SCUOLA
NORMALE
SUPERIORE

Classe di Scienze

Corso di perfezionamento in
Methods and Model for Molecular Sciences

XXXIII ciclo

Modelling Weak Interactions in the Gas Phase: From Rotational Spectroscopy to Reaction Rates Using Accurate Quantum-Chemical Approaches

Settore Scientifico Disciplinare **CHIM/02**

Candidata
dr.ssa Silvia Alessandrini

Relatrice

Prof.ssa Cristina Puzzarini

Supervisore interno

Prof. Vincenzo Barone

Anno accademico 2021/2022

List of Publications

1. D. A. Obenchain, L. Spada, S. Alessandrini, S. Rampino, S. Herbers, N. Tasinato, M. Mendolicchio, P. Kraus, J. Gauss, C. Puzzarini, J. -U. Grabow, V. Barone, "Unveiling the Sulfur–Sulfur Bridge: Accurate Structural and Energetic Characterization of a Homochalcogen Intermolecular Bond", *Angew. Chem. Int. Ed.*, **57**, 15822 (2018).

2. J. Wang, L. Spada, J. Chen, S. Gao, S. Alessandrini, G. Feng, C. Puzzarini, Q. Gou, J. -U. Grabow, V. Barone,"The Unexplored World of Cycloalkene–Water Complexes: Primary and Assisting Interactions Unraveled by Experimental and Computational Spectroscopy", *Angew. Chem. Int. Ed.*, **58**, 13935 (2019).

3. S. Alessandrini*, V. Barone, C. Puzzarini, "Extension of the "Cheap" Composite Approach to Noncovalent Interactions: The jun-ChS Scheme", *J. Chem. Theory Comput.*, **16**, 988 (2020).

4. S. Alessandrini*, V. Dell'Isola, L. Spada, V. Barone, C. Puzzarini, "A Computational Journey in the CH_2O_2S Land: an Accurate Rotational and Ro-vibrational Analysis of the Sulfene Molecule and the O,S-and O,O-Monothiocarbonic Acids", *Mol. Phys.*, **118**, e1766707 (2020).

5. C. Puzzarini, L. Spada, S. Alessandrini, V. Barone, "The challenge of non-covalent interactions: theory meets experiment for reconciling accuracy and interpretation", *J. Phys. Condens. Matter*, **32**, 343002 (2020).

6. J. Lei†, S. Alessandrini†, J. Chen, Y. Zheng, L. Spada, Q. Gou, C. Puzzarini, V. Barone, "Rotational Spectroscopy Meets Quantum Chemistry for Analyzing Substituent Effects on Non-Covalent Interactions: The Case of the Trifluoroacetophenone-Water Complex", *Molecules*, **25**, 4899 (2020).

7. S. Alessandrini*, F. Tonolo. C. Puzzarini, "In search of phosphorus in astronomical environments: The reaction between the CP radical ($X^2\Sigma^+$) and methanimine", *J. Chem. Phys.*, **154**, 054306 (2021).

8. V. Barone, S. Alessandrini, M. Biczysko, J. R. Cheeseman, D. C. Clary, A. B. McCoy, R. J. DiRisio, F. Neese, M. Melosso, C. Puzzarini, "Computational molecular spectroscopy", *Nat. Rev. Methods Primers*, **1**, 1 (2021).

9. S. Alessandrini*, M. Melosso, "Fate of the gas-phase reaction between oxirane and the CN radical in interstellar conditions", *Front. Astron. Space Sci.*, **8**, 159 (2021).

10. S. Alessandrini*, M. Melosso, N. Jiang, L. Bizzocchi, L. Dore, C. Puzzarini, "Conformational stability of cyclopropanecarboxaldehyde is ruled by vibrational effects", *Mol. Phys.*, **119**, e1955988 (2021).

11. J. Lupi[†], S. Alessandrini[†], V. Barone, C. Puzzarini, "junChS and junChS-F12 models: parameter-free efficient yet accurate composite schemes for energies and structures of non-covalent complexes", *J. Chem. Theory and Comput.*, **17**, 6974 (2021).

12. L. Bizzocchi, S. Alessandrini, M. Melosso, V. M. Rivilla, C. Puzzarini, "Ab Initio Study of Fine and Hyperfine Interactions in Triplet POH", *Molecules*, **27**, 302 (2022).

13. M. Melosso, L. Bizzocchi, H. Gazzeh, F. Tonolo, J.-C. Guillemin, S. Alessandrini, V. M. Rivilla, L. Dore, V. Barone, C. Puzzarini, "Gas-phase identification of (Z)-1,2-ethenediol, a key prebiotic intermediate in the formose reaction", *Chem. Commun.*, **58**, 2570 (2022).

14. X. Li, L. Spada, S. Alessandrini, Y. Zheng, K. G. Lengsfeld, J.-U. Grabow, G. Feng, C. Puzzarini, V. Barone, "Stacked but not Stuck: Unveiling the Role of $\pi \rightarrow \pi*$ Interactions with the Help of the Benzofuran–Formaldehyde Complex", *Angew. Chem. Int. Ed.*, **61**, 264 (2022).

15. F. Tamassia, L. Bizzocchi, M. Melosso, M. A. Martin-Drumel, O. Pirali, A. P. Charmet, E. Cané, L. Dore, I. E Gordon, J.-C. Guillemin, B. M. Giuliano, P. Caselli, S. Alessandrini, V. Barone, C. Puzzarini, "Synchrotron-based far-infrared spectroscopy of HC_3N: Extended ro-vibrational analysis and new line list up to 3360 cm^{-1}", *J. Quant. Spectrosc. Radiat.*, **279**, 108044 (2022).

16. N. Jiang, M. Melosso, L. Bizzocchi, S. Alessandrini, J.-C. Guillemin, L. Dore, C. Puzzarini "Spectroscopic and Computational Characterization of 2-Aza-1,3-butadiene, a Molecule of Astrochemical Significance", *J. Phys. Chem. A*, **126**, 1881 (2022)

17. V. M. Rivilla, L. Colzi, I. Jimenez-Serra, J. Martín-Pintado, A. Megías, M. Melosso, L. Bizzocchi, A. López-Gallifa, A. Martínez-Henares, S. Massalkhi, B. Tercero, P. de Vicente, J.-C. Guillemin, J. G. de la Concepcion, F. Rico-Villas, S. Zeng, S. Martín, M. A. Requena-Torres, F. Tonolo, S. Alessandrini, L. Dore, V. Barone, C. Puzzarini, "Precursors of the RNA-world in space: Detection of (Z)-1,2-ethenediol in the interstellar medium, a key intermediate in sugar formation", *Astrophys. J. Lett., in press https://doi.org/10.48550/arXiv.2203.14728*.

18. P. Recio†, S. Alessandrini†, G. Vanuzzo, G. Pannacci, D. Marchione, A. Caracciolo, V. Murray, P. Casavecchia, N. Balucani, A. Baggioli, C. Cavallotti, C. Puzzarini, "Intersystem-Crossing in the Entrance Channel without Heavy Atoms", *submitted*.

*As corresponding author
† These authors contributed equally

Abstract

During these four years of PhD, my research was focused on structural, energetic and spectroscopic characterisation of stable and reactive systems in the gas phase. A special focus has been put on the theoretical description of non-covalent interactions (NCIs) as they occur in the gas phase. The interest on these chemical bonds arises from the fact that they play a key role in many aspects of life. Indeed, they are responsible for the folding of proteins and characterise the shape of DNA and RNA. The same types of interactions drive self-assembling processes and the interaction between a receptor and its ligands. Furthermore, NCIs can influence chemical reactions by favouring one conformer with respect to others in a given pathway and they can also affect the structure of reactive intermediates and transition states. Usually, these phenomena are not due to a single interaction but to the sum of several hundreds (or thousands) non-covalent contacts occurring simultaneously and in a cooperative manner. Therefore, it is rather difficult to elucidate the type of interactions occurring and their effects on the molecular structures involved. However, one can aim at studying models of such weak bonds through the analysis of prototypical single NCIs occurring in an isolated environment. This idea can be exploited thanks to experimental methods based on rotational spectroscopy, which is an intrinsic high-resolution technique working exclusively in the gas phase, but also thanks to quantum chemistry. Rotational spectroscopy can unveil the interaction occurring in a binary system where two molecules interact and it is able to point out the effects of non-covalent interactions on the molecular structures of the fragments involved. On the other hand, quantum chemical simulations allow for: (i) exploration of the potential energy surface (PES) of the bimolecular system, thus identifying all the possible isomers that can arise from the contacts of two fragments, also in the case of reactive PESs; (ii) accurate energetic studies and decomposition of the energy to unveil the nature of the interaction; (iii) providing *ab initio* data useful to guide the interpretation of experimental spectra, which can be difficult to analyse due to several factors. Currently, state-of-the-art information on non-covalent complexes are obtained via a strong interplay of rotational spectroscopy and quantum chemistry. However, computational simulations show some limitations due to the challenge of accurately describing NCIs; indeed, they are extremely sensitive to the level of theory employed and an effective compromise between accuracy and computational cost is

always difficult. In this context, my PhD thesis aimed at developing accurate computational models to treat medium-sized systems (20-30 atoms) dominated by NCIs such as intermolecular (binary) complexes in the gas phase, to support and/or complement experimental rotational spectroscopy, as well as reactive intermediates and transition states, to accurately describe reaction pathways. The developed models are able to provide reliable estimates for both energies and geometries. These approaches are based on coupled-cluster techniques including single and double excitations and a perturbative treatment of triples (CCSD(T) method). To reduce the computational cost without degrading the accuracy, the CCSD(T) method is employed in conjunction with Møller-Plesset perturbation theory to the second order (MP2) to account for the extrapolation to the complete basis set (CBS) limit and the core-valence correlation effects. Standard methods and their explicitly correlated counterparts, i.e., CCSD(T)-F12 and MP2-F12, have been employed. The thesis will describe how these new computational models have been built based on accurate reference data reported in the literature and the fundamental role played by diffuse functions in the basis sets. Then, the focus will move on the discussion of several examples where the new computational models, namely junChS and junChS-F12, have been used to characterise energies and structures of non-covalent complexes such as the complex between sulfur dioxide and dimethyl sulfide (SO_2-DMS), the benzofuran-formaldehyde complex (C_8H_6O-H_2CO), and the trifluoroacetophenone-water ($CF_3COC_6H_5-H_2O$) complex. The thesis will also address the performances of the junChS model in the case of astrochemically relevant reaction pathways, where the energetic barriers play a key role in establishing reliable reaction rate coefficients. Also in this case, a few examples will be given considering the reactions between methanimine (CH_2NH) and the CP radical, between oxirane (c-C_2H_4O) and the CN radical, and between propene (C_3H_6) and the C_3N radical. In these cases, the developed models had to provide accurate energies, not only for closed-shell species (all paired electrons), but also for structures with an unpaired electron (doublet state, open-shell), which are troublesome electronic configurations to describe from the theoretical point of view.

Contents

Chapter 1

Introduction

The topic of this *Philosophiæ Doctor* (PhD) Thesis is the accurate modeling of weak interactions, a broad topic with some of the aspects due to/or causing non-covalent Interactions (NCIs) that are not fully understood yet. Still, these molecular contacts are essential as they shape the world surrounding us in a nearly unnoticeable, but substantial way. The different hints of a non-covalent bond are difficult to fit in the present chapter which aims at explaining the essence of the *ab initio* chemical modelling of MCIs and, at the same time, at giving an introduction on the works that will be presented in the other chapters.

In general terms, a non-covalent system is identified by the presence of electric multipole–electric multipole interactions; including in this definition not only permanent, but also induced, and time-dependent multipoles [1]. Thus, this definition includes (*i*) Hydrogen bonds (H-bond), (*ii*) van der Waals (vdW) interactions and (*iii*) all the other mixed interactions occurring between species that do not have a permanent charge in the non-covalent system, such as π-interactions, chalcogen bonds, and London forces. While this definition is quite broad, it is a clear representation of the fact that NCIs are very different in their nature [1, 2]. However, a key point worth of notice is that non-covalent bonds usually operate in a cooperative manner. Indeed, their tangible effects are usually not due to a single interaction, but to a large amount of NCIs occurring simultaneously so that the total strength of these bonds is amplified [3]. For example, adhesion capability of Geckos are due to the (very) large amount of spatulas of the Gecko's soles that increases the contact with the surface and allows a large number of vdW interactions [4, 5]. This natural phenomenon is of inspiration in several fields, from the development of climbing robots, to the development of medical adhesive tapes and, more broadly, whenever an adhesive surface is required [6]. Another simple example involves the structure assumed by frozen water (ice). Here, the possibility to form a large amount of H-bonds between H_2O molecules leads to a well-organised structure where a very little amount of air remains trapped. Thanks to this, ice is able to float on water [7, 8]. The coexistence of different types of NCIs is also extremely common in nature, and a typical example in this regard is provided by the

structures of bio-molecules. For example, proteins are folded to minimise the interaction between the hydrophobic portions of their chain and hydrophilic environment surrounding them. This is accomplished in a two-step procedure where the proteins first maximise the intramolecular H-bonds that can occur between amino-acids close to each other in the chain. Then, the structure further folds to allow interactions between parts of the chain that are far apart but have a chemical affinity, thus leading to a large set of NCIs like $\pi - \pi$ stacking. [9]. The result is then structures with hydrophilic side-chains of the protein on the external surface and a hydrophobic core on the inner side. This process also involves the so-called hydrophobic effect, which is sometimes considered a NCI, but is a merely equilibrium between enthalpy and entropy [10]. It should also be mentioned that this process can be reversed in the case that membrane proteins are surrounded by a hydrophobic environment due to the tails of phospholipids. A large number of other examples could be made to underline the importance of NCIs, starting from how they drive catalysis, crystal growth, and self-assembling processes to how they influence conformational arrangements of simple chemical species, but also molecular recognition mechanisms [11–18]. Therefore, it seems that evolution has somewhat built an efficient machine to maximise these weak interactions able to confer stability to chemical species and also processes ensuring a certain flexibility and reversibility.

Since NCIs play such an important role in several chemical fields, their characterisation either theoretical or experimental is of great relevance and their interplay is also of fast-growing interest [19]. As template models to study NCIs, one can use non-covalent (nc)-complexes, where two (or more) species are bonded by weak interactions which can be investigated in the gas phase, a matrix- and solvent-free environment. Since the present thesis will only consider binary systems, the term non-covalent complex will implicitly indicate only complexes involving two molecules [20].

Aiming at an experimental investigation of nc-complexes, several spectroscopic techniques are available, which are based on the analysis of vibrational and/or rotational spectra, and thus on the determination of the corresponding energy levels [1,2]. While structural and energetic information is potentially achievable from experimental measurements, several constrains prevent from the full experimental determination of such quantities, the exception being extremely small nc-complexes of very limited interest from the biological point of view [19,20].

Still, the effects that NCIs induce on the two isolated molecules upon non-covalent bonding, especially in medium- to large-sized clusters, is a mandatory starting point for achieving a better description of bio-molecules and biological processes, but also to improve — for example — drug design or crystal engineering and to provide better estimate of physical-chemical proprieties of condensed state systems and of newly developed materials [1,19,21–23]. For these reasons, in recent years, the *ab initio* study of NCIs has gained an increasing interest and several efforts have been made to develop

methodologies able to provide the identification and description of such NCIs [19,24]. This thesis will present two new approaches for the accurate determination of energetic proprieties and structural parameters for medium-sized molecular complexes: the junChS approach and its explicitly correlated variant, junChS-F12 [25,26]. These approaches are composite schemes; thus, they include various contributions, each computed at the best level of theory possible according to the size of the system, with the final goal of maximising the accuracy reachable. The name of these schemes derive from the word "cheap" that stresses the cost-effectiveness of the approach. This scheme is generally abbreviated as "ChS" and was introduced to describe medium-sized molecules [27–29]. The extension of this approach to nc-complexes has led to the so-called junChS developed by the present author, where the prefix "jun-" indicates the use of the jun-cc-pVnZ seasonal basis sets by Papajak et al. [30] within the ChS approach. Traditionally, these schemes are based on methods that include the correlation energy via a definition of a multi-determinal wave function, such as those based on coupled-cluster theory and perturbation theory [31]. However, in order to reduce the intrinsic errors related to these theoretical approaches, a variant of the junChS scheme based on explicitly correlated methods has been developed, i.e. the junChS-F12 model. These methods introduce directly the inter-electronic distance into the wave function, thus improving the overall accuracy of the approach [32]. The theoretical background behind the development of such schemes will be briefly described in Chapter 2 and a full account on the development of the two schemes together with the statistical analysis of their performance for energies and structural parameters with respect to an accurate data set available in the literature will be reported in Chapter 3.

As the thesis presents the junChS and junChS-F12 schemes as powerful methodologies to describe nc-complexes as well as isolated systems, allowing an accurate investigation of the effects due to non-covalent bonding, applications and examples are also in order. These will be the focus of the last part of the thesis (Chapter 4), which will deal with the application of the junChS scheme on test cases that have been encountered during my four years of PhD. In particular, the applications explored can be classified in two large sub-groups:

- Investigation of nc-complexes to support and complement experimental measurements carried out using high-resolution microwave spectroscopy.

- Refinement of electronic energies for reactions that are of relevant interest in astrochemistry

Each work presented in Chapter 4 will have its own specific introduction but a general background for such applications is required.

As already mentioned, experimental measurements can be useful to derive energetic and structural properties of nc-complexes in the gas phase, and the experimental

technique of reference in such cases is rotational spectroscopy, especially when performed in supersonic expansion [20, 33, 34]. The latter is a high-resolution technique that can distinguish between different conformers, but also between isotopologues of the same species. On a general basis, the rotational transitions of a nc-complex lie in the microwave frequency-region and their assignment in terms of an effective Hamiltonian leads to the corresponding rotational parameters, i.e. Rotational Constants (RCs) of the vibrational ground state, centrifugal distortion constants and nuclear quadrupole coupling constants [33, 35, 36]. The peculiarity of this spectroscopic technique is that RCs directly depend on the atomic composition and the geometry of observed system, thus potentially allowing the derivation of the vibrationally averaged molecular structure [33, 37]. In this regard, nuclear quadrupole coupling constants are also important as they hold information on the electronic arrangement around the quadrupolar nuclei, thus being different for different conformers [37, 38]. However, difficulty in deriving molecular structures from purely experimental RCs increases with increasing size of the system. As a consequence, it is only feasible for small molecules, far from biological complexity [35,37,39]. Another complication in this procedure is the assignment of the rotational transitions which is, in the case of nc-complex, particularly difficult. Indeed, these bi-molecular clusters have to be formed *in situ*, often with a poor efficiency, and thus their concentrations are low and consequently the signals weak. Their intensity can also decrease because of (*i*) hyperfine couplings that split a single transition in more components and (*ii*) the presence of more than one isomer due to the flexibility of the nc-complex. Lastly, the spectrum observed in these cases is further complicated by the signals due to the presence in the experimental set-up of monomers as well as nc-complexes different from that of interest. All these facts lead to very congested spectra that could be misassigned [19, 20, 34].

This complicated picture drastically changes if theoretical studies are taken into account. Indeed, quantum-chemical simulations can be used to compute *ab initio* rotational parameters with great accuracy, providing an optimal initial guess of the rotational transitions that should be searched in the experimental spectrum. Theory can also be used to explore the Potential Energy Surface (PES) of a nc-complex, thus elucidating the possible conformers and their relative stability, consequently deriving the structures that could be experimentally observed. Furthermore, quantum-chemical computations are also of great help for the derivation of experimentally-based molecular geometries via the semi-experimental (SExp) approach [35, 37, 40].

The thesis will present a strong interplay between theory and high-resolution microwave spectroscopy to investigate molecular complexes of particular importance, thereby adopting the junChS and/or junChS-F12 approaches. The works addressed will involve the search of possible isomers and the computational spectroscopic characterisation of the most abundant, i.e. stable, ones. The SExp equilibrium structure of such nc-complexes will be then derived by employing also the so-called template

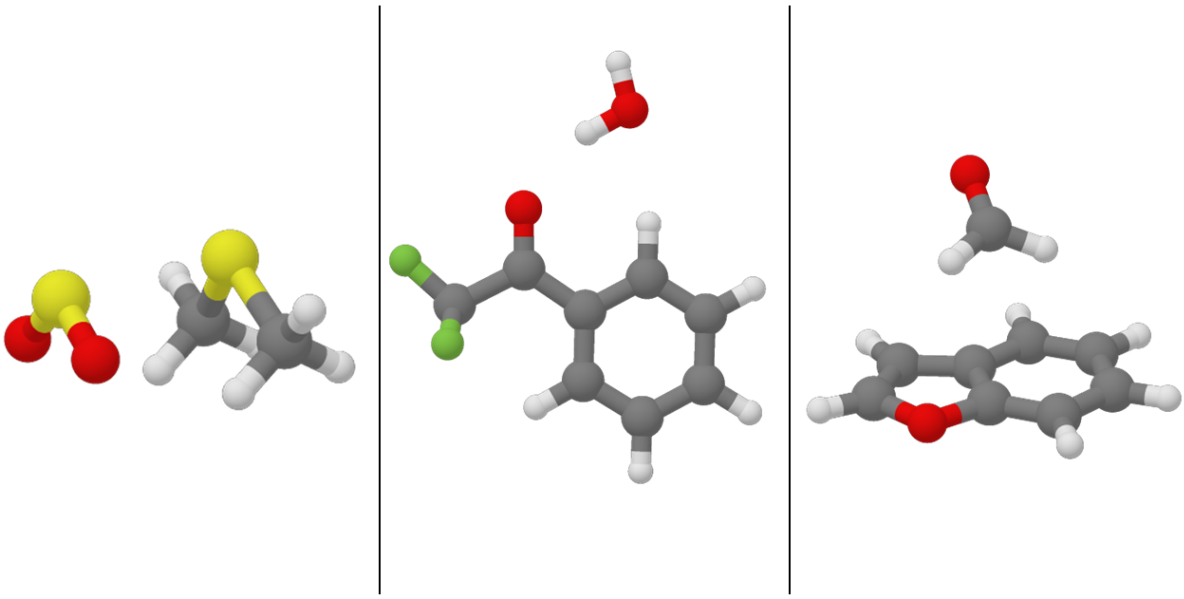


Figure 1.1: The three non-covalent complexes chosen as test cases in the present thesis.

approach together with the experimental parameters [26, 41]. To cover different types of NCIs, three different test cases, reported in fig. 1.1, have been chosen as examples:

- The complex formed by sulfur dioxide and dimethyl sulfide ($SO_2\cdots DMS$), where a genuine $S\cdots S$ non-covalent bond (chalcogen bond) was first unveiled via the joint approach of theory and rotational spectroscopy [42];

- The cluster formed by trifluoroacetophenone and water ($CF_3COC_6H_5\cdots H_2O$) that is characterised by a H-bond and also a weak $C-H\cdots O$ interactions. This complex also allows for the investigation of the affects due to fluorination with respect to the acetophenone-water complex [43].

- The benzofuran-formaldehyde complex (c-$C_8H_6O\cdots H_2CO$), where a $\pi \longrightarrow \pi^*$ interaction stabilises the complex and is favoured among other possible interactions, such as H-bonds and lone-pair interactions [44];

The second type of application for the composite schemes developed in this thesis involves the computation of accurate reaction pathways of astrochemical interest. Astrochemistry is a recently developed research field that aims at understanding the chemical evolution of the Interstellar Medium (ISM), i.e. the matter and radiation between the star systems in a galaxy [45]. Starting from diatomic species in 1930, molecules of increasing chemical complexity have been observed in the ISM [45, 46]. The molecules detected up to date are reported in table 1.1 and comprehend aromatic [47] and chiral species [48] together with bi-functional molecules such as glycolonitrile [49], ethanolamine [50], and Z-1,2-ethendiol [51, 52].

Table 1.1: Molecules detected in the interstellar medium and circumstellar shells as of March 2022. In bold orange the molecules first detected in 2021, while in bold blue those first detected in 2022.

2 atoms	3 atoms	4 atoms	5 atoms	6 atoms	7 atoms	8 atoms	9 atoms	10 atoms	11 atoms	12 atoms	>12atoms
H_2	C_3	c-C_3H	C_5	C_5H	C_6H	CH_3C_3N	CH_3C_4H	CH_3C_5N	HC_9N	c-C_6H_6	C_{60}
AlF	CCH	l-C_3H	C_4H	l-H_2C_4	CH_2CHCN	$HCOOCH_3$	CH_3CH_2CN	$(CH_3)_2CO$	CH_3C_6H	n-C_3H_7CN	C_{70}
AlCl	CCO	C_3N	C_4Si	C_2H_4	CH_3C_2H	CH_3COOH	$(CH_3)_2O$	$(CH_2OH)_2$	C_2H_5OCHO	i-C_3H_7CN	C_{60}^+
C_2	CCS	C_3O	c-C_3H_2	CH_3CN	HC_5N	C_7H	CH_3CH_2OH	CH_3CH_2CHO	CH_3OCOCH_3	$C_2H_5OCH_3$	c-C_6H_5CN
CH	CH_2	C_3S	l-C_3H_2	CH_3NC	CH_3CHO	C_6H_2	HC_7N	CH_3CHCH_2O	CH_3COCH_2OH	**1-c-C_5H_5CN**	$HC_{11}N$
CH^+	HCN	C_2H_2	H_2CCN	CH_3OH	CH_3NH_2	CH_2OHCHO	C_8H	CH_3OCH_2OH	**c-C_5H_6**	**2-c-C_5H_5CN**	**1-$C_{10}H_7CN$**
CN	HCO	NH_3	CH_4	CH_3SH	c-C_2H_4O	l-HC_6H	CH_3CONH_2	**c-C_6H_4**	**$HOCH_2CH_2NH_2$**	**CH_3C_7N**	**2-$C_{10}H_7CN$**
CO	HCO^+	HCCN	HC_3N	C_3NH^+	H_2CCHOH	CH_2CHCHO	C_8H^-	**H_2CCCHC_3N**			**c-C_9H_8**
CO^+	HCS^+	$HCNH^+$	HCCNC	HCCCHO	C_6H^-	CH_2CCHCN	C_3H_6	**C_2H_4NCO**			**1-c-C_5H_5CCH**
CP	HOC^+	HNCO	HCOOH	NH_2CHO	CH_3NCO	H_2NCH_2CN	CH_3CH_2SH				**2-c-C_5H_5CCH**
SiC	H_2O	HNCS	H_2CNH	C_5N	HC_5O	CH_3CHNH	CH_3NHCHO				
HCl	HNC	$HOCO^+$	H_2C_2O	l-HC_4H	$OHCH_2CN$	CH_3SiH_3	HC_7O				
KCl	HNO	H_2CO	H_2NCN	l-HC_4N	HCCCHNH	H_2NCONH_2	**HCCCHCHCN**				
NH	MgCN	H_2CN	HNC_3	c-H_2C_3O	HC_4NC	$HCCCH_2CN$	**H_2CCHC_3N**				
NO	MgNC	H_2CS	SiH_4	H_2CCNH	**c-C_3HCCH**	HC_5NH^+	**$H_2CCCHCCH$**				
NS	N_2H^+	H_3O^+	H_2COH^+	C_5N^-	**l-H_2C_5**	**CH_2CHCCH**					
NaCl	N_2O	c-SiC_3	C_4H^-	HNCHCN	**MgC_5N**	**MgC_6H**					
OH	NaCN	CH_3	HNCNH	SiH_3CN	**CH_2C_3N**	**$C_2H_3NH_2$**					
PN	OCS	C_3N^-	CH_3O	C_5S							
SO	SO_2	PH_3	NH_4^+	MgC_4H							
SO^+	c-SiC_2	HCNO	H_2NCO^+	**CH_3CO^+**							
SiN	CO_2	HOCN	$NCCNH^+$	**C_3H_3**							
SiO	NH_2	HSCN	CH_3Cl	**H_2C_3S**							
SiS	H_3^+	H_2O_2	MgC_3N	**HCCCCHS**							
CS	SiCN	C_3H^+	NH_2OH	**C_5O**							
HF	AlNC	HMgNC	HC_3O^+	**C_5H^+**							
HD	H_2S	HCCO	HCOCN								
O_2	SiNC	CNCN	**H_2C_2S**								
CF^+	HCP	HONO	**C_4S**								
SiH	CCP	MgC_2H	**HCOSH**								
PO	AlPH	**HCCS**	**HCSCN**								
AlO	H_2O^+	**HNCN**	**HCCCO**								
OH^+	H_2Cl^+	**H_2NC**	**HC_3S^+**								
CN^-	KCN	**$HCCS^+$**									
SH^+	FeCN										
SH	OH_2										
HCl^+	TiO_2										
TiO	CCN										
ArH^+	Si_2C										
N_2	HS_2										
NO^+	HCS										
NS^+	HSC										
HeH^+	NCO										
FeO	CaNC										
	NCS										

Many of the species observed in the ISM have a prebiotic interest, which means that are precursors of bio-molecule building-blocks such as amino-acids and nucleobases [53]. This somewhat supports the exogenous delivery as theory for the origin of life on Earth. While this theory affirms that the bio-molecule precursors were delivered on Earth by meteorites and comets, the rough path that has led to the formation of life is most probably the result of several mechanisms, combining exogenous and endogenous phenomena [54, 55]. The observation of molecules in the ISM is based on the laboratory measurement of their rotational transitions. As already mentioned in the previous section, they are extremely specific and provide a reliable way to claim the presence of molecules in the ISM [45, 46, 56]. In this context, quantum-chemical computations are useful to simulate the accurate rotational transitions of exotic species and/or isotopologues that are difficult to obtain in the experimental set-up, due to the instability of the molecule/precursor [35, 57]. The comparison between computed

and experimental spectra collected by space-probes and radio-interferometers can allow the assignment of unknown lines and the observation of exotic species such as C_5N^- [57, 58]. The role of theory in astrochemistry is not limited to spectra simulations, but also involves the *ab initio* investigation of the reactive PESs for the formation of new and more complex species that could be detected or have already been observed [57]. The ISM is characterised by particular physical conditions: very low temperatures, ranging from 10 to 200 K depending on the region considered, and very low number density, which implies low probability of collisions between reactive species [45]. Therefore, astrochemical reactions must be simulated nearly in absence of thermal energy, only considering reactants that should lead to efficient collisions, with a high probability to evolve in products. From a quantum-chemical simulation point of view, this translates into exploring the reactive PES of those reaction mechanisms that are exothermic and characterised by submerged barriers with respect to the reactants. This latter constrain implies at least one reactive species, such as a radical and/or an ion [45, 59]. Once the interesting paths and the relative products are outlined, the following step is two simulate the kinetics of the total reaction. In this case, one has to search for the low-pressure limit of the reaction and run several simulations at different temperatures to understand the T-dependence of the reaction and extrapolate the rate constants to values outside the experimentally reachable range, such as those between 0 and 10 K, for which quantum-chemical programs generates numerical instabilities and singularities [59, 60]. These simulations are typically used directly to build up chemical networks connecting species observed in the ISM that should be able to somewhat reproduce the relative abundances derived from experimental observation [61]. A drawback of the computational approaches is the accuracy of the methodology, which should be able to provide accurate estimates of the reaction barriers for species with an open-shell character that have also loose bonds, such as the transitions states [57, 59]. The accuracy is mainly needed to verify the exothermic nature of the reaction pathway with respects of the reactants, which is used as discriminating factor for inclusion/exclusion of a reaction mechanism, and to accurately determine the reaction barriers. For this reason, the junChS scheme has been chosen in the present thesis to evaluate the electronic energies of the reactive PESs considered, which are:

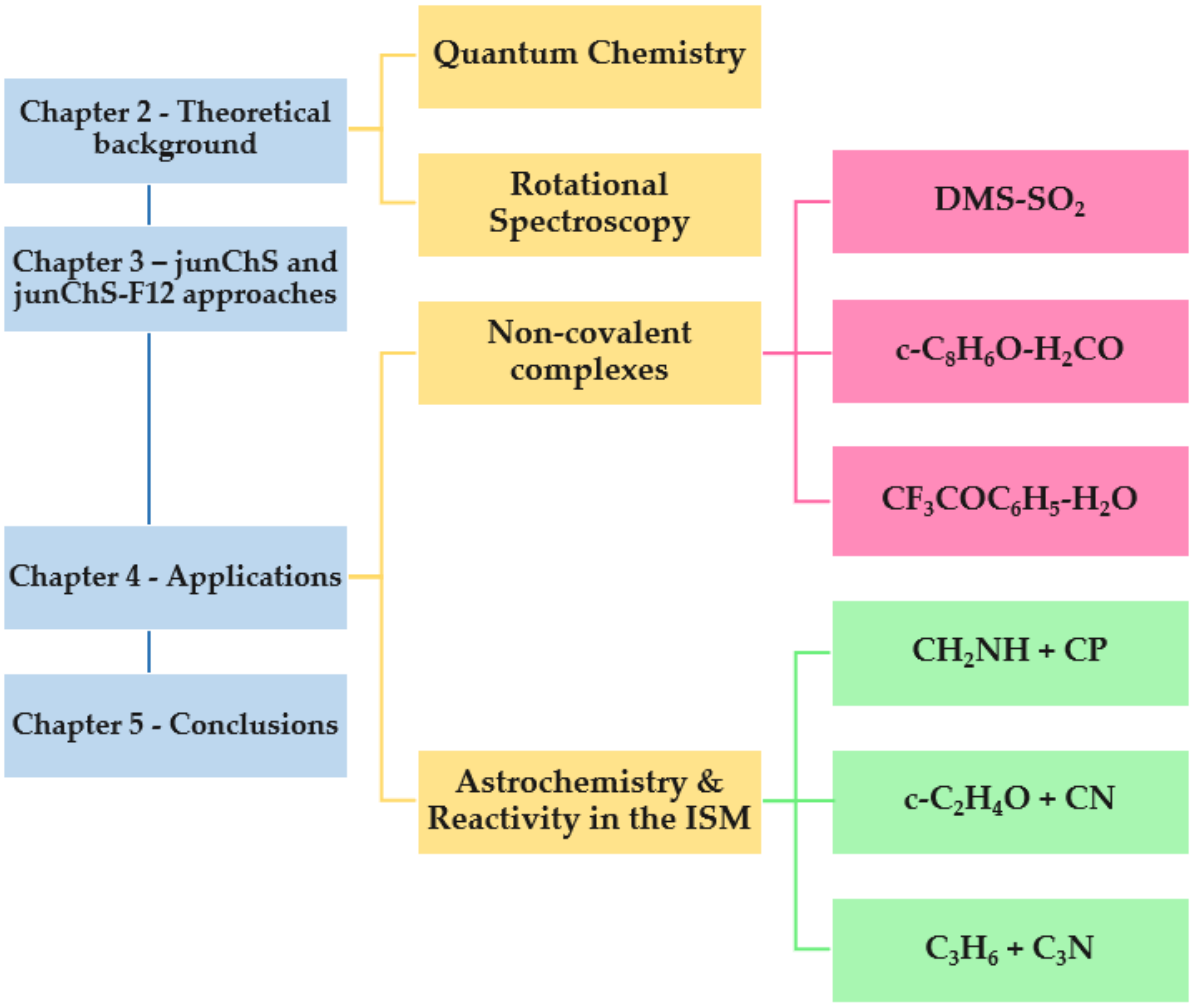
- The reaction between the CP radical and methanimine (CH_2NH). This was chosen for its chemical similarity with the reactions between methanimine and the CN and CCH radicals, which are plausible gas-phase formation routes for cyanomethanimine and propargylimine, respectively. Since, both species have been observed in the ISM, its possible that the products of the CP+CH_2NH reaction could be observed in the ISM. This led to the computational spectroscopic characterisation of the products, i.e. HNCHCP and CH_2NCP [62].

- The pathways involved in the oxirane (c-C_2H_4O) + the CN radical reaction. This is a promising mechanism for the formation of isomers belonging to the C_3H_3NO family of compounds. This family represents an interesting case in astrochemistry, since none of its isomers has ever been detected in the ISM. The work aims at pointing out which is the most abundant product of this family for the reaction above, thereby combining thermodynamic and kinetic studies [63].

- The formation route of c-C_5H_5CN starting from propene (C_3H_6) and the C_3N radical. The interest here revolves around 1- and 2-cyanocyclopentadiene and vinylcyanoacetylene, which have been detected in the ISM in 2021. Although a formation route for such species has been already reported, the work will report a more detailed exploration of the reactive PES, starting from two common molecules in the ISM, propene and the cyanoethynyl radical (see Table 1.1).

It should also be mentioned that experimental techniques can be employed to study astrochemical reactions such as molecular crossed beams and the CRESU technique [64–68]. Both of them are considered powerful experimental methodologies to study the reactions of astrochemical interest, but they are not able to reproduce the real ISM conditions. The CRESU technique, where the acronym stands for "Cinétique de Réaction en Ecoulement Supersonique Uniforme" (i.e. reaction kinetics in uniform supersonic expansion) is able to work at temperature typical of the ISM, but using higher pressures, therefore more collisions are typically involved together with the "three-body stabilitazion" effect. On the contrary, molecular crossed beams are able to reproduce the collison-free condition of the ISM, but thermal energy is present in the collimated beam because of the working temperature (usually room temperature).

Now that the general background of this thesis have been outlined, a short and schematic representation can be useful to the reader to understand the scaffold of the thesis. The flowchart reported in the following shows the overall organisation of the thesis with the main topics that will encountered in each Chapter, starting from the theoretical background (Chapter 2) and proceeding through the different applications to the conclusions (Chapter 5).


Chapter 2 - Theoretical background
Chapter 3 – junChS and junChS-F12 approaches
Chapter 4 - Applications
Chapter 5 - Conclusions
Quantum Chemistry
Rotational Spectroscopy
Non-covalent complexes
Astrochemistry & Reactivity in the ISM
DMS-SO2
c-C8H6O-H2CO
CF3COC6H5-H2O
CH2NH + CP
c-C2H4O + CN
C3H6 + C3N

Chapter 2

Theoretical Background

In the following, the theoretical foundations for the works reported in the Results sections (Chapters 3 and 4) are described. Different quantum-chemical methods, ranging from Density Functional Theory (DFT) to Coupled Cluster (CC) techniques and F12 methodologies, have been extensively used to describe NCIs, but the full theoretical treatment of such methods is out of the scope of the present thesis. Thus, the next section focuses mainly on the accuracy issuing from the aforementioned theories and on the rationale behind the development of composite schemes. The second part of the chapter discusses the basic concepts of rotational spectroscopy that are required to understand the spectroscopic characterisation of nc-complexes and the *ab initio* calculations of the rotational parameters, and how experiment and theory can complement each other to derive the structure of nc-complexes.

2.1 Quantum chemistry

In quantum chemistry, the accurate description of molecules from *ab initio* considerations is entirely based on the choice of the level of theory, which is defined as the combination of the quantum-chemical model used for the description of the wave function and the basis set employed. For each system, the computational chemist has to define the best computational level and this essentially means finding the right compromise between accuracy and computational cost. The accuracy of a level of theory is influenced by different error sources and the main goal is to minimise them without using an extended amount of computer resources. For this reason, the issues of basis set completeness and inclusion of many-body effects will be addressed. This thesis will not report a detailed theoretical treatment of the methodologies employed, but mainly the rationale behind the calculation of the electronic energy and the computational cost, highlighting pros and cons. Regarding the basis sets, the thesis will discuss how these are generally built and how an hierarchical series of basis sets can be used to extrapolate electronic energies and molecular properties. This will open the route towards the composite schemes.

2.1.1 Theoretical Models

In quantum chemistry, the molecular description starts from the resolution of the Schrödinger equation (SE) or the Kohn-Sham (KS) equation in the case of DFT. Since all the methodologies involved in the composite schemes are wave-function based, the focus will mainly be on the former equation, with DFT methodologies being addressed at the end of this section.

In the time-independent form of SE (2.1), the complete set of molecular energies $E(r, R)$ depending on electronic (r) and nuclear (R) coordinates is obtained considering the time-independent molecular Hamiltonian $\mathbf{H}(r, R)$ and the wave function $\Psi(r, R)$.

$$\mathbf{H}(r, R)\Psi(r, R) = E(r, R)\Psi(r, R) \quad . \tag{2.1}$$

The Hamiltonian $\mathbf{H}(r, R)$ term incorporates (*i*) the electronic and nuclear kinetic energy operators, (*ii*) the attractive potential between nuclei and electrons and, (*iii*) the repulsive potential between nuclei and that between electrons. The first step to solve eq. (2.1) is to decouple the nuclear and electronic motions. To this end, the most effective approach is rooted in the Born-Oppenheimer (BO) approximation that leads to the formal partition of the molecular Hamiltonian, and thus of the SE, into a nuclear and electronic part. In this manner, the total energy ($E(r, R)$) of eq. (2.1) is obtained in a two-step procedure. First, one solves the electronic SE for a fixed set of nuclear coordinates, obtaining the corresponding eigenvalues. The resolution of the electronic SE for different sets of nuclear arrangements leads to an hyper-surface of ($3N$-6) internal coordinates. This is the potential energy surface (PES) whose two or three dimensional cuts are often shown. The PES contains several points of chemical interest, among which we find the minima, which are the stable chemical structures (either structural isomers and diastereomers, or reactants, products, and intermediates), and the transition states, i.e. a local maximum of the energy with respect to one of the $3N$-6 internal coordinates. The PES is then used to solve the nuclear SE from which, for example, the rotational and vibrational energies are obtained. To solve both the electronic and nuclear SEs, the wave function $\Psi(r, R)$ needs to be known, which is not generally the case for a molecular system. For the electronic part, the problem is addressed by providing an initial guess that is refined during the computational procedure. This guess is based on chemical intuition and usually involves Atomic Orbitals (AOs) represented by mathematical functions or combinations of functions. The number and type of functions used to built the AOs of an atom is defined by the *basis set* employed. A linear combination of AO is then defined to construct the Molecular Orbitals (MOs), which are employed to build the the initial guess of the wave function by considering the electron occupation dictated by the Pauli's exclusion principle. The choice of the basis set strongly influences the quality of the results and, in principle, the number of functions employed should be infinite, which is not feasible in practice. This leads to

the Basis Set Error (BSE) and several procedures to estimate the *infinite/complete basis set limit* have been developed. In the following discussion, we will assume that the energy is not affected by the BSE and more details on this error will be presented later in the text.

Having described in broad terms how the wave function is built, the major issue of the electronic SE remains. Indeed, the bi-electronic repulsion potential ($V_{ee} \propto 1/r_{12}$) appearing in the $\mathbf{H}(r, R)$ term, leads to a non-exactly solvable SE. Different approaches have been developed to obtain an *approximate solution* of this equation, with the simplest option being to consider each electron surrounded by a mean repulsive potential field due to the presence of all the other electrons. This is the so-called Hartree-Fock (HF) method, where the operator used to obtain the electronic energy (E_{HF}) is expressed using only mono- and pseudomono-electronic terms and is known as Fock operator. The latter is expressed as follows:

$$\mathbf{F}_i = \mathbf{h}_i + \sum_{i}^{N_{elec}} (\mathbf{J}_i - \mathbf{K}_i) \quad , \tag{2.2}$$

where the mono-electronic $\mathbf{h}_i$ operator accounts for the kinetic energy of the electrons as well as their attraction to the nuclei. The $\mathbf{J}$ and $\mathbf{K}$ operators are denoted as the Coulomb and Exchange operators, respectively, and depend on the two-electron operator ($\mathbf{g}_{ij}$) which represent the electron-electron repulsion contribution:

$$\begin{aligned} \mathbf{J}_i|\psi_j(2)\rangle &= \langle\psi_i(1)|\mathbf{g}_{12}|\psi_i(1)\psi_j(2)\rangle \\ \mathbf{K}_i|\psi_j(2)\rangle &= \langle\psi_i(1)|\mathbf{g}_{12}|\psi_j(1)\psi_i(2)\rangle \quad , \end{aligned} \tag{2.3}$$

where the ψ_j and ψ_i are two MOs. Therefore, the actual form of the Fock operator depends on the MOs and an initial guess of the wave function (Ψ_{HF}) is thus required to build up the pseudo mono-electronic operators $\mathbf{J}$ and $\mathbf{K}$, and derive the energy. Using the variational principle [69], the coefficients of the wave function are optimised to compute the lowest energy possible, thus defining a new wave function that can be used to re-build the pseudomono-electronic operators and re-computed the energy. This iterative procedure is denoted as "Self Consistent Field" (SCF) and is repeated until the convergence criteria, established at the beginning of the procedure, are met. The final wave function represents the ground state electronic configuration of the molecular system and is a single Slater-determinant [69, 70]. The E_{HF} usually recovers the 99% of the real energy for a molecular system [71]. Surprisingly, this high percentage is not enough for a quantitative (and accurate) description and the remaining energy, denoted as *correlation energy* (E_{corr}), has to be included. The mean-field approximation is not a good representation of the electronic picture. Indeed, for a given electron the probability to find another electron in its proximity is lower than the mean average and this probability becomes even lower if the two electrons have the same spin compared

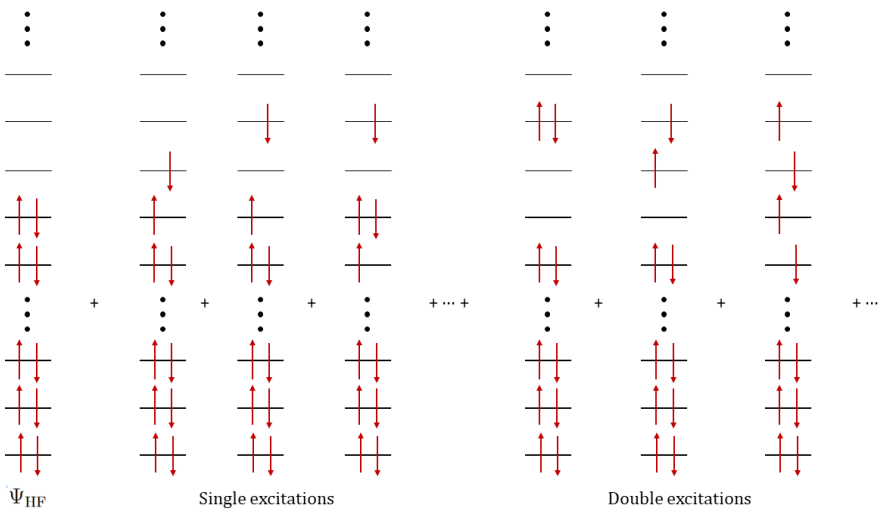


Figure 2.1: Schematic representation of a multi-determinants wave function containing single and double excitations together with the the HF reference.

to opposite spin. This leads to the definition of the so-called Fermi and Coulomb *holes* for opposite and same spin electrons, respectively. Thus, the E_{corr} can be divided in two contributions, one due to electrons with the same spin (Coulomb correlation) and another due to the opposite spin correlation (Fermi correlation), with the former being the larger term as a sequence of intra- and inter-orbitals interactions [71].

The easiest way to recover the correlation energy is to set-up a more flexible wave function including more than one Slater-determinant, which means, in practical terms, to consider that electrons are not fixed in specific MOs, thus opening the route to other (excited) configurations, as shown in fig. 2.1. Since, the HF method provides the best mono-determinant wave function, its use as starting point for the inclusions of other determinants seems reasonable. The present thesis will deal in particular with two correlated methods: Møller-Plesset (MP) Perturbation Theory (PT) and CC methodologies [31, 72]. When multi-determinant wave functions are introduced one has to define which electrons can generate the excited configurations. Usually, one refers to *frozen-core* (fc) computations when only valence electrons are considered, while the notation *all electrons* (ae) indicates that also the inner electrons are being excited. Since more electrons imply more configurations to include, ae computations are intrinsically more computationally demanding.

MP-PT is a specific type of many-body PT based on the idea that the real solution is close to that of an ideal system for which exact solutions are known, thus meaning that only a small perturbation needs to be introduced. In PT, the real Hamiltonian is defined as the sum of an unperturbed (zero-order) Hamiltonian ($\mathbf{H}_0$) and a perturbation term ($\mathbf{H}'$) [71,73]. In MP-PT, the $\mathbf{H}_0$ is the sum of the pseudomono-electronic Fock operators and the $\mathbf{H}'$ is equal to the fluctuation potential of eq. (2.4), where $\mathbf{V}_{ee}$ and $\langle \mathbf{V}_{ee} \rangle$ are the exact and mean repulsion potentials, respectively [71,72]:

$$\mathbf{H}' = \mathbf{H} - \mathbf{H}_0 = \mathbf{V}_{ee} - 2\langle \mathbf{V}_{ee} \rangle \quad . \tag{2.4}$$

The sum of $\mathbf{H}_0$ and $\mathbf{H}'$ is the exact Hamiltonian $\mathbf{H}$ [71, 72]. The zero-order wave function is Ψ_{HF} and the zero-order energy is just is the sum of the molecular orbital energies (MP0). The first-order perturbation (MP1) results in the HF energy and the first possible inclusion of the correlation energy is obtained only when second-order perturbation is considered, thus resulting in the MP2 method [71]. Skipping the theoretical derivation, the MP2 contribution to the energy is the following:

$$E(\mathrm{MP2}) = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\left(\langle \psi_i \psi_j | \psi_a \psi_b \rangle - \langle \psi_i \psi_j | \psi_b \psi_a \rangle\right)}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \quad , \tag{2.5}$$

where the bra-ket notations is adopted for the two-electrons integrals over the MOs (ψ) and the MO energies are indicated with ϵ. The subscripts $i, j, k, ...$ indicate the occupied MOs, while $a, b, c, ...$ are used for the virtual (unoccupied) orbitals. In eq. (2.5), only double excitations from occupied orbitals to virtual ones are included, but they account for a large part of the correlation energy, about 80-90% [71]. Further orders of perturbation can be considered, thus resorting to the MP3 and MP4 methods. However, the MP3 electronic energy is degraded compared to the MP2 one and an improvement is seen only with MP4. Since the energy does not benefit from higher-level of perturbations and the computational cost increases by two orders of magnitude when moving from MP2 to MP4, the MP4 and MP3 methods are rarely employed in literature [71].

As previously mentioned, CC methodologies are also employed extensively in the present thesis. The idea behind the CC formalism is to include a *type* of excitation to an infinite order, e.g. incorporating in the multi-Slater determinant all possible single excitations that can be generated from the reference wave function considering all the unoccupied orbitals. This is actually achieved via the definition of the *cluster* operator, $\mathbf{T}$ as the sum of operators ($\mathbf{T}_i$) that are able to generate the i-th type of excitation, where i is the number of excited electrons. For example, the $\mathbf{T}_1$ operator applied to the HF wave function generates all the Slater determinants Ψ_i^a resulting from single excitations as indicated in the following:

$$\mathbf{T}_1 \Psi_{\mathrm{HF}} = \sum_{i}^{occ} \sum_{a}^{vir} t_i^a \Phi_i^a \quad , \tag{2.6}$$

where t_i^a are expansion coefficients, historically called *amplitudes*. The multi-determinant wave function in CC is built via the $\mathbf{T}$ operator as follows:

$$\Psi_{\mathrm{CC}} = e^{\mathbf{T}} \Psi_{\mathrm{HF}} \quad , \tag{2.7}$$

where the exponential can be expanded in Taylor-series and, in view of the form of $\mathbf{T}$,

it can be expressed as:

$$e^{\mathbf{T}} = 1 + \mathbf{T}_1 + \left(\mathbf{T}_2 + \frac{1}{2}\mathbf{T}_1^2\right) + \left(\mathbf{T}_3 + \mathbf{T}_2\mathbf{T}_1 + \frac{1}{6}\mathbf{T}_1^3\right) + \dots \quad . \tag{2.8}$$

Introducing eq. (2.8) in eq. (2.7), the first term on the right side of the former equation generates the HF wave function, the second term all the singly excited states, the third term (i.e., the first parenthesis) the doubly excited states and so on. Starting from the third term, it can be noticed the presence of a connected term (direct double excitations, $\mathbf{T}_2$) and a disconnected term (double excitations obtained via two times the single excitation operator, $\mathbf{T}_1^2$) and, similarly, the same occurs for the third excitations and all the higher terms.

In this case, the SE to obtain the CC correlation energy is solved by projecting the Ψ_{CC} onto the Ψ_{HF} as follows:

$$\langle \Psi_{\mathrm{HF}} | \mathbf{H} e^{\mathbf{T}} | \Psi_{\mathrm{HF}} \rangle \quad , \tag{2.9}$$

where the Ψ_{CC} in the form of eq. (2.7) is considered. By expanding the exponential, it can be easily seen that non-null terms in eq. (2.9) are obtain only when considering single and doubles excitations, which can be counterbalanced the effect of the one- and two-electron operators of the Hamiltonian. Therefore, the CC energy becomes:

$$E_{CC} = E_{\mathrm{HF}} + \sum_{i<j}^{occ} \sum_{a<b}^{vir} \left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a\right) \left(\langle \psi_i \psi_j | \psi_a \psi_b \rangle - \langle \psi_i \psi_j | \psi_b \psi_a \rangle\right) \quad . \tag{2.10}$$

As evident from the previous equation, the energy depends only on single and double excitations via the corresponding amplitudes, which in turn are coupled with higher-order excitations in the amplitudes equation:

$$\langle \Psi_{\mathrm{CC}}^* | e^{-\mathbf{T}} \mathbf{H} e^{\mathbf{T}} | \Psi_{\mathrm{HF}} \rangle = 0 \quad . \tag{2.11}$$

Here, Ψ_{CC}^* represents a generic excited state generated by the cluster operator and a large number of equations has to be solved to obtain the amplitudes, and their contribute to the energy.

The number of equation to be solved increases by increasing the size of the molecule, because (i) more electrons lead ideally to more $\mathbf{T}_i$ operators (higher excitations) and (ii) the number of generated excited states increases consequently. Therefore, complete CC expansions are only feasible for very simple systems and one has to truncate the cluster operator to a certain order of excitations. The most important excitations to consider are the doubles (D), whose consideration leads to the CCD method [74], which however does not include the disconnected double excitations previously mentioned. Indeed, the latter ones depend on $\mathbf{T}_1$ and, in order to incorporate them, one has to

include also single (S) excitations, thus obtaining the CCSD method [75]. If the triples are also included, the CCSDT method is defined [76,77], while the addition of Q and P stands for quadruple and pentuple excitations, respectively [71,78,79].

The truncation of the cluster operator leads to several vanishing terms in the amplitudes equations derived from eq. (2.11), thus resulting in approximated amplitudes, i.e. an approximated wave function. The latter then leads to an approximated electronic energy [71,79].

CC theory and MP-PT are strictly correlated [31,71]. The former includes a specific type of excitation to the infinite order, while the latter incorporates all the types of excitations up to a specific order, starting from the second (MP2, MP3, MP4, etc.). For example, MP2 and MP3 include only doubly excited determinants, but MP4 includes singles, doubles, triples and quadruples excitations. MP-PT gives an indication of the most important excitation class for the description of the wave function, thus pointing out that, in addition to doubles and singles, one has to include at least triple excitations. At the CC level, this means to consider the CCSDT method, which is computationally unfeasible already for medium-sized systems due to the large number of determinants incorporated in the wave function. Therefore, one can resort to a mixed approach where the amplitudes of the CCSD method are used to compute the contribution of triples via the MP4 formula. In this procedure, an additional term not belonging to the fifth-order perturbation is also included due to its minor computational cost with respect to the other step of the computation. This term describes the coupling between single and triple excitations and, in total, one has the CCSD(T) method, with the effects of triple excitations is included in a single step "(T)", avoiding the iterative process of the full CCSDT model [71,73,80].

The theoretical approaches outlined above neglect the explicit two-electron interaction in the Hamiltonian and try to recover its effect by inclusion of different excited configurations in the wave function. However, the r_{12} term has an impact on the wave function, which should present a singularity at $r_{12} = 0$, i.e., when the electrons coalesce [71,79]. This singularity is represented by a cusp which is not well described by traditionally correlated wave functions and for this reason the MPn and CC approaches slowly converge to the exact energy [71,79]. At the Coulomb cusp the wave function is non-differentiable; this implies that it should depend linearly on r_{12} for small values of such distance. The inclusion of the r_{12} term in the wave function leads to the definition of the so-called explicitly correlated methods [71,79,81]. The parent method of such methodology is the R12 (Ψ_{R12}) method which includes, on top of the HF reference, corrections by r_{ij} factors. The R12 method converges faster to the exact energy, but the great disadvantage is that the energy depends on three- and four-electrons integrals [81]. This is by far more computationally demanding than the computation of the bi-electronic terms required in CC theory (eq. (2.10)). Consequently, this method is less used than standard CC methodologies. A way-out is offered by the Resolution of the

Identity (RI), which reduces the three and four-electron integrals to a sum of integrals involving only two particles [81]. This methodology is exact in the limit of the infinite basis set , but in real cases it represents an approximation that greatly reduces the cost and it is extensively applied in explicitly correlated methods without introducing a large computational error.

In recent years, the linear dependence in r_{12} was replaced by a Slater-type function [81]:

$$F(r_{12}) = \frac{1}{\gamma} exp^{-\gamma r_{12}} \quad , \tag{2.12}$$

thus defining the so-called F12 methods, where γ is usually referred to as geminal exponent. The latter methods have extensively replaced the R12 counterparts because of their lower computational cost. In fact, the Slater function of eq. (2.12) can be approximated using Gaussian functions that simplify the overall integral computation [82] and this can be combined with the RI approximation. Density Fitting (DF) procedures have also been introduced to heavily reduce the computational cost [81]. It should be mentioned that the theoretical backgrounds of R12 and F12 methodologies are not different from the standard methods already outlined, but the additional operator $F(r_{12})$ leads to a more complicated theoretical treatment that will not be expanded here. In the thesis work, F12 approaches have been employed together with MP2 and CCSD(T) methods, i.e. MP2-F12 and CCSD(T)-F12, respectively. For the former, the general orbital-invariant MP2-F12 approach is used and the ansatz is based on the 3*C approximation, which assumes the extended Brillouin condition (EBC) and uses the following operator to define the explicitly correlated wave-function:

$$\mathbf{O}_{12} = (1 - \mathbf{o}_1)(1 - \mathbf{o}_2)(1 - \boldsymbol{\nu}_1\boldsymbol{\nu}_2) \quad . \tag{2.13}$$

The operator above generates doubly excited determinants due to the $\boldsymbol{\nu}_1\boldsymbol{\nu}_2$ term and the coefficients of such configurations are kept fixed, thus defining the so-called FIX ansatz [83]. A short note is deserved to understand the EBC condition. The RI approximation is computed typically with an *ad hoc* basis denoted as auxiliary basis set (ABS). If this ABS is chosen to be the complementary of the basis set used to build the orbitals, i.e. the orthogonal counterpart, one has the complementary ABS (CABS) approach which is extensively used in F12 methods and combines both types of basis sets. The RI approximation generates a matrix term that couples the Fock operator with the full CABS orbitals and the EBC approximation avoids the possibility of having these terms, but also of having coupling terms between virtual orbitals and the other CABS orbitals [83].

Concerning the CCSD(T)-F12 method, the explicit correlation is only introduced in the CCSD term, while the (T) contribution is sill computed with the standard approach. Also in this case the FIX ansatz is employed [82, 84] and the treatment of the

amplitudes arising from the explicitly correlated term can lead to different approximations denoted as *a*, *b*, and *c*. All these three approximations have been used in the present thesis, with the latter usually indicated as CCSD(F12*)(T) [82,84].

Due to these different approximations, the F12 methods do not convergence to the exact value (in the limit of infinite basis set) as rapidly as the R12 methods, but they still represents an improvement compared to the standard counterparts.

A conceptually different methodology is DFT, which is based on the idea that molecular information can be derived from the electronic density (ρ) and instead of the molecular wave-function. Indeed, it has been proven that there is an unique correspondence between the molecular electronic density and its energy [85]. Using KS theory [86], the overall energy can be partitioned as follows:

$$E_{DFT}[\rho] = T[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}\rho \quad , \tag{2.14}$$

where the term on the left-hand side is the exact energy, $E_{ne}[\rho]$ and $J[\rho]$ are the potentials due to the attraction between nuclei and electrons and the classical Coulomb repulsion between electrons, respectively [71]. The term $T[\rho]$ is the kinetic energy of non-interacting electrons, i.e. a non-real system. The last term on the right-hand side of eq. (2.14), i.e. the exchange-correlation term $E_{xc}[\rho]$, includes non-classical terms due to exchange (E_x) and correlation (E_c) of the electrons, but also accounts for the correction to the kinetic energy due to the fact that electrons are actually interacting. Although it can be proven that the exact functional form of $E_{xc}[\rho]$ exists, it is still (and will probably remain) unknown. Several estimates of such term have been developed over the years, starting from the Local Density Approximation (LDA) which considers the electron density as approximated to that of an homogeneous electron gas for each point in space. This idea was further improved considering the separated electronic density for the two α and β spins, leading to the Local Spin Density Approximation (LSDA), whose exchange term is approximated to that of a uniform electron gas for each spin [71]:

$$E_x^{LSDA} = -2^{1/3} C_x \int (\rho_\alpha^{4/3} + \rho_\beta^{4/3}) d\mathbf{r} \quad . \tag{2.15}$$

The analytical form of E_c is available for particular density conditions, and different formula, depending on several fitting parameters, have been developed to simulate the analytical values. The exchange and correlation energies obtained within the LSDA approximation are typically overestimated, with results differing orders of magnitude from chemical accuracy [71]. For this reason, the GGA approaches have been developed, and they are based on the General Gradient Approximation, which considers that the energy depends on the local value of the density but also on its gradient. A very well-known GGA exchange functional was proposed by Becke [87]:

$$E_x^B = E_x^{LSDA} + \Delta E_x^B \quad \text{with} \quad \Delta E_x^B = -\beta\rho^{1/3}\frac{x^2}{1+6\beta x sinh^{-1}x} \quad . \tag{2.16}$$

In eq. (2.16) the x term includes the dependence of the energy on the density gradient and is equal to $|\nabla\rho|/\rho^{4/3}$, while β is a fitting parameter. The latter was optimised by fitting the exact exchange energy of six noble gases, from He to Rn. For the correlation term, the most common functional employed is the Lee, Yang, Parr (LYP), E_c^{LYP} [88]. Once combined with the Becke exchange functional, the BLYP functional is obtained. Other approximations for the correlation and the exchange functionals were developed by Perdew-Burje-Ernzerhof and the corresponding method is abbreviated as PBE [89].

GGA functionals are still limited in accuracy and one can improve the DFT methodology by exploiting the Adiabatic Connection Model (ACM). According to it, the total exchange-correlation energy is expressed as a function of the GGA E_{xc} energy with an additional contribution from the HF exchange energy. As already mentioned, the HF exchange energy accounts for the non-classical repulsion between electrons, i.e. it tunes the probability also considering the spin of the electron. The combination of these terms can involve several parameters; in particular, the B3LYP approach, extensively used in this thesis, requires three fitting parameters (a, b and c) that are used to combine (i) the Becke exchange GGA functional, (ii) the LYP correlation functional mixed with a LSDA based functional and (iii) the HF exchange energy, as follows [71,90]:

$$E_{xc}^{B3LYP} = (1-a)E_x^{LSDA} + aE_x^{\mathrm{HF}} + b\Delta E_x^B + (1-c)E_c^{LSDA} + cE_c^{LYP} \quad . \tag{2.17}$$

The fitting parameters a, b and c are equal to 0.20, 0.72, and 0.81, respectively. The a parameter tunes the HF exchange contribution, which in general is always between 10-25%. The methods containing HF energy are sometimes referred to as hybrid or hyper-GGA functionals, among which another important example is the PBE1PBE (or PBE0) functional, which contains only one fitting parameter (g), kept fixed at a value of 0.25 based on perturbation theory. In the Gaussian suite of programs [91] employed in the present thesis, a different parametrization of the LSDA correlation is employed which is based on the the Vosko, Wilk and Nusair (VWN) functional [92].

Similarly to wave-function-based methods, one can expect an improvement by the electronic treatment of hybrid functional using perturbation theory. In this thesis, two important methods that are based on such theoretical means have been used: the B2PLYP [93] and revDSD-PBEP86 [94] method. In the former, the total energy is expressed as:

$$E_{xc} = (1-a_x)E_x^{GGA} + a_xE_x^{\mathrm{HF}} + bE_c^{GGA} + (1-b)E_c^{PT2} \quad , \tag{2.18}$$

where the form of term E_c^{PT2} is the same as that reported in eq. (2.5), but it is computed

using the KS orbitals instead of the HF ones. The other GGA terms in eq. (2.18) are those of the BLYP functional. The fitting parameters are a_x equal to 0.53 and b equal to 0.27.

The revDSD-PBEP86 functional is based on the spin-component scaled (SCS) MP2 energy that separates the opposite and same spin contributions to the energy. This separation gives flexibility to the method, balancing short- (same spin) and long-range (opposite spin) interactions [94–96]. In this case, the exchange-correlation term can be written as:

$$E_{xc} = c_x E_x^{\mathrm{HF}} + (1 - c_x) E_x^{PBE} + c_c E_c^{P86} + c_o E_o^{MP2} + c_s E_s^{MP2} \quad , \tag{2.19}$$

where the correlation part is computed using the P86 functional in combination with the PBE exchange correlation. The E_o^{MP2} and E_s^{MP2} terms indicate the opposite and same-spin contributions of MP2 energy, respectively, and E_x^{HF} is the usual HF exchange energy. Each contribution has the corresponding fitting parameter and the values are: c_x=0.69, c_c=0.4296, c_o=0.5785, c_s=0.0799. If the PT2 energy is included in the total DFT energy, the methodology is said to be based on a double-hybrid density functional.

A weakness of DFT is that, by definition, poorly describes the dispersion interactions, thus often providing unreliable descriptions of weakly bounded systems. This behaviour is corrected via an empirical dispersion term, which means that the DFT energy is corrected for an additional term, E_D [97]. In the case of the D3BJ correction employed in this work [98,99], the term is given by:

$$E_D = \frac{s_n}{2} \sum_{N \neq L} \sum_{n=6,8} \frac{\frac{c_n^{NL}}{R_{NL}^n}}{1 + \frac{a_1 R_{NL}^0 a_2}{R_{NL}}} \quad . \tag{2.20}$$

In the previous equation, s_6, s_8, a_1 and a_2 are fitting parameters. R_{NL} is the interatomic distance, R_{NL}^0 is the sum of van der Waals radii, and C_n^{NL} are the dispersion coefficients [98]. The value of these quantities can be found in the literature, but it is important to notice that the fitting parameters change for different functionals. Indeed, for B3LYP these parameters are: $s_6 = 1.0000$, $s_8 = 1.9889$, $a_1 = 0.3981$ and $a_2 = 4.4211$. For B2PLYP the same parameters are 0.6400, 0.9147, 0.3065 and 5.0570. In the case revDSD-PBEP86 functional, s_8 is set to zero, s_6 is 0.4377, $a_1 = 0$ and $a_2 = 5.5$. The performance of DFT is greatly enhanced by empirical dispersion. For this reason, it will always be included in the work considered in this thesis.

2.1.2 Basis Sets

As outlined above, the wave function has to be constructed since it is not known *a priori*. The first step in this process is to build the MOs which are typically obtained via Linear Combination of AOs, i.e. via the so-called LCAO approach. In the HF-

SCF procedure, the coefficients of the LCAO approach that define the initial MOs are variationally optimised to give the lowest HF energy and the HF wave function. The latter is then used as starting point to include electron correlation in CC theory, MP-PT and other methods. This is schematically represented in fig. 2.2 and it should also be noted that a similar process, starting from AOs, is used in DFT to construct the KS orbitals.

In this procedure, one question is still open: which mathematical functions should be used as atomic orbitals? The question is simple, but there is not a correct or unique answer. Theoretical chemists have focused their effort over the years on two main functional forms: the Slater-type orbitals (STOs) and the Gaussian-type orbitals (GTOs). The former are functions based on the resolution of the SE for the hydrogen atom and well represent the behaviour of the electron density. However, STOs are computationally disfavoured because the analytical form of the multi-particle integrals is not known and they need to be computed numerically [71]. This becomes even more complicated if one considers that STOs do not have radial nodes and linear combination of STOs have to be used, thus enlarging the total number of functions. From the computational point of view, a better choice is offered by GTOs that have an analytical form for the multi-index integrals. However, as detailed in eq. (2.21) [79, 100], they behave differently with respect to STOs:

$$\begin{aligned} \chi_{\zeta,n,l,m}(r,\theta,\psi) &= NY_{l,m}(\theta,\psi)r^{n-1}e^{-\beta r} \qquad \text{Slater}-\text{type} \\ \chi_{\zeta,n,l,m}(r,\theta,\psi) &= NY_{l,m}(\theta,\psi)r^{2n-2-l}e^{-\beta r^2} \qquad \text{Gaussian}-\text{type} \end{aligned} \tag{2.21}$$

Such a difference is emphasised in panel (a) of fig. 2.3 for the radial part. In eq. (2.21), N is a normalisation constant, $Y_{l,m}$ are the spherical harmonic functions and r is the distance between the nucleus and the electron. The β parameter controls the width of the function and n is a natural number, which indicates the electronic shell [71]. A solution to recover such a difference is to linearly combine more GTOs with different widths (exponents) to simulate a single STO. The minimal number of GTOs used to simulate one STO is three, which leads to the definition of the STO-3G basis set [101] (see

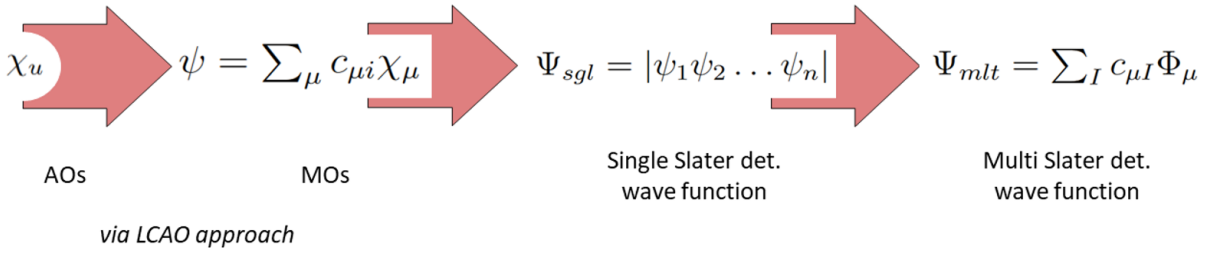


Figure 2.2: Construction of a correlated wave function (Ψ_{mlt}) starting from AOs. Ψ_{slg} indicates a single-determinant wave function, such as the HF one.

panel (b) of fig. 2.3). Over the years, basis sets have extensively developed around this concept and contracted basis sets have been developed. In this case, one uses a fixed linear combination of primitive GTOs (pGTOs) to obtain a smaller set of contracted GTOs (cGTOs), which is then used as atomic orbitals. Within the general contraction scheme, all the primitives of one atom are used with different contraction coefficients to obtain the cGTOs required.

The size of the basis set is indicated in terms of the "ζ" and the minimal basis set (single-ζ) is made up only by one GTO (or cGTO) for describing one orbital. For example, for helium only the $1s$ orbital is required, while for the carbon atom $1s$, $2s$ and three $2p$ orbitals are used. Moving to a double-ζ basis set, always considering He, the $1s$ orbital is described by two GTOs (or cGCTOs) and the carbon atom has in addition other two GTOs/cGTOs describing the $2s$ orbital and 6 GTO/cGTOs describing the p orbitals.

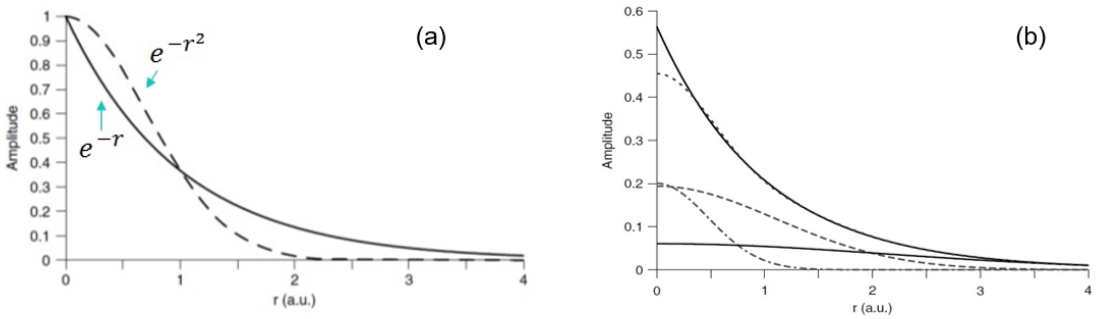


Figure 2.3: Panel (a) shows a comparison between a GTO (dashed line) and a STO (solid line) for the radial part of a $1s$ orbital. In panel (b) three Gaussian functions are contracted to obtain a single GTO (STO-3G basis set). The resulting GTO (dashed line) better represents the single STO (solid line) for the radial part of a $1s$ orbital.

This thesis employs the correlation-consistent (cc) basis sets, developed by Dunning and co-workers [102], that have contraction coefficients optimised to include the correlation energy. This type of basis set is indicated as "cc-pVnZ", where VnZ indicates that nZ functions are introduced only for the valence orbitals, while the inner one are described by a single GTO or cGTO. Therefore, such basis set should be used to correlate only valence electrons. The size of the basis is expressed in term of ζ, as explained above, with n equal to D, T, Q, 5, and 6. The "p" term stands for "polarized", thus meaning that the basis includes polarisation functions, i.e. functions with an higher angular momentum than the last occupied atomic orbital. For a H atom, p-type orbitals are polarisation functions, while for a carbon atom (where p orbitals are occupied) the d orbitals are polarisation functions. Similar basis sets have also been developed to describe the correlation of inner electrons, i.e. the core-core and core-valence electron correlation, and are indicated as cc-pCVnZ [103]. This family of basis sets has additional tight functions, i.e. functions with a large exponent and thus very

narrow. Another possible addition to these basis sets is the inclusion of diffuse Gaussian functions, which have small exponent; thus, they are loose and of great relevance when aiming at describing weak interactions or particular proprieties like the dipole moment, which are both sensitive to the electron density in regions away from the nucleus [37, 104]. The number of diffuse functions added depends on the prefix, for example, a diffuse function added for each angular momentum present in the basis set corresponds to the full augmentation, which is denoted as "aug-" [104]. If the diffuse functions are removed from the H atoms, one has the "jul-cc-pVTZ" set. The additional removal of the highest angular momentum diffuse functions from non-H atoms leads to the "jun-cc-pVnZ" basis set and so on along the months of the calendar (may-, apr-) if more the diffuse functions are eliminated. If only the s and p shell are augmented by diffuse functions, one has the minimally augmented basis set indicated by the suffix "maug-" [30, 105]. The interesting propriety of the correlation-consistent basis sets is that they constitute a hierarchical series, thus the energy is improving systematically with increasing the size of the basis set.

As correlation consistent basis sets have been developed to use methods that incorporates electron correlation in the standard manner, similar basis sets have been optimised to be employed with explicitly correlated methods and they are denoted by the suffix "F12": "cc-pVnZ-F12" [106]. These family is constructed by using s and p functions of the cc-pV(n+1)Z basis set and the diffuse function of the s and p shells of the aug-cc-pV(n+1)Z basis set. They further include a tight p function for each size of the basis set. The standard set is then augmented by 2d functions at the DZ level, 3d2f for the TZ basis and 4d3f2g at the QZ size. The F12 methodology also requires basis sets to perform the resolution of identity and density fitting procedures; these sets depend on the orbital basis set employed and will be illustrated in each specific case [81].

2.1.3 Accuracy in Quantum Chemistry

As the basic concepts of theoretical methods and basis sets have been illustrated, the natural step forward is the discussion of the overall accuracy of a given level of theory. Its error with respect to the exact solution is the so-called apparent error, but what is at the basis of such error? In the previous discussion, two important truncations have been made. For a N-particle model, considering a specific basis set, the number of excitations considered was truncated to a certain level, e.g. only double excitations for MP2. Such a truncation leads to an error, denoted as N-electron error. As the number of excitations considered in the computation increases, the N-electron error decreases and, already at the CCSD(T) level of theory, it is considered quite small.

The second truncation regards the orbital basis set. Indeed, whenever expressing the MOs as a sum of one-electron functions (AOs), the summation should be infinite or, in other word, the basis set should be complete. As already stated, this is not feasible

and the construction of MOs is only based on a finite number of AOs. The BSE is due to this incompleteness and is represented by the difference between the basis set employed and the solution at the Complete Basis Set (CBS) limit for a given theoretical method. The residual N-electron error remaining at the CBS limit is denoted as *intrinsic error*. The relation between these errors is shown in fig. 2.4. Reducing the error of a computation translates into reducing the apparent error and this should be done in two, complementary, ways. First, one can benefit of extrapolation formulas to estimate the CBS limit of a N-electron model. Secondly, since the N-electron error is due to the type of model, one can combine different theoretical approaches to partially reduce also the N-electron error. Approaches based on the combination of both the previous points are denoted as composite schemes, and they strongly enhance the accuracy of computational models.

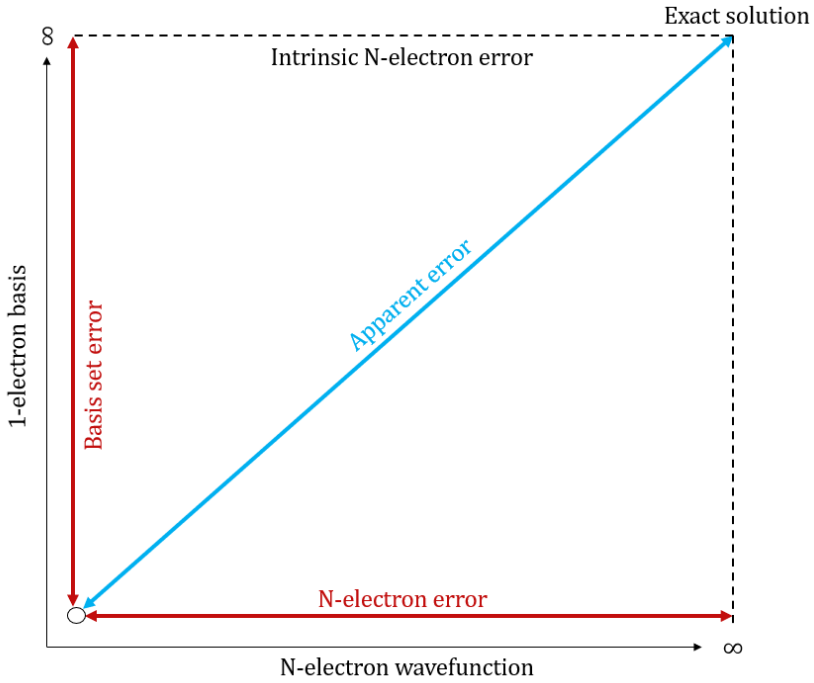


Figure 2.4: Representation of the N-electron error, BSE error, and the apparent error with respect to the exact solution of the SE equation.

How can one estimate the CBS limit? What are the contributions to consider in a composite scheme? The latter question does not have a unique answer, because it depends on the computational resources and on the propriety of interest. The former question however is quite simple and the basic idea is that the energy (or a propriety) converges monotonically to the CBS limit by enlarging the basis set, provided that one deals with a hierarchical series such as the correlation-consistent family [102]. In the following, we will illustrate how the CBS limit can be estimated for the energy but the same schemes can be applied to geometrical parameters and/or other molecular proprieties.

Fig. 2.5 shows the generic convergence of a propriety to the CBS limit, which can be represented by a specific functional form. The latter depends on the basis sets employed and on the type of energy considered. This functional form can be hypothesised on theoretical basis, but the trend is unavoidably affected by the numerical artefacts of the computation and the approximations made. Therefore, the trends are actually identified owing to the systematic analysis of the convergence for a large data set of molecules. In particular, for the HF energy, an extrapolation formula, which is largely employed, was obtained by Feller [107]:

$$E_{\mathrm{HF}}^{CBS} = E_{\mathrm{HF}} + ae^{(-bn)} \quad , \tag{2.22}$$

where a, b are the parameters tuning the extrapolation formula. E_{HF} is the HF energy obtained using basis set of the $n\zeta$-order, where n also denotes the cardinal number of the exponent. Therefore, three computations with basis sets of increasing dimension are required to estimate the CBS limit of the HF energy, E_{HF}^{CBS}. This formula is sometimes referred to as three-point extrapolation formula, because calculations with three different bases are required to estimate the CBS limit.

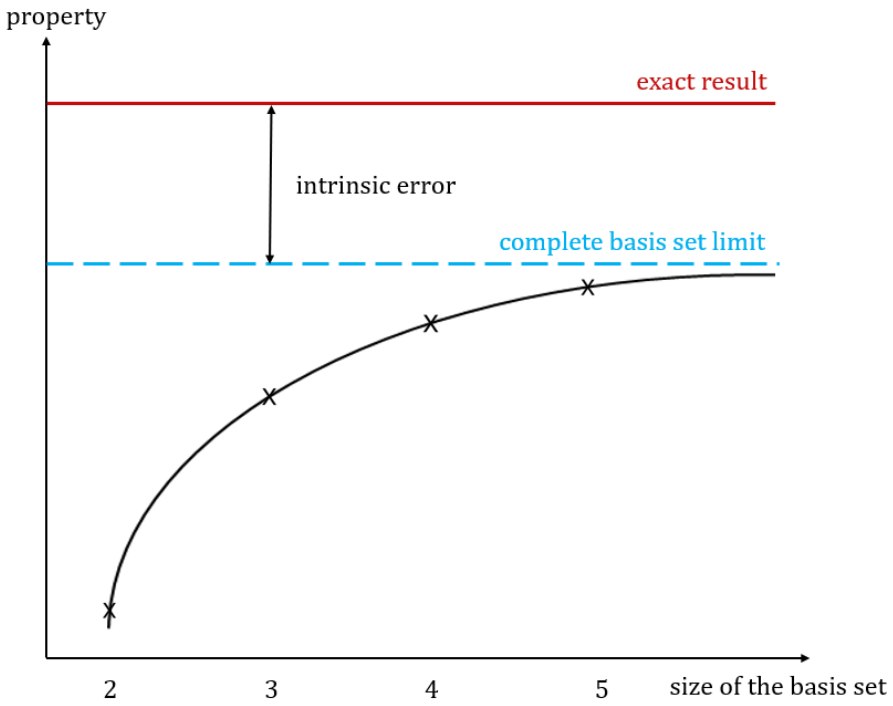


Figure 2.5: Trend of convergence for a generic propriety with respect to the size of the basis set.

Several extrapolation formulas have been proposed to extrapolate the contribution of the correlation energy. In particular, one can refer to the two-point formula by Helgaker and co-workers [108] or to an extrapolation formula requiring three different computations, similarly to that of the HF energy [109]. The latter is maybe more ac-

curate, but it is only useful for small systems for which computations with large basis sets are still feasible in terms of memory and computational time. For this reason, the two-point formula is probably the most employed in the literature and is the only one considered in the present thesis. The correlation energy is extrapolated as follows:

$$\Delta E_{corr}^{CBS} = \Delta E_{corr}(n) - \frac{c}{n^3} \quad , \tag{2.23}$$

where ΔE_{corr}^{CBS} and c are the extrapolation parameters and are obtained by means of a two-point fit. As before, n is the cardinal number of the basis set employed. This formula is commonly used in conjunction with MP2 and CCSD(T) methods.

The trend of the explicitly correlated energy with respect to the size of the basis sets is different because it one includes directly the two-particles repulsion. Theoretically, the energy should converge as n^{-7}. However, the convergence is somewhat different due to the approximations that plague the F12 computations and this topic will be tackled in Chapter 3. Basis set extrapolations lead to large improvements and are a mandatory requirement when aiming at the so-called chemical accuracy [110–112]. The remaining error, i.e. the intrinsic error, can only be lowered by improving the methodology employed, but this severely hampers the computational cost. Indeed, if *M* is the basis set size, the simple HF method scales as M^4, while MP2 scales as M^5. The CCSD method scales as M^6 and the CCSD(T) method as M^7, with the difference that the former scaling is fully iterative while the latter has, in addition to the M^6 iterative scaling, non-iterative M^7 steps, still being doable for medium-sized molecules. The CCSDT scales with the eighth power of *M*. The inclusion of full triple excitations does not represent a large improvement compared to CCSD(T) since perturbative triples tend to overestimate their contribution, thus often providing results lying between CCSDT and CCSDTQ in terms of accuracy. Therefore, one should employ the CCSDTQ method to notice a real improvement in the electronic description. Iterative quadruples scale as M^{10}, thus being very prohibitive in conjunction with large basis sets and/or for large molecules with more than 3-4 heavy atoms. One can resort to the CCSDT(Q) method which scales as M^9, where only non-iterative steps scale with the ninth-power and the iterative with scale as M^8.

Since employing higher-order wave functions means increasing the computational cost, it is reasonable to combine different methods where the basis sets are tuned to reduce the overall computational cost. For example, one might compute the CCSD/cc-pVQZ energy and combine it with the (T) contribution obtained using a triple-ζ basis set and the quadruples contribution with a double-ζ set. Another way to save computational time is to consider separately the correlation due to valence electrons and that due to of inner (core) electrons. For example, one can estimate the CBS limit for the energy contributions of valence electrons and add a single basis correction due to the correlation of inner (core) electrons [110–112]. The latter is computed at the considered level of theory as the difference between ae and fc calculations and is often denoted as

the Core-Valence (CV) term.

The combination of these ideas led to the definition of composite schemes, widely employed in the literature to determine energies, molecular geometries, vibrational frequencies and other properties. Two main schemes will be used throughout the thesis: (*i*) the HEAT-like scheme and (*ii*) the "cheap" scheme, abbreviated as ChS. Concerning the former, the energy is computed as [113]:

$$E_{\mathrm{HEAT-like}} = E_{\mathrm{HF}}^{\mathrm{CBS}} + \Delta E_{\mathrm{CCSD(T)}}^{\mathrm{CBS}} + \Delta E_{\mathrm{CV}} + \Delta E_{\mathrm{fullT}} + \Delta E_{\mathrm{pQ}} \tag{2.24}$$

where the extrapolation to the CBS limit of the HF ($E_{\mathrm{HF}}^{\mathrm{CBS}}$) and CCSD(T) ($\Delta E_{\mathrm{CCSD(T)}}^{\mathrm{CBS}}$) energies is carried out with the formulas previously discussed. The ΔE_{CV} indicates the contribution due to the correlation of inner electrons and is computed as the difference between ae- and fc-CCSD(T) computation in the same basis set. Indeed, the $E_{\mathrm{CCSD(T)}}^{\mathrm{CBS}}$ term is evaluated with the inner-shell electrons frozen. The other terms on the right-hand side of eq. (2.24) are the full account of triple excitations computed with the CCSDT method, and the contribution of quadruple excitations by means of a perturbative treatment, the CCSDT(Q) method [114–116].

This composite scheme reassembles the HEAT approach [110, 117, 118], but some different, time-saving, steps are taken. Indeed, the original approach carries out also the extrapolation of CV contribution and employs the CCSDTQ method instead of the perturbative one. Further smaller differences are also present and depend on the basis sets employed for the computations. However, the methodology still guarantees sub-kJ·mol^{-1} accuracy [113,119] with a reduced computational cost. Furthermore, in the scheme other two terms can be incorporated: the diagonal Born-Oppenheimer correction (DBOC) (ΔE_{DBOC}) [120–122] and the scalar relativistic correction (the mass-velocity and Darwin terms) computed with perturbative techniques (ΔE_{rel}) [123].

A simpler composite scheme, computationally less demanding than the HEAT and HEAT-like approaches, is the "cheap" scheme (ChS). The name points out to its low computational cost and it can be employed for medium-sized molecules. The energy of the ChS model is obtained as follows:

$$E(\mathrm{ChS}) = E(\mathrm{CCSD(T)}) + \Delta E_{\mathrm{MP2}}^{\infty} + \Delta E_{\mathrm{MP2}}^{\mathrm{CV}} \quad . \tag{2.25}$$

The first contribution on the right-hand side of this equation is the energy obtained at the CCSD(T) level of theory in conjunction with a triple-zeta basis set. The second term is the HF+MP2 energy extrapolated to the CBS limit using the two-point formula of eq. (2.23), thus extrapolating the total energy. The basis sets used are of triple- and quadrupole-zeta quality, with the smallest basis set used in the extrapolation procedure being the same basis set employed for the CCSD(T) term. This is important to avoid inconsistency in the extrapolated HF energy present in term $\Delta E_{\mathrm{MP2}}^{\infty}$. The last term accounts for the CV contribution as the first two terms of eq. (2.25) are evaluated

within the fc- approximation, and it is computed with the MP2 method. This scheme will be further developed in this thesis to treat nc-complexes and further details will be given in Chapter 3.

It should also be stressed that, while these composite schemes have been illustrated for the energy, they can be applied to molecular geometries and other proprieties [27,110,111,124–129]. Apart from the composite approaches already introduced, other important schemes belong to Weizmann-n (Wn) [127,129–133] and Correlation consistent composite approach (ccCA) [134, 135] families. In the case of molecular structure, one can build the energy gradient to be minimised on the basis of the chosen composite scheme or can apply directly them to the geometrical parameters. Both approaches are equally valid, but the second case results in a large saving of computational resources [136,137].

2.1.4 Vibrational and Rotational Motions

In the previous sections, the electronic energy was considered. However, when comparing computational data with experimental results, one has also to account for vibrational and rotational motions. This means somewhat solving the nuclear SE after the resolution of the electronic SE for obtaining a description of the PES (or the portion of interest). This process starts from the definition of an appropriate reference system which is derived imposing Eckart's conditions and working within the principal inertia system. In the latter, a general Cartesian axis α is denoted as a, b and c axes. In this case, one can arrive to the definition of normal coordinates that are linear combinations of internal coordinates and are obtained via the diagonalization of the force constant matrix. The latter contains all the second derivatives of the potential energy with respect to the possible nuclear displacements, thus being a $3N \times 3N$ matrix, with N the number of atoms. This matrix can be called also Hessian matrix and its diagonalization leads to $3N$-6 eigenvalues that are directly related to the vibrational frequencies (ω_r) of the molecule within the harmonic approximation, being the potential of a quadratic form with respect to Q_r [33,138]. The corresponding $3N$-6 eigenstates are the normal modes of the molecule, while the null eigenstates correspond to rotational and translational motions. Indeed, owing to the Eckart conditions, one can rule out the contributions of translation (3 coordinate) and rotation (3 coordinates, or 2 in the specific case of linear molecules).

An important Hamiltonian used to solve the nuclear problem was developed by Watson [33, 36, 37] and is expressed in dimensionless mass-weighted coordinates, i.e. $q_r = \sqrt{\omega_r} Q_r$:

$$\mathbf{H}_N = \frac{1}{2} \sum_{\alpha,\beta} \left(\mathbf{J}_\alpha - \boldsymbol{\pi}_\alpha \right) \mu_{\alpha\beta} \left(\mathbf{J}_\beta - \boldsymbol{\pi}_\beta \right) + \frac{1}{2} \sum_r \omega_r \mathbf{p}_r^2 + V(\mathbf{q}) + \frac{1}{8} \sum_\alpha \mu_{\alpha\alpha} \quad . \tag{2.26}$$

Here $\mathbf{J}_\alpha$ stands for the projection of the rotational-angular momentum operator along the α-axis. The other operator, i.e. π_α is the projected vibrational angular momentum on the α axis, while $\mathbf{p}_r^2$ is the r-th normal coordinate conjugate vibrational momentum. The remaining terms of eq. (2.26) are the inverse of the inertia moment ($\mu_{\alpha\beta}$) and the potential V($\mathbf{q}$) that can be both expressed in a Taylor series:

$$\mu_{\alpha\beta} = \mu^e_{\alpha\beta} + \sum_r \mu^r_{\alpha\beta} q_r + \frac{1}{2}\sum_{r,s} \mu^{rs}_{\alpha\beta} q_r q_s + \dots \tag{2.27}$$

$$V(\boldsymbol{q}) = \frac{1}{2}\sum_r \omega_r q_r^2 + \frac{1}{6}\sum_{r,s,t} \phi_{rst} q_r q_s q_t + \frac{1}{24}\sum_{r,s,t,u} \phi_{rstu} q_r q_s q_t q_u + \dots \quad . \tag{2.28}$$

In eq. (2.27), the leading term is the inverse principal moments of inertia at the equilibrium structure:

$$\mu^{\mathrm{e}}_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{I_\alpha{}^{\mathrm{e}}} \tag{2.29}$$

with the other terms of eq. (2.27) being its first and second partial derivatives w.r.t the dimensionless normal coordinates. The $I_\alpha{}^{\mathrm{e}}$ term is the α-component of the inertia equilibrium tensor ($\mathbf{I}$), which is diagonal in the principal inertia system. This tensor is straightforwardly derived once the equilibrium molecular geometry is known:

$$\mathbf{I} = \sum_N M_N \left(R_N^2 \mathbf{1} - \mathbf{R}_K \mathbf{R}_K^T\right) \quad , \tag{2.30}$$

and it is related to B_e^α via:

$$B^\alpha_{\mathrm{e}} = \frac{1}{2I^{\mathrm{e}}_\alpha} \quad \text{with } \alpha = a,\ b \text{ and } c \quad . \tag{2.31}$$

B_e^α takes the denomination of A_e, B_e and, C_e depending on the α-axis. These are the equilibrium rotational constants and are the main parameters in rotational spectroscopy.

In eq. (2.28), ϕ_{rst} and ϕ_{rstu} are the cubic and quartic force constants, i.e. the third and quartic derivatives of the potential with respect to the dimensionless normal coordinates evaluated at the equilibrium.

Noted is that the last term, denoted as the Watson term, on the right-hand side of eq. (2.26) is negligible and does not influence the results representing a very small shift of the energy values obtained.

If the terms in eq. (2.27) and (2.28) are truncated to the first term on the right-hand side, one has the ro-vibrational Hamiltonian within the Rigid-rotor/Harmonic-oscillator (RRHO) approximation:

$$\mathbf{H}_{\mathrm{RRHO}} = \sum_\alpha B_e^\alpha \mathbf{J}_k^2 + \frac{1}{2}\sum_r \omega_r (\mathbf{p}_r^2 + \mathbf{q}_r^2) \quad . \tag{2.32}$$

The RRHO approximation is well-known and the eigenvalue problems can lead to relatively simple formulas for the energy depending on two main quantum numbers J and v, the rotational and vibrational quantum number, respectively. However, the approximations made with the truncation of eq. (2.27) and (2.28) are limited and the corresponding simplifications lead to results not in good agreement with experimental data. This leads to the introduction of the additional terms in the Watson Hamiltonian, which are typically treated via contact transformation or PT, resorting to Rayleigh-Schödinger PT (RSPT) in the latter case.

There are several consequences from the introduction of higher-order terms in the Hamiltonian, all improving the description of *ab initio* quantities. In particular, for the rotational Hamiltonian, the incorporation of higher-order terms in eq. (2.27) leads to the definition of an effective Hamiltonian which depends on an effective rotational constants. The latter parameters are expressed as:

$$B_v^\alpha = B_\mathrm{e}^\alpha - \sum_r a_r^\alpha \left(v_r + \frac{1}{2} \right) \tag{2.33}$$

where a_r^α are the vibration-rotation interaction constants, whose expression is:

$$\alpha_r^\alpha = -2B_\mathrm{e}^{2,\alpha} \left[\sum_\beta \frac{3(a_r^{\alpha\beta})^2}{4I_\beta^e} + \sum_s \frac{(\zeta_{r,s}^k)^2 (3(\omega_r)^2 + (\omega_s)^2}{\omega_r((\omega_r)^2 - (\omega_s)^2)} + \frac{1}{2} \sum_s \frac{\phi_{rrs} a_s^{\alpha\alpha}}{\omega_s^{3/2}} \right]. \tag{2.34}$$

In the equation above, the first term on the right-hand side is a correction to the moment of inertia, the second is due to Coriolis interactions that can occur between vibrational modes similar in frequency. The latter term is an anharmonic correction which depends on the semi-diagonal cubic force constants. The correction to B_e^α in eq. (2.33) can be used to derive the rotational constants of any vibrational state and the correction is present even for $v = 0$, i.e. the vibrational ground state. In fact, the latter is:

$$\Delta B_0^\alpha = B_0^\alpha - B_\mathrm{e}^\alpha = -\frac{1}{2} \sum_r a_r^\alpha \quad , \tag{2.35}$$

having ΔA_0, ΔB_0, and ΔC_0 for $\alpha = a, b \quad \text{and} \quad c$, respectively. The latter terms are typically small, accounting for a few percents (1-3%) of the ground state rotational constant. The greatest contribution is due to the equilibrium value, which accounts for a 97-99% of B_0^α.

This treatment somewhat introduces the effect of vibrations on the rotational energies thanks to the effective rotational constants. However, this is not enough to have a quantitative picture. Indeed, as the molecule is distorted by the vibrational motions, also its own rotational motion affects the overall structure and one needs to account for this fact [33, 35, 37, 138]. From definition of a semi-rigid Hamiltonian, one can retrieve

the corresponding rotational energies, that depends on rotational constants (as for the rigid rotor), but also on several centrifugal distortion terms. According to the order of the corrections considered, one can have quartic, sextic, octic (etc.) centrifugal distortion constants, but only the former two can be obtained via *ab initio* procedures [35,37].

Moving to the vibrational Hamiltonian, the introduction of correction terms able to incorporate anharmonicity leads to the possibility of having combinations bands ($\nu_i\nu_j$) and overtones ($2\nu_i$) other than anharmonic vibrational frequencies (ν_i). The expressions are:

$$\nu_r = \omega_r + 2\xi_{rr} + \frac{1}{2}\sum_{s\neq r}\xi_{rs} \tag{2.36}$$

$$2\nu_r = 2\nu_r + 2\xi_{rr} \tag{2.37}$$

$$\nu_r\nu_s = \nu_r + \nu_s + \epsilon_{rs} \quad ; \tag{2.38}$$

where the ξ's the so-called are anharmonic constants. The zero-point energy correction (ZPE) in anharmonic terms then becomes:

$$\mathrm{ZPE} = \frac{1}{2}\sum_r \omega_r + \frac{1}{2}\sum_r \nu_r + \xi_0 - \frac{1}{4}\sum_r \xi_{rr} \quad , \tag{2.39}$$

where the first term on the right-hand side is the harmonic ZPE. The ξ_0, ξ_{rr}, and ξ_{rs} terms are combination of second-, third- and, semi-diagonal fourth-derivatives of the energy with respect to the normal coordinates and include also Coriolis coupling constants. The detailed expressions are reported in refs. [139, 140].

Lastly, it should also be mentioned that rotational energies are also influenced by electric and/or magnetic interactions that cause important features (hyperfine structure) of the rotational spectrum because they split the rotational energy levels. Indeed, the hyperfine structure is very distinctive of the electronic arrangement of the molecule [35, 37] and therefore of its conformation [38]. Considering species with paired electrons (closed-shell species) and with one nucleus with nuclear spin (I) equal or grater than 1, the nuclear quadrupole coupling occurs, which is described by the nuclear quadrupole coupling constants that computationally are obtained from the product between the electric field gradient at the nucleus and its quadrupole moment. Another possible interaction in closed-shell species is that between the magnetic field generated by the rotation of the molecule and the magnetic field generated by nuclei with $I \geq \frac{1}{2}$ [141]. In this case, the nuclear spin-rotation interaction arises. In the end, the third interaction is the dipolar spin–spin coupling occurring when the molecule possesses more than one non-vanishing nuclear spins [35, 37, 142]. All the spectroscopic parameters associated to these interactions can be computed *ab initio* and are relevant for the present thesis since rotational spectroscopy is the main experimental counterpart in the computational spectroscopic characterisation of nc-complexes. A

summary of the computational tasks required to obtain the quantities required in rotational spectroscopy is given in table 2.1.

Table 2.1: Computational tasks required to obtain the most important rotational parameters in the case of closed-shell species.

Parameters	Symbol	Computational tasks
Equilibrium rotational constants	A_e, B_e, C_e	Equilibrium geometry, Geometry optimisation
Vibrational corrections to the rotational constants	ΔA_v, ΔB_v, ΔC_v	Cubic force field
Quartic centrifugal distortion constants	$\tau_{\alpha\beta\gamma\delta}$, D's	Harmonic force field
Sextic centrifugal distortion constants	$\tau_{\alpha\beta\gamma\delta\epsilon\nu}$, H's	Cubic force field
Nuclear quadrupole coupling constants	$\chi^K_{\alpha\beta}$, q^K	Electric field gradient as 1$^{\mathrm{st}}$ derivative of the energy w.r.t. the nuclear quadrupole moment
Nuclear spin-rotation tensor	C^K	2$^{\mathrm{nd}}$ derivative of energy w.r.t. angular momentum and nuclear spin
Dipolar spin-spin coupling constant	D^{KL}	Molecular geometry

Chapter 3

The junChS and junChS-F12 approaches

This chapter presents the development and validation of the junChS and junChS-F12 schemes, which have been briefly introduced in Chapter 2 (for the acronyms, see Chapter 1). The discussion is based on the corresponding research articles, published in the *Journal of Chemical Theory and Computation* [25,26], where the methodology is presented and validated. These composite schemes were developed with the aim of providing accurate energies and geometries for nc-complexes, thus being *ad hoc* approaches to provide a quantitative description of their energetics and properties. The PES ruling NCIs shows several challenging features related to the balanced description of several weak interactions (multipole-multipole, polarisation, dispersion, etc.) and, therefore, it is strongly dependent on the level of theory employed. For example, one could easily miss a key species in the analysis, then losing important information for the modelling. Accuracy in the description of the PES is also a mandatory requirement to enhance the interplay between theory and experiment; indeed, spectroscopic proprieties of a given species are strongly dependent on its electronic structure, especially in the case of rotational spectroscopy.

First, the chapter will illustrate the evaluation of interaction energies, giving details on the reference data set. Then, the discussion will move to nc-geometries that influence (i) the energetics, especially if computed at a non-suitable level of theory, and (ii) the rotational constants, i.e. the main quantity of interest when dealing with rotational spectroscopy [143]. All the coupled-cluster calculations involving geometry optimisations that are described in the following, have been performed with the CFOUR quantum-chemical program package [144, 145]. DFT and MP2 computations as well as CC single-point energy calculations up to the CCSD(T) level, have been obtained with the Gaussian suite of programs [91], while the F12 computations have been carried out with the Molpro program [146].

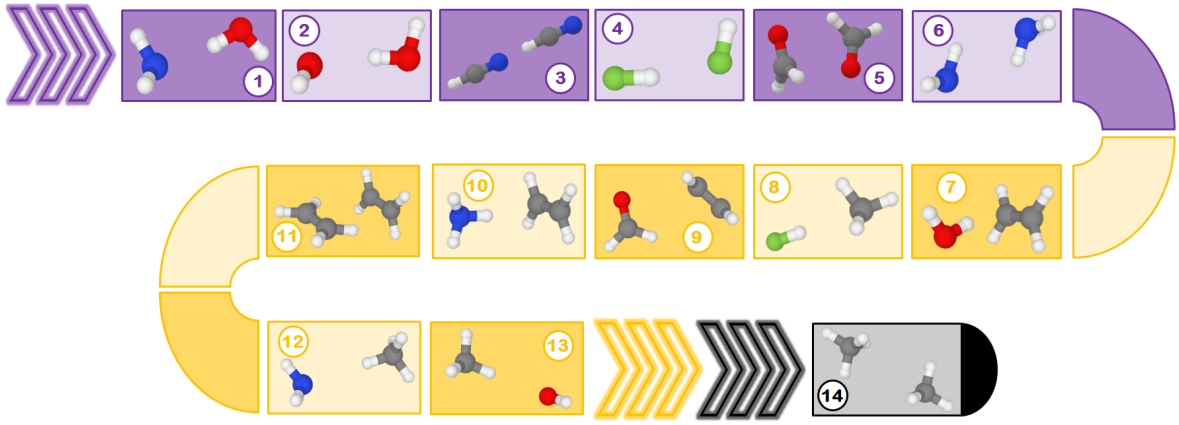


Figure 3.1: The A14 data set: 14 nc-complexes selected from the A24 data set used to develop the junChS and junChS-F12 composite schemes.

3.1 Interaction Energy

In the case of molecular systems formed by two molecular species, like the case of nc-complexes, the interaction energy is defined as the difference between the energy of the total system (complex, E_{AB}) and the energy of the separated partners (E_{A} and E_{B}), hereafter referred to as fragments [1,2]:

$$E_{\mathrm{int}} = E_{\mathrm{AB}} - (E_{\mathrm{A}} + E_{\mathrm{B}}) \; . \tag{3.1}$$

If E_{A} and E_{B} are computed at the geometry the fragments assume in the complex, i.e. a distorted geometry, and using the full basis set of the molecular complex, one obtains the so-called counterpoise (CP) corrected interaction energy [100]. Instead, the non-CP corrected (NCP) energy is obtained if the basis set of the fragment is employed on top of the distorted geometry. The arithmetic mean of the CP and NCP energies provides the half-CP energy. On a theoretical basis, the CP correction should be employed whenever a finite basis set is used and its importance should reduce with the extrapolation to the CBS limit. Indeed, the CP correction is introduced to correct the Basis Set Superposition Error (BSSE), which arises from the fact that the basis set of each fragment (in the nc-complex) is somewhat extended by the basis set functions belonging to the other fragment. However, at the CBS limit, the BSSE error should vanish.

The reference data set for the development of a new computational model for nc-complexes is a subset of the A24 data set [147]. In particular, the selected fourteen nc-complexes are shown in fig. 3.1 were selected, thus leading to the definition of the A14 data set. The reference geometries for these complexes were reported in ref. [147], while the accurate interaction energies were later refined in ref. [148], also considering the contributions from ref. [149]. To summarize, the reference structural parameters of the A14 set were derived from the minimisation of an energy gradient at the HF-

SCF/aug-cc-pVQZ level to which a correction due to the extrapolation of the MP2 energy to the CBS limit is added. The latter is obtained with the two-point extrapolation formula of eq. (2.23) and using the aug-cc-pVTZ and aug-cc-pVQZ basis sets. The minimised energy gradient also includes the energy difference between the CCSD(T) and MP2 energies computed using the aug-cc-pVDZ basis set. Each step was also CP-corrected in order to account for the BSSE error. This resulting geometry is indicated as "CBS-georef" in the following. On top of these geometries, accurate interaction energies were evaluated using a different composite scheme. This considers the correction due to the CBS limit of both the fc- and the CV contributions. The former term is taken from ref. [149], where a three-point extrapolation formula [147] was used together with large basis sets to reach a high accuracy. The CV term was instead extrapolated using a two-point formula and the aug-cc-pCVTZ and aug-cc-pCVQZ basis sets. In both cases, the CCSD(T) method was employed. These extrapolated corrections were added to an initial HF-SCF energy computed in conjunction with the aug-cc-pV5Z set. These energies should well fulfil the so-called chemical accuracy and will be shortly denoted as "ref" in the following.

Ideally, the ChS scheme (see eq. 2.25) for nc-complexes has to reproduce the "ref" energies employing basis sets of triple- and quadruple-zeta quality, also exploiting the systematic error compensation that occurs in the case of weakly bound compounds when using the MP2 method [150]. The aimed accuracy with respect to the "ref" values is 1% in relative terms, with maximum deviations up to 3%, and without absolute deviations larger than 0.2 kJ·mol^{-1}. This should lead to a methodology comparable to the reference one and able to uniformly describe the different types of NCIs, from hydrogen bonds (10-30 kJ·mol^{-1}) to weaker dispersion interactions (1-10 kJ·mol^{-1}). Therefore, the final goal of the new ChS scheme is to provide accurate and uniform interaction energies as the "ref" ones, but at a much lower computational cost, thus allowing its application to larger systems.

The first part of the discussion focuses on the definition of the best possible ChS scheme when considering methods dealing with canonical correlation. Then, explicitly correlated methods will be introduced in the development of a ChS approach. The "CBS-georef" will be used as reference for the first part of the discussio; the effects of using reference geometries based on DFT will be investigated before moving to the discussion to the explicitly correlated methods.

3.1.1 The junChS

The employed methodology was briefly described in the discussion of eq. (2.25), where it was also mentioned that triple- and quadruple-zeta quality basis sets are usually employed. To build a model able to describe correctly nc-complexes, different families of basis sets have been tested within the ChS scheme, also taking into account that diffuse functions are particularly relevant for the description of such systems. The full

incorporation of diffuse functions is achieved by employing the aug-cc-pVnZ set [104]. Starting from this, the removal of diffuse functions on the H atom leads to the first "seasonal" family of basis sets, i.e. jul-cc-pVnZ, also known as heavy-augmented (ha) basis sets [105]. Further removal of the highest angular momentum diffuse functions from non-H atoms defines the jun-cc-pVnZ family. Subsequently, further elimination of next angular momentum diffuse functions leads to the may-cc-pVnZ basis sets and then to the apr-cc-pVnZ family. The minimally augmented (maug-) basis set is obtained when only s and p diffuse functions are retained in the basis set, independently of its size. In fact, the maug-cc-pVnZ set is equivalent to jun-cc-pVDZ when n=D, but it is equal to the may-cc-pVTZ and apr-cc-pVQZ sets for n=T and Q, respectively [30]. All these families have been considered within the ChS scheme and employing two different approaches:

1. The basis sets containing diffuse functions regardless of being fully- or partially-augmented bases, are used to compute the $\Delta E^{\infty}_{\mathrm{MP2}}$ term and the CCSD(T) energy, thus directly including the effects of diffuse functions in the scheme. To simplify the discussion, a prefix indicating the type of augmented basis set used is added to the "ChS" acronym. For example, the "augChS" model is defined when the aug-cc-pVnZ family is used within the ChS approach.

2. An additional term is introduced in eq. (2.25), the $\Delta\alpha$ term. This is computed as the difference between a fc-MP2 computation in conjunction with a fully or partially augmented triple-zeta basis set and the same calculation with the corresponding non-augmented set. For example, $\Delta\alpha$ can be evaluated as the difference between the fc-MP2/aug-cc-pVTZ and fc-MP2/cc-pVTZ computations.

The $\Delta\alpha$ term being not justified by any theoretical consideration, a proper analysis of its effects is needed. According to the literature, the $\Delta\alpha$ term seems to be relevant for structural determinations [29, 136], but less reliable for energy evaluations, for which it is often removed [151, 152]. The CV contribution present in the ChS is computed either with the cc-pCVTZ set or the cc-pwCVTZ basis set [153], the implication of this choice will be discussed in the framework of each specific scheme. Lastly, in the following only CP-corrected energies will be addressed, but for the sake of completeness the NCP and half-CP errors will be also reported. As already mentioned, the "CBS-georef" structures are employed as reference for the following comparison and the errors introduced by the use of a cost-effective geometry (using DFT, *vide infra*) will be also discussed.

The extrapolation to the CBS limit of the ChS energies can be examined by considering the CV term fixed at the value obtained with the cc-pCVTZ basis set and inspecting the trend for all the species of the A14 data set. The ChS energies obtained with the cc-pV(T,Q)Z basis sets are reported in table 3.1 and the corresponding relative errors point

Table 3.1: ChS energies (kJ·mol^{-1}) using the cc-pVnZ basis sets.

Complex	"ref"	fc-CCSD(T)/TZ	ΔE^{∞}_{MP2}/(T,Q)Z	ΔE^{CV}_{MP2}	Energy	Relative Error	Error	$Error_{NCP}$	$Error_{half\text{-}CP}$
H_2O-H_2O	-21.0832	-18.3868	-1.9570	-0.1354	-20.4792	-2.86%	-0.60	-0.29	-0.44
NH_3-NH_3	-13.2131	-11.3498	-1.6052	-0.0733	-13.0283	-1.40%	-0.18	-0.40	-0.29
HF-HF	-19.2213	-17.2997	-1.2275	-0.1049	-18.6321	-3.07%	-0.59	-0.65	-0.62
H_2CO-H_2CO	-18.9284	-13.5169	-5.2213	-0.0987	-18.8369	-0.48%	-0.09	-0.55	-0.32
HCN-HCN	-19.9828	-18.2582	-1.3439	-0.0733	-19.6754	-1.54%	-0.31	-0.41	-0.36
C_2H_4-C_2H_4	-4.5647	-3.0318	-1.5070	-0.0439	-4.5827	0.39%	0.02	0.07	0.05
CH_4-CH_4	-2.2301	-1.2588	-0.8478	-0.0068	-2.1134	-5.23%	-0.12	-0.08	-0.10
H_2O-NH_3	-27.3759	-23.5124	-2.9325	-0.1677	-26.6126	-2.79%	-0.76	-0.40	-0.58
H_2O-C_2H_4	-10.7696	-8.0903	-2.4127	-0.0875	-10.5905	-1.66%	-0.18	0.37	0.09
C_2H_4-H_2CO	-6.7948	-4.5478	-2.1112	-0.0600	-6.7190	-1.12%	-0.08	-0.36	-0.22
C_2H_4-NH_3	-5.7865	-4.1166	-1.5012	-0.0488	-5.6666	-2.07%	-0.12	0.23	0.06
HF-CH_4	-6.9162	-4.7205	-2.0688	-0.0860	-6.8753	-0.59%	-0.04	0.07	0.02
H_2O-CH_4	-2.8242	-2.1376	-0.4039	-0.0286	-2.5701	-9.00%	-0.25	-0.39	-0.32
NH_3-CH_4	-3.2175	-2.7172	-0.1872	-0.0352	-2.9396	-8.64%	-0.28	-0.19	-0.23
MAE						2.91%	0.26	0.32	0.26

Table 3.2: augChS energies (kJ·mol^{-1}).

Complex	"ref"	fc-CCSD(T)/augTZ	ΔE^{∞}_{MP2}/aug(T,Q)Z	ΔE^{CV}_{MP2}	Energy	Relative Error	Error	$Error_{NCP}$	$Error_{half\text{-}CP}$
H_2O-H_2O	-21.0832	-19.9250	-1.1894	-0.1354	-21.2498	0.79%	0.17	0.44	0.30
NH_3-NH_3	-13.2131	-12.5692	-0.6682	-0.0733	-13.3107	0.74%	0.10	0.12	0.11
HF-HF	-19.2213	-17.9928	-1.3570	-0.1049	-19.4547	1.21%	0.23	0.75	0.49
H_2CO-H_2CO	-18.9284	-17.6285	-1.5474	-0.0987	-19.2746	1.83%	0.35	0.43	0.39
HCN-HCN	-19.9828	-19.3043	-0.5778	-0.0733	-19.9554	-0.14%	-0.03	-0.11	-0.07
C_2H_4-C_2H_4	-4.5647	-4.3068	-0.3787	-0.0439	-4.7294	3.61%	0.16	0.06	0.11
CH_4-CH_4	-2.2301	-2.0849	-0.1707	-0.0068	-2.2624	1.45%	0.03	-0.02	0.004
H_2O-NH_3	-27.3759	-26.1081	-1.2329	-0.1677	-27.5087	0.49%	0.13	0.36	0.25
H_2O-C_2H_4	-10.7696	-10.1895	-0.6334	-0.0875	-10.9104	1.31%	0.14	0.07	0.11
C_2H_4-H_2CO	-6.7948	-6.3544	-0.5124	-0.0600	-6.9268	1.94%	0.13	0.13	0.13
C_2H_4-NH_3	-5.7865	-5.4694	-0.3740	-0.0488	-5.8922	1.83%	0.11	-0.06	0.02
HF-CH_4	-6.9162	-6.4952	-0.4980	-0.0860	-7.0792	2.36%	0.16	0.56	0.36
H_2O-CH_4	-2.8242	-2.6011	-0.2057	-0.0286	-2.8354	0.40%	0.01	0.01	0.01
NH_3-CH_4	-3.2175	-3.0165	-0.2344	-0.0352	-3.2861	2.13%	0.07	0.06	0.06
MAE						1.38%	0.13	0.23	0.15

out that the energies are underestimated – on average – by 2.9%, with a maximum discrepancy of 9%. Large errors are observed especially for complexes containing CH_4, which are indeed expected to benefit from the inclusion of diffuse functions. This is confirmed by the data of table 3.2, which collects the augChS energies. Interestingly, the errors for the methane-containing systems are reduced to values well within the target accuracy, i.e. below the 3%. Furthermore, it is noted that the energies of the augChS approach are all overestimated, thus being lower (more negative) than the reference energies. The fully augmented approach seems promising, although the description of π systems slightly deteriorates when diffuse functions are included. For example, the relative error for the ethene dimer increases from 0.39% to 3.6% and a worsening is noted also for the C_2H_4-H_2CO complex. As demonstrated by the data of tables 3.1 and 3.2, there is no advantage in using NCP energies and the half-CP error is still worse than that associated to the CP corrected counterpart.

Since the interaction energy is overestimated by the augChS model and underesti-

Table 3.3: julChS energies (kJ·mol^{-1})

Complex	"ref"	fc-CCSD(T)/julTZ	ΔE^{∞}_{MP2}/jul(T,Q)Z	ΔE^{CV}_{MP2}	Energy	Relative Error	Error	Error$_{NCP}$	Error$_{half\text{-}CP}$
H_2O-H_2O	-21.0832	-19.6966	-1.2103	-0.1354	-21.0423	-0.19%	-0.04	0.09	0.02
NH_3-NH_3	-13.2131	-12.2108	-0.9287	-0.0733	-13.2128	0.00%	-0.001	0.09	0.04
HF-HF	-19.2213	-17.8393	-1.2781	-0.1049	-19.2223	0.01%	0.001	0.17	0.09
H_2CO-H_2CO	-18.9284	-17.4291	-1.6128	-0.0987	-19.1406	1.12%	0.21	0.33	0.27
HCN-HCN	-19.9828	-19.1640	-0.6978	-0.0733	-19.9351	-0.24%	-0.05	-0.15	-0.10
C_2H_4-C_2H_4	-4.5647	-4.0271	-0.5995	-0.0439	-4.6705	2.32%	0.11	0.10	0.11
CH_4-CH_4	-2.2301	-1.8353	-0.3506	-0.0068	-2.1927	-1.67%	-0.04	-0.18	-0.11
H_2O-NH_3	-27.3759	-25.6680	-1.5765	-0.1677	-27.4122	0.13%	0.04	0.10	0.07
H_2O-C_2H_4	-10.7696	-9.8531	-0.8998	-0.0875	-10.8404	0.66%	0.07	0.07	0.07
C_2H_4-H_2CO	-6.7948	-6.1293	-0.6630	-0.0600	-6.8523	0.85%	0.06	0.15	0.11
C_2H_4-NH_3	-5.7865	-5.1967	-0.5967	-0.0488	-5.8422	0.96%	0.06	0.08	0.07
HF-CH_4	-6.9162	-6.1143	-0.7931	-0.0860	-6.9934	1.12%	0.08	0.08	0.08
H_2O-CH_4	-2.8242	-2.4392	-0.3136	-0.0286	-2.7814	-1.52%	-0.04	-0.03	-0.04
NH_3-CH_4	-3.2175	-2.7902	-0.4066	-0.0352	-3.2320	0.45%	0.01	0.02	0.02
MAE						0.80%	0.05	0.12	0.09

mated by the ChS model, a possible way out might be offered by a partial augmentation of the basis sets. The results of the julChS approach are reported in table 3.3 and show a further reduction of the relative error, from 1.4% for augChS to 0.8% for julChS. For the methane-containing species, the results are still in good agreement with the "ref" values, the error – in absolute terms – being lower than 0.1 kJ·mol^{-1} for all the systems. The largest error is observed for the C_2H_4 dimer, this being 2.3% in relative terms and 0.11 kJ·mol^{-1} in absolute terms. The julChS approach describes uniformly all complexes and very small absolute errors, i.e. 0.001 kJ·mol^{-1}, are observed for H-bonded adducts such as the NH_3 and HF dimers.

The jul-cc-pVnZ family of basis sets allows a large saving of computational times with respect to the fully augmented sets, its use in highly accurate composite approaches being recently reported [154]. Since the removal of diffuse functions of H atoms permits the reduction of the computational cost without affecting the accuracy, the following step is to remove diffuse functions also from heavy atoms, thus obtaining the junChS approach. The results for this scheme are collected in table 3.4, where a minor worsening of the error is observed. Indeed, the Mean Absolute Error (MAE) for CP energies is 0.1 kJ·mol^{-1} in absolute terms, to be compared with 0.05 kJ·mol^{-1} for julChS and 0.13 kJ·mol^{-1} for augChS. The target accuracy is still matched, the relative error being of about 1%. Furthermore, the only outlier (with an error larger than 3%) has a deviation in absolute terms below 0.2 kJ·mol^{-1}. While the julChS model provides energies that are both over- and under-estimated, the junChS approach provides mainly underestimated values. For this reason, the NCP energies do not lead to any improvement. Lastly, it should also be mentioned that the methane-containing species are still well described when employing the jun-cc-VnZ family, with errors below 1.5%. Since junChS provides a good accuracy with a further lowering of the computational cost with respect to julChS, it represents the most promising approach for the description of medium-sized nc-molecular adducts.

Table 3.4: junChS energies (kJ·mol^{-1}).

Complex	"ref"	fc-CCSD(T)/junTZ	ΔE^{∞}_{MP2}/jun(T,Q)Z	ΔE^{CV}_{MP2}	Energy	Relative Error	Error	$Error_{NCP}$	$Error_{half\text{-}CP}$
H_2O-H_2O	-21.0832	-19.2986	-1.5414	-0.1354	-20.9754	-0.51%	-0.11	0.17	0.03
NH_3-NH_3	-13.2131	-11.8413	-1.3507	-0.0733	-13.2653	0.40%	0.05	0.09	0.07
HF-HF	-19.2213	-17.3511	-1.7673	-0.1049	-19.2233	0.01%	0.002	0.23	0.12
H_2CO-H_2CO	-18.9284	-16.1715	-3.0240	-0.0987	-19.2942	1.93%	0.37	0.56	0.46
HCN-HCN	-19.9828	-18.8230	-1.0026	-0.0733	-19.8989	-0.42%	-0.08	-0.23	-0.16
C_2H_4-C_2H_4	-4.5647	-3.7570	-0.9481	-0.0439	-4.7489	4.04%	0.18	0.24	0.21
CH_4-CH_4	-2.2301	-1.6317	-0.6187	-0.0068	-2.2572	1.22%	0.03	-0.003	0.01
H_2O-NH_3	-27.3759	-25.2160	-1.9219	-0.1677	-27.3056	-0.26%	-0.07	0.07	0.002
H_2O-C_2H_4	-10.7696	-9.5029	-1.2704	-0.0875	-10.8608	0.85%	0.09	0.25	0.17
C_2H_4-H_2CO	-6.7948	-5.6830	-1.2112	-0.0600	-6.9543	2.35%	0.16	0.31	0.23
C_2H_4-NH_3	-5.7865	-4.9695	-0.8644	-0.0488	-5.8828	1.66%	0.10	0.20	0.15
HF-CH_4	-6.9162	-5.8334	-1.0948	-0.0860	-7.0142	1.42%	0.10	0.12	0.11
H_2O-CH_4	-2.8242	-2.3273	-0.4490	-0.0286	-2.8049	-0.68%	-0.02	-0.005	-0.01
NH_3-CH_4	-3.2175	-2.6783	-0.5145	-0.0352	-3.2280	0.33%	0.01	0.01	0.01
MAE						1.14%	0.10	0.18	0.12

However, further tests are in order to analyse if the number of diffuse functions in the basis sets can be further reduced. Therefore, the maug-cc-pVnZ family of bases was also considered, thus defining the maugChS approach. As evident from the values reported in appendix A, the maugChS energies are worse than those from all previous schemes, with a MAE for CP energies of 6.8% and maximum errors up to 21%. This behaviour can be explained by the fact that the maug-cc-pVnZ family is not hierarchical because, as mentioned above, the triple- and quadruple-zeta basis sets do not belong to the "same month of the calendar". Thus, they do not represent a smoothly convergent basis set in terms of diffuse functions. This is easily understood from the analysis of the mayChS results, also provided in Appendix A. In the latter scheme, the CCSD(T) contribution is the same as in the maugChS model, but the combination may-cc-pV(T,Q)Z is employed for the extrapolation to the CBS limit. The relative error of the mayChS scheme drops from 8.8% (maugChS) to 2.4%. Using mayChS, the energy of the methane dimer is reproduced with an error slightly below 2%, while the discrepancy for the mixed cluster, i.e. CH_4-H_2O, CH_4-NH_3 and CH_4-HF, is around 3%, a marked improvement with respect to the maugChS scheme, whose associated errors are around 10-20%. Although the mayChS scheme performs well, the errors are out of the target range and the methodology of choice remains the junChS approach.

A summary of all previous data is schematically provided in fig. 3.2, where the CP, NCP and half-CP errors are reported for all the above mentioned approaches. It is readily seen how the maugChS model performs inadequately compared to the other composite schemes. The accuracy of the junChS and julChS models bis similar and both represent an improvement compared to the non-augmented and fully augmented counterparts, i.e. ChS and augChS, respectively. The trend is similar for CP, NCP and half-CP energies and a note in favor of the junChS scheme is that the half-CP energies are quite close the CP counterparts, thus implying that the CBS limit is only slightly underestimated. In the following discussion, which focuses only on the junChS scheme,

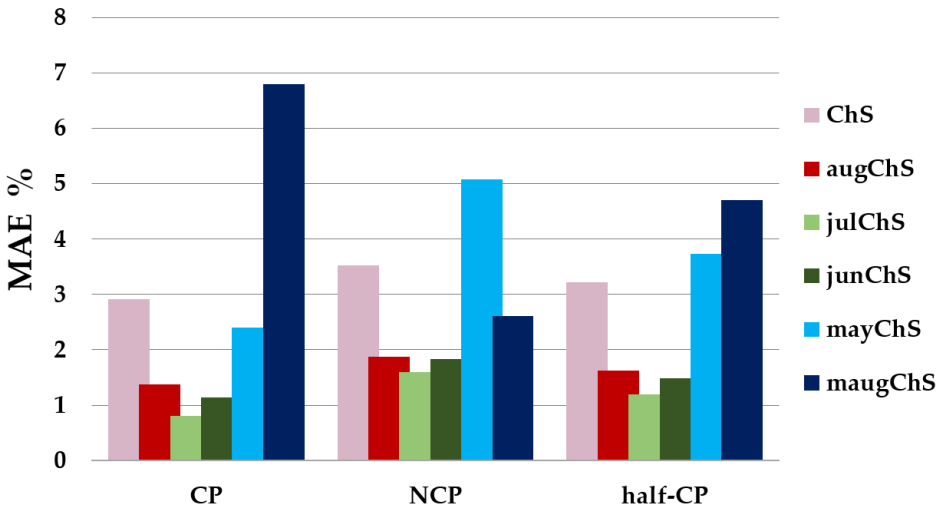


Figure 3.2: Mean Absolute Error (%) for the following schemes: ChS (pink bars), augChS (red bars), julChS (green bars), junChS (dark-green bars), mayChS (blue bars) and, maugChS (dark blue bars). The errors are computed w.r.t. the "ref" interaction energies.

only the CP energies will be considered since they appear as the most promising.

Once established that the presence of diffuse functions is needed to provide accurate results, the next step is to investigate if this contribution can be added via the separate $\Delta\alpha$ term previously mentioned. In this case, one exploits the ChS model, thereby employing the cc-pVnZ basis sets with n=(T,Q) and adds the contribution due to the difference between MP2 computations with the aug-cc-pVTZ and cc-pVTZ basis sets, the latter implying a computation already required in the scheme. The results of the ChS+$\Delta\alpha$(aug) approach are reported in table 3.5 and should be compared with those of table 3.2 to address the effectiveness of the additive $\Delta\alpha$ term. Already from the comparison of the MAEs, i.e. 12% for ChS+$\Delta\alpha$(aug) and 1.38% for augChS, it is apparent that the diffuse functions influence the energies in a manner that can not be considered as additive. Deviations of about 20% for the dimers of formaldehyde, methane and ethene are observed and errors of the same magnitude are also obtained for the majority of the mixed clusters. The differences between augChS and ChS+$\Delta\alpha$(aug) are not recovered by other types of basis sets. Indeed, in the reference paper [25], the partially augmented jun-cc-pVTZ and jul-cc-pVTZ basis sets were also employed for the computation of the $\Delta\alpha$ term, without however observing significant improvement, thus confirming the impossibility to consider the contribution of diffuse functions in an additive manner.

Lastly, a comment on the core-valence (CV) term within the ChS approach is deserved. As evident from the previous tables, the CV term is generally small but non-negligible. Indeed, when aiming at a quantitative accuracy, this term should always be included since its value is comparable or larger than the overall absolute error. For example, the CV term leads to contributions larger than the error for 8 systems out

Table 3.5: ChS+$\Delta\alpha$(aug) energies (kJ·mol^{-1})

Complex	fc-CCSD(T)/TZ	$\Delta E^{\infty}_{\rm MP2}$/(T,Q)Z	$\Delta E^{\rm CV}_{\rm MP2}$	$\Delta\alpha$(aug)	Energy	"ref"	Relative Error	Error
H_2O-H_2O	-18.3968	-1.9570	-0.1354	-1.0176	-21.4968	-21.0832	1.96%	0.41
NH_3-NH_3	-11.3498	-1.6052	-0.0733	-0.9211	-13.9494	-13.2131	5.57%	0.74
HF-HF	-17.2997	-1.2275	-0.1049	-0.0722	-18.7043	-19.2213	-2.69%	-0.52
H_2CO-H_2CO	-13.5169	-5.2213	-0.0987	-3.7534	-22.5903	-18.9284	19.36%	3.66
HCN-HCN	-18.2582	-1.3439	-0.0733	-0.8891	-20.5645	-19.9828	2.91%	0.58
C_2H_4-C_2H_4	-3.0318	-1.5070	-0.0439	-1.1774	-5.7600	-4.5647	26.19%	1.20
CH_4-CH_4	-1.2588	-0.8478	-0.0068	-0.6666	-2.7801	-2.2301	24.67%	0.55
H_2O-NH_3	-23.5124	-2.9325	-0.1677	-2.1569	-28.7695	-27.3759	5.09%	1.39
H_2O-C_2H_4	-8.0903	-2.4127	-0.0875	-1.8544	-12.4449	-10.7696	15.56%	1.68
C_2H_4-H_2CO	-4.5478	-2.1112	-0.0600	-1.6154	-8.3344	-6.7948	22.66%	1.54
C_2H_4-NH_3	-4.1166	-1.5012	-0.0488	-1.1854	-6.8521	-5.7865	18.42%	1.07
HF-CH_4	-4.7205	-2.0688	-0.0860	-1.5040	-8.3793	-6.9162	21.16%	1.46
H_2O-CH_4	-2.1376	-0.4039	-0.0286	-0.2821	-2.8522	-2.8242	0.99%	0.03
NH_3-CH_4	-2.7172	-0.1872	-0.0352	-0.0938	-3.0334	-3.2175	-5.72%	-0.18
MAE							12.21%	1.07

of 14 in the case of the julChS approach and the same applies to 6 complexes in the junChS scheme. From an inspection of table 3.2, it is evident that by error compensation one can remove the CV energy from the augChS model and obtain better statistics, but the same behaviour is not observed for the julChS and junChS schemes (see tables 3.3 and 3.4 and comments in ref. [25]). Since the CV term is relevant even for light nc-complexes, an improved basis set has also been investigated for its computation, i.e. the cc-pwCVTZ set. The latter should improve the description of intershell correlation, which is the most relevant contribution in the $\Delta E^{\rm CV}_{\rm MP2}$ term. For the junChS scheme, a nearly negligible worsening of the MAE (0.04% and 0.01 kJ·mol^{-1} in relative and absolute terms, respectively) is noted from the comparison of table 3.4 and table 3.6, the latter collecting the junChS values for the CV term computed with the cc-pwCVTZ basis set. Since the weighted core-valence basis set should provide more accurate energies and geometries for atoms belonging to the third-row of the periodic table [124, 153], it was chosen as the appropriate basis set for the computation of the CV term.

Based on the discussion above, the junChS model is the model of choice for the accurate characterisation of nc-complexes. Therefore, its formulation is here summarised:

$$E(\text{junChS}) = E(\text{fc-CCSD(T)/junTZ}) + \Delta E^{\infty}_{\rm MP2}/\text{jun(T,Q)Z} + \Delta E^{\rm CV}_{\rm MP2}/\text{wCVTZ}\ . \quad (3.2)$$

In details, the energy in eq. (3.2) in composed by:

1. The fc-CCSD(T)/jun-cc-pVTZ energy evaluation provides the leading energetic term: fc-CCSD(T)/junTZ.

Table 3.6: junChS energies (kJ·mol^{-1}) using the cc-pwCVTZ basis set for the CV term.

Complex	fc-CCSD(T)/junTZ	$\Delta E^{\infty}_{\rm MP2}$/jun(T,Q)Z	$\Delta E^{\rm CV}_{\rm MP2}$	Energy	"ref"	Relative Error	Error
H_2O-H_2O	-19.2986	-1.5414	-0.1444	-20.9844	-21.0832	-0.47%	-0.10
NH_3-NH_3	-11.8413	-1.3507	-0.0801	-13.2721	-13.2131	0.45%	0.06
HF-HF	-17.3511	-1.7673	-0.1106	-19.2290	-19.2213	0.04%	0.01
H_2CO-H_2CO	-16.1715	-3.0240	-0.0895	-19.2850	-18.9284	1.88%	0.36
HCN-HCN	-18.8230	-1.0026	-0.0652	-19.8908	-19.9828	-0.46%	-0.09
C_2H_4-C_2H_4	-3.7570	-0.9481	-0.0495	-4.7546	-4.5647	4.16%	0.19
CH_4-CH_4	-1.6317	-0.6187	-0.0069	-2.2573	-2.2301	1.22%	0.03
H_2O-NH_3	-25.2160	-1.9219	-0.1814	-27.3193	-27.3759	-0.21%	-0.06
H_2O-C_2H_4	-9.5029	-1.2704	-0.0963	-10.8696	-10.7696	0.93%	0.10
C_2H_4-H_2CO	-5.6830	-1.2112	-0.0656	-6.9598	-6.7948	2.42%	0.17
C_2H_4-NH_3	-4.9695	-0.8644	-0.0545	-5.8884	-5.7865	1.76%	0.10
HF-CH_4	-5.8334	-1.0948	-0.0976	-7.0258	-6.9162	1.59%	0.11
H_2O-CH_4	-2.3273	-0.4490	-0.0323	-2.8086	-2.8242	-0.55%	-0.02
NH_3-CH_4	-2.6783	-0.5145	-0.0401	-3.2329	-3.2175	0.48%	0.02
MAE						1.18%	0.11

2. The energies obtained with the jun-cc-pVnZ (n=T,Q) basis sets and employing the MP2 method are extrapolated to the CBS limit using eq. (2.23). The corresponding contribution ($\Delta \mathrm{E}^{\infty}_{\rm MP2}/\mathrm{jun(T,Q)Z}$) is obtained as follows:

$$\Delta E^{\infty}_{\rm MP2} = E^{\infty}_{\rm MP2} - E^{\rm junTZ}_{\rm MP2} \quad , \tag{3.3}$$

   the latter term being the fc-MP2/jun-cc-pVTZ energy.

3. The CV term is computed using the cc-pwCVTZ basis set and the MP2 method: $\Delta E^{\rm CV}_{\rm MP2}/\mathrm{wCVTZ}$.

Before moving to the re-elaboration of the ChS scheme in terms of explicitly correlated methodologies, the effects due to the employment of a different reference geometry than the "CBS-georef" have to be addressed. Since the junChS model aims at being applied to medium-sized nc-complexes, it would be computationally effective to introduce a reference geometry based on DFT, instead of resorting to expensive composite approaches as that employed in ref. [147]. In particular, two double-hybrid functionals have been tested: B2PLYP-D3BJ and revDSD-PBEP-D3BJ, hereafter denoted as B2 and revDSD, respectively. In panel (a) of fig. 3.3 (white background), the MAEs for the CP- and NCP-junChS energies computed on top of the reference geometries ("CBS-georef") as well as on top of the CP-B2/may-cc-pVTZ and NCP-revDSD/jun-cc-pVTZ geometries are reported. The use of a different reference geometry in the junChS model slightly increases the statistical errors, which are still very similar between CP-B2 and "CBS-georef", while NCP-revDSD geometries give somewhat higher errors when CP-energies are considered. However, NCP geometries are far less computationally demanding even if a larger basis set, i.e. jun-cc-pVTZ, is used instead of the may-cc-pVTZ. Another point in favour of NCP-revDSD geometries is given by the statistics

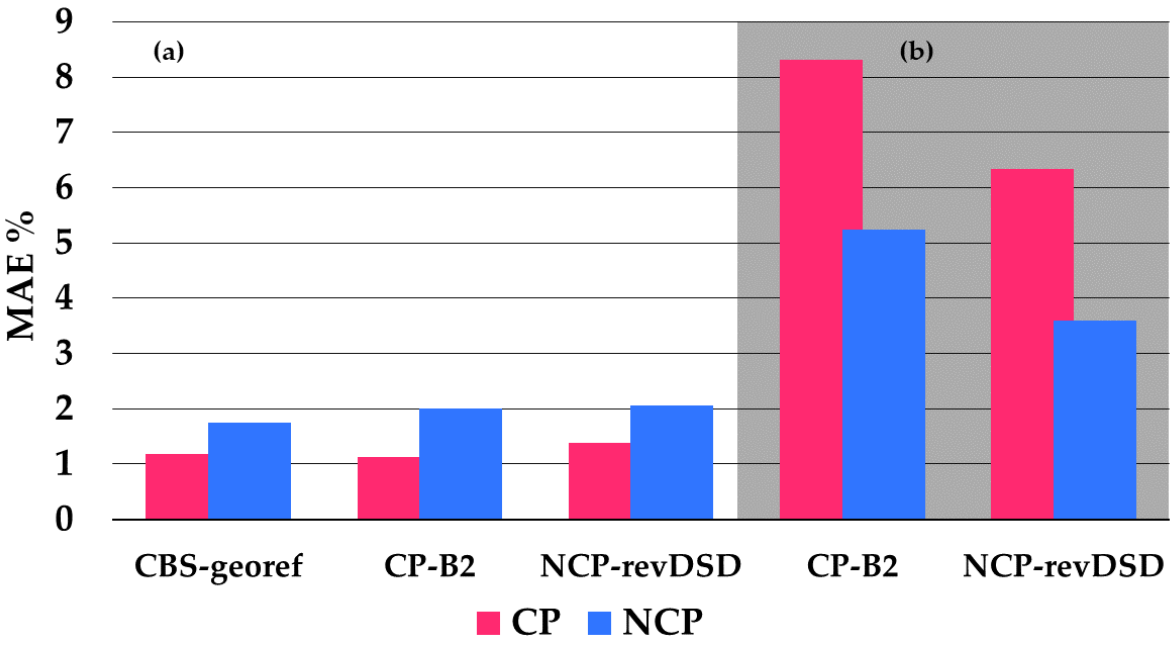


Figure 3.3: Panel (a): Comparison of CP and NCP energies obtained via the junChS approach on top of "CBS-georef", CP-B2, and NCP-revDSD geometries. Panel (b): CP and NCP energies obtained at the B2 and the revDSD levels on top of CP-B2 and NCP-revDSD geometries, respectively.

reported in panel (b) (gray background) of fig. 3.3, where the CP- and NCP- energies directly obtained from B2 and revDSD calculations are shown. In both cases, the errors associated to the revDSD energies are a few percent smaller than the B2 counterparts, thus implying that the corresponding equilibrium geometries should also be more reliable. In both cases, the NCP MAEs are lower than the CP counterparts, a clear indication that the interaction energies are strongly underestimated, as expected for a DFT technique. Therefore, DFT methodologies based on double-hybrid functionals can be used to retrieve the structure of nc-complexes, but not to compute accurate interaction energies. For the junChS approach, the reference equilibrium structures are those provided by the NCP-revDSD/junTZ model which results is a large computational saving with respect to the CP-B2/mayTZ ones and leads only to a minor worsening of the errors.

3.1.2 The junChS-F12

Explicitly correlated F12 methods are particularly promising in quantum chemistry in view of their fast convergence to the CBS limit. Theoretically, as mentioned in Chapter 2, the explicitly correlated energy should converge as $(L+1)^{-7}$, with L being the order of the highest angular momentum function present in the basis set. However, due to the approximation introduced in their formulation, the convergence is somewhat slower. Therefore, the set-up of an explicitly correlated ChS model needs to investi-

gate (i) the appropriate extrapolation formula to correctly describe the convergence to the CBS limit and (ii) the best family of basis sets to be used. The accurate reference data employed are still the "ref" energies [147]. The same energies without the CV contribution (hereafter "ref-CBS") are also used in the following. As the NCP-revDSD geometries have been chosen as reference for junChS, it is somewhat natural to exploit the same structures in the F12 counterpart. The families of basis sets tested are: may-cc-pVnZ, jun-cc-pVnZ, aug-cc-pVnZ, and cc-pVnZ-F12. The latter family was specifically developed for F12 methods, as explained in Chapter 2. In the following, in addition to the extrapolation to the CBS limit within the ChS approach (i.e. at the MP2-F12 level), the results obtained from the extrapolation to the CBS limit of the CCSD(T)-F12 energy are also reported, thus leading to the "CBS-F12" scheme. As before, a prefix is used to indicate the family of basis set employed, e.g. mayCBS-F12 indicates the CCSD(T)-F12 energy extrapolated to the CBS limit in conjunction with the may-cc-pV(T,Q)Z basis sets. If the CV correction is incorporated in the CBS scheme, then the "CBS+CV" approach is obtained and the prefix rule still holds. The CV term is computed in nearly all cases with the cc-pwCVTZ set; only when the cc-pVnZ-F12 family is used, the basis set employed is cc-pCVTZ-F12.

It is noted that CCSD(T)-F12 actually refers to the CCSD-F12 method to which the conventional (T) contribution is added. The CCSD(T)-F12a ansatz was used for single-point calculations, while the CCSD(T)-F12b variant was employed for energy extrapolations. In all computations, the geminal exponent γ was fixed to 1.0 a_0^{-1} and, in the case of cc-pVnZ-F12, the basis sets used for RI and DF were the default ones in the Molpro program [146]. For all the other cases, the DF and RI basis sets were set to the augmented versions of the orbital ones.

Extrapolation to the CBS limit

The discussion on the ChS-F12 schemes starts from the extrapolation to the CBS limit, since there is no well established formula for such contribution in the literature. After preliminary investigations, two possible extrapolation formula have been considered. The first one is that of eq. (2.23), i.e. the n^{-3} formula, while the second possibility is the same formula but with "-5" as exponent. The comparison between these two extrapolations is carried out using the "ref-CBS" energies as reference, thus only focusing on the CBS part of the reference energies. In order to perform a better analysis of the CBS limit in the case of F12 methods, also the CCSD(T)-F12 extrapolated energies are considered for comparison. Table 3.7 collects the CBS-F12 energies obtained with different basis sets, while the ChS counterparts (without CV term) are reported in table 3.8. Focusing on the former, the CBS methodologies better recover the "ref-CBS" energies if the n^{-5} formula is employed, with the corresponding CP-corrected energies deviating by only 0.08 kJ·mol^{-1} for three schemes out of the four tested. In fact, the (D,T) combination of the cc-pVnZ-F12 basis sets performs better if the "-3" exponent

is employed, probably because of some error compensation. Similar discrepancies are also observed for the CBS extrapolation of the ChS-F12 schemes of table 3.8. However, for the MP2-F12 method, the convergence seems to be more effective when the n^{-3} formula is employed, regardless of the family of basis sets. Relative MAEs (%) as low as 0.66% are observed for the junChS-F12 CBS scheme and similar results are obtained only if, within the ChS-F12 approach, the extrapolation is carried out with the cc-pV(T,Q)Z-F12 basis sets. Further hints can be obtained by comparing CP and NCP values of both the previously mentioned tables. In fact, on a theoretical basis, the BSSE should vanish at the CBS limit and the CP and NCP energies should be identical. This is the case for the cc-pV(T,Q)Z-F12 combination of basis sets, as well as for the may- and jun- families, when employing the n^{-5} formula in the CBS-F12 scheme (table 3.7), with the absolute error of NCP energies being only 0.03 kJ·mol^{-1} larger than the CP one. A similar behaviour is also observed for the ChS schemes with the cc-pV(T,Q)Z-F12 and jun-cc-pV(T,Q)Z combinations and using the n^{-3} extrapolation formula. Indeed, the junChS-F12 model has a CP absolute error of 0.05 kJ·mol^{-1}, to be compared with 0.07 kJ·mol^{-1} obtained for the NCP counterpart. A similar small difference is also noted for the ChS-F12 approach exploiting the (T,Q) combination of the cc-pVnZ-F12 family. The latter scheme is also the only scheme to have (minor) benefits from the n^{-5} extrapolation when the NCP energy is considered.

It can be concluded that, for the CBS-F12 schemes, the "-5" exponent should be preferred, while the "n^{-3}" formula is still retained for the ChS-F12 schemes.

Table 3.7: Relative and absolute errors[a,b] for theCBS-F12 approach in combination with different basis sets on top of the revDSD geometries.

		n^{-3}		n^{-5}	
Basis Set		Relative error	Absolute error	Relative error	Absolute error
	CP	0.86%	0.07	1.68%	0.11
cc-pV(D,T)Z-F12	NCP	1.77%	0.19	1.51%	0.16
	half-CP	1.17%	0.12	1.07%	0.09
	CP	1.22%	0.13	0.77%	0.08
cc-pV(T,Q)Z-F12	NCP	0.93%	0.11	0.85%	0.11
	half-CP	1.03%	0.12	0.81%	0.09
	CP	2.80%	0.24	0.81%	0.08
may-cc-pV(T,Q)Z	NCP	2.13%	0.14	1.23%	0.10
	half-CP	1.77%	0.15	0.93%	0.08
	CP	1.64%	0.16	0.76%	0.08
jun-cc-pV(T,Q)Z	NCP	1.21%	0.11	1.18%	0.11
	half-CP	1.25%	0.13	0.91%	0.09

[a]Errors evaluated with respect to "ref-CBS" reference energies. [b]Absolute errors in kJ·mol^{-1}.

Table 3.8: Relative and absolute errors[a,b] of the extrapolation to the CBS limit within ChS-F12 approaches.

		n^{-3}		n^{-5}	
Basis Set		Relative error	Absolute error	Relative error	Absolute error
	CP	2.27%	0.18	2.88%	0.23
cc-pV(D,T)Z-F12	NCP	1.10%	0.10	1.35%	0.18
	half-CP	1.30%	0.09	1.62%	0.15
	CP	0.73%	0.07	0.70%	0.09
cc-pV(T,Q)Z-F12	NCP	0.90%	0.09	0.79%	0.06
	half-CP	0.75%	0.07	0.71%	0.07
	CP	1.25%	0.10	1.90%	0.14
may-cc-pV(T,Q)Z	NCP	2.87%	0.19	1.59%	0.09
	half-CP	2.00%	0.14	1.65%	0.10
	CP	0.66%	0.05	1.03%	0.07
jun-cc-pV(T,Q)Z	NCP	1.05%	0.07	1.17%	0.10
	half-CP	0.84%	0.06	0.77%	0.05

[a]Errors evaluated with respect to "ref-CBS" reference energies. [b]Absolute errors in kJ·mol^{-1}.

The CV contribution

At this point the statistical performances of the schemes containing also the CV term have to be analysed. In this case, the reference energies used incorporate it and are those indicated as "ref". Table 3.9 reports the energies of the CBS+CV-F12 composite schemes in conjunction with the may- and jun-cc-pV(T,Q)Z combinations of basis sets as well as the cc-pVnZ-F12 basis sets, with n=T and Q. The lowest errors are observed for the junCBS+CV-F12 scheme, with the relative MAE being 0.79% when the CP correction is included. In absolute terms, the averaged error on the interaction energies is 0.07 kJ·mol^{-1}. These good results are followed by those of the mayCBS+CV-F12 scheme that shows a MAE of 0.87% when the CP correction is included. Instead, a relative error of 1.36% is obtained for the CBS+CV-F12 scheme, which is slightly improved by the consideration of half-CP values.

The results of the ChS-F12 approaches are reported in table 3.10. For the may-cc-pVnZ family of basis sets a worsening with respect to the corresponding CBS+CV-F12 scheme is observed, regardless of the CP correction. For example, the NCP energies of the mayCBS+CV-F12 scheme results in a MAE of 1.18%, while for the mayChS-F12 approach the error is 2.8%; in absolute terms, we have 0.09 kJ·mol^{-1} vs 0.19 kJ·mol^{-1}. For the CP energies, the same trend is noted, with the error increasing from 0.87% to 1.23%. The junChS-F12 and ChS-F12 approaches reproduce the reference energies with errors below 1% and show statistics that are better than their CBS+CV counterparts. In details, the MAE of the junChS-F12 scheme is 0.68%, i.e. 0.11% smaller than that of the junCBS+CV-F12 model, if one considers CP energies. An improvement of 0.4% is also

Table 3.9: Relative and absolute errors[a,b] of different CBS+CV-F12 composite schemes.

Model		Relative error	Absolute error
mayCBS+CV-F12	CP	0.87%	0.08
	NCP	1.18%	0.09
	half-CP	0.92%	0.08
junCBS+CV-F12	CP	0.79%	0.07
	NCP	0.92%	0.09
	half-CP	0.75%	0.07
CBS+CV-F12[c]	CP	1.36%	0.10
	NCP	1.46%	0.16
	half-CP	1.00%	0.08

[a]Errors evaluated with respect to "ref" reference energies. [b]Absolute errors in kJ·mol^{-1}. [c]The extrapolation to the CBS limit using n=T,Q of the cc-pVnZ-F12 family.

observed for the ChS-F12 model with respect to the CBS+CV-F12 counterpart. From an overall comparison of the ChS-F12 schemes with the CBS+CV-F12 counterparts, it can be seen that the former show larger differences between CP and NCP results than the latter. This further confirms the previous discussion, i.e. the CBS+CV-F12 energies recover very well the CBS limit and do not show major improvements if the CP correction is included, while the opposite trend is observed for the ChS-F12 schemes. Indeed, the inclusion of the CP correction in the junChS-F12 model leads to the lowest MAE available in the statistics. Finally, it can be noticed that the MAEs of the junCBS+CV-F12 and junChS-F12 approaches are quite similar, with a difference in absolute terms of only 0.02 kJ·mol^{-1}. Therefore, the junChS-F12 model represents a very promising alternative to schemes entirely based on the CCSD(T)-F12 method.

In conclusion, it is confirmed that, even for explicitly correlated composite approaches, the jun-cc-pVnZ family of basis sets is the best choice for the description of NCIs and that the junCBS+CV-F12 and junChS-F12 approaches are the methodologies of choice together with junChS. The junCBS+CV-F12 model should be preferred for small nc-complexes involving only atoms of the first and second rows, while the junChS(-F12) alternatives represent the best choice for the medium-sized adducts also including heavier atoms, such those belonging to the third row of the periodic table.

Tha A14 data set and its extension

To analyse in details the different contributions of the junCBS+CV-F12 and junChS-F12 approaches, as well as to provide an accurate data set of interaction energies, tables 3.11 and 3.12 collect the values of the different terms present in each scheme, for the A14 data set. In both tables, the CP corrected energies are reported, while the NCP counterparts are reported in the Supporting Information of ref. [26]. In these tables, CC

Table 3.10: Relative and absolute errors[a,b] for ChS-F12 variants.

Model		Relative error	Absolute error
mayChS-F12	CP	1.23%	0.10
	NCP	2.81%	0.19
	half-CP	1.93%	0.14
junChS-F12	CP	0.68%	0.05
	NCP	1.10%	0.08
	half-CP	0.88%	0.06
ChS-F12[c]	CP	0.93%	0.09
	NCP	0.86%	0.09
	half-CP	0.82%	0.09

[a]Errors evaluated with respect to "ref" reference energies. [b]Absolute errors in kJ·mol^{-1}. [c]Employing cc-pVTZ-F12 and cc-pVQZ-F12 basis sets.

Table 3.11: junCBS+CV-F12 CP-energies (kJ·mol^{-1}): the various contributions for the A14 complexes.

	"ref"	CC-CBS/jun(T,Q)Z	CC-CV/wCTZ	Total	Rel. error (%)	Abs. error
$H_2O\cdots H_2O$	-21.0832	-21.0524	-0.1200	-21.1724	0.42	-0.09
$NH_3\cdots NH_3$	-13.2131	-13.1601	-0.0605	-13.2206	0.06	-0.01
$HF\cdots HF$	-19.2213	-19.3738	-0.0818	-19.4556	1.22	-0.23
$CH_2O\cdots CH_2O$	-18.9284	-18.9566	0.0382	-18.9184	-0.05	0.01
$HCN\cdots HCN$	-19.9828	-19.8441	-0.0811	-19.9253	-0.29	0.06
$C_2H_4\cdots C_2H_4$	-4.6024	-4.5555	-0.0068	-4.5623	-0.87	0.04
$CH_4\cdots CH_4$	-2.2301	-2.2058	0.0049	-2.2009	-1.31	0.03
$H_2O\cdots NH_3$	-27.3759	-27.4233	-0.1633	-27.5865	0.77	-0.21
$H_2O\cdots C_2H_4$	-10.7696	-10.6645	-0.0532	-10.7177	-0.48	0.05
$C_2H_4\cdots CH_2O$	-6.7948	-6.7604	-0.0010	-6.7704	-0.36	0.02
$NH_3\cdots C_2H_4$	-5.7865	-5.7393	-0.0243	-5.7636	-0.40	0.02
$HF\cdots CH_4$	-6.9162	-7.0176	-0.0749	-7.0925	2.55	-0.18
$H_2O\cdots CH_4$	-2.8242	-2.7537	-0.0206	-2.7743	-1.77	0.05
$NH_3\cdots CH_4$	-3.2175	-3.2065	-0.0274	-3.2338	0.51	-0.02
MAE					0.79	0.07

and MP2 indicate the explicitly correlated versions of these methods., i.e. CCSD(T)-F12 and MP2-F12, respectively.

From tables 3.11 and 3.12, it is evident that, in absolute terms, the CCSD(T)-F12/jun-cc-pVTZ contributions underestimate the corresponding "ref" values by a quantity ranging from 0.1 to 0.4 kJ·mol^{-1}. If augmented by the MP2-F12/CBS corrections, then they are very close to the corresponding junCBS+CV-F12 values, thus further confirming the effectiveness of the MP2-F12 extrapolation. The CV contribution appears to be similar, regardless of the method (either CCSD(T)-F12 or MP2-F12). The CV term is — as expected — negative, and is again non negligible, indeed being larger than the absolute error associated to the model considered. The energetics of tables 3.11 and 3.12 represent a reference benchmark to test other F12 methodologies and composite

Table 3.12: junChS-F12 CP-energies (kJ·mol^{-1}): the various contributions for the A14 complexes.

Complex	"ref"	CC/junTZ	ΔMP2$^{\infty}$/jun(T,Q)Z	MP2-CV/wCTZ	Total	Rel. error (%)	Abs. error
$H_2O\cdots H_2O$	-21.0832	-20.8822	0.0344	-0.1512	-20.9990	-0.40	0.08
$NH_3\cdots NH_3$	-13.2131	-12.9057	-0.2132	-0.0807	-13.1997	-0.10	0.01
HF···HF	-19.2213	-19.1430	0.0014	-0.1078	-19.2494	0.15	-0.03
$CH_2O\cdots CH_2O$	-18.9284	-18.5310	-0.3690	-0.0552	-18.9553	0.14	-0.03
HCN···HCN	-19.9828	-19.7537	-0.0085	-0.0787	-19.8410	-0.71	0.14
$C_2H_4\cdots C_2H_4$	-4.6024	-4.3114	-0.3000	-0.0493	-4.6607	1.27	-0.06
$CH_4\cdots CH_4$	-2.2301	-1.9832	-0.2083	-0.0055	-2.1970	-1.48	0.03
$H_2O\cdots NH_3$	-27.3759	-27.1443	-0.1294	-0.2003	-27.4740	0.36	-0.10
$H_2O\cdots C_2H_4$	-10.7696	-10.4381	-0.2219	-0.1092	-10.7692	-0.004	0.0004
$C_2H_4\cdots CH_2O$	-6.7948	-6.4974	-0.2677	-0.0620	-6.8271	0.48	-0.03
$NH_3\cdots C_2H_4$	-5.7865	-5.5505	-0.2083	-0.0603	-5.8190	0.56	-0.03
$HF\cdots CH_4$	-6.9162	-6.7403	-0.1943	-0.1072	-7.0418	1.82	-0.13
$H_2O\cdots CH_4$	-2.8242	-2.6431	-0.0979	-0.0335	-2.7746	-1.76	0.05
$NH_3\cdots CH_4$	-3.2175	-3.0678	-0.1160	-0.0428	-3.2266	0.28	-0.01
MAE						0.68	0.05

schemes. In fact, the MAEs with respct to the reference values are as small as 0.79% and 0.68% for the junCBS+CV-F12 and junChS-F12, respectively. In absolute terms, this means an error of 0.07 kJ·mol^{-1} for the former and 0.05 kJ·mol^{-1} for the latter. Lastly, it should be mentioned that ref. [26] investigates the possibility of removing the d and f diffuse functions from the jun-cc-pVTZ and jun-cc-pVQZ basis sets, respectively. Even with this basis set reduction, the composite scheme performs as well as junChS-F12 (same statistical parameters), and the approach can be used to further save computational resources.

Having proposed the junChS and junChS-F12 approaches as standard models for medium-sized nc-complex, a larger data set should be considered to extend the representativeness of the benchmark study. Furthermore, this also offers the occasion to compare the junChS and junChS-F12 schemes. Two data sets including small nc-complexes with third-row atoms, but also larger systems of biological interest, have been considered in addition to the A14 data set. These are the B9 and C6 data sets, composed as follows:

- B9 data set: $FH_2P\cdots NH_3$, $FH_2P\cdots H_2S$, $H_2O\cdots H_2S$, $H_2O\cdots PH_3$, $OCS\cdots H_2O$, $OCS\cdots CH_4$, $CH_3NH_2\cdots HCl$, $(CH_3)_2S\cdots SO_2$, $SO_2\cdots H_2S$.
- C6 data set: c-$C_5H_8\cdots H_2O$ (Z isomer), $CH_3NH_2\cdots C_5H_5N$, $CH_3OH\cdots C_5H_5N$, $H_2O\cdots C_3H_7NO$, $C_5H_5N\cdots C_5H_5N$, $CH_3NH_2\cdots CH_3NH_2$.

These complexes have been chosen based on previous work [25, 42, 155], but also on other literature collections, such as the S66 data set [156]. The B9 and C6 sets are shown in fig. 3.4 and the corresponding junChS and junChS-F12 energies (CP, NCP and half-CP) are compared with the available literature data in table 3.13. For atoms belonging to the third-row of the periodic table, the basis sets employed include an additional tight d function, e.g. jun-cc-pV(n+d)Z instead of jun-cc-pVnZ.

The $FH_2P\cdots H_2S$, $FH_2P\cdots NH_3$ and $DMS\cdots SO_2$ complexes were already considered in ref. [25] and the small discrepancies (<0.1 kJ·mol^{-1}) observed between the present data and those there reported are probably due to the use of a different reference geometry (B2PLYP-D3/maug-cc-pVTZ in ref. [25]).

For complexes belonging to the S66 data set, i.e. $CH_3NH_2\cdots CH_3NH_2$, $CH_3NH_2\cdots C_5H_5N$, $C_5H_5N\cdots C_5H_5N$ and, $CH_3OH\cdots C_5H_5N$, larger differences with respect to the literature data are observed. These discrepancies can be explained taking into account the level of theory employed in the S66 data set. Indeed, in ref. [156], the energies were computed employing the same composite scheme used for the determination of the reference geometries of the A14 data set. Briefly, the energy was obtained as the sum of: (i) the HF/aug-cc-pVQZ energy, (ii) the CBS contribution evaluated at the MP2 level using the aug-cc-pVTZ and aug-cc-pVQZ basis sets, and (iii) the effect of triple excitations incorporated via the CCSD(T)-MP2 energy difference computed using the aug-cc-pVDZ basis set. This level of theory is intrinsically less accurate than the junChS and junChS-F12 approaches, since the CV correction is not included and a very small basis set (double-ζ) is used to incorporate the effects of triples excitations.

The cyclopentene-water complex was previously investigated at the ChS level [155] without inclusion of diffuse functions. Therefore, the differences in table 3.13 are mainly due to their balanced inclusion.

Comparing the junChS and junChS-F12 schemes, the inspection of table 3.13 points out that the two approaches provide very similar results, with a maximum difference of 0.6 kJ·mol^{-1} for CP-corrected interaction energies and 0.8 kJ·mol^{-1} for the NCP counterparts. The CP-corrected junChS interaction energies are usually larger than the junChS-F12 counterparts by about 0.1-0.2 kJ·mol^{-1}. However, the average CP correction of $\sim$0.3 kJ·mol^{-1} for the junChS model can not be overlooked because it is of the

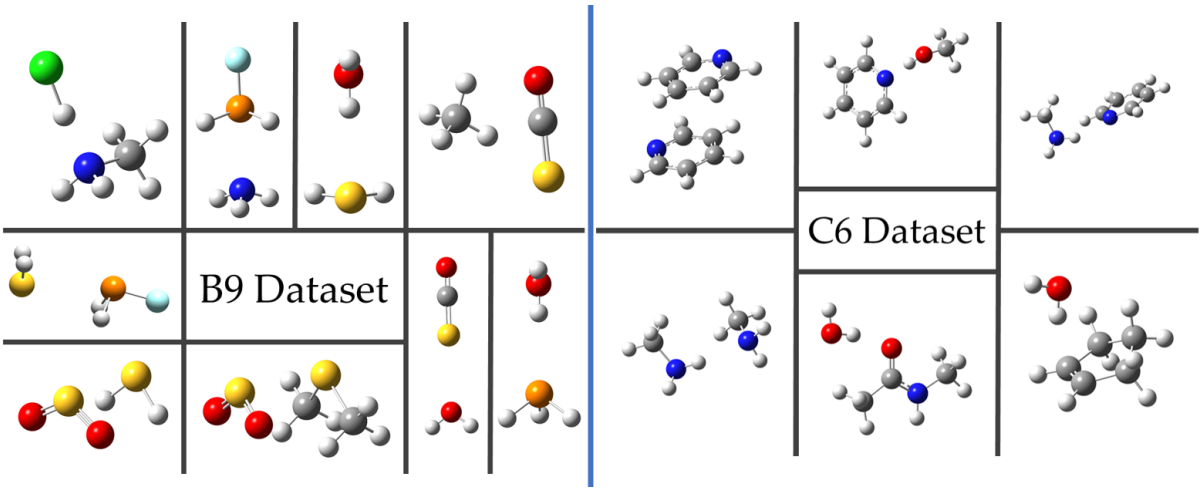


Figure 3.4: Panel (a), B9 data set: small systems containing atoms of the first three rows of the periodic table. Panel (b), C6 data set: larger systems containing atoms of the first two rows of the periodic table.

Table 3.13: junChS and junChS-F12 interaction energies[a] (revDSD reference geometry) for the B9 and C6 data sets.

	junChS			junChS-F12			
	CP	NCP	half-CP	CP	NCP	half-CP	Literature
B9 data set							
$FH_2P\cdots H_2S$	-14.92	-15.42	-15.17	-14.80	-14.89	-14.85	-14.94[b]
$FH_2P\cdots NH_3$	-29.46	-30.03	-29.74	-29.08	-29.24	-29.16	-29.54[b]
$H_2O\cdots H_2S$	-12.18	-12.41	-12.30	-12.23	-12.25	-12.24	
$H_2O\cdots PH_3$	-10.77	-10.95	-10.86	-10.76	-10.77	-10.76	
$CH_3NH_2\cdots HCl$	-52.27	-52.81	-52.54	-52.32	-52.41	-52.37	
$OCS\cdots CH_4$	-4.36	-4.45	-4.41	-4.25	-4.29	-4.27	
$OCS\cdots H_2O$	-7.96	-8.09	-8.02	-7.87	-7.89	-7.88	
$SO_2\cdots H_2S$	-12.22	-12.45	-12.34	-12.11	-12.20	-12.15	
$DMS\cdots SO_2$	-33.53	-34.40	-33.97	-32.97	-33.36	-33.17	-34.04[b]
C6 data set							
$Cyclopentene\cdots H_2O$	-16.59	-16.71	-16.65	-16.37	-16.44	-16.40	-12.8[c]
$H_2O\cdots Peptide$	-35.16	-35.54	-35.35	-34.91	-35.05	-34.98	-33.89[d]; -32.31[d]
$CH_3NH_2\cdots CH_3NH_2$	-17.82	-17.94	-17.88	-17.64	-17.67	-17.66	-17.41[d]; -15.23[e]
$CH_3NH_2\cdots Pyridine$	-20.17	-20.27	-20.22	-19.94	-19.99	-19.96	-16.61[d]; -13.19[e]
$CH_3OH\cdots Pyridine$	-28.18	-28.65	-28.42	-28.08	-28.21	-28.15	-31.00[d]; -28.57[e]
$Pyridine\cdots Pyridine$	-16.32	-16.58	-16.45	-16.12	-16.26	-16.19	-16.32[d]; -9.97[e]

[a]Values in kJ·mol^{-1}. [b]Ref. [25]: at the junChS level. [c]Vibrational ground-state dissociation energy taken from Ref. [155]: ChS CP-corrected electronic energy augmented by harmonic zero-point energy at the B2PLYP-D3BJ/maug-cc-pVTZ-*d*H level. [d]Ref. [156]: CCSD(T)/CBS computed as the HF/aug-cc-pVQZ energy corrected for: (*i*) the CBS contribution evaluated at the MP2 level using the aug-cc-pVTZ and aug-cc-pVQZ basis sets and (*ii*) the effect of triple excitations is incorporated via the CCSD(T)-MP2 energy difference, computed using the aug-cc-pVDZ basis set. [e]Ref. 157: CCSD/CBS energy computed as the sum of the HF/aug-cc-pVQZ and correlation PNO-CCSD-F12/aug-cc-pVTZ energies.

same order of, if not greater than, the accuracy expected for this approach. Instead, the CP correction can be safely neglected when employing the junChS-F12 model, with a remarkable saving of computer time, since the differences between CP and NCP energies is nearly negligible. A complex particularly affected by the CP correction is DMS-SO_2, for which the CP-NCP difference for the junChS interaction energy is 0.9 kJ·mol^{-1} and it halves when the junChS-F12 model is employed.

The A14, B9 and C6 data sets have been used to demonstrate the effectiveness of the junChS and junChS-F12 approaches, which should ideally become routine methodologies for the description of NCIs. In fact, these models can describe uniformly molecular complexes involving light molecules, as well as those containing third-row atoms and relative large systems. In conclusion, the proposed data sets can be safely used for future benchmark studies on NCIs.

3.2 Geometries

The accurate characterisation of molecular complexes is often limited to their energetics and benchmark studies on *ab initio* structural information are very limited in

the literature. However, the importance of accurate geometries was already mentioned and it is two-fold: (i) the energetics depends on the structure and reliable results are obtained only when reliable reference geometries are employed [158,159]; (ii) the spectroscopic characterisation of molecular complexes, carried out via rotational spectroscopy [1,35], depends on parameters that are strongly related to the molecular structure [35,143]. As previously mentioned, the ChS scheme has already been used to describe isolated molecules and nc-complexes [27,136,137], but an in-depth investigation is needed, also extending the systematic analysis to the explicitly-correlated counterpart. Indeed, since junChS and junChS-F12 have been suggested as accurate models for the description of interaction energies, it is somewhat natural to expect good results also for structural determinations. The composite approaches are applied directly to the geometrical parameters, instead of the minimisation of a composite energy gradient, therefore resorting to the "geometry" approach. The composite schemes employed are identical to those used for the energetic and thus no clarification is required in this respect. Ten different methodologies have been investigated and the full accounts of results is reported in ref. [26]. In the present thesis, only those retrieved from junChS, junCBS+CV, and their explicitly correlated counterparts (junChS-F12 and junCBS+CV-F12) will be considered together with those obtained from CCSD(T)-F12 computations in conjunction with the jun-cc-pVTZ basis set. All these approaches provide quite similar results for the intramolecular geometrical parameters, which means that differences of few mÅ for distances and changes smaller than 1 degree for angles are observed. For example, in the water dimer, the two intramolecular H−O−H angles are predicted to be 104.98° and 104.88° using the junCBS+CV-F12 model. The same angles are 104.90° and 104.84°, respectively, at the junChS level, and 104.72° and 104.92°, respectively, using the junChS-F12 approach. The same behaviour is also observed for the intermolecular angles. Again considering the H_2O dimer, the intramolecular O···O−H angle spans from 171.15° to 172.68°, with a maximum difference between all the considered schemes of 1.5°. Since intermolecular angles have only a minor influence on the electronic energy because the PES is rather flat along these coordinates, it is somewhat difficult to predict the correct value for nc-complexes and small changes of 1-2° lie within the uncertainty of the structural determination. Similar conclusions can be made for all the complexes of the A14 data set, regardless of their nature. A full account of the intramolecular data for the complexes of the A14 data set is reported ref. [26].

Based on the discussion above, the focus is on the intermolecular distances, which tune the strength of the interaction. The latter parameters are reported in table 3.14, where the following levels of theory are reported:

1. CCSD(T)-F12 optimisation in conjunction with the jun-cc-pVTZ basis set (CC-F12)

2. CCSD(T)-F12 computation with the jun-cc-pVTZ basis set augmented for the CV contribution computed at the MP2-F12 level employing the cc-pwCVTZ basis set (CC-F12+CV).

3. ChS

4. junChS

5. junCBS+CV-F12

6. junChS-F12

7. revDSD/jun-cc-pVTZ

From the inspection of the data collected in table 3.14, it is noted that the junChS and junChS-F12 schemes give comparable distances, which are either longer or shorter than those obtained with the junCBS+CV-F12 approach. However, it is observed that the convergence to the CBS limit for the CCSD(T)-F12 method does not follow a smooth behaviour. For the $NH_3\cdots H_2O$, $H_2O\cdots C_2H_4$, HCN$\cdots$HCN, and $CH_4\cdots$HF complexes, the intermolecular bond distance increases when going from the jun-cc-pVTZ to the jun-cc-pVQZ basis set, a behaviour which is opposite to that systematically obtained for the conventional CCSD(T) and MP2 methods as well as for MP2-F12. Therefore, the conclusion is that CCSD(T)-F12 geometries do not benefit from the extrapolation to the CBS limit and that this can lead instead to unreliable corrections. This further supports the idea that F12 methods converge rapidly to the CBS limit and do not even benefit from the MP2-F12 corrective term included ChS-F12 schemes. The MP2-F12 CBS contribution often results in corrections smaller than 0.1 mÅ and the differences between the CC-F12 and junCBS+CV-F12 geometries are negligible or can easily be attributed to the CV term. This means that the simplest and "safest" route to obtain accurate structural parameters when relying on CCSD(T)-F12 calculations is to add CV corrections, evaluated at either the MP2-F12 or CCSD(T)-F12 level, to CC-F12 geometries, thus having the CC-F12+CV model.

With respect to the "CBS-georef" structures, the junChS and junChS-F12 models show the best performances with average deviation of -0.01 Å. Half of such a discrepancy can be attributed to the CV term, which is not included in the "CBS-georef" reference structures. Indeed, the CV correction accounts for -0.005 Å regardless of the method employed, i.e. MP2-F12 or CCSD(T)-F12. The ChS scheme, which does not account for diffuse functions, shows distances slightly shorter than those obtained using the junChS and junChS-F12 approaches. However, for troublesome cases like the $CH_4\cdots NH_3$ and $CH_4\cdots H_2O$ complexes, differences above 50 mÅ are observed, thus making the ChS less reliable for weakly-bonded complexes. The last comment concerns the revDSD geometries that represent the "cheapest" method among those reported in table 3.14. This methodology provides intermolecular parameters that are in

Table 3.14: Intermolecular distances (Å) for the A14 complexes obtained at different levels of theory.

Complex	Parameter[a]	CC-F12	CC-F12+CV	ChS	junChS	junCBS+CV-F12	junChS-F12	revDSD	CBS-georef
HCN···HCN	r(H4-C5)	1.0729	1.0718	1.0718	1.0714	1.0717	1.0716	1.0740	1.0743
HF···HF	r(F2-H3)	1.8280	1.8257	1.8244	1.8176	1.8212	1.8225	1.8261	1.8327
H_2O···H_2O	r(H3-O4)	1.9541	1.9503	1.9477	1.9507	1.9513	1.9487	1.9557	1.9618
NH_3···H_2O	r(H3-N4)	1.9733	1.9696	1.9731	1.9789	1.9738	1.9753	1.9707	1.9850
CH_2O···CH_2O	r(O5-H3)	2.3949	2.3870	2.3776	2.3707	2.3839	2.3781	2.4143	2.3827
CH_4···NH_3	r(H2-N3)	2.8003	2.7948	2.8220	2.7713	2.7818	2.7751	2.7964	2.7933
CH_4···HF	r(H2-C3)	2.3022	2.2980	2.3191	2.3103	2.3077	2.3133	2.3079	2.3195
NH_3···NH_3	r(H2-N4)	2.1760	2.1703	2.1627	2.1598	2.1612	2.1651	2.1802	2.1717
CH_4···CH_4	r(C2-C3)	3.6659	3.6605	3.6244	3.6231	3.6385	3.6373	3.6642	3.6380
C_2H_4···C_2H_4	r(X2-X3)	3.8347	3.8232	3.7915	3.7816	3.8142	3.7994	3.8216	3.8116
C_2H_4···CH_2O	r(H4-C5)	2.8650	2.8572	2.8344	2.8472	2.8488	2.8748	2.8706	2.8607
C_2H_4···H_2O	r(H3-X4)	2.4168	2.4105	2.4199	2.4252	2.4242	2.4283	2.4006	2.4224
C_2H_4···NH_3	r(C2-H3)	2.6778	2.6698	2.6953	2.6616	2.6661	2.6694	2.6706	2.6835
H_2O···CH_4	r(H3-O4)	2.6270	2.6232	2.8410	2.6434	2.6238	2.6229	2.6445	2.6279

[a]For atom labelling see Appendix B.

agreement with the high-level counterparts, further supporting the use of this level of theory for the computation of reference structures to be employed for accurate interaction energies.

Chapter 4

Applications

This present chapter considers six applications of the junChS and junChS-F12 schemes. The first section describes the application of such models to the characterisation of nc-complexes, while the second part explores how these schemes can be used to provide accurate energetics for formation reactions of astrochemically-relevant species.

4.1 Non-covalent complexes

The importance of nc-complexes as probes for the investigation of NCIs in a matrix- and solvent-free environment, i.e. the gas phase, was already discussed in the Introduction. In the following, three key case studies are considered to demonstrate the strong interplay of theory and experiment in rotational spectroscopy and also to show the relevance of junChS(-F12) in such integrated approach. Each example here reported will address the relevance of the complex under consideration and will illustrate the results achieved.

4.1.1 The dimethyl sulfide-sulfur dioxide complex

Since NCIs are of different nature and operate in a cooperative manner, the unequivocal definition of a type of non-covalent bond is not-trivial. For example, only in 2019, the International Union of Pure and Applied Chemistry (IUPAC) has proposed a definition for the "chalcogen bond" [160], thus introducing guidelines for its identification and also recognising its importance as key NCI together with the hydrogen- and halogen-bond. According to IUPAC, a chalcogen bond is *a net attractive interaction between an electrophilic region associated with a chalcogen atom in a molecular entity and a nucleophilic region in another, or the same, molecular entity*. Therefore, a chalcogen bond occurs whenever an atom belonging to the sixteenth group of the periodic table, i.e. oxygen, sulfur, selenium and tellurium, can accept a partial negative charge in its surrounding. Aiming at the analysis of a chalcogen bond in the gas phase, complexes involving a sulfur or an oxygen atom are particularly relevant, as both elements are

Table 4.1: Rotational constants[a] of seven isotopologues of the DMS···SO_2 complex together with the computed vibrational corrections at the B3-D3/SNSD level.

Parameter	SO_2···$S(CH_3)_2$	$^{34}SO_2$···$S(CH_3)_2$	SO_2···$^{34}S(CH_3)_2$	$^{33}SO_2$···$S(CH_3)_2$	SO_2···$^{33}S(CH_3)_2$	$S^{18}O^{16}O$···$S(CH_3)_2$	SO_2···$S(^{13}CH_3)(^{12}CH_3)$
A^b	3490.1920(16)	3484.0255(41)	3473.9912(39)	3487.028(10)	3481.9380(90)	3414.9948(81)	3438.8559(56)
B^b	1497.76387(44)	1477.70534(28)	1475.64069(27)	1487.57426(31)	1486.51807(47)	1475.98433(27)	1484.77801(20)
C^b	1216.39695(35)	1203.85113(20)	1203.70117(19)	1210.03633(20)	1209.96356(26)	1194.47542(27)	1203.33928(20)
ΔA_0	16.521	15.482	16.329	16.415	16.416	15.168	16.293
ΔB_0	38.182	37.604	37.376	37.820	37.774	39.927	37.485
ΔC_0	27.910	27.316	27.536	27.719	27.720	26.571	27.364

[a]All values in MHz. [b]In parentheses 1σ error in units of the last digit.

quite abundant in nature. However, the former is somewhat more intriguing than the latter as it can act either as donor or acceptor of a chalcogen bond [161,162]. Aiming at the characterisation of a pure chalcogen bond between two sulfur atoms of two different molecular entities, several sulfur-bearing compounds were considered, focusing on those species able to reduce the possibility of observing other types of NCIs. An appropriate fragment is represented by dimethyl sulfide (DMS, CH_3-S-CH_3), which does not have any S−H bond, thus disfavouring the formation of H-bonds. Additionally, the two methyl groups of DMS partially donate negative charge to the sulfur atom by induction. Thus, the S atom of DMS is electron-rich and can act as a nucleophilic centre. The S-containing partner should contain an electrophilic sulfur atom, with sulfur dioxide (SO_2) being a good candidate. Indeed, in the latter molecule, the oxygen atoms attract the electronic charge and determine on the sulfur atom a partially positive charge. Therefore, the DMS···SO_2 complex has been investigated as potential candidate for the characterisation of a pure chalcogen bond in the gas phase [42].

Experimentally, the complex was formed in a supersonic jet of Ar (rotational T = 4 K), containing 1.4-2% of DMS and 1.0-1.3% of SO_2 at a stagnation pressure of 1 bar. The first measurements were carried out in the 7-16 GHz frequency range with the In-phase/quadrature-phase-Modulation Passage- Acquired-Coherence Technique (IMPACT) spectrometer [163], and were further extended with additional data obtained with the COaxially Aligned Beam Resonator Arrangement) (COBRA) Fourier Transform Microwave spectrometer [164]. The rotational transitions of seven different isotopologues of the DMS-SO_2 complex were identified in natural abundance. These are: the parent species (i.e. that containing the most abundant isotopes) and the singly substituted ^{34}S, ^{33}S, ^{13}C, and ^{18}O molecular complexes. For each species, the fit of the recorded transitions using the appropriate form of the Watson's Hamiltonian, led to the determination of the experimental rotational constants and other spectroscopic parameters. A full list of the rotational transitions measured and of the rotational parameters obtained is reported in ref. [42], while the rotational constants of all the species investigated are reported in table 4.1.

All experimental (ground-state) rotational constants were found to be consistent with the same molecular structure, which was elucidated owing to quantum chemistry. In fact, the exploration of the PES at the B2PLYP-D3BJ/maug-cc-pVTZ-dH level

Table 4.2: Comparison between computed and experimental rotational parameters.[a]

Parameter	Exp. Value	B2-D3/maugTZ-dH[b]	junChS[b]	junChS-F12[b]	junChS+$\Delta B_0^{\alpha c,d}$	junChS-F12+$\Delta B_0^{\alpha c,d}$
A	3490.1920(16)	3448.96	3524.75	3504.50	3508.23 (0.52%)	3487.98 (-0.06%)
B	1497.76387(44)	1559.92	1547.44	1540.58	1509.28 (0.77%)	1502.42 (0.31%)
C	1216.39695(35)	1251.91	1250.91	1246.17	1223.00 (0.54%)	1218.26 (0.15%)

[a]All Values in MHz. [b]Equilibrium values. [c]In parentheses the error with respect to the corresponding experimental value. [d]ΔB_0^{α} from table 4.1.

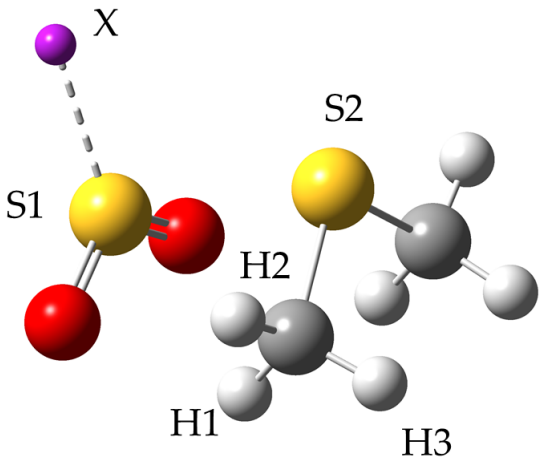


Figure 4.1: The DMS$-SO_2$ complex with atom labelling. X indicates a dummy atom located on the symmetry plane.

(shortly, B2-D3/maugTZ-dH)[1] pointed out the presence of only one stable conformer for the DMS$-SO_2$ complex, which is represented in fig. 4.1. The corresponding equilibrium rotational constants for the main isotopic species are reported in table 4.2, where they are compared with the experimental ground state ones for the main isotopologue. The agreement between the two sets of values is remarkable but it is not satisfactory enough to claim that the observed conformer is the one computationally predicted. Therefore, further computational effort has been put to obtain a better estimate of the rotational parameters. In particular, two actions have been taken. First, recalling that the equilibrium values recover 97-99% of the ground-state rotational constants (see Chapter 2) and that the former only depend on the equilibrium structure, this latter has been improved employing the junChS approach. Second, according to eqs. (2.33) and (2.35), one needs to incorporate the vibrational corrections to move from the equilibrium to the ground state. Based on eq. (2.34), these corrections require an anharmonic force-field calculation, which has been performed at the B3LYP-D3BJ/SNSD level (B3-D3/SNSD), and are collected in table 4.1 for all isotopologues considered. The SNSD basis set is of double-zeta quality and it is derived from the N07D basis set [165, 166] by inclusion of s diffuse functions on all atoms.

From the data reported in table 4.2 it can be readily seen how the junChS and

[1]The -dH label denotes the maug-cc-pVTZ basis set without d-type functions on the H atom

junChS-F12 provide equilibrium rotational constants that are in better agreement with experiment than those provided by the B2-D3/maugTZ-dH level of theory. In particular, for the junChS scheme the largest error is observed for the B rotational constant, this being 3.32% when the equilibrium value is considered and 0.77% when the vibrational correction is introduced. Smaller errors characterise the junChS-F12 approach, for which the largest deviation is still noted for B and it is 2.86% at the equilibrium and 0.31% when the ΔB_0 contribution is added (see table 4.2). Furthermore, for the junChS-F12 approach, a discrepancy as low as 0.06% is evident for the A rotational constant. According to these results, it can be confirmed that the observed molecular complex is the one shown in fig. 4.1. A further confirmation is also obtained by the comparison of the computed quartic centrifugal distortion and nuclear quadrupole-coupling constants with the experimental counterparts. The nuclear quadrupole coupling occurs for the two ^{33}S-bearing isotopologues. A full account of these data is reported in the ref. [42].

To understand the nature of the NCIs stabilising the DMS$-$SO$_2$ complex, a Natural Bond Orbital (NBO) analysis was carried out. The latter is able to point out the electron transfer between the two fragments of the complex by exploiting the chemically useful natural orbitals. The analysis was carried out at the B3LYP/maug-cc-pVTZ-dH level and shows that, at the second order of perturbation, the stability of the complex is dominated by a charge transfer occurring between a donating lone pair of the S atom belonging to the DMS fragment and an accepting S$-$O antibonding orbital of SO$_2$ (see table 4.3). This charge flow is confirmed also by the comparison of the NBO charges of the complex with those of the isolated monomers. Indeed, the oxygen atoms of the SO$_2$ species increase their negative charge when moving from the isolated monomer to the complex [42]. This charge transfer was estimated via the Natural Orbital for Chemical Valence/Charge Displacement (NOCV/CD) scheme, which indicates a global charge flow of 0.12 e along the axis connecting the two sulfur atoms. The electrostatic interaction occurring in the DMS$\cdots$SO$_2$ complex is apparent and confirms that the main interaction established between the two fragments is a chalcogen bond.

The dissociation energy (D_0) of this nc-complex was originally obtained in ref. [42] using the ChS approach on top of the ChS structure and it resulted in a value of 23.5 kJ·mol^{-1}, with the corresponding equilibrium value (D_e) being of 28.5 kJ·mol^{-1}. D_e was later refined employing the junChS approach on top of a NCP-revDSD geometry, leading to an energy of 31.1 kJ·mol^{-1}. For the latter, a contribution of 34.0 kJ·mol^{-1} is due to the interaction energy, with the remaining difference being due to the deformation term [25]. The interaction energy was also estimated at the junChS-F12 level, again using the NCP-revDSD geometry, the resulting value of 33.2 kJ·mol^{-1} (half-CP energy) being in good agreement with the previous one [26].

Back to the structural characterisation of the complex, since several sets of experimental rotational constants are available for the DMS$-$SO$_2$ complex, a SExp equilib-

Table 4.3: Second-order stabilisation energy contributions (kJ·mol^{-1}) for the DMS···SO_2 complex.[a] Only contributions larger than 0.84 kJ·mol^{-1} are reported. LP indicates a valence lone pair, while BD* and RY* denote 2-center antibond and Rydberg antibond orbitals, respectively.

Donor NBO	Accpetor NBO	E(2)	Donor NBO	Accpetor NBO	E(2)
From DMS to SO_2			From SO_2 to DMS		
LP(1) S2	BD*(2) S1-O1	2.72	BD(2) S1-O1	BD*(1) C1-H1b	0.92
LP(2) S2	RY*(2) S1	1.09	BD(2) S1-O1	BD*(1) C1-H3b	1.17
LP(2) S2	RY*(4) S1	3.89	LP(1) S1	RY*4 S2	1.63
LP(2) S2	BD*(1) S1-O1	4.56	LP(1) O1	BD*(1) C1-H1b	0.87
LP(2) S2	BD*(2) S1-O1	51.14	LP(1) O2	BD*(1) C2-H1a	0.87
LP(2) S2	BD*(1) S1-O2	2.09	LP(3) O2	BD*(1) C2-H1a	1.50
			LP(3) O2	BD*(1) C2-H3a	1.34
			BD*(2) S1-O1	RY*(4) S2	1.38

[a] Atom labelling according to fig. 4.1.

rium structure can be derived. Within the SExp approach, the equilibrium geometry is obtained via the least-square fit of the SExp equilibrium rotational constants (or the corresponding moment of inertia). The latter are obtained from ground state rotational constants by subtracting the computed vibrational corrections (see eqs. (2.33) and (2.35)). The isotopologues considered are those reported in table 4.1, whose vibrational corrections have been computed at the B2-D3/maugTZ-dH level of theory employing VPT2, the corresponding values being collected in table 4.1. Considering the number of isotopic substitutions available, a full SExp structure is not feasible and two different procedures have been adopted. In the first one, all the rotational constants are used to compute a partial SExp equilibrium structure, with the parameters involving H atoms being kept fixed at their computed value. In the second case, only the intermolecular parameters are fitted in the SExp approach, with the intramolecular ones fixed at improved values. To accomplish thus, the so-called "template molecule (TM) approch" (TMA) is used. In this case, each fixed intramolecular parameter ($r_{\mathrm{e}}^{\mathrm{fixed}}$) is obtained using the following formula:

$$r_{\mathrm{e}}^{\mathrm{fixed}} = r_{\mathrm{e}}^{\mathrm{comp,complex}} + \Delta r_{\mathrm{e}}(\mathrm{TM}) \tag{4.1}$$

with

$$\Delta r_{\mathrm{e}}(\mathrm{TM}) = r_{\mathrm{e}}^{\mathrm{acc,monomer}} - r_{\mathrm{e}}^{\mathrm{comp,monomer}} \quad . \tag{4.2}$$

In the previous equations, $r_{\mathrm{e}}^{\mathrm{comp,complex}}$ and $r_{\mathrm{e}}^{\mathrm{comp,monomer}}$ represent the same intramolecular parameter, the former in the complex and the latter in the monomer. These two terms refers to geometry optimisations performed at the same level of theory. The $\Delta r_e(TM)$ term represents the correction to be applied and is computed as the differ-

Table 4.4: Structural parameters[a,b] for the DMS···SO_2 complex and its molecular fragments (distances in Å and angles in degrees).

$(CH_3)_2S$	CP-B2-D3/junTZ	ChS	CCSD(T)/CBS+CV	r_{SExp}[c]
r(C-S)	1.8075	1.7998	1.7970	1.79863(13)
θ(CSC)	99.03	98.41	98.55	98.58000(81)
r(H2C)	11.0870	1.0865	1.0873	1.08857(38)
θ(H2CS)	107.28	107.44	107.44	107.4196(69)
r(H1C)	1.0882	1.0879	1.0885	1.08972(47)
θ(H1CS)	110.93	110.73	110.80	110.688(29)
ϕ(H1CSH2)	-118.85	-118.96	-118.95	-119.053(44)
ϕ(H_1CSC)	-61.15	-61.04	-61.05	-60.95
SO_2	CP-B2-D3/junTZ	ChS	CCSD(T)/CBS+CV	r_{SExp}[d]
r(SO)	1.4443	1.4331	1.4245	1.4307858(15)
θ(OSO)	119.28	119.11	120.45	119.329872(81)

		r_{SExp}				
		TMA				
$(CH_3)_2S\cdots SO_2$	CP-B2-D3/junTZ	CP-B2-D3/junTZ	ChS	CCSD(T)/CBS+CV	SExp	Partial[e]
r**(S1S2)**	2.9344	2.932(4)	2.946(1)	2.953(2)	2.948(2)	2.947(3)
r(S1O)	1.4535	1.4535	1.4423	1.4337	1.4401	1.446(6)
θ(OS1S2)	95.21	95.46(2)	95.11(7)	95.0(1)	95.07(7)	95.0(2)
ϕ**(OS1S2X)**	-121.10	-121.10	-121.18	-120.73	-121.08	-118.2(9)
r(CS2)	1.8040	1.8040	1.7963	1.7935	1.7951	1.790(5)
θ**(CS2S1)**	91.33	91.4(2)	91.58(7)	91.6(1)	91.59(7)	91.7(2)
ϕ**(CS2S1X)**	129.93	132.5(8)	130.7(3)	129.9(4)	130.5(3)	9.7(7)
r(H1C)	1.0885	1.0885	1.088	1.0887	1.0899	1.0883
θ(H1CS3)	110.68	110.68	110.48	110.55	110.44	110.67
ϕ(H1CS3C)	-65.53	-65.53	-65.41	-65.43	-65.32	26.06
r(H2C)	1.0865	1.0832	1.086	1.0867	1.0880	1.0864
θ(H2CS3)	107.14	107.14	106.94	107.30	107.28	107.14
ϕ(H2CS2H1)	-118.65	-118.65	-118.77	-118.75	-118.86	-92.56
r(H3C)	1.0871	1.0871	1.0868	1.0874	1.0886	1.0870
θ(H3CS2)	109.93	109.93	109.73	109.79	109.68	109.97
ϕ(H3CS2C)	57.03	57.03	56.92	56.93	56.83	148.63
σ[f]		1.52×10^{-2}	1.47×10^{-3}	3.41×10^{-3}	1.66×10^{-3}	0.02[g]

[a]Atom labelling according to fig. 4.1. [b]Intermolecular parameters in bold. [c]Taken from ref. [167]. [d]Taken from ref. [168]. [e]Taken from ref. [42] [f]Standard deviation of the fit using moments of inertia [g]Standard deviation of the fit using rotational constants

ence between $r_e^{comp,monomer}$ and $r_e^{acc,monomer}$, with the latter being the accurate intramolecular parameter of monomer either derived from high-level computations or from the SExp approach.

In the case of the DMS−SO_2 complex, the partial SExp equilibrium structure is reported in table 4.4 in the column denoted as "Partial" and it was obtained by keeping fixed at the B2-D3/maugTZ-*d*H level all the parameters involving the H atoms [42]. In the same table, the results of the TM approach are also given. In this case, the computed intramolecular parameters for the monomers and the complex have been obtained by a CP-B2PLYP-D3BJ/jun-cc-pVTZ (CP-B2-D3/junTZ) geometry optimisation and the Δ(TM) has been derived using accurate *ab initio* geometries. In particular, the CCSD(T)/CBS+CV and ChS approaches have been considered [19]. As already mentioned, the former is based on a gradient scheme and it extrapolates the HF and

Table 4.5: Structural parameters[a,b] for the DMS···SO_2 complex and its molecular fragments (distances in Å and angles in degrees).

$(CH_3)_2S$	junChS	junChS-F12	r_{SExp}[c]
r(C-S)	1.7984	1.7977	1.79863(13)
θ(CSC)	98.43	98.49	98.58000(81)
r(H2C)	1.0867	1.0870	1.08857(38)
θ(H2CS)	107.45	107.46	107.4196(69)
r(H1C)	1.0879	1.088	1.08972(47)
θ(HCS)	110.75	110.79	110.688(29)
ϕ(H1CSH2)	-118.98	-118.96	-119.053(44)
ϕ(H2CSC)	-61.02	-61.04	-60.95
SO_2	junChS	junChS-F12	r_{SExp}[d]
r(S-O)	1.4298	1.4288	1.4307858(15)
θ(OSO)	119.29	119.24	119.329872(81)

			r_{SExp}		
				TMA	
$(CH_3)_2S\cdots SO_2$	junChS	junChS-F12	fixed[e]	junChS	junChS-F12
***r*(S1S2)**	2.9570	2.9632	2.944(2)	2.945(2)	2.944(2)
r(S1O)	1.4266	1.4358	-	1.4276	1.4378
θ(OS1S2)*	94.18	94.21	95.5(1)	95.1(1)	95.4(1)
***φ*(OS1S2X)**	-120.76	-121.08	-	-120.11	-120.81
r(CS2)	1.7928	1.7958	-	1.7930	1.7968
***θ*(CS2S1)**	91.16	91.1	91.5(1)	91.6(1)	91.6(1)
***φ*(CS2S1X)**	130.30	130.32	130.7(3)	130.7(3)	130.6(3)
r(H1C)	1.0886	1.0887	-	1.0904	1.0904
θ(H1CS2)	110.66	110.63	-	110.60	110.53
ϕ(H1CS2C)	-64.34	-64.00	-	-64.26	-63.90
r(H2C)	1.0866	1.0865	-	1.0885	1.0881
θ(H2CS2)	107.35	107.35	-	107.32	107.31
ϕ(H2CS2H1)	-118.74	-118.73	-	-118.81	-118.82
r(H3C)	1.0873	1.0872	-	1.0891	1.0889
θ(H3CS2)	109.94	109.94	-	109.89	109.84
ϕ(H3CS2C)	58.05	58.36	-	57.98	58.27
σ			4.8×10^{-3}	4.4×10^{-3}	5.1×10^{-3}

[a]Atom labelling according to fig. 4.1. [b]Intermolecular parameters in bold. [c]Taken from ref. [167]. [d]Taken from ref [168]. [e]The parameters not reported are fixed at the SExp values of the monomers.

CCSD(T) energies using the cc-pVnZ basis sets, with the n=T,Q,5 for the HF energy and n=T,Q for the CCSD(T) correlation contribution. Instead, the ChS approach is a "geometry" scheme, with the combination n=T,Q of the cc-pVnZ family of basis set used to extrapolate the MP2 parameters. The cc-pVTZ set is employed for the CCSD(T) leading term. In both cases, the CV term is obtained in conjunction with the cc-pCVTZ basis set. The intermolecular parameters fitted in all the TMA structures are: the S−S distance, the OSS and CSS angles, and the dihedral angle formed by the C−S−S plane and the bisector plane of the complex containing the two sulfur atoms and the dummy atom (SSX; see fig. 4.1).

From the the values of table 4.4 it is observed that the quality of the different fits of TMA structures is similar, with the best performance showed by the ChS-TMA structure. The latter is consistent with the TMA structure that employs the SExp of the monomers to improve the parameters to be kept fixed, while the standard deviation

when using the CCSD(T)/CBS+CV level to template is slightly larger. For the latter, the SS bond distance is predicted 5-6 mÅ longer compared to the ChS-TM geometry, while the difference of the intramolecular angles is in agreement within the associated errors. By comparing these structures with the partial SExp one, it is observed that the intermolecular parameters of the TMA structures well reproduce those of the partial SExp, all being consistent within the associated errors. Finally, to understand the effect of the TM approach, one can compare the best TMA structure (ChS) with the one obtained without TM approach, where the intramolecular parameters are fixed at the CP-B2-D3/junTZ values. The difference is remarkable: the SS bond distance is 2.932(4) Å for the CP-B2-D3/junTZ fit and it is thus 14 mÅ longer than the value in the ChS-TM structure. The dihedral angle, i.e. CSSX, is also somewhat different from the TMA structures, indeed being 2 degrees larger. Therefore, it can be concluded that the TM approach leads to reliable results, greatly improved and at a price of a limited computational cost. Indeed, one can employ the SExp equilibrium structures of the monomers, which might be available in the literature [41]; alternatively, the equilibrium geometry of the monomers can be obtained via cost-effective approaches such as the ChS model.

Still considering the TM approach, a further analysis is needed to understand the importance of the level of theory used to compute the $r_e^{\mathrm{comp,complex}}$ and $r_e^{\mathrm{comp,monomer}}$ terms in eqs. (4.1) and (4.2). For this analysis, $r_e^{\mathrm{acc,monomer}}$ is fixed at the SExp structure which gives the best values according to the previous discussion, and the computed geometries are evaluated using the junChS and junChS-F12 approaches. The results of this analysis are collected in table 4.5, where the term "fixed" indicates that the fit is carried out with the intramolecular parameters fixed at the corresponding SExp value of the monomer. All the fitted SExp equilibrium parameters of table 4.5 well compare with the Partial SExp structure of table 4.4, indeed being consistent within the predicted errors. This implies that the TM approach strongly benefits from the use of SExp equilibrium parameters for the monomers, even if fixed and not templated. Further information can be gained from the comparison of the TMA structures in table 4.5 and the "SExp-TMA" structure of table 4.4, where the latter uses the CP-B2-D3/junTZ level of theory for the $r_e^{\mathrm{comp,complex}}$ and the $r_e^{\mathrm{comp,monomer}}$ terms. In this comparison, all the intermolecular parameters are in good agreement, thus confirming the reliability of using double-hybrid DFT methods within the TM approach.

Summarising, a S···S chalcogen bond was unveiled in the DMS$-SO_2$ complex thanks to the joint (experimental) rotational spectroscopy and quantum chemistry study. The former provided precise rotational constants that have been unambiguously assigned to the structure reported in fig. 4.1 by means of accurate computations based on the junChS and junChS-F12 approaches (see table 4.2). The nature of the NCI was investigated via NBO analysis that pointed out a charge transfer from the S atom of DMS to the S$-$O antibonding orbital of SO_2 fragment. Then, the accurate intramolecular pa-

rameters of the DMS$-SO_2$ complex were also derived by means of the SExp approach, employing the experimental data of seven isotopic species and the computed vibrational corrections obtained at the B3-D3/SNSD level. This nc-complex was also useful to test the combination of the SExp approach with the TM one, which is, according to the results, a reliable model to provide accurate intermolecular parameters at a limited computational cost and without requiring a large number of experimental data. The S$\cdots$S bond distance is predicted at 2.947(3) Å from the partial SExp approach (table 4.4) and 2.945(2) Å from the junChS-TMA approach (table 4.5). This bond length is $\sim$0.65 Å shorter than the sum of the corresponding van der Waals radii, thus further confirming the chalcogen NCI.

4.1.2 The trifluoroacetophenone-water complex

As mentioned in the Introduction, H-bonds are strong NCIs showing the particular property of being directional. This leads to specific conformations in isolated fragments and nc-complexes. Since H-bonds have this specificity, it is interesting to analyse the rearrangement occurring in a nc-complex characterised by at least one H-bond when one or more H atoms are substituted. Indeed, this kind of analysis is useful for the design of new drugs or, more generally, whenever one has to introduce a minor modification in a nc-system to slightly change its chemical properties, tuning —for example— the strength of NCIs. In this regard, an interesting substitution to consider is H $\longrightarrow$ F, which induces a change in the electronic arrangement of the molecule, this becoming progressively more significant as the number of substituted H atoms increases [169,170]. The effects of this electronic modification reflect directly in the chemical properties of the nc-complex and on the type of NCIs characterising the complex. An example is provided by the benzene-water complex which is characterised by an O$-$H$\cdots\pi$ NCI [171], while its mono- and di-fluorinated derivatives form nc-complexes involving both O$-$H$\cdots$F and C$-$H$\cdots$O bonds [172]. If one moves to the adduct formed by water and the fully fluorinated form of benzene (C_6F_6), a different NCI is observed: the one between the oxygen lone pair and the π hole of the benzene ring due to the very electronegative fluorine atoms [173]. A common fluorinated substituent in drugs is $-CF_3$ moiety [174] that, once inserted in a molecule, induces strong conformational modifications, thus influencing its interaction upon non-covalent bonding. The effects of the $-CF_3$ group on a molecule have been tested on anisole ($PhOCH_3$). In this molecule the methyl group lies on the same plane as the phenyl ring, but in $PhOCF_3$ the fluorinated group is perpendicular to the ring plane [175–177]. The same conformational behaviour is also observed in the case of O $\longrightarrow$ S substitution, i.e. moving from $PhOCH_3$ and $PhOCF_3$ to $PhSCH_3$ and $PhSCF_3$ [178,179], thus pointing out that the conformational change is only due to the H $\longrightarrow$ F substitution and quite independent of the chalcogen atom involved. If one considers these species involved a nc-complex formed with water, a similar conformational behaviour is observed. The main NCIs

observed, which are the O−H···O and the C−H···O H-bonds, lie in the plane for the $PhOCH_3$ and $PhSCH_3$ systems, while they are out-of-plane for the fluorinated counterparts. For biologically-relevant species, a common functional group is the C=O moiety and would be of particular importance the analysis of the H $\longrightarrow$ F substitution in their presence. As prototype of a biological molecule containing the "CH_3C=O" group, acetophenone (AP) can be considered, for which accurate rotational spectroscopic data are available in the literature together with those for its water complex [180,181]. The rotational spectrum of the trifluorinated counterpart (2,2,2-trifluoroacetophenone, TFAP) has also been studied [182], but the corresponding water adduct has never been investigated. For this reason, in ref. [43], the nc-complex formed by TFAP with H_2O was considered, thereby exploiting the synergism of quantum chemistry and rotational spectroscopy. This strategy was employed to provide an accurate characterisation of the main NCIs occurring in TFAP, and to highlight the role of trifluoromethylation in tuning their behaviour with respect to the AP-water complex.

The stable isomers of the TFAP-H_2O complex were obtained from sampling its PES at the revDSD level of theory (sill employing the jun-cc-pVTZ basis set). The optimised structures are shown in fig. 4.2, and it is evident that they mainly differ for the type of H-bonds formed. In fact, water can interact with the phenyl ring, the $-CF_3$ group, and/or the C=O bond, thus forming the O−H···π, O−H···F and O−H···O interactions, respectively. To understand the relative stability of these conformers, the energies were refined employing the junChS approach on top of the revDSD geometries, also accounting for the ZPE correction at the same level of theory. The stability order was then confirmed at the junChS level and the equilibrium energies, together with the corresponding ZPE corrected values, are reported in table 4.6. It is noted that the isomers ***I*** and ***II*** are very close in energy, these being separated by only 4.5 kJ·mol^{-1} at the junChS+ZPE level. This small energy difference is explained by the fact that both isomers form a O−H···O=C H-bond, but only isomer ***I*** is able to form a secondary NCI, namely the weak C−H···O−H bond. The remaining isomers, i.e. ***III*** and ***IV***, do not show the presence of any H-bond, and their stability is due to C−H···O and O−H···F contacts for isomer ***III*** and to O−H···F and O−H···π interactions in the case of isomer ***IV***. Since isomer ***I*** is the most stable structure, it is likely that it will be the only species observed in the experiment. The latter was carried out by producing in a supersonic expansion the TFAP-H_2O complex and recording its rotational spectrum with a COBRA-type spectrometer [164]. To form the supersonic jet, a He buffer gas at a stagnation pressure of 0.1 MPa was passed over a mixture of TFAP heated at 313 K and water, which was then expanded through a valve into the Fabry–Pérot cavity at a pressure of about 10^{-5} Pa.

The number of rotational lines measured and assigned is $\sim$160 and they are all consistent with only one species, belonging to the C_s symmetry point group. Therefore, considering the structures of fig. 4.2 and the computed dipole moments reported in

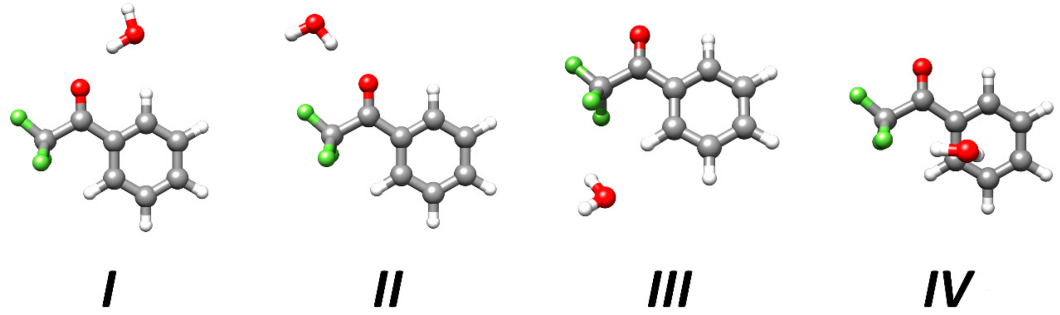


Figure 4.2: The stable conformational isomers of the TFAP-H_2O complex at the revDSD level.

Table 4.6: CP-corrected interaction energy[a], equilibrium dissociation energy[a], equilibrium rotational constants[b], and the dipole moment components[b] of the TFAP-H_2O isomers.

	Unit	*I*	*II*	*III*	*IV*
ΔE	kJ·mol^{-1}	0.0	5.0	9.3	12.6
ΔE_0	kJ·mol^{-1}	0.0	4.5	6.9	9.4
E_{int}	kJ·mol^{-1}	-23.24	-17.56	-13.62	-10.65
D_{e}	kJ·mol^{-1}	-22.36	-17.29	-13.11	-9.96
A_{e}	MHz	891.23	1277.58	1027.75	1193.21
B_{e}	MHz	611.55	472.14	603.12	520.70
C_{e}	MHz	387.60	367.15	408.30	483.30
μ_{a}	D	1.8	5.5	1.5	2.6
μ_{b}	D	1.7	2.2	2.1	0.7
μ_{c}	D	0.0	0.0	0.5	2.5

[a] Based on junChS values on top of revDSD geometries. [c] Obtained from revDSD optimisation.

table 4.6, the presence of Isomer ***IV*** can be ruled out. At this point, the observed transitions could belong to isomer ***I***, ***II***, or ***III*** and the discrimination can be made thanks to the comparison of computed and experimental rotational constants. In particular, the computed values (revDSD level, see table 4.6) of isomer ***I*** are in very good agreement with the experimental ones of table 4.7, thus pointing out the observation of only this isomer for the TFAP-H_2O adduct.

In a following step, the focus moved on the recording of the rotational spectra of four different isotopologues containing mono- and bi-deuterated water (TFAP-D_2O, TFAP-DOH and, TFAP-HOD), and $H_2^{18}O$. The rotational parameters for these isotopic species are reported in table 4.7. Owing to the rotational constants of these four isotopologues, a SExp equilibrium determination of the intermolecular parameters can be obtained employing the TM approach described in section 4.1.1 in combination with the computed values obtained at the revDSD level. As accurate structure for the TMA, the SExp structures of water and TFAP have been used. The former is reported in ref. [41], while the one of TFAP was purposely derived using the experimental data

Table 4.7: Experimental spectroscopic parameters of different isotopic species for isomer ***I*** of the TFAP-H_2O complex. For comparison, the theoretical values of the parent species are also reported.

	Theory[a]	Experiment[b,c]				
	parent	TFAP-H_2O	TFAP-$H_2^{18}O$	TFAP-D_2O	TFAP-HOD	TFAP-DOH
A_0 (MHz)	879.39	878.0858(1)	833.9848(2)	830.3953(1)	845.3957(1)	861.6862(1)
B_0 (MHz)	605.16	609.4679(1)	607.227(1)	608.7236(6)	608.6732(9)	609.5202(9)
C_0 (MHz)	383.35	384.50355(4)	374.9407(1)	374.82029(6)	377.82212(9)	381.3495(1)
D_J(kHz)	0.09	0.121(1)	0.117(2)	0.1115(9)	0.115(1)	0.115(1)
D_{JK}(kHz)	-0.11	-0.147(3)	-0.143(3)	-0.138(1)	-0.140(2)	-0.141(1)
D_K (kHz)	0.03	0.035(2)				
d_1 (kHz)	0.04	0.0557(5)				
d_2 (Hz)	-6.6	-6.3(3)				
σ (kHz)		2.4	2.7	2.1	2.4	2.8

[a] Vibrationally corrected rotational constants obtained from the equilibrium revDSD rotational constants (table 4.6) and the vibrational corrections at the B3-D3/SNSD level; quartic centrifugal distortion constants are also computed at the same level. [b] Standard error in parentheses in unit of the last digit. [c] D_K, d_1, and d_2 values were kept fixed at the value of the parent species for all the isotopologue considered.

Table 4.8: Vibrational corrections to the rotational constants of TFAP-H_2O at the B3-D3/SNSD level.

	TFAP-H_2O	TFAP-$H_2^{18}O$	TFAP-D_2O	TFAP-HOD	TFAP-DOH
ΔA_0 (MHz)	11.872	10.716	10.945	11.961	10.787
ΔB_0 (MHz)	6.397	6.466	6.246	6.355	6.265
ΔC_0 (MHz)	4.256	4.099	4.025	4.256	4.008

reported in ref. [182], with the resulting structure being reported in ref. [43]. The vibrational corrections employed in the SExp procedure reported in table 4.8 and the intermolecular parameters fitted are those involving the oxygen atom of water and the oxygen atom of the carbonyl group, together with the angle formed by these atoms with the C atom of the C=O moiety. The purpose of choosing only heavy atoms in the fitting procedure is merely due to the fact that the rotational constants carry more accurate information on such atoms than on the lighter H atoms involved in the intermolecular bond. However, from the resulting structure the O−H···O and O−H···H intermolecular distances as well as the corresponding angles have been derived.

The comparison of their SExp structural values with the revDSD counterpart is reported in fig. 4.3. A good agreement between the two structures is noted, but different trends are evident for the bond distances. Indeed, the SExp equilibrium parameters predict a O−H···O bond shorter by about 0.02 Å than to the revDSD value. Similarly, the SExp O···H−C contact is longer by about 0.06 Å than the revDSD counterpart. Therefore, the SExp procedures somewhat strengthens the H-bond between water and

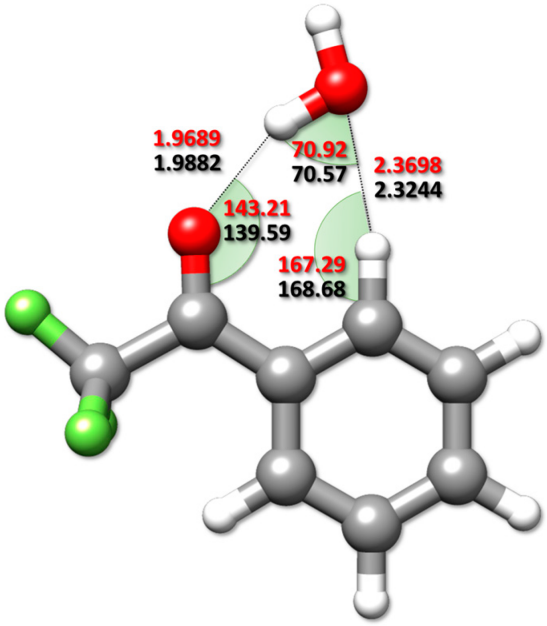


Figure 4.3: Equilibrium intramolecular parameters of the TFAP-H_2O complex, in red those obtained from the SExp+TMA procedure and in black those at the revDSD level. Distances in Å and angles in degrees.

the carbonyl group compared to the computed geometry. For angles, the maximum discrepancy of 4° is observed for H···O=C, which varies from 139.59° (revDSD) to 143.21° (r_{SExp}), thus towards a more linear H-bond, further confirming the previous discussion about the strengthening of the O−H···O bond.

According to the literature [182], trifluoromethylation in TFAP introduces a strong electronic density rearrangement compared to AP. In particular, the $-CF_3$ group increases its electron density, thus becoming more negative and inducing a depletion in the aromatic ring and in the C=O moiety. In view of these changes, the NCIs formed by TFAP/AP with water are also expected to change. The effects of fluorination on NCIs can be unravelled by comparing the TFAP-H_2O and AP-H_2O complexes. The most stable isomers of the former complex are reported in fig. 4.4 (panels (a) and (b)) together with the experimentally observed molecular complexes for the AP-H_2O system (panels (c) and (d)). The experimentally characterised species, i.e. isomers ***I*** and ***II*** of AP-H_2O and isomer ***I*** of TFAP-H_2O, show a strong O−H···O H-bond, which largely stabilises the complex, but at the same time this is not the only driving force. Indeed, isomer ***II*** of TFAP-H_2O also presents the same type of NCI, but it was not experimentally observed. This can be explained by the fact that, in the latter isomer, the H-atom is too far from the F atom to form a second NCI, which however contributes to the stabilisation of all the other species mentioned. Indeed, for isomer ***I*** of TFAP-H_2O (panel (b)), a secondary weak H-bond, i.e. O···H−C, is present and the same NCI also occurs in isomer ***II*** of AP-H_2O (panel (d)). Isomer ***I*** of AP-H_2O also shows a secondary weak O···H−C contact involving the CH_3 group, which is –for obvious reasons– not possi-

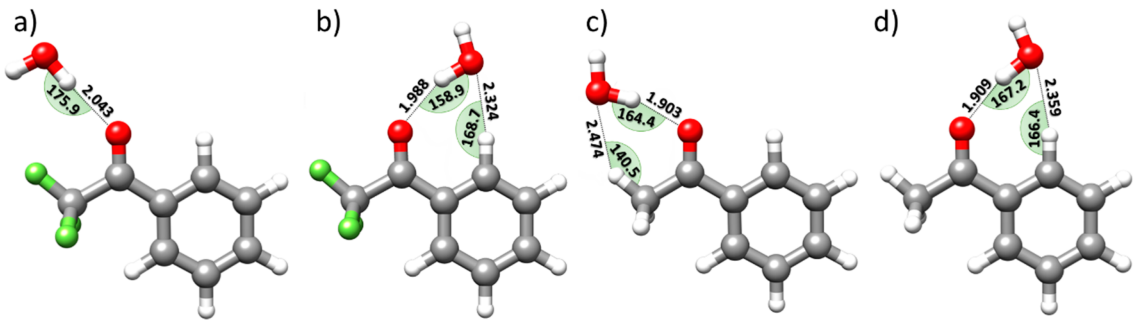


Figure 4.4: revDSD equilibrium structures for the two stable isomers of TFAP-H_2O (isomer ***II*** (a) and isomer ***I*** (b)) together with those of the AP-H_2O complex (isomers ***I*** and ***II***, in panel (c) and (d), respectively).

ble in the case of TFAP. These considerations are also supported by the non-covalent intermolecular parameters reported in fig. 4.4. Indeed, all the experimentally observed species show a O−H···O distance shorter than 2 Å at the revDSD level, which is a sort of diagnostic of a strong H-bond. In the non-observed isomer of TFAP (isomer ***II***), not only there is only one H-bond, but also it is characterised by a longer intermolecular distance (2.043 Å). From the previous discussion and the data of fig 4.4, it is evident that the main H-bond of the AP-H_2O isomers is stronger than that occurring in the TFAP-H_2O adduct, an outcome that is consistent with the previously mentioned electron density depletion of the carbonyl group in the case of TFAP.

To summarise, the NCIs present in the TFAP-H_2O complex were characterised owing to the strong interplay between of quantum chemistry and rotational spectroscopy. The leading interaction is a O−H···O H-bond supported by a weaker C−H···O contact. The measurement of the rotational transitions for five different isotopologues allowed for retrieving accurate equilibrium intermolecular parameters by exploitation of the SExp+TMA procedure. Finally, the effect of trifluoromethylation was also addressed in comparison with the AP-H_2O complex.

4.1.3 The benzofuran-formaldehyde complex

Particularly relevant, but often undervalued, NCIs occurring in biological environments are those involving aromatic and heteroaromatics moieties. Indeed, these interactions rule photosynthetic processes, DNA replication and other important biological functions. Additionally, heteroaromatics are also present in nucleobases and aminoacids and they are often scaffolds of drugs. Key heteroaromatic compounds are benzofurans (BZFs), whose derivatives show important pharmacological activities [183,184]. The NCIs that BZFs can establish with other biologically relevant groups are of great interest in order to understand their pharmacological activity and how they contribute to the stabilisation of complex systems [185, 186]. In the human activated blood coagulation factor X (FXa) [187], the methylbenzofuran group of the (*S*)-2-cyano-

1(2-methylbenzofuran-5-yl)-3-(2-oxo-1-(2-oxo-2(pyrrolidin-1-yl)ethyl)azepan-3-yl)guanidine ligand[2] interacts with three C=O moieties (see fig. 4.5), which are largely present in proteins. The NCIs occurring in such adduct are difficult to describe and disentangle because of the dimension of the ligand and the co-factor. Therefore, aiming at its characterisation, one could consider a prototype model like the BZF-formamide or BZF-formaldehyde adducts. Since BZF has two main competitive sites to form NCIs, either the π-system or the lone pair of oxygen, formaldehyde (FA) is the preferred system (with respect to formamide) because it cannot establish strong H-bonds with BZF. For this reason, the experimental and theoretical study carried out in ref. [44] aimed at the characterisation of the BZF-H_2CO complex, thereby exploiting the interplay of quantum chemistry and rotational spectroscopy.

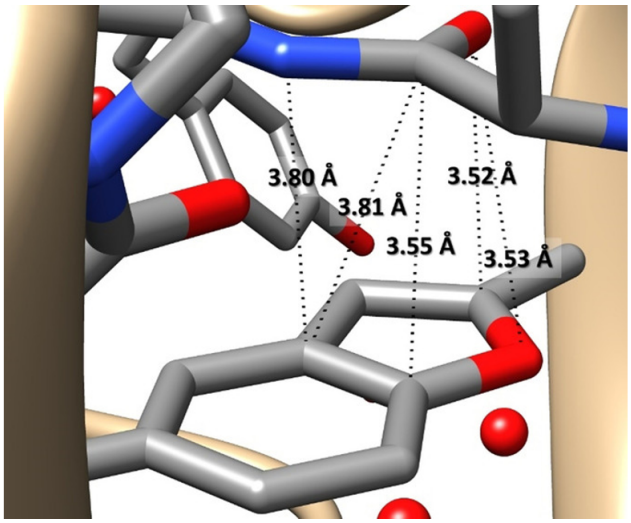


Figure 4.5: Region of the human activated blood coagulation factor X (FXa) showing close intermolecular contacts between a (2-methyl)-BZF moiety of the YET 2.D ligand and a carbonyl group.

Experimentally, the complex between BZF and FA was formed employing helium as carrier gas. The gas-line was in contact with two reservoirs, one contaning BFZ and the other polyformaldehyde (PFA), with the latter being heated to 355 K to produce FA. The mixture at about 3 bar was then supersonically expanded into the cell of a COBRA type jet Fourier transform microwave (FTMW) spectrometer [164]. The spectrum was recorded in the 8-10 GHz frequency range and 204 rotational transitions have been measured, which could all be assigned to only one structure of the BZF-FA complex. The structure of the observed species was assigned thanks to theoretical estimates. Indeed, the PES was investigated employing the B2PLYP-D3BJ/jun-cc-pVTZ level of theory and accounting for the CP correction in each step of the optimisation procedure (CP-B2-D3/junTZ). Fourteen low-energy minima were found on the PES and their electronic energy was further refined at the junChS level. Since all the minima lie within 4 kJ·mol^{-1}, their interconversion barriers were also investigated. This

[2](PDB:3HPT YET 2.D ligand) [188]

Table 4.9: Experimental ground-state rotational constants[a] and the corresponding computed vibrational corrections[a] for all the BFZ-FA isotopologues.

Isotopologue[b]	A_0	B_0	C_0	ΔA_{vib}	ΔB_{vib}	ΔC_{vib}
parent	1180.9038(2)	1096.1994(2)	788.2780(1)	-9.224	-20.572	-9.628
$^{13}C_{FA}$	1178.526(4)	1082.367(8)	781.4010(6)	-8.736	-16.717	-8.269
$^{18}O_{FA}$	1179.608(2)	1047.2712(6)	762.0013(3)	-10.032	-20.691	-10.023
$^{13}C2$	1165.146(6)	1095.96(1)	781.1750(8)	-10.074	-18.838	-8.669
$^{13}C3$	1169.799(4)	1092.838(6)	785.1122(5)	-9.194	-21.380	-9.969
$^{13}C4$	1172.73(1)	1090.81(1)	786.9611(4)	-9.292	-21.356	-10.185
$^{13}C5$	1168.897(3)	1092.213(5)	782.1475(4)	-10.363	-21.303	-10.371
$^{13}C6$	1171.869(8)	1089.31(1)	781.581(1)	-10.697	-20.589	-10.086
$^{13}C7$	1175.922(4)	1088.025(7)	785.4454(5)	-9.909	-21.479	-10.075
$^{13}C_{\alpha}$	1179.780(5)	1093.874(8)	787.1308(6)	-8.366	-18.725	-9.128
$^{13}C_{\beta}$	1178.786(7)	1094.24(1)	788.0239(9)	-7.914	-18.958	-9.166

[a]Values in MHz. [b]Atom labelling according to fig. 4.7.

pointed out that nearly all the structures can relax to the most stable minimum, i.e. isomer ***I***. All these stationary points are shown in fig. 4.6, where the CP-B2-D3/junTZ and junChS energy values are compared. To understand which is the experimentally observed structure, a comparison between experimental and computed rotational constants was required. The former are reported in table 4.9, while the computed equilibrium computational rotational constants are collected in table 4.10. The experimental parameters lie in between the equilibrium values of isomers ***I*** and ***IV***, thus requiring a deeper inspection. Due to the size of the system, the derivation of junChS geometries was excluded, but the incorporation of the vibrational corrections to the RCs led to an unequivocal assignment. The latter were computed at the CP-B3-D3/SNSD level and, once added to the equilibrium rotational constants, they allowed to undoubtedly confirm the presence of isomer ***I*** in the resonant cavity of the experimental set-up (see tables 4.10 and 4.9). Therefore, even if other isomers are formed, they relax to the most stable species, i.e. isomer ***I*** which is characterised by an interaction energy of -16.15 kJ·mol^{-1} at the junChS level (see table 4.10). Isomer ***II*** lies 0.15 kJ·mol^{-1} above ***I*** and it is further shifted to higher energies when the harmonic ZPE energy at the CP-B2-D3/junTZ level is introduced. The highest isomer observed is ***XIV***, which is 3.10 kJ·mol^{-1} above ***I*** and has a E_{int} of -12.67 kJ·mol^{-1}. As evident from fig. 4.6, isomers ***V*** and ***II*** directly interconvert to ***I***, while all the other interconversions involve two- or three-steps paths, like the ***XI*** $\longrightarrow$ ***IX*** $\longrightarrow$ ***II*** $\longrightarrow$ ***I*** relaxation.

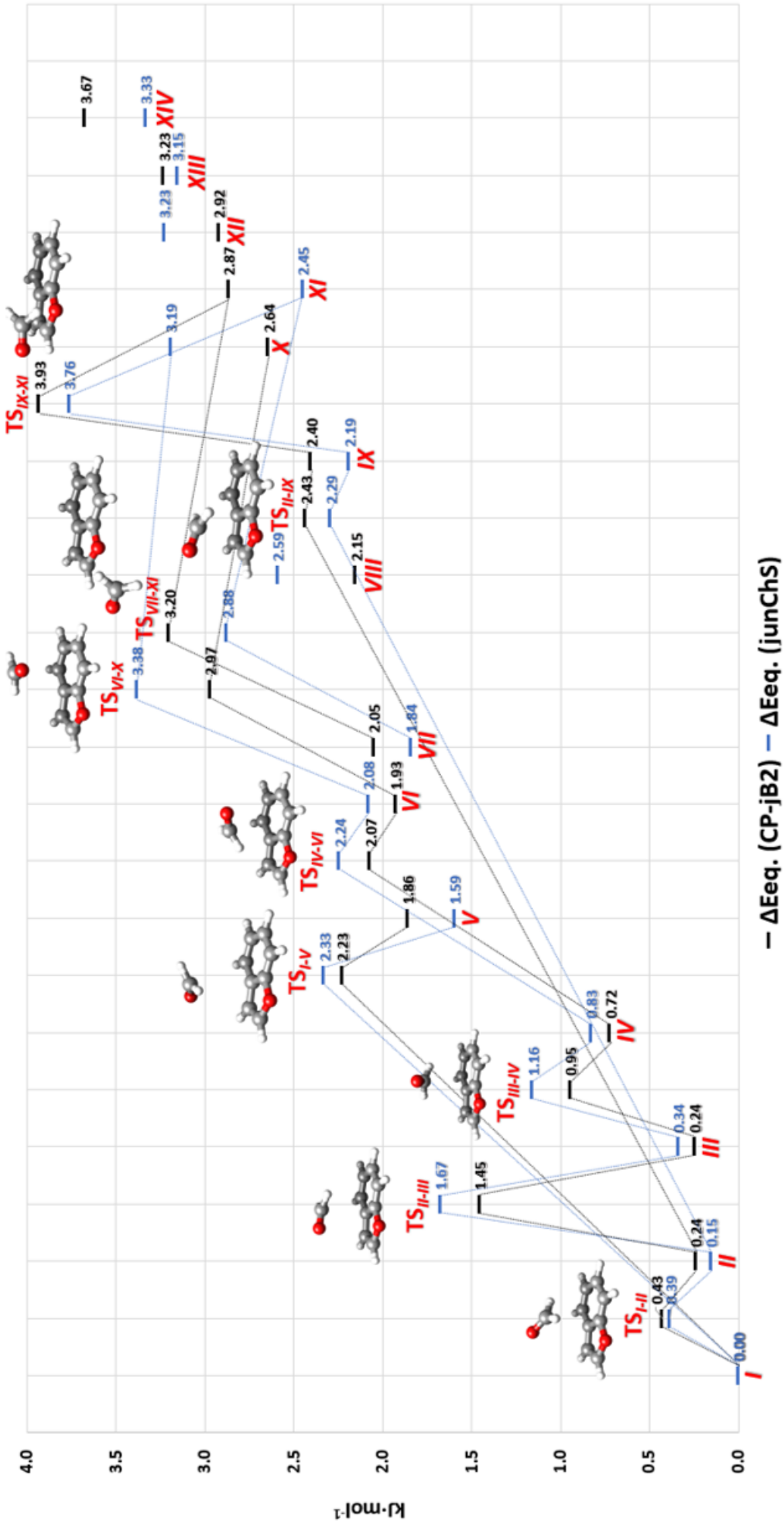


Figure 4.6: The fourteen minima of the BZF-FA complex together with the possible transition states for their interconversion. The relative junChS and CP-B2-D3/junTZ equilibrium energies (ΔE_{eq} in kJ·mol^{-1}) are reported.

A further confirmation that the isomer experimentally observed is ***I*** was achieved by recording and assigning the rotational transitions for all the mono-substituted ^{13}C species of BFZ and FA, together with the BFZ-$H_2C^{18}O$ isotopologue. The experimental rotational constants are reported in table 4.9 together with the computed vibrational corrections (CP-B3-D3/SNSD level). These quantities were used to derive SExp equilibrium rotational constants to be employed in a SExp fitting procedure combined with the TM approach. To apply the latter methodology, the accurate equilibrium structures of the monomers were required. In this case, the FA SExp equilibrium structure was available in the SMART Lab database [41, 189], while that of BZF was purposely derived using the data of ref. [190]. For the isolated BZF, the anharmonic force fields required to retrieve the vibrational corrections to rotational constants were computed using the B3-D3/SNSD level. The full list of data for BZF are reported in refs. [44, 190] and, due to the lack of the deuterated substitution, the C-H lengths were kept fixed at the B2-D3/junTZ level. The SExp equilibrium structure of BFZ is reported in ref. [44], while fig. 4.7 shows the SExp equilibrium parameters obtained for the BFZ-FA complex.

Table 4.10: CP-B2-D3/junTZ equilibrium rotational constants (in MHz), equilibrium relative energies (ΔE; in kJ·mol^{-1}) at the CP-B2-D3/junTZ and junChS levels, harmonic ZPE-corrected energies (ΔE_0; in kJ·mol^{-1}) obtained by adding the CP-B2-D3/junTZ harmonic ZPE contributions, and interaction energies (E_{int}; in kJ·mol^{-1}), for all the fourteen minima of the BFZ-FA system.

Isomer	A_e	B_e	C_e	ΔE		ΔE_0		E_{int}
				CP-B2-D3/junTZ	junChS	CP-B2-D3/junTZ	junChS	junChS
I	1189.4	1123.68	798.64	0	0	0	0	-16.15
II	1419.94	1044.06	878.27	0.24	0.15	0.35	0.26	-16.01
III	1373.64	1065.63	879.58	0.24	0.34	0.43	0.53	-15.75
IV	1211.55	1068.19	774.45	0.72	0.83	0.63	0.74	-15.32
V	1464.02	899.74	625.5	1.86	1.59	1.04	0.78	-14.48
VI	1381.13	1045.21	864.11	1.93	2.08	1.79	1.95	-14.07
VII	2883.8	542.31	456.47	2.05	1.84	1.68	1.47	-14.47
VIII	1429.24	866.93	720.62	2.15	2.59	1.69	2.12	-13.51
IX	1691.92	805.34	698.2	2.4	2.19	1.83	1.62	-13.89
X	1425.68	1044.97	894.59	2.64	3.19	2.27	2.82	-12.91
XI	2667.46	606.9	496.12	2.87	2.45	2.36	1.94	-13.57
XII	1394.38	1047.5	878.67	2.92	3.23	2.11	2.42	-12.86
XIII	1625.79	760.93	518.33	3.23	3.15	3.01	2.93	-13.08
XIV	1624.89	848.89	577.87	3.67	3.33	3.44	3.10	-12.67

The SExp equilibrium intermolecular distance between the C_{FA} and C_β atoms is of 3.2257$\pm$0.0006 Å, which is shorter than the sum of the carbon van der Waals radii (3.40 Å). A good agreement between the CP-B2-D3/junTZ and SExp values is noted, thus confirming the reliability of the DFT methodology in providing good estimate for nc-complex structures. The nature of the interaction was studied via NBO analysis, which pointed out to the clear presence of a $\pi \longrightarrow \pi^*$ interaction. According to junChS

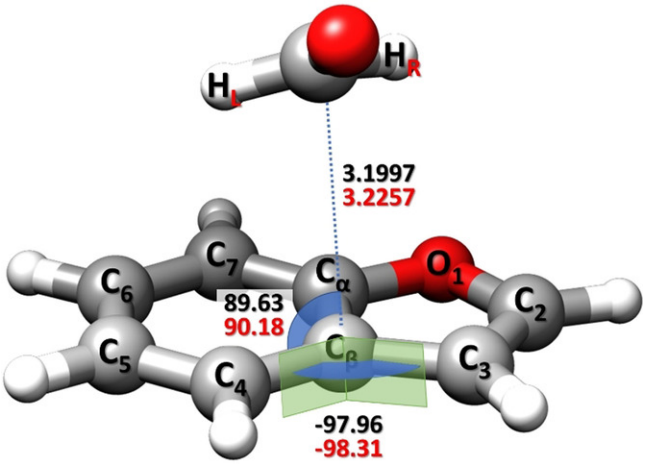


Figure 4.7: SExp equilibrium intramolecular parameters for the BFA-FA complex (red) in comparison to those computed at the CP-B2/junTZ (black).

results, the NCI has an interaction energy higher than that of the benzene dimer, i.e. -16.2 kJ·mol^{-1} for the former vs -11.6kJ·mol^{-1} for the latter.

To conclude, a comparison between the BFZ-FA structure and the more complex structure of fig. 4.5 is deserved. In particular, the matrix/solvent free structure of isomer ***I*** is not observed in the PDB:3HPT (YET 2.D) system, which instead shows a structure similar to that of isomer ***II***. Indeed, the latter two species have similar O(FA)···O(BFZ) and O(FA)···C2(BFZ) bond distances, i.e. 3.53 Å and 3.64 Å vs 3.52 Å and 3.59 Å. However, a large discrepancy in the C(FA)···Cα contact is also noted, which is 3.09 Å for isomer ***II*** and 3.55 Å for the PDB:3HPT (YET 2.D) structure. Finally, thanks to the NBO analysis of a selected portion of the PDB:3HPT (YET 2.D) crystal, computations were able to unvail that the $\pi \longrightarrow \pi*$ interaction plays a minor role in the solid phase compared to the gas phase, with the C−H···H NCI increasing its importance in the crystal.

4.2 Astrochemistry & Reactivity in the ISM

The chemical modelling of astrochemical environments has become particularly relevant in recent years as the number of molecules observed has been increasing at a rapid pace. The modellings aims at: (*i*) explaining the formation of molecules already observed in the ISM; (*ii*) proposing new exotic species of astrochemical interest via plausible formation routes, and (*iii*) supporting the astrochemical models via computation of accurate rate constants.

This thesis considers only reaction pathways in the gas phase, whose limiting factors are mainly two: low temperatures (from 10 to 200 K) and very low number density (1-10^8 cm^{-3}, where 10^4 cm^{-3} is approximate 3.8×10^{-10} Pa). The former affects the energy available to the reaction, which has to occur in the ISM without any external energy. Instead, low-pressures imply a low probability of collisions between reactive species and only two-body reactions are possible [45]. In practical terms, this means that the products and all the reaction steps to reach them must be submerged with respect to the reactants. In addition, to increase the probability that the collision is effective, one should consider very reactive species, like radicals or anions. The latter constrains is also an energetic requirement because barrierless approaches are required. Furthermore, they help in having exothermic processes. Therefore, the energetics establishes which are the potential reaction pathways that can occur in the ISM. In a second step, kinetics is able to evaluate the fastest process among those that are energetically allowed. However, kinetic estimates are strongly affected by the reaction barriers and one has to compute such quantities with great accuracy. For this reason, in the following, the junChS(-F12) schemes are proposed as ideal composite approaches to computed accurate reaction energies, even for an extended PES, whose TS and vdW intermediates can be considered weakly-bonded systems.

4.2.1 The methanimine + CP radical reaction

The presence of P-bearing species in the ISM is a key aspect that needs to be elucidated in astrochemistry. Phosphorus is quite ubiquitous on the Earth, also being an essential constituent of life, but its cosmic abundances are still unclear and its chemistry is poorly understood [191–193]. The high depletion of gas-phase phosphorus can be explained by consideration of its high desorption temperature, according to which a large quantity of P atoms should reside adsorbed on dust grains [192]. However, as reported in table 1.1, some small P-bearing species have been observed in the ISM, among which the PO and CP radicals [194, 195]. The latter is most probably produced via photodissociation of the HCP molecule, also detected in the ISM in 2007 [196]. The CP radical ($X^2\Sigma^+$) ha been identified in the carbon-rich circumstellar shell IRC+10216 [195], a very well studied astrochemical object because of its chemical richness. The chemical diversity of IRC+10216 provides an ideal environment to speculate on new and exotic

P-bearing species that might be present in the ISM. In this regard, one can consider the reaction pathway between the CP radical and methanimine (CH_2NH), which was also observed in IRC+10216 in 2010 [197]. The CP + CH_2NH reaction is particularly promising in view of other gas-phase formation routes proposed in the literature. In fact, the CP radical is isoelectronic with the CN and CCH radicals, which can efficiently react with CH_2NH to form E-/Z-cyanomethanimine and E-/Z-propargylimine, respectively [60, 198]. These species have been detected in the ISM and the above mentioned reactions are considered their main gas-phase formation routes.

As first step, the reactive PES of the CP + CH_2NH system was investigated by means of the B2-D3/junTZ method. For P atoms, the jun-cc-pVTZ basis set augmented by an additional tight d function (jun-cc-pV(T+d)Z) was used. This level of theory allows a rapid exploration of the PES, also ensuring a good estimate of the energetic barriers with respect to the reactants, which are used to include or exclude a reaction pathway for further consideration. Initially, the investigation focused on the location of the minima formed after the approach between the two species. Then, the inverse route from such minima to the reactants was considered, thus pointing out three open approaches, which are all barrierless, similarly to the reactions involving the CN and CCH radicals. In fact, the C-end of the CP radical can attack methanimine on the carbon atom, on the nitrogen atom or on the C=N π system. Starting from these minima, several TSs were hypothesised, and for those actually located on the PES, an Intrinsic Reaction Coordinate (IRC) analysis [199, 200] was carried out. In turn, IRCz allowed for finding new intermediates (minima), from which new TSs could be hypothesised in a recursive manner until the final bimolecular products were obtained. For each point on the PES, the harmonic force field was computed at the B2-D3/junTZ level, thus retrieving the corresponding ZPE and ensuring the nature of the stationary point considered. At this stage, more accurate estimates of the reaction barriers were required and the electronic energies were refined employing the junChS scheme. Since some stationary points could present spin-contamination and the wave function is not an eigenstate of the total spin operator ($\mathbf{S}^2$), the junChS energies were obtained using the restricted open-shell version of both CCSD(T) and MP2 computations. The reactive PES is illustrated in fig. 4.8, where the dotted orange lines represent the barrierless approaches. The following discussion is based on the junChS+ZPE energies (in blue in fig. 4.8), but the differences with the B2-D3/junTZ level of theory will also be addressed.

The approach of the CP radical on the C-end of methanimine results in MIN1. The latter has two conformers, MIN1z and MIN1e, arising from the different orientation of the N−H bond with respect to the C−CP moiety. These two conformers are quite close in energy, with MIN1z being located at -175.5 kJ·mol^{-1} and MIN1e being only 5 kJ·mol^{-1} above. The intercoversion between the two can also occur through to TSi, at -163.8 kJ·mol^{-1} with respec to the reactants. MIN1z and MIN1e can evolve directly into products via the cleavage of a C-H bond (TS1z and TS1e), thus leading to Z- or

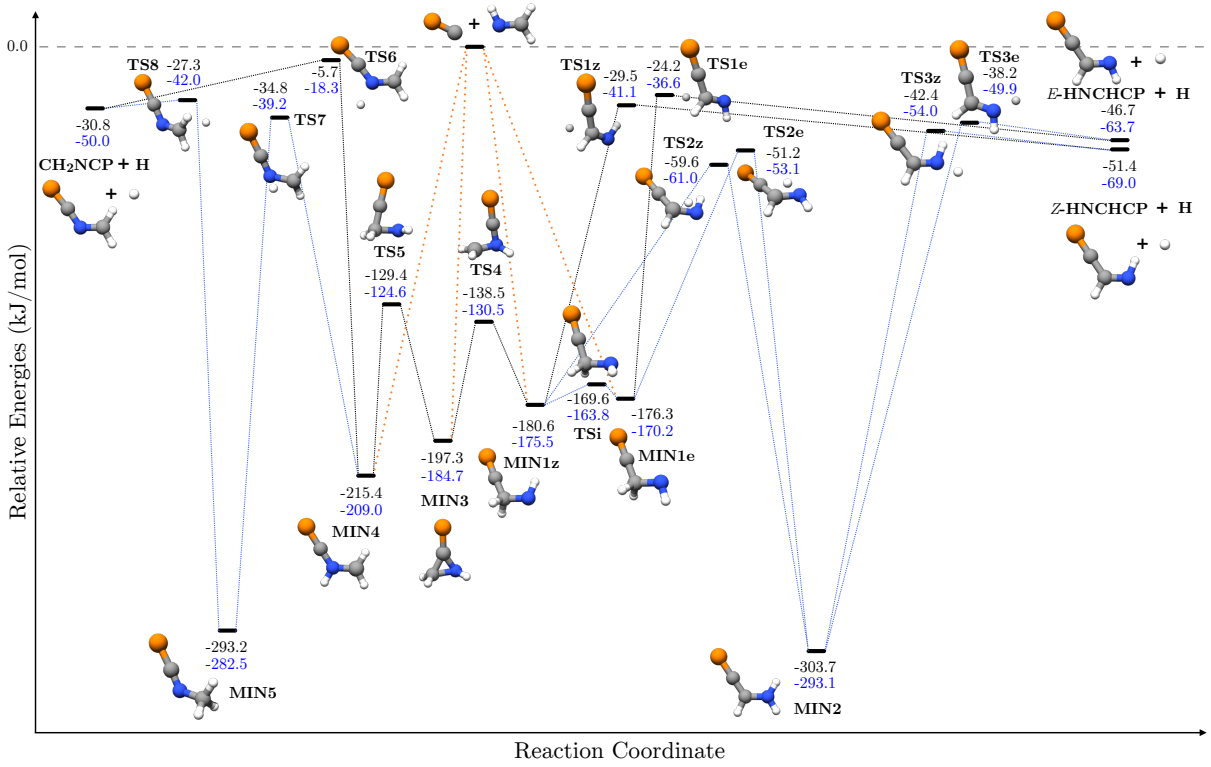


Figure 4.8: CP + CH_2NH reaction pathways: junChS relative energies (**black**) and their harmonic ZPE-corrected (**blue**) counterparts.

E-2-phosphanylidyneethan-1-imine (HN=CHCP), respectively and a H atom. These products can also be obtained from the same intermediates in a two-step procedure. Starting from MIN1z and MIN1e, a H atom of the CH_2 group can migrate on the NH group via TS2z and TS2e, respectively. From both the intermediate formed is MIN2, which is also the most stable minimum of the PES, being located at -293.1 kJ·mol^{-1} with respect to the reactants. MIN2 can evolve, via TS3z and TS3e, to the previously mentioned products, i.e. *Z*- and *E*-HN=CHCP, respectively. The barrier involved in the one-step procedure is slightly above 130 kJ·mol^{-1}, while in the two-step procedure, the first barrier is 114.5 kJ·mol^{-1} for the *Z* isomer and 117.1 kJ·mol^{-1} for the *E* isomer. MIN1z can also form, via TS4 located at -130 kJ·mol^{-1}, a cyclic intermediate (MIN3) that corresponds to the second barrierless approach of the CP radical, i.e. the one on the π system of methanimine. MIN3 is located at -184.7 kJ·mol^{-1} and can only evolve back to MIN1 or form MIN4 via a barrier of $\sim$60 kJ·mol^{-1} (TS5). MIN4 can also be formed directly from the barrierless approach of the CP radical on the N-end of methanimine, and it is located at -209.0 kJ·mol^{-1} with respect to the reactants. Similarly to MIN1, MIN4 can evolve to the bimolecular product, i.e. N-(phosphaneylidynemethyl)methanimine (H_2C=NCP) + H, in two ways. The direct process involves TS6 at an energy of -18.3 kJ·mol^{-1}, while the two-step synthesis requires overcoming a first barrier of 166.2 kJ·mol^{-1} (TS7), which corresponds to the H-migration from the NH group to the CH_2 group. Then, the intermediate formed (MIN5) evolves into products via TS8 lo-

cated at an energy of -42.0 kJ·mol^{-1} with respect to the reactants. Also in this case, the two-step process involves a lower reaction barriers compared to the single-step mechanism. Therefore, two-step pathways are probably the preferred ones from the kinetics point of view. According to energetics, the most exothermic product is *Z*-HN=CHCP, lying 69.0 kJ·mol^{-1} below the reactants. The latter is followed by its isomer, i.e. *E*-HN=CHCP, at -63.7 kJ·mol^{-1} and then CH_2NCP at -50.0 kJ·mol^{-1}.

Table 4.11: Relative electronic energies (kJ mol^{-1}) of the stationary points for the CP + CH_2NH reaction.

	B2-D3/junTZ		junChS	
	Energy	ZPE Corrected[a]	Energy	ZPE Corrected[a]
Reactants	0.00	0.00	0.00	0.00
MIN1z	-201.48	-196.38	-180.59	-175.49
TSi	-189.85	-184.06	-169.57	-163.79
MIN1e	-196.65	-190.57	-176.29	-170.22
TS1e	-42.40	-54.73	-24.24	-36.58
TS1z	-47.15	-58.82	-29.46	-41.12
TS2z	-79.03	-80.38	-59.62	-60.98
TS2e	-69.92	-71.80	-51.19	-53.08
MIN2	-326.45	-315.89	-303.66	-293.10
TS3e	-60.41	-72.16	-38.18	-49.92
TS3z	-64.43	-75.96	-42.45	-54.00
E-HNCHCP+H	-73.10	-90.17	-46.67	-63.74
Z-HNCHCP+H	-68.14	-85.71	-51.41	-68.97
TS5	-147.84	-143.05	-129.44	-124.62
MIN3	-214.53	-201.92	-197.30	-184.75
TS4	-152.28	-144.29	-138.47	-130.48
MIN4	-241.05	-234.78	-215.36	-209.02
TS6	-36.00	-48.47	-5.68	-18.26
TS7	-58.84	-63.19	-34.76	-39.19
MIN5	-321.65	-311.06	-293.22	-282.53
TS8	-50.70	-65.57	-27.29	-42.04
CH_2NCP+H	-58.87	-78.08	-30.77	-49.96

[a]Harmonic ZPE corrections at the B2-D3/junTZ level.

The junChS and B2-D3/junTZ energies are compared in table 4.11, where an average energy difference of about 20 kJ·mol^{-1} is noted between the two levels. The improvement due to junChS does not affect the overall barrier heights but leads to a shift of the PES at higher, slightly more unstable, energies. The maximum difference is observed for TS6, which is predicted at -36.0 kJ·mol^{-1} by B2-D3/junTZ and at -5.7 kJ·mol^{-1} by the junChS approach. Although the point is still submerged with respect to the reactants, its junChS energy is quite close to that of the reactants and leads

Table 4.12: Relative electronic energies (in kJ·mol^{-1}) for the CH_2NH + CP → MIN1e/z → TS1e/z → Z/E-HNCHCP+H path.

	junChS	HEAT-like	$\lvert\Delta E\rvert$[a]
MIN1z	-180.59	-179.86	0.73
MIN1e	-176.29	-175.50	0.79
TS1e	-24.24	-24.67	0.43
TS1z	-29.46	-29.90	0.44
Z-HNCHCP+H	-51.41	-51.12	0.29
E-HNCHCP+H	-46.67	-45.86	0.81

[a] Absolute junChS – HEAT-like energy difference.

to an unfavourable path from the kinetic point of view.

To validate the junChS model for the accurate computation of reaction barriers, a small portion of the PES was further refined employing the HEAT-like approach mentioned in Chapter 2. The latter provides energies with an accuracy of 1-2 kJ·mol^{-1} and thus well within the chemical accuracy. This methodology is entirely based on CC theory including up to quadruple excitations. For the comparison, the one-step mechanism from MIN1z/e to the E/Z-HNCHCP products was considered and the results are collected in table 4.12. The largest deviation observed is of 0.8 kJ·mol^{-1} for MIN1e and E-HNCHCP, while on average an absolute deviation of 0.6 kJ·mol^{-1} is noted. This confirms the reliability and accuracy of the junChS energies, which are obtained at a much lower computational cost than the HEAT-like energies.

Thermodynamically, the three products of the CH_2NH + CP reaction can all be formed and, to understand which is the most favoured process, kinetics studies are currently under considerations. The latter will also consider the H-abstraction channel leading to the formation of HCP.

The presence of E-/Z-HNCHCP and/or CH_2NCP in astronomical environments, such as IRC+10216, can only be confirmed by observation of the rotational transitions of such species in the spectra collected by radio-telescopes and/or interferometers. In turn, these spectra can only be analysed in terms of accurate rotational transitions derived from experimental measurements in the laboratory. This step can be complicated by several factors, from the formation of the species of interest to the analysis of the recorded lines in terms of an effective Hamiltonian. In the case of the above mentioned products, the synthesis is challenging because they are all unstable species and have to be produced *in situ*. Based on the strategy employed for cyanomethanimine and propargylimine, a suitable precursor for pyrolysis is a N-containing phosphine or a P-bearing amine [201, 202]. Pyrolysis also produces other contaminant species inside the cell and, to interpret the laboratory measurements, an accurate and preliminary *ab initio* simulation is crucial. For this reason, all the main rotational parameters required to predict the rotational transitions of each product have been accurately computed.

The rotational constants of the vibrational ground state have been obtained starting

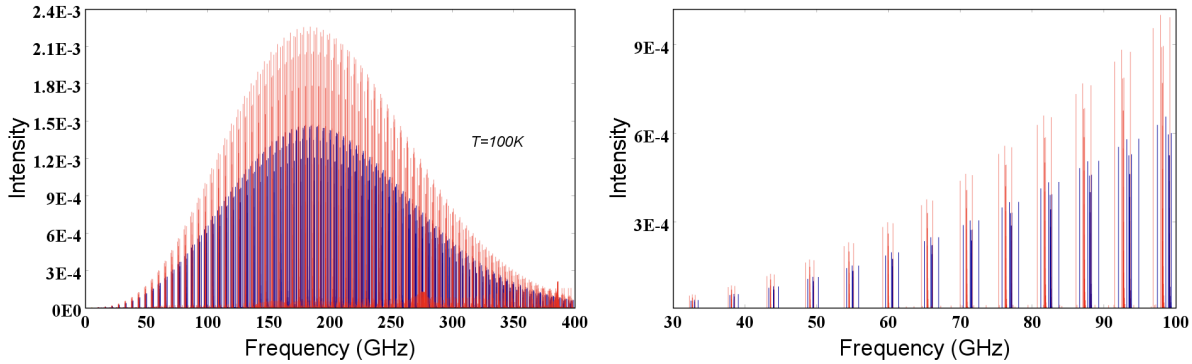


Figure 4.9: Left panel: Simulated rotational spectra of Z- and E-HNCHCP, blue and red lines, respectively, up to 400 GHz. Right panel: Zoom of the rotational spectra between 30 and 100 GHz. Intensity in nm^{2}MHz

from the equilibrium counterparts evaluated with the CCSD(T)/CBS+CV scheme. In the latter, the T,Q,5 combination of the cc-pVnZ family of basis sets was employed for the extrapolation to the CBS limit of the HF energy and the combination n=Q,5 for the extrapolation of the CCSD(T) correlation energy. The CV term was computed using the cc-pwCVTZ set. The vibrational corrections, required to correct the equilibrium rotational constants, were computed at the B2-D3/junTZ level of theory. The results are reported in table 4.13. The quartic centrifugal distortion constants have been determined from an ae-CCSD(T)/cc-pwCVQZ harmonic force field and are also reported in table 4.13 together with the absolute value of the dipole moment components. Since all the species are planar, μ_c is equal to zero by symmetry. Since rotational transitions are permitted only when a dipole moment is present, the allowed rotational transitions will depend only on the μ_a and μ_b components.

Using the rotational parameters of table 4.13, the rotational spectrum of E- and Z-HNCHCP was simulated in the range between 0 and 400 GHz and considering a rotational temperature of 100 K. The two spectra are superimposed in fig. 4.9. Both isomers show the maximum of intensity at 180 GHz, with the transitions of the E isomer (red lines in fig. 4.9) being more intense due to its larger dipole moment. The b-type transitions, i.e. those depending on the μ_b component, are not visible for the Z isomer due its small μ_b component, while they start to be visible at frequency higher than 250 GHz for the E isomer. Fig. 4.9 also shows a zoom for the 0-100 GHz frequency region, where the computational predictions should be more accurate. Indeed, the effects of centrifugal distortion on the line position is less pronounced at these frequencies. The rotational spectrum of CH_2NCP was simulated in the 0-500 GHz ranges and it is depicted in fig. 4.10. Due to its comparable μ_a and μ_b components, both a- and b-type transitions are visible, with the maximum of intensity being above the range considered.

The rotational constants obtained at the CCSD(T)/CBS+CV can be compared with those obtained from the junChS scheme. For the CH_2NCP species, the junChS values are: A_e=61608.63 MHz, B_e=3001.55 MHz and C_e=2862.11 MHz, thus differing only by 0.18% from the more

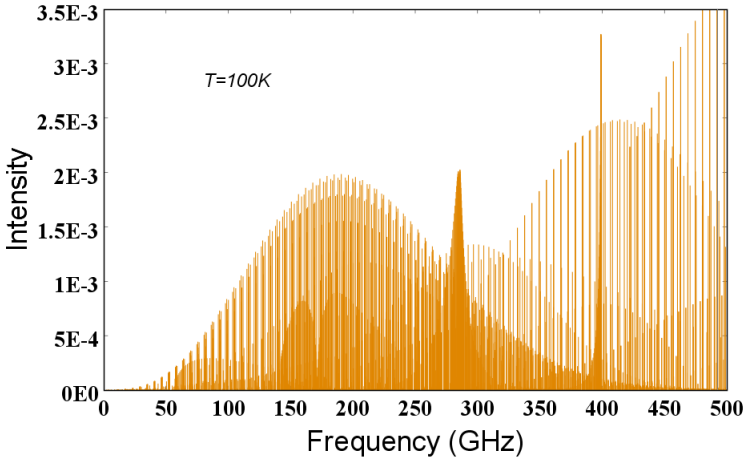


Figure 4.10: Simulated rotational spectrum of CH_2NCP in the 0-500 GHz frequency range. Intensity in nm^2MHz.

Table 4.13: Rotational spectroscopic parameters (MHz, if not otherwise stated) of *E*-/*Z*-HNCHCP and CH_2NCP.

Parameter[a,b]	*Z*-HNCHCP	*E*-HNCHCP	CH_2NCP
A_e	50319.38	58098.90	61454.73
B_e	2846.04	2796.71	2997.13
C_e	2693.69	2668.27	2857.76
ΔA_{vib}	272.77	52.49	1372.17
ΔB_{vib}	-11.55	-6.82	-17.25
ΔC_{vib}	-12.34	-8.66	-16.54
A_0	50592.15	58151.39	62826.90
B_0	2834.49	2789.89	2979.88
C_0	2681.35	2659.61	2841.22
D_J	0.65×10^{-3}	0.54×10^{-3}	0.63×10^{-3}
D_{JK}	-0.056	-0.066	-0.080
D_K	3.9	6.2	7.7
d_1	-0.90×10^{-4}	-0.70×10^{-4}	-0.88×10^{-4}
d_2	-0.47×10^{-5}	-0.32×10^{-5}	-0.45×10^{-5}
$\lvert\mu_a\rvert$ / D	1.58	1.90	1.68
$\lvert\mu_b\rvert$ / D	0.06	0.37	1.39
$\lvert\mu_c\rvert$ / D	0.00	0.00	0.00

[a]Watson's S reduction, I^r representation. [b]Equilibrium rotational constants at the fc-CCSD(T)/CBS(Q,5)+CV/(pwCVTZ) level. Vibrational corrections (ΔB_{vib^α}) at the B2-D3/junTZ level. Quartic centrifugal distortion constants at the ae-CCSD(T)/cc-pwCVQZ level. Equilibrium dipole moment components at the ae-CCSD(T)/cc-pwCVQZ level corrected for vibrational contributions at the B2-D3/junTZ level.

accurate CCSD(T)/CBS+CV values. A similar error is observed for the E and Z isomers of the HCNHCP species, for which the junChS rotational constants are A_e=50358.39 MHz, B_e= 2852.18 MHz and C_e=2699.30 MHz (Z-isomer) and A_e=58177.76 MHz, B_e= 2801.54 MHz and C_e=2672.83 MHz (E isomer).

To conclude, the proposed reaction between CH_2NH and CP represents a plausible formation route for new and exotic species bearing a P atom. This is also confirmed by the detection of E/Z-cyanomethanimine and Z-propargylimine, which can be possibly formed in the gas phase by the reaction of methanimine with the CN or the CCH radical, respectively. In addition, the CH_2NH + CP reaction is a another example of the general reaction mechanism involving methanimine and a radical species, which was found in all the above mentioned cases as well as for the OH radical [203]. Further developments include the comparison of the reaction mechanism of methanimine with that of ethene.

4.2.2 The oxirane + CN radical reaction

As already mentioned, accurate investigations of reactive PESs having as reactants species already detected in the ISM, can be used to propose new, potentially detectable, molecules. However, a reactive PES could also be used to unravel which is, for a given family of compounds, the isomer most likely present in the ISM. This possibility was explored for the C_3H_3NO family in ref. [63]. None of the species having this chemical formula was detected in the ISM, despite its simple atomic composition. Indeed, the astronomical searches for cyanooxirane (in 1996 and 2007 [204,205]) and propiolamide (in 2021 [206]), were unsuccessful and the class of C_3H_3NO isomers is still elusive as of March 2022.

Two plausible reactants for the formation of C_3H_3NO isomers in the gas phase, are oxirane and the CN radical ($X^2\Sigma^+$). In fact, the latter is one of the very first species observed in the ISM and it is widespread in such environment [207–209]. In addition, the CN radical is often employed as reactant in gas-phase astrochemical reactions. For example, it was invoked in the formation of cyanocyclopentadiene [210,211], benzonitrile [212, 213], and cyanomethanimine [198]. The other reactant, i.e. oxirane (also known as ethylene oxide), is also present in several astronomical environments, this being detected in molecular clouds [214, 215], star-forming regions [216] and protostars [217]. Oxirane is an ideal partner in astronomical reactions because: (i) its symmetry reduces the number of possible attacks, (ii) it offers the possibility of forming open-chain species as well as rings, and (iii) it is a well-known reactive species that should lead to exothermic products. Focusing on C_3H_3NO isomers, they can be obtained via simple loss of a H atom.

The reactive PES was investigated using the revDSD model. The strategy employed for the exploration of the reactive PES is the same described for the CH_2NH + CP reaction and it is detailed in section 4.2.1. The PES obtained for the c-C_2H_4O + CN reaction

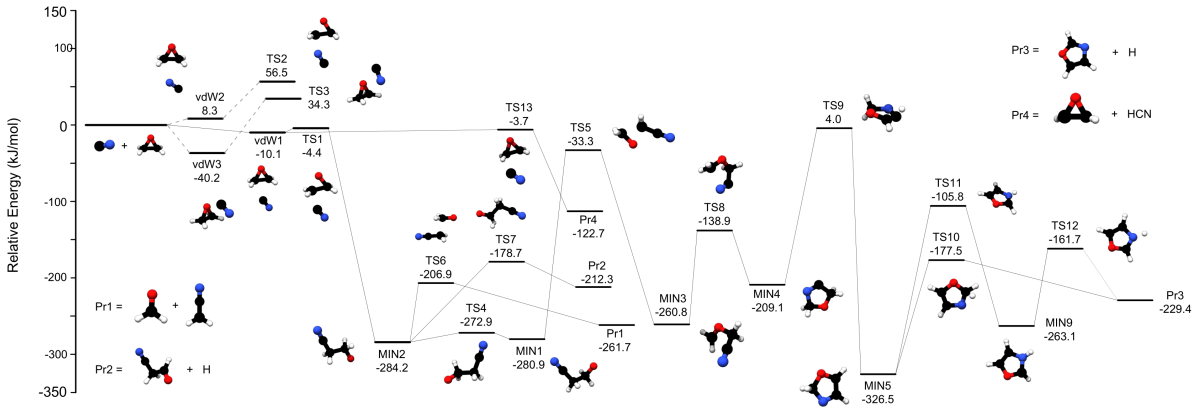


Figure 4.11: The most relevant stationary points located on the PES for the reaction between *c*-C_2H_4O and CN. Energies (in kJ·mol^{-1}) are given with the respect to the reactants.

is shown in fig. 4.11, where the electronic energies at the revDSD level, corrected for harmonic ZPE at the same level of theory, are reported. A summary of the energies is also given in table 4.14. In the following discussion, the energy in relative terms with respect to the reactants will be used.

The possible approaches of the CN radical on the oxirane ring are:

- The C-end of the radical attacks on the C atom of c-C_2H_4O, with formation of an initial precomplex vdW1 at -10.1 kJ·mol^{-1}. This can overcome a barrier of about 6 kJ·mol^{-1} (TS1) and forms MIN2.

- The N-end of the radical approaches the C atom of oxirane, thus leading to vdW2 at 8.3 kJ·mol^{-1}. The associated TS is located at 56.5 kJ·mol^{-1}, therefore this path was not further considered in view of the emerged barrier.

- The CN radical attacks the oxygen atom of the ring, thus forming a precomplex located at -40.2 kJ·mol^{-1} (vdW3). The following step involves a barrier of 74.5 kJ·mol^{-1} via TS3, located at 34.3 kJ·mol^{-1}. Therefore, this path was not further considered as well.

- The CN radical attacks on the H atoms of oxirane, thus leading to H-abstraction and forming HCN and 2-oxiranyl radical (Pr4). This is a one-step process involving TS13 located only 6.4 kJ·mol^{-1} above vdW1 (thus still being submerged).

Based on the discussion above, starting from the reactants, the reaction can either form Pr4 or proceed via MIN2 (located at -284.2 kJ·mol^{-1}). Focusing on the latter pathway, a rotation around the H_2C-CH_2 bond of MIN2 can occur to form its conformer, i.e. MIN1 at -280.9 kJ·mol^{-1}, or the same bond can break homolytically, then forming formaldehyde and the H_2CCN radical (Pr1, -261.7 kJ·mol^{-1}). Both processes involve a barrier: 11.3 kJ·mol^{-1} for the former, 77.3 kJ·mol^{-1} for the latter. However, MIN2

Table 4.14: Equilibrium and ZPE-corrected relative energies[a] for the most relevant stationary points located on the PES and reported in fig. 4.11.[b]

	Equilibrium[b]	ZPE-corrected[c]
Oxirane	-153.5760 E_h	-153.5182 E_h
CN	-92.5751 E_h	-92.5696 E_h
Reactants	0.00	0.00
vdW1	-12.96	-10.74
vdW2	1.64	8.32
TS1	-5.06	-4.36
TS2	57.73	56.49
TS3	34.61	34.29
vdW3	-45.05	-40.25
MIN1	-285.15	-280.87
MIN2	-288.29	-284.16
MIN3	-265.82	-260.85
MIN4	-224.42	-209.08
MIN5	-340.97	-326.52
MIN9	-277.47	-263.07
TS5	-31.77	-33.30
TS6	-204.68	-206.87
TS4	-276.66	-272.95
TS8	-145.41	-138.88
TS7	-162.54	-178.75
TS9	2.21	4.00
TS11	-107.84	-105.75
TS10	-172.69	-177.47
TS12	-156.37	-161.77
TS13	-2.04	-3.72
Pr1 (H_2CCN +H_2CO)	-248.25	-261.71
Pr2 (cyanoacetaldehyde + H atom)	-189.93	-212.34
Pr3 (oxazole + H atom)	-217.32	-229.44
Pr4 (HCN + 2-oxiranyl radical)	-114.78	-122.66

[a] All values in kJ·mol^{-1} if not otherwise specified. [b] For completeness, the absolute energies in Hartree units of the two reactants are also reported. [c] Equilibrium electronic energies obtained at the revDSD level of theory. [d] Electronic energies corrected by harmonic zero-point energy contributions, both computed at the revDSD level of theory.

also leads to Pr2, via cleavage of a C−H bond of the terminal H_2CO group. Pr2 is cyanoacetaldehyde, one of the C_3H_3NO isomer, with an H atom, and it is located at -212.3 kJ·mol^{-1}. The formation of Pr2 involves a reaction barrier higher than that involved in the production of Pr1 of about 30 kJ·mol^{-1}.

Lastly, the oxygen atom of MIN1 can attack on the C atom of the CN moiety, thus giving rise to MIN3 via a barrier of 247.6 kJ·mol^{-1} (TS5). MIN3 closely resembles cyanooxirane but the formation of such species was not observed on the PES. Indeed, the preferred path seems a ring-expansion, thanks to the attack of the terminal N atom on the other terminal group, CH_2. This process involves a barrier of about 120 kJ·mol^{-1} and forms the 5-memebered cyclic intermediate, MIN4. The latter can only evolve into MIN5 via an H-migration, whose TS (TS9) is located 4.0 kJ·mol^{-1} above the reactants. Then, MIN5 produces Pr3, i.e. oxazole + H, via two different processes. The first is a single step path via TS10 and requires overcoming a barrier of about 150 kJ·mol^{-1} due to a C-H bond cleavage. The second process involves two steps: first a H migration which occurs via TS11, with a barrier of 220.7 kJ·mol^{-1}; then, the MIN9 intermediate formed produces Pr3 via TS12, thus overcoming a barrier of 101.4 kJ·mol^{-1}. As mentioned above, the production of Pr3 involves a TS located 4 kJ·mol^{-1} above the the reactants, which was still considered in the present investigation in view of the error associated to double-hybrid DFT methodologies such as revDSD, which is about 20 kJ·mol^{-1} [62]. Therefore, the TS might be located even 10 (or more) kJ·mol^{-1} below the reactants.

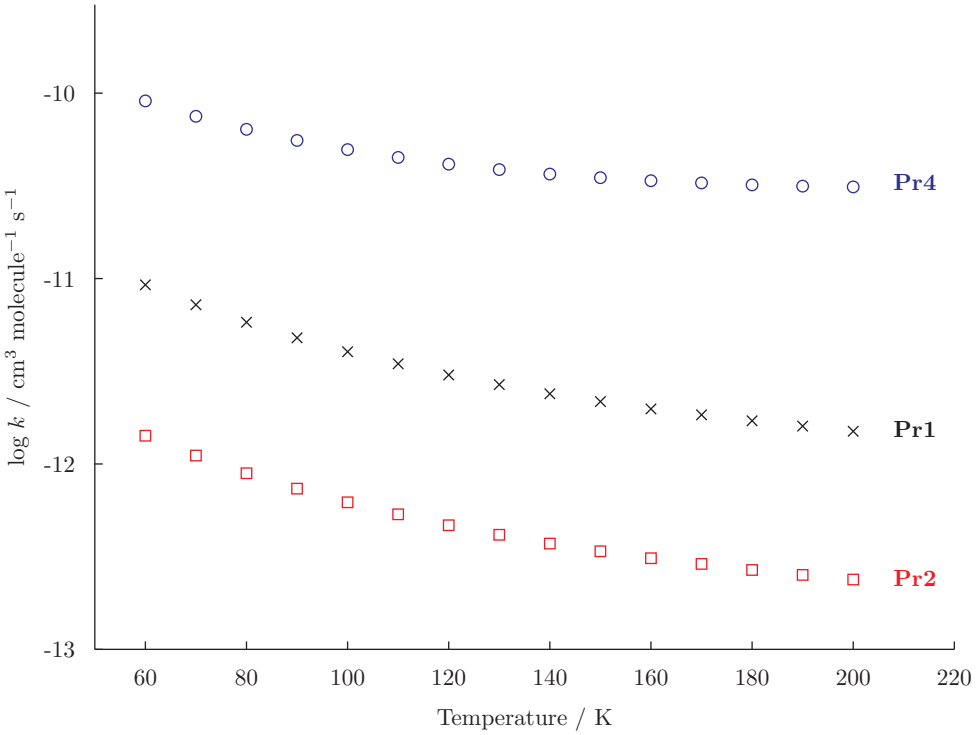


Figure 4.12: Rate constants (k) for the formation of Pr1 (black crosses), Pr2 (red squares), and Pr4 (blue dots) computed in the 60-200 K temperature range.

The kinetic modelling of the reaction between c-C_2H_4O and the CN radical was carried out to understand which is the most abundant product. To this aim, the MESS program [218,219] was employed and the reaction channel from the reactants to vdW1 was described using the phase-space theory [220, 221]. According to the latter, the

potential of the entrance channel was modelled as [222]:

$$f(R) = f(R_0) - \frac{C_6}{R^6} \tag{4.3}$$

where R is the distance between the two approaching fragments and, in this case, it was assumed to be equal to the distance between the two carbon atoms of the fragments ($C_{CN}\cdots C_{ox}$). The potential ($f(R)$) was obtained from a rigid scan of such coordinate from 3 to 10 Å, with $f(R_0)$ being the potential of the two non-interacting fragments. The fitted C_6 coefficient, required for the kinetic modelling, is 104.11 $a_0^6 E_h$. The remaining part of the reaction was then modelled using TS Theory within the RRHO approximation and accounting for tunnelling via the Eckart model. The other quantities required in the kinetic modelling, together with the energy barriers of table 4.14, are: (i) the rotational constants, (ii) the vibrational frequencies, and (iii) the spin multiplicity of the electronic state, for each stationary point. The former two points were obtained via geometry optimisations and harmonic force field calculations at the revDSD level, respectively. Instead, the spin multiplicity is equal to 2 (doublet) for open-shell species (radicals, intermediate and TS) and to 1 (singlet) for the closed-shell molecules (products and oxirane). All these quantities were then used by the program to derive the molecular partition function of each species and, then, the kinetic constant via resolution of the master equation. The latter equation correlates, via an appropriate matrix, the several differential equations that are required to describe the overall kinetic process. The diagonalization of the correlation matrix leads to the rate constants for each step and thus to the preferred kinetic path.

The rate constants were computed in the 60-200 K temperature range and at the pressure of 1×10^{-12} atm, i.e. in the low-pressure limit of the reaction. The results are schematically reported in fig. 4.12 for the considered temperature range. The rate constant for the formation of Pr4 is on the order of 10^{-10} cm^{3}molecule^{-1}s^{-1} and, kinetically, it is the most favoured path. The formations of Pr1 and Pr2 are comparable, with rate constants on the order of 10^{-11}-10^{-12} cm^{3}molecule^{-1}s^{-1}. The production of Pr3 is negligible in all simulations, the rate constant being on the order of 10^{-24} cm^{3}molecule^{-1}s^{-1}. According to these data, the reaction c-C_2H_4O + CN mainly leads to the "2-oxiranyl radical + HCN" product, thus suggesting the potential presence of 2-oxiranyl radical in the ISM. However, the latter is a highly unstable species which is not straightforwardly produced *in situ* and its rotational spectrum has never been measured in the laboratory, this preventing its detection in any astronomical source. In this regard, its first accurate computational spectroscopic characterisation is currently in progress.

The predominant formation of 2-oxiranyl radical does not preclude the formation of other products like, e.g. cyanoacetaldehyde. Indeed, the 2-oxiranyl radical could likely re-form oxirane by re-combination with other species, among which the most abundant species in the ISM, i.e. H_2. The newly formed oxirane molecule might react again with the CN radical and might follow a different reaction path. However, a clear

picture in this sense can only be obtained with a full chemical modelling accounting for all possible reacting species, such as H_2.

Concerning the C_3H_3NO family of compounds, the reaction between oxirane and the CN radical can only form cyanoacetaldehyde in the gas phase. However, the role of the solid phase (grains and ices) can not be ruled out and, thanks to their catalytic effect, other channels of the title reaction, such as the one forming oxazole, might be open in ISM conditions. The rotational spectrum of cyanoacethaldehyde was investigated in 2012 [223], but this species has never been searched for in the ISM and its abundances in any astronomical source is unknown.

To have a more reliable estimate of the rate constants, more accurate reaction barriers should be used and their derivation is currently in progress. In fact, some intermediates of the reaction have a multi-reference character and their energies can not be computed using the junChS(-F12) approaches described in the present thesis. The relevance of the static correlation for such stationary points was established by the $\mathbb{T}_1$ diagnostics, using the amplitudes of the single excitations as they are obtained from the CCSD method. For example, the $\mathbb{T}_1$ value obtained for TS1 is 0.044, with the limit for negligible static correlation being $\mathbb{T}_1 \leq 0.04$ in the case of open-shell species.

4.2.3 The propene + C_3N radical reaction

Among the observed molecules in 2021 (see table 1.1), a somewhat complex species is vinylcyanoacetylene CH_2CHC_3N [224], detected in the TMC-1 molecular cloud. For this species, the authors of ref. [224] propose two main formation routes: (i) the reaction between the CN radical and CH_3CHCCH [225] and (ii) the addition of the C_3N radical to C_2H_4 [226]. In the following, a different pathway for the formation of vinylcyanoacetylene (VCA) is presented and it is based on the reaction between the C_3N radical and propylene ($CH_2{=}CH{-}CH_3$). The C_3N radical ($X^2\Sigma^+$) was first observed in the TCM-1 cloud by Guélin and co-workers [227], while the presence of propylene was first reported in 2007 [228]. According to ref. [224], VCA could be and important intermediate in the synthesis of cyclic molecules, like pyrrole and pyridine, but also cyanocyclopentadiene, a five-membered ring whose isomers, 1- and 2-C_5H_5CN, have also been detected toward the TMC-1 molecular cloud [210, 211]. For this reason, the propene + C_3N reaction was also considered in view of a potential connection that between VCA and cyanocyclopentadiene. The only proposed formation route for cyanocyclopentadiene is the barrierless reaction between the CN radical and cyclopentadiene. The presence of the latter species was recently confirmed in the TCM-1 cloud [229], shortly after the reaction was the hypothesised.

The reactive PES of the "C_3H_6 + C_3N" system was investigated using an extremely convenient level of theory based on the hybrid B3LYP functional, including the D3BJ empirical dispersion and using the cc-pVTZ basis set (hereafter, B3-D3/TZ). For each stationary point, the corresponding harmonic ZPE correction was obtained at the same

level. To improve the electronic energy of the key stationary points, the geometries were refined employing the B2-D3/junTZ level. The final electronic energies were obtained, on top of the B2-D3/junTZ geometries, using the CCSD(F12*)(T) method and the cc-pVTZ-F12 basis set, indicated in the following as CCTZ-F12. Analogously to the CH_2NH + CP reaction, the ROHF reference wave function was employed to computed CCSD(F12*)(T) energies. Since this work was carried out during my 3-month stay in David Tew's research group at the University of Oxford, the Turmobole program was employed for the F12 computations [230].

A summary of the B2-D3/junTZ and CCSD(F12*)(T) electronic energies is reported in table 4.15, while fig. 4.13 shows the most relevant paths of this reactive PES. Going into details, the C_3N radical can approach C_3H_6 without any barrier on two different sites: either the CH_2 terminal group or the central sp^2 carbon atom. The former attack leads to MIN2, which is located 257.5 kJ·mol^{-1} below the reactants. The latter can evolve directly into vinylcyanoacetylene via cleavage of the $C(sp^3)-C(sp^2)$ bond, overcoming TS4 at -137.8 kJ·mol^{-1}. The other approach (i.e. attack on the central carbon atom of C_3H_6) leads to MIN1, which is connected to MIN2 via TS0 at -57.4 kJ·mol^{-1} (with respect to the reactants). This barrier is in competition with the formation of MIN3 that involves a reaction barrier of about 170.0 kJ·mol^{-1} (TS1). MIN3 evolves into MIN26b via H-migration and, from the latter, via a series of conformational rearrangements (TS26a-b, TS25a-b) and H-migrations (TS49 and TS52), all submerged with respect to the reactants energy, MIN21d is formed. From MIN21d, two paths can be followed. First, MIN21b is obtained via simple rotation around a C−C bond and, then, via formation of a new bond that involves a barrier of 242.1 kJ·mol^{-1}, the cyclic MIN5a2 is reached. This minimum forms MIN8 in four steps, two conformational rearrangements and two H-migrations. MIN8 is a key intermediate that leads to a 5-membered ring (MIN12) via TS17, thus overcoming a barrier of 79.5 kJ·mol^{-1}. However, this cyclic minimum can be obtained more straightforwardly from MIN21d, indeed the path MIN21d → TS43 → MIN22 → TS48 → MIN12, with MIN22 being already a 5-membered cycle similar to MIN12, is also present on the PES. The cyclic structures formed in these paths have the C_5H_6CN chemical formula and closely reassembles cyanocyclopentadiene. Via a large number of H-migrations, MIN12 and MIN22 can interconvert to other structural isomers: MIN13, MIN14, MIN16, and MIN15.

Among these species, MIN22 and MIN13, connected via TS1ii, can lead without barrier to Pr3, i.e. cyclopentadiene + the CN radical, which is located 171.9 kJ·mol^{-1} below the reactants. In addition, MIN22, MIN14, and MIN16 can form Pr1 via TS46, TS24, and TS28, respectively. Pr1 is 1-cyanocyclopentadiene and a H atom. MIN14 can also produce Pr2, i.e. 2-cyanocyclopentadiene + H atom, and a similar process is also possible from MIN15 and MIN13. Interestingly, the portion of the PES from Pr3 to Pr1 and Pr2 is actually the formation of 1- and 2-cyanocyclopentadiene from cyclopentadiene + the CN radical and it can be compared to the pathways reported

in ref. [210]. Such a comparison shows that the investigation of ref. [210] is actually partial and several minima were left out from that study. Moreover, the production of 1- and 2-cyanocyclopentadiene from cyclopentadiene is barrierless according to ref. [210], while in the present PES, several TSs for the cleavage of the appropriate C−H bonds have been found. Finally, the present investigation has also pointed out the possibility of forming MIN4ii, which is a reaction intermediate where the CN radical approaches cyclopentadiene with its N-end. MIN4ii is obtained barrierless from CN+C_5H_6 and through high reaction barriers from MIN13. The latter process involves TS59, MIN32 and TS65, with MIN32 being a bicyclic structure. Similarly to what occurs for MIN22, MIN4ii can also produce other structural isomers via H-migrations, i.e. MIN6ii, MIN7ii, and MIN8ii. These form, with reaction barriers, Pr4 and Pr5 which are the combination of a H atom with 1- and 2-isocyanocyclopentadiene, respectively.

1- and 2-isocyanocyclopentadiene might be detectable in the ISM, but two main issues could prevent their detection. First, based on the C_3H_6 + C_3N or C_5H_6 + CN reactions, they are about 90 kJ·mol^{-1} higher in energy than the cyanocyclopentadiene counterparts. In addition, isocyanates usually have a smaller dipole moment than cyanates; therefore, their rotational transitions are usually less intense than those of 1 and 2-C_5H_5CN. This renders particularly difficult to observe them in the spectra collected from TMC-1 or other astronomical sources characterised by similar very low temperatures ($\sim$10 K).

Concluding, the C_3N + C_3H_6 reaction provides an effective mechanisms for the formation of vinylcyanoacetylene, but also pathways for the formation of: (i) cyclopentadiene, (ii) cyanocyclopentadiene and (iii) isocyanocyclopentadiene. Indeed, this reaction also contains, among its open routes, the paths that characterise the reaction of the CN addition to cyclopentadiene, only partially explored in ref. [210]. The more detailed characterisation here provided points out the presence of exit barriers for the formation of both isomers of cyanocylopentadiene. These might be useful to explain the slightly different abundances of the two isomers in TMC-1. Lastly, it should also be mentioned that a hypothetical reaction where vinylcyanoacetylene and the CH_3 radical (Pr10) are the reactants to form cyanocylopentadiene (Pr4 and Pr5) would not be possible in the ISM conditions, due to the emerged reaction barriers ruling the PES. For the C_3N + C_3H_6 system, the refinement of the electronic energies is still in progress and it will involve the ChS-F12 scheme. In addition, the H-abstraction by the C_3N radical could also be an important path to consider and the optimisation of this path is currently under investigation.

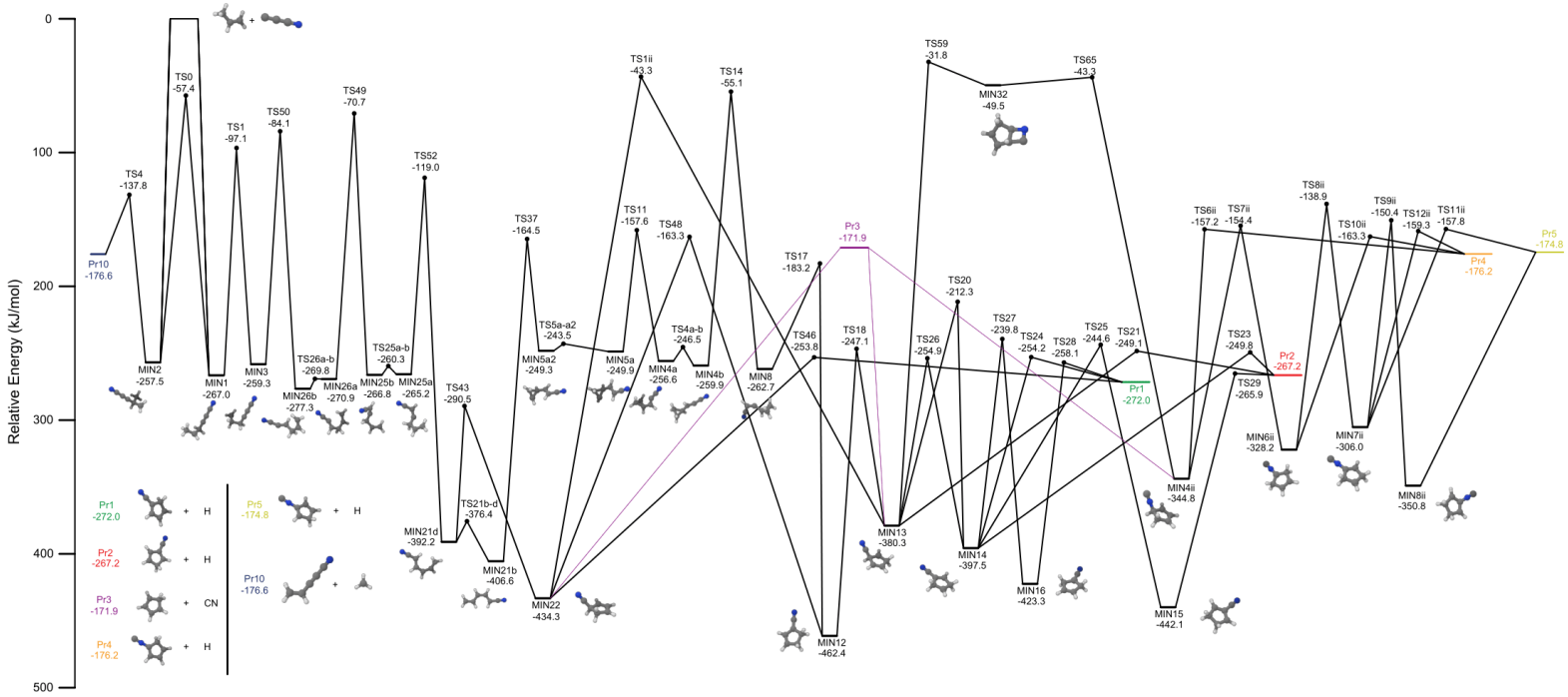


Figure 4.13: The reactive PES of the propene + C_3N reaction. CCTZ-F12 energies augmented by B3-D3/TZ harmonic ZPE.

Table 4.15: Relative energies (kJ·mol^{-1}) of the stationary points involved in the C_3N + C_3H_6 reaction[a]

	ZPE harm	B2-D3/junTZ	CCTZ-F12		ZPE harm	B2-D3/junTZ	CCTZ-F12
C3H6	0.0794	-117.8367 Ha	-117.7284 Ha	TS21	0.0927	-240.64	-249.11
C3N	0.0132	-168.7756 Ha	-168.6122 Ha	TS23	0.0925	-254.13	-249.76
MIN1	0.0978	-287.59	-267.01	TS24	0.0927	-259.03	-254.19
MIN2	0.0976	-275.97	-257.49	TS25	0.0963	-248.65	-244.61
MIN3	0.0977	-278.29	-259.32	TS26	0.0966	-261.67	-254.92
MIN4a	0.0996	-265.42	-256.58	TS27	0.0954	-247.94	-239.81
MIN4b	0.0998	-269.34	-259.90	TS28	0.0925	-262.75	-258.08
MIN5a	0.0991	-257.61	-249.93	TS29	0.0921	-270.19	-265.94
MIN5a2	0.0991	-257.93	-249.26	TS37	0.0975	-206.29	-194.53
MIN8	0.1003	-258.72	-262.68	TS43	0.0986	-287.11	-290.45
MIN12	0.1020	-466.33	-462.39	TS46	0.0931	-257.43	-253.82
MIN13	0.1006	-379.60	-380.26	TS48	0.0959	-160.95	-163.26
MIN14	0.1004	-401.79	-397.54	TS49	0.0929	-86.77	-70.72
MIN15	0.1012	-443.38	-442.05	TS50	0.0926	-100.94	-84.07
MIN16	0.1014	-426.34	-423.33	TS52	0.0944	-136.48	-119.01
MIN21b	0.0996	-419.47	-406.63	TS59	0.1011	-16.94	-31.77
MIN21d	0.0998	-404.50	-392.19	TS65	0.1010	-30.09	-43.32
MIN22	0.1015	-433.18	-434.33	TS4a-b	0.0992	-253.57	-246.45
MIN25a	0.0985	-273.16	-265.19	TS26a-b	0.0981	-280.21	-269.80
MIN25b	0.0990	-274.53	-266.76	TS25a-b	0.0982	-268.28	-260.26
MIN26a	0.0981	-282.41	-270.87	TS5a-a2	0.0976	-253.05	-243.92
MIN26b	0.0985	-287.69	-277.27	TS1ii	0.1010	-30.09	-43.32
MIN32	0.1022	-34.33	-49.46	TS6ii	0.0926	-158.58	-157.20
MIN4ii	0.1010	-342.11	-344.84	TS7ii	0.0962	-157.57	-154.42
MIN6ii	0.1008	-330.59	-328.18	TS8ii	0.0944	-144.27	-138.90
MIN7ii	0.0998	-307.80	-305.96	TS9ii	0.0957	-152.38	-150.42
MIN8ii	0.1005	-350.57	-350.84	TS10ii	0.0918	-166.25	-163.43
TS0	0.0961	-86.20	-57.43	TS11ii	0.0918	-160.17	-157.79
TS1	0.0938	-117.79	-97.05	TS12ii	0.0920	-161.68	-159.28
TS4	0.0938	-162.69	-137.83	Pr1 (1-cyanocyclopentadiene + H atom)	0.0915	-278.18	-272.02
TS11	0.0950	-164.97	-157.58	Pr2 (2-cyanocyclopentadiene + H atom)	0.0912	-271.43	-267.17
TS14	0.0940	-70.66	-55.08	Pr3 (cyclopentadiene + CN)	0.0973	-161.71	-171.87
TS17	0.0979	-187.51	-183.15	Pr4 (1-isocyanocyclopentadiene + H atom)	0.0908	-179.33	-176.17
TS18	0.0967	-251.25	-247.08	Pr5 (2-isocyanocyclopentadiene + H atom)	0.0906	-176.80	-174.77
TS20	0.0940	-216.02	-212.27	Pr10 (vinylcyanoacetylene + CH_3)	0.0905	-203.05	-176.55

[a]For the two reactants the absolute energies are reported in Hartree.

Conclusions

This thesis has presented two new chemical models for the accurate description of NCIs and weakly bonded systems: the junChS and junChS-F12 composite schemes. The junChS model was developed and validated in ref. [25], with the aim of reproducing the accurate interaction energies of the A14 data set, first reported in ref. [147]. In its final formulation, the junChS energies accounts for triple excitations thanks to the CCSD(T) method, but also incorporates the contribution due to the extrapolation to the CBS limit and the effects of core-valence correlation using the MP2 method. The basis sets employed belong to the jun-cc-pVnZ family [105], with n=T being employed for the CCSD(T) leading term and the combination n=T,Q used for the extrapolation to the CBS limit. The CV term is computed in conjuction with the cc-pwCVTZ basis set [153]. According to ref. [25], the junChS model provides CP-corrected energies that are very close to those of the A14 data set, with a relative error of 1.18%, i.e. 0.11 kJ·mol^{-1} in relative terms, using the "CBS-georef" as reference structure. This values increases to 1.38% when the NCP-revDSD geometries are employed. However, this increase in counterbalanced by a large saving in terms of computational times. Indeed, the NCP-revDSD geometries are less computationally demanding then the "CBS-georef"ones.

Shortly after, the junChS scheme was redesigned with the idea of using explicitly correlated methods. These should provide a faster convergence to the CBS limit and, thus, more accurate results at nearly the same computational cost of the junChS scheme. As detailed in ref. [26], explicitly correlated methods converge with the power -5 for the CCSD(T)-F12 model, but if the MP2-F12 method is employed, the convergence is still better described by the n^{-3} formula. Within these explicitly correlated methodologies, the best-performing basis sets are the cc-pVnZ-F12 ones together with those of the jun-cc-pVnZ family. In details, if a scheme analogous to the junChS model is employed in conjunction with the explicitly correlated methods, the relative error for the interaction energies of the A14 data set drops from 1.38% (revDSD geometries, see above) to 0.68%, which means 0.05 kJ·mol^{-1}, on average, in absolute terms. If the cc-pVnZ-F12 family of bases is used, the error is slightly larger, being of 0.09 kJ·mol^{-1} in absolute terms (relative MAE is 0.93%). To extend the statistical analysis, additional computations were carried out on top of the B9 and C6 data sets, thus providing a larger collection of systems for which the junChS and junChS-F12 energies are available and that can be used for future benchmark studies.

In ref. [26], the possibility of using the junChS and junChS-F12 models for the derivation of accurate structural parameters — fundamental in the interplay of theory and experimental rotational spectroscopy — was explored, still considering the A14 data set. Both models provide similar results for both intra- and inter-molecular parameters, which are consistently shorter than the reference geometries "CBS-georef" that do not incorporate the CV term. The performances of these two schemes in providing accurate structures were tested, in Chapter 4, for the DMS$-SO_2$ complex. In particular, the two models were used to retrieve accurate equilibrium rotational constants, which were corrected for the vibrational contributions computed using the B3LYP-D3/SNSD level of theory. The junChS scheme provides rotational constants that differ, on average, by 0.6% from the experimental values, while the error reduces to 0.2% for the junChS-F12 approach. For both models, the errors are always below 1%, thus confirming the possibility of using such approaches to (i) obtain accurate rotational constants to guide experimental studies and (ii) gain information on the geometrical parameters of the nc-complex. The latter information is fundamental to understand the electronic rearrangement occurring upon formation of the complex starting from the isolated partners.

The junChS scheme was also successfully applied to obtain accurate interaction energies of quite large nc-complexes, as detailed in refs. [43] and [44]. In these works, the trifluoroacetophenone-water and benzofuran-formaldehyde complexes have been studied by means of a joint quantum chemistry-rotational spectroscopy approach. In the first case, the role of trifluoromethylation in tuning the NCIs was investigated in comparison with the acetophenone-water complex. The work pointed out the formation of a stable seven-membered ring of TFAP with water, via the weak C(sp^2)$-$H$\cdots$O linkage and the strong C$-$H$\cdots$O H-bond. Using the SExp+TMA approach, the main intermolecular parameters were derived, with the equilibrium H$\cdots$O bond distance being of 1.969 Å. In the second case, the study aimed at understanding the possible NCIs that a heteroaromatic compound can form in biological environments, using H_2CO as model for the C=O moieties often present in biological systems. The study, carried out in the gas phase, pointed out a dominant π-π^* interaction between the benzofuran molecule and H_2CO, which is preferred with respect to other possible NCIs, such as H-bonds.

The junChS scheme was also applied to retrieve accurate reaction barriers for the reaction between CH_2NH and the CP radical [62]. This model provides energies that are in very good agreement with those obtained by the very accurate, but much more expensive HEAT-like protocol, the difference being well below 1 kJ·mol^{-1}. The CH_2NH + CP system forms E/Z-HNCHCP and CH_2NCP, which are new exotic P-bearing species, of potential astrochemical interest. The kinetic modelling of this reaction is currently under development and will aim at pointing out the must abundant products among the above mentioned. The junChS approach could not be applied to the reaction be-

tween oxirane and the CN radical because of the multireference character of some intermediates and the pathways were only investigated at the revDSD/junTZ level. Among the products located on the corresponding reactive PES, cyanoacetaldehyde and oxazole are two isomers of the C_3H_3NO family, that could be potentially present in the ISM. The kinetic modelling, at 60 K and in the low pressure limit suggests that the formation of oxazole is negligible, the rate constant being 10^{-24} $cm^3molecule^{-1}s^{-1}$. Additionally, the formation of cyanoacetaldehyde is competition with the dissociation of the system to H_2CCN and H_2CO. Indeed, for the latter products, the rate constant is in the order of 10^{-11}-10^{-12} $cm^3molecule^{-1}s^{-1}$.

Lastly, the refinement of the electronic energies for the reaction between the C_3N radical and propylene, is currently in progress. The reaction was investigated first at the B3-D3/TZ level and then refined at the B2-D3/junTZ. However, in view of applying the ChS-F12 scheme for the energetics, the CCSD(F12*)(T) energy was computed in conjunction with the cc-pVTZ-F12 basis set. The results at this last level of theory, also corrected for B3-D3/TZ harmonic ZPE, show that the products of the reaction are vinylcyanoacetylene + CH_3, 1- and 2-cyanocyclopentadiene + H, 1- and 2-isocyanocyclopentadiene + H, and cyclopentadiene + the CN radical. After the refinement of the energetics, the goal will be the kinetic simulation of the reaction, in order to unveil the most abundant product formed.

Appendices

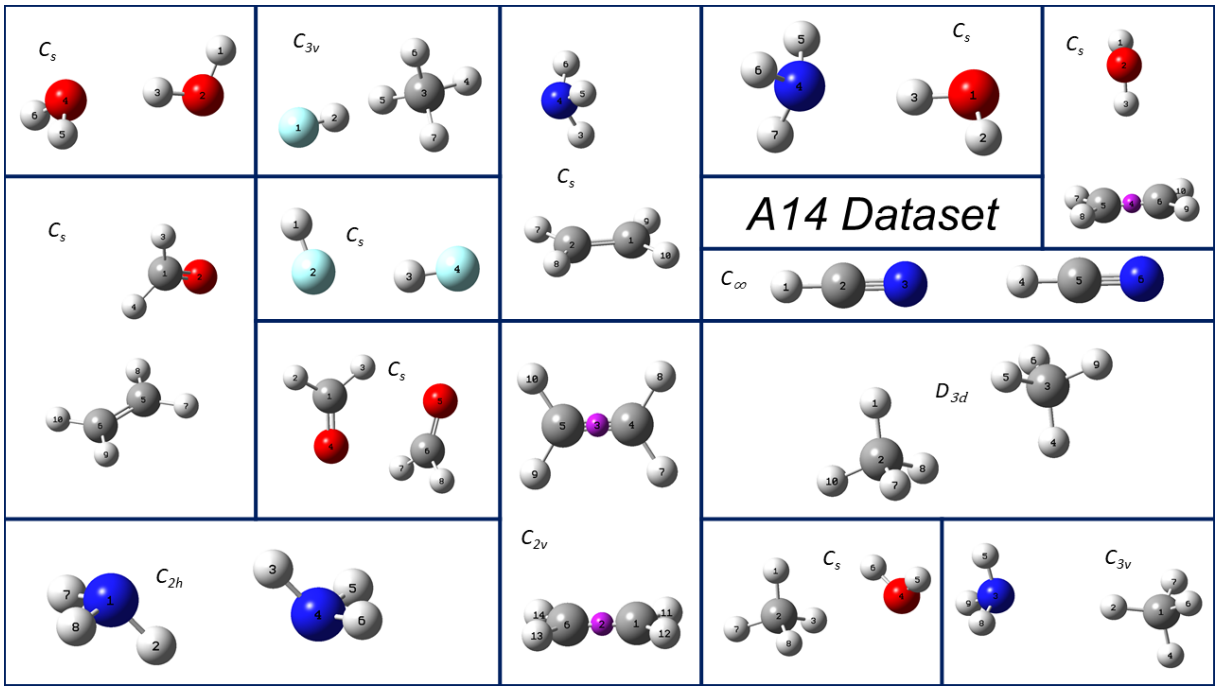
Appendix A

maugChS and mayChS energies (kJ·mol^{-1}).

Complex	"ref"	fc-CCSD(T)/ mayTZ[a]	$\Delta E^{\infty}_{\mathrm{MP2}}$/ maug(T,Q)Z	$\Delta E^{\infty}_{\mathrm{MP2}}$/ may(T,Q)Z	$\Delta E^{\mathrm{CV}}_{\mathrm{MP2}}$	Energy maugChS	Relative Error	Energy mayChS	Relative Error
H_2O-H_2O	-21.0832	-18.8584	-1.6990	-1.7405	-0.1354	-20.6928	-1.85%	-20.7343	-1.65%
NH_3-NH_3	-13.2131	-11.8607	-0.7661	-0.7364	-0.0733	-12.7001	-3.88%	-12.6704	4.11%
HF-HF	-19.2213	-16.3966	-2.1345	-2.7880	-0.1049	-18.6360	-3.05%	-19.2895	0.36%
H_2CO-H_2CO	-18.9284	-14.5364	-3.6000	-4.5578	-0.0987	-18.2351	-3.66%	-19.1929	1.40%
HCN-HCN	-19.9828	-18.4257	-1.1215	-1.3843	-0.0733	-19.6205	-1.81%	-19.8838	-0.50%
C_2H_4-C_2H_4	-4.5647	-3.1670	-1.0029	-1.6162	-0.0439	-4.2138	-7.69%	-4.8271	5.75%
CH_4-CH_4	-2.2301	-1.3146	-0.4327	-0.8652	-0.0068	-1.7541	-21.34%	-2.1866	-1.95%
H_2O-NH_3	-27.3759	-24.9552	-1.5671	-1.8348	-0.1677	-26.6900	-2.51%	-26.9577	-1.53%
H_2O-C_2H_4	-10.7696	-8.9416	-1.3045	-1.8674	-0.0875	-10.3336	-4.05%	-10.8965	1.18%
C_2H_4-H_2CO	-6.7948	-4.8713	-1.3947	-2.0543	-0.0600	-6.3260	-6.90%	-6.9856	2.81%
C_2H_4-NH_3	-5.7865	-4.5153	-0.8650	-1.3476	-0.0488	-5.4291	-6.18%	-5.9117	2.16%
HF-CH_4	-6.9162	-5.3603	-0.9313	-1.6636	-0.0860	-6.3776	-7.79%	-7.1099	2.80%
H_2O-CH_4	-2.8242	-2.0514	-0.3803	-0.6405	-0.0286	-2.4603	-12.89%	-2.7205	3.67%
NH_3-CH_4	-3.2175	-2.4662	-0.3387	-0.6360	-0.0352	-2.8401	-11.73%	-3.1374	-2.49%
MAE							6.81%		2.40%

[a] The may- and maug-cc-pVTZ basis sets are equivalent.

Appendix B



Atom Labelling for the nc-complex for the A14 data set.

Acronyms

ABS auxiliary basis set. 18

ACM Adiabatic Connection Model. 20

AOs Atomic Orbitals. 12

BO Born-Oppenheimer. 12

BSE Basis Set Error. 13

BSSE Basis Set Superposition Error. 36

CABS complementary ABS. 18

CBS Complete Basis Set. 25

CC Coupled Cluster. 11

ccCA Correlation consistent composite approach. 29

COBRA COaxially Aligned Beam Resonator Arrangement). 58

CRESU Cinétique de Réaction en Ecoulement Supersonique Uniforme. 8

CV Core-Valence. 28

DF Density Fitting. 18

DFT Density Functional Theory. 11

EBC extended Brillouin condition. 18

GGA General Gradient Approximation. 19

GTOs Gaussian-type orbitals. 22

HF Hartree-Fock. 13

IMPACT In-phase/quadrature-phase-Modulation Passage- Acquired-Coherence Technique. 58

IRC Intrinsic Reaction Coordinate. 77

ISM Interstellar Medium. 5

IUPAC International Union of Pure and Applied Chemistry. 57

KS Kohn-Sham. 12

LCAO Linear Combination of AOs. 21

LDA Local Density Approximation. 19

LSDA Local Spin Density Approximation. 19

MAE Mean Absolute Error. 40

MOs Molecular Orbitals. 12

MP Møller-Plesset. 14

NBO Natural Bond Orbital. 60

nc non-covalent. 2

NCIs non-covalent Interactions. 1

NOCV/CD Natural Orbital for Chemical Valence/Charge Displacement. 60

PES Potential Energy Surface. 4

PT Perturbation Theory. 14

RCs Rotational Constants. 4

RI Resolution of the Identity. 17, 18

RRHO Rigid-rotor/Harmonic-oscillator. 30

RSPT Rayleigh-Schödinger PT. 31

SCF Self-Consistent Field. 13

SE Sch-rödinger equation. 12

SExp semi-experimental. 4

STOs Slater-type orbitals. 22

vdW van der Waals. 1

VWN Vosko, Wilk and Nusair. 20

Wn Weizmann-n. 29

Bibliography

[1] P. Hobza and K. Müller-Dethlefs, *Non-Covalent Interactions*, ser. Theoretical and Computational Chemistry Series. The Royal Society of Chemistry, 2009.

[2] P. Hobza, R. Zahradník, and K. Müller-Dethlefs, "The world of non-covalent interactions: 2006," *Collect. Czechoslov. Chem. Commun.*, vol. 71, pp. 443–531, 2006.

[3] A. S. Mahadevi and G. N. Sastry, "Cooperativity in noncovalent interactions," *Chem. Rev.*, vol. 116, pp. 2775–2825, 2016.

[4] K. Autumn, M. Sitti, Y. A. Liang, A. M. Peattie, W. R. Hansen, S. Sponberg, T. W. Kenny, R. Fearing, J. N. Israelachvili, and R. J. Full, "Evidence for van der Waals adhesion in gecko setae," *Proc. Natl. Acad. Sci. USA*, vol. 99, pp. 12 252–12 256, 2002.

[5] K. Autumn, Y. A. Liang, S. T. Hsieh, W. Zesch, W. P. Chan, T. W. Kenny, R. Fearing, and R. J. Full, "Adhesive force of a single gecko foot-hair," *Nature*, vol. 405, pp. 681–685, 2000.

[6] W. Wang, Y. Liu, and Z. Xie, "Gecko-like dry adhesive surfaces and their applications: a review," *J. Bionic Eng.*, vol. 18, pp. 1011–1044, 2021.

[7] V. F. Petrenko and R. W. Whitworth, *Physics of ice*. Oxford University Press, Oxford, 2002.

[8] E. Brini, C. J. Fennell, M. Fernandez-Serra, B. Hribar-Lee, M. Luksic, and K. A. Dill, "How water's properties are encoded in its molecular structure and energies," *Chem. Rev.*, vol. 117, pp. 12 385–12 414, 2017.

[9] R. W. Newberry and R. T. Raines, "Secondary forces in protein folding," *ACS Chem. Biol.*, vol. 14, pp. 1677–1686, 2019.

[10] M. B. Hillyer and B. C. Gibb, "Molecular shape and the hydrophobic effect," *Annu. Rev. Phys. Chem.*, vol. 67, pp. 307–329, 2016.

[11] J.-M. Lehn, "Perspectives in supramolecular chemistry—from molecular recognition towards molecular information processing and self-organization," *Angew. Chem. Int. Ed.*, vol. 29, pp. 1304–1319, 1990.

[12] J. J. Novoa, Ed., *Intermolecular Interactions in Crystals*. The Royal Society of Chemistry, Croydon, UK, 2018.

[13] R. R. Knowles and E. N. Jacobsen, "Attractive noncovalent interactions in asymmetric catalysis: Links between enzymes and small molecule catalysts," *Proc. Natl. Acad. Sci. USA*, vol. 107, pp. 20 678–20 685, 2010.

[14] K. T. Mahmudov, A. V. Gurbanov, F. I. Guseinov, and M. F. C. G. da Silva, "Noncovalent interactions in metal complex catalysis," *Coord. Chem. Rev.*, vol. 387, pp. 32–46, 2019.

[15] W. Tang, S. Johnston, J. A. Iggo, N. G. Berry, M. Phelan, L. Lian, J. Bacsa, and J. Xiao, "Cooperative catalysis through noncovalent interactions," *Angew. Chem. Int. Ed.*, vol. 125, pp. 1712–1716, 2013.

[16] V. Georgakilas, J. N. Tiwari, K. C. Kemp, J. A. Perman, A. B. Bourlinos, K. S. Kim, and R. Zboril, "Noncovalent functionalization of graphene and graphene oxide for energy materials, biosensing, catalytic, and biomedical applications," *Chem. Rev.*, vol. 116, pp. 5464–5519, 2016.

[17] A. C. Coxon, D. A. Laidler, R. B. Pettman, and J. F. Stoddart, "Stereochemistry of noncovalent interactions in organic and metal cationic complexes," *J. Am. Chem. Soc.*, vol. 100, pp. 8260–8262, 1978.

[18] Y. Dangat, S. Popli, and R. B. Sunoj, "Unraveling the importance of noncovalent interactions in asymmetric hydroformylation reactions," *J. Am. Chem. Soc.*, vol. 142, pp. 17 079–17 092, 2020.

[19] C. Puzzarini, L. Spada, S. Alessandrini, and V. Barone, "The challenge of noncovalent interactions: theory meets experiment for reconciling accuracy and interpretation," *J. Phys. Condens. Matter*, vol. 32, p. 343002, 2020.

[20] W. Caminati and J.-U. Grabow, "Chapter 17 - Advancements in microwave spectroscopy," in *Frontiers and Advances in Molecular Spectroscopy*, J. Laane, Ed. Elsevier, 2018, pp. 569–598.

[21] H. Fenniri, P. Mathivanan, K. L. Vidale, D. M. Sherman, K. Hallenga, K. V. Wood, and J. G. Stowell, "Helical rosette nanotubes: Design, self-assembly, and characterization," *J. Am. Chem. Soc.*, vol. 123, pp. 3854–3855, 2001.

[22] T. B. Becher, C. B. Braga, D. L. Bertuzzi, M. D. Ramos, A. Hassan, F. N. Crespilho, and C. Ornelas, "The structure–property relationship in LAPONITE® materials: from wigner glasses to strong self-healing hydrogels formed by non-covalent interactions," *Soft Matter*, vol. 15, pp. 1278–1289, 2019.

[23] J. Solà and I. Alfonso, *Non-covalent Interactions for the Preparation of Pseudopeptidic Synthetic Compounds and Materials*. John Wiley & Sons, Ltd, 2016, ch. 22, pp. 391–412.

[24] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, and W. Yang, "Revealing noncovalent interactions," *J. Am. Chem. Soc.*, vol. 132, pp. 6498–6506, 2010.

[25] S. Alessandrini, V. Barone, and C. Puzzarini, "Extension of the "cheap" composite approach to noncovalent interactions: The jun-ChS scheme," *J. Chem. Theory Comput.*, vol. 16, pp. 988–1006, 2019.

[26] J. Lupi, S. Alessandrini, C. Puzzarini, and V. Barone, "junChS and junChS-F12 Models: Parameter-free efficient yet accurate composite schemes for energies and structures of noncovalent complexes," *J. Chem. Theory Comput.*, pp. 6974–6992, 2021.

[27] C. Puzzarini and V. Barone, "Extending the molecular size in accurate quantum-chemical calculations: the equilibrium structure and spectroscopic properties of uracil," *Phys. Chem. Chem. Phys.*, vol. 13, pp. 7189–7197, 2011.

[28] C. Puzzarini, M. Biczysko, V. Barone, L. Largo, I. Peña, C. Cabezas, and J. L. Alonso, "Accurate characterization of the peptide linkage in the gas phase: A joint quantum-chemical and rotational spectroscopy study of the glycine dipeptide analogue," *J. Phys. Chem. Lett.*, vol. 5, pp. 534–540, 2014.

[29] C. Puzzarini, M. Biczysko, V. Barone, I. Pena, C. Cabezas, and J. L. Alonso, "Accurate molecular structure and spectroscopic properties of nucleobases: a combined computational–microwave investigation of 2-thiouracil as a case study," *Phys. Chem. Chem. Phys.*, vol. 15, pp. 16 965–16 975, 2013.

[30] E. Papajak, H. R. Leverentz, J. Zheng, and D. G. Truhlar, "Efficient diffuse basis sets: cc-pV*x*Z+ and maug-cc-pV*x*Z," *J. Chem. Theory Comput.*, vol. 5, pp. 1197–1202, 2009.

[31] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*, ser. Cambridge Molecular Science. Cambridge University Press, Cambridge, 2009.

[32] A. Grüneis, S. Hirata, Y.-y. Ohnishi, and S. Ten-No, "Perspective: Explicitly correlated electronic structure theory for complex systems," *J. Chem. Phys.*, vol. 146, p. 080901, 2017.

[33] W. Gordy and R. L. Cook, *Microwave molecular spectra*, 2nd ed. Wiley, New York, 1984.

[34] W. Caminati, "Microwave spectroscopy of large molecules and molecular complexes," *Handbook of High-resolution Spectroscopy*, 2011.

[35] V. Barone, S. Alessandrini, M. Biczysko, J. R. Cheeseman, D. C. Clary, A. B. McCoy, R. J. DiRisio, F. Neese, M. Melosso, and C. Puzzarini, "Computational molecular spectroscopy," *Nat. Rev. Methods Primers*, vol. 1, pp. 1–27, 2021.

[36] M. A. J. Watson, "Higer-order effects in the vibration-rotation spectra of semirigid molecules." *Molecular Spectroscopy: Modern Research*, vol. 3, p. 1, 2012.

[37] C. Puzzarini, J. F. Stanton, and J. Gauss, "Quantum-chemical calculation of spectroscopic parameters for rotational spectroscopy," *Int. Rev. Phys. Chem.*, vol. 29, pp. 273–367, 2010.

[38] I. Peña, M. E. Sanz, J. C. López, and J. L. Alonso, "Preferred conformers of proteinogenic glutamic acid," *J. Am. Chem. Soc.*, vol. 134, pp. 2305–2312, 2012.

[39] C. Puzzarini and V. Barone, "Diving for accurate structures in the ocean of molecular systems with the help of spectroscopy and quantum chemistry," *Acc. Chem. Res.*, vol. 51, pp. 548–556, 2018.

[40] J. Demaison, "Experimental, semi-experimental and ab initio equilibrium structures," *Mol. Phys.*, vol. 105, pp. 3109–3138, 2007.

[41] M. Piccardo, E. Penocchio, C. Puzzarini, M. Biczysko, and V. Barone, "Semi-experimental equilibrium structure determinations by employing B3LYP/SNSD anharmonic force fields: Validation and application to semirigid organic molecules," *J. Phys. Chem. A*, vol. 119, pp. 2058–2082, 2015.

[42] D. A. Obenchain, L. Spada, S. Alessandrini, S. Rampino, S. Herbers, N. Tasinato, M. Mendolicchio, P. Kraus, J. Gauss, C. Puzzarini *et al.*, "Unveiling the sulfur–sulfur bridge: Accurate structural and energetic characterization of a homochalcogen intermolecular bond," *Angew. Chem. Int. Ed.*, vol. 57, pp. 15 822–15 826, 2018.

[43] J. Lei, S. Alessandrini, J. Chen, Y. Zheng, L. Spada, Q. Gou, C. Puzzarini, and V. Barone, "Rotational spectroscopy meets quantum chemistry for analyzing substituent effects on non-covalent interactions: The case of the trifluoroacetophenone-water complex," *Molecules*, vol. 25, p. 4899, 2020.

[44] X. Li, L. Spada, S. Alessandrini, Y. Zheng, K. G. Lengsfeld, J.-U. Grabow, G. Feng, C. Puzzarini, and V. Barone, "Stacked but not stuck: Unveiling the role of $\pi \rightarrow \pi^*$ interactions with the help of the benzofuran–formaldehyde complex," *Angew. Chem. Int. Ed.*, vol. 61, p. e202113737, 2022.

[45] S. Yamamoto, *Introduction to Astrochemistry: Chemical Evolution from Interstellar Clouds to Star and Planet Formation*. Springer, Tokyo, 2017.

[46] A. M. Shaw, *Astrochemistry: From astronomy to astrobiology*, 1st ed. John Wiley & Sons, 2007.

[47] B. A. McGuire, R. A. Loomis, A. M. Burkhardt, K. L. K. Lee, C. N. Shingledecker, S. B. Charnley, I. R. Cooke, M. A. Cordiner, E. Herbst, S. Kalenskii, M. A. Siebert, E. R. Willis, C. Xue, A. J. Remijan, and M. C. McCarthy, "Detection of two interstellar polycyclic aromatic hydrocarbons via spectral matched filtering," *Science*, vol. 371, pp. 1265–1269, 2021.

[48] B. A. McGuire, P. B. Carroll, R. A. Loomis, I. A. Finneran, P. R. Jewell, A. J. Remijan, and G. A. Blake, "Discovery of the interstellar chiral molecule propylene oxide (ch3chch2o)," *Science*, vol. 352, pp. 1449–1452, 2016.

[49] S. Zeng, D. Quénard, I. Jiménez-Serra, J. Martín-Pintado, V. Rivilla *et al.*, "First detection of the pre-biotic molecule glycolonitrile ($HOCH_2CN$) in the interstellar medium," *Mon. Not. R. Astron. Soc. Lett.*, vol. 484, pp. L43–L48, 2019.

[50] V. M. Rivilla, I. Jiménez-Serra, J. Martín-Pintado, C. Briones, L. F. Rodríguez-Almeida, F. Rico-Villas, B. Tercero, S. Zeng, L. Colzi, P. de Vicente, S. Martín, and M. A. Requena-Torres, "Discovery in space of ethanolamine, the simplest phospholipid head group," *Proc. Natl. Acad. Sci.*, vol. 118, 2021.

[51] M. Melosso, L. Bizzocchi, H. Gazzeh, F. Tonolo, J.-C. Guillemin, S. Alessandrini, V. M. Rivilla, L. Dore, V. Barone, and C. Puzzarini, "Gas-phase identification of (Z)-1,2-ethenediol, a key prebiotic intermediate in the formose reaction," *Chem. Commun.*, vol. 58, pp. 2750–2753, 2022.

[52] V. M. Rivilla, L. Colzi, I. Jiménez-Serra, J. Martín-Pintado, A. Megías, M. Melosso, L. Bizzocchi, Á. López-Gallifa, A. Martínez-Henares, S. Massalkhi *et al.*, "Precursors of the rna-world in space: Detection of (z)-1, 2-ethenediol in the interstellar medium, a key intermediate in sugar formation," *arXiv preprint arXiv:2203.14728*, 2022.

[53] C. Puzzarini and V. Barone, "Interstellar complex organic molecules: A step toward biomolecule building blocks in the skies," in *Prebiotic Photochemistry*, 2021, pp. 195–218.

[54] K. Ruiz-Mirazo, C. Briones, and A. de la Escosura, "Prebiotic systems chemistry: new perspectives for the origins of life," *Chem. Rev.*, vol. 114, pp. 285–366, 2014.

[55] N. Kitadai and S. Maruyama, "Origins of building blocks of life: A review," *Geosci. Front.*, vol. 9, pp. 1117–1153, 2018.

[56] A. G. G. M. Tielens, "The molecular universe," *Rev. Mod. Phys.*, vol. 85, pp. 1021–1081, 2013.

[57] M. Biczysko, J. Bloino, and C. Puzzarini, "Computational challenges in astrochemistry," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 8, p. e1349, 2018.

[58] J. Cernicharo, M. Guélin, M. Agúndez, M. C. McCarthy, and P. Thaddeus, "Detection of C_5N^- and vibrationally excited C_6H in IRC+10216," *Astrophys. J.*, vol. 688, pp. L83–L86, 2008.

[59] C. Puzzarini and V. Barone, "A never-ending story in the sky: The secrets of chemical evolution," *Phys. Life Rev.*, vol. 32, pp. 59–94, 2020.

[60] J. Lupi, C. Puzzarini, and V. Barone, "Methanimine as a key precursor of imines in the interstellar medium: The case of propargylimine," *Astrophys. J.*, vol. 903, p. L35, 2020.

[61] V. Wakelam, I. Smith, E. Herbst, J. Troe, W. Geppert, H. Linnartz, K. Öberg, E. Roueff, M. Agúndez, P. Pernot *et al.*, "Reaction networks for interstellar chemical modelling: improvements and challenges," *Space Sci. Rev.*, vol. 156, pp. 13–72, 2010.

[62] S. Alessandrini, F. Tonolo, and C. Puzzarini, "In search of phosphorus in astronomical environments: The reaction between the CP radical ($X^2\Sigma^+$) and methanimine," *J. Chem. Phys.*, p. 054306, 2021.

[63] S. Alessandrini and M. Melosso, "Fate of the gas-phase reaction between oxirane and the CN radical in interstellar conditions," *Front. Astron. Space Sci.*, vol. 8, p. 159, 2021.

[64] I. R. Sims and I. W. M. Smith, "Gas-phase reactions and energy transfer at very low temperatures," *Annu. Rev. Phys. Chem.*, vol. 46, pp. 109–138, 1995.

[65] I. W. M. Smith and B. R. Rowe, "Reaction kinetics at very low temperatures: Laboratory studies and interstellar chemistry," *Acc. Chem. Res.*, vol. 33, pp. 261–268, 2000.

[66] A. Potapov, A. Canosa, E. Jiménez, and B. Rowe, "Uniform supersonic chemical reactors: 30-years of astrochemical history and future challenges," *Angew. Chem. Int. Ed.*, vol. 56, pp. 8618–8640, 2017.

[67] P. Casavecchia, "Chemical reaction dynamics with molecular beams," *Rep. Prog. Phys.*, vol. 63, p. 355, 2000.

[68] P. Recio, D. Marchione, A. Caracciolo, V. J. Murray, L. Mancini, M. Rosi, P. Casavecchia, and N. Balucani, "A crossed molecular beam investigation of the N(^{2}D) + pyridine reaction and implications for prebiotic chemistry," *Chem. Phys. Lett.*, p. 138852, 2021.

[69] C. C. J. Roothaan, "New developments in molecular orbital theory," *Rev. Mod. Phys.*, vol. 23, p. 69, 1951.

[70] A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.

[71] F. Jensen, *Introduction to computational chemistry*. John wiley & sons, 2017.

[72] C. Møller and M. S. Plesset, "Note on an approximation treatment for many-electron systems," *Phys. Rev.*, vol. 46, p. 618, 1934.

[73] R. J. Bartlett, "Many-body perturbation theory and coupled cluster theory for electron correlation in molecules," *Annu. Rev. Phys. Chem.*, vol. 32, pp. 359–401, 1981.

[74] R. J. Bartlett and G. D. Purvis, "Many-body perturbation theory, coupled-pair many-electron theory, and the importance of quadruple excitations for the correlation problem," *Int. Rev. Phys. Chem.*, vol. 14, pp. 561–581, 1978.

[75] G. D. Purvis III and R. J. Bartlett, "A full coupled-cluster singles and doubles model: The inclusion of disconnected triples," *J. Chem. Phys.*, vol. 76, pp. 1910–1918, 1982.

[76] J. Noga and R. J. Bartlett, "The full CCSDT model for molecular electronic structure," *J. Chem. Phys.*, vol. 86, pp. 7041–7050, 1987.

[77] G. E. Scuseria and H. F. Schaefer III, "A new implementation of the full CCSDT model for molecular electronic structure," *Chem. Phys. Let.*, vol. 152, pp. 382–386, 1988.

[78] S. A. Kucharski and R. J. Bartlett, "Recursive intermediate factorization and complete computational linearization of the coupled-cluster single, double, triple, and quadruple excitation equations," *Theor. Chim. Acta*, vol. 80, pp. 387–405, 1991.

[79] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular electronic-structure theory*. John Wiley & Sons, 2014.

[80] J. F. Stanton, "Why CCSD(T) works: a different perspective," *Chem. Phys. Lett.*, vol. 281, pp. 130–134, 1997.

[81] C. Hattig, W. Klopper, A. Kohn, and D. P. Tew, "Explicitly correlated electrons in molecules," *Chem. Rev.*, vol. 112, pp. 4–74, 2012.

[82] G. Knizia, T. B. Adler, and H.-J. Werner, "Simplified CCSD(T)-F12 methods: Theory and benchmarks," *J. Chem. Phys.*, vol. 130, p. 054104, 2009.

[83] H.-J. Werner, T. B. Adler, and F. R. Manby, "General orbital invariant MP2-F12 theory," *J. Chem. Phys.*, vol. 126, p. 164102, 2007.

[84] C. Hättig, D. P. Tew, and A. Köhn, "Communications: Accurate and efficient approximations to explicitly correlated coupled-cluster singles and doubles, CCSD-F12," *J. Chem. Phys.*, vol. 132, p. 231102, 2010.

[85] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, vol. 136, p. B864, 1964.

[86] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.*, vol. 140, p. A1133, 1965.

[87] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Phys. Rev. A*, vol. 38, p. 3098, 1988.

[88] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B*, vol. 37, p. 785, 1988.

[89] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.*, vol. 77, p. 3865, 1996.

[90] A. D. Becke, "A new mixing of Hartree-Fock and local density-functional theories," *J. Chem. Phys.*, vol. 98, pp. 1372–1377, 1993.

[91] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas,

J. B. Foresman, and D. J. Fox, "Gaussian~16 Revision C.01," 2016, gaussian Inc. Wallingford CT.

[92] S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis," *Can. J. Phys.*, vol. 58, pp. 1200–1211, 1980.

[93] S. Grimme, "Semiempirical hybrid density functional with perturbative second-order correlation," *J. Chem. Phys.*, vol. 124, p. 034108, 2006.

[94] G. Santra, N. Sylvetsky, and J. M. Martin, "Minimally empirical double-hybrid functionals trained against the gmtkn55 database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4," *J. Phys. Chem. A*, vol. 123, pp. 5129–5143, 2019.

[95] S. Kozuch and J. M. Martin, "DSD-PBEP86: in search of the best double-hybrid dft with spin-component scaled MP2 and dispersion corrections," *Phys. Chem. Chem. Phys.*, vol. 13, pp. 20 104–20 107, 2011.

[96] ——, "Spin-component-scaled double hybrids: an extensive search for the best fifth-rung functionals blending DFT and perturbation theory," *J. Comput. Chem.*, vol. 34, pp. 2327–2344, 2013.

[97] T. Schwabe and S. Grimme, "Double-hybrid density functionals with long-range dispersion corrections: higher accuracy and extended applicability," *Phys. Chem. Chem. Phys.*, vol. 9, pp. 3397–3406, 2007.

[98] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," *J. Chem. Phys.*, vol. 132, p. 154104, 2010.

[99] S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," *J. Comput. Chem.*, vol. 32, pp. 1456–1465, 2011.

[100] S. F. Boys and M. V. Wilkes, "The integral formulae for the variational solution of the molecular many-electron wave equation in terms of gaussian functions with direct electronic correlation," *Proc. R. Soc. A*, vol. 258, pp. 402–411, 1960.

[101] W. J. Hehre, R. F. Stewart, and J. A. Pople, "Self-consistent molecular-orbital methods. I. use of gaussian expansions of slater-type atomic orbitals," *J. Chem. Phys.*, vol. 51, pp. 2657–2664, 1969.

[102] T. H. Dunning Jr, "Gaussian basis sets for use in correlated molecular calculations. I. the atoms boron through neon and hydrogen," *J. Chem. Phys.*, vol. 90, pp. 1007–1023, 1989.

[103] D. E. Woon and T. H. Dunning Jr, "Gaussian basis sets for use in correlated molecular calculations. V. core-valence basis sets for boron through neon," *J. Chem. Phys.*, vol. 103, pp. 4572–4585, 1995.

[104] R. A. Kendall, T. H. Dunning Jr, and R. J. Harrison, "Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions," *J. Chem. Phys.*, vol. 96, pp. 6796–6806, 1992.

[105] E. Papajak, J. Zheng, X. Xu, H. R. Leverentz, and D. G. Truhlar, "Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions," *J. Chem. Theory Comput.*, vol. 7, pp. 3027–3034, 2011.

[106] K. A. Peterson, T. B. Adler, and H.-J. Werner, "Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, B-Ne, and Al-Ar," *J. Chem. Phys.*, vol. 128, p. 084102, 2008.

[107] D. Feller, "The use of systematic sequences of wave functions for estimating the complete basis set, full configuration interaction limit in water," *J. Phys. Chem.*, vol. 98, pp. 7059–7071, 1993.

[108] T. Helgaker, W. Klopper, H. Koch, and J. Noga, "Basis-set convergence of correlated calculations on water," *J. Chem. Phys.*, vol. 106, p. 9639, 1997.

[109] A. Varandas, "Basis-set extrapolation of the correlation energy," *J. Chem. Phys.*, vol. 113, pp. 8880–8887, 2000.

[110] A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, and J. F. Stanton, "HEAT: High accuracy extrapolated ab initio thermochemistry," *J. Chem. Phys.*, vol. 121, pp. 11 599–11 613, 2004.

[111] M. Heckert, M. Kállay, and J. Gauss, "Molecular equilibrium geometries based on coupled-cluster calculations including quadruple excitations," *Mol. Phys.*, vol. 103, pp. 2109–2115, 2005.

[112] M. Heckert, M. Kállay, D. P. Tew, W. Klopper, and J. Gauss, "Basis-set extrapolation techniques for the accurate calculation of molecular equilibrium geometries using coupled-cluster theory," *J. Chem. Phys.*, vol. 125, p. 044108, 2006.

[113] C. Puzzarini and V. Barone, "The challenging playground of astrochemistry: an integrated rotational spectroscopy – quantum chemistry strategy," *Phys. Chem. Chem. Phys.*, vol. 22, pp. 6507–6523, 2020.

[114] Y. J. Bomble, J. F. Stanton, M. Kállay, and J. Gauss, "Coupled-cluster methods including noniterative corrections for quadruple excitations," *J. Chem. Phys.*, vol. 123, p. 054101, 2005.

[115] M. Kállay and J. Gauss, "Approximate treatment of higher excitations in coupled-cluster theory," *J. Chem. Phys.*, vol. 123, p. 214105, 2005.

[116] ——, "Approximate treatment of higher excitations in coupled-cluster theory. II. Extension to general single-determinant reference functions and improved approaches for the canonical Hartree–Fock case," *J. Chem. Phys.*, vol. 129, p. 144101, 2008.

[117] Y. J. Bomble, J. Vázquez, M. Kállay, C. Michauk, P. G. Szalay, A. G. Császár, J. Gauss, and J. F. Stanton, "High-accuracy extrapolated ab initio thermochemistry. II. Minor improvements to the protocol and a vital simplification," *J. Chem. Phys.*, vol. 125, p. 064108, 2006.

[118] M. E. Harding, J. Vázquez, B. Ruscic, A. K. Wilson, J. Gauss, and J. F. Stanton, "High-accuracy extrapolated ab initio thermochemistry. III. Additional improvements and overview," *J. Chem. Phys.*, vol. 128, p. 114111, 2008.

[119] J. Lupi, C. Puzzarini, C. Cavallotti, and V. Barone, "State-of-the-art quantum chemistry meets variable reaction coordinate transition state theory to solve the puzzling case of the H_2S + Cl system," *J. Chem. Theory Comput.*, vol. 16, pp. 5090–5104, 2020.

[120] N. C. Handy, Y. Yamaguchi, and H. F. Schaefer III, "The diagonal correction to the born–oppenheimer approximation: its effect on the singlet–triplet splitting of ch2 and other molecular effects," *J. Chem. Phys.*, vol. 84, pp. 4481–4484, 1986.

[121] N. C. Handy and A. M. Lee, "The adiabatic approximation," *Chem. Phys. Lett.*, vol. 252, pp. 425–430, 1996.

[122] H. Sellers and P. Pulay, "The adiabatic correction to molecular potential surfaces in the scf approximation," *Chem. Phys. Lett.*, vol. 103, pp. 463–465, 1984.

[123] R. D. Cowan and D. C. Griffin, "Approximate relativistic corrections to atomic radial wave functions," *J. Opt. Soc. Am.*, vol. 66, pp. 1010–1014, 1976.

[124] S. Coriani, D. Marchesan, J. Gauss, C. Hättig, T. Helgaker, and P. Jørgensen, "The accuracy of ab initio molecular geometries for systems containing second-row atoms," *J. Chem. Phys.*, vol. 123, p. 184107, 2005.

[125] S. Alessandrini, J. Gauss, and C. Puzzarini, "Accuracy of rotational parameters predicted by high-level quantum-chemical calculations: Case study of sulfur-containing molecules of astrochemical interest," *J. Chem. Theory Comput.*, vol. 14, pp. 5360–5371, 2018.

[126] C. Puzzarini, M. Heckert, and J. Gauss, "The accuracy of rotational constants predicted by high-level quantum-chemical calculations. I. molecules containing first-row atoms," *J. Chem. Phys.*, vol. 128, p. 194108, 2008.

[127] N. Sylvetsky, K. A. Peterson, A. Karton, and J. M. Martin, "Toward a W4-F12 approach: Can explicitly correlated and orbital-based ab initio CCSD(T) limits be reconciled?" *J. Chem. Phys.*, vol. 144, p. 214101, 2016.

[128] C. Puzzarini, M. Biczysko, and V. Barone, "Accurate anharmonic vibrational frequencies for uracil: the performance of composite schemes and hybrid CC/DFT model," *J. Chem. Theory Comput.*, vol. 7, pp. 3702–3710, 2011.

[129] A. Karton and J. M. Martin, "Performance of W4 theory for spectroscopic constants and electrical properties of small molecules," *J. Chem. Phys.*, vol. 133, p. 144102, 2010.

[130] A. Karton, S. Daon, and J. M. Martin, "W4-11: a high-confidence benchmark dataset for computational thermochemistry derived from first-principles w4 data," *Chem. Phys. Lett.*, vol. 510, pp. 165–178, 2011.

[131] A. D. Boese, M. Oren, O. Atasoylu, J. M. Martin, M. Kállay, and J. Gauss, "W3 theory: Robust computational thermochemistry in the kj/mol accuracy range," *J. Chem. Phys.*, vol. 120, pp. 4129–4141, 2004.

[132] J. M. Martin and G. de Oliveira, "Towards standard methods for benchmark quality ab initio thermochemistry—W1 and W2 theory," *J. Chem. Phys.*, vol. 111, pp. 1843–1856, 1999.

[133] A. Karton, E. Rabinovich, J. M. Martin, and B. Ruscic, "W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions," *J. Chem. Phys.*, vol. 125, p. 144108, 2006.

[134] N. J. DeYonker, T. R. Cundari, and A. K. Wilson, "The correlation consistent composite approach (ccCA): An alternative to the gaussian-n methods," *J. Chem. Phys.*, vol. 124, p. 114104, 2006.

[135] N. J. DeYonker, B. Mintz, T. R. Cundari, and A. K. Wilson, "Application of the correlation consistent composite approach (ccCA) to third-row (ga- kr) molecules," *J. Chem. Theory Comput.*, vol. 4, pp. 328–334, 2008.

[136] C. Puzzarini, M. Biczysko, V. Barone, L. Largo, I. Pena, C. Cabezas, and J. L. Alonso, "Accurate characterization of the peptide linkage in the gas phase: A joint quantum-chemical and rotational spectroscopy study of the glycine dipeptide analogue," *J. Phys. Chem. Lett.*, vol. 5, pp. 534–540, 2014.

[137] C. Puzzarini, "Accurate molecular structures of small-and medium-sized molecules," *Int. J. Quantum Chem.*, vol. 116, pp. 1513–1519, 2016.

[138] H. W. Kroto, "Molecular rotation spectra," *Molecular rotation spectra, Dover*, 1992.

[139] V. Barone, "Vibrational zero-point energies and thermodynamic functions beyond the harmonic approximation," *J. Chem. Phys.*, vol. 120, pp. 3059–3065, 2004.

[140] ——, "Anharmonic vibrational properties by a fully automated second-order perturbative approach," *J. Chem. Phys.*, vol. 122, p. 014108, 2005.

[141] R. L. Cook and F. C. De Lucia, "Application of the Theory of Irreducible Tensor Operators to Molecular Hyperfine Structure," *Am. J. Phys.*, vol. 39, pp. 1433–1454, 1971.

[142] G. Cazzoli and C. Puzzarini, "The Lamb-dip spectrum of the J+1←J (J=0, 1, 3–8) transitions of $H^{13}CN$: The nuclear hyperfine structure due to H, ^{13}C, and ^{14}N," *J. Mol. Spectrosc.*, vol. 233, pp. 280–289, 2005.

[143] C. Puzzarini, J. Bloino, N. Tasinato, and V. Barone, "Accuracy and interpretability: the devil and the holy grail. new routes across old boundaries in computational spectroscopy," *Chem. Rev.*, vol. 119, pp. 8131–8191, 2019.

[144] D. A. Matthews, L. Cheng, M. E. Harding, F. Lipparini, S. Stopkowicz, T.-C. Jagau, P. G. Szalay, J. Gauss, and J. F. Stanton, "Coupled-cluster techniques for computational chemistry: The CFOUR program package," *J. Chem. Phys.*, vol. 152, p. 214108, 2020.

[145] J. F. Stanton, J. Gauss, L. Cheng, M. E. Harding, D. A. Matthews, and P. G. Szalay, "CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package," With contributions from A.A. Auer, R.J. Bartlett, U. Benedikt, C. Berger, D.E. Bernholdt, Y.J. Bomble, O. Christiansen, F. Engel, R. Faber, M. Heckert, O. Heun, M. Hilgenberg, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, T. Kirsch, K. Klein, W.J. Lauderdale, F. Lipparini, T. Metzroth, L.A. Mück, D.P. O'Neill, D.R. Price, E. Prochnow, C. Puzzarini, K. Ruud, F. Schiffmann, W. Schwalbach, C. Simmons, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J.D. Watts and the integral packages MOLECULE (J. Almlöf and P.R. Taylor), PROPS (P.R. Taylor), ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de.

[146] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, "Molpro: a general-purpose quantum chemistry program package," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 2, pp. 242–253, 2012.

[147] J. Rezac and P. Hobza, "Describing noncovalent interactions beyond the common approximations: how accurate is the "gold standard", CCSD(T) at the complete basis set limit?" *J. Chem. Theory Comput.*, vol. 9, pp. 2151–2155, 2013.

[148] J. Řezáč, M. Dubeckỳ, P. Jurečka, and P. Hobza, "Extensions and applications of the A24 data set of accurate interaction energies," *Phys. Chem. Chem. Phys.*, vol. 17, pp. 19 268–19 277, 2015.

[149] L. A. Burns, M. S. Marshall, and C. D. Sherrill, "Comparing counterpoise-corrected, uncorrected, and averaged binding energies for benchmarking noncovalent interactions," *J. Chem. Theory Comput.*, vol. 10, pp. 49–57, 2014.

[150] P. Hobza and J. Rezac, "Introduction: noncovalent interactions," *Chem. Rev.*, vol. 116, pp. 4911–4912, 2016.

[151] W. Li, L. Spada, N. Tasinato, S. Rampino, L. Evangelisti, A. Gualandi, P. G. Cozzi, S. Melandri, V. Barone, and C. Puzzarini, "Theory meets experiment for noncovalent complexes: The puzzling case of pnicogen interactions," *Angew. Chem. Int. Ed.*, vol. 57, pp. 13 853–13 857, 2018.

[152] L. Spada, N. Tasinato, G. Bosi, F. Vazart, V. Barone, and C. Puzzarini, "On the competition between weak OH$\cdots$F and CH$\cdots$F hydrogen bonds, in cooperation with CH$\cdots$O contacts, in the difluoromethane–tert-butyl alcohol cluster," *J. Mol. Spectr.*, vol. 337, pp. 90–95, 2017.

[153] K. A. Peterson and T. H. Dunning Jr, "Accurate correlation consistent basis sets for molecular core–valence correlation effects: The second row atoms Al-Ar, and the first row atoms B-Ne revisited," *J. Chem. Phys.*, vol. 117, pp. 10 548–10 560, 2002.

[154] J. H. Thorpe, C. A. Lopez, T. L. Nguyen, J. H. Baraban, D. H. Bross, B. Ruscic, and J. F. Stanton, "High-accuracy extrapolated ab initio thermochemistry. iv. a modified recipe for computational efficiency," *J. Chem. Phys.*, vol. 150, p. 224102, 2019.

[155] J. Wang, L. Spada, J. Chen, S. Gao, S. Alessandrini, G. Feng, C. Puzzarini, Q. Gou, J.-U. Grabow, and V. Barone, "The unexplored world of cycloalkene–water complexes: Primary and assisting interactions unraveled by experimental and computational spectroscopy," *Angew. Chem. Int. Ed.*, vol. 131, pp. 14 073–14 079, 2019.

[156] J. Řezáč, K. E. Riley, and P. Hobza, "S66: a well balanced database of benchmark interaction energies relevant to biomolecular structures," *J. Chem. Theory Comput.*, vol. 7, pp. 2427–2438, 2011.

[157] G. Schmitz, C. Hättig, and D. P. Tew, "Explicitly correlated pno-mp2 and pno-ccsd and their application to the s66 set and large molecular systems," *Phys. Chem. Chem. Phys.*, vol. 16, pp. 22 167–22 178, 2014.

[158] P. R. Spackman, D. Jayatilaka, and A. Karton, "Basis set convergence of ccsd(t) equilibrium geometries using a large and diverse set of molecular structures," *J. Chem. Phys.*, vol. 145, p. 104101, 2016.

[159] A. Karton and P. R. Spackman, "Evaluation of density functional theory for a large and diverse set of organic and inorganic equilibrium structures," *J. Comp. Chem.*, vol. 42, pp. 1590–1601, 2021.

[160] C. B. Aakeroy, D. L. Bryce, G. R. Desiraju, A. Frontera, A. C. Legon, F. Nicotra, K. Rissanen, S. Scheiner, G. Terraneo, P. Metrangolo, and G. Resnati, "Definition of the chalcogen bond (IUPAC Recommendations 2019)," *Pure Appl. Chem.*, vol. 91, pp. 1889–1892, 2019.

[161] P. Politzer and J. S. Murray, "σ-holes and π-holes: Similarities and differences," *J. Comput. Chem.*, vol. 39, pp. 464–471, 2018.

[162] Y. Geboes, E. De Vos, and W. A. Herrebout, "S···S and S···P chalcogen bonding in solution: a cryospectroscopic study of the complexes of 2,2,4,4-tetrafluoro-1,3-dithietane with dimethyl sulfide and trimethylphosphine," *New J. Chem.*, vol. 42, pp. 10 563–10 571, 2018.

[163] M. K. Jahn, D. A. Dewald, D. Wachsmuth, J.-U. Grabow, and S. C. Mehrotra, "Rapid capture of large amplitude motions in 2,6-difluorophenol: High-resolution fast-passage FT-MW technique," *J. Mol. Spectr.*, vol. 280, pp. 54–60, 2012.

[164] J.-U. Grabow, W. Stahl, and H. Dreizler, "A multioctave coaxially oriented beam-resonator arrangement fourier-transform microwave spectrometer," *Rev. Sci. Instrum.*, vol. 67, pp. 4072–4084, 1996.

[165] V. Barone, P. Cimino, and E. Stendardo, "Development and validation of the B3LYP/N07D computational model for structural parameter and magnetic tensors of large free radicals," *J. Chem. Theory Comput.*, vol. 4, pp. 751–764, 2008.

[166] V. Barone and P. Cimino, "Accurate and feasible computations of structural and magnetic properties of large free radicals: The PBE0/N07D model," *Chem. Phys. Lett.*, vol. 454, pp. 139–143, 2008.

[167] J. Demaison, L. Margules, and H. D. Rudolph, "Accurate determination of an equilibrium structure in the presence of a small coordinate: The case of dimethylsulfide," *J. Mol. Struct.*, vol. 978, pp. 229–233, 2010.

[168] J. Demaison and J. Liv́in, "Accuracy of the equilibrium structure of sulphur dioxide," *Mol. Phys.*, p. e1950857, 2021.

[169] M. Cametti, B. Crousse, P. Metrangolo, R. Milani, and G. Resnati, "The fluorous effect in biomolecular applications," *Chem. Soc. Rev.*, vol. 41, pp. 31–42, 2012.

[170] S. Potenti, L. Spada, M. Fusè, G. Mancini, A. Gualandi, C. Leonardi, P. G. Cozzi, C. Puzzarini, and V. Barone, "4-Fluoro-Threonine: From diastereoselective synthesis to ph-dependent conformational equilibrium in aqueous solution," *ACS Omega*, vol. 6, pp. 13 170–13 181, 2021.

[171] S. Suzuki, P. G. Green, R. E. Bumgarner, S. Dasgupta, W. A. Goddard III, and G. A. Blake, "Benzene forms hydrogen bonds with water," *Science*, vol. 257, pp. 942–945, 1992.

[172] K. Brendel, H. Mäder, Y. Xu, and W. Jäger, "The rotational spectra of the fluorobenzene···water and p-difluorobenzene···water dimers: Structure and internal dynamics," *J. Mol. Spectr.*, vol. 268, pp. 47–52, 2011.

[173] L. Evangelisti, K. Brendel, H. Mäder, W. Caminati, and S. Melandri, "Rotational spectroscopy probes water flipping by full fluorination of benzene," *Angew. Chem. Int. Ed.*, vol. 129, pp. 13 887–13 891, 2017.

[174] S. Purser, P. R. Moore, S. Swallow, and V. Gouverneur, "Fluorine in medicinal chemistry," *Chem. Soc. Rev.*, vol. 37, pp. 320–330, 2008.

[175] L. Kang, S. E. Novick, Q. Gou, L. Spada, M. Vallejo-Lopez, and W. Caminati, "The shape of trifluoromethoxybenzene," *J. Mol. Spectr.*, vol. 297, pp. 32–34, 2014.

[176] M. Onda, A. Toda, S. Mori, and I. Yamaguchi, "Microwave spectrum of anisole," *J. Mol. Struct.*, vol. 144, pp. 47–51, 1986.

[177] O. Desyatnyk, L. Pszczółkowski, S. Thorwirth, T. Krygowski, and Z. Kisiel, "The rotational spectra, electric dipole moments and molecular structures of anisole and benzaldehyde," *Phys. Chem. Chem. Phys.*, vol. 7, pp. 1708–1715, 2005.

[178] Y. Jin, J. Wang, Q. Gou, Z. Xia, and G. Feng, "Fluorination effect on conformational preferences of trifluorothioanisole," *J. Mol. Struct.*, vol. 1156, pp. 230–234, 2018.

[179] B. Velino, S. Melandri, W. Caminati, and P. Favero, "Free-jet absorption millimeter-wave spectrum of thioanisole," *Gazz. Chim. Ital.*, vol. 125, pp. 373–376, 1995.

[180] M. Onda, Y. Kohama, K. Suga, and I. Yamaguchi, "Microwave spectrum and molecular planarity of acetophenone," *J. Mol. Struct.*, vol. 442, pp. 19–22, 1998.

[181] J. Lei, J. Zhang, G. Feng, J.-U. Grabow, and Q. Gou, "Conformational preference determined by inequivalent n-pairs: Rotational studies on acetophenone and its monohydrate," *Phys. Chem. Chem. Phys.*, vol. 21, pp. 22 888–22 894, 2019.

[182] J. Lei, J. Chen, G. Feng, Z. Xia, and Q. Gou, "Rotational spectrum of 2,2,2-trifluoroacetophenone," *J. Mol. Spectr.*, vol. 351, 2018.

[183] Y.-h. Miao, Y.-h. Hu, J. Yang, T. Liu, J. Sun, and X.-j. Wang, "Natural source, bioactivity and synthesis of benzofuran derivatives," *RSC Adv.*, vol. 9, pp. 27 510–27 540, 2019.

[184] H. Khanam *et al.*, "Bioactive benzofuran derivatives: A review," *Eur. J. Med. Chem.*, vol. 97, pp. 483–504, 2015.

[185] A. Poblotzki, J. Altnöder, and M. A. Suhm, "Subtle solvation behaviour of a biofuel additive: the methanol complex with 2,5-dimethylfuran," *Phys. Chem. Chem. Phys.*, vol. 18, pp. 27 265–27 271, 2016.

[186] H. Sasaki, S. Daicho, Y. Yamada, and Y. Nibu, "Comparable strength of OH···O versus $OH^{-}\pi$ hydrogen bonds in hydrogen-bonded 2, 3-benzofuran clusters with water and methanol," *J. Phys. Chem. A*, vol. 117, pp. 3183–3189, 2013.

[187] Y. Shi, J. Zhang, M. Shi, S. P. O'Connor, S. N. Bisaha, C. Li, D. Sitkoff, A. T. Pudzianowski, S. Chong, H. E. Klei *et al.*, "Cyanoguanidine-based lactam derivatives as a novel class of orally bioavailable factor xa inhibitors," *Bioorg. Med. Chem. Lett.*, vol. 19, pp. 4034–4041, 2009.

[188] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Res.*, vol. 28, pp. 235–242, 2000.

[189] E. Penocchio, M. Piccardo, and V. Barone, "Semiexperimental equilibrium structures for building blocks of organic and biological molecules: The B2PLYP route," *J. Chem. Theory Comput.*, vol. 11, pp. 4689–4707, 2015.

[190] A. Maris, B. M. Giuliano, S. Melandri, P. Ottaviani, W. Caminati, L. B. Favero, and B. Velino, "Structure, dipole moment and large amplitude motions of 1-benzofuran," *Phys. Chem. Chem. Phys.*, vol. 7, pp. 3317–3322, 2005.

[191] E. Maciá, "The role of phosphorus in chemical evolution," *Chem. Soc. Rev.*, vol. 34, pp. 691–701, 2005.

[192] V. Lebouteiller, Kuassivi, and R. Ferlet, "Phosphorus in the diffuse interstellar medium," *Astron. Astrophys.*, vol. 443, pp. 509–517, 2005.

[193] J. Chantzos, V. M. Rivilla, A. Vasyunin, E. Redaelli, L. Bizzocchi, F. Fontani, and P. Caselli, "The first steps of interstellar phosphorus chemistry," *Astron. Astrophys.*, vol. 633, p. A54, 2020.

[194] E. D. Tenenbaum, N. J. Woolf, and L. M. Ziurys, "Identification of phosphorus monoxide ($X^2\Pi_r$) in VY canis majoris: Detection of the first P–O bond in space," *Astrophys. J.*, vol. 666, pp. L29–L32, 2007.

[195] M. Guélin, J. Cernicharo, G. Paubert, and B. Turner, "Free CP in IRC+10216," *Astron. Astrophys.*, vol. 230, pp. L9–L11, 1990.

[196] M. Agúndez, J. Cernicharo, and M. Guélin, "Discovery of phosphaethyne (HCP) in space: Phosphorus chemistry in circumstellar envelopes," *Astrophys. J.*, vol. 662, pp. L91–L94, 2007.

[197] E. D. Tenenbaum, J. L. Dodd, S. N. Milam, N. J. Woolf, and L. M. Ziurys, "Comparative spectra of oxygen-rich vs carbon-rich circumstellar shells: VY canis majoris and IRC+10216 at 215-285 Hz," *Astrophys. J.*, vol. 720, pp. L102–L107, 2010.

[198] F. Vazart, C. Latouche, D. Skouteris, N. Balucani, and V. Barone, "Cyanomethanimine isomers in cold insterstellar clouds: insights from electronic structure and kinetic calculations," *Astrophys. J.*, vol. 810, p. 111, 2015.

[199] K. Fukui, "The path of chemical reactions - the IRC approach," *Acc. Chem. Res.*, vol. 14, pp. 363–368, 1981.

[200] H. Hratchian and H. Schlegel, *Theory and Applications of Computational Chemistry: The First 40 Years*. Elsevier, Amsterdam, 2005.

[201] L. Bizzocchi, D. Prudenzano, V. M. Rivilla, A. Pietropolli-Charmet, B. M. Giuliano, P. Caselli, J. Martín-Pintado, I. Jiménez-Serra, S. Martín, M. A. Requena-Torres *et al.*, "Propargylimine in the laboratory and in space: millimetre-wave spectroscopy and its first detection in the ISM," *Astron. & Astrophys.*, vol. 640, p. A98, 2020.

[202] M. Melosso, A. Melli, C. Puzzarini, C. Codella, L. Spada, L. Dore, C. Degli Esposti, B. Lefloch, R. Bachiller, C. Ceccarelli *et al.*, "Laboratory measurements and astronomical search for cyanomethanimine," *Astron. & Astrophys.*, vol. 609, p. A121, 2018.

[203] F. Vazart, D. Calderini, C. Puzzarini, D. Skouteris, and V. Barone, "State-of-the-art thermochemical and kinetic computations for astrochemical complex organic molecules: Formamide formation in cold interstellar clouds as a case study," *J. Chem. Theory Comput.*, vol. 12, pp. 5385–5397, 2016.

[204] J. Dickens, W. M. Irvine, M. Ohishi, G. Arrhenius, S. Pitsch, A. Bauder, F. Müller, and A. Eschenmoser, "A search for interstellar oxiranecarbonitrile (C_3H_3NO)," *Orig. Life Evol. Biosph.*, vol. 26, pp. 97–110, 1996.

[205] E. Wirström, P. Bergman, Å. Hjalmarson, and A. Nummelin, "A search for prebiotic molecules in hot cores," *Astron. Astrophys.*, vol. 473, pp. 177–180, 2007.

[206] E. Alonso, L. Kolesniková, A. Belloche, S. Mata, R. Garrod, A. Jabri, I. León, J.-C. Guillemin, H. Müller, K. Menten, and J. L. Alonso, "Rotational spectroscopic study and astronomical search for propiolamide in SgrB2(N)," *Astron. Astrophys.*, vol. 647, p. A55, 2021.

[207] B. A. McGuire, "2018 census of interstellar, circumstellar, extragalactic, protoplanetary disk, and exoplanetary molecules," *Astrophys. J., Suppl. Ser.*, vol. 239, p. 17, 2018.

[208] A. McKellar, "Evidence for the molecular origin of some hitherto unidentified interstellar lines," *Pub. Astron. Soc. Pac.*, vol. 52, pp. 187–192, 1940.

[209] K. B. Jefferts, A. A. Penzias, and R. W. Wilson, "Observation of the CN radical in the Orion nebula and W51," *Astrophys. J. Lett.*, vol. 161, p. L87, 1970.

[210] M. C. McCarthy, K. L. K. Lee, R. A. Loomis, A. M. Burkhardt, C. N. Shingledecker, S. B. Charnley, M. A. Cordiner, E. Herbst, S. Kalenskii, E. R. Willis *et al.*, "Interstellar detection of the highly polar five-membered ring cyanocyclopentadiene," *Nat. Astron.*, vol. 5, pp. 176–180, 2021.

[211] K. L. K. Lee, P. B. Changala, R. A. Loomis, A. M. Burkhardt, C. Xue, M. A. Cordiner, S. B. Charnley, M. C. McCarthy, and B. A. McGuire, "Interstellar detection of 2-cyanocyclopentadiene, C_5H_5CN, a second five-membered ring toward TMC-1," *Astrophys. J. Lett.*, vol. 910, p. L2, 2021.

[212] D. E. Woon, "Modeling chemical growth processes in titan's atmosphere: 1.Theoretical rates for reactions between benzene and the ethynyl (C_2H) and cyano (CN) radicals at low temperature and pressure," *Chem. Phys.*, vol. 331, pp. 67–76, 2006.

[213] I. R. Cooke, D. Gupta, J. P. Messinger, and I. R. Sims, "Benzonitrile as a proxy for benzene in the cold ISM: Low-temperature rate coefficients for CN + C_6H_6," *Astrophys. J.*, vol. 891, p. L41, 2020.

[214] J. E. Dickens, W. M. Irvine, M. Ohishi, M. Ikeda, S. Ishikawa, A. Nummelin, and Å. Hjalmarson, "Detection of interstellar ethylene oxide c-C_2H_4O," *Astrophys. J.*, vol. 489, pp. 753–757, 1997.

[215] M. A. Requena-Torres, J. Martin-Pintado, S. Martin, and M. R. Morris, "The galactic center: The largest oxygen-bearing organic molecule repository," *Astrophys. J.*, vol. 672, pp. 352–360, 2008.

[216] A. Nummelin, J. E. Dickens, P. Bergman, A. Hjalmarson, W. M. Irvine, M. Ikeda, and M. Ohishi, "Abundances of ethylene oxide and acetaldehyde in hot molecular cloud cores," *Astron. Astrophys.*, vol. 337, pp. 275–286, 1998.

[217] J. M. Lykke, Coutens, A., Jørgensen, J. K., van der Wiel, M. H. D., Garrod, R. T., Müller, H. S. P., Bjerkeli, P., Bourke, T. L., Calcutt, H., Drozdovskaya, M. N., Favre, C., Fayolle, E. C., Jacobsen, S. K., Öberg, K. I., Persson, M. V., van Dishoeck, E. F., and Wampfler, S. F., "The ALMA-PILS survey: First detections of ethylene oxide, acetone and propanal toward the low-mass protostar IRAS 16293-2422," *Astron. Astrophys.*, vol. 597, p. A53, 2017.

[218] Y. Georgievskii, J. A. Miller, M. P. Burke, and S. J. Klippenstein, "Reformulation and solution of the master equation for multiple-well chemical reactions," *J. Phys. Chem. A*, vol. 117, pp. 12 146–12 154, 2013.

[219] Y. Georgievskii and S. Klippenstein, "MESS. 2016.3.23," *Argonne National Laboratory*, 2016.

[220] P. Pechukas and J. C. Light, "On detailed balancing and statistical theories of chemical kinetics," *J. Chem. Phys.*, vol. 42, pp. 3281–3291, 1965.

[221] W. J. Chesnavich, "Multiple transition states in unimolecular reactions," *J. Chem. Phys.*, vol. 84, pp. 2615–2619, 1986.

[222] A. Fernández-Ramos, J. A. Miller, S. J. Klippenstein, and D. G. Truhlar, "Modeling the kinetics of bimolecular reactions," *Chem. Rev.*, vol. 106, pp. 4518–4584, 2006.

[223] H. Møllendal, L. Margulés, R. A. Motiyenko, N. W. Larsen, and J.-C. Guillemin, "Rotational spectrum and conformational composition of cyanoacetaldehyde, a compound of potential prebiotic and astrochemical interest," *J. Phys. Chem. A*, vol. 116, pp. 4047–4056, 2012.

[224] K. L. K. Lee, R. A. Loomis, A. M. Burkhardt, I. R. Cooke, C. Xue, M. A. Siebert, C. N. Shingledecker, A. Remijan, S. B. Charnley, M. C. McCarthy *et al.*, "Discovery of interstellar trans-cyanovinylacetylene (HC≡CCH=CHC≡N) and vinylcyanoacetylene (H_2C=CHC_3N) in GOTHAM observations of TMC-1," *Astrophys. J. Lett.*, vol. 908, p. L11, 2021.

[225] I. R. Sims, J.-L. Queffelec, D. Travers, B. R. Rowe, L. B. Herbert, J. Karthäuser, and I. W. Smith, "Rate constants for the reactions of CN with hydrocarbons at low and ultra-low temperatures," *Chem. Phys. Lett.*, vol. 211, pp. 461–468, 1993.

[226] J. Moon and J. Kim, "Reaction mechanism of cyanoethynyl radical (C_3N) with ethylene (C_2H_4) to form C_5H_3N and H: a theoretical investigation," *Theo. Chem. Acc.*, vol. 136, pp. 1–10, 2017.

[227] M. Guélin, P. Friberg, and A. Mezaoui, "Astronomical study of the c3n and c4h radicals-hyperfine interactions and rho-type doubling," *Astron. Astrophys.*, vol. 109, pp. 23–31, 1982.

[228] N. Marcelino, J. Cernicharo, M. Agúndez, E. Roueff, M. Gerin, J. Martín-Pintado, R. Mauersberger, and C. Thum, "Discovery of interstellar propylene (CH_2CHCH_3): Missing links in interstellar gas-phase chemistry," *Astrophys. J.*, vol. 665, p. L127, 2007.

[229] J. Cernicharo, M. Agúndez, C. Cabezas, B. Tercero, N. Marcelino, J. R. Pardo, and P. de Vicente, "Pure hydrocarbon cycles in TMC-1: Discovery of ethynyl cyclopropenylidene, cyclopentadiene and indene." *Astron. Astrophys.*, vol. 649, 2021.

[230] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, "Turbomole," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 4, pp. 91–100, 2014.