}^{\mathrm{Tot}} = \sum_{\mathrm{A}} E_{\mathrm{self}}^{\mathrm{A}}\,, \qquad (3)$$

and by summing up all unique diatomic interaction energies between atoms A and B one obtains the total interaction energy of a molecular system, Eq. 4

$$E_{\mathrm{int}}^{\mathrm{Tot}} = 0.5\sum_{\mathrm{A}}\sum_{\mathrm{B}\neq\mathrm{A}} E_{\mathrm{int}}^{\mathrm{A,B}}\,. \qquad (4)$$

Note that the total self-molecular energy and total interaction energy of a molecular system recovers the electronic energy of that system, Eq. 5,

$$E = E_{\mathrm{IQA}} = E_{\mathrm{self}}^{\mathrm{Tot}} + E_{\mathrm{int}}^{\mathrm{Tot}} . \quad (5)$$

IQA is a powerful tool that partitions self-atomic and interaction energies into many important components that are extremely useful in theoretical studies, e.g., in understanding a nature and strength of chemical bonding. For instance, whereas $E_{\mathrm{int}}^{\mathrm{A,B}}$ quantifies strength of interaction or chemical bonding of any nature between two atoms A and B, by partitioning this energy term - Eq. 6,

$$E_{\mathrm{int}}^{\mathrm{A,B}} = V_{\mathrm{XC}}^{\mathrm{A,B}} + V_{\mathrm{cl}}^{\mathrm{A,B}} \quad (6)$$

to the exchange-correlation ($V_{\mathrm{XC}}^{\mathrm{A,B}}$, the interaction energy due to purely quantum effects) and classical ($V_{\mathrm{cl}}^{\mathrm{A,B}}$, the classical electrostatic Coulomb interaction) components a chemists gains an instant quantitative description of bonding and its nature in terms of the degree of covalent (XC-term) or electrostatic contributions. To learn more about other IQA-defined energy components (we will not make use of them in this work) an interested reader is referred to relevant literature on IQA and its applications.

2.2. Fragment attributed molecular system energy change (FAMSEC)

It is trivial to state that it would be of great importance and assistance to a chemist if one was able to understand and quantify changes taking place throughout a molecular system when it is exposed to a new environment. A change of environment can be seen as a broad spectrum of chemically relevant events, such as:

(a) Conformational change (to understand relative stability of conformers and role played by intramolecular interactions as well steric clashes).

(b) Formation of adducts and clusters (what drives them to form, molecular fragments interacting strongest in adducts).

(c) Reaction pathway from reactants through a transition state to products (this covers inter- and intramolecular interactions, bond breaking and new bond formation; all needed to explain reaction mechanism and preferential substitution sites).

(d) Formation of metal complexes and their relative stability (*e.g.* in terms of formation of 5- and 6-membered coordination rings and their influence on strength of coordination bonds), and many more.

Moreover, it would be highly beneficial to explain these changes in terms of classical thinking as it should be useful in designing a chemical process leading to a desired output (product). Clearly, to gain an insight on a complex chemical process, one must compare the properties of atoms, chemically meaningful molecular fragments or even entire molecules between two states of a molecular system, *i.e.*, when it changes form a particular initial state (it can be used as a reference state, *ref*) to a state of a system that is of interest (final state, *fin*). Two major and useful for interpretation approaches are used in FAMSEC; they focus on:

(1) Properties confined to a 3D space occupied by a selected (on purpose) *n*-atom fragment $\mathcal{G}$ of a system and related to it energetic effects when the *ref* $\rightarrow$ *fin* structural transformation or chemical change takes place. This can be seen as focusing on a localised to within a fragment $\mathcal{G}$ event and *loc*-FAMSEC energy term applies

$$\textit{loc}\text{-FAMSEC} = \Delta E_{\text{self}}^{\mathcal{G}} + \Delta E_{\text{int}}^{\mathcal{G}} \ . \qquad (7)$$

The $\Delta E_{\text{self}}^{\mathcal{G}}$ term accounts for self-fragment energy change, i.e., a sum of self-atomic energy changes of atoms constituting a molecular fragment $\mathcal{G}$. The $\Delta E_{\text{int}}^{\mathcal{G}}$ term quantifies the intra-fragment interaction energy change and when $\mathcal{G}$ is made of two atoms it quantifies a diatomic interaction energy change. From this follows that *loc*-FAMSEC might be useful in identifying parts of a molecule that experienced most significant decrease/increase of their energies on a

ref $\rightarrow$ fin environmental change that can be interpreted as being most stabilised/strained, respectively, in *fin* relative to *ref*.

(2) How changes in properties of $\mathcal{G}$ and remaining atoms of a molecule (typically treated as another molecular fragment $\mathcal{H}$) impact on entire molecule when *ref* $\rightarrow$ *fin* occurs and what are energetic consequences in terms of stability of a molecule. This can be seen as a global, on a molecular scale, event and it can be quantified by use of the *mol*-FAMSEC energy term,

mol-FAMSEC = *loc*-FAMSEC + $\Delta E_{\text{int}}^{\mathcal{G},\mathcal{H}}$ (8)

where the $\Delta E_{\text{int}}^{\mathcal{G},\mathcal{H}}$ energy term quantifies the inter-fragment interaction energy change between $\mathcal{G}$ (a fragment of interest) and $\mathcal{H}$ (remaining atoms of a molecular system).

Moreover, when $\Delta E_{\text{int}}^{\mathcal{G},\mathcal{H}} < 0$ then it implies that $\mathcal{G}$ found itself, relative to the *ref* state, in more attractive (stabilizing) molecular environment when in the *fin* state. The interplay between the two components, *loc*-FAMSEC and $\Delta E_{\text{int}}^{\mathcal{G},\mathcal{H}}$, decides whether the molecular fragment $\mathcal{G}$ has added to stability of the *fin* state of a molecular system (then *mol*-FAMSEC < 0) or contributed in a destabilizing manner.

It is important to stress that the *loc*- and *mol*-FAMSEC terms can be computed for all unique, 2-, 3-, …, *n*-atom, fragments. From that one can establish which fragments were most locally (de)stabilized and which ones (de)stabilized a molecule the most, *etc.* This is very useful information in interpreting many chemical phenomena and also puts the energies attributed to a selected fragment on a molecular-scale perspective.

2.3. FAMSEC-based protocol designed for the study of reaction mechanism

Firstly, let us point at several aspects that must be brought to the attention of a classical chemist and we will make use of adduct made of S-proline (lowest energy conformer) and acetone to illustrate points specified below:

1) The E_{self}^{Tot} energy term always contributes most to molecular electronic energy *E* regardless of the level of theory used. The computed at a B3LYP/6-311++G(d,p) level of theory *E*(adduct) is –594.5415085 a.u. that translates to hundreds of thousands of kcal mol^{-1}, namely –373080.4 kcal mol^{-1}, of which 98.32 % comes from the total self-molecular energy, E_{self}^{Tot} (adduct). Very much comparable values apply to the components of adduct; 98.33 and 98.31% of the total energy of S-proline and acetone comes from E_{self}^{Tot}(S-proline) and E_{self}^{Tot} (acetone), respectively. It means that less than 1.7 % of *E* comes from all interactions and this also includes all covalent bonds! These %-fractions are typical in many molecular systems and do not very significantly with the level of theory.

2) When a synthetic process is considered (from reactants, through adduct formation, structural re-arrangements leading to a transition state, formation of intermediates, …, formation of a final product and by-products) rather small changes in molecular system energy ΔE are observed when compared with the total energy of a system. Typically, ΔE does not exceed ±30 kcal mol^{-1} at a single step of a chemical process and this would constitute just 0.008% of this adduct energy.

3) Obviously, the expression $\Delta E = \Delta E_{self}^{Tot} + \Delta E_{int}^{Tot}$ holds at any point along reaction coordinates but the changes in self-atomic energies do not typically exceed ±10 kcal mol^{-1} for an individual atom. For instance, 22 (out of 27) atoms of adduct experienced $\left| \Delta E_{self}^{A} \right| < 2$ kcal mol^{-1} and the largest change found on the adduct formation was +8.3 kcal mol^{-1}.

4) The number of unique atom-pairs {A,B} in any molecular system is $(n\times(n\text{-}1))/2$ where *n* is the number of atoms (e.g., $n = 27$ in the S-proline adduct with acetone). Hence, not only there are many more diatomic interactions than atoms in a molecule (the number of unique diatomic pairs in the adduct is 351) but their interactions can vary extensively; the $\Delta E_{int}^{A,B}$ values are often over an order of magnitude larger when compared with either ΔE_{self}^{A} or ΔE_{self}^{B}. To

illustrate this point, the most significant changes $\Delta E_{\text{int}}^{\text{A,B}}$ were found to be –161.3 and +138.3 kcal mol^{-1} for {O19,C14} and {C18,C14} atom pairs, respectively, on this adduct formation.

The above observations inspired us in designing a protocol where a general concept of FAMSEC, i.e., monitoring changes in selected on purpose energy terms (rather than values themselves) is used to explain ΔE for each consecutive step (hence the overall reaction mechanism) with a main focus on interaction energies as they vary most and can be seen as a driving force for a chemical change. This protocol can be seen as open-ended as one can pursue many strategies in monitoring and explaining a chemical process. To this effect, one can consider all possible 2,3,4-…*n*-atom fragments in order to identify parts of a molecule that play the leading role. Moreover, this protocol is perfectly suited for making best use of chemical intuition and general knowledge in selecting atoms constituting classical functional groups or specific fragments of molecules, or even entire molecules that interact with each other. Clearly, there is no specific protocol to follow as each synthetic route involving different reactants containing specific functionalities might require a unique set of descriptors needed to explain the role played by uniquely selected molecular fragments. However, as a good starting point, *n*-atomic fragments might/should be selected for which most significant change in the intra- and inter-fragment interaction energies were computed between consecutive steps. Furthermore, from a change in e.g., di-atomic interaction energies it is highly informative and useful to identify atoms of these fragments that facilitate or obstruct the progress of reaction most. To guide a chemist in selecting energy terms that might be most appropriate in explaining the computed ΔE values for a particular step along the reaction coordinates, a set of approaches (far from being exhaustive) and what knowledge can be gained from them is included in PART S2 of the ESI.

3. Results and discussion

There is a growing evidence that chemical bonding has a multicentre character not only in the case of classical intramolecular H-bonds but also in the case of typical covalent bonds, such as C–C, with numerous atom contributing to electron density into the inter-nuclear region where bond is thought to be formed.[57–62] Furthermore, diatomic interaction energies computed for all unique atom pairs show that many of them, even when considered as being non-bonded, are indeed involved in very strong interactions. We realised that more fruitful and informative approach should focus on molecular fragments containing atoms involved in most significant inter-fragment and inter-molecular interactions. This concept, is implemented in the present study and applied to each consecutive step identified from computational modelling of a reaction mechanism.

3.1. The origin of relative stability of *S*-proline conformers

Molecular graphs of *S*-proline conformers **1** reported in the literature[37] (Fig. 1, part A) show the same kind of classical H-bond (O16–H17···N13) in **1a** and **1b** (lower and higher energy conformers, LEC and HEC, respectively); in each case a well-defined density bridge (or Bader's bond path) is linking N13 and H17. From MP2 data (E_{ZPVE}, H and G), **1a** is lower in energy by ~ –6.7 kcal mol^{-1}.

Classically, by an eye inspection of structural features, the higher stability of **1a** would be attributed to the presence of the significantly shorter H-bond (d(N13,H17) = 1.78637 Å in **1a**; ~ 0.3388 Å shorter than in **1b**) and/or (in)availability of the lone electron-pair on N13 to form an intramolecular H-bond with H17. This might be the case, but would have to be proven, and this is not an easy (if at all possible) task. Let us start then with the two-atom (classical) approach: we computed $E_{int}^{N13,H17}$ of –132.8 kcal mol^{-1} in **1a** that is stronger, by –33.5 kcal mol^{-1}, than in **1b.** In both conformers, the diatomic $E_{int}^{N13,H17}$ interaction energy is dominated, in accord with classical thinking, by the electrostatic component with $V_{cl}^{N13,H17}$ of –113.8 and –92.5 kcal mol^{-1} in **1a** and

1b, respectively. Importantly, the exchange-correlation energy term, $V_{XC}^{N13,H17}$ = –19.1 kcal mol^{-1}, in **1a** is not only significant (it constitutes 14.3 % of the total interaction energy) but it is also stronger, by –12.2 kcal mol^{-1}, when compared with **1b**. This seems to correlate well with the reviewer's comment on the better exposure of a free electron pair on N13 in **1a**, but it does not explain fully (as will be shown below) a relative stability of the two conformers.

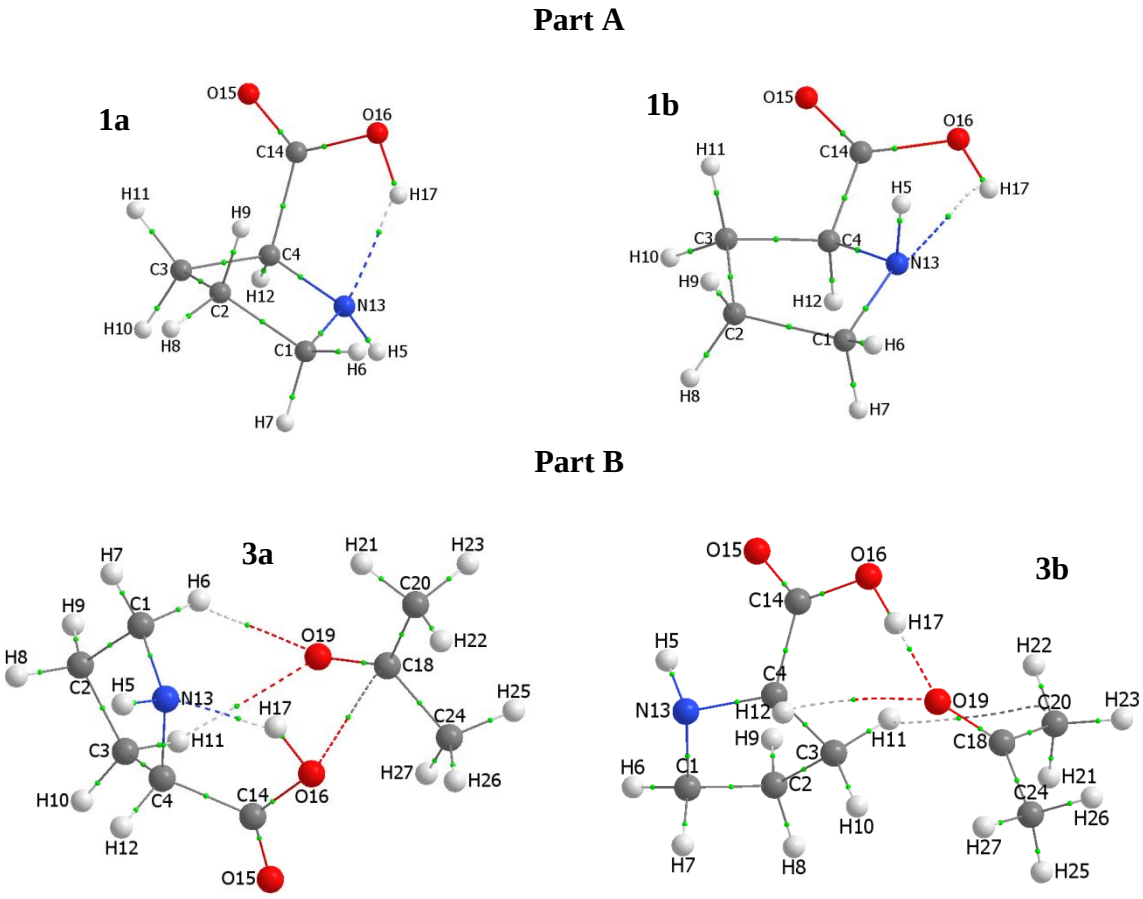


Fig. 1 Molecular graphs of: part A - lower (**1a**) and higher (**1b**) energy conformers of *S*-proline; part B - the global minimum structures of adducts **3a** (**1a** and acetone, **2**) and **3b** (**1b** and **2**).

Hence, instead of focusing on a single interaction, we analysed entire molecules and relevant data are placed in PART S3 of the ESI. Analysis of all 136 unique atom pairs (119 non-covalent interactions and 17 covalent bonds) shows that:

1 Covalent bonds (*Cov*-bonds) are stronger in **1a** by $\Delta E_{int}^{Cov\text{-bonds}}$ = –39.2 kcal mol^{-1} that is more significant than $\Delta E_{int}^{N13,H17}$. Moreover, the components of $\Delta E_{int}^{Cov\text{-bonds}}$, namely $\Delta V_{XC}^{Cov\text{-bonds}}$ =

+15.0 kcal mol^{-1} and $\Delta V_{\mathrm{cl}}^{Cov\text{-bonds}}$ = –54.2 kcal mol^{-1} show that, quite unexpectedly, an increase in strength of all covalent bonds in **1a** is entirely due to a large contribution of stabilizing nature made by these bonds' electrostatic (classical) components.

2 The sum of all (covalent and long distance) 136 diatomic interaction energies ($E_{\mathrm{int}}^{\mathrm{Tot}}$) is more negative in **1a,** by $\Delta E_{\mathrm{int}}^{\mathrm{Tot}}$ = –49.6 kcal mol^{-1}, with a classical component $\Delta V_{\mathrm{cl}}^{\mathrm{Tot}}$ of –49.8 kcal mol^{-1}, meaning that the same set of interactions contributes to stability of the LEC much more then in HEC

3 Atoms of the {C14,N13} fragment are involved in the strongest attractive intramolecular diatomic interaction in both conformers, $E_{\mathrm{int}}^{\mathrm{C14,N13}}$ = –188.2 and –176.9 kcal mol^{-1} in **1a** and **1b**, respectively, with 96.8 % coming from the electrostatic nature of these interactions ($V_{\mathrm{cl}}^{\mathrm{C14,N13}}$ of –182.2 and –171.1 kcal mol^{-1} in **1a** and **1b**, respectively). The N13···H17 H-bond is only the second strongest interaction that is weaker, by 55.3 and 77.5 kcal mol^{-1} in **1a** and **1b**, respectively, when compared with the C14···N13 interaction. In accord with classical thinking, however, the intramolecular N13···H17 H-bonding interaction strengthened most (by –33.5 kcal mol^{-1}) among all di-atomic interactions on the **1b** → **1a** structural change closely followed by the covalently bonded {C14,O16} atom-pair with $\Delta E_{\mathrm{int}}^{\mathrm{C14,O16}}$ of –33.0 kcal mol^{-1} ($\Delta V_{\mathrm{XC}}^{\mathrm{C14,O16}}$ = –6.6 and $\Delta V_{\mathrm{cl}}^{\mathrm{C14,O16}}$ = –26.4 kcal mol^{-1}).

One can also gain an additional insight on relative stability of conformers by analysing changes in specific energy components on the structural transformation of **1b** (HEC) to **1a** (LEC). The FAMSEC method is perfectly suited for the purpose and it revealed that:

1) Out of 17 atoms of *S*-proline, 11 became involved in stronger intramolecular interactions as measured by the $E_{\mathrm{int}}^{\mathrm{A,}\mathcal{R}}$ term where $\mathcal{R}$. is a molecular fragment made of all the atoms of *S*-proline except A. This means that most of atoms found molecular environment of the LEC of S-proline favourable with N13, C14, O16 and H17 strengthening their interactions most for

which we obtained the $\Delta E_{\rm int}^{\rm A,\mathcal{R}} / \Delta V_{\rm cl}^{\rm A,\mathcal{R}}$ values of –45.8/–36.7, –27.2/–29.2, –21.9/–22.7 and –9.8/–9.6 kcal mol^{-1}, respectively. Notably, strengthening/weakening of most (but not all) interactions (regardless whether being considered as covalent bonds or long-distance) is predominantly due to changes in the electrostatic components as no new bonds are formed or broken on the **1b** $\rightarrow$ **1a** structural transformation.

2) The {H17,N13} fragment became most stabilised (*loc*-FAMSEC = –18.3 kcal mol^{-1}) whereas the {O16,N13} fragment stabilised the entire **1a** molecule the most (*mol*-FAMSEC = –49.7 kcal mol^{-1}). It is important to understand that there are two possible ways any *n*-atom fragment can stabilize the *fin* (here **1a**) relative to *ref* (here **1b**) state of a molecular system when interactions are considered: interactions can become either more attractive or less repulsive. To illustrate this let us first follow energy contributions made by an 'obvious' $\mathcal{G}$ = {H17,N13} fragment. Its di-atomic interaction strengthened most (with $\Delta E_{\rm int}^{\rm N13,H17}$ of –33.5 kcal mol^{-1}) among all 136 atom-pairs, self-fragment energy increased ($\Delta E_{\rm self}^{\rm N13,H17}$ = 15.2 kcal mol^{-1}) and as a result this fragment became stabilised in **1a** (eq. 7) with *loc*-FAMSEC = –18.3 kcal mol^{-1}. New nuclear positions of N13 and H17 in **1a** resulted in the overall weakening of these atoms interactions with remaining atoms of S-proline treated as a fragment $\mathcal{H}$ ($\Delta E_{\rm int}^{\mathcal{G},\mathcal{H}}$ =11.3 kcal mol^{-1}). Summing up *loc*-FAMSEC and $\Delta E_{\rm int}^{\mathcal{G},\mathcal{H}}$ (eq. 8) gives *mol*-FAMSEC energy term of –7.0 kcal mol^{-1} showing that this atom-pair does indeed adds to overall stability of **1a** but its contribution is seven times smaller than that made by the $\mathcal{G}$ = {O16,N13} fragment. To explain this unexpected finding we will follow the same protocol as for the intramolecular H-bond. Predictably, N13 and O16 are involved in highly repulsive interaction in **1b** ($E_{\rm int}^{\rm N13,O16}$ = 128.0 kcal mol^{-1}) that became even more repulsive in **1a** by +12.9 kcal mol^{-1} (with $\Delta V_{\rm XC}^{\rm N13,O16}$ and $\Delta V_{\rm cl}^{\rm N13,O16}$ of –5.0 and +17.9 kcal mol^{-1}, respectively) due to d(N13,O16) of 2.5535 Å in **1a**

being shorter by 0.1577 Å than in **1b**. The self-fragment energy increased by $\Delta E_{\text{self}}^{\text{N13,O16}}$ = 30.8 kcal mol^{-1} and the sum of $\Delta E_{\text{int}}^{\text{N13,O16}}$ and $\Delta E_{\text{self}}^{\text{N13,O16}}$ gave *loc*-FAMSEC = +43.7 kcal mol^{-1} resulting in the $\mathcal{G}$ = {N13,O16} fragment being most destabilised on **1b** → **1a**. However, the new placement of these two atoms in **1a** facilitated these atoms interactions with the remaining atoms (fragment $\mathcal{H}$) such that $\Delta E_{\text{int}}^{\mathcal{G},\mathcal{H}}$ = –93.5 kcal mol^{-1} was obtained. Clearly, the inter-fragment interactions strengthened much more when compared with an increased (i) repulsive interaction between the two atoms and (ii) their self-atomic energies. Summing up *loc*-FAMSEC and $\Delta E_{\text{int}}^{\mathcal{G},\mathcal{H}}$ terms we obtained the net energy contribution, *mol*-FAMSEC = –49.7 kcal mol^{-1}, made by this fragment to a molecular energy that is the most significant contribution of stabilizing nature among all unique atom pairs in **1a**.

The above observations (together with additional data in ESI) would not be easily predicted (most likely not even considered) by a classical organic chemist, but on the other hand they provide a wealth of information and, most importantly, fully explain the relative stability of the *S*-proline conformers (and molecular systems in general).

3.2. Proline-Acetone adduct formation

To understand reaction mechanism fully we have analysed all possible structural changes leading to consecutive steps and relevant energy profiles computed at two levels of theory are depicted in Fig. 2. Such a detailed approach is not used in classical interpretations of reaction mechanisms (see Scheme 1); hence, they cannot explain, e.g., chemical reactivity of conformers or necessary structural re-arrangements as well as forces driving a chemical change. In sections that follow we will explain all energy changes along reaction coordinates as well as energy differences between consecutive steps computed for the lowest and higher energy conformers of S-proline.

S-proline **1** and acetone **2** readily form adducts **3**; the global minimum structures discovered, **3a** and **3b,** are shown in Fig. 1, part B. Importantly, the energy of **3b** is only 2.5 kcal mol^{-1} higher

relative to **3a** at the MP2 level (Fig. 2). Clearly, on the adduct formation, the energy of molecular system **3b** made of **1b**+**2** must have decreased more significantly relative to **3a**.

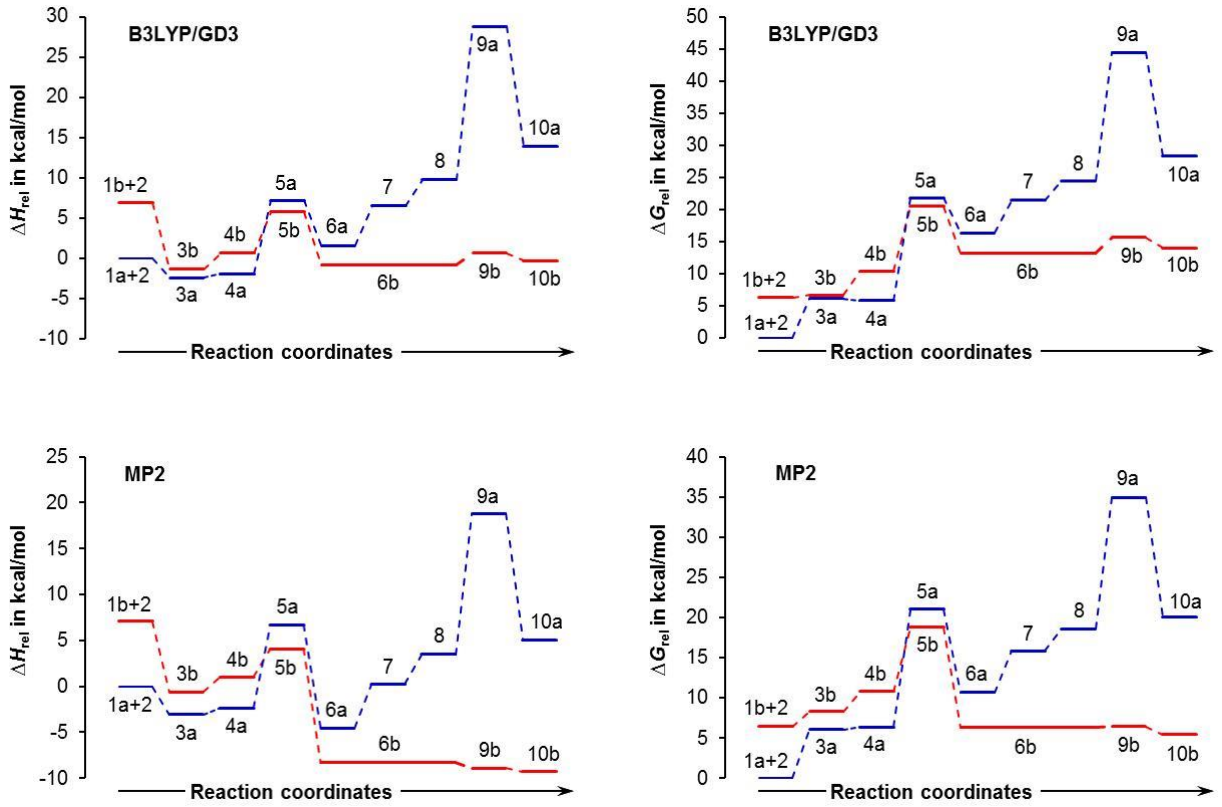


Fig. 2 Relative to the initial states, either **1a**+**2** or **1b**+**2**, enthalpy and Gibbs free energy changes computed at the indicated levels of theory for all intermediate structures leading to the product of H-transfer, **10a** and **10b**.

To explain this, let us focus on what we consider being most pertinent to a classical chemist (for more details and relevant data see PART S4 in the ESI):

1) The two molecules, **1** and **2**, can be seen as molecular fragments of the molecular system **3**. These molecules show high affinity to each other as measured by the total inter-molecular interaction energy, $E_{\rm int}^{\mathbf{1,2}}$, *i.e.*, the sum of *inter-molecular* diatomic interaction energies computed for all unique 170 atom-pairs. We found $E_{\rm int}^{\mathbf{1a,2}}$ and $E_{\rm int}^{\mathbf{1b,2}}$ of –34.3 (with $V_{\rm XC}^{\mathbf{1a,2}}/V_{\rm cl}^{\mathbf{1a,2}}$ = –26.7/–7.6 kcal mol^{-1}) and –53.6 kcal mol^{-1} (with $V_{\rm XC}^{\mathbf{1b,2}}/V_{\rm cl}^{\mathbf{1b,2}}$ = –40.7/–12.9 kcal mol^{-1}) in

3a and **3b**, respectively. Therefore the HEC interacts with **2** stronger by –19.3 kcal mol^{-1} with main contribution coming from the XC-term.

2) The combined intra- and inter-molecular diatomic interactions became stronger in **3**; hence, they stabilized both molecular systems but significantly more, by –24.8 kcal mol^{-1}, in the case of **3b**.

3) Molecular fragments $\mathcal{G}$ and $\mathcal{H}$ (Fig. 3) containing atoms of **1** and **2**, respectively, can be seen as driving the adduct formation. This is because these atoms are involved in strongest diatomic inter-molecular interactions in **3a** and **3b** with $\left|E_{\mathrm{int}}^{\mathrm{A,B}}\right|$ > 10 kcal mol^{-1}; among them, 10 and 5 atom-pairs are involved in very strong interactions with $\left|E_{\mathrm{int}}^{\mathrm{A,B}}\right|$ above 50 and 100 kcal mol^{-1}, respectively - see Tables S7–S9, PART S4 in the ESI. The inter-fragment interaction energy, $E_{\mathrm{int}}^{\mathcal{G},\mathcal{H}}$ of –5.3 and –43.2 kcal mol^{-1} computed for **3a** and **3b,** respectively, supports the much higher affinity between **1b** and **2**.

4) We have also identified individual atoms playing most significant role – Table S7 in PART S4 in the ESI. The formation of adducts is driven mainly by attraction between entire fragment $\mathcal{G}$ of **1** and (i) C18 of **2** with $E_{\mathrm{int}}^{\mathcal{G},\mathrm{C18}}$ = –19.5 kcal mol^{-1} ($V_{\mathrm{cl}}^{\mathcal{G},\mathrm{C18}}$ = –17.4 kcal mol^{-1}) in **3a** and (ii) quite unexpectedly, O19 of **2** with over three times stronger interaction energy $E_{\mathrm{int}}^{\mathcal{G},\mathrm{O19}}$ of –62.0 kcal mol^{-1} ($V_{\mathrm{XC}}^{\mathcal{G},\mathrm{O19}}$ = –28.3 kcal mol^{-1} and $V_{\mathrm{cl}}^{\mathcal{G},\mathrm{O19}}$ = –33.7 kcal mol^{-1}) in **3b**.

5) The leading role of the {H17,O19} atom-pair destined to form a new covalent bond is already apparent on the adduct formation. Relative to **3a**, an order of magnitude larger $V_{\mathrm{XC}}^{\mathcal{G},\mathrm{O19}}$ in **3b** is mainly due to the exchange-correlation term $V_{\mathrm{XC}}^{\mathrm{H17,O19}}$ of –18.3 kcal mol^{-1} that is in contrast to ~0 kcal mol^{-1} (in both adducts, **3a** and **3b**) computed for the atom-pair destined to make a new C18–N13 bond.

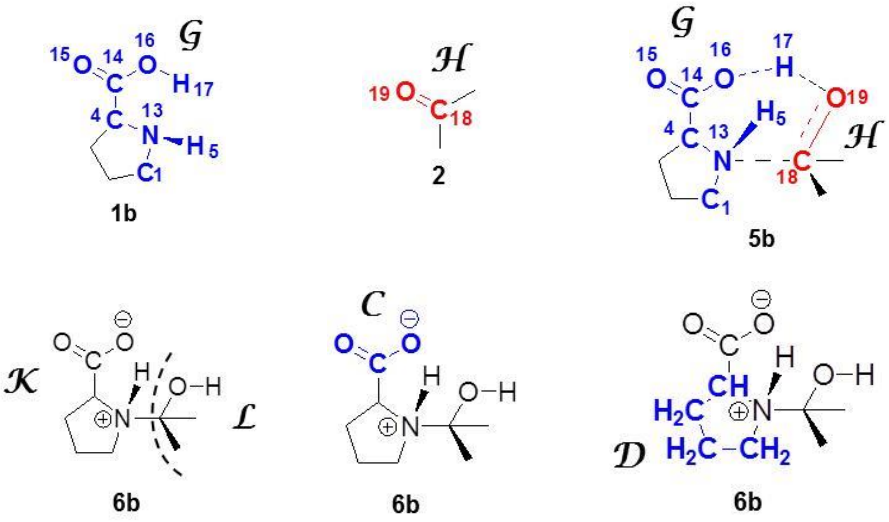


Fig. 3 Schematic presentation of selected molecular fragments used in this study.

To conclude, regardless of the approach taken, a consistent picture emerges pin-pointing the origin of a more significant energy decrease for the **3b** formation and the higher affinity of the HEC (**1b**) to **2**.

3.3. CN-bond formation

In a first step, proline-acetone adducts **3** must overcome an energy barrier of about 3 kcal mol^{-1} to form pre-organised structures **4** (Fig. 4); a complete set of data pertaining to the CN bond formation is included in PART S5 of the ESI. As found for **3**, **1** and **2** are also involved in overall attractive inter-molecular interactions in **4** but nearly three times stronger in **4b** ($E_{\text{int}}^{\mathbf{1,2}}$ = –70.3 kcal mol^{-1}). Surprisingly, however, the $E_{\text{int}}^{\mathbf{1,2}}$ energy term changed in the opposite direction on the **3a** → **4a** and **3b** → **4b** structural re-arrangements as we found $\Delta E_{\text{int}}^{\mathbf{1,2}}$ of +10.4 and –16.7 kcal mol^{-1}, respectively. Therefore, **1** and **2** interact much stronger in **4b** (by –46.4 kcal mol^{-1}) and **4b** appears to be much better pre-organised for progression of the reaction as (i) the interatomic distances d(N13,C18) and d(H17,O19) of 2.9278 and 1.6806 Å in **4b** are much shorter than in **4a** (by 0.41 and 0.64 Å, respectively). Moreover, considering the **3** → **4** structural re-arrangements we found that:

a) The same sixteen {A,B} atom-pairs (constituting the $\mathcal{G}$ and $\mathcal{H}$ molecular fragments shown in Fig. 3) are involved in either attractive or repulsive inter-molecular interactions with $\left|E_{\text{int}}^{\text{A,B}}\right|$ > 10 kcal mol^{-1} in **3** and **4**; the same atoms also experienced the most significant change in interaction energies on the preorganization process – Tables S13-S15 in PART S5 of the ESI.

b) Inter-molecular interactions between atoms of the {N13,C18} and {H17,O19} fragments in **3a** (note that these atom-pairs are to form new covalent bonds in **6a**) are highly attractive and entirely dominated by the classical term (with $E_{\text{int}}^{\text{N13,C18}} \approx V_{\text{cl}}^{\text{N13,C18}}$ = –76.9 kcal mol^{-1} and $E_{\text{int}}^{\text{H17,O19}} \approx V_{\text{cl}}^{\text{H17,O19}}$ = –65.6 kcal mol^{-1}, respectively) due to a large difference in these atoms net atomic charges ΔQ(C18,N13) = Q(C18) – Q(N13) = 0.9685 – (–0.9831) = 1.9516e and ΔQ(H17,O19) = Q(H17) – Q(O19) = 0.6213 – (–1.1537) = 1.7750e. However, they are only fourth and fifth, respectively, among most attractive interactions (the strongest interaction in **3a** was found between C14 (Q(C14) = +1.5236e) and O19 (Q(O19) = –1.1579e) atoms with $E_{\text{int}}^{\text{C14,O19}}$ and $V_{\text{cl}}^{\text{C14,O19}}$ of –161.3 and –161.0 kcal mol^{-1}, respectively; clearly, the large electrostatic interaction is due to a large difference in these atoms net charges, ΔQ(C14,O19) = 2.6815e. Interactions between atoms of {N13,C18} and {H17,O19} became even stronger and of comparable strength in **4a** (with $E_{\text{int}}^{\text{A,B}}$ of about –103 kcal mol^{-1}) but $E_{\text{int}}^{\text{C14,O19}}$ of –139.6 kcal mol^{-1} is still the strongest. We also noted that $\Delta E_{\text{int}}^{\text{H17,O19}}$ of –37.1 kcal mol^{-1} was most significant among 170 intermolecular diatomic interaction changes on **3a** → **4a** (the $\Delta E_{\text{int}}^{\text{N13,C18}}$ term is only third among most significant changes among attractive diatomic interactions) and the {H17,O19} fragment became most stabilised in **4a** with *loc*-FAMSEC = –33.4 kcal mol^{-1} whereas {N13,C18} with *loc*-FAMSEC = –24.1 kcal mol^{-1} is only fourth.

c) In **3b**, H17 and O19 are involved in an extremely strong interaction of $E_{\text{int}}^{\text{H17,O19}}$ = –143.9 kcal mol^{-1} that is (i) nearly three times stronger than that between N13 and C18 and (ii) second

strongest as it is about 'only' 16 kcal mol^{-1} weaker than that found between C14 and O19. The interaction between N13 and C18 strengthened the most, by –54.1 kcal mol^{-1} when in **4b** ($E_{\text{int}}^{\text{N13,C18}}$ = –112.2 kcal mol^{-1}) but it is still 34.3 kcal mol^{-1} weaker when compared with the {H17,O19} fragment. Here again, atoms of the {C14,O19} fragment are involved in the strongest interaction of –165.3 kcal mol^{-1}.

d) Importantly, the largest XC-term of –18.3 kcal mol^{-1} is observed between H17 and O19 in **4b** in contrast to $V_{\text{XC}}^{\text{N13,C18}}$ of –4.7 kcal mol^{-1}. Recalling that the XC-term describes a degree of covalent contribution to an interaction, it is clear that the process of a covalent O19–H17 bond formation is more advanced relative to N13–C18. Moreover, the XC-term is only –0.9 and –3.8 kcal mol^{-1} for the {N13,C18} and {H17,O19} atom-pairs in **4a**, respectively, pointing at the HEC as most likely conformer to be involved in bonds formation with S-proline.

From the above, which can be seen as a picture recovered from the diatomic inter-molecular interaction perspective, it is obvious that (i) the interaction between N13 and C18 cannot be seen as the leading driver in forming either **4a** or **4b** and (ii) between the {H17,O19} and {N13,C18} atom pairs, the former plays by far more important role in leading to **4b**.

An additional and important insight one can gain from a single atom A perspective when its interactions with entire oncoming molecule are considered, either $E_{\text{int}}^{\text{A,1}}$ or $E_{\text{int}}^{\text{A,2}}$. To this effect we discovered (Table S16 in PART S5 of the ESI) that on **3** → **4**:

a) Totally unexpectedly, C18 that is destined to form a covalent bond with N13 of **1**, is involved in most significant overall repulsive interactions dominated by electrostatic repulsion as we obtained $E_{\text{int}}^{\text{C18,1}}/V_{\text{cl}}^{\text{C18,1}}$ of +7.8/+9.8 and +11.2/+17.2 kcal mol^{-1} in **4a** and **4b**, respectively; hence, it is opposing oncoming **1a** and **1b**. Even more surprising is the fact that C18 interactions with **1** changed from slightly attractive in **3a** to repulsive in **4a.** In contrast, O19 that is destined to form a covalent bond with H17 of **1**, is involved in most attractive interactions with about

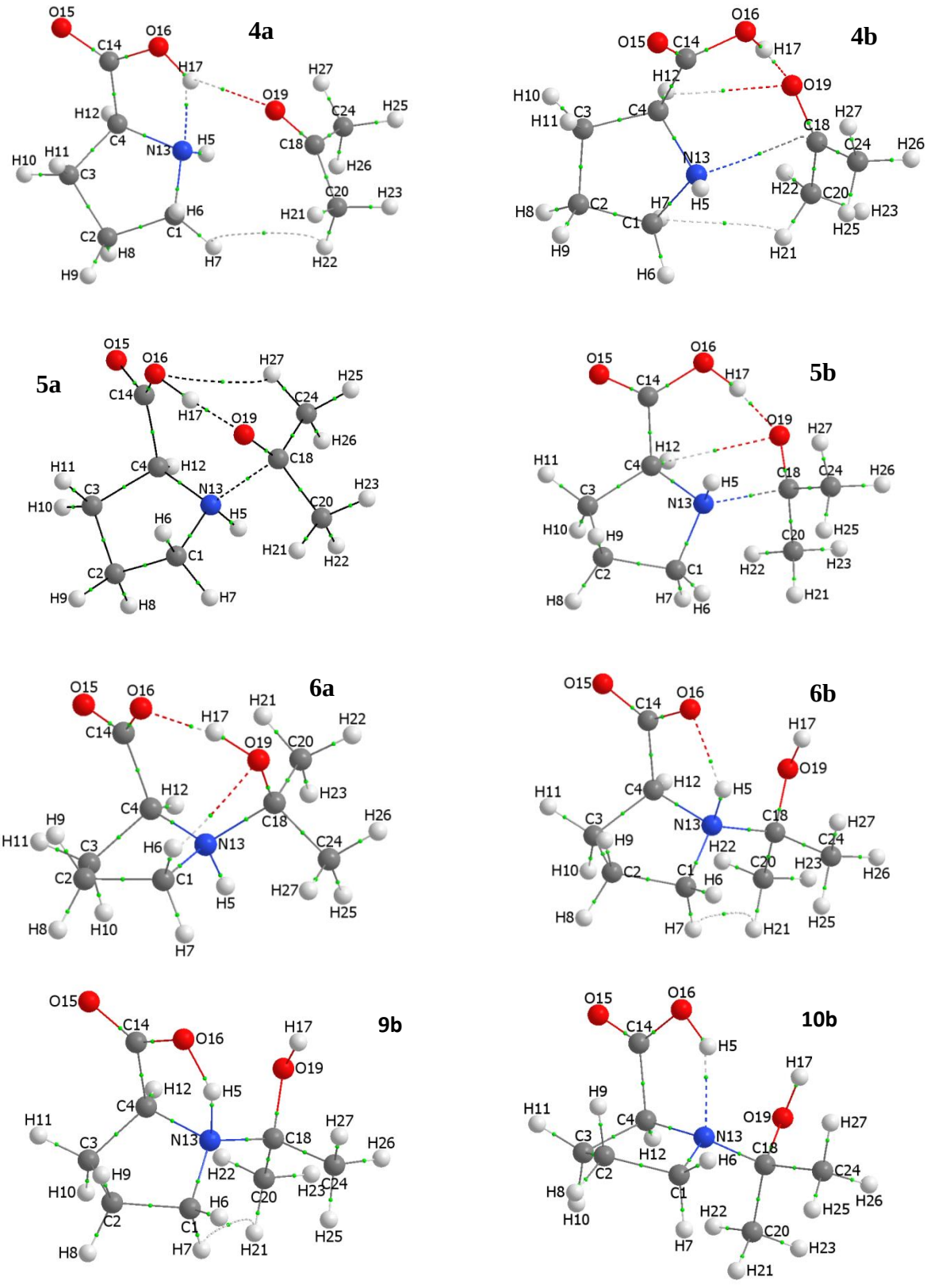


Fig. 4 Molecular graphs of pre-organised adducts (**4a** and **4b**), TS structures for the CN-bond formation (**5a** and **5b**), products after the C–N bond formation (**6a** and **6b**), TS structure for the first H-transfer (**9b)** and product after the first H-transfer (**10b**).

50% covalent contribution; we obtained $E_{\text{int}}^{\text{O19},\mathbf{1}}/V_{\text{XC}}^{\text{O19},\mathbf{1}}$ of –25.8/–13.3 and –69.3/–35.6 kcal mol^{-1} in **4a** and **4b**, respectively. Hence O19 facilitates the process leading to the formation of these two bonds most among atoms of **2**.

b) Focusing on **1**, there are three atoms that attract oncoming **2** most, namely N13, H17 and C14. H17 of both S-proline conformers is involved in the strongest attractive interactions that are predominantly of electrostatic nature in **4a** but showing 50% covalent character in **4b** ($E_{\text{int}}^{\text{H17},\mathbf{2}}$/ $V_{\text{XC}}^{\text{H17},\mathbf{2}}$ of –17.0/–4.0 (in **4a**) and –38.4/–19.2 (in **4b**) kcal mol^{-1}). Interactions involving N13 and **2** are stronger in **4b** and are dominated by the XC-term ($E_{\text{int}}^{\text{N13},\mathbf{2}}/V_{\text{XC}}^{\text{N13},\mathbf{2}}$ of –5.8/–7.3 (in **4a**) and –24.6/–15.2 (in **4b**) kcal mol^{-1}). This illustrates an important role played by H17 in both conformers of S-proline and N13 in the HEC that, as discussed above, shows much higher overall affinity to **2**.

Let us discus a picture that immerged on the **4** → **5** step (*i.e.*, reaching the TS) starting from the inter-molecular diatomic interactions. As one would expect, atoms of the {N13,C18} and {H17,O19} fragments are involved in strongest interactions in **5** (Table S19 in PART S5 of the ESI). However, $E_{\text{int}}^{\text{H17,O19}}$ of –217.0 kcal mol^{-1} is still stronger (by –9.4 kcal mol^{-1}) than that involving N13 and C18 in **5a**. In contrast, $E_{\text{int}}^{\text{N13,C18}}$ of –224.8 kcal mol^{-1} is now stronger (by –26.7 kcal mol^{-1}) than that involving H17 and O19 in **5b**. Notably, only at the TS the covalent contribution to interaction between N13 and C18 became larger ($V_{\text{XC}}^{\text{N13,C18}}$ = –74.2 kcal mol^{-1}) relative to the that found for the {H17,O19} atom-pair ($V_{\text{XC}}^{\text{H17,O19}}$ = –44.9 kcal mol^{-1}). Interestingly, the new bonds formed, C18N13 and O19H17, are of comparable strength in **6b**, as indicated by interaction energies of –311.9±0.1 kcal mol^{-1}, but the O19H17 bond in **6a** is stronger by –30.4 kcal mol^{-1}.

From the 1-atom perspective, interactions between C18 and **1** changed from overall repulsive to attractive ($E_{\text{int}}^{\text{C18},\mathbf{1}}$ of –32.3 and –43.4 kcal mol^{-1} were computed in TSs **5a** and **5b**, respectively,

Table S21 in PART S5 of the ESI) and they are several times weaker when compared with interactions between O19 and **1** for which we found $E_{\text{int}}^{\text{O19},\mathbf{1}}$ of –174.9 and –167.6 kcal mol^{-1} in **5a** and **5b**, respectively. Furthermore, from the classical electrostatic component point of view, C18 is strongly obstructing oncoming **1** in the TSs as $V_{\text{cl}}^{\text{C18},\mathbf{1}}$ became more repulsive from +9.8/+17.2 to +47.8/+48.3 kcal mol^{-1} in **4a**/**5a** and **4b**/**5b**, respectively. In contrast, O19 interactions with **1** are characterised by largely strengthened electrostatic attractions that changed from –12.5/–33.8 kcal mol^{-1} in **4a**/**4b** to –93.1/-94.5 kcal mol^{-1} in **5a**/**5b**.

Considering atoms of **1**, the interactions between either N13 or H17 and **2** are by far the strongest as we found $E_{\text{int}}^{\text{N13},\mathbf{2}}$ of –120.0/–131.9 kcal mol^{-1} and $E_{\text{int}}^{\text{H17},\mathbf{2}}$ of –100.0/–93.8 kcal mol^{-1} in **5a**/**5b**. Importantly, these atoms attractive nature of electrostatic interactions with **2** strengthened and particularly so in the case of H17 as $V_{\text{cl}}^{\text{H17},\mathbf{2}}$ changed from –12.8/–19.2 (in **4a**/**4b**) to –53.8/–56.8 (in **5a**/**5b**) kcal mol^{-1}.

From the above and data in the ESI we came to several important conclusions:

1. This stage of the catalytic process is always reflected as the CN-bond formation implying that it is driven by the interacting $\{N^{\delta-}\cdots C^{\delta+}\}$ atom pair. Our results contradict this generally accepted view as they conclusively showed that many diatomic and atom-molecule interactions (excluding those involving C18) can be seen as responsible for relative orientation of S-proline (**1**) and acetone (**2**) already on the adduct (**3**) formation and its better pre-organised structure (**4**) leading to the TS. As a matter of fact, C18 is involved in repulsive classical interactions with an on-coming **1** even at the TS. Hence, we have concluded that binding of S-proline and acetone via the C18–N13 bond formation can be seen as kind of a 'by-product' of other and much stronger, hence leading interactions.
2. From the perspective of a single atom interacting with an oncoming molecule it follows that: (i) O19 of **2** (due to interactions with atoms of oncoming **1**, either **1a** or **1b**) and (ii) N13 and H17 of **1** (due to these atoms interactions with atoms of oncoming **2**) drive the process from **4** to **6**.

3. From the 2-atom perspective we discovered that the interaction between C14 and O19 is the strongest in **3** and **4** and is closely followed by strength of interaction between H17 and O19 in **3b** and **4b**. Hence, the {C14$^{\delta+}\cdots$O19$^{\delta-}$} fragment (with $\Delta\delta$(C14,O19) = 2.69*e* in **4b**) can be seen as a driving force that is highly assisted by the {H17$^{\delta+}\cdots$O19$^{\delta-}$} fragment (with $\Delta\delta$(H17,O19) = 1.79*e* in **4b**) in process leading to the CN as well as OH bonds formation in **6**.
4. It is evident that O19 plays a very special and deceive role at this stage of a catalytic process as it is a major player regardless of the perspective taken.
5. It is apparent that **2** should preferentially form CN and OH bonds with **1b** rather than **1a**.

Let us briefly discuss an energy difference between transition states **5a** and **5b**. The energy barrier, when moving from **4** to TSs **5,** is much lower (by ~6 kcal mol^{-1}) for **5b** resulting in **5b** having a lower energy than **5a** by ~2.5 kcal mol^{-1} (MP2 data). The lower energy barrier computed for **5b** can be attributed to **4b** being better pre-organised when compared with **4a**, as discussed above. A slightly lower energy of **5b** relative to **5a** can be explained from the total self-molecular energy, $E_{\text{self}}^{\text{Tot}}$, and the total interaction energy of the molecular system, $E_{\text{int}}^{\text{Tot}}$. Since $E_{\text{self}}^{\text{Tot}}$ increased in **5**, relative to **4**, by about the same value of 26 kcal mol^{-1}, the lower energy of **5b** when compared with **5a** must be attributed to the change in $E_{\text{int}}^{\text{Tot}}$ (recall that energy of a system is the sum of $E_{\text{self}}^{\text{Tot}}$ and $E_{\text{int}}^{\text{Tot}}$). Indeed, the overall change in all interaction energies was found to be –8.5 and –11.0 kcal mol^{-1} for **5a** and **5b**, respectively, and this compares very well with the difference in the energy barrier at the TS being ~2.5 kcal mol^{-1} lower in the case of **5b**. Clearly, the small energy difference between **5a** and **5b** is, in this instance, a result of combined large increases and decreases in interaction energies between many atoms on the **4** $\rightarrow$ **5** step.

Finally, combined MP2 data (for details see Table S4 in PART S4 and Table S8 in PART S5 in the ESI) shows that the energy of **6b** is lower, relative to **1b+2**, by ΔH = –15.4 kcal mol^{-1} and ΔG = –0.2 kcal mol^{-1}. In contrast, a small decrease in *H* and significant increase in *G* of –4.6

and +10.6 kcal mol^{-1}, respectively, was computed for **6a**. From this, one can conclude that on the CN-bond formation the reaction path involving **1a** should be eliminated. In other words, any molecular system that has two equivalent (chemistry-wise) states will proceed towards the lower energy structure, here **6b**, provided this does not require overcoming a large energy barrier as is indeed the case for the **1b**+**2** pathway. As this work has demonstrated, a selection based on relative energies of conformers or molecules in general (a common practice in the field[6]) in considering their involvement in synthetic rout might lead to wrong conclusions.

3.4. First proton transfer

Energy profile diagrams in Fig. 2 show that **6a**, even if it were present in the reaction environment, cannot be involved in consecutive steps. This is because (i) **1a+2** would have to overcome (ΔG_{rel} at the MP2 level) a total energy barrier of ~35 kcal mol^{-1} to reach **9a** (TS) and (ii) the product of H-transfer **10a** is higher in energy than **1a**+**2** by 20 kcal mol^{-1} (note significantly larger values at B3LYP/GD3). From the ΔG_{rel} perspective, a reverse process is thermodynamically driven, from **10a**, via quite small 15 kcal mol^{-1} energy barrier at **9a**, to initial reactants **1a**+**2**.

A very different and significantly more favourable energy profile is observed for the path involving **1b**+**2**. These reactants must overcome a small energy barrier (from **1b**+**2** to **5b**) of 12.3 kcal mol^{-1} (ΔG_{rel} at MP 2) to form **6b** and the proton transfer is essentially 'energy-free' (ΔG_{rel} at MP2 between **6a** and **9b** is 0.1 kcal mol^{-1}). Hence, our overall interpretation is as follows. We found an energy barrier of 10.8 kcal mol^{-1} (at B3LYP/GD3) applicable to the **1a** $\rightarrow$ **1b** conformational re-arrangement. This means that starting from **1a**+**2** it is less energy demanding to reach **5b** via **1b** than **5a** via **3a** and **4a**. Furthermore, ΔE_{rel}(**6b**) << ΔE_{rel}(**6a**) regardless of whether ΔG_{rel} or ΔH_{rel} at MP2 is considered. It is then clear that **1a** is eliminated from the reaction environment already at the first major step of this catalytic process;

nonetheless, for those interested we have provided relevant data to explain step-wise hypothetical changes from **6a** to **10a** in the ESI.

Typically, the energy profiles shown in Fig. 2, together with accompanied analysis of energy differences between consecutive steps, would fully satisfied a physical organic chemist in interpreting (i) relative catalytic properties of S-proline and (ii) the proton transfer as nearly a spontaneous process, hence not rate determining step. However, having all diatomic interactions at hand, we noticed with a great concern that H5 is attracted much more to N13 (by –135 kcal mol^{-1}) than to O16 in **6b**. Then why does H5 leave N13 as easily as revealed by the enthalpy and Gibbs free energy changes shown in Fig. 2? Clearly, the related energy profile cannot be explained by way of classical thinking and it was of paramount importance to analyse molecular system from **6b**, *via* **9b**(TS) to **10b**.

Looking at computed diatomic interactions we established that H5 is involved in attractive interactions only with 4 atoms, namely: N13 to which it is bonded to in **6b**, O15 and O16 of the COO functional group of molecular fragment $\mathcal{K}$ and O19 of the $\mathcal{L}$ fragment (see Fig. 3). Due to the fact that on a proton transfer process H5 is heading not only towards O16 but rather in the direction of the entire COO group, it makes perfect sense to partition $\mathcal{K}$ into two fragments, one containing COO atoms (we will call it $\mathcal{C}$) and remaining atoms of $\mathcal{K}$ with exclusion of N13 (let us call it $\mathcal{D}$). We decided to make use of our approach to gain an insight on a plausible origin of the proton transfer. To this effect, we considered specifically selected for the purpose interactions:

a) The interaction energies $E_{\text{int}}^{\text{H5,N13}}$ and $E_{\text{int}}^{\text{H5,O16}}$ are –254.3 and –119.1 kcal mol^{-1} in **6b**, respectively. Clearly, this cannot lead to an energy 'free' transfer of H5 to O16.

b) H5 is being attracted by atoms of the $\mathcal{C}$ fragment with $E_{\text{int}}^{\text{H5},\mathcal{C}}$ of –65.5 (in **6b**) and –148.0 (in **9b**) kcal mol^{-1} and this is not sufficient for the spontaneous proton transfer either

because $E_{int}^{H5,N13}$ = –203.8 kcal mol^{-1} was obtained in **9b**. Therefore, there must be other interactions that facilitate the chemical event leading to the TS **9b**.

c) Notably, H5 is being repelled by atoms of $\mathcal{L}$ with $E_{int}^{H5,\mathcal{L}}$ of +46.2 (in **6b**) and +43.5 (in **9b**) kcal mol^{-1}. One must stress that the computed $E_{int}^{H5,\mathcal{L}}$ term includes the attractive interaction between H5 and O19 in $\mathcal{L}$ of –68.1 and –75.8 kcal mol^{-1} in **6b** and **9b**, respectively.

d) H5 is also being repelled by atoms of $\mathcal{D}$ with $E_{int}^{H5,\mathcal{D}}$ of +85.3 (in **6b**) and +86.7 (in **9b**) kcal mol^{-1}.

Importantly, these repulsive interactions, $E_{int}^{H5,\mathcal{L}}$ and $E_{int}^{H5,\mathcal{D}}$, do not change significantly on **6b** → **9b** and, by repelling H5, can be seen as counteracting the H5 attraction to N13 by 'pushing' H5 towards O16 (or C in general); hence, atoms of $\mathcal{L}$ and $\mathcal{D}$ facilitate H5 transfer. One can get a rough estimate of the corrected (for repulsive contributions) interaction energy $^{corr}E_{int}^{H5,N13}$ between H5 and N13 by summing up $E_{int}^{H5,N13}$, $E_{int}^{H5,\mathcal{L}}$ and $E_{int}^{H5,\mathcal{D}}$; it gives a product of –122.8 (in **6b**) and –73.7 kcal mol^{-1} in **9b**. Finally, accounting for the attraction of H5 to the C fragment ($^{corr}E_{int}^{H5,N13} - E_{int}^{H5,C}$) gives us –57.3 and +74.3 kcal mol^{-1} of the net interaction energies between N13 and H5 in **6b** and **9b**. This shows that a small movement of H5 between N13 and O16, *e.g.*, due to numerous vibrational modes, will change the balance from being attracted more either to N13 or to O16.

Finally, the computed $^{corr}E_{int}^{H5,N13}$ and $E_{int}^{H5,C}$ interaction energy terms in **10b** of –16.4 and –223.3 kcal mol^{-1}, respectively, show that the proton transfer from **6b** to **10b** can be seen as overall favourable. Furthermore, one must note that (i) **10b** is perfectly pre-organised for the next step, *i.e.*, water elimination and (ii) according to our study, H5 must be on O16 for the water elimination to take place. From this follows that even when one assumes some kind of

equilibrium N13 ← H5 → O16 with H5 oscillating between N13 and O16 in **6b**, **9b** and **10b**, the reaction will proceed with ease due to **10b** being used up by the water elimination process.

4. Conclusions

The reaction energy profile (REP) computed for the assumed reaction mechanism illustrates how the energy of a molecular system varies along the reaction coordinates. Small (or large) energy differences between consecutive steps are typically used in support (or rejection) of the proposed mechanism but they provide no insight on the origin of processes taking place. In order to identify atoms and molecular fragments leading to a chemical change (with the associated computed an energy change of a molecular system) we have implemented the general concept of the fragment attributed molecular system energy change (FAMSEC) method combined with analysis of interactions between fragment of different sizes (they range between a single atom to entire molecule). Hence, a new method is proposed (called the REP-FAMSEC method) that represents a shift from a commonly used 2-atom approach (involving interacting atoms of reactants with most negative $A^{\delta-}$ and positive $B^{\delta+}$ partial charges) to interacting poly-atomic fragments of a molecular system. Focusing on initial steps of the proline catalysed aldol reaction (used here as a case study), we have (i) identified atoms and molecular fragments that lead to every incremental step along the reaction pathway and (ii) quantified their energy contributions in terms of relevant intra- and inter-fragment/molecular interaction energies and their changes between consecutive stages of the reaction progression. The proposed REP-FAMSEC approach proved to be of general nature as we were able to fully explain (i) relative stability of the S-proline conformers, (ii) why only higher energy conformer (HEC) can act as a catalyst; it has been demonstrated that the involvement of the lower energy conformer (LEC) in the proline catalysed aldol reaction is terminated already in the first major step commonly called the CN-bond formation, (iii) the CN-bond formation mechanism showing that it is not driven by the interacting $\{N^{\delta-}\cdots C^{\delta+}\}$ atom pair but rather the O-atom of acetone plays a very special and

deceive role at this stage of a catalytic process, (iv) nearly energy-free the intra-molecular proton transfer (taking place after the CN-bond formation) from N to O atoms of proline moiety even though initially H-atom interacts three times stronger with N- than O-atom as well as (v) small differences in the ΔE values computed for reaction pathways involving LEC and HEC.

It has been shown recently[63] that the FAMSEC energy terms (more generally, changes in the IQA-defined energy terms) computed at the computationally affordable B3LYP/IQA combination produced exact qualitative description of relative stability of glycol conformers and, quantitatively, perfectly comparable values with CCSD/BBC1/IQA data. This, together with the protocol described in this work, paves the way for studying many reaction mechanisms even when a significant number of atoms is involved.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

The authors gratefully acknowledge the Centre for High Performance Computing (CHPC), South Africa, for providing computational resources to this research project and National Research Foundation of South Africa, Grant Number 105855.

References

1 R. Koch, and T. Clark, *Linear Combination of Atomic Orbitals*, The Chemist's Electronic Book of Orbitals. Springer, Berlin, Heidelberg 1999, pp. 5–22.

2 C. Jonathan, N. Grevees, S. Warren, and P. Wothers, *Organic Chemistry*, 1st ed., Oxford University Press 2001, pp. 123–133

3 W. O. Kermack, R. Robinson, *J. Chem. Soc. Trans.*, 1922, **121**, 427–440.

4 N. P. Grove, M. M. Cooper, and K. M. Rush, *J. Chem. Edu.*, 2012, **89**, 844–849.

5 G. Abhik, and B. Steffen, *Arrow Pusshing in Inorganic Chemistry*, A logical Approach to the Chemistry of Main Group Elements. John Wiley & Sons 2014.

6 L. Jarrige, D Glavač, G. Levitre, P. Retailleau, G. Bernadat, L. Neuville and G. Masson, *Chem. Sci*., 2019, **10**, 3765–3769.

7 H. Mitsunuma, S. Tanabe, H. Fuse, K. Ohkubo and M. Kanai, *Chem. Sci*., 2019, **10**, 3459–3465.

8 R. B. Woodward and R. Hoffmann, *J. Am. Chem. Soc*., 1965, **87**, 395–397.

9 R. B. Woodward and R. Hoffmann, *J. Am. Chem. Soc*., 1965, **87**, 2511–2513.

10 R. Hoffmann and R. B. Woodward, *Acc. Chem. Res*., 1968, **1**, 17–22.

11 K. Fukui, *Acc. Chem. Res*., 1971, **4**, 57–64.

12 R. D. Harcourt, *J. Mol. Struct*., (THEOCHEM) 1997, **398**, 93–100.

13 A. C. Pavão, C. A. Taft, T. C. F. Guimarães, M. B. C Leão, J. R. Mohallem and W. A. Lester, Jr., *J. Phys. Chem. A*, 2001, **105**, 5–11.

14 S. Shaik and P. C. Hiberty, Valence bond theory, its history, fundaments, and applications: A primer. In *Reviews in Computational Chemistry*; K. B. Lipkowitz, R. Larter and T. R. Cundari, Eds.; Wiley-VCH: Hoboken, New Jersey, 2004, Vol. 20, Chapter 1.

15 P. C. Hiberty and S. Shaik, *J. Comput. Chem*., 2007, **28**, 137–151.

16 E. D. Glendening and F. Weinhold, *J. Comput. Chem*., 1998, **19**, 593–609.

17 E. D. Glendening and F. Weinhold, *J. Comput. Chem*., 1998, **19**, 610–627.

18 E. D. Glendening, J. K. Badenhoop and F. Weinhold, *J. Comput. Chem*. 1998, 19, 628–646.

19 X. Li, Y. Zeng, L. Meng and S. Zheng, *J. Phys. Chem. A*, 2007, **111**, 1530–1535.

20 X. Li, H. Fan, L. Meng, Y. Zeng and S Zheng, *J. Phys. Chem. A*, 2007, **111**, 2343–2350.

21 P. Macchi and A. Sironi, *Coord. Chem. Rev*., 2003, **238**, 383–412.

22 N. O. J. Malcolm and P. L. A. Popelier, *J. Phys. Chem. A*, 2001, **105**, 7638–7645.

23 J. Andrés, S. Berski, M. Feliz, R. Llusar, F. Sensato and B. Silvi, *C. R. Chimie*, 2005, **8**, 1400–1412.

24 L. R. Domingo, *RSC Adv.*, 2014, **4**, 32415–32428.

25 J. Andrés, S. Berski, L. R. Domingo, V. Polo and B. Silvi, *Current Org. Chem.*, 2011, **15**, 3566–3575.

26 L. R. Domingo, M. Ríos-Gutiérrez, B. Silvi and P. Pérez, *Eur. J. Org. Chem.*, 2018, **9**, 1107–1120.

27 L. R. Domingo, *Molecules*, 2016, **21**, 1319.

28 V. Tognetti, S. Bouzbouz and L. Joubert, *J. Mol. Model.*, 2017, **23**, 5.

29 R. Inostroza-Rivera, M Yahia-Ouahmed, V. Tognetti, L. Joubert, B. Herrera and A. Toro-Labbé, *Phys. Chem. Chem. Phys.*, 2015, **17**, 17797–17808.

30 Z. G. Hajos and D. R. Parrish, *J. Org. Chem.*, 1974, **39**, 1615–1621.

31 U. Eder, G. Sauer and R. Wiechert, *Angew. Chem., Int. Ed.*, 1971, **10**, 496–497.

32 B. List, R. A. Lerner and C. F. Barbas, *J. Am. Chem. Soc.*, 2000, **122**, 2395–2396.

33 R. B. Sunoj, *Wiley. Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 920–931.

34 M. M. Heravi, V. Zadsirjan, M. Dehghani and N. Hosseintash, *Tetrahedron: Asymmetry.*, 2017, **28**, 587–707.

35 T. D. Machajewski and C. H. Wong, *Angew. Chem., Int. Ed.*, 2000, **39**, 1352–1375.

36 Y. Nobakht and N. Arshadi, *J. Mol. Model.*, 2018, **24**, 334.

37 M. J. Ajitha and C. H. Suresh, *J. Mol. Catal. A: Chem.*, 2011, **345**, 37–43.

38 G. Yang, and L. Zhou, *Catal. Sci. Technol.*, 2016, **6**, 3378–3385.

39 W. Notz and B. List, *J. Am. Chem. Soc.*, 2000, **122**, 7386–7387.

40 S. Mukherjee, J. W. Yang, S. Hoffmann and B. List, *Chem. Rev.*, 2007, **107**, 5471–5569.

41 H. Yang, X. Zhang, S. Li, X. Wang and J. Ma, *RSC Adv.*, 2014, **4**, 9292–9299.

42 C. Agami, C. Puchot and H. Sevestre, *Tetrahedron Lett.*, 1986, **27**, 1501–1504.

43 L. Hoang, S. Bahmanyar, K. Houk and B. List, *J. Am. Chem. Soc.*, 2003, **125**, 16–17.

44 B. List, L. Hoang and H. J. Martin, *Proc. Natl. Acad. Sci. U. S. A.* 2004, **101**, 5839–5842.

45 K. N. Rankin, J. W. Gauld and R. J. Boyd, *J. Phys. Chem. A.*, 2002, **106**, 5155–5159.

46 G. Yang, Z. Yang, L. Zhou, R. Zhu and C. Liu, *J. Mol. Catal. A: Chem.*, 2010, **316**, 112–117.

47 M. Arnó and L. R. Domingo, *Theor. Chem. Acc.*, 2002, **108**, 232-239.

48 M. Arnó, R. J. Zaragozá and L. R. Domingo, *Tetrahedron: Asymmetry.*, 2005, **16**, 2764–2770.

49 C. Allemann, J. M. Um and K. Houk, *J. Mol. Catal. A: Chem.*, 2010, **324**, 31–38.

50 S. Bahmanyar and K. Houk, *J. Am. Chem. Soc.*, 2001, **123**, 12911–12912.

51 F. R. Clemente and K. Houk, *J. Am. Chem. Soc.*, 2005, **127**, 11294–11302.

52 M. Blanco, M. A. Pendás and E. Francisco, *J. Chem. Theory Comput.*, 2005, **1**, 1096–1109.

53 E. Francisco, M. A. Pendás and M. A. Blanco, *J. Chem. Theory. Comput.*, 2006, **2**, 90–102.

54 I. Cukrowski, *Comput. Theoret. Chem.*, 2015, **1066**, 62–75.

55 I. Cukrowski, F. Sagan and M. P. Mitoraj, *J. Comput. Chem.*, 2016, **37**, 2783–2798.

56 I. Cukrowski, D. M. E. van Niekerk and J. H. de Lange, *Struct. Chem*., 2017, **28**, 1429–1444.

57 B. Silvi, *J. Mol. Struct.*, 2002, **614**, 3–10.

58 L. R. Lavine and W. N. Lipscomb, *J. Chem. Phys.*, 1954, **22**, 614–620.

59 J. S. Miller, and J. J. Novoa, *Acc. Chem. Res.*, 2007, **40**, 189–196.

60 R. Ponec, G. Lendvay and J. Chaves, *J. Comput. Chem.*, 2008, **29**, 1387–1398.

61 W. Wang, Y. Kan, L. Wang, S. Sun and Y. Qiu, *J. Phys. Chem. C.*, 2014, **118**, 28746–28756.

62 J. H. de Lange, D. M. E. van Niekerk and I. Cukrowski, *J. Comp. Chem.*, 2018, **39**, 973–985.

63 I. Cukrowski, *Phys. Chem. Chem. Phys.*, 2019, **21**, 10244–10260.